The association of polymorphisms in 5-fluorouracil metabolism genes with outcome in adjuvant treatment of colorectal cancer

Aim: The purpose of this study was to investigate whether specific combinations of polymorphisms in 5-fluorouracil (5-FU) metabolism-related genes were associated with outcome in 5-FU-based adjuvant treatment of colorectal cancer. **Methods:** We analyzed two cohorts of 302 and 290 patients, respectively, one cohort for exploratory analyses and another cohort for validating the exploratory analyses. A total of ten polymorphisms in genes involved in 5-FU pharmacodynamics and pharmacokinetics were studied. End points were disease-free survival (DFS) and overall survival. Multifactor dimensionality reduction was used to identify genetic interaction profiles associated with outcome. **Results:** Low-expression alleles in thymidylate synthase (*TYMS*) were associated with decreased DFS and overall survival (DFS:hazard ratio [HR] exploration 2.65 [1.40–4.65]; p = 0.004, HR validation 1.69 [1.03–2.66]; p = 0.03). A specific multifactor dimensionality reduction derived combination of dihydropyrimidine dehydrogenase and *TYMS* polymorphisms was associated with increased DFS (HR exploration 0.69 [0.49–0.98]; p = 0.04, HR validation 0.66 [0.45–0.95]; p = 0.03). Specific combinations of functional polymorphisms in *DPYD* and *TYMS* were demonstrated to be associated with DFS and overall survival in patients receiving adjuvant 5-FU-based treatment. Specifically high *TYMS* expression alleles seem to be associated with decreased DFS.

Original submitted 13 April 2011; Revision submitted 10 June 2011

KEYWORDS: 5-fluorouracil ■ adjuvant chemotherapy ■ colorectal cancer ■ multifactor dimensionality reduction method ■ pharmacogenetics

Shoaib Afzal†, Milena Gusella, Søren Astrup Jensen, Ben Vainer, Ulla Vogel, Jon Trærup Andersen, Kasper Brødbæk, Morten Petersen, Espen Jimenez-Solem, Vilmos Adleff, Barna Budai, Erika Hitre, István Láng, Enikő Orosz, Laura Bertolaso, Carmen Barile, Roberto Padrini, Judit Kralovánszky, Felice Pasini & Henrik Enghusen Poulsen

†Author for correspondence:
Rigshospitalet, Copenhagen, Denmark
and
Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
and
Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark
Tel.: +45 5134 4875
shoaibafzal@hotmail.com
For a full list of affiliations please see the back page

5-fluorouracil (5-FU) is widely used to treat solid tumors, including colorectal cancers. 5-FU cytotoxicity depends primarily on two active metabolites. Fluorodeoxyuridine monophosphate inhibits the thymidylate synthase (TYMS) enzyme [1]. The other metabolite, fluorouridine triphosphate, impairs RNA function and thereby induces cell toxicity [2,3]. The inhibition of the TYMS enzyme is dependent on and enhanced by intracellular 5,10-methylenetetrahydrofolate (Supplementary Figure 1; www.futuremedicine.com/doi/suppl/10.2217/pgs.11.83) [1,4].

Increased sensitivity of cancer cell lines to 5-FU is correlated with decreased expression or activity of dihydropyrimidine dehydrogenase (*DPYD*), methylenetetrahydrofolate reductase (*MTHFR*) and *TYMS* and increased activity or expression of oratate phosporibosyltransferase (*OPRT* or *UMPS*) [5–10]. Studies investigating the association of *TYMS, DPYD, OPRT* and *MTHFR* polymorphisms or expression with survival in adjuvant 5-FU-based treatment of colorectal cancer have yielded contradictory results, especially regarding *TYMS* and *MTHFR* [11–24].

Most studies investigate the association of individual polymorphisms with disease-free survival (DFS) or overall survival (OS). Theoretically, the 5-FU metabolic phenotype is better explained by multigene and pathway-oriented analysis rather than single gene analysis. Systematic multipolymorphism combinations can be constructed by two methods: phenotypic classifications of gene expression or enzyme activity based on combinations of functional polymorphisms, that is, if two genotypes at two different loci in the same gene decrease enzyme activity their simultaneous presence will decrease activity more than if either one is present alone, or through investigation of gene–gene interactions. One method of studying gene–gene interactions is the multifactor dimensionality reduction (MDR) method, which reveals interactions between genetic profiles and response. The basic concept behind the MDR methodology is that polymorphisms in one gene are only functional in the presence of other polymorphisms. This concept is known as nonlinear interaction or epistasis.

The purpose of this study was to investigate whether individual polymorphisms, haplotypes, phenotypic classifications based on functional polymorphisms and specific gene–gene interactions were associated with DFS. We used two independent cohorts treated with adjuvant 5-FU-based chemotherapy; one cohort

part of future medicine fsg

for exploratory analyses and another cohort for validating the exploratory analyses. This analysis strategy has been successfully applied in a previous publication [25].

Methods

Populations

Exploration cohort

This cohort consisted of two prospectively collected patient groups from Italy (161 patients) and Hungary (129 patients). The Hungarian patient group was recruited from May 1995 to June 2004. The Italian patient group was recruited from 1999 to 2005 and follow-up ended in 2009. In both patient groups DNA was isolated from peripheral blood mononuclear cells, all patients were Dukes' stage B2 and C and all patients were treated with surgery and the Mayo regimen (folinate 20 mg/m^2, 5-FU 425 mg/m^2/day, days 1–5, every 28 days, six cycles). The starting dose could be reduced by 25% in the oldest patients. Each patient provided written informed consent and both studies were approved by their respective local ethics committees [15,22].

Validation cohort

A total of 302 Caucasian patients with Dukes' stage B2 and C treated at Rigshospitalet (Copenhagen University Hospital, Denmark) with surgery and the Mayo regimen (Levofolinate 10 mg/m^2, 5-FU 425 mg/m^2/day, days 1–5, every 28 days, six cycles) from 1996 to 2003 were eligible for our study. DNA was isolated from formalin-fixed paraffin-embedded tumor tissue, with maximum 50% normal tissue. Clinical data and tumor pathology was reviewed retrospectively. The last follow-up date was 30 August 2007. This study was approved by the local ethics committee [21].

Genotyping

Extraction and genotyping methods for tissue and samples have been described elsewhere [15,21,22]. *TYMS* was genotyped using RFLP methods as detailed previously elsewhere [15,22].

Table 1 lists the ten studied polymorphisms, which were chosen based on functional effects. Genotypes were determined using the fluorogenic 5′-nuclease assay (TaqMan® SNP Genotyping Assay made-to-order on an ABI 7900 HT, Applied Biosystems, CA, USA) [21].

The reaction mix was as follows: 25 µl containing 10 ng DNA, 14 µl primer/probe mix with TaqMan® Universal PCR Master mix (Applied Biosystems, CA, USA) according to the manufacturer's instructions. PCR amplification was carried out with an initial step of 95°C for 10 min followed by 40 cycles of 92°C for 15 s and 60°C for 1 min (Applied Biosystems 7900HT Sequence Detection System). The fluorescence profile of each well was measured in an Applied Biosystems 7900HT Sequence Detection System, and the results were analyzed with Sequence Detection Software (SDS 2.3, Applied Biosystems). Controls were included on each plate. Reproducibility was checked by regenotyping 10% of the samples with 100% agreement.

The greater number of failed genotyping attempts in the validation cohort is due to DNA

Table 1. Polymorphisms included in this work.

Gene (location)	rs ID	Alleles	Protein/mRNA	Functional effects
MTHFR (1p36.3)	1801133	c.665C>T (c.677C>T)	Ala222Val	Decreased activity
	1801131	c.1286A>C (c.1298A>C)	Glu429Ala	Decreased activity
DPYD (1p22)	3918290 (DPYD2a)	c.1905+1G>A	Del (exon 14)	No activity
	1801265 (DPYD9a)	c.85T>C	Cys29Arg	Decreased activity?
	2297595	c.496A>G	Met166Val	Decreased activity
	1801159 (DPYD5)	c.1627A>G	Ile543Val	Decreased activity?
OPRT (UMPS (3q13)	1801019	c.638G>C	Gly213Ala	Increased activity
TYMS (18p11.32)	45445694	CCGCGCCACTTGGCCT-GCCTCCGTCCCG 2/3/4/7/8/9 (vast majority two to three repeats)	5′-UTR VNTR 28 bp, (mRNA start position 43)	Increased expression with increasing number of repeats
	34743033†	CCGCGCCACTTCGCCTG-CCTCCGTCCCG)2/3/4	5′-UTRG>C 12th bp in VNTR repeat	C-allele: decreased transcription
	34489327	-/TTAAAG	3′-UTR 6 bp (mRNA start position 1530)	Deletion: decreased expression

†*The variable number of tandem SNP was only available for the Italian and Danish cohorts.*
DPYD: Dihydropyrimidine dehydrogenase; MTHFR: Methylenetetrahydrofolate reductase; OPRT: Orotate phosphoribosyltransferase; TYMS: Thymidylate synthase; VNTR: Variable number of tandem repeats.

being isolated from formalin-fixed paraffin-embedded tissue in contrast to the exploration cohort where DNA was isolated from peripheral blood mononuclear cells.

Phenotypic classification of enzyme activity based on functional polymorphisms

Two common *MTHFR* SNPs cause reduced enzyme activity in homozygous individuals [26–28]: the *MTHFR* 677C>T polymorphism induces an Ala-to-Val substitution in the catalytic domain (70% reduction in activity), whereas the *MTHFR* 1298A>C polymorphism induces a Glu-to-Ala substitution in a regulatory domain (30–40% reduction in activity). Compound heterozygosity leads to a 40–50% reduction in enzyme activity [27,28]. Based on the presence or absence of both variant alleles *MTHFR* activity was defined as normal or low, whereas the presence of two or more variant alleles was defined as low activity (Table 2).

Based on the variable number of tandem repeats (VNTRs), the G>C SNP in the 5'-UTR (VNTR SNP) and the 3'-UTR 6 bp insertion/deletion (ins/del) polymorphism, *TYMS* expression was classified as high, intermediate or low expression (Table 2). Studies have demonstrated that the *TYMS* VNTR and an insertion/deletion polymorphism are associated with mRNA stability and *TYMS* expression, with the three repeat and insertion alleles, respectively, being associated with increased *TYMS* expression, whereas the G>C base change in the second VNTR decreases the transcriptional activity of the *TYMS* gene [29–31].

The polymorphisms in *DPYD* were combined additively, meaning that increasing the number of variant alleles was thought to correlate with decreasing *DPYD* activity.

MDR method

The MDR method, a nonparametric method, was used to study interaction between the polymorphisms in relation to toxicity [32]. This method is implemented in the MDR software, which was used for interaction analyses (version 2.0 β 8). Patients with missing data for polymorphisms were excluded from the analysis. We assumed that the patients with beneficial or detrimental genetic profiles had a DFS much longer or much shorter than median DFS. We chose top and bottom quartiles as they can be regarded as having exceptionally long or short survival, while still containing enough patients for meaningful analysis. We carried out two separate analyses, where patients in the top or bottom quartiles for DFS were considered as affected patients in the genetic interaction analysis. The ratio between patients in the top or bottom survival quartile to the rest of the patients for each genotype combination was evaluated. Combinations with more patients in shortest or longest quartile than other combinations were considered to be associated with a high chance of short or long DFS. This procedure was carried

Table 2. Explicit definition of theoretical phenotypic consequences of different genotype combinations based on *a priori* knowledge of the functional consequences of each polymorphism.

Gene	Genotype 1	Genotype 2	Enzyme effects	Effect on 5-fluoruracil activity
MTHFR	*MTHFR677*	*MTHFR1298*		
	CCCT	AA, AC, AA	Normal activity	
Two variant alleles	TT, CT, CC	AA, AC, CC, AC, CC, CC	Low activity	Increased inhibition of TYMS
TYMS	TYMS 5'-UTR	TYMS 3'-UTR	Expression	
	2/3, 3/3, 3/3	ins/ins, ins/del	High	
	3/3, 2/3, 2/2	del/del, ins/del, ins/ins	Intermediate	
	2/2, 2/3	del/del, ins/del, del/del	Low	Increased inhibition of TYMS
DPYD (c.85T>C, c.496A>G and c.1627A>G)	Number of variant alleles 1 2 3 4 5 6		Increasing number of variant alleles associated with decreasing DPYD enzyme activity	Decreased catabolism

This table represents within gene combinations only.
del: Deletion; DPYD: Dihydropyrimidine dehydrogenase; ins: Insertion; MTHFR: Methylenetetrahydrofolate reductase; TYMS: Thymidylate synthase.

out across tenfold cross-validation samples to avoid over-fitting and was repeated for all possible combinations of two and three polymorphisms. The genotype combinations with the highest test sample accuracy and cross-validation consistency (fraction of correctly classified patients) were considered the combinations that best predicted outcome and were selected for further analysis.

Statistics

The primary end point was DFS, that is, recurrence of disease or death from disease, whatever occurred first. The secondary end point, OS, was defined as death from any cause. DFS and OS were measured from the date of intended curative resection to the date of event. Patients who were alive or had not relapsed at the last follow-up were censored at that time.

Differences in clinical variables between the two cohorts were analyzed using the chi-squared test or Mann–Whitney U test.

The associations of polymorphisms and all genetic classifiers with DFS and OS were assessed using univariable and multivariable Cox proportional hazards models. Adjustment was made for known predictive factors (age, sex, tumor grade and tumor stage) and cohort in the multivariable Cox proportional hazards models. Significance of individual factors and interaction terms were tested by log-likelihood statistics with nested models. Assumptions for the Cox analysis, that are linearity and proportional hazards, were checked for each variable using cumulative residuals; variables not fulfilling the assumptions were used as stratification variables. Survival differences between different groups were visualized using Kaplan–Meier curves.

All variables, including the genetic classifiers met the assumptions of the model except tumor stage; adjustment for tumor stage was carried out using it as a stratification variable. The basic design was that all statistical analyses were carried out on the exploration cohort and any significant associations were tested in the validation cohort.

Inferred haplotypes with more than 50% missing genotypes were set as missing data in analysis.

In the bootstrap procedure, the original set of data of size N became a parent population from which samples of size N were randomly drawn with replacement. All samples had identical stratification to the original sample in regards to the specific genotype combination. A total of 2000 bootstrap samples were created from the combined cohort and the aforementioned multivariable Cox regression was applied to each sample. Percent inclusion was used to determine the predictive importance of a variable because it was expected that an important variable would be included in the model for a majority of the bootstrap samples, defined as an inclusion rate of >65% [33–35]. The bootstrap-T interval was used to construct bootstrap confidence intervals (CIs) for the sample mean regression parameter for each genotype combination and expressed as hazard ratios (HRs).

Linkage disequilibrium between the relevant polymorphisms and Hardy–Weinberg equilibrium were calculated using Haploview v4.1. Phase 2.1.1. software for inferring individual haplotypes. All reported p-values are two-sided. All analyses were performed using the SAS Statistical Package Version 9.2 (SAS Institute Inc, NC, USA).

Results

Population characteristics & genotyping

The main differences in clinical variables between the two cohorts were in tumor characteristics (Table 3). The tumors in the validation cohort had higher stage and grade; follow-up and the numbers of events were similar between the two cohorts.

The genotype distribution of *MTHFR677*, *DPYD 85T>C* and the *TYMS* polymorphisms were different between the cohorts. *DPYD 1905 +1G>A* was excluded from the analysis due to too low allele frequency. All the genotypes in the exploration cohort were in Hardy–Weinberg equilibrium.

Genotype & haplotype associations

None of the individual polymorphisms were associated with DFS (Table 4 & Supplementary Table 1). There was linkage disequilibrium between *MTHFR677* and *MTHFR1298* and to a lesser degree between *DPYD* 85T>C and *DPYD* 496A>G and between the *TYMS* VNTR and ins/del polymorphisms (Supplementary Table 2). These haplotypes were chosen for further analysis. None of the haplotypes were associated with DFS. Diplotype, defined as a pair of haplotypes, analysis revealed that patients harboring either the *TYMS* 2R-del/2R-del or 2R-del/2R-ins diplotypes had worse DFS compared with other patients ($DFS_{exploration}$ HR: 2.65 [1.40–4.65]; p = 0.004) (Supplementary Figure 2). These two diplotypes were combined due to only one patient harboring the 2R-del/2R-del in the exploration cohort (Table 5). This finding was confirmed in the validation cohort ($DFS_{validation}$ 1.69 [1.03–2.66];

Table 3. Clinical data on the two studied cohorts.

Characteristics	Exploration cohort (n = 290)	Validation cohort (n = 302)	p-value
Age at diagnosis (median years, range)	63 (30–86)	61 (19–85)	0.07
Sex			***0.03***
Male	171 (59%)	151 (50%)	
Female	119 (41%)	151 (50%)	
Median follow-up (years)	4.8 (0.5–10.3)	5.3 (0.1–11.3)	0.16
Disease-free survival			***0.17†***
Events	142 (49%)	142 (47%)	
Censored	148 (51%)	160 (53%)	
Overall survival			***0.53†***
Events	108 (37%)	127 (42%)	
Censored	182 (63%)	175 (58%)	
Stage			***<0.0001***
B	101 (35%)	37 (12%)	
C	189 (65%)	265 (88%)	
Tumor grade			***<0.0001***
1	22 (8%)	91 (30%)	
2	228 (78%)	129 (43%)	
3	32 (11%)	82 (26%)	
Missing	8 (3%)	2 (1%)	
Tumor site			***0.15***
Colon	239 (86%)	246 (81%)	
Rectum	41 (14%)	56 (19%)	

†Based on Cox regression.

p = 0.03). None of the other diplotypes were associated with DFS. The inclusion of the *TYMS* VNTR G>C SNP was not informative with regard to defining predictive markers.

Gene–gene interactions

The functional classifications of *MTHFR*, *TYMS* or *DPYD* expression were not associated with DFS (Table 2).

In the MDR analysis all of the polymorphisms and the functional classifications of *MTHFR*, *TYMS* and *DPYD* were included as variables. As mentioned previously, the best two-way and three-way classifiers for long-term and short-term DFS were evaluated. Only a two-way genetic classifier (Figure 1), including the number of variant alleles in *DPYD* and the *TYMS* VNTR polymorphism, was associated with improved DFS in both cohorts ($DFS_{exploration}$ 0.69 [0.49–0.98]; p = 0.04; $DFS_{validation}$ 0.66 [0.45–0.95]; p = 0.03) (Supplementary Figure 2). The MDR classifier, is a dichotomous variable, where all the shaded cells in Figure 1, are considered to be the same group.

Table 4. Associations of individual polymorphisms with disease-free survival.

Exploration cohort	Exploration cohort, disease-free survival (HR [95% CI])†	p-value	Validation cohort, disease-free survival (HR [95% CI])†	p-value
MTHFR 677C>T	1.08 (0.82–1.41)	0.61		
MTHFR 1298A>C	1.00 (0.74–1.35)	0.99		
*DPYD9*a	0.68 (0.46–1.01)	0.06	1.07 (0.81–1.41)	0.65
DPYD 496A>G	1.00 (0.68–1.46)	0.98		
DPYD5	0.84 (0.58–1.25)	0.40		
UMPS 638G>C	0.65 (0.43–0.98)	0.04	0.97 (0.70–1.34)	0.86
TYMS VNTR	0.96 (0.92–1.01)	0.06	1.00 (0.98–1.02)	0.94
TYMS 6 bp ins/del	0.97 (0.72–1.32)	0.86		

†Allelic HR adjusted for age, sex, tumor grade, tumor stage and cohort.
del: Deletion; DPYD: Dihydropyrimidine dehydrogenase; ins: Insertion; HR: Hazard ratio; MTHFR: Methylenetetrahydrofolate reductase; TYMS: Thymidylate synthase; UMPS: Uridine monophosphate synthetase; VNTR: Variable number of tandem repeats.

Table 5. Associations of genotype combinations with disease-free survival adjusted for potential confounders.

Cohorts	Patients (n)	Multiple regression DFS†	
		HR (95% CI)	*p-value*
Exploration			
***TYMS* diplotype**			
–	273	1	
2/2 + del/del or ins/del	17	2.65 (1.40–4.65)	0.004
Missing	0		
MDR classifier‡			
–	135	1	
+	152	0.69 (0.49–0.98)	0.04
Missing	3		
Validation			
***TYMS* diplotype**			
–	257	1	
2/2 + del/del or ins/del	36	1.69 (1.03–2.66)	0.03
Missing	10		
MDR classifier			
–	158	1	
+	111	0.66 (0.45–0.95)	0.03
Missing	33		
Combined			
***TYMS* diplotype**			
–	530	1	
2/2 + del/del or ins/del	53	1.89 (1.29–2.70)	0.001
MDR classifier			
–	293	1	
+	263	0.68 (0.53–0.86)	0.002

†*Adjusted for age, sex, tumor grade, tumor stage and cohort.*
‡*The MDR classifier has been defined in* Figure 1.
del: Deletion; DFS: Disease-free survival; ins: Insertion; HR: Hazard ratio; MDR: Multifactor dimensionality reduction; TYMS: Thymidylate synthase.

Combined cohort & bootstrap analysis

We undertook a bootstrap analysis on the combined cohort to supplement our results. Estimates for the inclusion rate, mean HR and 95% CI for the original HR confirmed that the genotype combinations were associated with DFS (Table 6). This added further support to the external validation results detailed above; the random bootstrap samples were different from the original population in regard to nongenetic variables and only provides internal validation. The model, on which these estimates were based, was identical to the multiple Cox regression model in the ordinary analysis.

Additional analyses

The genetic classifiers did not show interactions with the cohort variable (consisting of three levels: Danish, Italian and Hungarian cohort) in the combined or exploration cohort.

The *TYMS* diplotype variable was also associated with OS in both cohorts ($OS_{exploration}$ 2.40 [95% CI: 1.24–4.24]; p = 0.005 $OS_{validation}$ 1.92 [95% CI: 1.17–3.03]; p = 0.007). The MDR classifier showed a similar tendency but was only significant for one cohort ($OS_{exploration}$ 0.67 [0.47–0.97]; p = 0.03 $OS_{validation}$ 0.70 [0.47–1.02]; p = 0.06).

Discussion

We were able to confirm the hypothesis that combinations of genotypes may create predictive variables associated with DFS and OS that are more robust than individual polymorphisms as demonstrated by the independent validation of our genotype combinations and the irreproducible associations of individual SNPs with DFS (Supplementary Table 1). We found that low-expression diplotypes in TYMS were associated with decreased DFS and OS in both cohorts with a doubling of the HR.

The MDR-derived classifier was associated with improved DFS in both cohorts with a 40% decrease in the HR compared with the reference groups.

At first, the association of low expression of *TYMS* with decreased DFS seems counterintuitive. However, several clinical studies evaluating mRNA expression or protein expression corroborate this finding. High *TYMS* expression has been associated with early disease recurrence and death in patients receiving surgery only [11,14,36–39]. Accordingly, comparisons of patients treated with surgery only and patients who also received 5-FU-based adjuvant therapy have shown that patients with high *TYMS* expression have improved outcome from 5-FU-based chemotherapy [11,36,37,39,40]. By contrast, patients with low tumoral TYMS have no improved outcome from adjuvant 5-FU therapy [36,37,39,40] or even worse outcome from such treatment [11]. Investigations on polymorphisms have shown diverging results, some indicate that high-expression alleles are associated with improved DFS or OS [15,16,41], while others have found opposite results [19,42,43]. This only applies to adjuvant treatment of colorectal cancer, as in metastatic cancer, high *TYMS* mRNA expression is associated with decreased survival [14]. The *TYMS* VNTR SNP did not enhance classification based on *TYMS* VNTRs and 3′UTR polymorphisms, which may be expected since only the VNTR and 3′UTR polymorphisms have been associated with increased *TYMS* expression in tumor tissues [29,31]. In summary, TYMS is both a prognostic and predictive marker; patients with low *TYMS* expression may not benefit substantially from 5-FU-based adjuvant treatments and have a better outcome than patients with high *TYMS* expression. The regimens used in our cohorts were bolus regimens; studies indicate that the mechanisms of action may be different depending on whether a bolus or continuous infusion regimen is used. Bolus regimens may predominantly work through perturbation of RNA function whereas continuous regimens primarily work through TYMS inhibition [44]. Thus the relevance of our findings for all regimens cannot be deduced directly from our study.

The ratio of TYMS:DPYD has been investigated in several clinical studies assuming to represent the ratio between TYMS concentration and catabolism. Low *TYMS* and *DPYD* expression together has been associated with improved DFS or OS in three studies [5,12,45–47]; a high TYMS:DPYD ratio has been associated with decreased survival in patients treated with surgery alone and in patients treated adjuvantly [12,48] and the combination of low *TYMS* and high *DPYD* expression together has been associated with decreased survival in adjuvantly treated patients in one study [49]. The presence of genetic interactions in the 5-FU metabolism pathway are evidenced by the previously referenced studies in a general way, but validation of specific genetic interactions require specific reports aimed at evaluating these specific

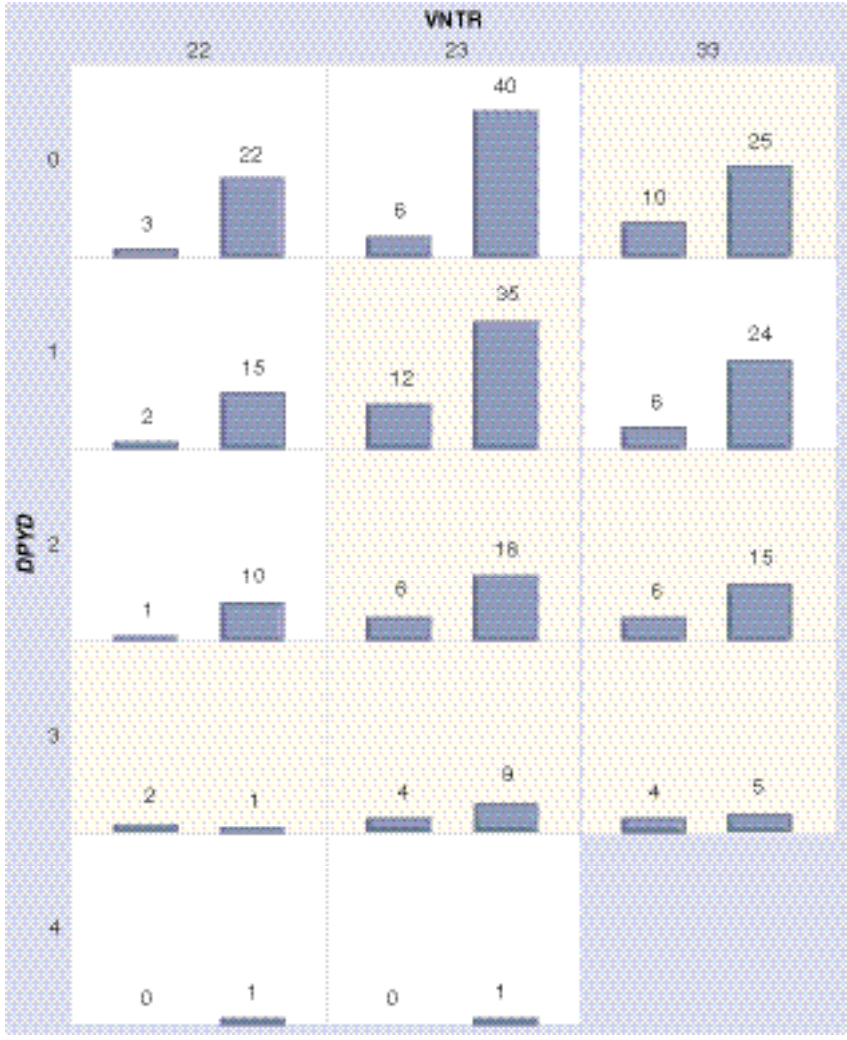

Figure 1. Genetic classifier derived from the multifactor dimensionality reduction algorithm in the exploration cohort. *DPYD* indicates the number of variant alleles when counting alleles cumulatively through the *DPYD* c.85T>C (rs1801265), c.496A>G (rs2297595) and c.1627A>G (rs1801159) polymorphisms. VNTR shows the genotypes for the thymidylate synthase VNTR polymorphism. Shaded cells represent the combinations associated with increased disease-free survival. The bars on the left are patients with disease-free survival in the top 25% and bars on the right are the patients with a disease-free survival of less than the top 25%. The mutifactor dimensionality reduction classifier is a dichotomous variable were all the shaded cells in the figure are considered to be the same group and the cells considered together as the reference group.
VNTR: Variable number of tandem repeats.

Table 6. Bootstrap estimates, based on 2000 samples with replacement compared with TABLE 5.

RFS	Inclusion	Mean HR	Bootstrap (95% CI)
TYMS diplotype	92%	1.98	1.30–2.66
MDR classifier	90%	0.67	0.53–0.86

HR: Hazard ratio; MDR: Multifactor dimensionality reduction; TYMS: Thymidylate synthase.

interactions. The associations revealed by applying the MDR algorithm are statistical interactions that need further validation in biological systems and have to be replicated in other patient populations.

The strengths in our study design include large populations, homogenous treatments and the replication of our results in an independent validation cohort. Using an exploration versus validation cohort design should minimize the detection of false or random associations.

The limitations of the study are that we cannot account for whether tumor genotypes are representative of germline genotypes in the validation cohort, however, several studies have demonstrated excellent concordance between germline and tumor tissue genotypes [50]. In our study the only indication of loss of heterogeneity is for *TYMS* genotypes, but the consequent erroneous genotype assignment would only decrease the association. The differences in frequency for the two cohorts concerning *MTHFR* 677C>T and *DPYD* 85T>C polymorphisms do not seem to be due to loss of heterozygosity as the other genotypes within these genes are similar. This discrepancy could be explained by patients being from different geographical regions or random differences. For *MTHFR*677, it has been shown that the genotype frequencies vary with geographical regions both within and outside of Europe [51]. Other possible sources of bias are differences in reference groups in the two cohorts; the composite nature of the validation cohort; differences in tumor staging and grade; and in relation to OS treatment heterogeneity after relapse.

The bootstrap procedure can, to some degree, address these latter weaknesses and provide an internal validation strategy for the combined cohort. In the bootstrap samples we had the same genotype combinations with the same number of patients in each genotype stratum as the original population, but the random samples with replacement create samples with different cohort characteristics. This gives an opportunity to test whether the associations remain stable under different cohort characteristics. The high inclusion rates of the genotype combinations and almost identical estimates compared with the original analysis support the discernment that we have identified independent predictive markers.

The biological significance of the genotype combinations cannot be directly derived from our results, but high *TYMS* expression in normal tissues may be protective against certain toxicities. Furthermore, the MDR combination suggests that decreasing DPYD activity with decreasing TYMS activity, defined by the *TYMS* VNTR SNP, is associated with higher DFS. However, as already mentioned, MDR identifies statistical interaction and not necessarily a biological interaction, so care should be applied in making an overt biological interpretation of the MDR classifier.

In conclusion, we have provided evidence for three findings. First, individual polymorphisms were not reproducibly associated with DFS and OS. Second, low-expression diplotypes in *TYMS* were associated with decreased DFS and OS. Third, specific genotype interactions in *DPYD* and *TYMS* were associated with improved DFS.

Executive summary

- The clinical efficacy of 5-fluorouracil may depend on interactions of polymorphisms rather than individual polymorphisms.

Methods

- Combinations of polymorphisms were derived from haplotype structures and the multifactor dimensionality reduction method.
- Survival analyses using disease-free survival and overall survival as outcomes were conducted.
- These associations were analyzed in an exploration cohort and confirmed in a replication cohort.

Results

- Our results suggest that low-expression diplotypes in *TYMS* and the interaction between *DPYD* allele score and the *TYMS* variable number of tandem repeats polymorphism is associated with disease-free survival and possibly overall survival in adjuvant treatment of colorectal cancer.

Conclusion

- Specific combinations of functional polymorphisms in *DPYD* and *TYMS* were demonstrated to be associated with disease-free survival and overall survival in patients receiving adjuvant 5-fluorouracil-based treatment. Specifically, high *TYMS* expression alleles seem to be associated with decreased disease-free survival.

Financial & competing interests disclosure

Shoaib Afzal has been supported by grants from Copenhagen University Faculty of Health Sciences, Danish Agency for Science, Technology and Innovation, and Dansk Kræft Forsknings Fond. Laura Bertolaso was supported by CARIPARO foundation, Rovigo-Padova, Italy. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Bibliography

Papers of special note have been highlighted as:
- of interest

1 Danenberg PV, Danenberg KD. Effect of 5, 10-methylenetetrahydrofolate on the dissociation of 5-fluoro-2'-deoxyuridylate from thymidylate synthetase: evidence for an ordered mechanism. *Biochemistry* 17(19), 4018–4024 (1978).

- **Important paper revealing the mechanics of thymidylate inhibition by 5-fluorouracil and 5,10-methylenetetrahydrofolate.**

2 Armstrong RD, Takimoto CH, Cadman EC. Fluoropyrimidine-mediated changes in small nuclear RNA. *J. Biol. Chem.* 261(1), 21–24 (1986).

3 Doong SL, Dolnick BJ. 5-fluorouracil substitution alters pre-mRNA splicing *in vitro*. *J. Biol. Chem.* 263(9), 4467–4473 (1988).

4 Houghton JA, Maroda SJ Jr, Phillips JO, Houghton PJ. Biochemical determinants of responsiveness to 5-fluorouracil and its derivatives in xenografts of human colorectal adenocarcinomas in mice. *Cancer Res.* 41(1), 144–149 (1981).

5 Beck A, Etienne MC, Cheradame S *et al.* A role for dihydropyrimidine dehydrogenase and thymidylate synthase in tumor sensitivity to fluorouracil. *Eur. J. Cancer* 30A(10), 1517–1522 (1994).

6 van Triest B, Pinedo HM, van Hensbergen Y *et al.* Thymidylate synthase level as the main predictive parameter for sensitivity to 5-fluorouracil, but not for folate-based thymidylate synthase inhibitors, in 13 nonselected colon cancer cell lines. *Clin. Cancer Res.* 5(3), 643–654 (1999).

7 Takebe N, Zhao SC, Ural AU *et al.* Retroviral transduction of human dihydropyrimidine dehydrogenase cDNA confers resistance to 5-fluorouracil in murine hematopoietic progenitor cells and human $CD34^+$-enriched peripheral blood progenitor cells. *Cancer Gene Ther.* 8(12), 966–973 (2001).

8 Etienne MC, Ilc K, Formento JL *et al.* Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphisms: relationships with 5-fluorouracil sensitivity. *Br. J. Cancer* 90(2), 526–534 (2004).

9 Sohn KJ, Croxford R, Yates Z, Lucock M, Kim YI. Effect of the methylenetetrahydrofolate reductase C677T polymorphism on chemosensitivity of colon and breast cancer cells to 5-fluorouracil and methotrexate. *J. Natl Cancer Inst.* 96(2), 134–144 (2004).

10 Sakamoto E, Nagase H, Kobunai T *et al.* Orotate phosphoribosyltransferase expression level in tumors is a potential determinant of the efficacy of 5-fluorouracil. *Biochem. Biophys. Res. Commun.* 363(1), 216–222 (2007).

11 Edler D, Glimelius B, Hallstrom M *et al.* Thymidylate synthase expression in colorectal cancer: a prognostic and predictive marker of benefit from adjuvant fluorouracil-based chemotherapy. *J. Clin. Oncol.* 20(7), 1721–1728 (2002).

12 Kornmann M, Schwabe W, Sander S *et al.* Thymidylate synthase and dihydropyrimidine dehydrogenase mRNA expression levels: predictors for survival in colorectal cancer patients receiving adjuvant 5-fluorouracil. *Clin. Cancer Res.* 9(11), 4116–4124 (2003).

13 Tsuji T, Hidaka S, Sawai T *et al.* Polymorphism in the thymidylate synthase promoter enhancer region is not an efficacious marker for tumor sensitivity to 5-fluorouracil-based oral adjuvant chemotherapy in colorectal cancer. *Clin. Cancer Res.* 9(10 Pt 1), 3700–3704 (2003).

14 Popat S, Matakidou A, Houlston RS. Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis. *J. Clin. Oncol.* 22(3), 529–536 (2004).

15 Hitre E, Budai B, Adleff V *et al.* Influence of thymidylate synthase gene polymorphisms on the survival of colorectal cancer patients receiving adjuvant 5-fluorouracil. *Pharmacogenet. Genomics* 15(10), 723–730 (2005).

16 Dotor E, Cuatrecases M, Martinez-Iniesta M *et al.* Tumor thymidylate synthase 1494del6 genotype as a prognostic factor in colorectal cancer patients receiving fluorouracil-based adjuvant treatment. *J. Clin. Oncol.* 24(10), 1603–1611 (2006).

17 Ochiai T, Nishimura K, Noguchi H *et al.* Prognostic impact of orotate phosphoribosyl transferase activity in resectable colorectal cancers treated by 5-fluorouracil-based adjuvant chemotherapy. *J. Surg. Oncol.* 94(1), 45–50 (2006).

18 Terrazzino S, Agostini M, Pucciarelli S *et al.* A haplotype of the methylenetetrahydrofolate reductase gene predicts poor tumor response in rectal cancer patients receiving preoperative chemoradiation. *Pharmacogenet. Genomics* 16(11), 817–824 (2006).

19 Lurje G, Zhang W, Yang D *et al.* Thymidylate synthase haplotype is associated with tumor recurrence in stage II and stage III colon cancer. *Pharmacogenet. Genomics* 18(2), 161–168 (2008).

20 Soong R, Shah N, Salto-Tellez M *et al.* Prognostic significance of thymidylate synthase, dihydropyrimidine dehydrogenase and thymidine phosphorylase protein expression in colorectal cancer patients treated with or without 5-fluorouracil-based chemotherapy. *Ann. Oncol.* 19(5), 915–919 (2008).

21 Afzal S, Jensen SA, Vainer B *et al. MTHFR* polymorphisms and 5-FU-based adjuvant chemotherapy in colorectal cancer. *Ann. Oncol.* 20(10), 1660–1666 (2009).

22 Gusella M, Frigo AC, Bolzonella C *et al.* Predictors of survival and toxicity in patients on adjuvant therapy with 5-fluorouracil for colorectal cancer. *Br. J. Cancer* 100(10), 1549–1557 (2009).

23 Zintzaras E, Ziogas DC, Kitsios GD, Papathanasiou AA, Lau J, Raman G. *MTHFR* gene polymorphisms and response to chemotherapy in colorectal cancer: a meta-analysis. *Pharmacogenomics* 10(8), 1285–1294 (2009).

24 Park CM, Lee WY, Chun HK *et al.* Relationship of polymorphism of the tandem repeat sequence in the thymidylate synthase gene and the survival of stage III colorectal

cancer patients receiving adjuvant 5-flurouracil-based chemotherapy. *J. Surg. Oncol.* 101(1), 22–77 (2010).

25 Afzal S, Gusella M, Vainer B *et al.* Combinations of polymorphisms in genes involved in the 5-fluorouracil metabolism pathway are associated with gastrointestinal toxicity in chemotherapy-treated colorectal cancer patients. *Clin. Cancer Res.* 17(11), 3822–3829 (2011).

26 Frosst P, Blom HJ, Milos R *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat. Genet.* 10(1), 111–113 (1995).

27 Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (*MTHFR*) associated with decreased enzyme activity. *Mol. Genet. Metab.* 64(3), 169–172 (1998).

28 van der Put NM, Gabreels F, Stevens EM *et al.* A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *Am. J. Hum. Genet.* 62(5), 1044–1051 (1998).

29 Kawakami K, Salonga D, Park JM *et al.* Different lengths of a polymorphic repeat sequence in the thymidylate synthase gene affect translational efficiency but not its gene expression. *Clin. Cancer Res.* 7(12), 4096–4101 (2001).

▪ **The first paper to show the functional consequence of the variable tandem number repeat polymorphism in the thymidylate synthethase promotor region.**

30 Mandola MV, Stoehlmacher J, Muller-Weeks S *et al.* A novel single nucleotide polymorphism within the 5′ tandem repeat polymorphism of the thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. *Cancer Res.* 63(11), 2898–2904 (2003).

▪ **The first paper to show the functional consequence of the SNP in the variable tandem number repeat in the thymidylate synthethase promotor region.**

31 Mandola MV, Stoehlmacher J, Zhang W *et al.* A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. *Pharmacogenetics* 14(5), 319–327 (2004).

▪ **The first paper to show the functional consequence of the 6 bp insertion/deletion polymorphism in the thymidylate synthethase 3′-UTR.**

32 Moore JH, Gilbert JC, Tsai CT *et al.* A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J. Theor. Biol.* 241(2), 252–261 (2006).

33 Chen CH, George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat. Med.* 4(1), 39–46 (1985).

34 Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat. Med.* 11(16), 2093–2109 (1992).

35 Altman DG, Andersen PK. Bootstrap investigation of the stability of a Cox regression model. *Stat. Med.* 8(7), 771–783 (1989).

36 Aguiar S Jr, Lopes A, Soares FA *et al.* Prognostic and predictive value of the thymidylate synthase expression in patients with non-metastatic colorectal cancer. *Eur. J. Surg. Oncol.* 31(8), 863–868 (2005).

37 Ohrling K, Edler D, Hallstrom M, Ragnhammar P, Blomgren H. Detection of thymidylate synthase expression in lymph node metastases of colorectal cancer can improve the prognostic information. *J. Clin. Oncol.* 23(24), 5628–5634 (2005).

38 Sugiyama Y, Kato T, Nakazato H *et al.* Retrospective study on thymidylate synthase as a predictor of outcome and sensitivity to adjuvant chemotherapy in patients with curatively resected colorectal cancer. *Anticancer Drugs* 13(9), 931–938 (2002).

39 Tomiak A, Vincent M, Earle CC *et al.* Thymidylate synthase expression in stage II and III colon cancer: a retrospective review. *Am. J. Clin. Oncol.* 24(6), 597–602 (2001)

40 Johnston PG, Fisher ER, Rockette HE *et al.* The role of thymidylate synthase expression in prognosis and outcome of adjuvant chemotherapy in patients with rectal cancer. *J. Clin. Oncol.* 12(12), 2640–2647 (1994).

41 Paez D, Pare L, Altes A *et al.* Thymidylate synthase germline polymorphisms in rectal cancer patients treated with neoadjuvant chemoradiotherapy based on 5-fluorouracil. *J. Cancer Res. Clin. Oncol.* 136(11), 1681–1689 (2010).

42 Fernandez-Contreras ME, Sanchez-Hernandez JJ, Guijarro M *et al.* Influence of thymidylate synthase DNA polymorphisms and gender on the clinical evolution of patients with advanced colorectal cancer. *Oncol. Rep.* 23(5), 1393–1400 (2010).

43 Iacopetta B, Grieu F, Joseph D, Elsaleh H. A polymorphism in the enhancer region of the thymidylate synthase promoter influences the survival of colorectal cancer patients treated with 5-fluorouracil. *Br. J. Cancer* 85(6), 827–830 (2001).

44 Aschele C, Sobrero A, Faderan MA, Bertino JR. Novel mechanism(s) of resistance to 5-fluorouracil in human colon cancer (HCT-8) sublines following exposure to two different clinically relevant dose schedules. *Cancer Res.* 52(7), 1855–1864 (1992).

45 Ciaparrone M, Quirino M, Schinzari G *et al.* Predictive role of thymidylate synthase, dihydropyrimidine dehydrogenase and thymidine phosphorylase expression in colorectal cancer patients receiving adjuvant 5-fluorouracil. *Oncology* 70(5), 366–377 (2006).

46 Ishibiki Y, Kitajima M, Sakamoto K, Tomiki Y, Sakamoto S, Kamano T. Intratumoral thymidylate synthase and dihydropyrimidine dehydrogenase activities are good predictors of 5-fluorouracil sensitivity in colorectal cancer. *J. Int. Med. Res.* 31(3), 181–187 (2003).

47 Salonga D, Danenberg KD, Johnson M *et al.* Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clin. Cancer Res.* 6(4), 1322–1327 (2000).

48 Lassmann S, Hennig M, Rosenberg R *et al.* Thymidine phosphorylase, dihydropyrimidine dehydrogenase and thymidylate synthase mRNA expression in primary colorectal tumors – correlation to tumor histopathology and clinical follow-up. *Int. J. Colorectal Dis.* 21(3), 238–247 (2006).

49 Jensen SA, Vainer B, Sorensen JB. The prognostic significance of thymidylate synthase and dihydropyrimidine dehydrogenase in colorectal cancer of 303 patients adjuvantly treated with 5-fluorouracil. *Int. J. Cancer* 120(3), 694–701 (2007).

50 McWhinney SR , McLeod HL. Using germline genotype in cancer pharmacogenetic studies. *Pharmacogenomics* 10(3), 489–493 (2009).

51 Wilcken B, Bamforth F, Li Z *et al.* Geographical and ethnic variation of the 677C>T allele of 5,10 methylenetetrahydrofolate reductase (*MTHFR*): findings from over 7000 newborns from 16 areas world wide. *J. Med. Genet.* 40(8), 619–625 (2003).

Affiliations

▪ **Shoaib Afzal**
Rigshospitalet, Copenhagen, Denmark
and
Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
and
Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

- **Milena Gusella**
  Laboratory of Pharmacology & Molecular Biology, Oncology Department, Azienda-ULSS 18-Rovigo, Trecenta, Italy
- **Søren Astrup Jensen**
  Rigshospitalet, Copenhagen, Denmark
- **Ben Vainer**
  Rigshospitalet, Copenhagen, Denmark
- **Ulla Vogel**
  National Research Centre for the Working Environment, Copenhagen, Denmark & Institute for Science, Systems & Models, University of Roskilde, Denmark
- **Jon Trærup Andersen**
  Rigshospitalet, Copenhagen, Denmark
  and
  Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
- **Kasper Brødbæk**
  Rigshospitalet, Copenhagen, Denmark
  and
  Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
- **Morten Petersen**
  Rigshospitalet, Copenhagen, Denmark
  and
  Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
- **Espen Jimenez-Solem**
  Rigshospitalet, Copenhagen, Denmark
  and
  Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
- **Vilmos Adleff**
  National Institute of Oncology, Budapest, Hungary
- **Barna Budai**
  National Institute of Oncology, Budapest, Hungary
- **Erika Hitre**
  National Institute of Oncology, Budapest, Hungary
- **István Láng**
  National Institute of Oncology, Budapest, Hungary
- **Eniko Orosz**
  National Institute of Oncology, Budapest, Hungary
- **Laura Bertolaso**
  Laboratory of Pharmacology & Molecular Biology, Oncology Department, Azienda-ULSS 18-Rovigo, Trecenta, Italy
- **Carmen Barile**
  Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark
- **Roberto Padrini**
  Department of Clinical & Experimental Medicine, University of Padova, Padova, Italy
- **Judit Kralovánszky**
  National Institute of Oncology, Budapest, Hungary
- **Felice Pasini**
  Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark
- **Henrik Enghusen Poulsen**
  Rigshospitalet, Copenhagen, Denmark
  and
  Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark
  and
  Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark