Carcinogenesis vol.30 no.6 pp.982–986, 2009
doi:10.1093/carcin/bgp086
Advance Access publication April 8, 2009

Association of common genetic variants in SMAD7 and risk of colon cancer

Cheryl L.Thompson[1,2,3,7], Sarah J.Plummer[4], Louise S.Acheson[1,3,5], Thomas C.Tucker[6], Graham Casey[4] and Li Li[1,2,3,7,*]

[1]Department of Family Medicine, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, OH 44106, USA, [2]Case Center for Transdisciplinary Research on Energetics and Cancer, [3]Case Comprehensive Cancer Center, Cleveland, OH 44106, USA, [4]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA, [5]Department of Reproductive Biology, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, OH 44106, USA, [6]Cancer Control Program, Markey Cancer Center, University of Kentucky, Lexington, KY 40504, USA and [7]Department of Epidemiology and Biostatistics, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, OH 44106, USA

[*]To whom correspondence should be addressed. Department of Family Medicine, Case Western Reserve University, 11001 Cedar Avenue, Suite 306, Cleveland, OH 44106, USA. Tel: +1 216 368 5437; Fax: +1 216 368 4348; Email: li.li@case.edu

Two recent genome-wide association studies (GWAS) identified three common variants in SMAD7 (rs4464148, rs4939827 and rs12953717) that confer modest susceptibility to colorectal cancer. Here, we replicated the association of rs4464148 with colon cancer in a population-based case–control study (561 cases and 721 controls). Compared with the TT genotype, those with CT and CC had an adjusted odds ratio (OR) and 95% confidence interval of 1.06 (0.82–1.38) and 1.86 (1.17–2.96), respectively ($P_{trend} = 0.04$). However, stratified analyses revealed that this association was limited to women only [OR = 1.25 (0.88–1.78) for CT and OR = 2.76 (1.53–4.98) for CC, $P_{trend} = 0.002$, $P_{interaction} = 0.08$], which was not noted in any GWAS. Similarly, we found evidence for association with both rs4939827 and rs12953717 in women only ($P = 0.007$ in dominant rs4939827 model and $P = 0.015$ in recessive rs12953717 model), but not in men ($P > 0.05$) and evidence of an interaction with gender ($P = 0.015$ for rs4939827 and $P = 0.061$ for rs12953717). Similar effect modification was found in haplotype analyses. Our data add evidence supporting these genetic variants as markers predisposing to colon cancer, specifically in women.

Introduction

The transforming growth factor beta (TGF-β) signaling pathway plays an important role in cancer initiation and progression (1). This pathway regulates inflammation and exhibits tumor suppressor properties in the early stages of tumorigenesis and pro-oncogenic properties in later stages (2,3). It has been reported that increased TGF-β1 expression correlates with tumor progression and recurrence in colorectal cancer (4). The importance of the TGF-β pathway in colorectal cancer has also been shown through the discovery of somatic mutations in *TGFBR2, SMAD2* and *SMAD4* as well as the association of a germ line variant in *TGFBR1* (4).

SMAD7 is an inhibitory SMAD and a negative regulator of the TGF-β signaling pathway that promotes the anti-inflammatory effects of TGF-β signaling via binding to TAB2 and TAB3 and inhibiting TAK1 (5). Although SMAD7 has been shown to induce hepatic metastasis in colorectal cancer (6), its role in cancer development, particularly colorectal cancer, has not been fully explored.

Several genetic variants within SMAD7, located on chromosome 18, have recently been reported to be associated with colorectal cancer in two genome-wide association studies (GWAS) (7,8). In both studies, a highly significant association with colorectal neoplasia was found for a single nucleotide polymorphism (SNP) in intron 3 of *SMAD7*, rs4939827, and a nearby, intronic SNP rs12953717. Broderick *et al.* (7) additionally found a significant association for intronic SNP rs4464148. We reported here associations of these SNPs with risk of colon cancer in a population-based case–control study and further explored potential effect modification by age, gender and family history of colorectal cancer.

Abbreviations: CI, confidence interval; GWAS, genome-wide association studies; NSAID, non-steroidal anti-inflammatory drug; OR, odds ratio; SNP, single nucleotide polymorphism; TGF-β, transforming growth factor beta.

Materials and methods

Study population

The details of study design and data collection methods have been described elsewhere (9). Briefly, 561 incident colon cancer cases diagnosed within 6 months prior to recruitment were systematically enrolled through the population-based Surveillance, Epidemiology and End Results Kentucky Cancer Registry and 721 population controls were recruited from the state of Kentucky through random-digit dialing. Controls were required to be 30 years or older, free of known cancer (except non-melanoma skin cancer), inflammatory bowel diseases, family history of familial adenomatous polyposis and hereditary non-polyposis colorectal cancer. We generated a list of four-digit random numbers and combined them with the area codes and prefix (first three digits of phone number) of the cases and systematically dialed these numbers to recruit controls representative of the general population of Kentucky rather than matched to the cases.

After informed consent, the subjects were arranged to go to a nearby medical facility for blood draw after overnight fasting. The blood samples were shipped overnight to the research laboratory at Case Western Reserve University with a frozen ice pack and immediately processed and stored frozen at −80°C until DNA extraction. Each participant then received a risk factor questionnaire developed by the National Cancer Institute Colon Cancer Familial Cancer Registry (http://epi.grants.cancer.gov/documents/CFR/center_questionnaires/Colon/LA/ColonRiskFactor_USC.pdf) to record detailed information on family history of colorectal cancer, lifestyle and behavioral risk factors. The response rates were 72.2% for the cases and 62.5% for eligible controls. The study was approved by the Institutional Review Boards of the University of Kentucky, Lexington, and Case Western Reserve University/University Hospitals of Cleveland.

We defined a positive family history of colorectal cancer when the participant reported colorectal cancer in one or more first-degree relatives on the risk factor questionnaire. Age was defined as age at colon cancer diagnosis for cases and age at recruitment for controls. Body mass index was calculated based on self-reported current weight (kg) divided by height in meters squared (kg/m^2). Regular non-steroidal anti-inflammatory drug (NSAID) use was defined as self-reported use of ibuprofen or aspirin at least twice a week for 6 months or longer.

Genotyping

Genomic DNA was extracted from frozen buffy coat aliquots using the Biorobot EZ1 (Qiagen, Valencia, CA) and quantitated using the Quant-It picogreen kit (Invitrogen, Carlsbad, CA). The Taqman allelic discrimination assay was used for genotyping. Assays were performed in 384-well plates with 1.25 ng of genomic DNA, specific primer/probe set and RealMasterMix Probe + ROX (5 Prime) according to the manufacturer's instructions. Predesigned primer/probe sets were used for rs4939827 (C_27913406_10), rs4464148 (C_27989234_10) and a custom designed set for rs12953717 (Applied Biosystems, Carlsbad, CA) (sequence for custom set provided on request). The 7900HT Sequence Detection System with SDS 2.2 software from Applied Biosystems was used to read the assays and assign genotypes. The no call rate for rs4939827 was 21 (1.6%), 2 for rs4464148 (0.16%) and 3 for rs12953717 (0.23%). For quality assurance, laboratory personnel were blinded to the case–control status of all samples, and two percent of the samples were independently re-genotyped. The concordance call rate was 100% in our study.

Statistical methods

We evaluated the association for each of the *SMAD7* genotypes and haplotypes using unconditional logistic regression models under unrestricted, additive, dominant and recessive genetic modes of inheritance. In all analyses, the lower frequency allele was coded as the 'risk' allele. For the additive model, individuals were assigned a 0, 1 or 2 representing the number of risk alleles they possessed for that SNP. For the dominant model, individuals were coded as 1 if

they carried at least 1 risk allele and 0 otherwise; for the recessive model, individuals were coded as 1 if they were homozygous for the risk allele (two copies) and 0 otherwise.

Since we do not 1:1 match our controls to cases, and, on average, our controls were slightly younger than the cases, we statistically adjusted for age in our base models. In our full models, we additionally controlled for sex, race, family history of colorectal cancer, body mass index and NSAID use.

We further explored potential effect modification by age, stratified on the median age of the cases ($\leq$65 years or $>$65 years), as well as gender, and family history of colorectal cancer with each SNP. For each effect modification, we added the main effect of the best fitting (lowest *P*-value) SNP model (additive, dominant or recessive) and the categorical effect modifier (young versus old, male versus female, positive versus negative family history of colorectal cancer) as well as a multiplicative term of these two variables to the logistic regression.

Due to the physical proximity of the SNPs, we also inferred haplotypes using DECIPHER (10), which implements a maximum likelihood method to estimate the most likely haplotypes for each individual. Since over 98% of the participants had a probability of 95% or higher for each of the inferred haplotypes, we chose the most likely combination of haplotypes for each individual. Four haplotypes had a population frequency of 5% or higher and were included in haplotype association analyses. We created a variable corresponding to the number of copies (0, 1 or 2) of each of the four haplotypes inferred for each individual and used them in the logistic regression analyses.

Statistical significance was assessed via both Wald test and likelihood ratio test comparing full and reduced models (i.e. with and without the cross-product term). All *P*-values reported here are two sided. All analyses were undertaken using SAS software (version 9.1; SAS Institute, Cary, NC). To account for potential bias due to population stratification, we also repeated the analyses after restricting to Caucasians (93.9% of the sample).

Results

All three SNPs conformed to Hardy–Weinberg proportions in the controls ($P > 0.1$). The majority of our study sample are Caucasians (93.9%), consistent with the general population of Kentucky. Association results were very similar when using the entire sample or when restricting to Caucasians only. For brevity, all results reported in the tables here are for the entire study population. Table I summarizes the descriptive characteristics and allele frequencies.

The rs4464148 CC genotype was strongly associated with colon cancer in both crude and adjusted analyses (Table II). The odds ratio (OR) estimates we observed were very similar to those reported by Broderick *et al.* (6). However, our stratified analyses revealed that this association was largely limited to women in a monotonic, gene–dose response manner ($P_{trend} = 2.2 \times 10^{-3}$), with an almost 3-fold increase of risk for women homozygous for the C allele (OR = 2.76, CI = 1.53–4.98). The recessive model showed an approximately equal fit ($P = 1.7 \times 10^{-3}$). A test for interaction of rs4464148 (using the best fitting recessive genetic model) with gender was marginally significant ($P = 0.081$ in the full model). The results were very similar in the Caucasian-only analyses ($P_{trend} = 2.3 \times 10^{-3}$ in women; OR = 2.87, CI = 1.57–5.26 for additive model, $P_{recessive} = 1.3 \times 10^{-3}$; $P_{interaction} = 0.055$). This gender-specific effect was not reported in the original GWAS (6).

In contrast to the two GWAS analyses, we found no association of rs4939827 with colon cancer in our overall study population nor when limited to Caucasians (data not shown). However, as with rs4464148, we observed a substantial gender difference in disease association in stratified analyses. In women, the rs4939827 C allele was statistically significantly associated with a decreased risk of colon cancer in a gene–dose response manner ($P_{trend} = 0.041$). This is consistent with the overall analyses from both GWAS (7,8). In contrast, a statistically significant increase of risk was observed in the crude analysis in men; further adjustment for other covariates reduced the OR to non-significance. Test for interaction revealed significant effect modification by gender ($P_{interaction} = 0.015$ using the best fitting dominant genetic model). As with the other SNPs, we obtained very similar results in the Caucasian-only analyses ($P_{interaction} = 0.012$). Tenesa *et al.* (8) reported no evidence for such a gender differential effect in their GWAS analysis for this SNP.

For rs12953717, we again observed an appreciable gender difference that was not observed by others (8), with statistically significant increased risk among women homozygous for the A allele, but non-significant decrease of risk among men. The recessive genetic model was the best fitting model, particularly in the women ($P = 0.015$), and test for interaction suggested potential effect modification by gender ($P_{interaction} = 0.061$ in the entire sample; $P_{interaction} = 0.040$ in Caucasians only). This is in contrast to the additive models suggested by both Broderick *et al.* (7) and Tenesa *et al.* (8). The gender-specific effect was not reported for this SNP in either original GWAS (7,8).

None of these three SNPs showed evidence of interaction with age or family history of colorectal cancer (data not shown).

The three SNPs are physically close and are highly correlated in our sample ($D' = 0.92$ between rs4939827 and rs12953717, $D' = 0.83$ between rs12953717 and rs4464148 and $D' = 0.79$ between rs4939827 and rs4464148). We estimated four haplotypes with a frequency $>$5% in our study population (Table III). Consistent with our SNP analyses, one rs4939827-rs12953717-rs4464148 haplotype

Table I. Descriptive characteristics of Kentucky colon cancer study population

	Controls ($n = 721$)	Cases ($n = 561$)	*P*
Mean age (SD)	58.9 (10.9)	64.1 (10.8)	$<1 \times 10^{-4}$
Sex (%)			7×10^{-4}
Female	447 (62.0)	295 (52.6)	
Male	274 (38.0)	266 (47.4)	
Race (%)			0.24
African-American	23 (3.2)	27 (4.8)	
Caucasian	678 (94.7)	526 (93.8)	
Other	15 (2.1)	8 (1.4)	
Regular NSAID use (%)[a]			0.16
Regular	445 (61.7)	309 (55.1)	
Never	213 (29.5)	187 (33.3)	
Missing	63 (8.7)	65 (11.6)	
Mean BMI (kg/m^2) (SD)[b]	28.3 (6.2)	29.3 (6.1)	7.7×10^{-3}
Family history of colorectal cancer (%)[c]			3×10^{-4}
Yes	166 (23.0)	175 (34.8)	
No	494 (68.5)		
Missing	61 (8.5)		
rs4939827 (%)			0.42
CC	146 (20.6)	125 (22.6)	
CT	378 (53.3)	275 (49.6)	
TT	185 (26.1)	154 (27.8)	
MAF (C allele)			
Entire population	0.47	0.47	
Males	0.45	0.51	
Females	0.49	0.44	
rs4464148 (%)			0.066
CC	53 (7.4)	61 (10.9)	
CT	324 (45.0)	231 (41.2)	
TT	342 (47.6)	269 (48.0)	
MAF (C allele)			
Entire population	0.30	0.31	
Males	0.31	0.29	
Females	0.29	0.34	
rs12953717 (%)			0.039
AA	129 (17.9)	116 (20.7)	
AG	370 (51.5)	248 (44.3)	
GG	22 (30.6)	196 (35.0)	
MAF (G allele)			
Entire population	0.44	0.43	
Males	0.46	0.40	
Females	0.43	0.46	

MAF, minor allele frequency; BMI, body mass index.
[a]Regular NSAID use was defined as reporting ever use of NSAIDs at least twice a week for at least 1 month.
[b]BMI was calculated based on self-reported current weight and height (kg/m^2). Seventy cases and 69 controls were missing on weight or height.
[c]A positive family history of colorectal cancer was defined as self-report of colorectal cancer in $\geq$1 first-degree relative.

Table II. Logistic regression analyses of SMAD7 SNPs and colon cancer

Genotype	Cases (*n*)	Controls (*n*)	Crude		Age adjusted		Full model	
			OR (95% CI)[a]	*P*	OR (95% CI)[b]	*P*	OR (95% CI)	*P*
rs4464148								
CC	61	53	1.46 (0.98–2.19)	0.39[c]	1.58 (1.04–2.41)	0.21[c]	1.86 (1.17–2.96)[d]	0.039[c]
CT	231	324	0.91 (0.72–1.14)		0.95 (0.75–1.21)		1.06 (0.82–1.38)	
TT	269	342	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	61 versus 500	53 versus 666	1.55 (1.05–2.29)	0.026	1.61 (1.07–2.42)	0.022	1.81 (1.16–2.83)	8.8×10^{-3}
CC/CT versus TT	292 versus 269	377 versus 342	0.99 (0.80–1.24)	0.95	1.04 (0.82–1.30)	0.77	1.16 (0.90–1.49)	0.24
Males								
CC	23	20	1.04 (0.55–1.99)	0.32[c]	1.11 (0.56–2.20)	0.45[c]	0.98 (0.47–2.08)[e]	0.62[c]
CT	107	130	0.75 (0.53–1.07)		0.79 (0.55–1.14)		0.85 (0.57–1.27)	
TT	136	124	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	23 versus 243	20 versus 254	0.99 (0.50–1.97)	0.97	1.05 (0.51–2.13)	0.90	1.06 (0.51–2.19)	0.87
CC/CT versus TT	130 versus 136	150 versus 124	0.80 (0.56–1.15)	0.22	0.83 (0.57–1.20)	0.31	0.87 (0.59–1.28)	0.48
Females								
CC	38	33	1.95 (1.16–3.27)	0.022[c]	2.04 (1.19–3.49)	0.015[c]	2.76 (1.53–4.98)[e]	2.2×10^{-3}[c]
CT	124	194	1.06 (0.78–1.45)		1.10 (0.80–1.52)		1.25 (0.88–1.78)	
TT	133	218	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	38 versus 257	33 versus 412	1.98 (1.17–3.37)	0.011	2.03 (1.17–3.50)	0.011	2.47 (1.41–4.34)	1.7×10^{-3}
CC/CT versus TT	162 versus 133	227 versus 218	1.27 (0.93–1.74)	0.14	1.32 (0.96–1.82)	0.093	1.44 (1.03–2.01)	0.033
$P_{interaction}$					**0.15**[f]		**0.081**[f]	
rs4939827								
CC	125	146	1.02 (0.74–1.41)	0.98[c]	0.97 (0.70–1.36)	0.73[c]	0.91 (0.63–1.30)[d]	0.50[c]
CT	275	378	0.90 (0.67–1.13)		0.85 (0.65–1.12)		0.78 (0.58–1.05)	
TT	154	185	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	125 versus 429	146 versus 563	1.13 (0.85–1.52)	0.39	1.08 (0.80–1.45)	0.62	1.07 (0.79–1.45)	0.68
CC/CT versus TT	400 versus 154	524 versus 185	0.90 (0.69–1.17)	0.41	0.87 (0.66–1.15)	0.32	0.82 (0.62–1.08)	0.16
Males								
CC	66	48	1.80 (1.10–2.97)	0.029[c]	1.57 (0.94–2.64)	0.089[c]	1.49 (0.85–2.62)[e]	0.17[c]
CT	136	142	1.27 (0.85–1.92)		1.18 (0.77–1.81)		1.13 (0.72–2.62)	
TT	61	80	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	66 versus 197	48 versus 222	1.53 (0.97–2.40)	0.067	1.41 (0.89–2.25)	0.15	1.37 (0.85–2.21)	0.19
CC/CT versus TT	202 versus 61	190 versus 80	1.43 (0.95–2.16)	0.089	1.32 (0.87–2.03)	0.19	1.22 (0.79–1.88)	0.37
Females								
CC	59	98	0.67 (0.44–1.03)	0.058[c]	0.68 (0.44–1.05)	0.058[c]	0.63 (0.39–1.02)[e]	0.041[c]
CT	139	236	0.65 (0.46–0.93)		0.67 (0.46–0.96)		0.59 (0.40–0.88)	
TT	93	105	1.0 (referent)		1.0 (referent)		1.0 (referent)	
CC versus CT/TT	59 versus 232	98 versus 341	0.92 (0.63–1.36)	0.68	0.89 (0.60–1.33)	0.58	0.89 (0.59–1.33)	0.56
CC/CT versus TT	198 versus 93	334 versus 105	0.63 (0.45–0.90)	0.010	0.64 (0.45–0.92)	0.015	0.60 (0.42–0.88)	7.7×10^{-3}
$P_{interaction}$					**0.010**[f]		**0.015**[f]	
rs12953717								
AA	116	129	1.01 (0.74–1.39)	0.74[c]	1.04 (0.75–1.45)	0.86[c]	1.14 (0.80–1.64)[d]	0.70[c]
AG	248	370	0.75 (0.59–0.97)		0.74 (0.57–0.96)		0.83 (0.63–1.11)	
GG	196	220	1.0 (referent)		1.0 (referent)		1.0 (referent)	
AA versus AG/GG	116 versus 444	129 versus 590	1.15 (0.85–1.55)	0.36	1.18 (0.87–1.61)	0.29	1.28 (0.93–1.75)	0.13
AA/AG versus GG	364 versus 196	499 versus 220	0.85 (0.66–1.09)	0.20	0.86 (0.67–1.11)	0.25	0.91 (0.70–1.19)	0.49
Males								
AA	49	53	0.68 (0.42–1.11)	0.047[c]	0.74 (0.44–1.22)	0.083[c]	0.70 (0.40–1.22)[e]	0.15[c]
AG	114	144	0.59 (0.40–0.87)		0.61 (0.41–0.91)		0.70 (0.45–1.08)	
GG	103	77	1.0 (referent)		1.0 (referent)		1.0 (referent)	
AA versus AG/GG	49 versus 217	53 versus 221	0.83 (0.52–1.31)	0.42	0.86 (0.53–1.38)	0.53	0.88 (0.54–1.42)	0.59
AA/AG versus GG	163 versus 103	197 versus 77	0.64 (0.43–0.94)	0.021	0.66 (0.44–0.98)	0.039	0.70 (0.46–1.05)	0.085
Females								
AA	67	76	1.38 (0.91–2.11)	0.18[c]	1.37 (0.89–2.13)	0.18[c]	1.62 (1.00–2.62)[e]	0.094[c]
AG	134	226	0.91 (0.65–1.27)		0.86 (0.61–1.22)		0.94 (0.64–1.37)	
GG	93	143	1.0 (referent)		1.0 (referent)		1.0 (referent)	
AA versus AG/GG	67 versus 227	76 versus 369	1.47 (0.99–2.17)	0.055	1.49 (1.00–2.24)	0.052	1.69 (1.11–2.57)	0.015
AA/AG versus GG	201 versus 93	302 versus 143	1.07 (0.77–1.50)	0.68	1.07 (0.76–1.51)	0.71	1.09 (0.76–1.56)	0.64
$P_{interaction}$					**0.082**[f]		**0.061**[f]	

[a]Crude OR and 95% CI estimates for genotype effect on entire sample (561 cases and 721 controls).
[b]OR and 95% CI estimates for unconstrained genotype effect adjusted for age on those with available data (554 cases and 704 controls).
[c]*P* for trend (additive model).
[d]OR and 95% CI estimates and *P*-value for unconstrained genotype effect adjusted for age, race, gender, family history of colorectal cancer, body mass index and NSAID use (479 cases and 640 controls with data available).
[e]OR and 95% CI estimates and *P*-value for unconstrained genotype effect adjusted for age, race, family history of colorectal cancer, body mass index and NSAID use (479 cases and 640 controls with data available).
[f]*P* for interaction between variable and genotype via a likelihood ratio test comparing the models with and without the interaction term from best fitting model (recessive for rs4464148, dominant for rs4939827 and recessive for rs12953717).

Table III. SMAD7 haplotypes and risk of colon cancer

Haplotype	Estimated frequency (%)	OR (95% CI)[a]		*P* for trend
		Heterozygous (one copy)	Homozygous (two copies)	
All subjects				
C-G-T	43.8	0.82 (0.62–1.09)	0.87 (0.60–1.27)	0.35
T-A-C	27.0	0.99 (0.76–1.28)	2.23 (1.32–3.78)	0.060
T-A-T	14.3	0.91 (0.68–1.22)	0.41 (0.16–1.09)	0.14
T-G-T	9.8	0.86 (0.62–1.20)	1.86 (0.63–5.49)	0.84
Males				
C-G-T	44.0	1.08 (0.70–1.66)	1.26 (0.72–2.24)	0.44
T-A-C	26.4	0.81 (0.54–1.20)	0.76 (0.33–1.76)	0.27
T-A-T	14.4	0.85 (0.55–1.32)	0.54 (0.13–2.26)	0.31
T-G-T	10.1	0.98 (0.59–1.62)	2.30 (0.43–12.19)	0.68
Females				
C-G-T	43.6	0.67 (0.46–0.98)	0.68 (0.41–1.11)	0.069
T-A-C	27.4	1.16 (0.82–1.66)	4.49 (2.26–8.95)	6×10^{-4}
T-A-T	14.2	0.94 (0.64–1.40)	0.33 (0.09–1.25)	0.26
T-G-T	9.5	0.77 (0.49–1.20)	1.95 (0.46–8.33)	0.58

[a]OR and 95% CI estimates and *P*-value for unconstrained genotype effect adjusted for age, race, gender, family history of colorectal cancer, body mass index and NSAID use (479 cases and 640 controls with data available), compared with no copies of that haplotype.

(T-A-C) showed strong association with colon cancer in women, with an estimated 60% increase of risk per copy of the haplotype [full model OR = 1.60 (1.22–2.10), $P = 6 \times 10^{-4}$], but not in men [OR = 0.84 (0.61–1.15), $P = 0.27$]. A test for interaction supported a significant effect modification by gender ($P_{\text{interaction}} = 5.5 \times 10^{-3}$ in the full model). Similarly for the C-G-T haplotype, there is suggestive evidence for a gender differential effect with a borderline statistically significant inverse association in women ($P = 0.069$), but not in men ($P = 0.44$) ($P_{\text{interaction}} = 0.069$ in the full model).

Discussion

Due to the potential for false positives resulting from the large number of tests, replication of GWAS findings in independent study populations is an extremely important step of disease susceptibility gene discovery (11,12). Here, we present further evidence of SNPs in SMAD7 being risk loci for colon cancer in a relatively large population-based study sample and further note a substantial differential effect by gender that was not observed in the original GWAS (8).

We noted an appreciable increase in OR estimates, particularly for the rs4464148 SNP, when we further adjusted for other covariates (full model) in addition to age (Table II), suggesting confounding by one or more of the variables included in the full model. Indeed, analyses of associations of the SNPs with these additionally adjusted covariates in the controls revealed several significant correlations. Not unexpectedly, we found significant genotype frequency differences for rs12953717 ($P = 5.5 \times 10^{-3}$) and borderline significant genotype frequency differences for rs4464148 ($P = 0.080$) across the race groups, but did not note a difference in rs4939827 genotype frequencies between races ($P = 0.79$). We also found evidence for association of NSAID use with rs4464148 ($P = 0.048$) and rs12953717 ($P = 0.046$), but not with rs4939827 ($P = 0.38$). As such, we further adjusted for these covariates in our full models to statistically control for potential confounding and hence more accurate estimates of the OR. We did not find differences in Body mass index or family history by genotype ($P > 0.1$).

It is important to note that we did have a higher than usual no call rate for rs4939827 (1.6%). However, this no call rate is still fairly low and is consistent across cases (1.2%) and controls (1.7%) as well as across males (1.3%) and females (1.6%); thus, we do not expect this to significantly affect our results.

Although the mechanisms underlying possible gender-specific effects are unclear, SMAD7 is an intracellular TGF-β type 1 receptor antagonist, thereby blocking the TGF-β1 signaling pathway (13,14). The TGF-β1 signaling pathway functions as both a tumor suppressor in early stage cancers as well as an oncogene in advanced cancers and metastasis (2,3) and this pathway is well known to be influenced by sex steroid hormones. Testosterone decreases TGF-β secretion in rats (15). Estrogen increases TGF-β mRNA expression in mouse osteoblasts and osteosarcoma cells (16), whereas estradiol treatment decreases (17) or does not change (18) TGF-β2 and TGF-β3 mRNA levels in breast cancer cell lines. While estradiol did not increase TGF-β expression levels in prostate carcinoma cell lines, an increase in TGF-β secretion was observed (18). SMAD7 expression is increased and TGF-β signaling inhibited by gonadotropin-releasing hormone agonists in both myometrium (19) and endometrium (20). If gonadotropin-releasing hormone or gonadal hormones influence SMAD7 expression in other tissues as well, then this is a possible mechanism for gender-specific effects of SMAD7 variants as we observed here. Further evidence for the potential for gender-specific effects comes from a study by Dixon and Maric (21) in which 17β-estradiol supplementation attenuated the decrease in SMAD7 signaling associated with diabetes in rats. It is thus conceivable that the SMAD7 genetic variants studied here may indeed exert gender differential effects on colon cancer development, although the functionality of these SNPs remains largely unknown at present.

One may also speculate that gender-specific effects that we observed may be due to the existence of some unmeasured environmental factors that are correlated with SNPs under investigation and differ between males and females in our study population, but not in the population of the study by Tenesa *et al.* (8).

The TGF-β1 pathway acts in a cell type-dependent manner which further complicates predictions of its role in a given tumor type (2). SMAD7 has been shown to be functionally involved in intestinal inflammation through TGF-β signaling (13) and to be amplified in colon cancers with poor prognosis (14). Broderick *et al.* (7) observed that the risk alleles of rs12953717 and of rs4464148 were associated with lower SMAD7 mRNA expression in lymphoblastoid cell lines. Although opposite the expected result, this may reflect the effect of SMAD7 on other signaling pathways such as the Wnt pathway (7,22) or simply that expression in the cell lines is not indicative of expression in colon cells. Thus, although the mechanism is not clear, the relevance of SMAD7 to colon cancer suggests that the observed genetic variations (or unknown causal variants in linkage disequilibrium (LD) with those reported here) are likely to affect colon cancer risk.

Confirmation of disease–genotype associations found in genome-wide scans in independent populations provides strong evidence that the association is robust (11). Although we did not replicate association with colon cancer for all three SMAD7 variants in our entire study population, we found evidence of the associations of all three SNPs to colon cancer among women. Indeed, we found an opposite, although statistically insignificant, effect of the rs4939827 SNP in men. It is possible the lack of statistical significance in the men is due to the small sample size (284 cases and 266 controls) when we limit the analyses to males. Caution must be exercised in the interpretation of the gender-specific effects observed in our study population. Nevertheless, the consistency of the direction and magnitude of the associations for all three SNPs in women in our current analyses with that reported by the GWAS, and our previous replication of the GWAS association of an 8q24 SNP (rs6983267) with colon cancer risk in our study population support the validity of our observed associations (23). Moreover, our findings of a protective effect of NSAID use (OR = 0.78, 95% CI = 0.61–1.00, $P = 0.05$) and a positive association with family history of colorectal cancer (OR = 1.61, 95% CI = 1.25–2.09, $P < 0.001$) are in agreement with their well-documented associations with colon cancer (24), lending credibility to our results. These data provide further evidence that common genetic variants in SMAD7 may confer susceptibility to colon cancer, particularly among women. More research is warranted to confirm these findings and functionally characterize the SMAD7 variants. All the three SNPs are all intronic, and if they are indeed the causal

variants, their function remains to be elucidated. Furthermore, the gender differential effect is an interesting avenue for future work.

Funding

National Cancer Institute (U54 CA116867, K22 CA120545, R25T CA094186); Damon-Runyon Cancer Research Foundation, Clinical Investigator Award CI-8; U.S. Public Health Service Resource, National Center for Research Resources (RR03655).

Acknowledgements

Some of the results of this paper were obtained by using the program package S.A.G.E.

Conflicts of Interest Statement: None declared.

References

1. Siegel,P.M. *et al*. (2003) Cytostatic and apoptotic actions of TGF-β in homeostasis and cancer. *Nat. Rev. Cancer*, **3**, 807–821.
2. Roberts,A.B. *et al*. (2003) The two faces of transforming growth factor β in carcinogenesis. *Proc. Natl Acad. Sci. USA*, **100**, 8621–8623.
3. Glick,A.B. (2004) TGFβ1, back to the future: revisiting its role as a transforming growth factor. *Cancer Biol. Ther.*, **3**, 276–283.
4. Xu,Y. *et al*. (2007) TGF-β signaling alterations and susceptibility to colorectal cancer. *Hum. Mol. Genet.*, **16**, R14–R20.
5. Hong,S. *et al*. (2007) Smad7 binds to the adaptors TAB2 and TAB3 to block recruitment of the kinase TAK1 to the adaptor TRAF2. *Nat. Immunol.*, **8**, 504–513.
6. Halder,S.K. *et al*. (2008) Smad7 induces hepatic metastasis in colorectal cancer. *Br. J. Cancer*, **99**, 957–965.
7. Broderick,P. *et al*. (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, **39**, 1315–1317.
8. Tenesa,A. *et al*. (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.*, **40**, 631–637.
9. Li,L. *et al*. (2008) Association between phosphatidylinositol 3-Kinase regulatory subunit p85α Met326Ile genetic polymorphism and colon cancer risk. *Clin. Cancer Res.*, **14**, 633–637.
10. S.A.G.E. (2007) http://darwin.case.edu. (18 November 2008, date last accessed).
11. Pearson,T.A. *et al*. (2008) How to interpret a genome-wide association study. *JAMA*, **299**, 1335–1344.
12. Vineis,P. *et al*. (2008) Expectations and challenges stemming from genome-wide association studies. *Mutagenesis*, **23**, 429–444.
13. Monteleone,G. *et al*. (2004) Smad7 in TGF-beta-mediated negative regulation of gut inflammation. *Trends Immunol.*, **25**, 513–517.
14. Boulay,J.L. *et al*. (2003) SMAD7 is a prognostic marker in patients with colorectal cancer. *Int. J. Cancer*, **104**, 446–449.
15. Lucia,M.S. *et al*. (1998) The role of transforming growth factor-beta1, -beta2, and -beta3 in androgen-responsive growth of NRP-152 rat prostatic epithelial cells. *J. Cell. Physiol.*, **175**, 184–192.
16. Komm,B.S. *et al*. (1988) Estrogen binding, receptor mRNA, and biologic response in osteoblast-like osteosarcoma cells. *Science*, **241**, 81–84.
17. Arrick,B.A. *et al*. (1990) Differential regulation of expression of three transforming growth factor beta species in human breast cancer cell lines by estradiol. *Cancer Res.*, **50**, 299–303.
18. Roberts,A.B. *et al*. (1992) Mechanistic interrelationships between two superfamilies: the steroid/retinoid receptors and transforming growth factor-β. *Cancer Surv.*, **14**, 205–219.
19. Chegini,N. *et al*. (2003) The expression of Smads and transforming growth factor beta receptors in leiomyoma and myometrium and the effect of gonadotropin releasing hormone analogue therapy. *Mol. Cell. Endocrinol.*, **209**, 9–16.
20. Luo,X. *et al*. (2003) Gonadotropin releasing hormone analogue (GnRHa) alters the expression and activation of SMAD in human endometrial epithelial and stromal cells. *Reprod. Biol. Endocrinol.*, **1**, 125.
21. Dixon,A. *et al*. (2007) 17β-Estradiol attenuates diabetic kidney disease by regulating extracellular matrix and transforming growth factor-β protein expression and signaling. *Am. J. Physiol. Renal Physiol.*, **293**, F1678–F1690.
22. Tang,Y. *et al*. (2008) Smad7 stabilizes β-catenin binding to E-cadherin complex and promotes cell-cell adhesion. *J. Biol. Chem.*, **283**, 23956–23963.
23. Li,L. *et al*. (2008) A common 8q24 variant and the risk of colon cancer: a population-based case–control study. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 339–342.
24. Potter,J.D. (1999) Colorectal cancer: molecules and populations. *J. Natl Cancer Inst.*, **91**, 916–932.

Received November 21, 2008; revised March 26, 2009; accepted April 4, 2009