Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Clinical Cancer Research

Common Cancer Stem Cell Gene Variants Predict Colon Cancer Recurrence

Armin Gerger[1], Wu Zhang[1], Dongyun Yang[2], Pierre Bohanes[1], Yan Ning[1], Thomas Winder[1], Melissa J. LaBonte[1], Peter M. Wilson[1], Leonor Benhaim[1], David Paez[1], Rita El-Khoueiry[1], Anthony El-Khoueiry[1], Michael Kahn[3], and Heinz-Josef Lenz[1,3]

Abstract

Purpose: Recent evidence suggests that cancer stem cells (CSC) are responsible for key elements of colon cancer progression and recurrence. Germline variants in CSC genes may result in altered gene function and/or activity, thereby causing interindividual differences in a patient's tumor recurrence capacity and chemoresistance. We investigated germline polymorphisms in a comprehensive panel of CSC genes to predict time to tumor recurrence (TTR) in patients with stage III and high-risk stage II colon cancer.

Experimental Design: A total of 234 patients treated with 5-fluorouracil–based chemotherapy at the University of Southern California were included in this study. Whole blood samples were analyzed for germline polymorphisms in genes that have been previously associated with colon CSC (CD44, Prominin-1, DPP4, EpCAM, ALCAM, Msi-1, ITGB1, CD24, LGR5, and ALDH1A1) by PCR-RFLP or direct DNA-sequencing.

Results: The minor alleles of CD44 rs8193 C>T, ALCAM rs1157 G>A, and LGR5 rs17109924 T>C were significantly associated with increased TTR (9.4 vs. 5.4 years; HR, 0.51; 95% CI: 0.35–0.93; $P = 0.022$; 11.3 vs. 5.7 years; HR, 0.56; 95% CI: 0.33–0.94; $P = 0.024$, and 10.7 vs. 5.7 years; HR, 0.33; 95% CI: 0.12–0.90; $P = 0.023$, respectively) and remained significant in the multivariate analysis stratified by ethnicity. In recursive partitioning, a specific gene variant profile including LGR5 rs17109924, CD44 rs8193, and ALDH1A1 rs1342024 represented a high-risk subgroup with a median TTR of 1.7 years (HR, 6.71, 95% CI: 2.71–16.63, $P < 0.001$).

Conclusion: This is the first study identifying common germline variants in colon CSC genes as independent prognostic markers for stage III and high-risk stage II colon cancer patients. *Clin Cancer Res; 17(21); 1–10. ©2011 AACR.*

Introduction

Adjuvant treatment is recommended for patients with stage III and high-risk stage II colon cancer. The risk of tumor recurrence can be significantly reduced by treating these patients with 5-fluorouracil (5-FU)–based chemotherapy. The addition of oxaliplatin to 5-FU–based chemotherapy is now a standard adjuvant treatment for colon cancer, with a higher 5-year disease-free survival (DFS) rate, compared with 5-FU–based treatment alone (73.3% vs. 67.4%; ref. 1). However, a considerable number of patients will relapse despite adjuvant treatment (2).

Tumor recurrence after curative surgery remains a major obstacle for improving overall cancer survival, which may be in part due to the existence of cancer stem cells (CSC). Growing evidence suggests that human cancers are stem cell diseases and only a small subpopulation of cancer cells, endowed with stem cell-like features, might be responsible for tumor initiation, progression and chemoresistance (3). Cancer cells with the properties of stem cells possess the ability to self-renew, to undergo multilineage differentiation, and to survive an adverse tissue microenvironment.

Putative CSC populations have been identified in colon cancer on the basis of the expression of specific markers and their functional properties; however, phenotypic characterization of colon CSCs is still a matter of debate and ongoing research studies (4). Ideally, definitive markers should be gene products that are coupled to the function of the stem cell. CSC markers in colon cancer include CD133, CD44, and CD166 (5). More recently, EpCAM, CD26, Msi-1, CD29, CD24, LGR5, and ALDH1A1 have been added to the list of putative stem cell markers for colon cancer (6, 7).

Authors' Affiliations: [1]Division of Medical Oncology; [2]Department of Preventive Medicine; Norris Comprehensive Cancer Center, [3]USC Center for Molecular Pathways and Drug Discovery, Keck School of Medicine, University of Southern California, Los Angeles, California

Corresponding Author: Heinz-Josef Lenz, Sharon A. Carpenter Laboratory, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, 1441 Eastlake Avenue, Los Angeles, CA 90033. Phone: 1-323-865-3967; Fax: 1-323-865-0061; E-mail: lenz@usc.edu

doi: 10.1158/1078-0432.CCR-11-1180

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Translational Relevance

Germline variants in cancer stem cell (CSC) genes have recently been linked to cancer outcome and chemoresistance. Here, we investigated germline polymorphisms in a comprehensive panel of genes that have been previously associated with colon CSC to predict time to tumor recurrence (TTR) in patients with stage III and high-risk stage II colon cancer treated with 5-fluorouracil–based adjuvant chemotherapy. We hypothesized that these gene variants may alter the gene's function and/or activity, thereby causing interindividual differences in a patient's tumor recurrence capacity and chemoresistance. Our findings suggest that TTR varies according to CD44, ALCAM, LGR5, and ALDH1A1 genotypes, thus representing valuable biomarkers to separate high-risk from low-risk colon cancer patients.

These colon CSC markers are representative of a range of pathways including the Wnt-target genes, cell adhesion molecules, RNA-binding proteins, and detoxifying enzymes, and play distinct roles in a variety of processes including cell differentiation, proliferation, migration, apoptosis, adhesion, lymphocyte homing, angiogenesis, and cellular response to chemotherapy (8). Current therapies target populations of rapidly growing and differentiated tumor cells, but have been shown to lack activity against CSCs (4). CSCs therefore may have an important role in tumor recurrence despite adjuvant chemotherapy. Thus far, preclinical studies in colon cancer have identified that CSCs are capable of initiating tumor development, however, little is known about the role of CSCs in colon cancer tumor recurrence.

There is substantial germline genetic variability within the genes used as markers to identify CSCs, including multiple single nucleotide polymorphisms (SNP). These common DNA-sequence variations may alter the gene function and/or activity including transcription, translation or splicing, thereby causing interindividual differences in relation to tumor recurrence capacity and chemoresistance (9). We recently tested the impact of common gene variants in the cell surface glycoprotein, CD44, a gastric and colon CSC marker, on clinical outcome of patients with localized gastric adenocarcinoma and found that the minor allele of CD44 rs187116 was significantly associated with decreased time to tumor recurrence (TTR) and overall survival (OS), identifying a "high-risk" patient population based on a germline genetic variant (10).

In this study, we investigated 25 germline polymorphisms in a comprehensive panel of genes that have been previously associated with colon CSC to predict tumor recurrence in patients with stage III and high-risk stage II colon cancer. The analyzed CSC genes included CD44, Prominin-1 (CD133), dipeptidyl peptidase-4 (DPP4/CD26), epithelial cell adhesion molecule (EpCAM), activated leukocyte cell adhesion molecule (ALCAM/CD166), musashi homolog-1 (MSI-1), integrin beta-1 (ITGB1/CD29), CD24, leucine-rich repeat containing G protein-coupled receptor-5 (LGR5), and aldehyde dehydrogenase-1 family member A1 (ALDH1A1). To the best of our knowledge, this is the first study investigating common germline genetic variants in a comprehensive panel of colon CSC genes to predict tumor recurrence. This study was conducted adhering to the reporting recommendations for tumor marker prognostic studies (11–13).

Patients and Methods

Eligible patients

A total of 234 patients with stage III and high-risk stage II colon cancer were included in this study. High-risk stage II colon cancer patients were defined if they presented with at least one of the following features: lymph node sampling less than 12; poorly differentiated tumor; vascular, lymphatic or perineural invasion; tumor presentation with obstruction or perforation and pT4. All patients were treated with 5-FU–based adjuvant chemotherapy at the Norris Comprehensive Cancer Center/University of Southern California (NCCC/USC) or the Los Angeles County/USC-Medical Center (LAC/USCMC) from 1987 to 2007. All patients were included in the colon cancer surveillance program of NCCC/USC or LAC/USCMC, providing history and physical examination and CEA determination every 3 months for 3 years and every 6 months at years 4 and 5 after surgery, colonoscopy at year 1 and thereafter every 3 to 5 years and CT scans of chest and abdomen every 6 months for the first 3 years. Patient data were collected retrospectively through chart review. Whole blood was collected at the time of diagnosis and stored at −80°C. Blood samples from 216 patients were available for current genetic analyses. The study was approved by the Institutional Review Boards at USC and all study participants signed informed consent for the analysis of molecular correlates.

Candidate polymorphisms

Common and putatively functional polymorphisms within genes that have been previously associated with colon CSC were selected using public literature resources and databases including: NCBI-PubMed, dbSNP, Ensembl, GeneCards, and the Pharmacogenomics-Knowledge-Base. Stringent and predefined selection criteria used were: (a) minor allele frequency 10% or more in Caucasians; (b) polymorphism that could alter the function of the gene in a biologically relevant manner [either published data or predicted function using Functional-Single-Nucleotide-Polymorphism (F-SNP) database, http://compbio.cs.queensu.ca/F-SNP (14, 15)]; and (c) published clinical associations (e.g., cancer risk/outcome or chemoresistance; Table 1).

Genotyping

Genomic DNA was extracted from peripheral blood using the QIAmp-kit (Qiagen). The majority of the samples were tested using PCR-based restriction fragment length

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Table 1. Analyzed CSC gene polymorphisms

Gene	rs-number	Base exchange	Region	Genotyping	Published genetic association studies (reference)
CD44	rs8193	C>T	3UTR	RE (BsrDI)	10
	rs187116	A>G	Intron	RE (MspI)	10
	rs4755392	T>A	3UTR	DS	10
	rs7116432	A>G	3UTR	RE (NlaIII)	10
Prominin-1	rs3130	A>G	3UTR	RE (EcorRI)	nd
	rs2240688	A>C	3UTR	DS	nd
	rs2286455	C>T	Splice Site	RE (MboI)	nd
DPP4	rs2300757	G>C	Intron	RE (TfiI)	nd
	rs1014444	A>G	Intron	RE (AluI)	nd
	rs2268894	A>G	Intron	DS	nd
EpCAM	rs17036526	G>C	Splice site	RE (DdeI)	nd
	rs1126497	C>T	Non-synonymous coding	RE (NiaIII)	32
	rs1421	T>C	3UTR	DS	32
ALCAM	rs6437585	C>T	5UTR	RE (Cac8I)	26
	rs1044240	A>G	Nonsynonymous coding	DS	ns
	rs1044243	G>A	Nonsynonymous coding	DS	nd
	rs1157	G>A	3UTR	RE (MspI)	nd
MSI-1	rs2522137	A>C	3UTR	DS	nd
ITGB1	rs2153875	T>G	Splice site	RE (BfaI)	nd
CD24	rs8734	C>T	Nonsynonymous coding	RE (AciI)	33
	rs3838646	–/CA	3UTR	RE (BsrI)	33
LGR5	rs17109924	T>C	Nonsynonymous coding	DS	nd
	rs17109926	G>A	3UTR	DS	nd
ALDH1A1	rs13959	G>A	Synonymous coding	RE (Hypch4III)	nd
	rs1342024	G>C	Upstream	DS	nd

Abbreviations: DS, direct DNA-sequencing; RE, restriction enzyme; UTR, untranslated region; nd, no data.

polymorphism (PCR-RFLP) analysis. Forward and reverse primers were used for PCR amplification, PCR products were digested by restriction enzymes (New England Biolabs), and alleles were separated on 4% NuSieve ethidium bromide stained agarose gel. If no matching restriction enzyme could be found, samples were analyzed by direct DNA-sequencing. For quality control purposes, a total of 5% PCR-RFLP analyzed samples were reanalyzed by direct DNA-sequencing. The investigator analyzing the germline polymorphisms was blinded to the clinical dataset.

Statistical analysis

The endpoint of the study was TTR. TTR was calculated from the date of diagnosis of colon cancer to the date of the first observation of tumor recurrence. TTR was censored at the time of death or at the last follow-up if the patient remained tumor recurrence-free at that time. With 216 patients, there was an 80% power to detect a minimum HR of 1.8 across the range of minor allele frequencies (0.2–0.5) for TTR using a dominant model. For the recessive model, the minimum HR was 3.1 when the allele frequency was 0.2 and approaches 1.8 when the allele frequency was 0.5. Allelic distribution of the polymorphisms by ethnicity was tested for deviation from Hardy–Weinberg equilibrium using χ^2 test. The distribution of polymorphisms across baseline demographic, clinical, and pathologic characteristics was examined using Fisher's exact test. The true mode of inheritance of all polymorphisms tested is not established yet and we assumed a codominant, additive, dominant or recessive genetic model where appropriate. The association of polymorphisms with TTR was analyzed using Kaplan–Meier curves and log-rank test. In the multivariate Cox-regression analysis, the model was adjusted for stage and type of adjuvant chemotherapy, and stratified by ethnicity. Interactions between polymorphisms and stage and gender on TTR were tested by comparing likelihood ratio statistics between the baseline and nested Cox proportional hazards models that include the multiplicative product term. *P* values for all polymorphisms were adjusted for multiple testing using a modified test of Conneely and Boehnke that was applied for the correlated tests due to linkage disequilibrium and the different modes of inheritance considered (16). Recursive partitioning (RP), including cross validation, was used to explore and identify polymorphism profiles associated with TTR using the rPart-function in S-plus (17, 18). Case-wise deletion for missing polymorphisms was used in univariate and multivariate analyses. In the RP-analysis, all patients with at least

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

one polymorphism result available were included. All analyses were done using SAS 9.2 (SAS Institute Inc.).

Results

The baseline characteristics of the 234 patients included in this analysis are summarized in Table 2. The median age at time of diagnosis was 59 years (range 22–87), with a median follow-up time of 4.4 years (range 0.4–16.8). Ninety (38.5%) patients showed tumor recurrence, with a stage III and high-risk stage II dependent probability of 3-year recurrence of 0.45 ± 0.047 and 0.21 ± 0.043, respectively.

Table 2. Baseline patient characteristics

	n	%
Sex		
Female	107	45.72
Male	127	54.28
Ethnicity		
Asian	34	14.53
African American	15	6.41
Caucasian	123	52.56
Hispanic	62	26.5
T		
T1	2	0.85
T2	14	5.98
T3	187	79.91
T4	27	11.54
Tx	4	1.72
Grade		
Well	11	2.18
Moderate	151	64.53
Poor/undifferentiated	54	23.08
Missing	18	10.21
N		
Negative	105	44.87
N1	72	30.77
N2	57	24.36
Stage		
High-risk II	105	44.87
III	129	55.13
N of resected lymph nodes		
≤12	70	29.91
>12	145	61.97
Missing	19	8.12
Tumor side		
Left	110	47.01
Right	115	49.15
Left and right	4	1.71
Missing	5	2.13
Adjuvant treatment		
5-FU	151	64.53
5-FU/LV/Oxaliplatin	60	25.64
5-FU/LV/Irinotecan	23	9.83

Median OS has not been reached yet. The genotyping quality control by direct DNA-sequencing provided a genotype concordance of 99% or more. Genotyping was successful in at least 90% of cases for each polymorphism analyzed, with the exception of CD44 rs8193 (88.4%). In failed cases, genotyping was not successful because of limited quantity and/or quality of extracted genomic DNA. The allelic frequencies for all polymorphisms were within the probability limits of Hardy–Weinberg equilibrium, with the exception of EpCAM rs17036526 (data not shown).

There were no significant associations between the polymorphisms and baseline demographic, clinical or pathologic characteristics (data not shown). In the univariate analysis, the minor alleles of CD44 rs8193 C>T, ALCAM rs1157 G>A and LGR5 rs17109924 T>C were significantly associated with an increased TTR. Patients carrying at least one T allele in CD44 rs8193 showed a median TTR of 9.4 years. In contrast, patients with homozygous C/C had a median TTR of 5.4 years (HR, 0.51; 95% CI: 0.35–0.93; $P = 0.022$). Patients harboring the minor allele of ALCAM rs1157 showed a median TTR of 11.3 years compared with 5.7 years for patients harboring the homozygous G/G (HR, 0.56; 95% CI: 0.33–0.94; $P = 0.024$). Patients carrying one C allele in LGR5 rs17109924 had a median TTR of 10.7 years compared with 5.7 years for those patients carrying the homozygous T/T (HR, 0.33; 95% CI: 0.12–0.90; $P = 0.023$; Fig. 1). The other tested gene variants did not show any statistically significant associations with TTR in the univariate analyses.

In the multivariate analysis stratified by ethnicity, the minor alleles of CD44 rs8193 C>T, ALCAM rs1157 G>A and LGR5 rs17109924 T>C remained significantly associated with increased TTR (Table 3). There was no significant interaction between the polymorphisms and tumor stage or gender on TTR (*P* values for interactions less than 0.05). In multiple testing including all polymorphisms analyzed, none of them remained significantly associated with TTR (adjusted-*P* for CD44 rs8193 = 0.142; adjusted-*P* for ALCAM rs1157 = 0.199; adjusted-*P* for LGR5 rs17109924 = 0.394)

When RP was used to construct a decision tree as a predictive model to classify patients based on the gene variants, high- and low-risk patient subgroups were identified. In the resultant tree, the most important factor that determined the TTR in these patients was LGR5 rs17109924. Patients carrying the combination of LGR5 rs17109924 wild type and at least one CD44 rs8193 wild-type allele and the mutant variant of ALDH1A1 rs1342024 showed a TTR of 1.7 years (node 5) compared with 10.7 years in patients with the minor allele of LGR5 rs17109924 (node 1) or the combination of LGR5 rs17109924 wild-type and CD44 rs8193 mutant variant (node 2; HR, 6.71, 95% CI: 2.71–16.63, $P < 0.001$; Fig. 2).

To evaluate whether the high-risk patients identified from our gene variant profile benefit from combination chemotherapy ($n = 47$) compared with 5-FU alone ($n = 66$), the cases from node 4 and 5 of the decision tree were combined for further analysis. According to the treatment regimen, no

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

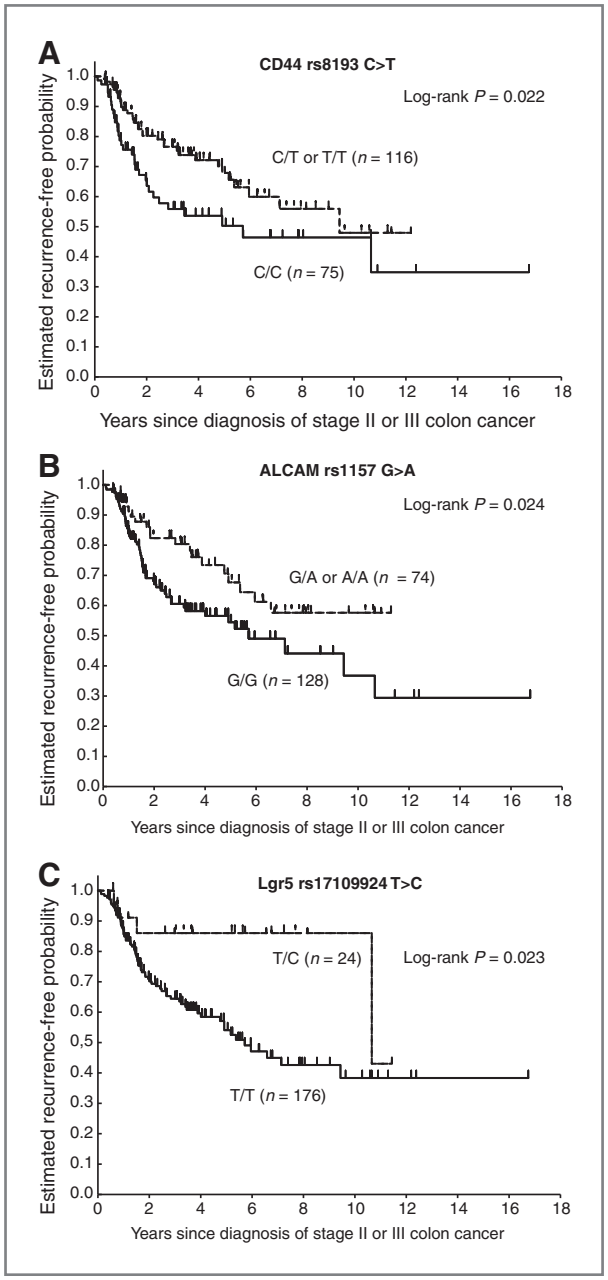

Figure 1. TTR by CD44 rs8193 C>T (A), ALCAM rs1157 G>A (B), and Lgr5 rs17109924 T>C (C).

significant difference in TTR were identified in this high-risk subgroup ($P > 0.05$).

Discussion

In this study, we investigated germline polymorphisms in a comprehensive panel of genes that have been previously associated with colon CSC to predict tumor recurrence in patients with stage III and high-risk stage II colon cancer. The results indicate that common CSC gene variants in CD44, ALCAM, LGR5, and ALDH1A1 may be valuable to separate high-risk from low-risk colon cancer patients.

The detailed molecular mechanisms involved in how the CD44 rs8193, ALCAM rs1157, LGR5 rs17109924, and ALDH1A1 rs1342024 polymorphisms exert effects on colon cancer are unclear. Nonsynonymous polymorphisms lead to amino acid changes and thus may affect the protein function (19). 3′UTRs have been implicated in the modulation of gene regulation at the transcriptional or translational level and function as regulators mainly through control of mRNA stability and/or translational efficiency, and therefore play an important role in the overall fate of gene expression. Furthermore, germline polymorphisms in the 3′UTRs have been shown to have functional effects on overall gene expression (20). We used the F-SNP database to predict the functional effects of the analyzed polymorphisms. F-SNP gathers computationally predicted functional information about polymorphisms, particularly aiming to facilitate identification of disease-related polymorphisms in association studies (14, 15). When used for this study set of polymorphisms, F-SNP predicted changes in transcription factor binding to the 3′UTR located CD44 rs8193 and ALCAM rs1157, and changes in splicing regulation and protein coding for the nonsynonymous LGR5 rs17109924, thus supporting the effects seen in our study. No prediction could be provided for the upstream located ALDH1A1 rs1342024 by the software.

CD44-signaling is crucial in cancer cell proliferation, motility, and migration. As a Wnt-target gene, CD44 promotes cell proliferation via the phosphatidylinositol 3-kinase (PI3K)/Akt pathway. CD44 positive colon cancer cells have been reported to possess the capacity for self-renewal, longevity, and multipotency (21). ALCAM belongs to the immunoglobulin superfamily of cell adhesion molecules involved in cell–cell interactions. ALCAM may regulate through cytoskeletal anchoring and the integrity of the extracellular immunoglobulin-like domains complex cellular properties in regard to cell adhesion, migration, and growth (22). Isolation of ALCAM/CD44 double-positive cells from human colon cancer cells can recapitulate tumorigenesis when xenografted at low numbers into immune-deficient mice which represents a hallmark of CSCs (5). Despite the potentially high clinical relevance of these CSC markers, little is known about their prognostic significance in colon cancer and contradictory findings have been reported. In a recent study based on 110 colorectal cancer (CRC) patients, membranous expression of CD44 and ALCAM did not predict survival in single-marker analyses, but gained significance when combined (23). In contrast, Weichert and colleagues found a correlation between membranous ALCAM expression and decreased survival in 111 CRC patients (24). Interestingly, loss, rather than overexpression, of membranous CD44 and ALCAM was correlated to outcome in an analysis including 101 CRC patients. The authors suggested that their results are in large part dependent on the cell adhesion function of CD44 and ALCAM with loss of cell adhesion representing a

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Table 3. Univariate and multivariate analysis of polymorphisms and TTR

		Univariate analysis				Multivariate analysis	
	n	Median TTR, y (95% CI)	Probability ± SE[a] of 3-year recurrence	HR (95% CI)	*P*[b]	HR (95% CI)	*P*[c]
CD44 rs8193					0.022		0.047
C/C	75	5.4 (2.1–16.8)	0.44 ± 0.06	1 (Reference)		1 (Reference)	
C/T[d]	92	9.4 (5.9–12.2)	0.23 ± 0.04	0.51 (0.35–0.93)		0.60 (0.36–0.99)	
T/T[d]	24						
CD44 rs187116					0.31		0.20
A/A	61	12.4 (9.4–12.4)	0.25 ± 0.06	1 (Reference)		1 (Reference)	
A/G	94	10.7 (4.9–11.3)	0.33 ± 0.05	1.40 (0.78–2.51)		1.43 (0.79–2.59)	
G/G	45	4.9 (3.5–16.8)	0.34 ± 0.08	1.65 (0.85–3.21)		1.92 (0.93–3.95)	
CD44 rs4755392					0.94		0.74
T/T	48	9.4 (4.9–16.8)	0.28 ± 0.07	1 (Reference)		1 (Reference)	
T/A	97	7.1 (4.9–11.4)	0.33 ± 0.05	1.10 (0.61–1.98)		1.22 (0.67–2.21)	
A/A	57	12.4 (3.4–12.4)	0.36 ± 0.07	1.04 (0.53–2.02)		1.02 (0.51–2.01)	
CD44 rs7116432					0.48		0.61
A/A	61	5.7 (2.5–16.8)	0.40 ± 0.07	1 (Reference)		1 (Reference)	
A/G	87	9.4 (5.2–9.6)	0.28 ± 0.05	0.75 (0.44–1.30)		0.80 (0.45–1.41)	
G/G	50	10.3 (4.9–10.3)	0.23 ± 0.06	0.73 (0.38–1.40)		0.72 (0.35–1.47)	
Prominin-1 rs3130					0.59		0.40
A/A	124	7.1 (5.2–12.2)	0.31 ± 0.05	1 (Reference)		1 (Reference)	
A/G	80	9.4 (3.4–16.8)	0.34 ± 0.06	1.14 (0.71–1.81)		1.24 (0.76–2.03)	
Prominin-1 rs2240688					0.86		0.91
A/A	104	6.6 (4.9–10.7)	0.27 ± 0.05	1 (Reference)		1 (Reference)	
A/C[d]	89	12.4 (3.2–12.4)	0.38 ± 0.05	1.04 (0.66–1.65)		0.97 (0.61–1.56)	
C/C[d]	10						
Prominin-1 rs2286455					0.45		0.21
C/C	155	10.7 (5.7–12.4)	0.32 ± 0.04	1 (Reference)		1 (Reference)	
C/T[d]	41	5.2 (3.9–16.8)	0.33 ± 0.07	1.22 (0.73–2.04)		1.41 (0.83–2.40)	
T/T[d]	5						
DPP4 rs2300757					0.79		0.71
G/G	75	5.4 (3.4–16.8)	0.35 ± 0.06	1 (Reference)		1 (Reference)	
G/C	96	10.7 (5.9–12.2)	0.30 ± 0.05	0.85 (0.52–1.41)		0.87 (0.52–1.45)	
C/C	30	5.7 (2.4–11.3)	0.33 ± 0.10	1.00 (0.50–2.01)		0.75 (0.37–1.54)	
DPP4 rs1014444					0.44		0.41
A/A	68	5.2 (3.4–16.8)	0.35 ± 0.06	1 (Reference)		1 (Reference)	
A/G	92	10.7 (5.9–12.2)	0.29 ± 0.05	0.72 (0.43–1.20)		0.73 (0.43–1.24)	
G/G	42	11.3 (2.4–11.3)	0.37 ± 0.08	0.90 (0.48–1.68)		0.70 (0.37–1.35)	
DPP4 rs2268894					0.81		0.79
A/A	55	5.7 (2.4–11.3)	0.37 ± 0.07	1 (Reference)		1 (Reference)	
A/G	98	10.7 (5.4–12.2)	0.32 ± 0.05	0.84 (0.49–1.46)		1.18 (0.67–2.10)	
G/G	44	5.2 (3.4–16.8)	0.30 ± 0.08	0.95 (0.51–1.79)		1.22 (0.64–2.33)	
EpCAM rs17036526					0.95		0.73
G/G	34	7.1 (2.5–7.1)	0.32 ± 0.09	1 (Reference)		1 (Reference)	
G/C	152	9.4 (5.2–16.8)	0.32 ± 0.04	1.03 (0.54–1.97)		0.99 (0.50–1.95)	
C/C	13	4.9 (1.7–10.9)	0.38 ± 0.15	1.19 (0.41–3.43)		1.45 (0.48–4.34)	
EpCAM rs1126497					0.69		0.56
C/C	50	7.1 (3.2–16.8)	0.32 ± 0.07	1 (Reference)		1 (Reference)	
C/T	104	5.9 (4.0–10.7)	0.34 ± 0.05	1.00 (0.57–1.75)		0.92 (0.52–1.64)	
T/T	49	9.4 (4.9–12.2)	0.30 ± 0.07	0.79 (0.40–1.54)		0.70 (0.34–1.41)	
EpCAM rs1421					0.10		0.24
T/T	157	6.6 (3.9–16.8)	0.37 ± 0.04	1 (Reference)		1 (Reference)	

(Continued on the following page)

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Table 3. Univariate and multivariate analysis of polymorphisms and TTR (Cont'd)

		Univariate analysis				Multivariate analysis	
	n	Median TTR, y (95% CI)	Probability ± SE[a] of 3-year recurrence	HR (95% CI)	*P*[b]	HR (95% CI)	*P*[c]
T/C[d]	45	7.1 (5.9–12.4)	0.18 ± 0.06	0.60 (0.32–1.12)		0.68 (0.36–1.29)	
C/C[d]	2						
ALCAM rs6437585					0.080		0.16
C/C	175	9.4 (5.4–16.8)	0.29 ± 0.04	1 (Reference)		1 (Reference)	
C/T	26	2.4 (1.5–10.3)	0.50 ± 0.11	1.73 (0.93–3.21)		1.62 (0.83–3.17)	
ALCAM rs1044240					0.95		0.77
A/A	166	9.4 (5.4–16.8)	0.32 ± 0.04	1 (Reference)		1 (Reference)	
A/G[d]	31	5.2 (3.2–12.4)	0.31 ± 0.08	1.02 (0.57–1.82)		0.91 (0.50–1.68)	
G/G[d]	6						
ALCAM rs1044243					0.30		0.18
G/G	170	6.6 (4.9–16.8)	0.34 ± 0.04	1 (Reference)		1 (Reference)	
G/A	32	10.9 (4.8–10.9)	0.22 ± 0.08	0.70 (0.36–1.38)		0.62 (0.31–1.25)	
ALCAM rs1157					0.024		0.027
G/G	128	5.7 (3.2–10.7)	0.39 ± 0.05	1 (Reference)		1 (Reference)	
G/A[d]	67	11.3 (5.9–11.3)	0.20 ± 0.05	0.56 (0.33–0.94)		0.55 (0.32–0.93)	
A/A[d]	7						
MSI-1 rs2522137					0.59		0.41
A/A	65	5.7 (4.0–12.4)	0.30 ± 0.06	1 (Reference)		1 (Reference)	
A/C	85	7.1 (3.2–12.2)	0.35 ± 0.06	0.87 (0.51–1.48)		0.80 (0.46–1.37)	
C/C	51	16.8 (4.9–16.8)	0.31 ± 0.07	0.73 (0.39–1.34)		0.65 (0.35–1.23)	
ITGB1 rs2153875					0.81		0.66
T/T	98	12.4 (5.2–12.4)	0.32 ± 0.05	1 (Reference)		1 (Reference)	
T/G	89	6.6 (4.9–10.7)	0.32 ± 0.05	1.14 (0.70–1.85)		1.23 (0.74–2.02)	
G/G	15	4.9 (2.7–16.8)	0.30 ± 0.12	1.24 (0.55–2.82)		1.34 (0.57–3.15)	
CD24 rs8734					0.62		0.33
C/C	81	5.4 (3.9–16.8)	0.33 ± 0.06	1 (Reference)		1 (Reference)	
C/T	62	6.6 (4.9–10.9)	0.31 ± 0.06	0.85 (0.50–1.46)		0.74 (0.43–1.30)	
T/T	56	12.4 (5.2–12.4)	0.33 ± 0.07	0.75 (0.42–1.36)		0.65 (0.35–1.19)	
CD24 rs3838646					0.39		0.53
(CA)1	182	9.4 (4.9–16.8)	0.35 ± 0.04	1 (Reference)		1 (Reference)	
(CA)2[d]	16	7.1 (5.9–7.2)	0.12 ± 0.08	0.67 (0.27–1.67)		0.73 (0.28–1.90)	
(CA)3[d]	1						
LGR5 rs17109924					0.023		0.035
T/T	176	5.7 (4.0–16.8)	0.36 ± 0.04	1 (Reference)		1 (Reference)	
T/C	24	10.7 (10.7–11.4)	0.14 ± 0.08	0.33 (0.12–0.90)		0.33 (0.12–0.93)	
LGR5 rs17109926					0.92		0.86
G/G	123	6.6 (5.2–11.4)	0.31 ± 0.05	1 (Reference)		1 (Reference)	
G/A[d]	73	10.7 (3.4–16.8)	0.35 ± 0.06	1.02 (0.64–1.64)		1.05 (0.64–1.70)	
A/A[d]	5						
ALDH1A1 rs13959					0.66		0.45
G/G	61	5.4 (4.8–9.4)	0.25 ± 0.06	1 (Reference)		1 (Reference)	
G/A	98	10.7 (2.7–16.8)	0.41 ± 0.06	0.97 (0.57–1.63)		0.96 (0.56–1.64)	
A/A	41	11.3 (3.4–11.3)	0.25 ± 0.07	0.74 (0.38–1.46)		0.65 (0.32–1.33)	
ALDH1A1 rs1342024					0.18		0.37
G/G	79	9.4 (4.9–11.4)	0.21 ± 0.05	1 (Reference)		1 (Reference)	
G/C	86	16.8 (3.2–16.8)	0.38 ± 0.06	1.20 (0.71–2.02)		1.04 (0.60–1.78)	
C/C	35	5.2 (1.7–7.8)	0.46 ± 0.09	1.79 (0.95–3.34)		1.54 (0.81–2.93)	

[a]Greenwood SE.

[b]Based on log-rank test.

[c]Based on Wald test within Cox proportional hazards model.

[d]In the dominant model.

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

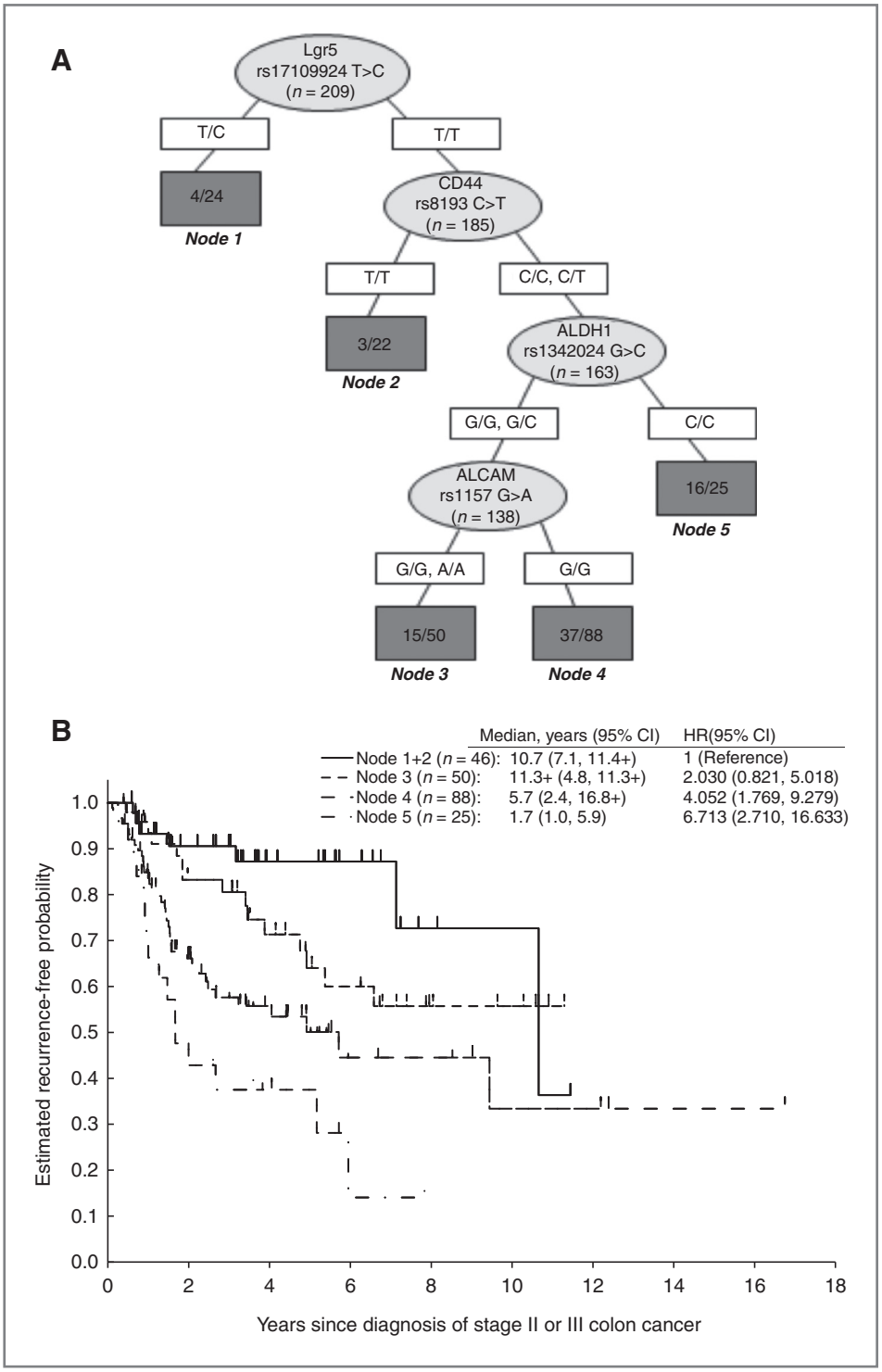

Figure 2. A, RPart analysis of TTR. The end nodes of the tree model represent subgroups of low- and high-risk patients based on either a single gene variant or combination of gene variants. Fractions within the end nodes indicate patients who recurred/total patients with this gene variant profile. B, TTR by tree model defined subgroups. Node 5 represents a high-risk subgroup based on a specific gene variant profile including LGR5 rs17109924, CD44 rs8193, and ALDH1A1 rs1342024.

fundamental step underlying the initiation of the metastatic process (25).

These conflicting results raised the question whether germline genetic variants putatively changing the gene's function rather than membranous evaluation of these proteins may predict colon cancer patient's outcome. We recently showed that the minor allele of CD44 rs187116 predicts decreased TTR and OS in patients with localized gastric adenocarcinoma (10). More recently, Zhou and colleagues analyzed 2 polymorphisms in ALCAM

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

investigating 1,033 breast cancer patients and 1,116 controls and found that individuals harboring the ALCAM rs6437585 C/T or T/T genotypes have an OR of 1.38 (95% CI: 1.11–1.72) for developing breast cancer, compared with the C/C genotype. Additional experiments showed that the T allele was associated with a higher transcriptional activity of the ALCAM gene (26). Both polymorphisms, CD44 rs187116 and ALCAM rs6437585, did not show any clinical associations in this study. However, these SNPs have been analyzed in other tumor entities and/or setting (risk assessment) and therefore might not exert their effects in colon cancer and tumor recurrence assessment. Because CD44 rs8193 and ALCAM rs1157 may change transcriptional factor binding to the polymorphic region of the gene as predicted by F-SNP, the findings in this study that these gene variants affect TTR in colon cancer are biologically plausible. Taking into account the clinical association found in our study and the predicted function by F-SNP, we hypothesize that the wild-type genotype of CD44 rs8193 is associated with a higher transcriptional activity of the CD44 gene leading to a lower TTR in our study cohort. For ALCAM expression and clinical outcome in colon cancer conflicting results can be found in the literature, therefore we cannot assess the molecular mechanisms involved in how the ALCAM rs1157 polymorphism exerts its effects on colon cancer and studies elucidating the exact biological function of the ALCAM rs1157 polymorphism should be under future consideration.

Interestingly, the tree analysis provided LGR5 rs17109924 as the first split indicating the most important factor determining TTR in this patient cohort. LGR5 is a member of the G-protein-coupled receptor family comprising proteins with 7 transmembrane domains. G-protein-coupled receptors function as receptors for various classes of ligands, including peptide hormones and chemokines; however, the ligand for and function of LGR5-related signaling remains unclear. LGR5, a Wnt-target gene, has been reported to be a marker for colon CSC, thus playing a putative role in the biological function of stem cells. In a recent study, high membranous LGR5 expression was shown to predict lower DFS in CRC patients (27–29). LGR5 rs17109924 represents a nonsynonymous SNP and was predicted to affect splicing regulation and protein coding by F-SNP. In addition, LGR5 rs17109924 predicted TTR in both the univariate and multivariate analysis and was incorporated in the tree analysis, strongly indicating that this SNP has functional significance. Taking into account the clinical association found in our study and the predicted function by F-SNP, we hypothesize that the LGR5 rs17109924 wild-type genotype is associated with a higher protein expression of the LGR5 gene leading to a lower TTR in our study cohort.

A combination of gene variants in the tree analysis defined a high-risk subgroup with significantly lower TTR by incorporating ALDH1A1 rs1342024 when compared with single marker analysis. ALDH1 is a detoxifying enzyme that oxidizes intracellular aldehydes. This detoxification capacity may protect stem cells against oxidative insult (30, 31). Taking into account the clinical association found in our study, we hypothesize that the mutant variant of ALDH1A1 rs1342024 is associated with a higher detoxification capacity of ALDH1 leading to a lower TTR in our study cohort when incorporated in the decision tree algorithm. In a recent study, membranous ALDH1 expression did not predict survival in CRC patients (25). Although the mechanism of ALDH1A1 rs1342024 remains unclear, our data suggest that a multigenic approach, which assesses the combined effects of gene variants, may detect synergistic interactions between individual SNPs thus enhancing the predictive power of the model.

This study further used multiple testing due to the large number of independent genetic variants evaluated. Application of a modified test of Conneely and Boehnke resulted in a nonsignificant adjusted *P* value for CD44 rs8193, ALCAM rs1157, and LGR5 rs17109924. Therefore, these data warrant further validation in a larger cohort. Nevertheless, the biological plausibility and our translational findings hold promise for further investigations in independent study populations. In a subanalysis combining high-risk patients based on our gene variant profile, no benefit for the addition of oxaliplatin or irinotecan to 5-FU–based chemotherapy could be shown. Because all patients included in this study represent stage III and high-risk stage II colon cancer treated with adjuvant therapy, it was not possible to correlate the genotypes with clinical outcome in an untreated control group. As a consequence, it could not be determined whether the high-risk patients, based on the gene variant profile, did not benefit from combination chemotherapy or from chemotherapy at all.

This study provides the first evidence that germline polymorphisms in CSC genes predict early tumor recurrence in patients with colon cancer. This may help to select subgroups of patients who may benefit from more aggressive treatment strategies or newly developed stem cell targeting drugs. Future biomarker-embedded translational trials are warranted to validate these findings.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was funded by the Dhont Foundation. A. Gerger is supported in part by a research grant from the Austrian Society of Hematology and Oncology, the Bank Austria Visiting Scientists Program, and the "Verein fuer Krebskranke" of the Medical University Graz; T. Winder is supported in part by a research grant of the Austrian Society of Hematology and Oncology and the "Kurt und Senta-Herrmann" foundation; P. Bohanes is supported in part by a research grant from the "Ligue Suisse contre le cancer."

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 4, 2011; revised July 29, 2011; accepted September 4, 2011; published OnlineFirst September 14, 2011.

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

References

1. Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal cancer. Lancet 2010;375:1030–47.
2. Tejpar S, Bertagnolli M, Bosman F, Lenz HJ, Garraway L, Waldmann F, et al. Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. Oncologist 2010;15:390–404.
3. Zeki SS, Graham TA, Wright NA. Stem cells and their implications for colorectal cancer. Nat Rev Gastroenterol Hepatol 2011;8:90–100.
4. Todaro M, Francipane MG, Medema JP, Stassi G. Colon cancer stem cells: promise of targeted therapy. Gastroenterology 2010;138: 2151–62.
5. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A 2007;104:10158–63.
6. Sanders MA, Majumdar AP. Colon cancer stem cells: implications in carcinogenesis. Front Biosci 2010;16:1651–62.
7. Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez-Alea M, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. Proc Natl Acad Sci U S A 2008;105:13427–32.
8. Saif MW, Chu E. Biology of colorectal cancer. Cancer J 2010;16: 196–201.
9. Coate L, Cuffe S, Horgan A, Hung RJ, Christiani D, Liu G. Germline genetic variation, cancer outcome, and pharmacogenetics. J Clin Oncol 2010;28:4029–37.
10. Winder T, Ning Y, Yang D, Zhang W, Power DG, Bohanes P, et al. Germline polymorphisms in genes involved in the CD44 signaling pathway are associated with clinical outcome in localized gastric adenocarcinoma (GA). Int J Cancer 2010;129:1096–104.
11. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 2006;100:229–35.
12. Alonzo TA. Standards for reporting prognostic tumor marker studies. J Clin Oncol 2005;23:9053–4.
13. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies. J Clin Oncol 2005;23:9067–72.
14. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. Nucleic Acids Res 2008;36: D820–4.
15. Lee PH, Shatkay H. An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics 2009;25:1048–55.
16. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. Am J Hum Genet 2007;81:1158–68.
17. Breiman L, Friedman JH, Olshen RA, Stone CJ. classification and regression trees. New York: Chapman&Hall (Wadsworth, Inc.); 1984.
18. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. Technical Report no 61, Rochester: Mayo Clinic; 1997.
19. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 2006;7:61–80.
20. Mandola MV, Stoehlmacher J, Zhang W, Groshen S, Yu MC, Iqbal S, et al. A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. Pharmacogenetics 2004;14:319–27.
21. Du L, Wang H, He L, Zhang J, Ni B, Jin H, et al. CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res 2008;14: 6751–60.
22. Levin TG, Powell AE, Davies PS, Sil AD, Dismuke AD, Anderson EC, et al. Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. Gastroenterology 2010;139:2072–82.
23. Horst D, Kriegl L, Engel J, Kirchner T, Jung A. Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. Cancer Invest 2009;27:844–50.
24. Weichert W, Knosel T, Bellach J, Dietel M, Kristiansen G. ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. J Clin Pathol 2004;57:1160–4.
25. Lugli A, Iezzi G, Hostettler I, Muraro MG, Mele V, Tornillo L, et al. Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. Br J Cancer 2010;103:382–90.
26. Zhou P, Du LF, Lv GQ, Yu XM, Gu YL, Li JP, et al. Functional polymorphisms in CD166/ALCAM gene associated with increased risk for breast cancer in a Chinese population. Breast Cancer Res Treat 2011;128:527–34.
27. Barker N, Clevers H. Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. Gastroenterology 2010; 138:1681–96.
28. Uchida H, Yamazaki K, Fukuma M, Yamada T, Hayashida T, Hasegawa H, et al. Overexpression of leucine-rich repeat-containing G protein-coupled receptor 5 in colorectal cancer. Cancer Sci 2010; 101:1731–7.
29. Takahashi H, Ishii H, Nishida N, Takemasa I, Mizushima T, Ikeda M, et al. Significance of Lgr5(+ve) Cancer Stem Cells in the Colon and Rectum. Ann Surg Oncol 2010;18:1166–74.
30. Huang EH, Hynes MJ, Zhang T, Ginestier C, Dontu G, Appelman H, et al. Aldehyde dehydrogenase 1 is a marker for normal and malignant human c8olonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. Cancer Res 2009;69: 3382–9.
31. Deng S, Yang X, Lassus H, Liang S, Kaur S, Ye Q, et al. Distinct expression levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. PLoS One 2010;5:e10277.
32. Jiang L, Zhang C, Li Y, Ju X, Zheng J, Zou P, et al. A non-synonymous polymorphism Thr115Met in the EpCAM gene is associated with an increased risk of breast cancer in Chinese population. Breast Cancer Res Treat 2011;126:487–95.
33. Diaz-Gallo LM, Medrano LM, Gomez-Garcia M, Cardena C, Rodrigo L, Mendoza JL, et al. Analysis of the influence of two CD24 genetic variants in Crohn's disease and ulcerative colitis. Hum Immunol 2011 (Epub ahead of print).

Published OnlineFirst September 14, 2011; DOI: 10.1158/1078-0432.CCR-11-1180

Clinical Cancer Research

Common Cancer Stem Cell Gene Variants Predict Colon Cancer Recurrence

Armin Gerger, Wu Zhang, Dongyun Yang, et al.

Clin Cancer Res Published OnlineFirst September 14, 2011.