Post-reproductive survival in a polygamous society in rural Africa

Bodegom, D. van

Chapter 6

Common CFTR gene variants influence body composition and survival in rural Ghana

M. Kuningas, D. van Bodegom, L. May, J.J. Meij, P.E. Slagboom, R.G.J. Westendorp

Published in:
Human Genetics **127**(2):201-206 (2010)

Abstract

Various studies in mice have found support for the hypothesis that heterozygous carriers of *CFTR* mutations have an increased resistance to fatal infection compared to both homozygous mutation carriers and non-carriers, while in humans such evidence is scarce. In this study, we assessed the *CFTR* heterozygotes survival advantage hypothesis in a contemporary rural population that lives under adverse environmental conditions in the Upper-East region of Ghana. We genotyped 30 SNPs throughout the *CFTR* gene in 4,230 participant and tested their influence on survival and on body composition in the population at large. With a sliding-window haplotype analysis, we identified a set of six common haplotypes that influenced survival probabilities (global $p=6.00x10^{-05}$). Individual haplotype analyses revealed two haplotypes of specific interest. One of these haplotypes was enriched (p=0.003) whereas the other was depleted (p=0.041) among people of old age (≥65 years) compared to young study participants (≤5 years). In addition, children (n=474) carrying the latter haplotype had lower body weight ($p_{trend}=0.020$) and height ($p_{trend}=0.010$) compared to non-carriers. For all these analyses, similar associations for heterozygous and homozygous *CFTR* haplotype carriers were observed, revealing an additive effect of haplotype alleles. In conclusion, we identified common haplotypes in the *CFTR* gene that influence survival and body composition in the population at large with no evidence for heterozygote advantage.

Introduction

The cystic fibrosis transmembrane conductance regulator (CFTR) gene contains a variety of mutations which all contribute to the development of cystic fibrosis (CF). In order to explain the high frequency of CF in various populations, it has been hypothesized that in our recent evolutionary past heterozygous carriers of *CFTR* mutations have had a survival advantage compared to both homozygous mutation carriers and non-carriers[1]. Proposed mechanisms that would lead to such advantage are an increased resistance to cholera, typhoid fever, or tuberculosis[2-7]. Several studies with mouse models have provided support for this hypothesis, where it has been shown that heterozygous *CFTR* mutation carriers are more resistant to cholera-induced diarrhea, and against bacteria that use CFTR for entering cells[2,5]. In case of humans however, support for this hypothesis is generally lacking[3,8]. Therefore, the aim of this study was to assess the *CFTR* heterozygotes advantage hypothesis in a contemporary rural population that lives under adverse environmental conditions in the Garu-Tempane district, a densely populated agricultural area in southeast of the Upper-East region of Ghana[9,10]. This region is highly endemic for malaria, typhoid fever, diarrheal diseases and intestinal helminth infections, whereas hospitals and medical services are only marginally available. As there is evidence that the mutation spectrum for CF in African populations is different than in European populations[11-13], we selected common variants from the *CFTR* gene and tested their influence on survival as well as on body composition in the population at large.

Materials and methods

Research area and study population

This study was conducted in the Garu-Tempane district, a densely populated agricultural area in the southeast of the Upper-East region of Ghana, which is inhabited by several tribes, mostly Bimoba (67%) and Kusasi (27%)[9]. The area is highly endemic for malaria, typhoid fever, meningococcal disease and intestinal helminth infections. Hospitals and medical services are only marginally available in the area. Vaccination of children was introduced in the early 1990s, but coverage amongst children is highly variable. It is estimated that about 50% of the children under the age of ten years have been vaccinated at least once against either measles, poliomyelitis, or diphtheria-tetanus-pertussis[9]. The region and study population have been described in more detail elsewhere[9,10]. The Medical Ethical Committee of the Ghana Health Service in Ghana, as well as the Medical Ethical

Committee of the Leiden University Medical Center in the Netherlands approved the study. Witness observed oral informed consent was obtained from all participants.

The measurement of body composition and socioeconomic status (SES)
Weight (kg) and height (cm) was measured in 2007 for 474 children who were equal to or younger than five years of age. In 2007 a DHS-type questionnaire was designed to assess the SES of the study participants using a free listing technique, whereby we asked people, both male and female, from different villages in the research area in focus group discussions to list the household items of most value[10]. The resulting list of valuable items was comparable to part of the core welfare indications questionnaire (CWIQ) from the World Bank and to the extended DHS asset list, adapted to our region.

SNP selection and genotyping
We selected 37 SNPs from the *CFTR* gene region covering 204.3 kbp (chr7:116897317-117101642) from the HapMap database release #21 (www.hapmap.org) using the Yoruba in Ibadan, Nigeria (Yoruba) data. The Haploview's program Tagger[14] was used to derive a set of tag SNPs from the whole gene region such that each common SNP (5%) in that set was captured with r^2 0.8. All SNPs were genotyped using mass spectrometry (Sequenom Inc, San Diego, CA, USA), according to the manufacturer's instructions. Altogether 4,336 participants were genotyped for 37 SNPs in the *CFTR* gene. Genotyping failed for 7 SNPs as determined by more than 20% missing individuals. From the 4,336 participants 106 (2.44%) were excluded due to >50% missing genotypes, leaving 4,230 participants for further analyses.

Statistical analysis
The program Haploview[14] was used to estimate allele frequencies, Hardy–Weinberg equilibrium and pair-wise linkage disequilibrium (LD) between the SNPs. Sliding window haplotype analysis was performed with the program Haplo.Stats (version 1.4.0)[15]. In this analysis, using a range of n.slide values, the region with the strongest association will consistently have low p-values for locus subsets containing the associated haplotypes. The global p-value measures significance of the entire set of haplotypes for the locus subset. Haplo.stats was also used to calculate and compare allele frequencies between young and old study

participants. Haplotypes and haplotype frequencies per individual were calculated using the program SNPHAP(http://wwwgene.cimr.cam.ac.uk/clayton/software). The prevalence of *CFTR* haplotype alleles in elderly compared to young was analyzed using logistic regression. For analyses with body weight and height, these variables were first converted into age-adjusted z-scores ([individual level-mean level]/SD), in order to provide comparable estimates for haplotype effects. The cross-sectional associations between *CFTR* haplotypes and body weight and height were performed using linear regression. The linear and logistic regression analyses were performed using STATA version 9 (StataCorp LP, TX, USA) statistical software. In all these analyses the posterior probabilities of pairs of haplotypes per participant, as estimated by SNPHAP, were used as weights. All analyses were adjusted for sex, SES and tribe. No evidence for population stratification or structure was detected when conducting non parametric clustering of genotypes for 147 SNP genotyped in controls (children ≤5 years of age) and cases (elderly ≥65 years of age) using the AWClust algorithm (http://awclust.sourceforge.net/).

Table 1. Characteristics of the study population

N	4,230
Women (n, %)	2,888 (68%)
Elderly (≥65 years, n %)	819 (19%)
Children (≤5 years, n %)	936 (22%)
Children ≤5 years measured (n, %)	474 (11%)
Weight (mean, SD)	13.5 (1.84)
Height (mean, SD)	94.7 (7.04)

Results

Altogether 30 single nucleotide polymorphisms (SNPs) in the *CFTR* gene were genotyped in 4,230 participants (supplementary table 1). From these participants 936 (22%) were children five years of age or younger (age range 0-5) and 819 (19%) were 65 years of age or older (age range 65-97) (table 1). To assess whether genetic variants in the *CFTR* gene influence survival in adverse environmental conditions, we compared allele frequencies between these young and old study participants. As linkage disequilibrium (LD) within the *CFTR* gene is not strong, we used a sliding window haplotype analysis approach. With a window length of four SNPs, strong associations for haplotype frequency differences between young and old study participants were observed for two locations in the *CFTR* gene: in the middle (global $p=6.00x10^{-05}$) and in the 3' end of the gene (global p-value of $8.20x10^{-4}$) (figure 1). However, only one very rare haplotype (population frequency 0.05%) contributed to the association in the 3' end of the gene, and therefore it was not investigated further.

In the middle of the *CFTR* gene, the rs213952, rs10281281, rs17140174 and rs3808185 SNPs constitute the window that associated with allele frequency differences between young and old study participants. These SNPs, give rise to six common haplotypes (frequency ≥5%) (figure 2). From the individual haplotypes, haplotype 1 (frequency 26%) was depleted in participants of old age (p=0.041) whereas haplotype 5 (frequency 13%) was enriched (p=0.003) (figure 2). In addition, an additive effect of haplotype alleles for lower ($p_{trend}=0.06$) and higher ($p_{trend}=0.004$) chances to reach an old age was observed for carriers of haplotype 1 and haplotype 5, respectively (table 2). Hence, no evidence for advantage or disadvantage for the heterozygous haplotype carriers over homozygotes was observed.

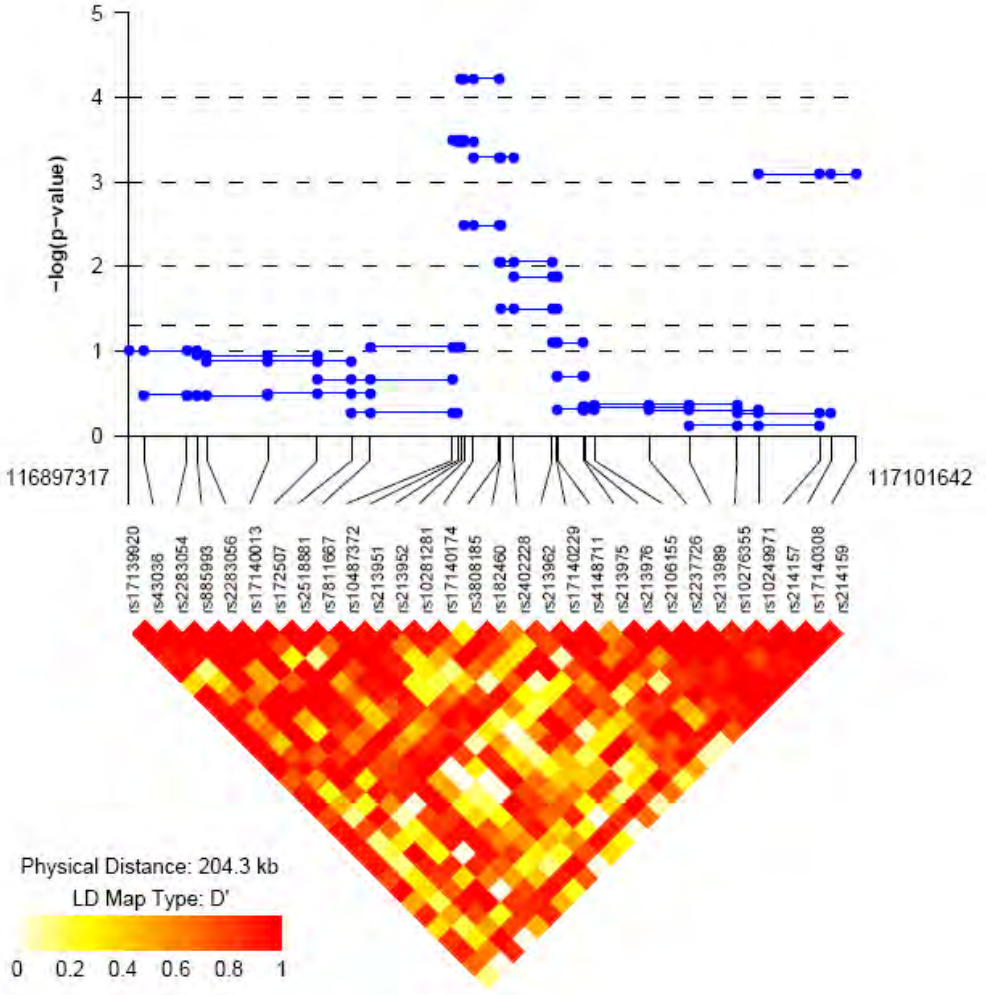

Figure 1. The genetic structure of CFTR gene and –log p-values for the global association of CFTR haplotype frequency differences between young (≤5 years, n=936) and old ((≥65 years, n=819) study participants, as assessed by sliding window haplotype analysis. Analysis was adjusted for sex, socioeconomic status and tribe.

Table 2. The prevalence of *CFTR* haplotypes in the group of old (≥65 years) study participants compared to young (≤5 years)

	Young versus old			
	0- copies	1-copy	2-copies	P_{trend}
	OR (95%CI)	OR (95%CI)	OR (95%CI)	
Haplotype 1	1 (reference)	0.91 (0.75-1.10)	0.69 (0.47-1.01)	0.06
Haplotype 2	1 (reference)	1.16 (0.96-1.41)	1.06 (0.66-1.71)	0.22
Haplotype 3	1 (reference)	1.15 (0.94-1.40)	0.85 (0.49-1.48)	0.47
Haplotype 4	1 (reference)	0.93 (0.76-1.14)	0.54 (0.27-1.08)	0.13
Haplotype 5	1 (reference)	1.26 (1.02-1.55)*	2.20 (1.07-4.52)*	**0.004**
Haplotype 6	1 (reference)	1.07 (0.81-1.40)	1.14 (0.12-10.5)	0.65

OR = odds ratio; CI = confidence interval; *-p<0.05. Odds ratios were calculated using sex, socioeconomic status and tribe adjusted logistic regression.

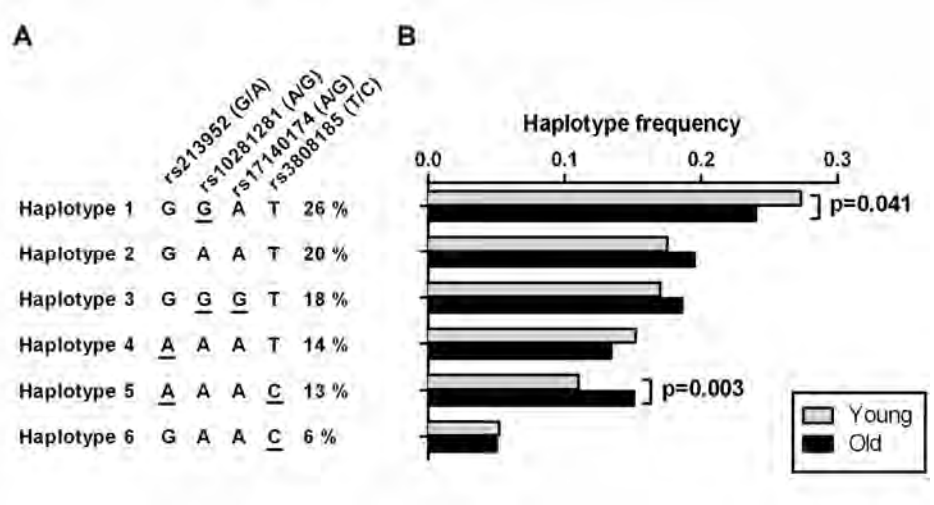

Figure 2. CFTR haplotypes and their frequencies in the general population (n=4230) (A), and in the young (≤5 years, n=936) and old (≥65 years, n=819) study participants (B). Underlining denotes the minor allele. Differences in haplotype frequencies between young and old were analyzed using sex, socioeconomic status and tribe adjusted Haplo.stats.

In addition to survival, we also assessed the influence of *CFTR* haplotypes on weight and height in 474 children who were equal to or younger than five years of

age (mean age 3.42 years, SD 0.61). Carriers of haplotype 1 had lower weight (p_{trend}=0.020) and height (p_{trend}=0.010), whereas children carrying haplotype 5 had only marginal differences in weight (p_{trend}=0.11) and height (p_{trend}=0.43) compared to non-carriers (figure 3; supplementary table 2). Besides these associations a beneficial influences on body composition were observed for haplotype 4 (frequency 14%). Carriers of this haplotype had higher weight (p_{trend}=0.013) and height (p_{trend}=0.12) compared to non-carriers (figure 3; supplementary table 2), even though they had similar survival probabilities. For all these analyses similar associations were observed for boys and girls, and for children from compounds with different socioeconomic status (data not shown).

All these analyses were repeated using the four individual SNPs: rs213952, rs10281281, rs17140174 and rs3808185. These analyses revealed associations with two of these SNPs (rs10281281 and rs3808185), which however were not as strong as in combination within a haplotype (supplementary tables 3 and 4).

Discussion

In this study we identified common *CFTR* haplotypes that influence body composition in children and survival in the population at large. In all our analyses we observed an additive effect of haplotype alleles and no selective advantage for heterozygous carriers over both, homozygous mutation carriers and non-carriers. This observation is in accordance with the few other studies that have been conducted[3,7]. Therefore, it has been hypothesized that other mechanisms, such as a past selective event or random genetic drift, would explain for the high frequency of CF in European populations. Interestingly, the prevalence of CF and its causal mutations are different in populations of different origin, and do not coincide with locations in the world where the proposed selective agents, such as cholera and typhoid fever have the highest prevalence[16]. In Africa the prevalence of CF is low and the mutation spectrum for CF in African populations has been shown to be different than in European populations[11-13]. One reason for the lower prevalence of CF in Africa could be that the affected individuals die in early childhood either due to the lack of proper diagnosis and/or medical care. This might also be the case in our research area.

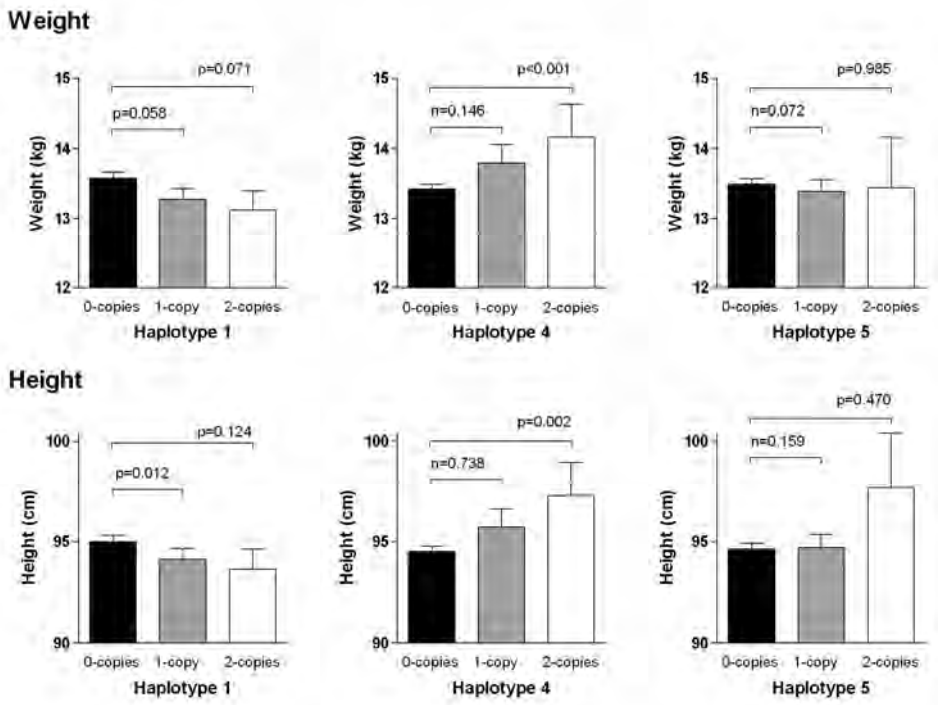

Figure 3. The influence of CFTR haplotypes on weight and height in children equal to or younger than five years of age (n=474). Linear regression adjusted for sex, socioeconomic status and tribe.

On the other hand, the lack of evidence for selective advantage for heterozygous carriers over both, homozygous mutation carriers and non-carriers in our study could rely on our selection of common variants in the *CFTR* gene, instead of mutations that have been associated with CF. It could be that the common variants contribute to mild functional differences of the CFTR protein, leading to additive effects. In case of *CFTR* mutations, such effects would be observed for the heterozygous carriers, whereas homozygous carriers would suffer CF.

Previously, it has been observed that children with untreated or poorly controlled CF have poor growth, reflected by lower weight and height[17,18]. Also in this study, genetic variants in the *CFTR* gene influenced body composition. Carriers of haplotype 1 had lower weight and height and this haplotype was depleted in people of old age. This detrimental effect could be because of pancreatic insufficiency that leads to malnutrition[19,20]. However, the mechanisms that lead to the enrichment of haplotype 5 still need to be elucidated, since carriers of this

haplotype had no differences in body compositions compared to non-carriers. The opposite was true for carriers of haplotype 4, who had better body compositions but no survival advantage. It is known that variation in the *CFTR* gene influences multiple phenotypes and it might be that some variants are more detrimental to some phenotypes than to others. In addition, the phenotypic influences of *CFTR* variants can be affected by other genes, as recently, several modifier genes for CF have been identified[21,22].

The strengths of this study are the large population size and the thorough evaluation of common variants in the *CFTR* gene, which through linkage disequilibrium capture the information for functional variants that were not genotyped. The limitations include the lack of data on body composition for grown-ups, on the prevalence of diseases, and on the specific mortality causes. Therefore, we could not directly assess the influence of *CFTR* variants on health. In addition, considering the number of tests performed, it cannot be excluded that some of the observed associations were due to chance. Nevertheless, we identified common haplotypes in the *CFTR* gene that influence survival and body composition in the population at large.

Acknowledgements
This research was supported by the Netherlands Foundation for the advancements of Tropical Research (WOTRO 93-467), the Netherlands Organization for Scientific Research (NWO 050-60-810), the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NGI/NWO 911-03-016), the EU funded Network of Excellence Lifespan (FP6 036894), Stichting Dioraphte and the Centre for Medical Systems Biology, which is a centre of excellence approved by the NWO in the Netherlands. We want to thank Dennis Kremer for assistance in the genetic studies.

References

1. Danks D., Allan J., Anderson C. A genetic study of fibrocystic disease of the pancreas. *Ann Hum Genet.* **3** 323-356 (1965).

2. Gabriel S.E. et al. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science.* **266** 107-109 (1994).

3. Hogenauer C. et al. Active intestinal chloride secretion in human carriers of cystic fibrosis mutations: an evaluation of the hypothesis that heterozygotes have subnormal active intestinal chloride secretion. *Am J Hum Genet.* **67** 1422-1427 (2000).

4. Pier G.B. et al. Role of mutant CFTR in hypersusceptibility of cystic fibrosis patients to lung infections. *Science.* **271** 64-67 (1996).

5. Pier G.B. et al. Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature.* **393** 79-82 (1998).

6. Poolman E.M., Galvani A.P. Evaluating candidate agents of selective pressure for cystic fibrosis. *J R Soc Interface.* **4** 91-98 (2007).

7. Van de Vosse E. et al. Susceptibility to typhoid fever is associated with a polymorphism in the cystic fibrosis transmembrane conductance regulator (CFTR). *Hum Genet.* **118** 138-140 (2005).

8. Jorde L.B., Lathrop G.M. A test of the heterozygote-advantage hypothesis in cystic fibrosis carriers. *Am J Hum Genet.* **42** 808-815 (1988).

9. Meij J.J. et al. Low-cost interventions accelerate epidemiological transition in Upper East Ghana. *Trans R Soc Trop Med Hyg.* **103** 173-178 (2009).

10. Van Bodegom D. et al. Socio-economic status by rapid appraisal is highly correlated with mortality risks in rural Africa. *Trans R Soc Trop Med Hyg.* **103** 795-800 (2009).

11. Carles S. et al. First report of CFTR mutations in black cystic fibrosis patients of southern African origin. *J Med Genet.* **33** 802-804 (1996).

12. Goldman A. et al. The molecular basis of cystic fibrosis in South Africa. *Clin Genet.* **59** 37-41 (2001).

13. Padoa C. et al. Cystic fibrosis carrier frequencies in populations of African origin. *J Med Genet.* **36** 41-44 (1999).

14. Barrett J.C. et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* **21** 263-265 (2005).

15. Schaid D.J. et al. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet.* **70** 425-434 (2002).

16. Fontelo P. Protection against cholera. *Science.* **267** 440 (1995).

17. Miller M. et al. Altered body composition and muscle protein degradation in nutritionally growth-retarded children with cystic fibrosis. *Am J Clin Nutr.* **36** 492-499 (1982).

18. Morison S. et al. Height and weight in cystic fibrosis: a cross sectional study. UK Cystic Fibrosis Survey Management Committee. *Arch Dis Child.* **77** 497-500 (1997).

19. Davies J.C., Alton E.W., Bush A. Cystic fibrosis. *BMJ.* **335** 1255-1259 (2007).

20. Zielenski J., Tsui L.C. Cystic fibrosis: genotypic and phenotypic variations. *Ann Rev Genet.* **29** 777-807 (1995).

21. Gu Y. et al. Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. *Nature.* **458** 1039-1042 (2009).

22. Levy H. et al. IL1B polymorphisms modulate cystic fibrosis lung disease. *Pediatr Pulmonol.* **44** 580-593 (2009).

Supplementary table 1. Genotyped *CFTR* SNPs

#	Name	Position	Alleles	MAF
1	rs17139920	promoter	G:A	0.377
2	rs43036	promoter	C:T	0.428
3	rs2283054	Intron 1	G:A	0.049
4	rs885993	Intron 1	C:T	0.076
5	rs2283056	Intron 1	A:G	0.182
6	rs17140013	Intron 3	T:G	0.075
7	rs172507	Intron 3	T:A	0.264
8	rs2518881	Intron 4	G:A	0.291
9	rs7811667	Intron 7	A:G	0.049
10	rs10487372	Intron 11	C:T	0.139
11	rs213951	Intron 11	A:G	0.208
12	rs213952	Intron 11	G:A	0.282
13	rs10281281	Intron 11	A:G	0.475
14	rs17140174	Intron 11	A:G	0.211
15	rs3808185	Intron 11	T:C	0.227
16	rs182460	Intron 11	A:G	0.075
17	rs2402228	Intron 11	A:T	0.169
18	rs213962	Intron 12	C:G	0.065
19	rs17140229	Intron 12	T:C	0.408
20	rs4148711	Intron 13	A:T	0.243
21	rs213975	Intron 15	C:T	0.096
22	rs213976	Intron 15	G:T	0.079
23	rs2106155	Intron 15	A:C	0.062
24	rs2237726	Intron 21	C:T	0.234
25	rs213989	Intron 21	C:A	0.215
26	rs10276355	Intron 22	C:G	0.024
27	rs10249971	Intron 23	A:T	0.085
28	rs214157	Intron 24	G:A	0.115
29	rs17140308	3' UTR	A:G	0.224
30	rs214159	3' UTR	A:C	0.407

MAF = minor allele frequency.

Supplementary table 2. The influence of *CFTR* haplotypes on weight and height in children equal to or younger than five years of age (n=474)

	Haplotype alleles			
	0-copies	1-copy	2-copies	
z-scores	Mean (SE)	Difference (SE)	Difference (SE)	P_{trend}
Haplotype 1				
Weight	reference	-0.15 (0.08)	-0.29 (0.16)	**0.020**
Height	reference	-0.19 (0.08)*	-0.18 (0.12)	**0.010**
Haplotype 2				
Weight	reference	0.15 (0.08)	-0.04 (0.20)	0.28
Height	reference	0.14 (0.08)	0.07 (0.19)	0.16
Haplotype 3				
Weight	reference	-0.01 (0.09)	0.07 (0.20)	0.89
Height	reference	-0.08 (0.09)	0.16 (0.15)	0.92
Haplotype 4				
Weight	reference	0.13 (0.09)	0.42 (0.12)*	**0.013**
Height	reference	0.03 (0.08)	0.43 (0.14)*	0.12
Haplotype 5				
Weight	reference	-0.16 (0.09)	0.01 (0.34)	0.11
Height	reference	-0.12 (0.08)	0.34 (0.47)	0.43
Haplotype 6				
Weight	reference	0.07 (0.14)	-0.09 (0.09)	0.64
Height	reference	0.06 (0.11)	0.64 (0.31)	0.56

Linear regression adjusted for sex, socioeconomic status and tribe; *p-value <0.05 in comparison to people carrying 0-copies of the respective haplotype.

Supplementary table 3. The prevalence of *CFTR* SNPs in the group of old (≥65 years) study participants compared to young (≤5 years)

	Young versus old			
	0- copies OR (95%CI)	1-copy OR (95%CI)	2-copies OR (95%CI)	P_{trend}
rs213952 (G/A)	1 (reference)	0.98 (0.80-1.20)	0.98 (0.66-1.44)	0.848
rs10281281 (A/G)	1 (reference)	0.92 (0.73-1.17)	0.72 (0.54-0.96)*	**0.030**
rs17140174 (A/G)	1 (reference)	1.15 (0.93-1.41)	0.91 (0.58-1.41)	0.536
rs3808185 (T/C)	1 (reference)	1.24 (1.01-1.53)*	1.46 (0.96-2.23)	**0.014**

Logistic regression adjusted for sex, SES and tribe.

Supplementary table 4. The influence of *CFTR* SNPs on weight and height in children equal to or younger than five years of age (n=474)

	Haplotype alleles			
	0-copies	1-copy	2-copies	
z-scores	Mean (SE)	Difference (SE)	Difference (SE)	P_{trend}
rs213952 (G/A)				
Weight	reference	-0.002 (0.09)	0.20 (0.17)	0.422
Height	reference	-0.07 (0.08)	0.28 (0.16)	0.468
rs10281281 (A/G)				
Weight	reference	-0.22 (0.10)*	-0.19 (0.12)	0.122
Height	reference	-0.28 (0.10)*	-0.19 (0.11)	0.090
rs17140174 (A/G)				
Weight	reference	-0.003 (0.09)	0.04 (0.18)	0.890
Height	reference	-0.07 (0.09)	0.14 (0.17)	0.960
rs3808185 (T/C)				
Weight	reference	-0.11 (0.09)	0.03 (0.20)	0.470
Height	reference	-0.07 (0.09)	0.10 (0.19)	0.837

Linear regression adjusted for sex, socioeconomic status and tribe; *p-value <0.05 in comparison to people carrying 0-copies of the respective haplotype.