npg

ARTICLE

Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis

Franziska M Gisler[1], Thomas von Kanel[1], Richard Kraemer[2], André Schaller[1] and Sabina Gallati*,[1]

There is growing evidence that the great phenotypic variability in patients with cystic fibrosis (CF) not only depends on the genotype, but apart from a combination of environmental and stochastic factors predominantly also on modifier gene effects. It has been proposed that genes interacting with CF transmembrane conductance regulator (CFTR) and epithelial sodium channel (ENaC) are potential modifiers. Therefore, we assessed the impact of single-nucleotide polymorphisms (SNPs) of several of these interacters on CF disease outcome. SNPs that potentially alter gene function were genotyped in 95 well-characterized p.Phe508del homozygous CF patients. Linear mixed-effect model analysis was used to assess the relationship between sequence variants and the repeated measurements of lung function parameters. In total, we genotyped 72 SNPs in 10 genes. Twenty-five SNPs were used for statistical analysis, where we found strong associations for one SNP in *PPP2R4* with the lung clearance index ($P \leq 0.01$), the specific effective airway resistance ($P \leq 0.005$) and the forced expiratory volume in 1 s ($P \leq 0.005$). In addition, we identified one SNP in *SNAP23* to be significantly associated with three lung function parameters as well as one SNP in *PPP2R1A* and three in *KRT19* to show a significant influence on one lung function parameter each. Our findings indicate that direct interacters with CFTR, such as SNAP23, PPP2R4 and PPP2R1A, may modify the residual function of p.Phe508del-CFTR while variants in KRT19 may modulate the amount of p.Phe508del-CFTR at the apical membrane and consequently modify CF disease.

INTRODUCTION

Cystic fibrosis (CF), the most common lethal autosomal recessive disease, is caused by mutations in the CF transmembrane conductance regulator (CFTR) gene.[1,2] So far, more than 1900 mutations have been reported to the Cystic Fibrosis Genetic Analysis Consortium (www.genet.sickkids.on.ca/cftr). The most common mutation is p.Phe508del leading to an abnormally folded protein.

CFTR encodes a cyclic adenosine monophosphate-regulated chloride channel at the apical membrane of epithelial cells.[3] Clinically, CF is characterized by elevated sweat chloride concentration, exocrine pancreatic insufficiency, male infertility and progressive obstructive lung disease, the latter being the primary cause of mortality.[4,5] The respiratory disease progression in CF is expressed by various pulmonary dysfunctions,[6–10] such as ventilation inhomogeneities,[8,11] pulmonary hyperinflation,[9,12] bronchial obstruction, trapped gas and gas exchange disturbances,[10] most of them occurring early in life,[7,12,13] and progressing even in the absence of clinical signs and symptoms.[11,14,15] Owing to the great phenotypic variability in lung disease even observed among patients carrying the same CFTR genotype, several association studies have been conducted to find modifying genetic factors and to study their influence on CF disease outcome.[16] So far, five genes showed an association in at least two independent populations with more than 500 participants in total, namely *MBL2*, *IFRD1*, *IL8* and *TGFB1* that are involved in the

immune response and *EDNRA* that might influence CF lung disease by altering smooth muscle tone in the airways and/or vasculature.[17–21] *IFRD1* was initially found in a genome-wide association study (GWAS) comprising 160 severe and 160 mild CF patients from the GMS (extremes of phenotype) cohort and replicated in the TSS cohort (family-based CF Twin and Sibling study cohort). Interestingly, *IFRD1* did not achieve genome-wide significance in another GWA and linkage study including 1978 individuals from the GMS and CGS (Canadian Consortium for Genetic Studies) cohort. However, the same study identified two loci at 11p13 and 20q13.2 modifying disease severity, and these findings were replicated in a separate family-based study in the TSS cohort ($n = 557$).[22] This example shows that GWAS in huge populations like the North American Cystic Fibrosis Gene Modifier Consortium, comprising the GMS, CGS and TSS cohorts, is a powerful tool, but that there is the danger of false-negative results due to the stringent corrections for multiple comparison in the statistical analysis and the heterogeneity of the cohorts. Therefore, being aware of the fact that large and small cohorts both have their pros and cons, and having a small but well-defined cohort at our disposal, we chose a candidate gene approach. Keeping in mind the danger of false-positive results, we evaluated the influence on lung disease of variants in 10 genes of the CF interactome, coding for proteins that either directly interact with CFTR or the epithelial sodium channel (ENaC), a counterplayer of

[1]Division of Human Genetics, Departments of Pediatrics and Clinical Research, University of Bern, Bern, Switzerland; [2]Division of Pediatric Respiratory Medicine, Department of Pediatrics, Inselspital, University of Bern, Bern, Switzerland
*Correspondence: Professor S Gallati, Division of Human Genetics, University Children's Hospital, Inselspital, Freiburgstrasse, CH-3010 Bern, Switzerland. Tel: +41 31 632 94 94; Fax: +41 31 632 94 84; E-mail: sabina.gallati@insel.ch

CFTR, or are involved in CFTR trafficking. The genes interacting with CFTR are *SNAP23* (synaptosomal-associated protein, 23 kDa (Annexin A5) and *PPP2R4, PPP2R1A, PPP2R5E*, that encode three regulatory subunits of protein phosphatase 2A.[23–29] The two genes interacting with EnaC are *Nedd4L* (neural precursor cell expressed, developmentally downregulated 4-like) and *PRSS8* (prostasin).[30–32]

The last group consists of three genes that are involved in CFTR trafficking, *AHSAI* (activator of heat shock 90-kDa protein ATPase homolog 1), *CALR* (calreticulin) and *KRT19* (cytokeratin 19).[33–36]

Most candidate gene and whole-genome association studies so far focused on time point measurements of only one or two lung function parameters. To get a better insight on the influence of our single-nucleotide polymorphisms (SNPs) on CF lung disease, we used a new approach and correlated our genotype data with a wider range of lung function parameters measured over a time period to evaluate the influence of a SNP on the progression of the lung disease. We were not only interested in lung function parameters quantifying the degree of bronchial obstruction and hence airway patency, but also in parameters representing pulmonary hyperinflation, ventilation inhomogeneities, and reduction of forced expiratory flow and volume characteristics. In addition, we decided to investigate the influence of variants on longitudinal data of lung function parameters to get an idea of the progression of the lung disease over time. To achieve this, we used the clinically very carefully characterized patient cohort of the Children's University Hospital in Bern documented over a time period of 13 years.[8–10] Owing to the fact that these patients were treated at the same hospital, environmental influences in the context of medical treatment are reduced in this cohort, making it well suited to study-modifying effects.

MATERIALS AND METHODS

Patients
We studied 95 p.Phe508del homozygous patients (53 male and 42 female; birth years 1975–2001; onset of chronic *Pseudomonas aeruginosa* infection until age of three 37.8% and after age of three 62.2%) attending the Department of Pediatrics, Inselspital, Bern, Switzerland.[8] Patients included in this study fulfilled the following criteria: (1) being homozygous for p.Phe508del, (2) having characteristic phenotypic features including positive sweat test and documented onset of chronic *P. aeruginosa* infection, and (3) providing lung function data from age 6 to 20 years. To investigate the influence of modifier genes on different aspects of CF lung disease, six lung function parameters were evaluated: (1) functional residual capacity (FRC$_{pleth}$; degree of pulmonary hyperinflation), (2) lung clearance index (LCI; degree of ventilation inhomogeneities), (3) effective specific airway resistance (sR$_{eff}$; degree of bronchial obstruction), (4) forced expiratory volume in 1 s (FEV1), (5) forced expiratory flow at 50% forced vital capacity (FEF$_{50}$) and (6) volume of trapped gas (V_{TG}). All participants gave their full informed consent and the study was approved by the local ethics committee.

Lung function testing
Whole-body plethysmography and multibreath nitrogen washout (MBNW) techniques provided data pertaining FRC$_{pleth}$, FRC$_{MBNW}$, LCI and sR$_{eff}$, whereas FEV1 and FEF$_{50}$ were assessed by spirometry. Measurement techniques have been described in detail previously.[8–10,37] All values were expressed by z-transformation in SD scores (SDS), based on gender- and age-specific regression equations.[38,39] Details regarding measurement techniques and clinical evidence of these target parameters are given in the online data supplement.

Single-nucleotide polymorphisms
SNPs that potentially alter gene function were chosen from the NCBI Database according to the following criteria: (1) missense SNPs, (2) intronic SNPs in proximity of splice sites (-50, respectively, $+15$ bp), (3) synonymous SNPs

located in exonic splicing enhancer (ESE) and silencer (ESS) motifs and (4) SNPs in the 5′- and 3′-untranslated region (UTR). SNPs with a minor allele frequency <10% were excluded owing to the small genotype counts leading to groups that are too small for our statistical model. For the complete list of SNPs including corresponding primers see online data Supplement Table S1.

Genotyping
Genomic DNA was isolated from EDTA blood using the Qiamp DNA Blood Maxi Kit (Qiagen Gmbh, Hilden, Germany), according to the supplier's protocol. Genotyping was performed by high-resolution melting (HRM) using the LigthCycler High Resolution Melting Master (Roche Diagnostics Gmbh, Mannheim, Germany) on the LigthCycler 480 (Roche Diagnostics Gmbh), by single-strand conformation polymorphism/heteroduplex analysis according to Liechti-Gallati *et al*[40] or by restriction fragment length polymorphism analysis as indicated in the online data supplement. The accuracy of the genotyping was verified in around 10% of the PCR products by Sanger sequencing (see online data supplement).

Statistical analysis
Linear mixed model analysis was performed to assess the influence of variants and haplotypes in the candidate genes in a recessive, codominant and dominant model on longitudinal measures of each lung function parameter using the PROC MIXED procedure of the SAS 9.2 software (SAS Institute, Cary, NC, USA). In addition, we investigated the cumulative effects of the SNPs that showed a significant influence individually. The lung function parameter of interest was included in the calculations as a dependent variable, and the SNP genotype, age at lung function measurement, birth year, test year and gender as explaining variables. The genotype model with the lowest Akaike information criterion fitting best for our statistical model was chosen. *P*-values of all genotype comparisons were obtained with a contrast statement. We performed a Bonferroni correction for 10 genes and considered associations with $P \le 0.005$ to be statistically significant. In case of an association to one lung function parameter with $P \le 0.005$, we also included associations of this variant to other lung function parameters with $P \le 0.01$ in our evaluation.

Graphical displays of all significant associations were done using R 2.9.0 (www.r-project.org). The corresponding codes are listed in the online data supplement.

Haplotypes were computed using the Arlequin 3.5 software.[41]

In silico tools
To analyze the influence of significant variants on transcription and translation, we used different *in silico* tools according to the location of the variants in the gene. To assess the pathogenicity of the missense SNP in *KRT19*, we used PolyPhen-2 and SIFT.[42,43] For intronic and synonymous SNPs, we used the Human splicing finder tool with default threshold levels and the SpliceAid 2 tool for alveolar cells.[44,45]

RESULTS

Genotyping
In total, we genotyped 72 SNPs in 10 candidate genes. Genotype and allele frequencies of all variants were in Hardy–Weinberg equilibrium in our population.

Seventy SNPs are known to the NCBI Database and two SNPs were found during our study (c.-131C>T in *AnxA5* and c.48 + 24C>T in *Nedd4L*). Of these SNPs, 47 showed an allele frequency <10% and were therefore excluded. Twenty-five SNPs were included in the statistical analysis (Table 1). SNPs in linkage disequilibrium are displayed in Table 2.

Influence of variants on lung function
We found six variants that presented with significant associations with at least one lung function parameter (Table 3 and Figure 1).

In PPP2R4:c.-185A>C, lung disease progression over years of age was significantly stronger for AA in comparison with AC/CC best

Table 1 Variants showing a minor allele frequency ⩾10%

Gene	Location	rs number	a.a. change	N	AA	Aa	aa	p	q
SNAP23	c.*902T>C/c.*906T>C	rs11022/1060265		94	53	36	5	0.755	0.245
	c.267-9T>C	rs9302112		82	47	29	6	0.75	0.25
	5′ near gene g.42787144C>A	rs1900921		86	49	30	7	0.744	0.256
	c.396G>A	rs1060291	Thr(ACG)-Thr(ACA)	93	53	35	5	0.758	0.242
PPP2R4	c.*350T>G	rs746489		95	51	38	6	0.737	0.263
	c.-185A>C	rs3118625		88	49	34	5	0.75	0.25
PPP2R1A	c.-73T>C	rs13344984		94	67	25	2	0.846	0.154
	c.*465T>A	rs2162779		93	39	45	9	0.661	0.339
PPP2R5E	c.-32T>C	rs1255720		91	48	35	8	0.72	0.28
	c.-8+8A>C	rs1271563		91	45	36	10	0.692	0.308
AnxA5	c.-64G>T	rs1050606		91	19	54	18	0.505	0.495
	c.-184T>C	rs28651243		93	56	36	1	0.796	0.204
	c.-210A>C	rs28717001		93	56	36	1	0.796	0.204
	c.-131C>T			93	70	22	1	0.871	0.129
	c.9+18G>C	rs2306417		92	70	21	1	0.875	0.125
AHSAI	c.272-27C>T	rs12436593		92	26	42	24	0.511	0.489
	c.690+7A>C	rs7146595		93	26	43	24	0.511	0.489
	c.793-42A>G	rs4903584		92	26	42	24	0.511	0.489
	c.133C>T	rs1061629	Leu(CTG)-Leu(TTG)	93	26	43	24	0.511	0.489
	c.294G>A	rs1061638	Gln(CAA)-Gln(CAG)	93	38	43	12	0.64	0.36
	c.945G>A	rs7250	Thr(ACG)-Thr(ACA)	93	26	43	24	0.511	0.489
PRSS8	c.538+10G>A	rs45619735		92	44	41	7	0.701	0.299
KRT19	c.90T>C	rs11550883	Ala(GCC)-Ala(GCT)	90	34	41	15	0.606	0.394
	c.179G>C	rs4602	Gly(GGC)-Ala(GCC)	92	37	40	15	0.62	0.38
	c.471C>T	rs4601	Asn(AAT)-Asn(AAC)	91	36	40	15	0.615	0.385

Abbreviations: AA, wild-type homozygote; Aa, heterozygote; aa, mutant homozygote; N, number of individuals genotyped; rs number, reference SNP accession ID on http://www.ncbi.nlm.nih.gov/snp.

Table 2 Variants in linkage disequilibrium

Gene	SNPs	Lewontin's D′	r²
SNAP23	c.*902T>C/c.*906T>C with c.267-9T>C	0.97	0.89
	c.*902T>C/c.*906T>C with g.42787144C>A	0.97	0.92
	c.*902T>C/c.*906T>C with c.396G>A	1	1
KRT19	c.90T>C with c.179G>C	0.95	0.95
	c.179G>C with c.471C>T	1	1
AHSAI	c.272-27C>T with c.690+7A>C	1	1
	c.690+7A>C with c.793-42A>G	1	1
	c.272-27C>T with c.945G>A	1	1
	c.133C>T with c.945G>A	1	1
AnxA5	c.-184T>C with c.-210A>C	1	1
	c.-131C>T with c.9+18G>C	0.95	0.9

reflected by sR_{eff} and FEV1, the former a measure of bronchial obstruction of larger and smaller airways, the latter an expression of loss of elastic recoil (Figure 1a). Moreover, LCI as a measure of ventilation inhomogeneities, and hence early outcome parameter, was already increased in both subgroups at age 6 and the SDS of genotype AA significantly deteriorated with age.

Calculated in a codominant model SNAP23:c.267-9T>C in *SNAP23* showed significant associations with the lung function parameters FRC_{pleth}, FEF_{50} and V_{TG} (Figure 1b). Whereas SNAP23:c.267-9T>C CT presented with lowest progression in pulmonary hyperinflation, this subgroup already started with a z-score of 1.6 at age 6 years. Significantly higher progression of pulmonary hyperinflation over age was found for SNAP23:c.267-9T>C CC. This finding presented with

a mirror picture for FEF_{50} presenting a most pronounced decline for SNAP23:c.267-9T>C CC as estimate of the small airway dysfunction. For V_{TG} there is a highly significant difference in the slope of SNAP23:c.267-9T>C TT compared with CT, indicating that genotype TT results in faster V_{TG} progression.

Three SNPs in *KRT19*, KRT19:c.90T>C, KRT19:c.179G>C and KRT19:c.471C>T show association with sR_{eff} ($P \leqslant 0.01$, $P \leqslant 0.005$ and $P \leqslant 0.01$) in a codominant model. Owing to the fact that these three SNPs are in linkage disequilibrium (Table 2) only the graph for KRT19:c.179G>C (Figure 1c) is displayed, where genotype GG shows a significantly steeper slope and therefore more pronounced progression.

Finally, concerning FRC_{pleth} genotype AA in PPP2R1A:c.*465T>A presented with a much milder course than genotypes TT and AT when calculated in a codominant model (Figure 1d). It can be hypothesized that homozygosity for the mutant allele may induce a certain protection from the development of pulmonary hyperinflation of the lungs.

Haplotype analysis did not result in significant associations to lung disease except for the two main haplotypes TGC and CCT in *KRT19* (KRT19:c.90T>C, KRT19:c.179G>C and KRT19:c.471C>T), where homozygosity for haplotype TGC results in a significant steeper slope ($P \leqslant 0.01$) (represented by the graph of KRT19:c.179G>C (Figure 1c).

Predicted influence of variants on transcription and translation
For KRT19:c.179G>C, a missense SNP in exon 1 of *KRT19*, we used PolyPhen-2 and SIFT to evaluate its pathogenicity. PolyPhen-2 predicted this variant to be benign with a score of 0.000, and SIFT predicted it to be tolerated with a score of 1, therefore this SNP does

Table 3 Results of random coefficient models (only variants with significant associations are shown)

| | Intercept | | | Slope | | |
| | Codominant model | | | Codominant model | | |
	mut/mut vs wt/wt	wt/mut vs wt/wt	mut/mut vs wt/mut	mut/mut vs wt/wt	wt/mut vs wt/wt	mut/mut vs wt/mut
SNAP23:c.267-9T>C						
FEF_{50}	0.0426	0.1698	0.0088	0.2412	0.0548	0.0364
FRC_{pleth}	0.0408	0.6836	0.0287	0.081	0.1312	0.011
VTG	0.8418	0.0391	0.2972	0.2424	0.0043	0.5307
KRT19:c.90T>C						
sR_{eff}	0.0297	0.0322	0.8424	0.0797	0.0093	0.3596
KRT19:c.179G>C						
sR_{eff}	0.032	0.0189	0.8749	0.0602	0.0052	0.28
KRT19:c.471C>T						
sR_{eff}	0.049	0.0259	0.8186	0.0647	0.0085	0.3644
KRT19: haplotype						
sR_{eff}	0.0574	0.0236	0.7541	0.0994	0.0097	0.3024
*PPP2R1A:c.*465T>A*						
FRC_{pleth}	0.6628	0.3737	0.9109	0.0033	0.1414	0.0473

| | Intercept | Slope | | | | |
	Dominant model	Dominant model				
PPP2R4:c.-185A>C						
FEV1	0.0364	0.0048				
LCI	0.0933	0.0059				
sR_{eff}	0.0064	0.0029				

Values shown are *P*-values.

not seem to have any harming effect at all. For SNAP23:c.267-9T>C, a SNP in an intronic region, and for KRT19:c.90T>C, KRT19:c.179G>C and KRT19:c.471C>T, Human Splicing Finder and SpliceAid 2 were used to assess the influence of the variants on splicing. SNAP23:c.267-9T>C does not seem to influence splice sites directly but the consensus value (CV) of the T allele sequence compared with the C allele sequence is reduced by 23.8% for a branch point motif leading to a loss of the splice site with high probability. For the three KRT19 SNPs, we looked for changes in ESSs and ESEs with the SpliceAid 2 software. KRT19:c.90T>C T allele leads to a change in the recognized sequence of the protein HuB, a RNA-binding protein that has an important role in neuronal differentiation. SNP KRT19:c.179G>C does not lead to any changes at all in ESS or ESE. KRT19:c.471C>T is the most likely of the three SNPs to influence splicing, because the C allele creates a sequence that is recognized by YB-1, a positive splicing factor (for more details see Supplement Table S2 in the online data supplement).

DISCUSSION
Disease severity in CF varies greatly among patients, even when they carry the same CFTR genotype.[16] We hypothesized that genes interacting with CFTR and epithelial sodium channel ENaC might be potential modifiers of CF disease severity. Therefore, we assessed the impact of variants of these interacters on CF disease outcome. In contrary to most association studies performed so far, we used longitudinal lung function data of six lung function parameters. In total, we found one variant in *PPP2R4*, one variant in *PPP2R1A*, one

variant in *SNAP23* and three variants in *KRT19* that presented with significant associations to at least one lung function parameter.

The pathogenicity of p.Phe508del CFTR is due to its inability to fold properly in the endoplasmatic reticulum (ER), destining it for ubiquitination and degradation by the proteasome. Nevertheless, when reaching the apical membrane, it shows activity. Interestingly, there is still a slight remaining CFTR function in CF patients being homozygous for p.Phe508del, indicating that some CFTR proteins pass the ER quality control and reach the apical membrane.[46] We think that variants in direct interacters like PP2A and SNAP23 might enhance the residual function of CFTR while variants in KRT19, which are involved in trafficking of CFTR, might raise the amount of functional CFTR at the apical membrane and therefore modify CF disease.

PP2A and PP2C are the major phosphatases involved in CFTR deactivation. Phosphorylated CFTR is a substrate for PP2A dephosphorylation and it was shown that inhibition of PP2A activity delays the closure of the CFTR channel, while an overexpression of PP2A results in prolonged deactivation of the channel.[27,29] Thelin *et al*[28] demonstrated that inhibition of PP2A in well-differentiated human bronchial epithelial cells increases the airway surface liquid (ASL) in a CFTR-dependent manner. In our study, we looked at *PPP2R4*, the protein phosphatase 2A activator, regulatory subunit 4, of PP2A and in accordance with Thelin *et al* at the two regulatory subunits *PPP2R1A* and *PPP2R5E*. SNP PPP2R4:c.-185A>C genotype AA was found to be associated with a worse progression for the three lung function parameters FEV1, LCI and sR_{eff} (Figure 1a).
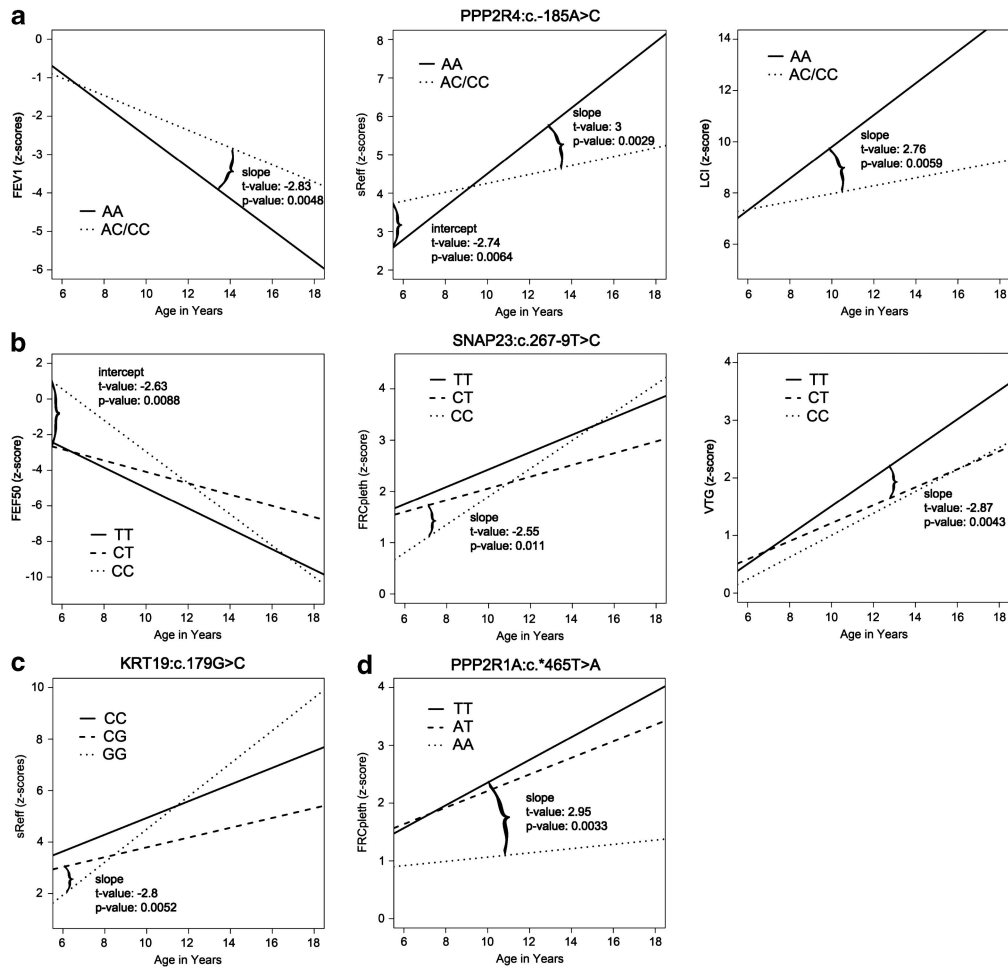
Figure 1 Results of random coefficient model: (**a**) For all three lung function parameters, genotype AA of PPP2R4:c.-185A>C has a significant different slope and overall a worse pulmonary outcome. (**b**) Genotype CT of SNAP23:c.267-9T>C shows a significant different intercept from genotype CC in FEF_{50} and a significant steeper slope in FRC_{pleth}. For VTG, genotype TT has a significantly steeper slope than genotype CT. (**c**) The slope of genotype GG is significantly different from genotype CG for KRT19:c.179G>C in association with sR_{eff}. (**d**) PPP2R1A:c.*465T>A genotype TT shows a significant steeper slope of the curve than genotype AA for FRC_{pleth}.

PPP2R4:c.-185A>C is a variant in the 5′ UTR of *PPP2R4*. We hypothesize that the C allele might lead to interferences with the translation machinery and therefore to reduced PP2A levels resulting in delayed closure of the CFTR channel and elevated ASL. Variant PPP2R1A:c.*465T>A is located in the 3′ UTR of *PPP2R1A* and therefore could affect the abundance of the mRNA by influencing the stability of the transcript or the site selection for polyadenadenylation. In addition, it is known that there are several mechanisms of translational control by the 3′ UTR.[47] Therefore it is expected that variants in the 3′ UTR might influence translation and therefore protein levels. Genotype AA in PPP2R1A:c.*465T>A shows a much better lung progression *vs* genotype TT for the lung function parameter FRC_{pleth} indicating that homozygosity for the mutant allele seems to interfere with translation leading to lower levels of PP2A and therefore to a better residual function of CFTR (Figure 1d).

SNAP23:c.267-9T>C, an intronic variant located −9 nucleotides from exon 6 in *SNAP23*, showed significant associations with the three lung function parameters FEF_{50}, FRC_{pleth} and V_{TG}. SNAP23 physically interacts with CFTR by binding to its amino-terminal tail and inhibits CFTR chloride currents by influencing channel gating.[23] Using the Human Splicing Finder tool, we found a change in a

potential branch point. The CV for the sequence with the C allele (0–100) is 80.91, whereas the CV for the sequence with the T allele is 57.07 leading to an abolished splice site. Therefore one could hypothesize that genotype T leads to an aberrant splicing and less functional protein resulting in a better clinical phenotype due to the loss of the inhibitory effect of SNAP23 on CFTR channel gating. According to this hypothesis, genotype TT shows a worse intercept for all three lung function parameters but slower progression for FEF_{50} and FRC_{pleth} than genotype CC. We hypothesize that in young years with relatively good lung function, the unrestricted interaction and regulation of SNAP23 and CFTR in CC genotype regulates the phenotype in a positive way. But with increasing progression of multifactorial lung disease, a decreased interaction might influence the residual CFTR function positively. Finally, heterozygous patients seem to have the best balance between full and reduced interaction resulting in least progression of lung disease (Figure 1b).

There are three variants KRT19:c.90T>C, KRT19:c.179G>C and KRT19:c.471C>T in *KRT19* that are in strong linkage disequilibrium resulting in the two main haplotypes TGC and CCT (Table 2). KRT19:c.90T>C is a synonymous SNP, KRT19:c.179G>C a missense SNP in exon 1 and KRT19:c.471C>T a synonymous SNP in exon 2.

According to PolyPhen-2 and SIFT, the missense SNP does not lead to pathogenic alterations in the protein, therefore we looked for changes in ESE and ESS motifs that might influence splicing of *KRT19* and therefore lead to changes in protein level. Especially the change from the C to T allele in SNP KRT19:c.471C>T leads to a loss of affinity for the positive splicing factor YB-1 and therefore probably to changes in exon inclusion/skipping as could be shown in the c.6792C>G mutation in the neurofibromatosis type 1 gene (Supplement Table S2).[48] KRT19 was shown to be a modulator of cellular differentiation processes at the apical membrane of epithelial cells.[35] Sun[36] (presented at the 23rd Annual North American Cystic Fibrosis Conference) illustrated that KRT19 affects CFTR function by stabilizing it at the plasma membrane and is able to increase its plasma membrane density by inhibiting CFTR endocytosis. It seems that homozygosity for haplotype TGC has a negative effect on protein levels and therefore worse progression of sR$_{eff}$ (Figure 1c).

In summary, regarding interaction of these SNPs with lung function, it can be concluded that the C allele in PPP2R4:c.-185A>C may exert a certain stabilization of the CFTR to be moved at the cell membrane, an effect that is expressed by a significant milder progression of airway mechanics. Heterozygosity for SNAP23:c.267-9T>C may be helpful in the prohibition to develop pulmonary hyperinflation and hence small airway dysfunction, an effect which is additionally also shown for PPP2R1A:c.*465T>A genotype AA. On the other hand, it seems that homozygosity for the wild-type allele of the three SNPS in *KRT19*, that modulates CFTR maturation, may lead to a faster development of pulmonary hyperinflation.

The results of our study show especially how important it is to assess genotype–phenotype associations by a whole set of lung function parameters, and not only by spirometry and to consider longitudinal data, because the SNPs might influence disease severity in different ways at different time points. We know that owing to the small patient cohort certain effects probably are under- or over-estimated in our study, therefore it needs further studies to evaluate these first promising results. We intend to confirm our results in a second bigger cohort but we were not able so far to find a CF center with comparable longitudinal lung function data. Currently, we are evaluating our candidate genes functionally by searching for differences in expression by comparative quantitative proteomics where we already found KRT19 to be upregulated twofold in a F508del homozygous compared with a wild-type bronchial cell line (publication in preparation).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Riordan JR, Rommens JM, Kerem B *et al*: Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989; **245**: 1066–1073.
2 Kerem B, Rommens JM, Buchanan JA *et al*: Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989; **245**: 1073–1080.
3 Collins FS: Cystic fibrosis: molecular biology and therapeutic implications. *Science* 1992; **256**: 774–779.
4 Rosenstein BJ, Cutting GR: The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel. *J Pediatr* 1998; **132**: 589–595.
5 Knowles MR, Durie PR: What is cystic fibrosis? *N Engl J Med* 2002; **347**: 439–442.
6 Corey M, Edwards L, Levison H, Knowles M: Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr* 1997; **131**: 809–814.
7 Gustafsson PM, Aurora P, Lindblad A: Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis. *Eur Respir J* 2003; **22**: 972–979.
8 Kraemer R, Blum A, Schibler A, Ammann RA, Gallati S: Ventilation inhomogeneities in relation to standard lung function in patients with cystic fibrosis. *Am J Respir Crit Care Med* 2005; **171**: 371–378.
9 Kraemer R, Baldwin DN, Ammann RA, Frey U, Gallati S: Progression of pulmonary hyperinflation and trapped gas associated with genetic and environmental factors in children with cystic fibrosis. *Respir Res* 2006; **7**: 138.
10 Kraemer R, Latzin P, Pramana I, Ballinari P, Gallati S, Frey U: Long-term gas exchange characteristics as markers of deterioration in patients with cystic fibrosis. *Respir Res* 2009; **10**: 106.
11 Aurora P, Bush A, Gustafsson P *et al*: Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis. *Am J Respir Crit Care Med* 2005; **171**: 249–256.
12 Kraemer R, Birrer P, Liechti-Gallati S: Genotype-phenotype association in infants with cystic fibrosis at the time of diagnosis. *Pediatr Res* 1998; **44**: 920–926.
13 Ranganathan SC, Stocks J, Dezateux C *et al*: The evolution of airway function in early childhood following clinical diagnosis of cystic fibrosis. *Am J Respir Crit Care Med* 2004; **169**: 928–933.
14 Konstan MW, Hilliard KA, Norvell TM, Berger M: Bronchoalveolar lavage findings in cystic fibrosis patients with stable, clinically mild lung disease suggest ongoing infection and inflammation. *Am J Respir Crit Care Med* 1994; **150**: 448–454.
15 Ranganathan SC, Dezateux C, Bush A *et al*: Airway function in infants newly diagnosed with cystic fibrosis. *Lancet* 2001; **358**: 1964–1965.
16 Vanscoy LL, Blackman SM, Collaco JM *et al*: Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med* 2007; **175**: 1036–1043.
17 Buranawuti K, Boyle MP, Cheng S *et al*: Variants in mannose-binding lectin and tumour necrosis factor alpha affect survival in cystic fibrosis. *J Med Genet* 2007; **44**: 209–214.
18 Bremer LA, Blackman SM, Vanscoy LL *et al*: Interaction between a novel TGFB1 haplotype and CFTR genotype is associated with improved lung function in cystic fibrosis. *Hum Mol Genet* 2008; **17**: 2228–2237.
19 Gu Y, Harley IT, Henderson LB *et al*: Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. *Nature* 2009; **458**: 1039–1042.
20 Hillian AD, Londono D, Dunn JM *et al*: Modulation of cystic fibrosis lung disease by variants in interleukin-8. *Genes Immun* 2008; **9**: 501–508.
21 Darrah R, McKone E, O'Connor C *et al*: EDNRA variants associate with smooth muscle mRNA levels, cell proliferation rates, and cystic fibrosis pulmonary disease severity. *Physiol Genomics* 2010; **41**: 71–77.
22 Wright FA, Strug LJ, Doshi VK *et al*: Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* 2011; **43**: 539–546.
23 Cormet-Boyaka E, Di A, Chang SY *et al*: CFTR chloride channels are regulated by a SNAP-23/syntaxin 1A complex. *Proc Natl Acad Sci USA* 2002; **99**: 12477–12482.
24 Tang BL, Gee HY, Lee MG: The cystic fibrosis transmembrane conductance regulator's expanding SNARE interactome. *Traffic* 2011; **12**: 364–371.
25 Le Drevo MA, Benz N, Kerbiriou M *et al*: Annexin A5 increases the cell surface expression and the chloride channel function of the DeltaF508-cystic fibrosis transmembrane regulator. *Biochim Biophys Acta* 2008; **1782**: 605–614.
26 Faria D, Dahimene S, Alessio L *et al*: Effect of Annexin A5 on CFTR: regulated traffic or scaffolding? *Mol Membr Biol* 2011; **28**: 14–29.
27 Luo J, Pato MD, Riordan JR, Hanrahan JW: Differential regulation of single CFTR channels by PP2C, PP2A, and other phosphatases. *Am J Physiol* 1998; **274**: C1397–C1410.
28 Thelin WR, Kesimer M, Tarran R *et al*: The cystic fibrosis transmembrane conductance regulator is regulated by a direct interaction with the protein phosphatase 2A. *J Biol Chem* 2005; **280**: 41512–41520.
29 Vastiau A, Cao L, Jaspers M *et al*: Interaction of the protein phosphatase 2A with the regulatory domain of the cystic fibrosis transmembrane conductance regulator channel. *FEBS Lett* 2005; **579**: 3392–3396.
30 Rotin D, Staub O: Role of the ubiquitin system in regulating ion transport. *Pflugers Arch* 2011; **461**: 1–21.
31 Caohuy H, Jozwik C, Pollard HB: Rescue of DeltaF508-CFTR by the SGK1/Nedd4-2 signaling pathway. *J Biol Chem* 2009; **284**: 25241–25253.
32 Myerburg MM, McKenna EE, Luke CJ, Frizzell RA, Kleyman TR, Pilewski JM: Prostasin expression is regulated by airway surface liquid volume and is increased in cystic fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2008; **294**: L932–L941.
33 Koulov AV, Lapointe P, Lu B *et al*: Biological and structural basis for Aha1 regulation of Hsp90 ATPase activity in maintaining proteostasis in the human disease cystic fibrosis. *Mol Biol Cell* 2010; **21**: 871–884.
34 Harada K, Okiyoneda T, Hashimoto Y *et al*: Calreticulin negatively regulates the cell surface expression of cystic fibrosis transmembrane conductance regulator. *J Biol Chem* 2006; **281**: 12841–12848.
35 Salas PJ, Rodriguez ML, Viciana AL, Vega-Salas DE, Hauri HP: The apical submembrane cytoskeleton participates in the organization of the apical pole in epithelial cells. *J Cell Biol* 1997; **137**: 359–375.
36 Sun F: Poster Session Abstracts. *Pediatric Pulmology* 2009; **44**: 213–435.
37 Kraemer R, Meister B: Fast real-time moment-ratio analysis of multibreath nitrogen washout in children. *J Appl Physiol* 1985; **59**: 1137–1144.

38 Kraemer R, Zehnder M, Meister B: Intrapulmonary gas distribution in healthy children. *Respir Physiol* 1986; **65**: 127–137.

39 Zapletal A, Samanek M, Paul T: *Lung function in Children and Adolescents: Methods, Reference Values*. Basel: Karger, 1987.

40 Liechti-Gallati S, Schneider V, Neeser D, Kraemer R: Two buffer PAGE system-based SSCP/HD analysis: a general protocol for rapid and sensitive mutation screening in cystic fibrosis and any other human genetic disease. *Eur J Hum Genet* 1999; **7**: 590–598.

41 Excoffier L, Lischer HE: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010; **10**: 564–567.

42 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.

43 Ng PC, Henikoff S: Predicting deleterious amino acid substitutions. *Genome Res* 2001; **11**: 863–874.

44 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.

45 Piva F, Giulietti M, Burini AB, Principato G: SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* 2012; **33**: 81–85.

46 Sharma M, Benharouga M, Hu W, Lukacs GL: Conformational and temperature-sensitive stability defects of the delta F508 cystic fibrosis transmembrane conductance regulator in post-endoplasmic reticulum compartments. *J Biol Chem* 2001; **276**: 8942–8950.

47 de Moor CH, Meijer H, Lissenden S: Mechanisms of translational control by the 3′ UTR in development and differentiation. *Semin Cell Dev Biol* 2005; **16**: 49–58.

48 Skoko N, Baralle M, Buratti E, Baralle FE: The pathological splicing mutation c.6792C>G in NF1 exon 37 causes a change of tenancy between antagonistic splicing factors. *FEBS Lett* 2008; **582**: 2231–2236.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)