Prevalence of *CFTR* variants in primary immunodeficiency patients with bronchiectasis is an important modifying cofactor

Dylan Lawless, PhD,[a] Hana Lango Allen, PhD,[b,c] James E. D. Thaventhiran, MD, PhD,[d,e] Sarah Goddard, MD, PhD,[f] Oliver S. Burren, PhD,[g] Evie Robson, MD,[h] NIHR BioResource–Rare Diseases Consortium,[b] Daniel Peckham, MBBS, DM,[h,i] Kenneth G. C. Smith, MD, PhD,[b,g,j] and Sinisa Savic, MD, PhD[k,l] *Lausanne, Switzerland; and Cambridge, Stoke-on-Trent, and Leeds, United Kingdom*

Background: Cystic fibrosis (CF) is one of the most common life-limiting autosomal-recessive disorders and is caused by genetic defects in the CF transmembrane conductance regulator *(CFTR)* gene. Some of the features of this multisystem disease can be present in primary immunodeficiency (PID).
Objective: We hypothesized that a carrier *CFTR* status might be associated with worse outcome regarding structural lung disease in patients with PID.
Methods: A within-cohort and population-level statistical genomic analysis of a large European cohort of PID patients was performed using genome sequence data. Genomic analysis of variant pathogenicity was performed.
Results: Compared to the general population, p.Phe508del carriage was enriched in lung-related PID. Additionally, carriage of several pathogenic *CFTR* gene variants were increased in PID associated with structural lung damage compared to PID patients without the structural lung damage. We identified 3 additional biallelic cases, including several variants not traditionally considered to cause CF.
Conclusion: Genome sequencing identified cases of *CFTR* dysfunction in PID, driving an increased susceptibility to infection. Large national genomic services provide an opportunity for precision medicine by interpreting subtle features of genomic diversity when treating traditional Mendelian disorders. (J Allergy Clin Immunol 2023;152:257-65.)

Key words: *Bronchiectasis, CFTR, cystic fibrosis, genomics, primary immunodeficiency*

Cystic fibrosis (CF) is one of the most common life-limiting autosomal-recessive disorders, with a prevalence of 1 in 2500 in European people. The causative gene, identified in 1989, is referred to as the CF transmembrane conductance regulator *(CFTR)*.[1] The most common pathogenic variant in the *CFTR* gene is the 3-nucleotide deletion NM_000492.3(*CFTR*):c.1521_1523del (p.Phe508del) (often called delta F508).[1] Today, more than 2000 different variants have been identified in the *CFTR* gene, many of which cause CF.[2] It is estimated that 1 in 25 individuals of European descent carries a heterozygous CF-causing gene variant. Previous studies have assessed the effect of heterozygous variants in the general population. Çolak et al[3] screened 108,034 individuals (3% heterozygous carriers), and using multivariable adjustment, they determined that carriers had increased risk for chronic bronchitis, bronchiectasis, and lung cancer despite similarities in lung function and other morbidity outcomes. Overall survival was similar between those who did and did not carry the *CFTR* gene. A number of studies have previously reported phenotypic associations for heterozygous carriage. However, few studies have used statistically robust analysis methods.

We wondered whether carrier *CFTR* status might be associated with worse outcome regarding structural lung disease in patients with primary immunodeficiency (PID). To test our hypothesis, we included a cohort of patients recruited into the National Institute for Health Research (NIHR) BioResource–Rare Diseases (NIHR BR-RD) study. This UK NIHR BR-RD study was set up with the aim of improving the clinical management of patients with rare diseases and gathering insight through large-scale genomics based on disease cohort phenotypes. At the time of this study, whole genome sequence data from about 8000 individuals had been collected. We focused on a cohort of 1046 PID patients and their relatives. We investigated the presence of *CFTR* carriage

From [a]the Global Health Institute, School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne; [b]the National Institute for Health Research (NIHR) BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge; [c]the MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge, Cambridge; [d]the Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge; [e]the MRC Toxicology Unit, School of Biological Sciences, University of Cambridge, Cambridge; [f]the University Hospitals of North Midlands NHS Trust, Stoke-on-Trent; [g]the Cambridge Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge; [h]the Leeds Cystic Fibrosis Trust Strategic Research Centre, University of Leeds, Leeds; [i]the Leeds Institute of Medical Research at St James's University Hospital, University of Leeds, Leeds; [j]the Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge; [k]the Department of Clinical Immunology and Allergy, St James's University Hospital, Leeds; and [l]the Leeds Institute of Rheumatic and Musculoskeletal Medicine, St James's University Hospital, Leeds.

Supported by the University of Leeds 110 Anniversary Research Scholarship and by the NIHR (grant RG65966). The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.
Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.
Received for publication August 6, 2022; revised January 21, 2023; accepted for publication January 25, 2023.
Available online February 23, 2023.
Corresponding author: Sinisa Savic, MD, PhD, Department of Clinical Immunology and Allergy, St James's University Hospital, Beckett Street, Leeds, United Kingdom. E-mail: s.savic@leeds.ac.uk.

0091-6749

https://doi.org/10.1016/j.jaci.2023.01.035

Abbreviations used

CF: Cystic fibrosis
CFSPID: CF screening positive, inconclusive diagnosis
CFTR: CF transmembrane conductance regulator
CFTR2: Clinical and Functional Translation of CFTR database (cftr2.org)
ENaC: Epithelial sodium channel
gnomAD: Genome Aggregation Database (gnomad.broadinstitute.org)
LD: Linkage disequilibrium
LoF: Loss of function
NIHR BR-RD PID: NIHR BioResource–Rare Diseases PID cohort
NIHR: National Institute for Health Research
PID: Primary immunodeficiency

for known CF variants in PID patients with and without structural lung disease and assessed global allele frequencies; we screened for additional biallelic variants that have been associated with mild disease; and we assessed clinical features for confirmed cases. Our analysis also included related genes—*SCNN1A, SCNN1B*, and *SCNN1G*—that encode the alpha, beta, and gamma subunits of an epithelial sodium channel complex (ENaC). Like CFTR, these epithelial cell surface sodium transport channels are found in many tissues, including kidneys, lungs, and sweat glands. Damaging variants causing residual function lead to a phenotype that has an overlap with CF, including bronchiectasis with or without elevated sweat chloride.

We were also interested to know if subclinical/undiagnosed antibody deficiency is associated with worse outcomes in patients with CF. For this purpose, we reinvestigated 574 clinically validated CF cases in our single-center local cohort by assessing complete immunoglobulin profile and long-term outcomes, such as survival and need for a lung transplant.

METHODS

Ethics

The study was performed in accordance with the Declaration of Helsinki. The NIHR BioResource projects were approved by research ethics committees in the United Kingdom.

Whole genome sequencing

As part of the NIHR BR-RD study, 1046 PID patients and relatives had their genomes sequenced, as described elsewhere.[4,5] Briefly, paired-end whole genome sequencing was performed by Illumina's HiSeq X Ten system (Illumina, San Diego, Calif); 95% of bases were covered by at least 15 reads. Substitutions and indels up to 50 bp were called by Isaac, then merged using the AGG3 tool, while structural variants were called by Manta and Canvas (all software by Illumina). Only variants with an overall pass rate of >80% were considered for further analysis. Additionally, structural variants were analyzed for gene deletions, but none was identified.

Sanger sequencing

Genomic DNA was prepared from peripheral blood by spin column purification (QIAamp DNA Blood Mini Kit; Qiagen, Hilden, Germany). Amplification of genomic DNA for Sanger sequencing was performed by the standard PCR method. PCR cleanup was performed by ExoSAP-IT (Affymetrix, Santa Clara, Calif). Sanger sequencing used the BigDye Terminator Cycle Sequencing Kit 3.1 (Applied Biosystems; Thermo Fisher Scientific, Waltham, Mass) and analysis on an ABI 3130XL DNA analyzer (Applied Biosystems).

Variant filtration

PID cohorts were assessed for the regions of GRCh37 as follows: *CFTR,* 7:117118017-117310718; *SCNN1A,* 12:6454009-6488523; *SCNN1B,* 16:23311591-23394620; and *SCNN1G,* 16:23192040-23230200 (approximately 2 kb of flanking sequences were included for analysis). Filtering and predicting functional consequences were performed by Variant Effect Predictor (VEP),[6] Exome Variant Server,[7] the Single Nucleotide Polymorphism database,[8] ClinVar,[9] the Exome Aggregation Consortium, and the Genome Aggregation Database.[10] Filtering of common variations and annotation was performed using 'vcfhacks'[11] and in-house scripts. Candidate variants were required to pass the following filtering conditions: frequency (count/coverage) between 20% and 100%. According to VEP annotation, at least 1 canonical transcript is affected with 1 of the following consequences: variants of the coding sequence, frameshift, missense, protein altering, splice acceptor, splice donor, or splice region; an in-frame insertion or deletion; a start lost, stop gained, or stop retained, or (according to VEP) an ExAC frequency of unknown, of 0.01 or less, or with clinical significance flagged "pathogenic."

Data preparation for analysis

Candidate variants were identified only in *CFTR* requiring further statistical analysis but not in *SCNN1A, SCNN1B,* or *SCNN1G.* The NIHR BR-RD PID cohort data were prepared by conversion of VCF to a data frame with annotation headers matching reference population data from the Genome Aggregation Database (gnomAD; gnomad.broadinstitute.org) and the CFTR2 database (cftr2.org/). gnomAD (version r2.0.2) was queried for the canonical transcripts of *CFTR* from population genetics data of approximately 146,000 individuals as well as ENSG00000001626, GRCh37 7:117118017-117310718 (including flanking sequence padding). Data were filtered to contain the variant effect identifiers: frameshift, in-frame deletion, in-frame insertion, missense, stop lost, or stop gained. Reference transcripts were sourced from Ensembl in the FASTA-format amino acid sequence for transcript *CFTR* ENST00000003084 [HGNC:1884]. Data were exported to CSV format and imported into R software (R Project; www.r-project.org) for combined analysis with cohort variant data. A CFTR2 database variant list (current version during data collection 08Dec2017) was prepared in the same format as the gnomAD and NIHR BR-RD PID cohort data. For all 3 data sets, cohort-specific allele frequencies were calculated.

Statistical analysis

To test for the frequency of variants within cohorts and compared to the general population, we used routine genomic analysis and filtering to produce a list of candidate variants. A total of 3160 variants that met quality-control criteria were identified in *CFTR.* Because CF-causing variants are well characterized, we used 374 known variants from the CFTR2 database to filter qualifying variants. This database also lists variants of unknown consequence. We found 118 qualifying variants in our cohort. Qualifying variants were required to be present in index cases from our cohort and present in those with lung disease. Seven variants were tested for enrichment in PID cases for those with and without lung disease. These were also compared to the allele frequency in the general population. The Fisher exact test with multiple test correction was used. Although at least 2 variants were in linkage disequilibrium (LD) ($D' = 1.0$, $R^2 = 0.4975$), all variants were tested individually, and Bonferroni correction was applied for all to reduce false-positive results. The significant threshold after multiple testing correction was (.05/7) = .007. *CFTR* genotype (homozygous, heterozygous, or other, coded as 2, 1, and 0, respectively) and immunoglobulin responses were assessed using a least squares regression with weights for repeated measures and Bonferroni correction for multiple testing.

RESULTS

Gene structure and variant frequency in *CFTR*

The genomic structure of *CFTR* coding exons was mapped to functional domains on cDNA in Fig E1, *A* (available in this article's

Online Repository available at www.jacionline.org) on the basis of canonical ENST00000003084.11, 6070 nt, 1480 aa, and 27 exons. CF-related *CFTR* variants are listed in the CFTR2 database.[12] By comparing CF-related variants to the normal background allele frequency within in the population, we ranked potential candidate variants by importance for public health. The population allele frequencies of variants reported on gnomAD are shown (Fig E1, *B*), followed by allele frequencies of known damaging variants in the CFTR2 database. Combining these data sets reveals variants that are known determinants of disease and are present at the highest frequency in the population, thus constituting the largest public health concern. In Fig 1, *E,* and Fig E1, *C,* the CFTR protein is shown in its role acting as a transmembrane, phosphorylation, and ATP-gated anion channel that provides increased conductance for anions, including Cl^-, according to their electrochemical gradient. The "opening" and "closing" of CFTR conformational change provides a transmembrane flow of anions.[13] Although CF-specific variant enrichment is noted for some protein domains, damaging variants can present throughout the gene and can occur as several classes of mutation—for example, nonfunctional (nonsense, deletion, ie, p.Gly542Ter), expressed misfolded (p.Phe508del, p.Asn1303Lys), gate function variants (p.Gly551Asp), channel conduction variants (p.Arg117His), or expressed truncated or splice variant (c.2657+5G>A). The function of the ENaC complex, the subunits of which are encoded by *SCNN1A, SCNN1B*, and *SCNN1* (as illustrated in Fig 1, *E*), is selectively permeable to sodium ions (Na^+).

NIHR BR-RD PID cohort

We investigated the role of CFTR on PID in a large European cohort using statistical genomic analysis. We included the genome sequences of 1046 participants from the cohort of PID patients who had their genomes sequenced as part of the NIHR BR-RD PID study.[4,5] Within our cohort, 66.5% (n = 696) were enrolled as index patients with unknown PID, 7.7% (n = 81) were relatives of index patients who also had a PID, and 25.7% (n = 269) were nonindex relatives with no features of PID reported. We subgrouped the cohort first into index versus nonindex cases of PID, and second into those with and without evidence of lung disease (Fig 1, *A*). The subgroups were categorized as follows: 12.2% (n = 128) PID lung related, 54.3% (n = 568) PID non–lung related, 0.7% (n = 7) relatives' lung related, and 32.8 (n = 343) relatives' non–lung related. We specify the latter group (269 of 343 relatives who also have PID) to confirm that analysis does not mistakenly count in duplicate a variant that is shared among 2 siblings with the same phenotype. Our routine analysis pipeline, described elsewhere,[4,5] identified 3125 variants for the *CFTR* gene (GrCh37 7:117118017-117310718). Tailored analysis with the predefined criteria described in the Methods section permitted us to reduce this set to a short list of rare candidate variants. We identified 118 functional coding variants. We screened our results against a reliable source of known damaging (or candidate) variants using the CFTR2 database. This contained 374 variants that occur in 89,052 confirmed cases of CF. Of these, 35 variants were shared within the NIHR-PID cohort (Fig 1, *B*). According to our predefined criteria, candidate variants were restricted to those that were (1) reported in the CFTR2 database, (2) occurred in NIHR-PID index cases, and (3) occurred in is those with lung disease. Our within-cohort testing therefore consisted of 7 functional variants passing the qualifying inclusion criteria.

p.Phe508del carriage is enriched in lung-related PID

We performed a within-cohort association test for variants occurring in PID cases with and without lung disease. We found that only p.Phe508del carriage was significantly enriched in PID-related lung disease, with $\chi^2(1, N = 258) = 7.8994$, $P = .004945$; and a Bonferroni correction for multiple testing significance threshold of .007 ($P_{adj} = .035$) (Fig 1, *C*). The remaining 6 variants were also present at higher-than-expected frequencies based on the European and global allele frequencies reported in gnomAD. According to our analysis plan, these were subsequently tested against the background population frequency.

Pathogenic *CFTR* variant carriers are enriched for PID-related lung disorder

We tested variant carriage enrichment in PID-related lung disorders compared to background population frequencies. From the 7 functional variants found in both PID-related lung disease cases and the CFTR2 database, we identified 2 significantly enriched variants that are known to be pathogenic (as a recessive disorder), NM_000492.3(*CFTR*):c.1521_1523del (p.Phe508del) and NM_000492.4(*CFTR*):c.3276C>A (p.Tyr1092Ter), and 1 variant that is not reported as known pathogenic: NM_000492.4:c.3080T>C (p.Ile1027Thr) (Fig 1, *D*). Three additional variants that were significantly enriched in our cohort that did not meet criteria defining pathogenicity according to the CFTR2 database are NM_000492.4:c.91C>T (p.Arg31Cys), NM_000492.4:c.[1727G>C (p.Gly576Ala);2002C>T (p.Arg668Cys)], and NM_000492.4:c.3705T>G (p.Ser1235Arg). The latter 3 were significantly increased in general PID (PID lung disease and nonlung PID) (Fig 1, *C* and *D*). We note that p.Gly576Ala is correlated with p.Arg668Cys, and neither is likely to be inherited independently (LDlink[14]-measured degree of correlation is $D' = 1.0$, $R^2 = 0.4975$—compared, for example, to p.Ser1235Arg, with $D' = 1.0$, $R^2 = 0.0001$[14]). However, for accuracy, multiple test correction was applied for all variants.

Normalized pathogenic variant frequency

We were interested in the most common CF-causing variants (based on the CFTR2 database) compared to the global carriage frequency (gnomAD). Accounting for both population and disease frequency, the most clinically relevant variants were mapped according to their normalized frequency rank (see Fig E2, *A,* in the Online Repository available at www.jacionline.org). A small number of residues were evident as a top priority; variants at p.Ser18ArgfsX16, p.Phe508del, p.Gly542Ter, and p.Arg553Ter have normalized frequencies of >0.4, making them most likely to be seen clinically. A concentration of pathogenic variants is seen in nucleotide binding domains 1 and 2 (aka NBD1/2). There were 150 variants of normalized frequency 0-0.2 that are known to be pathogenic but will be seen less frequently in patients. The top 13 of 14 residues (p.Leu671 is not available on structure) were mapped onto the CFTR protein structure (RCSB Protein Data Bank, www.rcsb.org, entry 5UAK)[15] in Fig E2, *B*. The less commonly reported variants are colored *black* on the ribbon structure. Annotation was limited to variants with the highest normalized frequency after binning into 5 groups, as shown in Fig E2, *C*.

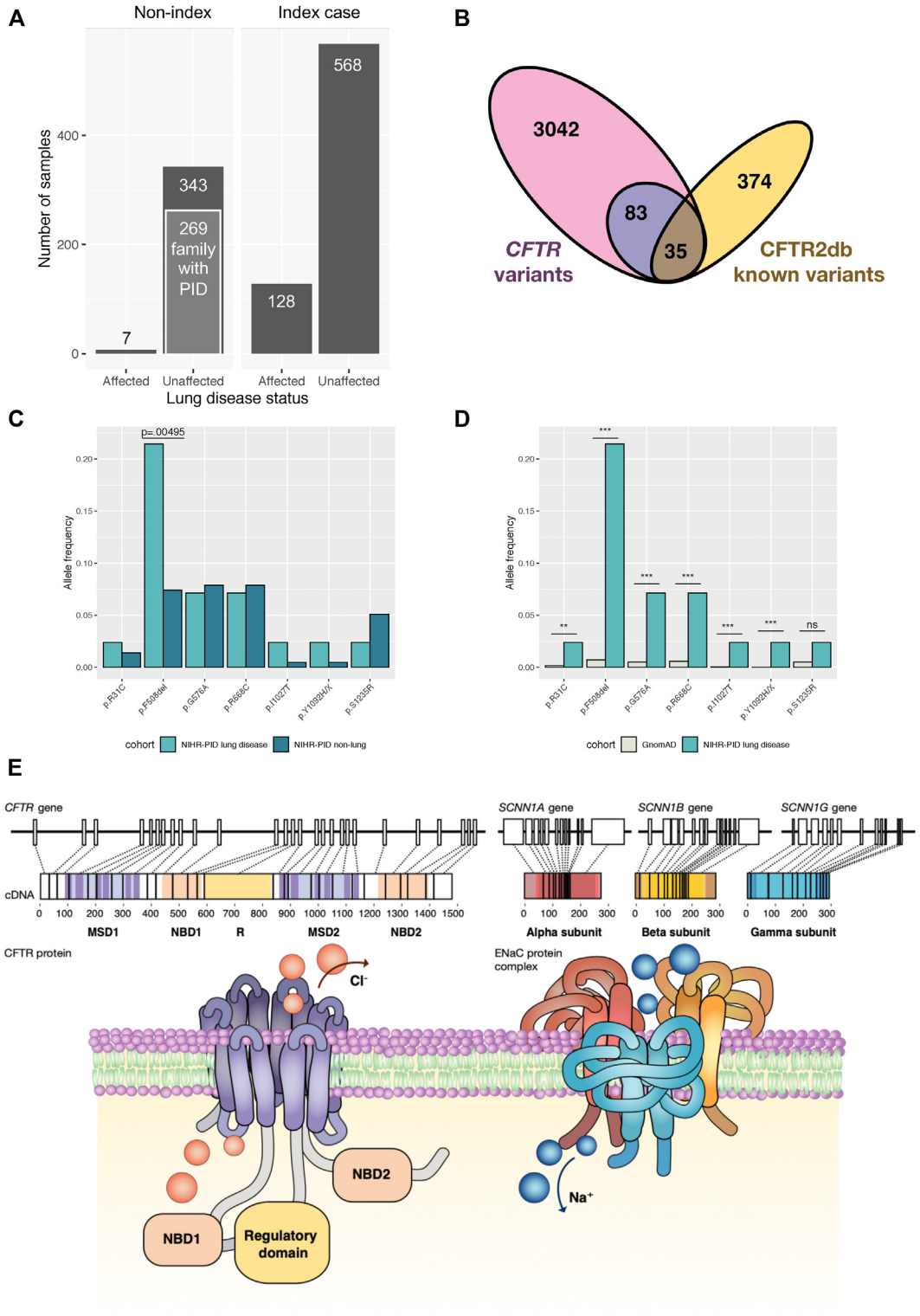

FIG 1. NIHR BR-RD PID candidate variants. **(A)** Number of PID patients reported as having lung-related disease status; index patients versus nonindex patients recruited either as healthy relative or with shared phenotype. Nonindex relatives unaffected by lung disease are illustrated as those with (n = 269) and without (n = 74) PID. **(B)** Routine analysis of variants produced 3160 total *CFTR* variants within the NIHR BR-RD PID cohort *(pink).* One hundred eighteen functional variants passed our filtering criteria *(purple).* The CFTR2 database listed variants affecting 374 coding residues *(yellow),* 35 of which were present as candidate variants in our cohort after filtering. Seven of these 35 met our inclusion criteria for subsequent testing. **(C)** p.Phe508del was significantly enriched for PID patients with lung-related disease (P = .00495, P_{adj} = .035). **(D)** Enrichment for variants in lung-related PID versus background population. p.Gly576Ala

Newly identified compound heterozygous cases

During investigation of the NIHR BR-RD PID cohort, we identified several potential cases of compound heterozygous pathogenic variants. Three sodium channel complex genes of ENaC were included in our screening as phenocopies of *CFTR* because they may produce an overlapping phenotype. A total of 6227 functional variants were identified in the cohort for the 4 screened genes: *SCNN1A, SCNN1B, SCNN1G,* and *CFTR*. For patients with more than 1 variant, we estimated the probability of *cis* and *trans* inheritance; we followed up to confirm and assess clinical relevance. Potentially pathogenic compound heterozygous variants were filtered if they were predicted to be nonbenign by at least 1 annotation tool including SIFT, PolyPhen, or Condel and if they occurred in combination with a second variant, resulting in a candidate list of variants: 25 in *SCNN1A,* 6 in *SCNN1B,* none in *SCNN1G,* and 52 in *CFTR*.

For all *SCNN1A, SCNN1B,* and *SCNN1G, trans*-inherited variant pairs were only found when including 1 common within-cohort variant that was not considered damaging for either gene, and therefore none of these was considered to be a cause of autosomal-recessive disease. Fifty-two candidate *CFTR* variants were filtered by allele frequency, LD, and functional annotation, and thereby reduced to a list of 5 candidate compound heterozygous variants from 5 families. After assessing pathogenicity potential based on functional annotation, population genetics, variant effect, and phasing estimate, the consulting physicians for each patient were contacted for detailed investigation of their history and follow-up testing. Two patients were excluded on the basis of these data, resulting in a final set of 3 newly identified biallelic cases of damaging *CFTR* variants from 3 families (Table I).

Several additional cases were ruled out as being CF-related disease after being thoroughly assessed for variants of unknown significance. Some patients who carried at least 1 known CF-causing variant were offered genetic consultation. A pair of variants within our cohort, p.Gly576Ala and p.Arg668Cys, occurs within a single haplotype block in LD that is not enriched in lung disease cases. However, we did find that this pair of variants occurred in combination with a third variant (ostensibly in *trans*). Table I lists the variants identified in our study along with supporting phenotype information; we also show the CFTR2 database labeling for each variant. Several variants are not considered CF causing and therefore would likely remain unidentified with a traditional, single-case approach. Table II summarizes the clinical features for each patient. More extensive information is reported in Table E1 in the Online Repository available at www.jacionline.org.

As shown in Table II, patient 1 had severe combined immunodeficiency with antibody deficiency, T-cell lymphopenia, recurrent infections, and lymphoma, and treatment included chemotherapy, radiotherapy, autologous immune enhancement therapy, and IgG replacement. Patient 1 had the variants NM_000492.4(*CFTR*):c.[91C>T (p.Arg31Cys)];[1521_1523del (p.Phe508del)]. Patient 2 had severe autoimmunity syndrome as well as recurrent infections including cytomegalovirus and *Candida* infection, and was treated with prednisolone and hydrochloroquine. Patient 2 had the variants NM_000492.4(CFTR):c.[220C>T (p.Arg74Trp);3808G>A (p.Asp1270Asn)];[293A>G (p.Gln98 Arg)]. Patient 3 had unclassified antibody deficiency (mild panhypogammaglobulinemia) and recurrent infections. Patient 3 had the variants NM_000492.4(*CFTR*):c.[3705T>G (p.Ser1235Arg)];[2900T>C (p.Leu967Ser)]. Presentation occurred at ages 13 (female), 12 (male), and 0 (male) for patients 1, 2, and 3, respectively.

Immunoglobulin levels in clinically confirmed CF

To investigate if lower serum immunoglobulin levels are associated with worse clinical outcomes in CF, we analyzed a cohort of 574 clinically and genetically confirmed cases of CF from a single center. We retrospectively collected serum immunoglobulin level data, including IgG, IgM, and IgA. The serum levels for all immunoglobulin classes were typically within reference ranges (Fig 2, *A*). High immunoglobulin levels observed during individual clinic visits typically returned to within reference ranges by the time of the patient's next visit.

We tested outcomes for patients with reduced immunoglobulin levels (IgM ≤ 0.5 and IgA < 0.8). Adult and pediatric population immunoglobulin levels were available for 574 subjects (381 and 193 subjects with normal and low immunoglobulin, respectively). There were 58 non–transplant-related deaths and 60 transplant-related deaths with normal IgA/IgM compared to 16 nontransplantation deaths and 9 transplant deaths with low IgA/IgM. Patents found to have a reduction in at least 1 immunoglobulin class did not have worse outcome when it came to survival with or without lung transplant [$\chi^2(1, N = 143) = 1.8$, $P = .2$]. We note that fewer deaths occurred for those with low immunoglobulin; however, this ought not suggest that low immunoglobulin is protective.

DISCUSSION

Newborn screening for CF is generally successful in high-prevalence areas, although the specific protocols may differ between countries.[16,17] An approximate measure of newborn screening across Europe is shown in Fig E2. Diagnosis usually follows newborn screening or occurs during early life; late diagnosis more often involves less damaging variants (eg, p.Arg117His).[18] A CF diagnosis can be designated as CFTR-related metabolic syndrome or CF screening positive, inconclusive diagnosis (CFSPID).[19] This definition may be applied for positive newborn screening results for CF and includes either (1) sweat chloride <30 mmol/L when at least 1 of 2 *CFTR* variants does not cause any physical symptoms; or (2) intermediate sweat chloride (30-59 mmol/L) with 1 or no CF-causing variants. As a CF-related disorder, *CFTR*-related metabolic syndrome or CFSPID may be an important factor that is overlooked when a more prominent determinant of PID is found.[20] Similarly, carrier screening does not routinely identify carrier status, which itself may be a gene

is strongly correlated with p.Arg668Cys (LDlink: $D' = 1.0$, $R^2 = 0.4975$). $^{**}P = .000348$ ($P_{adj} = .0024$), $^{***}P \leq .00001$ ($P_{adj} < .0001$); *ns,* does not pass significance threshold after multiple test correction, $P > .007$ ($P_{adj} > .05$). **(E)** Gene and protein structures. *Left,* CFTR protein acting as a transmembrane channel where it moves chloride ions (Cl^-) to the outside of the cell. *Right, SCNN1A, SCNN1B,* and *SCNN1G* encode subunits of ENaC, which is selectively permeable to sodium ions (Na^+). *Colors* match those on cDNA structure map. *MSD1/2,* Membrane spanning domain; *NBD1/2,* nucleotide binding domain; *R,* regulatory domain.

TABLE I. Additional *CFTR* variants in NIHR BR-RD PID cohort

Characteristic	Patient no. 1		2				3		
Family no.	1		2				3		
Member	Proband		Proband			Mother	Proband		Mother
Phenotype	Combined immunodeficiency		Autoimmunity			NA	Unclassified antibody deficiency		NA
Sex	F		M			F	M		F
Variant	A	B	A	B	C	C	A	B	B
Variant effect	Missense	In-frame deletion	Missense	Missense	Missense	Missense	Missense	Missense	Missense
cDNA	c.91C>T	c.1521_1523delCTT	c.220C>T	c.3808G>A	c.293A>G	c.293A>G	c.3705T>G	c.2900T>C	c.2900T>C
Amino acid	p.R31C	p.IF508I	p.R74W	p.D1270N	p.Q98R	p.Q98R	p.S1235R	p.L967S	p.L967S
Variant phase	A1*	A2*	A1†	A1†	A2‡	A2‡	A1§	A2§	A2
SNP ID	rs1800073	rs113993960	rs115545701	rs11971167	rs397508464	rs397508464	rs34911792	rs1800110	rs1800110
GRCh37 position	7:117144344	7:117199644	7:117149143	7:117282582	7:117170972	7:117170972	7:117267812	7:117243828	7:117243828
Exon	2/27	11/27	3/27	23/27	4/27	4/27	22/27	17/27	17/27
gnomAD									
Frequency	0.00164	0.00717	0.00172	0.00151	0.00001	0.00001	0.00504	0.00070	0.00070
Heterozygous count	463	2027	485	427	2	2	1409	199	199
Homozygous count	1	1	3	3	0	0	4	0	0
Allele no.	282310	282630	282362	282350	251054	251054	279632	282620	282620
CFTR2 database									
Allele count	23	99061	36	55	16	16	108	20	20
Allele frequency	0.00016	0.69744	0.00025	0.00039	0.00011	0.00011	0.00076	0.00014	0.00014
Pancreatic insufficient	30%	98%	15%	17%	33%	33%	38%	0	0
Final determination	Non–CF causing	CF causing	Varying clinical consequence	Varying clinical consequence	CF causing	CF causing	Non–CF causing	Varying clinical consequence	Varying clinical consequence

In addition to assessing the association between heterozygous carriage of known *CFTR* variants and lung disease, we screened the cohort for additional recessive disease. We identified 5 families with candidate *CFTR* variants. Currently variants of unknown significance cannot be assessed for their effect on lung function or treatment. Linkage was estimated before follow-up. *SNP*, Single nucleotide polymorphism.

*p.R31C and p.IF508I. LDlink: $D' = 1.0$, $R^2 = 0.0001$, $\chi^2 = 0.0207$, $P = .8855$. rs1800073 and rs199826652 are in linkage equilibrium. Estimated gnomAD probability that these variants occur in different haplotypes is 74%.

†p.R74W and p.D1270N. LDlink: p.Q98R (chr7:117170972) rs397508464 is not in 1000G reference panel. Different allele frequencies on gnomAD than p.R74W rs115545701 (chr7:117149143) and p.D1270N rs11971167 (chr7:117282582), estimated compound heterozygous in combination with maternal genotype evidence. Estimated gnomAD probability that these variants occur in different haplotypes is <1%.

‡[p.R74W/p.D1270N] and p.Q98R. LDlink: p.Q98R (chr7:117170972) rs397508464 is not in 1000G reference panel. Estimated gnomAD probability that these variants occur in different haplotypes is 100%.

§p.L967S and p.S1235R. LDlink: $D' = NA$, $R^2 = NA$, $\chi^2 = NA$, $P = NA$. rs1800110 and rs34911792 are in linkage equilibrium. Estimated gnomAD probability that these variants occur in different haplotypes is 83%.

modifier for diseases such as PID. In this cohort, we observed that p.Phe508del carriage was significantly enriched in PID lung disease compared to nonlung PID. Additionally, compared to the general population, 3 variants (p.Phe508del, p.Ile1027Thr, and p.Tyr1092His/Ter) were significantly increased in PID lung disease, and 3 variants (p.Arg31Cys, p.Gly576Ala-p.Arg668Cys, and p.Ser1235Arg) were increased in general PID (PID lung disease and nonlung PID). For the latter, these variants were significantly enriched with general PID but not exclusively in lung disease and were more difficult to interpret than those exclusive to lung disease. We observed no significant difference in serologic markers between genotypes, although other contributing factors that we could not control for may preclude the clinical application of association. We also compiled a list the most common CF-causing variants, which should be prioritized for analysis.

Individuals with hypomorphic variants may present with CFSPID or as a mild phenotype in childhood, and may in time develop typical features of bronchiectasis, pancreatitis, or infertility.[21,22] The survival rate in late diagnosis indicates a higher prevalence of hypomorphic variants and a phenotype of reduced severity.[17,21,22] The presence of damaging heterozygous variants in *CFTR* are closely associated with a wide spectrum of disease, including disseminated bronchiectasis,[23] allergic bronchopulmonary aspergillosis,[24] colorectal carcinoma, congenital bilateral absence of the vas deferens,[25] and chronic pancreatitis.[26] Examples of increased heterozygous variant carriage have been reported in children with chronic rhinosinusitis[27] and lung infection.[28] This impact of heterozygous variants in otherwise healthy individuals remains unclear.

A number of population-based publications have reported carrier status risk using unsatisfactory statistical methods (ie, failing to account for multiple testing or cherry-picking false discovery rate values). In time, large, well-curated cohort studies using natural quasi-random events may permit us to study the effect of *CFTR* on infection and/or lung disease in a randomly sampled population.

Our study relied on a well-curated cohort of PID cases—which, despite being a best-case scenario for a rare diseases study,

TABLE II. Characteristics of NIHR BR-RD PID subjects with additional *CFTR* variants

	Patient no.		
Characteristic	**1**	**2**	**3**
Genotype	p.R31C, p.IF508I	p.R74W / p.D1270N, p.Q98R	p.L967S, p.S1235R
Sex	F	M	M
Age at presentation (years)	13	12	0
Current age	40	24	15
Phenotype	Combined immunodeficiency	Autoimmunity syndrome	Unclassified antibody deficiency
Severity	Severe	Severe	Mild
Antibody deficiency	Yes	Yes	Yes
CD4 lymphopenia	Yes	Yes	Unknown
Autoimmunity	No	Yes	No
Malignancy	Yes	No	No
Pediatric onset	Yes	Yes	Yes
Immunoglobulin count	Low IgE	Low IgA	Panhypogammaglobulinemia
Cell count	Low CD4, high CD8, low B	Low CD4, low B	Unknown
Treatment			
IgG	Yes	No	Yes
Other	Chemotherapy, radiotherapy, ASCT, IgG	Prednisolone, hydrochloroquine	NA
Abpro	Yes	No	No
HSCT	No	No	No
GF	No	Yes	No
Recurrent infection	Yes	Yes	Yes
Bacteria	Yes	No	Yes
Fungus	No	Yes	No
Parasite	No	No	No
Virus	Yes	Yes	Yes
Infectious species	Unknown	CMV, *Candida*	Unknown
Lymphoma	Yes	No	No

A range of additional features of autoimmune, tissue abnormality, and neoplasm were negative for all 3 patients. Clinical features were reported by the original contributing physician. An extended reported is included in Table E1. *Abpro*, Antibody production; *ASCT,* autologous stem cell transplantation; *CMV,* cytomegalovirus; *GF*, graft failure; *HSCT,* hematopoietic stem cell transplantation.

demonstrates the difficulty in accumulating sufficient numbers of cases for statistical analysis. Gene candidate studies are by definition prone to detect false-positive results as a result of a preenrichment of genes or candidate variants. We report our findings with optimistic caution that further validation will be reported from other cohorts. The adoption of national-scale genomic services will in time increase the number of individuals with shared phenotypes and will permit novel discoveries that are otherwise not possible to detect.

The population genetics measure of LoF observed/expected upper bound fraction (aka LOEUF) for *CFTR* is 1.09 (0.91-1.31) —that is, more observed loss of function (LoF) than expected. Indeed, with 153 different LoF variants from 785 individuals, 0.64% of the global population sample (gnomAD) are heterozygous carriers:

$$\frac{\text{Heterozygous LoF alleles}}{\text{Number of individuals}} \times 100 = \frac{785 \text{ het LoF}}{\left(\frac{244786 \text{ alleles}}{2}\right)} \times 100 = 0.64\%.$$

Despite the relatively high frequency of LoF variants in gnomAD, no homozygous LoF variants are reported that would certainly manifest with CF and not be included in this database. While the distribution of variants changes according to population, the frequency of LoF variants is high enough to be considered an important factor for disease modulation in those already at risk of infection. Several interesting articles describe the potential selective advantage of heterozygous carriage of damaging variants in *CFTR*.[29-31]

The association between heterozygous carriage and lung disease in PID may reflect subtle changes in the host immune and inflammatory response, which have been described in CF and which may accentuate progression of lung disease. Recognizing pathogenic heterozygous *CFTR* variants as part of a PID diagnosis may help us identify individuals who would benefit from proactive, prophylactic clinical interventions to reduce disease progression. The effect of heterozygous *CFTR* variants in otherwise healthy individuals is usually mild or is not observed. Large cohort sizes are required to quantify an effect from heterozygous LoF in *CFTR* in addition to underlying PID. Untangling a compounded pathology is currently not possible, but with the increase of large population studies, a statistical genomics approach may elucidate the risk of lung disease introduced by a *CFTR* covariate in PID or other related genetic conditions. Other cohorts of genome sequencing in PID exist that may provide validation for our initial report; we estimate that at least >500 cases in a cohort with 1-2% heterozygous carriage would be required for sufficient power. We hope our report can contribute to improving preparedness for diagnosis of complex PID.

We gratefully acknowledge the participation of all NIHR BioResource volunteers and thank the NIHR BioResource centers and staff for their contribution. We thank the NIHR and NHS Blood and Transplant.

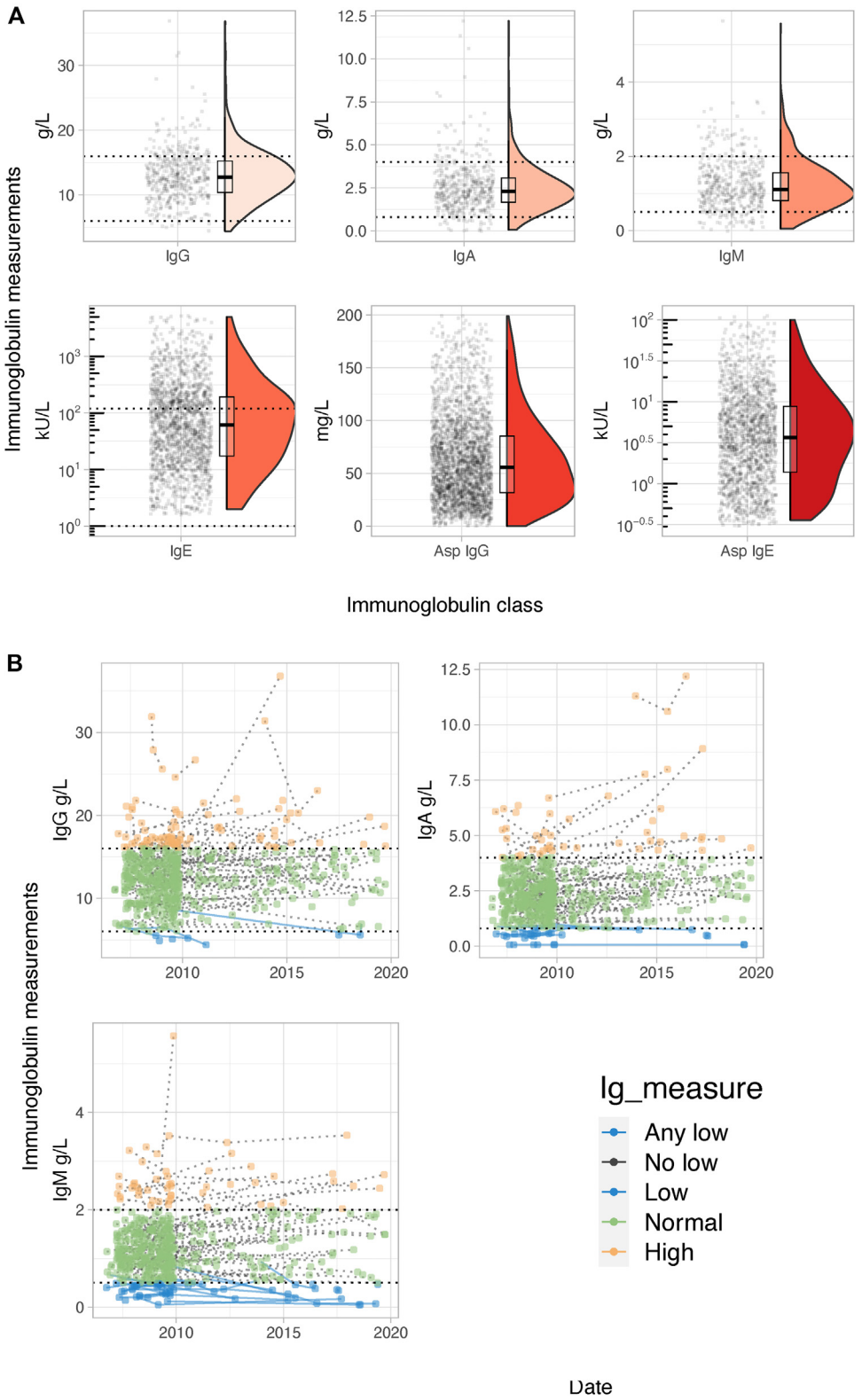

FIG 2. Serum immunoglobulins measurements in 574 confirmed CF cases. **(A)** Serum immunoglobulin levels for IgG, IgM, and IgA were retrospectively analyzed for all patients. *Dotted line* indicates reference ranges. **(B)** Immunoglobulin levels were assessed over time, as shown by a single *point* for each measurement and *dotted lines* joining each individual. Values outside the reference range are indicated by *color*. We subsequently tested for association between repeatedly low immunoglobulin and clinical outcomes. Patients with low immunoglobulin that otherwise returned to within reference ranges are additionally labeled as "any low." *Asp, Aspergillus.*

Clinical implication: *CFTR* genetic status should be included in diagnostic assessment of patients with PID because carrier status is associated with increased risk of developing structural lung damage.

REFERENCES

1. Riordan JR, Rommens JM, Kerem B, sheva, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989;245(4922):1066-73.
2. De Boeck K, Zolin A, Cuppens H, Olesen HV, Viviani L. The relative frequency of *CFTR* mutation classes in European patients with cystic fibrosis. J Cyst Fibros 2014;13:403-9.
3. Çolak Y, Nordestgaard BG, Afzal S. Morbidity and mortality in carriers of the cystic fibrosis mutation *CFTR* Phe508del in the general population. Eur Respir J 2020.
4. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. Nature 2020;583(7814):96-102.
5. Thaventhiran JE, Allen HL, Burren OS, Rae W, Greene D, Staples E, et al. Whole-genome sequencing of a sporadic primary immunodeficiency cohort. Nature 2020.
6. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol 2016;17:1-14.
7. National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP). Exome Variant Server. Available at: https://evs.gs.washington.edu/EVS/. Accessed December 8, 2017.
8. Kitts A, Phan L, Ward M, Holmes JB. The database of short genetic variation (dbSNP). Last updated April 3, 2014. In: The NCBI handbook. 2nd edition. Bethesda (Md): National Center for Biotechnology Information; 2013-. Available at: https://www.ncbi.nlm.nih.gov/books/NBK143764/.
9. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44(D1):D862-8.
10. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285.
11. Parry DA. VCFhacks v0.2.1 (command-line VCF programs programs for VCF filtering and manipulation). GitHub. Available at: https://github.com/david-a-parry/vcfhacks/releases, (8 December 2017).
12. Cutting G, Castellani C, Corey M, Lewis M, Penland C, Raraigh K, et al. Clinical and functional translation of CFTR (CFTR2). CFTR2. 2013. Available at: https://cftr2.org/.
13. Csanády L, Vergani P, Gadsby DC. Structure, gating, and regulation of the CFTR anion channel. Physiol Rev 2019;99:707-38.
14. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31(21):3555-7.
15. Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular structure of the human CFTR ion channel. Cell 2017;169:85-95.
16. Mayell S, Munck A, Craig J, Sermet I, Brownlee K, Schwarz M, et al. A European consensus for the evaluation and management of infants with an equivocal diagnosis following newborn screening for cystic fibrosis. J Cyst Fibros 2009;8:71-8.
17. Elborn JS. Cystic fibrosis. Lancet 2016;388(10059):2519-31.
18. McCloskey M, Redmond A, Hill A, Elborn J. Clinical features associated with a delayed diagnosis of cystic fibrosis. Respiration 2000;67:402-7.
19. Munck A, Mayell S, Winters V, Shawcross A, Derichs N, Parad R, et al. Cystic fibrosis screen positive, inconclusive diagnosis (CFSPID): a new designation and management recommendations for infants with an inconclusive diagnosis following newborn screening. J Cyst Fibros 2015;14:706-13.
20. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, et al. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies expert committee. J Clin Immunol 2020;40:24-64.
21. McKone EF, Velentgas P, Swenson AJ, Goss CH. Association of sweat chloride concentration at time of diagnosis and *CFTR* genotype with mortality and cystic fibrosis phenotype. J Cyst Fibros 2015;14:580-6.
22. Nick JA, Chacon CS, Brayshaw SJ, Jones MC, Barboa CM, Clair CGS, et al. Effects of gender and age at diagnosis on disease progression in long-term survivors of cystic fibrosis. Am J Respir Crit Care Med 2010;182:614-26.
23. Pignatti PF, Bombieri C, Marigo C, Benetazzo M, Luisetti M. Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. Hum Mol Genet 1995;4:635-9.
24. Miller PW, Hamosh A, Macek M Jr, Greenberger PA, MacLean J, Walden SM, et al. Cystic fibrosis transmembrane conductance regulator *(CFTR)* gene mutations in allergic bronchopulmonary aspergillosis. Am J Hum Genet 1996;59:45.
25. Chillón M, Casals T, Mercier B, Bassas L, Lissens W, Silber S, et al. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. N Engl J Med 1995;332:1475-80.
26. Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, et al. Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. N Engl J Med 1998;339:645-52.
27. Raman V, Clary R, Siegrist KL, Zehnbauer B, Chatila TA. Increased prevalence of mutations in the cystic fibrosis transmembrane conductance regulator in children with chronic rhinosinusitis. Pediatrics 2002;109:e13.
28. Casals T, De-Gracia J, Gallego M, Dorca J, Rodriguez-Sanchon B, Ramos M, et al. Bronchiectasis in adult patients: an expression of heterozygosity for *CFTR* gene mutations? Clin Genet 2004;65:490-5.
29. Gupta MK, Vadde R. Genetic basis of adaptation and maladaptation via balancing selection. Zoology 2019;136:125693.
30. Abramovs N, Brass A, Tassabehji M. Hardy-Weinberg equilibrium in the large scale genomic sequencing era. Front Genet 2020;11:210.
31. Bernut A, Dupont C, Ogryzko NV, Neyret A, Herrmann JL, Floto RA, et al. *CFTR* protects against *Mycobacterium abscessus* infection by fine-tuning host oxidative defenses. Cell Rep 2019;26:1828-40.