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

Patrick R Sosnay, Karen R Siklosi, Fredrick Van Goor, Kyle Kaniecki, Haihui Yu, Neeraj Sharma, Anabela S Ramalho, Margarida D Amaral, Ruslan Dorfman, Julian Zielenski, David L Masica, Rachel Karchin, Linda Millen, Philip J Thomas, George P Patrinos, Mary Corey, Michelle H Lewis, Johanna M Rommens, Carlo Castellani, Christopher M Penland, & Garry R Cutting

SUPPLEMENTARY NOTE

Public availability of data

A summary of data presented in this manuscript summarized by variant is available online at www.cftr2.org. This website is curated by the CFTR2 project team and will be updated as further data is collected. The data is linked by variant to the existing Cystic Fibrosis Mutation Database (http://www.genet.sickkids.on.ca/app). Additionally, variant information has been uploaded to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) and the Leiden Open Variation Database (LOVD; http://www.lovd.nl/3.0/home) with links to www.cftr2.org. The 159 variants in CFTR2 that were seen with an allele frequency greater than 0.01% are listed with associated clinical, functional and population/penetrance data in Supplementary Table 2 online. The final determination of the disease liability of the variant is listed in column "AC".

Data contribution

The researchers and clinicians who submitted patients to the CFTR2 database are acknowledged via microattribution, a process that provides links in publications to clinicians and investigators who contributed DNA variation data to a public resource[1, 2] (Supplementary Tables 1 and 4). Thirteen countries/registries contributed data only from a subset of their patient population such that patients with two p.Phe508del alleles or patients with two alleles found on the ACMG list of 23 common variants were excluded from submission. The 2,655 patients

submitted from these registries represent 7.5% of the total number of CFTR2 patients analyzed. Variant frequency calculations are based on all collected CFTR2 patient data.

Average sweat chloride concentration as basis for clinical assessment

The mean sweat chloride concentration in 10,106 patients homozygous for the common cystic fibrosis variant p.Phe508del was 102.5mmol/L. This value is 5 standard deviations above the mean sweat chloride of normal individuals (16 mmol/L across the age range represented by the CFTR2 population)[3-6]. Although this phenotype measurement is variable, the sweat chloride concentration is sufficiently raised in p.Phe508del homozygotes such that 99.5% of the patients are above the established cystic fibrosis diagnostic threshold of 60 mmol/L (Supplementary Figure 1a)[7]. Variation in sweat chloride concentration is exhibited for all CFTR variants and variants associated with mild exocrine pancreatic disease (pancreatic sufficiency or PS)[8] have a lower mean value. Hence, a fraction of patients carrying accepted cystic fibrosis causing variants have sweat chloride concentrations below 60 mmol/L. For example, the 277 patients carrying a single copy of *c.3717+12191C->T* (legacy name 3849+10kB C->T), whose other *CFTR* allele contained a variant previously identified as cystic fibrosis-causing, had a mean sweat chloride concentration of 67.3 mmol/L ± 22.6[9]. Sweat chloride concentrations in 29.5% of those 277 patients were below the diagnostic threshold of 60 mmol/L (Supplementary Figure 1b).

Functional Analysis: 5T and 7T variants.

There is known variation in the length of a polythymidine region of intron 9 of *CFTR* (*c.1210-12T[5]*; legacy name 5T, *c.1210-12T[7];* legacy name 7T, and *c.1210-12T[9];* legacy name 9T). The 5T variant has been described as a potentially deleterious variant with low penetrance in the setting of other intragenic modifiers[11-14] due to abnormal splicing (exon 9 skipping). The 5T variant also modifies the penetrance of the missense variant p.Arg117His[10]. The 7T variant is

the common allele in the polythymidine tract of intron 9 and does result in some exon 9 skipping but is not a cystic fibrosis-causing variant[15]. Given the abundance of published studies on these complex alleles, the 5T and 7T variants were excluded from functional analysis. The 9T variant was not reported with a frequency greater than 0.01% in the CFTR2 database and was therefore not included in analysis; however, it does not lead to *CFTR* exon 9 skipping and is considered a polymorphic variant[15].

Functional analysis: Premature termination variants

All 80 variants introducing a premature termination codon (PTC) (*in trans* with a known cystic fibrosis-causing variant) were associated with mean sweat chloride concentration unequivocally in the diagnostic range for cystic fibrosis (mean sweat chloride concentration for all variants introducing a PTC=102.73 mmol/L, 95%CI 101.22-104.25; Supplementary Figure 2a) and higher than for patients carrying *CFTR* variants not predicted to cause a PTC (also *in trans* with the same known cystic fibrosis-causing variants) (mean=84.70 mmol/L, 95%CI 79.68-89.72, $p<0.001$ vs. PTC group). Similarly, the prevalence of severe exocrine pancreatic disease (pancreatic insufficiency or PI) was higher in patients carrying a variant predicted to introduce a PTC (PTC mean=95.80%, 95%CI 93.49-98.11; non-PTC mean=59.63%, 95%CI=52.48-66.77; $p<0.001$ between groups; Supplementary Figure 2b).

Functional analysis: Splice site variants

Previously performed functional studies of five variants that alter *CFTR* splicing are summarized in Supplementary Table 3a. In four of five, *CFTR* RNA transcript was obtained from nasal epithelial cells and quantified by reverse-transcription PCR (RT-PCR); one variant was studied using minigene analysis.

Five previously unstudied variants predicted to affect RNA splicing were analyzed using *CFTR* minigenes (see methods) transfected into two human cell lines (Human Embryonic Kidney [HEK] 293 cells and Cystic Fibrosis Bronchial Epithelial [CFBE41o-]) homozygous for the p.Phe508del allele (a generous gift from Prof. D. Gruenert, University of California-San Francisco, San Francisco, CA) as previously described[22]. These cell lines were ideal for our study as neither exhibited detectable levels of endogenous CFTR transcript or protein (Sharma et al., manuscript in preparation). To evaluate concordance between functional data on these variants from samples obtained directly from patients and our *in vitro* minigene analysis, mutant minigenes were created bearing *c.579+1G>T* (legacy name 711+1G>T), *c.1585-1G>A* (legacy name 1717-1G>A) or *c.2657+5G>A* (legacy name 2789+5G>A). The mutant minigenes (Ramalho et al., manuscript submitted; Sharma et al., manuscript in preparation) had essentially the same splicing consequences that were observed from patient samples (Supplementary Table 3b). An additional transcript (*c.1717delG*) was detected on sequencing the products generated by the minigene bearing *c.1585-1G>A* (legacy name 1717-1G>A); however, even if the aberrant transcript avoided NMD, it would generate a truncated protein due to in-frame termination codon caused by the frameshift.

RT-PCR of RNA transcripts generated by *CFTR* minigenes containing each of the 5 putative splicing variants was performed. Amplification products were sequenced to determine the quantity of correctly spliced RNA, as previously described[26]. The amount of amplified product from *CFTR* minigenes bearing each variant was derived as a fraction of the amount of amplified product relative to the respective WT-*CFTR* minigene (Supplementary Table 3c). Western blot analysis of CFTR protein translated from the minigene RNA was performed to confirm the results of the RNA splicing analysis. The amount of mature complex glycosylated C-band was determined for each variant relative to the amount of protein generated by the WT minigene. RNA and protein studies were performed in triplicate in each cell line. Close correspondence

was observed between the level of correctly spliced transcript and fully processed protein (Supplementary Table 3c).

Assessment of the effect of variants upon CFTR processing

CFTR processing involves proper folding and core glycosylation in endoplasmic reticulum, complex glycosylation in the golgi apparatus and trafficking to the plasma membrane. *CFTR* variants that interrupt processing or alter folding are recognized and targeted for degradation[27]. Cells with *CFTR* variants that disrupt folding express much lower levels of mature complex glycosylated protein compared to WT-CFTR[28]. To assess the effect of amino acid substitution and in-frame deletions upon protein processing, the fraction of partially glycosylated (B-band) to fully glycosylated CFTR (C-band) was analyzed in transiently transfected HeLa cells (n=67 variants)[29] and stably transfected FRT cells (63 of the 67 variants)[30]. The ratio of complex glycosylated C-band to core glycosylated B-band + C-band (C/(B+C)) determined for each variant was generally comparable in the two cell lines (r^2=0.94; Supplementary Figure 3).

Functional analysis: Primary airway cells

When CFTR is correctly localized to the apical membrane, it acts as a chloride channel to regulate fluid and electrolyte levels in the apical environment. Chloride current is therefore one reasonable assessment of CFTR function. To evaluate the generalizability of the FRT cell line model to correlate CFTR function *in vitro* and clinical characteristics that reflect *CFTR* variant severity, human bronchial epithelial (HBE) cells were obtained (acknowledgements to Dr. Joseph W. Pilewski, University of Pittsburgh) from cystic fibrosis patients undergoing lung transplant. The forskolin-stimulated chloride short-circuit current of primary HBE cells from cystic fibrosis patients is expressed as a percentage of wild-type CFTR current generated by HBE cells from non-cystic fibrosis donors (cells harvested from patients undergoing lung transplantation for other reasons analyzed in the same fashion) in Supplementary Figure 4. In

each HBE culture, the variant of interest was present *in trans* with a variant known to generate less than 1% of wild-type CFTR chloride conductance (p.Phe508del or p.Gly551Asp). There was ordinal correlation of the currents in primary HBE cells when compared to currents generated by CFTR bearing the same variants expressed in FRT cells.

The relationship between CFTR processing (C/(B+C ratio)) and CFTR chloride current

To determine the threshold at which misprocessing of CFTR mutants caused severe reduction in CFTR chloride channel activity, the C/(B+C) ratio and corresponding chloride current (normalized to WT-CFTR) was plotted for 76 mutants (including 9 that do not occur in the group of 159 variants analyzed in this study) expressed in FRT cells (Supplementary Figure 5). Processing and chloride current demonstrate a non-linear relationship such that 10% of WT-CFTR function was not exceeded until the C/(B+C) ratio was greater than 0.6 (normalized to WT). Furthermore, the 10 cell lines that had C/(B+C) ratio less than 0.1 in FRT and HeLa cells (indicating a severe CFTR processing defect) all generated chloride currents less than 10% of WT-CFTR. None of the eleven variants exhibiting a severe processing defect in at least one cell line generated chloride currents greater than 10% of WT. On the basis of these results, we concluded that a variant that demonstrated a severe processing defect in HeLa or FRT cells (C/(B+C) <0.1) met functional criteria as deleterious, even if CFTR chloride channel function was not determined.

Genotyping of fathers of cystic fibrosis patients for 159 variants

Variants that did not meet both clinical and functional criteria to be cystic fibrosis-causing underwent penetrance screening to aid in determining disease liability. Fathers (n=2,188) of cystic fibrosis-offspring were genotyped for 159 variants using the assay previously described. Eleven samples were excluded due to contamination, a sample swap between father/offspring, or low quality/quantity DNA for analysis. Additional filtering yielded 2,062 fathers with at least

one identified variant for penetrance analysis (Supplementary Figure 6). There were 166 samples found to have more than one variant identified, including one of 22 ACMG variants (p.Arg117His was excluded from the ACMG list due to reduced penetrance, which has been well-described)[10]. When possible, offspring were genotyped to confirm the phase of variants in the fathers' samples. One hundred samples were confirmed to have a variant *in trans* with an ACMG variant, indicating non-penetrance of the non-ACMG variant.

Assay coverage (the percentage of variants successfully genotyped) was greater than 90% of fathers having better than 151 out of 159 variants successfully typed (95% assay coverage) (Supplementary Figure 7). Samples with <75% assay coverage were considered unsuccessful and their variant results were not interpreted.

The analysis of fathers and their offspring was valuable as it further specified *CFTR* variants that occur on the same chromosome as part of a complex allele. For example, p.Arg668Cys occurred as a second *CFTR* variant in 24 fathers. In 14 of these 24 fathers, an additional variant was present *in cis* with p.Arg668Cys: p.Gly576Ala in 12 fathers and *c.3717+12191C->T* (legacy name 3849+10kbC->T) in 2 fathers. An additional 6 fathers with p.Arg668Cys had a third variant, but phase could not be confirmed. One variant meeting clinical criteria only (p.Ile1027Thr) was confirmed to be on the same chromosome with p.Phe508del as seen previously[31] in 34/34 fathers carrying p.Ile1027Thr who had offspring DNA available for phase confirmation (data not shown).

Reference List

1. Giardine,B. *et al.* Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. *Nat. Genet.* **43**, 295-301 (2011).

2. Patrinos,G.P. *et al.* Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. *Hum. Mutat.* **33**, 1503-1512 (2012).

3. Farrell,P.M. & Koscik,R.E. Sweat chloride concentrations in infants homozygous or heterozygous for F508 cystic fibrosis. *Pediatrics* **97**, 524-528 (1996).

4. Massie,J., Gaskin,K., Van,A.P., & Wilcken,B. Sweat testing following newborn screening for cystic fibrosis. *Pediatr. Pulmonol.* **29**, 452-456 (2000).

5. Mishra,A., Greaves,R., & Massie,J. The limitations of sweat electrolyte reference intervals for the diagnosis of cystic fibrosis: a systematic review. *Clin. Biochem. Rev.* **28**, 60-76 (2007).

6. Jayaraj,R. *et al.* A reference interval for sweat chloride in infants aged between five and six weeks of age. *Ann. Clin Biochem.* **46**, 73-78 (2009).

7. 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).

8. 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).

9. Dreyfus,D., Bethel,R., & Gelfand,E. Cystic fibrosis 3849+10kb C>T mutation associated with severe pulmonary disease and male fertility. *Am. J. Resp. Crit. Care Med.* **153**, 858-860 (1996).

10. Kiesewetter,S. *et al.* A mutation in the cystic fibrosis transmembrane conductance regulator gene produces different phenotypes depending on chromosomal background. *Nature Genet.* **5**, 274-278 (1993).

11. Kerem,E. *et al.* A cystic fibrosis transmembrane conductance regulator splice variant with partial penetrance associated with variable cystic fibrosis presentations. *Am. J. Resp. Crit. Care Med.* **155**, 1914-1920 (1997).

12. Cuppens,H. *et al.* Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes: The polymorphic (TG)m locus explains the partial penetrance of the 5T polymorphism as a disease mutation. *J. Clin. Invest.* **101**, 487-496 (1998).

13. Noone,P.G. *et al.* Lung disease associated with the IVS8 5T allele of the CFTR gene. *Am. J Respir. Crit Care Med.* **162**, 1919-1924 (2000).

14. Lebo,R.V. & Grody,W.W. Variable penetrance and expressivity of the splice altering 5T sequence in the cystic fibrosis gene. *Genet. Test.* **11**, 32-44 (2007).

15. Chu,C.-S., Trapnell,B.C., Curristin,S., Cutting,G.R., & Crystal,R.G. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nature Genet.* **3**, 151-156 (1993).

16. 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).

17. Chillon,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).

18. 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).

19. 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.* **9**, 332-338 (1997).

20. 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.* **14**, 133-144 (1999).

21. 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 Mol. Biol.* **27**, 619-627 (2002).

22. Amaral,M.D. *et al.* Quantitative methods for the analysis of CFTR transcripts/splicing variants. *J Cyst. Fibros.* **3 Suppl 2**, 17-23 (2004).

23. Zielenski,J. *et al.* Analysis of CFTR transcripts in nasal epithelial cells and lymphoblasts of a cystic fibrosis patient with 621 + 1G -> T and 711 + 1G -> T mutations. *Hum. Mol. Genet.* **2**, 683-687 (1993).

24. Fonknechten,N., Chomel,J.-C., Kitzis,A., Kahn,A., & Kaplan,J.-C. Skipping of exon 5 as a consequence of the 711 + 1G->T mutation in the CFTR gene. *Hum. Mol. Genet.* **1(4)**, 281-282 (1992).

25. Hull,J., Shackleton,S., & Harris,A. Abnormal mRNA splicing resulting from three different mutations in the CFTR gene. *Hum. Mol. Genet.* **2**, 689-692 (1993).

26. Sheridan,M.B. *et al.* CFTR transcription defects in pancreatic sufficient cystic fibrosis patients with only one mutation in the coding region of CFTR. *J. Med. Genet.* **48**, 235-241 (2011).

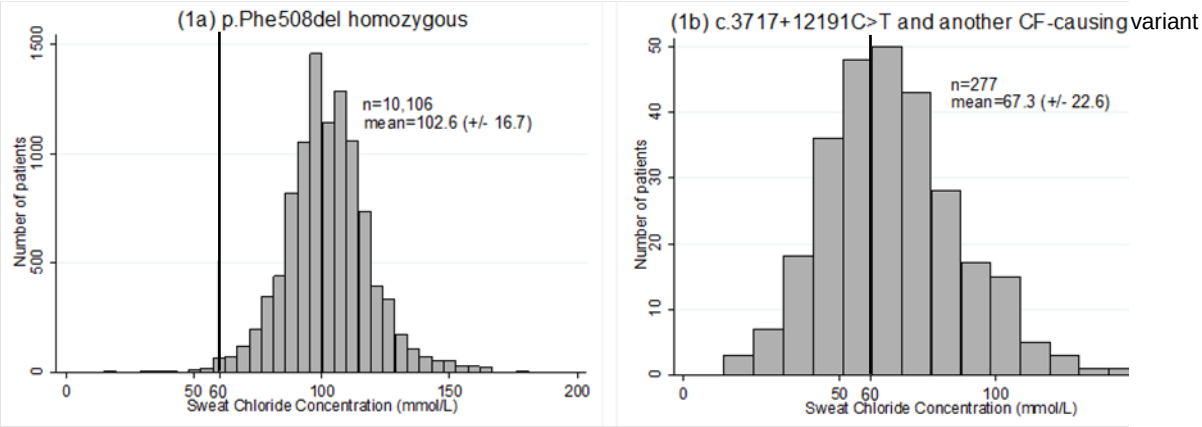
27. Ward,C.L., Omura,S., & Kopito,R.R. Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **83**, 121-127 (1995).

28. 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).

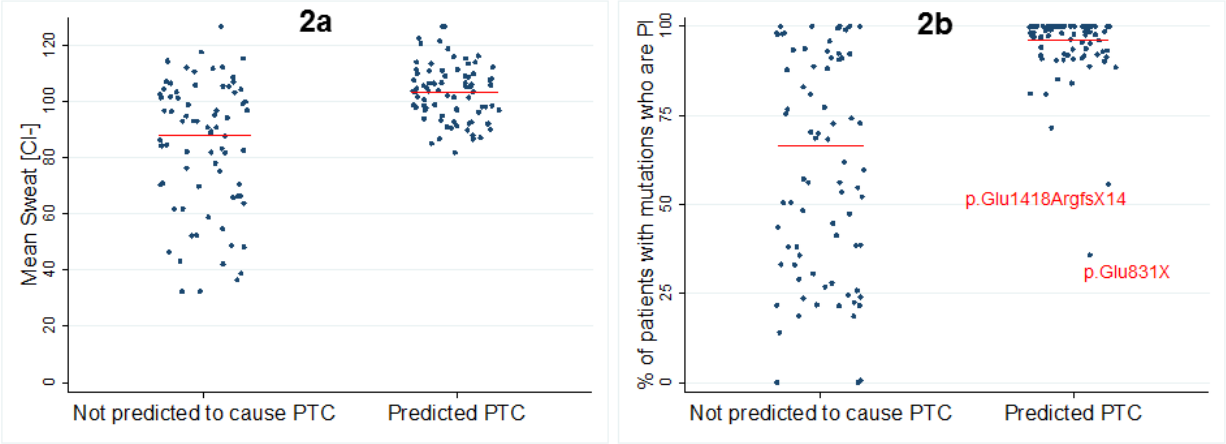
29. Qu,B.H. & Thomas,P.J. Alteration of the cystic fibrosis transmembrane conductance regulator folding pathway. *J. Biol. Chem.* **271**, 7261-7264 (1996).

30. 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. *Hum. Mutat.* **29**, 1364-1372 (2008).

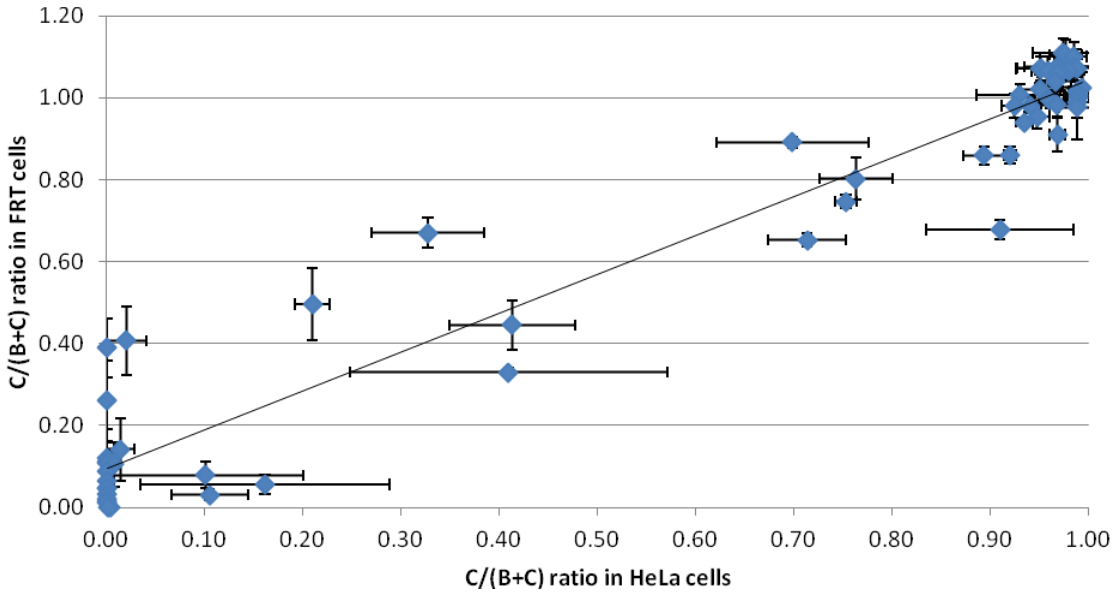
31. Fichou,Y. *et al.* Estimating the age of CFTR mutations predominantly found in Brittany (Western France). *J. Cyst. Fibros.* **7**, 168-173 (2008).



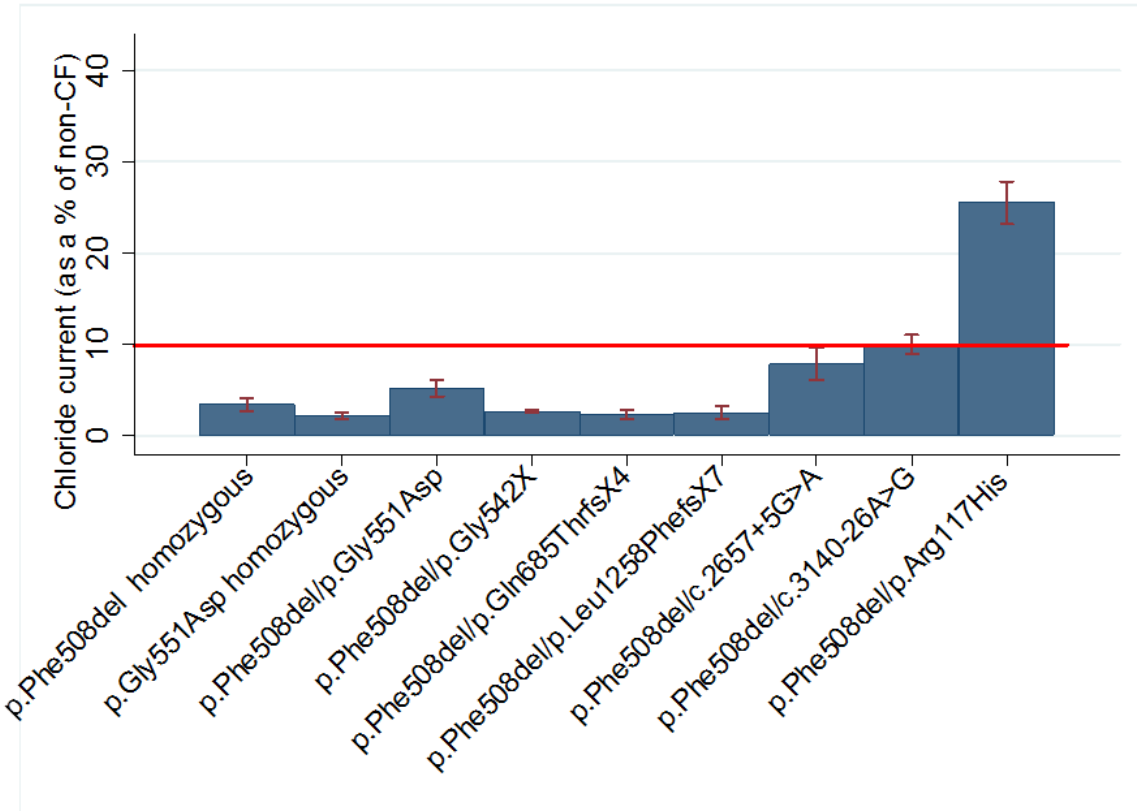
Supplementary Figure 1: Histograms showing the distribution of sweat chloride concentration of patients homozygous for p.Phe508del (1a) or carrying a single copy of *c.3717+12191C>T* (legacy name 3849+10kbC>T) with a cystic fibrosis-causing variant *in trans* (1b). Number of patients, mean, and standard deviation is reported for both figures. Sweat chloride concentration of 60 mmol/L is indicated by a line in each panel.



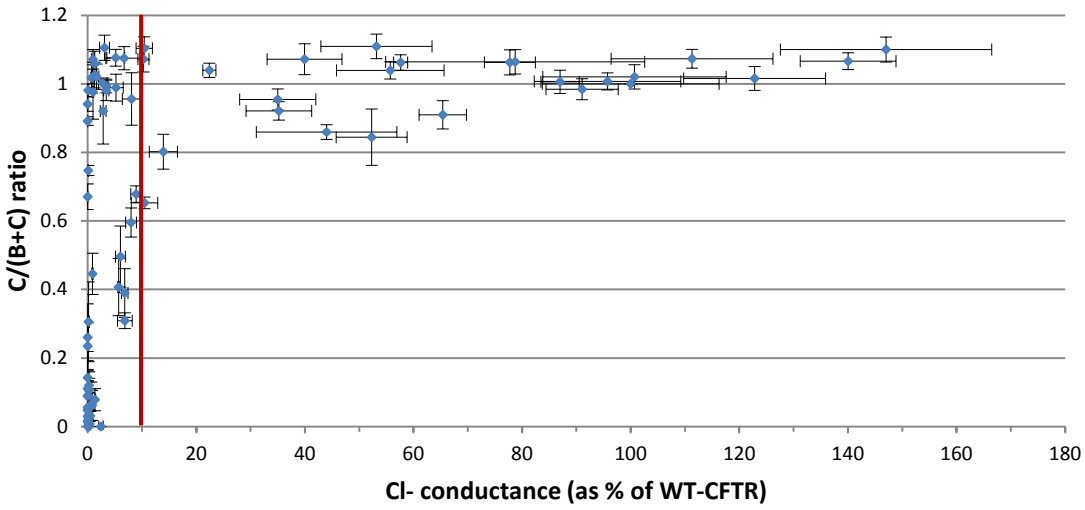
Supplementary Figure 2: The average sweat chloride (2a) and rate of PI (2b) for patients in CFTR2 that carry variants expected to introduce a PTC compared with variants not expected to cause a PTC. Each point represents the mean sweat chloride (2a) or the pancreatic insufficiency prevalence (2b) of all patients in the CFTR2 database with a single PTC variant (80 variants) vs. those with a single variant that does not cause a PTC (79 variants). All patients carried a previously reported cystic fibrosis-causing variant *in trans* with the reported variant. The red bars indicate mean. Variants predicted to cause a PTC, but with lower rates of PI are labeled in red.



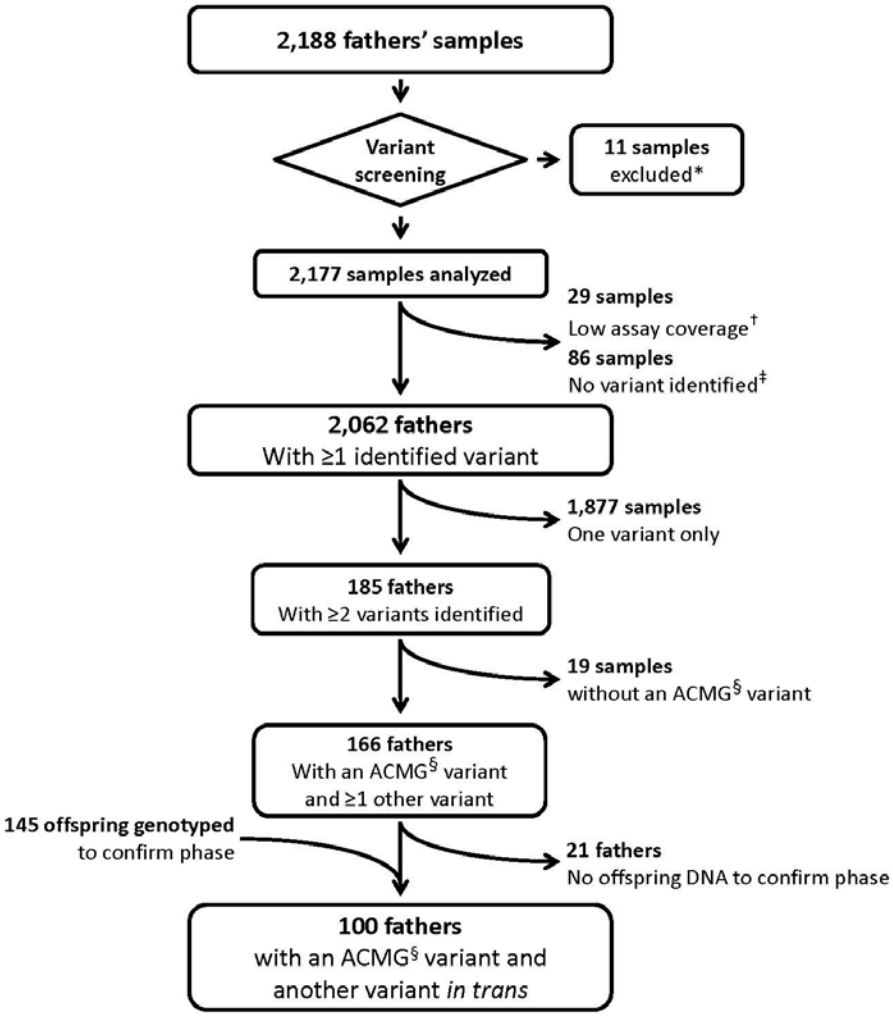
Supplementary Figure 3. The correlation of C-band to B-band + C-band ratio (C/(B+C)) for each variant expressed in HeLa cells and FRT cells. Each point represents an individual *CFTR* variant; error bars show one standard error of the mean C/(B+C) ratio in each direction. Ratios are normalized to the WT-CFTR C/(B+C) ratio determined during analysis. The C/(B+C) ratios in HeLa and FRT cells have a positive linear relationship ($y=0.95x+0.098$; $r^2=0.941$).



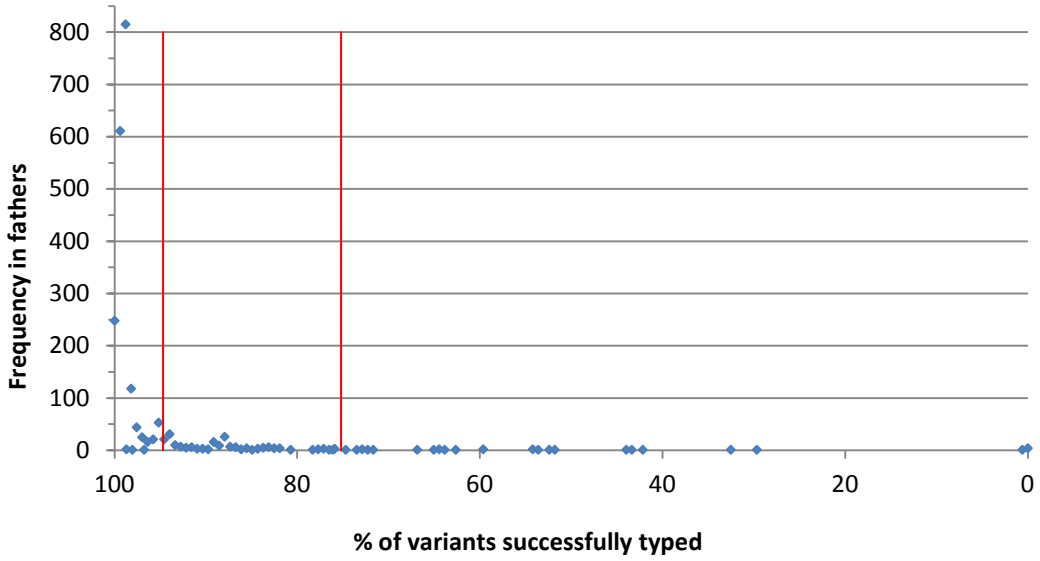
Supplementary Figure 4: Chloride current of selected primary airway epithelial cells from cystic fibrosis patients undergoing lung transplantation. Mean chloride current is plotted as a % of non-cystic fibrosis airway epithelial cells. Red line indicates 10% of WT-CFTR function. The p.Arg117His variant is known to be variably associated with cystic fibrosis[10]. Standard error is shown in red bars. As a comparison, chloride conductance of the following variants in FRT cells are as follows: p.Phe508del= 0.2 ± 0.28; p.Gly551Asp = 2.6 ± 0.81; p.Arg117His = 35.0 ± 7.01.



Supplementary Figure 5: Plot of chloride current expressed as a % of WT-CFTR compared to the maturation (C/(B+C)) ratio for each variant expressed in FRT cells, indicating a non-linear relationship. Each point represents an individual *CFTR* variant; error bars show one standard error of the mean chloride current and C/(B+C) ratio in each direction. C/(B+C) ratios are normalized to the WT-CFTR C/(B+C) ratio determined during analysis. The red line at 10% chloride current represents the threshold at which a variant is functionally deleterious; variants do not exceed this threshold unless the C/(B+C) ratio is at least 0.6.



Supplementary Figure 6: Flow diagram for penetrance analysis in fathers of cystic fibrosis offspring. 2,188 samples were genotyped for 159 variants. Samples were filtered until only those with >1 variant (including a known cystic fibrosis-causing variant from the AMCG list of 23, excluding p.Arg117His) remained for phase analysis. Offspring genotyping determined phase of the fathers' variants.*11 samples excluded due to contamination, father/offspring sample switch, or insufficient sample quality/quantity for analysis; †29 samples excluded due to less than 75% of variants typed. ‡Assay successful; variant assumed to be not on panel. §p.Arg117His excluded from ACMG list due to known reduced penetrance.



Supplementary Figure 7: Genotyping success rate for 159 variants in 2,177 fathers of cystic fibrosis offspring. Each point represents the number of fathers whose samples were typed at a specific level of assay coverage (as a % of variants successfully typed). Red lines indicate 75% and 95% of variants typed. 1,959 fathers had >95% of variants successfully typed, 189 had 75% to 95% typed while 29 had fewer than 75% of variants typed.