ORIGINAL INVESTIGATION

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

**Maris Kuningas · David van Bodegom · Linda May ·
Johannes J. Meij · P. Eline Slagboom ·
Rudi G. J. Westendorp**

Abstract Various studies in mice have found support for the hypothesis that heterozygous carriers of cystic fibrosis transmembrane conductance regulator (*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 participants 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 $p = 6.00 \times 10^{-05}$). 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 ($p_{\text{trend}} = 0.020$) and height ($p_{\text{trend}} = 0.010$) 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.

M. Kuningas · D. van Bodegom · L. May ·
R. G. J. Westendorp (✉)
Department of Gerontology and Geriatrics,
Leiden University Medical Center (LUMC),
C2-R, PO Box 9600, 2300 RC Leiden, The Netherlands
e-mail: R.G.J.Westendorp@lumc.nl

M. Kuningas
e-mail: M.Kuningas@lumc.nl

D. van Bodegom · J. J. Meij · R. G. J. Westendorp
Leyden Academy on Vitality and Ageing,
Leiden, The Netherlands

P. E. Slagboom
Department of Molecular Epidemiology,
Leiden University Medical Center,
Leiden, The Netherlands

P. E. Slagboom · R. G. J. Westendorp
Netherlands Consortium for Healthy Ageing,
Leiden, The Netherlands

Introduction

The cystic fibrosis transmembrane conductance regulator (CFTR) gene contains a variety of mutations all of which 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 (Danks et al. 1965). Proposed mechanisms that would lead to such advantages are an increased resistance to cholera, typhoid fever or tuberculosis (Gabriel et al. 1994; Hogenauer et al. 2000; Pier et al. 1996, 1998; Poolman and Galvani 2007; van de Vosse et al. 2005). 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 (Gabriel et al. 1994; Pier et al. 1998). In case of humans, however, support for this hypothesis is generally lacking (Hogenauer et al. 2000; Jorde and Lathrop 1988). 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 (Meij et al. 2009; van Bodegom et al. 2009). 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 (Carles et al. 1996; Goldman et al. 2001; Padoa et al. 1999), we selected common variants from the *CFTR* gene and tested its 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%) (Meij et al. 2009). 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 among children is highly variable. It is estimated that about 50% of the children under the age of 10 years have been vaccinated at least once against measles, poliomyelitis or diphtheria–tetanus–pertussis (Meij et al. 2009). The region and study population have been described in more detail elsewhere (Meij et al. 2009; van Bodegom et al. 2009). 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) were measured in 2007 for 474 children who were equal to or younger than 5 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 (van Bodegom et al. 2009). 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 Hap-Map database release #21 (http://www.hapmap.org) using the Yoruba in Ibadan, Nigeria (Yoruba) data. The Haploview's program Tagger (Barrett et al. 2005) was used to derive a set of tag SNPs from the whole gene region such that each common SNP ($\geq 5\%$) in that set was captured with $r^2 \geq 0.8$. 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 (Barrett et al. 2005) 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) (Schaid et al. 2002). 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

Table 1 Characteristics of the study population

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

of pairs of haplotypes per participant, as estimated by SNP-HAP, 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 $\leq$5 years of age) and cases (elderly $\geq$65 years of age) using the AWClust algorithm (http://awclust.sourceforge.net/).

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 of 5 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 $p = 6.00 \times 10^{-05}$) and in the 3′ end of the gene (global p-value of 8.20×10^{-4}) (Fig. 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 $\geq$ 5%) (Fig. 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 ($p = 0.003$) (Fig. 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
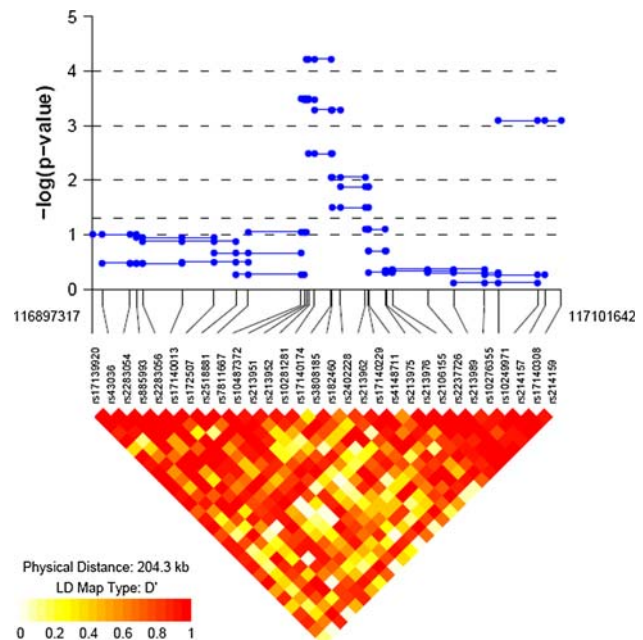

Fig. 1 The genetic structure of *CFTR* gene and $-$log p-values for the global association of *CFTR* haplotype frequency differences between young ($\leq$5 years, $n = 936$) and old ($\geq$65 years, $n = 819$) study participants, as assessed by sliding-window haplotype analysis. Analysis was adjusted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

for carriers of haplotypes 1 and 5, respectively (Table 2). Hence, no evidence for advantage or disadvantage for the heterozygous haplotype carriers over homozygotes was observed.

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 5 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 (Fig. 3; Supplementary Table 2). Besides these associations a beneficial influence on body composition was 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 (Fig. 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 that in combination within a haplotype (Supplementary Tables 3 and 4).

Fig. 2 *CFTR* haplotypes and its frequencies in the general population (*n* = 4,230) (**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
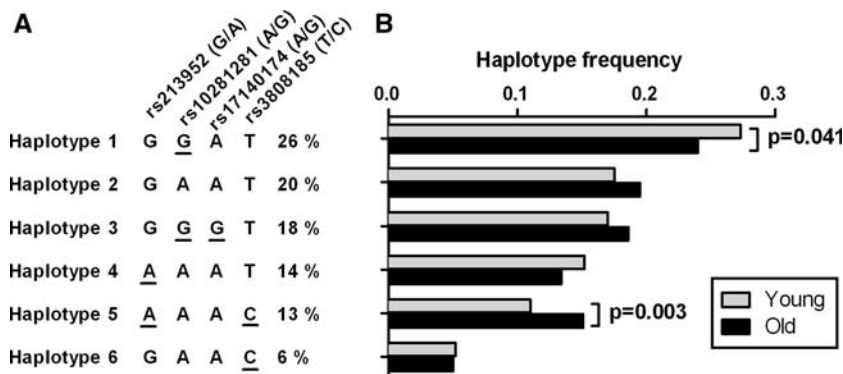

Table 2 The prevalence of *CFTR* haplotypes in the group of old (≥65 years) study participants compared to young (≤5 years)

	Young versus old			
	0-copies OR (95% CI)	1-copy OR (95% CI)	2-copies OR (95% CI)	p_{trend}
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

Odds ratios were calculated using sex, socioeconomic status and tribe adjusted logistic regression

OR odds ratio, *CI* confidence interval

* $p < 0.05$

Fig. 3 The influence of *CFTR* haplotypes on weight and height in children equal to or younger than 5 years of age (*n* = 474). Linear regression adjusted for sex, socioeconomic status and tribe
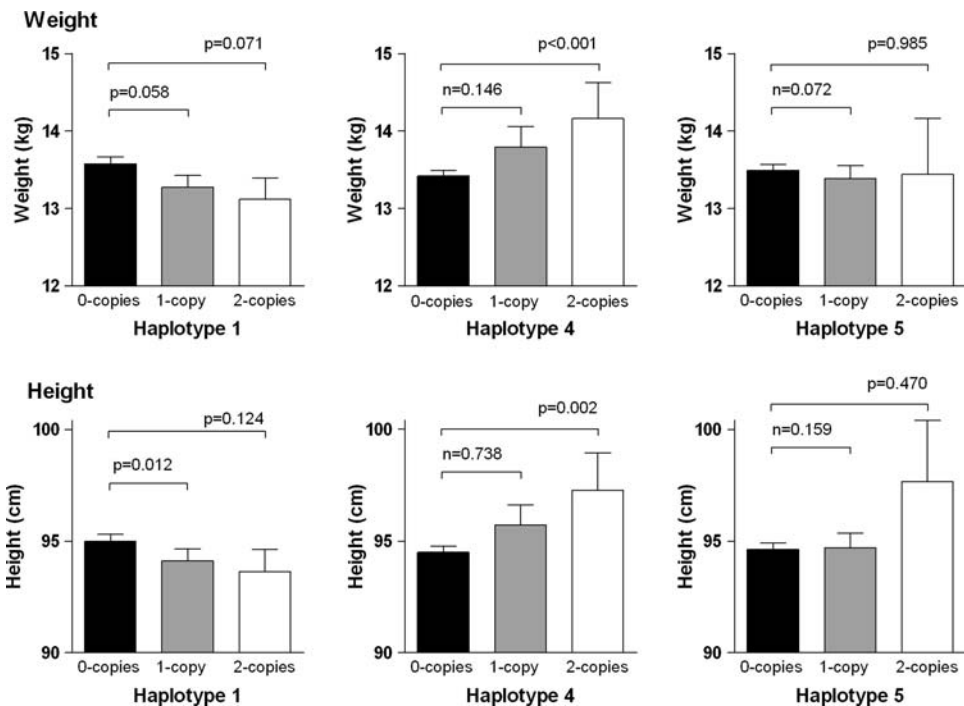
Discussion

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 (Hogenauer et al. 2000; van de Vosse et al. 2005). 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 (Fontelo 1995). In Africa the prevalence of CF is low and the mutation spectrum for CF in African populations has been shown to be different than that in European populations (Carles et al. 1996; Goldman et al. 2001; Padoa et al. 1999). 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.

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 (Miller et al. 1982; Morison et al. 1997). 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 (Davies et al. 2007; Zielenski and Tsui 1995). 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 (Gu et al. 2009; Levy et al. 2009).

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.

References

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Carles S, Desgeorges M, Goldman A, Thiart R, Guittard C, Kitazos CA, de Ravel TJ, Westwood AT, Claustres M, Ramsay M (1996) First report of CFTR mutations in black cystic fibrosis patients of southern African origin. J Med Genet 33:802–804

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

Davies JC, Alton EW, Bush A (2007) Cystic fibrosis. BMJ 335:1255–1259

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

Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ (1994) Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science 266:107–109

Goldman A, Labrum R, Claustres M, Desgeorges M, Guittard C, Wallace A, Ramsay M (2001) The molecular basis of cystic fibrosis in South Africa. Clin Genet 59:37–41

Gu Y, Harley IT, Henderson LB, Aronow BJ, Vietor I, Huber LA, Harley JB, Kilpatrick JR, Langefeld CD, Williams AH, Jegga AG, Chen J, Wills-Karp M, Arshad SH, Ewart SL, Thio CL, Flick LM, Filippi MD, Grimes HL, Drumm ML, Cutting GR, Knowles MR, Karp CL (2009) Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. Nature 458:1039–1042

Hogenauer C, Santa Ana CA, Porter JL, Millard M, Gelfand A, Rosenblatt RL, Prestidge CB, Fordtran JS (2000) 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

Jorde LB, Lathrop GM (1988) A test of the heterozygote-advantage hypothesis in cystic fibrosis carriers. Am J Hum Genet 42:808–815

Levy H, Murphy A, Zou F, Gerard C, Klanderman B, Schuemann B, Lazarus R, Garcia KC, Celedon JC, Drumm M, Dahmer M, Quasney M, Schneck K, Reske M, Knowles MR, Pier GB, Lange C, Weiss ST (2009) IL1B polymorphisms modulate cystic fibrosis lung disease. Pediatr Pulmonol 44:580–593

Meij JJ, de Craen AJ, Agana J, Plug D, Westendorp RG (2009) Low-cost interventions accelerate epidemiological transition in Upper East Ghana. Trans R Soc Trop Med Hyg 103:173–178

Miller M, Ward L, Thomas BJ, Cooksley WG, Shepherd RW (1982) Altered body composition and muscle protein degradation in nutritionally growth-retarded children with cystic fibrosis. Am J Clin Nutr 36:492–499

Morison S, Dodge JA, Cole TJ, Lewis PA, Coles EC, Geddes D, Russell G, Littlewood JM, Scott MT (1997) Height and weight in cystic fibrosis: a cross sectional study. UK Cystic Fibrosis Survey Management Committee. Arch Dis Child 77:497–500

Padoa C, Goldman A, Jenkins T, Ramsay M (1999) Cystic fibrosis carrier frequencies in populations of African origin. J Med Genet 36:41–44

Pier GB, Grout M, Zaidi TS, Olsen JC, Johnson LG, Yankaskas JR, Goldberg JB (1996) Role of mutant CFTR in hypersusceptibility of cystic fibrosis patients to lung infections. Science 271:64–67

Pier GB, Grout M, Zaidi T, Meluleni G, Mueschenborn SS, Banting G, Ratcliff R, Evans MJ, Colledge WH (1998) *Salmonella typhi* uses CFTR to enter intestinal epithelial cells. Nature 393:79–82

Poolman EM, Galvani AP (2007) Evaluating candidate agents of selective pressure for cystic fibrosis. J R Soc Interface 4:91–98

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70:425–434

van Bodegom D, May L, Kuningas M, Kaptijn R, Thomese F, Meij HJ, Amankwa J, Westendorp RG (2009) Socio-economic status by rapid appraisal is highly correlated with mortality risks in rural Africa. Trans R Soc Trop Med Hyg 103(8):795–800

van de Vosse E, Ali S, de Visser AW, Surjadi C, Widjaja S, Vollaard AM, van Dissel JT (2005) Susceptibility to typhoid fever is associated with a polymorphism in the cystic fibrosis transmembrane conductance regulator (CFTR). Hum Genet 118:138–140

Zielenski J, Tsui LC (1995) Cystic fibrosis: genotypic and phenotypic variations. Annu Rev Genet 29:777–807