Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene

Patrick R Sosnay1–3, Karen R Siklosi3, Fredrick Van Goor4, Kyle Kaniecki3,5, Haihui Yu4, Neeraj Sharma3, Anabela S Ramalho6,7, Margarida D Amaral6,7, Ruslan Dorfman8,9, Julian Zielenki8, David L Masica10, Rachel Karchin10, Linda Millen11, Philip J Thomas11, George P Patrinos12, Mary Corey13,14, Michelle H Lewis15, Johanna M Rommens8,16, Carlo Castellani17, Christopher M Penland18 & Garry R Cutting3,19

Allelic heterogeneity in disease-causing genes presents a substantial challenge to the translation of genomic variation into clinical practice. Few of the almost 2,000 variants in the cystic fibrosis transmembrane conductance regulator gene CFTR have empirical evidence that they cause cystic fibrosis. To address this gap, we collected both genotype and phenotype data for 39,696 individuals with cystic fibrosis in registries and clinics in North America and Europe. In these individuals, 159 CFTR variants had an allele frequency of ≥0.01%. These variants were evaluated for both clinical severity and functional consequence, with 127 (80%) meeting both clinical and functional criteria consistent with disease. Assessment of disease penetrance in 2,188 fathers of individuals with cystic fibrosis enabled assignment of 12 of the remaining 32 variants as neutral, whereas the other 20 variants remained of indeterminate effect. This study illustrates that sourcing data directly from well-phenotyped subjects can address the gap in our ability to interpret clinically relevant genomic variation.

The usefulness of genetic testing for both mendelian and polygenic disorders is limited by the substantial number of DNA variants of uncertain significance1–4. Next-generation sequencing in clinical laboratories will dramatically increase the number of variants of potential medical relevance5. Thus, an ever-widening gap is likely to occur between the ability to identify DNA variation and the ability to interpret its consequence6. One approach to address this gap is to aggregate variants identified by clinical and research laboratories into central repositories7,8. Observation of the same variant in individuals with the same phenotype supports the notion that the variant may be deleterious. However, physicians request clinical testing for a number of reasons, including confirmation or exclusion of a specific diagnosis. Aggregation of variants from testing facilities without robust phenotype and functional annotation can diminish the potential clinical value of repositories9,10. A prime example of the challenge of allelic heterogeneity is the gene responsible for cystic fibrosis, the cystic fibrosis transmembrane conductance regulator gene CFTR (NM_000492.3). Almost 2,000 variants have been reported in the CFTR coding and flanking sequences, but the disease liability of only a few dozen variants has been ascertained11. Consequently, sequence analysis of the CFTR gene for diagnostic purposes frequently uncovers variants of uncertain significance. The clinical implications of incomplete annotation of CFTR sequence variation extend well beyond the ~70,000 individuals with cystic fibrosis worldwide, particularly because CFTR genetic testing is frequently part of newborn screening12–15. Furthermore, population-based screening for carriers of cystic fibrosis has become progressively more common, with an estimated 1.2 million individuals tested each year in the United States16,17. In cases where one member of a couple is discovered to carry a known cystic fibrosis—causing variant, extensive CFTR analysis is often performed on the other member that identifies variants of uncertain significance18. Finally, the large number of non-experimentally verified disease-associated variants hampers the understanding of how structural changes in the CFTR protein lead to dysfunction and result in the cystic fibrosis phenotype. The gap in understanding of disease-causing versus neutral alleles presents a major challenge in the genomic sequencing era.

1Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA. 2Perdana University Graduate School of Medicine, Serdang, Malaysia. 3McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA. 4Vertex Pharmaceuticals, Inc., San Diego, California, USA. 5Department of Genetics and Development, Columbia University College of Physicians and Surgeons, New York, New York, USA. 6Centre for Biodiversity, Functional and Integrative Genomics (BioFIG), Faculty of Sciences, University of Lisboa, Lisbon, Portugal. 7Department of Genetics, National Institute of Health, Lisbon, Portugal. 8Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. 9Geneyouin, Inc., Maple, Ontario, Canada. 10Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, Maryland, USA. 11Department of Physiology, University of Texas Southwestern Medical Center, Dallas, Texas, USA. 12Department of Pharmacy, School of Health Sciences, University of Patras, University Campus, Patras, Greece. 13Program in Child Evaluative Health Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada. 14Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. 15Genetics and Public Policy Center, Berman Institute for Bioethics, Johns Hopkins University, Baltimore, Maryland, USA. 16Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. 17Cystic Fibrosis Center, Azienda Ospedaliera Universitaria Integrata, Verona, Italy. 18Cystic Fibrosis Foundation, Bethesda, Maryland, USA. 19Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Received 11 January; accepted 30 July; published online 25 August 2013; doi:10.1038/ng.2745
A central repository for CFTR variants termed the Cystic Fibrosis Mutation Database (CFMD; see URLs) began in 1990, shortly after the CFTR gene was identified. CFMD content was generated from discoveries in research laboratories, with additional contributions from genetic testing facilities. Although it provides an extensive collection of variation in CFTR, CFMD has little phenotypic annotation, and functional consequences of variation are primarily derived from predictions based on the nature of the nucleotide changes. Assessing the disease liability of CFTR variants with predictive algorithms has proven to be of limited usefulness.19,20 A key weakness in the development of more accurate algorithms is the paucity of variants with well-defined functional consequences.21

As the CFMD constituted an excellent existing repository of nucleotide variation in CFTR, we took a new approach to comprehensively address the phenotypic and functional implications of CFTR variants. Our Clinical and Functional Translation of CFTR (CFTR2) project assembled clinical data and accompanying CFTR variants from individuals with cystic fibrosis enrolled in national registries and large clinical centers from 24 countries. By focusing on variants present in individuals with a diagnosis of cystic fibrosis ascertained by expert clinicians, the project used a ‘phenotype-driven’ approach to data collection rather than the laboratory-based ‘genotype-driven’ approach. Second, microattribution recognition was used to identify the source and credit the contributors of the clinical and genetic data that constitute the CFTR2 database.22,23 To prioritize evaluation, the CFTR2 project started with the subset of CFTR variants exceeding an allele frequency of 0.01% in the collected individuals with cystic fibrosis. We used clinical features of subjects and functional assessment of each variant to define disease-causing variants. We evaluated variants not meeting clinical or functional thresholds for disease penetrance using a population-based approach. The phenotype-driven approach presented here could be used to inform the assignment of disease liability in a wide range of genetic disorders.

RESULTS

159 CFTR variants represent 96% of cystic fibrosis alleles

Data from the 39,696 individuals with cystic fibrosis in CFTR2 (Fig. 1) were collected from national cystic fibrosis patient registries or cystic fibrosis specialty clinics (Supplementary Table 1) and represent 57% of the estimated 70,000 individuals with cystic fibrosis.24 Informed consent was obtained by the participating registry or clinic according to local requirements. The vast majority (95% of the 31,727 individuals with ancestry data) are listed as Caucasian. In these individuals, 1,044 distinct CFTR variants were seen. The most common variant, p.Phe508del, accounted for 70% of the identified alleles in these individuals. Twenty-two additional variants previously defined as cystic fibrosis causing and reported to occur at a frequency of 0.1% or higher in individuals with cystic fibrosis by the American College of Medical Genetics (ACMG) represented 17.5% of the alleles.11 Another 136 variants occurred at a frequency exceeding 0.01% and were each reported on at least 9 alleles in the CFTR2 database (Supplementary Table 2). Together, these 159 variants accounted for 96.4% of the identified cystic fibrosis alleles in CFTR2. Our efforts focused on the evaluation of the disease liability of these 159 variants to maximize clinical sensitivity for cystic fibrosis genetic testing.

Phenotypic analysis

All individuals in the CFTR2 database were clinically diagnosed with cystic fibrosis; however, cystic fibrosis is a highly variable disorder.25 To evaluate individuals across the spectrum of cystic fibrosis severity, we used sweat chloride concentration, a biochemical phenotype that is integral to clinical diagnosis with cystic fibrosis. Sweat chloride analysis is widely performed in a standardized fashion, with well-defined differences in values from those observed in the population without cystic fibrosis.26–29 A variant was deemed disease causing by clinical criteria if the mean sweat chloride concentration derived from at least three individuals carrying the variant was ≥60 mM.28,30 The use of an average measure enabled accommodation of individual-level variability in sweat chloride concentration due to non-CFTR factors (Supplementary Fig. 1). When data were only available from two individuals, both sweat chloride concentrations had to exceed 90 mM. To attribute sweat chloride concentration to the variant under study, we analyzed individuals who carried a variant in their other CFTR gene that was known to cause complete or near-complete loss of CFTR function (Online Methods). Of the 159 variants under study, 140 met clinical criteria for causing cystic fibrosis (Fig. 2), of which 138 had sweat chloride concentrations derived from 3 or more individuals, whereas 2 variants each had measures exceeding 90 mM in 2 individuals (Supplementary Table 2). Thirteen of the 14 variants not meeting clinical criteria were associated with mean sweat chloride concentrations in the clinical ‘intermediate’ range from 40–58 mM; the remaining variant had an average measure of 39 mM. Individual variant data and details of other cardinal phenotypes of cystic fibrosis are shown in Supplementary Table 2.

Functional analysis

Two common variants (>5% frequency in the general population) in a polythymidine region in intron 9 (c.1210–12T(5) and c.1210–12T(7); legacy names 5T and 7T, respectively) that have been extensively studied were not reanalyzed here (Supplementary Note). Eighty of the remaining 157 variants are predicted to introduce a premature

---

Figure 1 Data collected for the CFTR2 project. The 159 variants seen in 9 or more alleles with an allele frequency of ≥0.01% in CFTR2 were prioritized for further analysis. aSweat chloride data were not reported for 10,170 affected individuals; 236 individuals had sweat chloride values outside the physiologic range (>150 mM or <5 mM) and were excluded. bLung function data recorded as forced expiratory volume in 1 s (FEV1) were not reported for 10,197 individuals; 5,633 individuals were under the age of 6 years and were excluded, if measurements were present, and 46 individuals had lung function measurements outside the physiological range (<3% or >150% predicted) and were excluded. cPancreatic status was characterized as sufficient (PS) or insufficient (PI); data were not reported for 5,083 individuals. dIncomplete clinical information available from the submitting registry at the time of analysis.
termination codon (PTC) into CFTR mRNA (nonsense variants (n = 35), variants in the canonical nucleotides of the splice donor or splice acceptor sites (GT-AG; n = 15) or insertion-deletion variants causing frameshifts (n = 30); Fig. 3). A common consequence of a PTC variant is nonsense-mediated decay (NMD) of mRNA, resulting in a severe reduction in mRNA levels and no protein produced31,32. In rare cases, variants affecting splicing can create stable in-frame transcripts through the skipping of in-frame exons; however, the translated protein is almost invariably non-functional (for example, c.1393–1G>A (legacy name 1525–1G>A))33. Thus, these 80 CFTR variants were predicted to be clinically deleterious4 and cystic fibrosis causing (Supplementary Fig. 2). Ten variants occurred within or near splice sites but did not alter canonical splice donor or acceptor sites (Fig. 3). Five of these variants have previously been evaluated and shown to express aberrant alternatively spliced transcripts in relevant tissues, leading to severe reduction in the levels of full-length CFTR mRNA (0–8% of the level for wild-type CFTR)34–39 (Supplementary Table 3a). The remaining five putative splice variants were studied using minigene analysis (Online Methods and Supplementary Table 3b). Aberrant splicing (resulting in <10% wild-type CFTR transcript) and reduced levels of mature CFTR protein (<10% of the levels of wild-type CFTR) were observed for four of the five variants (Supplementary Table 3c). Less than 10% of wild-type CFTR function has generally been accepted as a conservative threshold for the presence of cystic fibrosis features in the exocrine pancreas, sweat gland and lungs38,40,41. Together, nine of the ten variants that affected splicing had evidence of deleterious consequences consistent with disease (Fig. 3).

Sixty-seven variants are predicted to result in either an amino acid substitution (missense; n = 65) or the omission of a single amino acid (in-frame deletion; n = 2). As these variants permit the synthesis of stable mRNA and full-length protein, we performed experimental studies on each variant in isolation to determine its effect on CFTR biogenesis and function. We expressed CFTR bearing missense or in-frame changes in HeLa and Fischer rat thyroid (FRT) cells to assess glycosylation status with protein blotting, a well-established method to monitor CFTR maturation (Supplementary Note)32,42. We tested 63 (61 missense and 2 in-frame deletion) of the 67 variants in both cell lines for their effect on CFTR processing. Results from the two cell lines largely agreed (r² = 0.94; P < 0.001; Supplementary Fig. 3). The variants fell into three groups: those with minimal disruption in processing (>80% CFTR protein in the mature form in both cell lines; n = 32), those with intermediate disruption in processing (10–80% CFTR protein in the mature form in at least one cell line; n = 21) and those with a dramatic negative effect on processing (<10% CFTR protein in the mature form in both cell lines; n = 10). In the group with intermediate effects on processing, 11 variants caused a severe defect in processing in 1 cell line but not in the other; the remaining 10 variants caused a defect of intermediate severity in both cell lines.

To assess the effect of the missense variants on CFTR function, we measured chloride currents in FRT cells expressing CFTR bearing each of 63 variants individually (61 missense and 2 in-frame deletion). Chloride conductance was not determined for four missense variants. Functional analysis of primary airway cells obtained
from individuals with cystic fibrosis bearing nine different CFTR genotypes composed of established disease-causing variants was consistent with a threshold of 10% CFTR function being associated with cystic fibrosis (Supplementary Fig. 4). Forty-three variants (41 missense and 2 in-frame deletion) conducted chloride at a level less than 10% of that observed with wild-type CFTR and were deemed disease causing (Fig. 3). CFTR bearing each of the remaining 20 missense changes generated chloride conductance that ranged from 10.5% to 147% of that of wild-type CFTR. As such, the effects of these 20 variants on CFTR function were classified as inconsistent with cystic fibrosis, although these variants could contribute to other disease phenotypes. Comparison of CFTR processing and chloride current showed that a severe processing defect in HeLa cells was consistently associated with CFTR chloride channel function of less than 10% of wild-type function (Supplementary Fig. 5 and Supplementary Note). Of the four variants for which we did not measure chloride conduction, one (p.His199Tyr) exhibited a severe processing defect in HeLa or FRT cells (C band/(B band + C band) < 0.1) was consistently associated with CFTR chloride channel function of less than 10% of wild-type function (Supplementary Fig. 5 and Supplementary Note). Of the four variants for which we did not measure chloride conduction, one (p.His199Tyr) exhibited a severe processing defect in HeLa cells (<0.01) and was categorized as functionally deficient (Fig. 3). The remaining three variants (p.[Gln359Lys; Thr360Lys], p.Leu558Ser and p.Arg1070Gln) exhibited processing greater than 10% of that of wild-type CFTR and were not functionally classified.

Penetrance analysis
Of the 159 variants studied, 127 met clinical and functional criteria and were classified as cystic fibrosis–causing variants (Fig. 4 and Supplementary Table 2). To aid in the classification of variants not meeting clinical or functional criteria, we performed a penetrance study using 2,188 fathers of individuals with cystic fibrosis recruited from North America and Europe (Supplementary Table 4). The presence of a normally functioning CFTR gene is required in fathers of individuals with cystic fibrosis, as reduced CFTR function is associated with male infertility due to congenital bilateral absence of the vas deferens (CBAVD)44. Male infertility due to CBAVD affects 97–98% of males with cystic fibrosis45,46. Fathers of naturally conceived offspring with cystic fibrosis will transmit one pathogenic allele to their affected children. As those fathers are fertile, the non-transmitted allele should not contain a deleterious variant. Thus, any CFTR variants occurring on the non-transmitted allele in a fertile father was deemed non-penetrant for cystic fibrosis and CBAVD. To exclude errors that could have occurred during sample processing or if assisted reproductive technologies were used without our knowledge, we required that a variant be observed on the non-transmitted CFTR allele in at least two fathers.

Genotyping for the 159 CFTR variants yielded 2,062 samples suitable for penetrance analysis, of which 185 had 2 or more variants (Supplementary Figs. 6 and 7). After additional filtering, we found that 100 fathers carried at least 1 of the 159 variants in trans with a previously accepted cystic fibrosis–causing variant (Supplementary Fig. 6). Using data from these 100 fathers, we deemed 10 variants non-penetrant, as each occurred in the non-transmitted ‘healthy’ CFTR gene of at least 2 fathers (Table 1). To assess the validity of labeling these variants as non-penetrant, we compared the frequency of each variant in the fathers and in the CFTR2 database with its frequency in data available from the 1000 Genomes Project47. Our first premise was that non-penetrant and phenotypically irrelevant variants should occur in healthy cystic fibrosis–carrier fathers on the non-transmitted allele at the same frequency as observed in the general population. This was the case for nine non-penetrant variants for which 1000 Genomes Project data were available (Table 1). The second premise was that non-penetrant variants should occur at a much lower frequency in individuals with cystic fibrosis than in the general population. Indeed, the frequency of each non-penetrant variant in individuals with cystic fibrosis enrolled in CFTR2 was at least tenfold lower than the frequency in the general population where 1000 Genomes Project data were available. In addition to these ten variants, c.1210–12(7) (legacy name 7T) had already been reported to be non-penetrant48 and was identified as a second variant in numerous fathers, and a twelfth variant, p.Ile1027Thr, was deemed
non-penetrant, as it was observed exclusively in cis with the p.Phe508del change. The presence of non-penetrant variants in the CFTR2 database is likely due to incomplete genotyping and/or lack of analysis of allele assortment. Analysis of assortment is essential, as multiple examples of complex alleles were disclosed in the penetrance analysis of allele assortment. Analysis of assortment is essential, as multiple examples of complex alleles were disclosed in the penetrance study (Supplementary Note49).

We found no evidence of non-penetrance in the screen of fathers for 147 variants (all 127 that met clinical and functional criteria as well as 8 variants that met only clinical criteria, 6 variants that met only functional criteria and 6 variants that met neither criteria; Fig. 4). Included among the variants meeting neither clinical nor functional criteria are those that have previously been associated with variable penetrance (such as p.Arg75Gln), variants that have been reported as part of complex alleles in which the disease liability of each variant individually could not be determined (such as the pair p.Arg74Trp and p.Asp1270Asn) and variants with incomplete clinical or functional analysis.

**DISCUSSION**

Genetic testing of CFTR is widely employed for diagnosis in symptomatic individuals43, for carrier status in the general population47, increasingly as part of newborn screening50,51 and most recently for selection for treatment with variant-specific molecular therapy52. The primary goal of the CFTR2 project was to increase the fraction of variants in the CFTR gene that have been assessed for their propensity to cause disease. At the initiation of the project, 23 variants were defined as disease causing41. Combining phenotypic evidence with functional analysis enabled unambiguous assignment of pathogenicity to an additional 104 variants. Testing for all 127 variants is estimated to account for 95.4% of cystic fibrosis–confering alleles in our sample, leaving only 0.21% of affected individuals in our sample without at least one pathologic CFTR variant identified. Couples undergoing carrier screening will also benefit, as the sensitivity for detecting couples at a 1 in 4 risk of having a child with cystic fibrosis should increase from 72% to ~91% when screening for the 127 pathogenic variants. It should be noted that these estimated detection rates are subject to geographic and ancestry-based variability in variant distribution and frequency. This project illustrates the feasibility of translating allelic diversity into clinical application but also highlights the challenges in interpreting the disease implications of rare DNA changes.

The CFTR2 project gathered both genotype and phenotype data on individuals who were enrolled in registries and clinics. This approach enriched the pool of affected individuals, but we also used additional objective measures to differentiate variants causing life-shortening cystic fibrosis from those causing less severe disease44. Notably, 19 of 159 variants studied (12%) did not meet our clinical threshold, despite being reported in individuals diagnosed with cystic fibrosis by a medical professional familiar with the disease. This finding shows the degree of phenotypic heterogeneity existing even within well-annotated clinical data collections. Of the 140 variants identified as disease causing using clinical criteria, 13 (9.3%) did not meet functional criteria. Our study emphasizes the importance of using both phenotypic and functional analysis to clinically annotate variants found in affected individuals and demonstrates that the presence of a rare variant, even if reported in multiple unrelated affected individuals, does not ensure that it is deleterious or pathogenic.

Phenotypic and functional criteria for disease can be based on metrics that already exist for many genetic diseases. For this study, we chose sweat chloride concentration to define the phenotype because it is dependent on CFTR function, correlates with disease severity, is measured frequently in a standardized fashion and has well-validated cutoffs between normal and disease levels29,30. Similarly, assessment of the functional effects of variants can follow established guidelines4. For example, the assumption that variants predicted to introduce a PTC are deleterious is commonly accepted practice4. As noted here, the clinical features of individuals carrying predicted PTC variants are consistent with disease (Supplementary Fig. 2). Evaluation of the effect of missense variants poses the greatest hurdle; however, relatively straightforward assays such as protein blotting can disclose processing defects, a common consequence of amino acid substitutions. Expression of mutated protein in multiple cell lines, as employed here, minimizes cell type–specific effects. Perhaps the most challenging issue is the establishment of thresholds for both phenotypic and functional measures. In this study, the adopted thresholds were vetted by experts in the clinical and functional areas of cystic fibrosis research. The 10% threshold for protein expression and chloride conductance (both in comparison to wild-type CFTR) is not an absolute demarcation between disease and health but is a conservative threshold consistent with previous research correlating CFTR function with disease38,40,41. Provided that it is acknowledged that consensus opinions represent the current understanding of pathogenesis, thresholds can be modified if warranted by future studies, as some variants may influence CFTR in a manner not captured by our methods.

The 127 variants that met both clinical and functional criteria were designated cystic fibrosis causing; however, 32 remaining variants

---

**Table 1 Variants associated with incomplete penetrance**

| Variant       | Number of alleles in CFTR2 (out of 70,777 known alleles) | Frequency in CFTR2 | Number that occur in trans with a CF-causing variant in fathers | Number reported in 2,062 fathers | Frequency in fathers (out of 4,124 alleles) | Allele frequency in 1000 Genomes Project |
|---------------|----------------------------------------------------------|-------------------|-----------------------------------------------------------------|----------------------------------|------------------------------------------|----------------------------------------|
| p.Arg31Cys    | 13                                                       | 0.00018           | 4                                                               | 4                                | 0.00097                                  | 0.001--0.004                            |
| p.Ile148Thr   | 99                                                       | 0.00140           | 4                                                               | 9                                | 0.00218                                  | Not available                           |
| p.Met470Val   | 41                                                       | 0.00058           | Not analyzed                                                   | 1,412                            | 0.34239                                  | 0.087--0.647                            |
| p.Val754Met   | 9                                                        | 0.00013           | 4                                                               | 7                                | 0.00170                                  | 0--0.003                               |
| p.Arg75Gln    | 28                                                       | 0.00040           | 48                                                              | 74                               | 0.01794                                  | 0.009--0.033                            |
| p.Gly576Ala   | 42                                                       | 0.00059           | 12                                                              | 20                               | 0.00485                                  | 0.004--0.009                            |
| p.Arg668Cys   | 49                                                       | 0.00069           | 16                                                              | 29                               | 0.00703                                  | 0.004--0.009                            |
| p.Leu997Phe   | 28                                                       | 0.00040           | 5                                                               | 9                                | 0.00218                                  | 0.001--0.003                            |
| p.Arg1162Leu  | 9                                                        | 0.00013           | 2                                                               | 6                                | 0.00145                                  | 0.001                                 |
| p.Ser1235Arg  | 54                                                       | 0.00076           | 15                                                              | 21                               | 0.00509                                  | 0.005--0.016                            |
| p.Met470Thr   | 41                                                       | 0.00058           | Not analyzed                                                   | 1,412                            | 0.34239                                  | 0.087--0.647                            |
| p.Gly754Val   | 28                                                       | 0.00040           | 48                                                              | 74                               | 0.01794                                  | 0.009--0.033                            |
| p.Ser1235Thr  | 54                                                       | 0.00076           | 15                                                              | 21                               | 0.00509                                  | 0.005--0.016                            |

*Does not cause cystic fibrosis unless in cis with the known deleterious variant p.I1e1023.Val1024del46. In both the 1000 Genomes Project and in this study, this variant is always seen in cis with p.Gly576Ala. In the 1000 Genomes Project, this variant is always seen in cis with p.Gly1068Cys. In this study, it is seen both in cis and on its own.
(20%) required further analysis to determine whether they were neutral with respect to disease or associated with a milder phenotype or partial penetrance. Demonstration that a variant occurs in a sample of normal controls at the same frequency as observed in affected individuals has been a long-accepted method to determine neutrality.\textsuperscript{53,54} Under recessive inheritance conditions, this test can only be performed in ‘control’ individuals known to carry a deleterious allele in trans. Fathers of individuals with cystic fibrosis provide an ideal group to assess neutrality, as they carry a functional CFTR gene by virtue of their fertility and a disease-causing variant transmitted to their affected offspring. Demonstration that a variant under study occurs in the healthy (non-transmitted) CFTR genes of fertile fathers provides compelling evidence of neutrality or non-penetrance for cystic fibrosis. The power of this approach depends on the frequency of the alleles in the population and the number of ‘controls’ tested. In tests of the non-transmitted (non–cystic fibrosis) chromosome of fertile fathers, confidence that a given variant was not found because it is fully penetrant declines as the allele frequency declines. Therefore, we are more confident that more frequent variants such as p.Gly551Asp are fully penetrant than we are for variants such as p.[Gln359Lys; Thr360Lys], p.Phe1052Val and p.Gly1069Arg, which were seen with an allele frequency of less than 0.0002. Additional confidence in the assignment of variants is derived from the observation that variants that were non-penetrant for cystic fibrosis occurred at similar frequencies in individuals with cystic fibrosis and subjects of European ancestry in the 1000 Genomes Project. Penetrance analysis should become more useful for clinical applications as the frequencies of rare variants in the healthy population become more robust and complete (for example, with 100,000 genomes) and with more complete delineation of ancestry-based and geographic cohorts, to which we did not have access.

The instances in which the apparent disease liabilities determined by clinical, functional and penetrance criteria were discordant deserve special attention. For the 13 variants meeting clinical but not functional criteria, a common finding was the presence of additional variants in cis that ablated or modified CFTR function, thereby explaining the presence of these variants in individuals with cystic fibrosis. Recognizing that complex alleles may account for discordance between phenotype and genotype is critical in the clinical arena, as misidentification can lead to inappropriate medical actions. These findings emphasize that complete sequencing of the coding regions of genes bearing rare or novel alleles should be undertaken to identify all potentially deleterious alleles. Finally, penetrance analysis was helpful in distinguishing variants that might contribute to disease from those that were neutral. Included in the group not causing cystic fibrosis are known polymorphic variants such as p.Met470Val\textsuperscript{55,56} that seem to have been entered into patient registries owing to incomplete genotyping of CFTR. Individuals carrying variants not causing cystic fibrosis such as p.Met470Val with symptoms indicative of cystic fibrosis may benefit from being re-genotyped. Conversely, individuals diagnosed with cystic fibrosis on the basis of genetic findings of one or more variants now considered not to cause cystic fibrosis should be re-evaluated.

Although this work establishes disease liability for most of the alleles found in individuals with cystic fibrosis, 20 variants remain indeterminate. The ACMG has issued recommendations for the classification of unknown variants beyond those used in this study,\textsuperscript{4,57} however, probabilistic estimation may not be appropriate for all variants deemed indeterminate after extensive clinical, functional and penetrance analyses. As a purpose of this project is to definitively place CFTR variants into well-defined categories, further classification of indeterminate variants will require additional analysis to quantify the probability of causing or not causing disease. For example, it is possible that one or more of the indeterminate variants cause dysfunction of CFTR in a manner distinct from the functional assessments used in this study, as has been shown for two missense changes that also affect RNA splicing (p.Gly576Ala\textsuperscript{58} and p.Ile1234Val; A.S.R. and M.D.A., unpublished data).

The indeterminate variants as well as over 1,600 CFTR variants that are unclassified remain a diagnostic dilemma. Computational approaches predicting disease liability have been applied to splicing\textsuperscript{20} and missense\textsuperscript{19–21} changes to classify CFTR variants. These approaches lack the specificity needed for definitive clinical classification. Algorithms that predict splicing are useful for highly conserved sequences, but experimental studies are needed to analyze the effects of changes to nucleotides that are less well conserved or are located outside of consensus splice sites, as shown in the Supplementary Note\textsuperscript{59,60}. Given the large and diverse structure of the full-length CFTR protein (1,480 residues in length), annotation of more variants for algorithmic training should substantially improve the predictive performance of the classifier. To that end, machine learning approaches could prioritize future experimental testing.

Increased use of sequencing in the clinical setting has emphasized the medical challenges posed by rare variants. Although it appears to be a daunting task to determine the disease liability of all variants accounting for a mendelian disease, the CFTR2 project demonstrates the feasibility of the task using a phenotype-driven approach. Patient registries have been assembled for many genetic disorders\textsuperscript{61–63} that should enable the collation of patient genotype and associated phenotype data for detailed analysis. Microattribution can identify the data source and composition, while acknowledging the contributor and data integrity\textsuperscript{64}. For recessive disorders, the number of alleles of each gene in the human population is finite and stable (excepting extremely rare de novo variants). Thus, careful assignment of disease liability to the variants responsible for these disorders will be valuable for current and future generations of patients and their family members.

\textbf{METHODS}

Methods and any associated references are available in the online version of the paper.

\textbf{Accession codes}. All data can be accessed via the CFTR2 website. Individual variant rsIDs (created by dbSNP) are listed in Supplementary Table 2. Variants have been submitted to ClinVar and the Leiden Open Variant Database (LOVD) and are searchable under the NCBI CFTR gene ID 1080. The RefSeq accession for CFTR is NM\textunderscore 0004922.3, and the UniProt accession for CFTR is P13569.

\textit{Note}: Any Supplementary Information and Source Data files are available in the online version of the paper.
ACKNOWLEDGMENTS

The authors would like to thank the subjects who participated in their national/clinical registries. The authors thank D. Gruenewald (University of California, San Francisco) for CFBE 41o- cells and M. TMBB, University of Iowa for FRT cells. This work was supported by grants from the NIDDK (5R37DK044033 to G.R.C.) and the US NIH (DK49835 to P.T.) and from funding from Cystic Fibrosis Foundation Therapeutics, Inc. (to P.T.), the US Cystic Fibrosis Foundation (CUTTING08A, CUTTING09A and CUTTING10A to G.R.C. and SOSN1A91Q to F.R.S.) and FCT/Portugal (PI/CIC/85103/2007 and PEstOE/ BIA/UI4046/2011 to M.D.A. and Bioanalysis with statistical analysis was provided by E. Johnson, M.B. Drummond, D. Cutler and A. Arking. The authors received considerable guidance from the CFTR2 clinical expert panel: C. De Boeck, P. Durie, S. Elborn, P. Farrell, M. Knowles and I. Sermet; from the CFTR2 functional studies expert panel: R. Bridges, G. Lukacs and D. Sheppard; and from M. Sheridan who provided critical review.

DNA samples from fathers of individuals with cystic fibrosis were contributed by T. Casals (Bellvitge Biomedical Research Institute, Spain), G. Cutting (Johns Hopkins University, USA), C. Deechiuri (University of Verona, Italy), R. Dorfman (The Hospital for Sick Children, Canada), C. Ferec (Centre Hospitalier Universitaire, France), E. Girodon (GH Henri Mondor, France), M. Macek Jr. (Charles University, Czech Republic), D. Radjokovic (Institute of Molecular Genetics and Genetic Engineering, Serbia), M. Schwarz (St. Mary’s Hospital, UK), M. Stuhrmann (Medical School Hannover, Germany), M. Tzetis (National Kapodistrian University of Athens, Greece) and Z. Zielenski (The Hospital for Sick Children, Canada, with partial support from Genome Canada, through the Ontario Genomics Institute per research agreement 2004-OGI-3-05).

CFTR2 data were contributed by C. Barreto (Hospital Santa Maria, Portugal), D. Bilton (Royal Brompton and Harefield Hospital, UK), J. Borg (University of Malta, Malta), C. Colombo (University of Milan, Italy), S. Doudounakis (Aghia Sophia Children’s Hospital, Greece), H. Ellemunter (Innsbruck Medical University, Austria), G. Fletcher (Cystic Fibrosis Registry of Ireland, Ireland), I. Galeva (University Hospital Alesandroskaya, Bulgaria), S. Gartner (Hospital Vall de Hebron Unidad de Fibrosis Quística, Spain), V.A.M. Gullans (Dutch Cystic Fibrosis Foundation, The Netherlands), E. Hatzigourou (Aristotle University, Greece), L. Hjelte (Karolinska Institutet, Sweden), T. Kahre (University of Tartu, Estonia), N. Krulišová (Institute of Biomedical Sciences, Czech Republic, Russia), A. Katerali (Aghia Sophia Children’s Hospital, Greece), P. Laiseau (Universidad del Rosario, Colombia), L. Lemmonier (Association Vaincre La Mucoviscidose, France), A. Lindblad (Sahlgrenska University Hospital, Sweden), V. Lucidi (Ospedale Bambino Gesù, Italy), M. Macek Jr. (Charles University, Czech Republic), H. Makikh (Ukrainian Academy of Medical Sciences, Ukraine), B. Marshall (US Cystic Fibrosis Foundation, USA), I. McIntosh (Cystic Fibrosis Canada, Canada), M. Met-Zahav (Tbil University, Israel), P. Minic (Mother and Child Health Institute of Serbia, Serbia), H. Veberit Olsen (Aarhus University Hospital, Denmark), N. Petrova (Russian Academy of Medical Sciences, Russia), T. Pressler (University of Copenhagen, Denmark), D. Radiovjevic (Mother and Child Health Institute of Serbia, Serbia), S. Ravilly (Association Vaincre La Mucoviscidose, France), N. Regamey (University Hospital Bern, Switzerland), G. Repetto (Universidad de la Frontera, Chile), M.T. Sanseverino (Hospital de Clinicas de Porto Alegre, Brazil), C. Serceri (University of Malta, Malta), A. Stephenson (Cystic Fibrosis Canada, Canada), M. Stern (University of Tubingen, Germany), V. Svbke (Riga Stradinus University, Latvia), M. Thomas (Belgian Cystic Fibrosis Registry, Belgium), I. Tzanakas (Aristotle University, Greece), V. Vavrova (Charles University and University Hospital Motol, Czech Republic) and P. Wenzlaff (Centre for Quality and Management in Health Care, Germany).

AUTHOR CONTRIBUTIONS

P.R.S. jointly supervised research, collected and curated clinical data, performed statistical analysis, analyzed the data and wrote the manuscript. K.R.S. performed clinical data, jointly supervised research and analyzed the data. C.M.P. jointly supervised research and analyzed the data. C.C. coordinated the collection of clinical data, jointly supervised research and analyzed the data. C.G. Repetto (Universidad del Desarrollo, Chile), M.T. Sanseverino (Hospital de Clinicas de Porto Alegre, Brazil), C. Serceri (University of Malta, Malta), A. Stephenson (Cystic Fibrosis Canada, Canada), M. Stern (University of Tubingen, Germany), V. Svbke (Riga Stradinus University, Latvia), M. Thomas (Belgian Cystic Fibrosis Registry, Belgium), I. Tzanakas (Aristotle University, Greece), V. Vavrova (Charles University and University Hospital Motol, Czech Republic) and P. Wenzlaff (Centre for Quality and Management in Health Care, Germany).

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
29. Farrell, P.M. et al. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. J. Pediatr. 153, S4–S14 (2008).
30. Wilschanski, M. et al. Mutations in the cystic fibrosis transmembrane regulator gene and in vivo transspenhalial potentials. Am. J. Respir. Crit. Care Med. 174, 787–794 (2006).
31. Frischmeyer, P.A. & Dietz, H.C. Nonsense-mediated mRNA decay in health and disease. Hum. Mol. Genet. 8, 1893–1900 (1999).
32. Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W. & Kulozik, A.E. NMD: RNA biology meets human medicine. Biochem. J. 430, 365–377 (2010).
33. Ramalho, A.S. et al. Transcript analysis of the cystic fibrosis splicing mutation 1525–1G>A shows use of multiple alternative splicing sites and suggests a putative role of exonic splicing enhancers. J. Med. Genet. 40, e88 (2003).
34. Highsmith, W.E. Jr. et al. A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N. Engl. J. Med. 331, 974–980 (1994).
35. Chillón, M. et al. A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA->G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. Am. J. Hum. Genet. 56, 623–629 (1995).
36. Highsmith, W.E. Jr. et al. Identification of a splice site mutation (2789+5G>A) associated with small amounts of normal CFTR mRNA and mild cystic fibrosis. Hum. Mutat. 14, 133–144 (1999).
37. Beck, S. et al. Cystic fibrosis patients with the 3272–26A->G mutation have mild disease, leaky alternative mRNA splicing, and CFTR protein at the cell membrane. Hum. Mutat. 9, 332–338 (1997).
38. Ramalho, A.S. et al. Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. Am. J. Respir. Cell Biol. 27, 619–627 (2002).
39. Dujardin, G., Commandeur, D., Le Jossic-Corcos, C., Ferec, C. & Corcos, L. Splicing defects in the CFTR gene: minigene analysis of two mutations, 1811+1G>C and 1898+3A>G. J. Cyst. Fibros. 10, 212–216 (2011).
40. Chu, C.-S., Trapnell, B.C., Currin, S.M., Cutting, G.R. & Crystal, R.G. Extensive post-translational deletion of the coding sequences for part of nucleotide-binding fold I in intron 10 of abnormal mRNA transcripts of the cystic fibrosis transmembrane conductance regulator gene is not associated with the clinical manifestations of cystic fibrosis. J. Clin. Invest. 90, 785–790 (1992).
41. Johnson, L.G. et al. Efficiency of gene transfer for restoration of normal airway epithelial function in cystic fibrosis. Nat. Genet. 2, 21–25 (1992).
42. Cheng, S.H. et al. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell 63, 827–834 (1990).
43. Mendoza, J.L. et al. Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. Cell 148, 164–174 (2012).
44. Bombieri, C. et al. Recommendations for the classification of diseases as CFTR-related disorders. J. Cyst. Fibros. 10 (suppl. 2), S86–S102 (2011).
45. Tausig, L.M., Lobeck, C., di Sant’Agneese, P.A., Ackerman, D. & Kattwinkel, J. Fertility in males with cystic fibrosis. N. Engl. J. Med. 287, 586–589 (1972).
46. Anguiano, A. et al. Congenital bilateral absence of the vas deferens—a primarily genital form of cystic fibrosis. J. Am. Med. Assoc. 267, 1794–1797 (1992).
47. Abecasis, G.R. et al. An integrated map of genetic variation from 1,092 human genome. Nature 491, 56–65 (2012).
48. Chu, C.-S., Trapnell, B.C., Currin, S.M., Cutting, G.R. & Crystal, R.G. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. Nat. Genet. 3, 151–156 (1993).
49. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F. & South, S.T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. 13, 680–685 (2011).
50. Pagani, F. et al. New type of disease causing mutations: the example of the I148T mutation. Hum. Genet. 12, 1111–1120 (2003).
51. Estivill, X. Complexity in a monogenic disease. Nat. Rev. Genet. 3, 111–116 (2002).
52. Scott, A., Petrykowska, H.M., Hefferon, T., Gotea, V. & Elinski, L. Functional analyses of synonymous substitutions predicted to affect splicing of the CFTR gene. J. Cyst. Fibros. 11, 511–517 (2012).
53. Mailman, M.D. et al. The NCBI dbGaP database of genotypes and phenotypes. Nat. Genet. 39, 1181–1186 (2007).
54. Newcomb, P.A. et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol. Biomarkers Prev. 16, 2331–2343 (2007).
55. Tuffery-Giraud, S. et al. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum. Mutat. 30, 934–945 (2009).
56. Mons, B. et al. The value of data. Nat. Genet. 43, 281–283 (2011).
57. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F. & South, S.T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. 13, 680–685 (2011).
58. Pagani, F. et al. New type of disease causing mutations: the example of the I148T mutation. Hum. Genet. 12, 1111–1120 (2003).
59. Estivill, X. Complexity in a monogenic disease. Nat. Rev. Genet. 3, 111–116 (2002).
60. Scott, A., Petrykowska, H.M., Hefferon, T., Gotea, V. & Elinski, L. Functional analyses of synonymous substitutions predicted to affect splicing of the CFTR gene. J. Cyst. Fibros. 11, 511–517 (2012).
61. Mailman, M.D. et al. The NCBI dbGaP database of genotypes and phenotypes. Nat. Genet. 39, 1181–1186 (2007).
62. Newcomb, P.A. et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol. Biomarkers Prev. 16, 2331–2343 (2007).
63. Tuffery-Giraud, S. et al. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum. Mutat. 30, 934–945 (2009).
64. Mons, B. et al. The value of data. Nat. Genet. 43, 281–283 (2011).
65. Castellani, C. et al. Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. J. Cyst. Fibros. 7, 179–196 (2008).
66. Rohlfis, E.M. et al. The I148T CFTR allele occurs on multiple haplotypes: a complex allele is associated with cystic fibrosis. Genet. Med. 4, 319–323 (2002).
Analysis of variants expected to affect mRNA splicing. CFTR variants predicted to alter splicing efficiency that were not previously studied (Supplementary Table 2) were examined to confirm their deleterious nature using minigene constructs as previously described with some modifications 69. Briefly, a five-step strategy was employed. First, we amplified the 5′-acceptor and 3′-donor splice-site sequences of the intronic region of interest along with flanking exons from genomic DNA using KOD Hot-Start DNA polymerase (Novagen). Primer sequences are available upon request. Second, fusion PCR was performed on the amplicons generated in the first step using only the exonic primers, creating a fusion amplicon with 5′-acceptor and 3′-donor splice-site sequences with respective exons on either side. Third, we performed sticky end mutagenesis of pcDNAS/FRT/CFTR using the fusion PCR ampiclon as the primer to create a pcDNAS/FRT/CFTR minigene70. Fourth, we used site-directed mutagenesis using the fusion PCR amplicon as the primer to create a particular genotype (variant of interest/known cystic fibrosis–causing variant and without RNA were included. The amount of correctly spliced product from each CFTR–donor minigene was calculated from the sequencing data as described previously 72. Protein blot analysis was performed to evaluate the amount of complex glycosylated (C-band) CFTR. Mouse monoclonal antibodies 570 (R domain or 596; NB2D, UNC antibody distribution program sponsored by Cystic Fibrosis Foundation Therapeutics) and/or MM13-4 (N terminal; Chemicon) were used (1:5,000 dilution) to detect CFTR. GAPDH or tubulin was used as a loading control. Blots were quantified using ImageJ software (NIH) to determine the amount of processed CFTR (C band) for each experimental sample relative to the amount produced by the wild-type minigene.

Analysis of variants expected to alter protein processing and/or function. Variants causing an amino acid substitution or an in-frame deletion were introduced individually into CFTR cDNA using site-directed mutagenesis as previously described 73. The wild-type CFTR clone was obtained from an individual who did not have cystic fibrosis and encoded the known neutral variant p.ValI475Met. Transient expression of CFTR in HeLa cells (Clontech) was achieved as described previously 43. Stable expression of CFTR in FRT cells (a kind gift from M. Welsh) was achieved by integrating each mutated CFTR cDNA as a single copy into the same genomic location using the Invitrogen Flip-In system as described previously 73,74. After selection and confirmation of expression of the CFTR cDNA with the desired variant, the levels of heterologous human CFTR mRNA were determined for each cell line. Cell lines with mRNA levels of >0.5-fold or ≤0.3-fold the average level of four independent FRT cell lines expressing wild-type CFTR were tested. CFTR maturation and the amount of protein expression were quantified using the ratio of C-band CFTR to B-band + C-band CFTR (normalized to wild-type CFTR as described previously 43). Forskolin-activated, CFTR-dependent chloride secretion was measured in confluent FRT cells by short-circuit current (I_{sc}) in Ussing chambers. I_{sc} measurements were repeated 3 to 14 times per cell line and averaged. All readings were reported as a percentage of the average I_{sc} of wild-type CFTR–expressing FRT cell lines 75. Measurements from four separate FRT cell lines were used to establish the mean current and variance of wild-type CFTR. Human bronchial epithelial (HBE) cells were isolated from subjects with and without cystic fibrosis, and I_{sc} was measured as previously described 76.

MassARRAY assays. An assay to screen for the 159 CFTR variants seen in 20.01% of affected individuals was developed using open-platform matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry 77. The assay was validated using genomic DNA controls obtained from available study stocks from the United States (n = 99), France (n = 18), Canada (n = 22), the Czech Republic (n = 2) and Serbia (n = 1). When genomic DNA was not available (n = 11), plasmid DNA was generated using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies). We were unable to confirm two variants because of a problem with the extension primer (p.Gly330*) and an inability to derive a positive control (p.Glu1418Argfs*14). The assay was validated with Sanger sequencing. In validation studies, 35 of 35 variants were identified, and 29 of 29 wild-type chromosones were confirmed.

The multiplex assay design was initially accomplished using Assay Designer Software, Version 4.0 (Sequenom) to design both amplification and extension primers and was subsequently optimized using results with the positive controls. Multiplex PCR amplification of regions of up to 300 nt in length from genomic DNA, whole genome–amplified DNA or plasmid DNA at a concentration of 25, 50 or 5 ng/ml, respectively, was performed in 384-well plates. The reaction contained reagents from the iPLEX Gold SNP Genotyping kit (Sequenom). Because of variation in amplicon lengths and the likelihood of primer-dimer formation, differential primer concentrations were used. Primer multiplexes were evaluated for possible dimers and hairpins using NIST primer tools (see URLs) and subsequently adjusted. Details of primer sequences and the concentrations used are available upon request. Unincorporated dNTPs were neutralized using reagents from the iPLEX Gold Reagent kit (Sequenom).

Single-base extension products were generated from each purified PCR reaction using reagents from the iPLEX Gold Reagent kit. The resulting single-base extension PCR products were prepared for mass spectrometry and dispensed to a SpectroCHIP (Sequenom). Data were analyzed using a MALDI-TOF spectrometer (Sequenom). Data were generated with SpectroACQUIRE software version 3.0 on the MassARRAY spectrometer (Sequenom) and then analyzed using Typer software, version 4.0.22 (Sequenom). This assay was
employed to test DNA from the fathers of offspring with cystic fibrosis for variants suspected of being cystic fibrosis causing. Fathers provided informed consent for genetic study and were anonymized for analysis. Each father should carry only the deleterious \textit{CFTR} variants passed to his offspring. Additional variants detected in fathers represent either a complex allele in which two \textit{CFTR} variants are present on the same chromosome or a \textit{CFTR} variant in trans with the transmitted variant that is insufficiently deleterious to make the father infertile. Fathers with a cystic fibrosis–causing variant and a second variant had their offspring genotyped if available ($n = 145$) to delineate phase.

**Statistical analysis.** Statistical analyses were performed using Intercooled Stata version 11 (StataCorp) using the METAREG plug-in (see URLs). A meta-analysis approach was used to compare clinical features in subjects with PTC versus non-PTC variants and to perform regression analysis of aggregate data for each cell line. We incorporated the observed variance across subjects for each variant and allowed for heterogeneity of variance across variants to compare groups. Group means (for chloride current) were compared under a random-effects model accounting for statistical variance in the measurements both within and across the variant groups. All reported regression coefficients are unstandardized.

67. Wang, X., Dockery, D.W., Wypij, D., Fay, M.E. & Ferris, B.G. Jr. Pulmonary function between 6 and 18 years of age. \textit{Pediatr. Pulmonol.} \textbf{15}, 75–88 (1993).

68. Hankinson, J.L., Odencrantz, J.R. & Fedan, K.B. Spirometric reference values from a sample of the general U.S. population. \textit{Am. J. Respir. Crit. Care Med.} \textbf{159}, 179–187 (1999).

69. Cooper, T.A. Use of minigene systems to dissect alternative splicing elements. \textit{Methods} \textbf{37}, 331–340 (2005).

70. Clackson, T. & Winter, G. ‘Sticky feet’-directed mutagenesis and its application to swapping antibody domains. \textit{Nucleic Acids Res.} \textbf{17}, 10163–10170 (1989).

71. Ramalho, A.S., Clarke, L.A. & Amaral, M.D. Quantification of \textit{CFTR} transcripts. \textit{Methods Mol. Biol.} \textbf{741}, 115–135 (2011).

72. Sheridan, M.B. et al. \textit{CFTR} transcription defects in pancreatic sufficient cystic fibrosis patients with only one mutation in the coding region of \textit{CFTR}. \textit{J. Med. Genet.} \textbf{48}, 235–241 (2011).

73. Yu, H. et al. Ivacaftor potentiation of multiple \textit{CFTR} channels with gating mutations. \textit{J. Cyst. Fibros.} \textbf{11}, 237–245 (2012).

74. Krasnov, K.V., Tzetis, M., Cheng, J., Guggino, W.B. & Cutting, G.R. Localization studies of rare missense mutations in cystic fibrosis transmembrane conductance regulator (CFTR) facilitate interpretation of genotype-phenotype relationships. \textit{Hum. Mutat.} \textbf{29}, 1364–1372 (2008).

75. Van Goor, F. et al. Rescue of CF airway epithelial cell function in vitro by a \textit{CFTR} potentiator, VX-770. \textit{Proc. Natl. Acad. Sci. USA} \textbf{106}, 18825–18830 (2009).

76. Neuberger, T., Burton, B., Clark, H. & Van Goor, F. Use of primary cultures of human bronchial epithelial cells isolated from cystic fibrosis patients for the pre-clinical testing of \textit{CFTR} modulators. \textit{Methods Mol. Biol.} \textbf{741}, 39–54 (2011).

77. Thongnoppakhun, W. et al. Simple, efficient, and cost-effective multiplex genotyping with matrix assisted laser desorption/ionization time-of-flight mass spectrometry of hemoglobin $\beta$ gene mutations. \textit{J. Mol. Diagn.} \textbf{11}, 334–346 (2009).