Polymorphisms at the microRNA Binding-Site of the Stem Cell Marker Gene *CD133* Modify Susceptibility to and Survival of Gastric Cancer

Qiming Wang,[1,2] Hongliang Liu,[2] Huihua Xiong,[2] Zhensheng Liu,[2] Li-E. Wang,[2] Ji Qian,[2] Ramya Muddasani,[3] Victoria Lu,[3] Dongfeng Tan,[3] Jaffer A. Ajani,[2,4] and Qingyi Wei[2]*

[1]*Department of Internal Medicine, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China*
[2]*Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas*
[3]*Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas*
[4]*Department of GI Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas*

CD133 is one of the most common stem cell markers, and functional single nucleotide polymorphisms (SNPs) of *CD133* may modulate its gene functions and thus cancer risk and patient survival. We hypothesized that potentially functional *CD133* SNPs are associated with gastric cancer (GC) risk and survival. To test this hypothesis, we conducted a case-control study of 371 GC patients and 313 cancer-free controls frequency-matched by age, sex, and ethnicity. We genotyped four selected, potentially functional *CD133* SNPs (rs2240688A>C, rs7686732C>G, rs10022537T>A, and rs3130C>T) and used logistic regression analysis for associations of these SNPs with GC risk and Cox hazards regression analysis for survival. We found that compared with the miRNA binding site rs2240688 AA genotype, AC + CC genotypes were associated with significantly increased GC risk (adjusted OR = 1.52, 95% CI = 1.09–2.13); for another miRNA binding site rs3130C>T SNP, the TT genotype was associated with significantly reduced GC risk (adjusted OR = 0.68, 95% CI = 0.48–0.97), compared with CC + CT genotypes. In all patients, the risk rs3130 TT variant genotype was significantly associated with overall survival (OS) (adjusted $P_{trend} = 0.016$ and 0.007 under additive and recessive models, respectively). These findings suggest that these two *CD133* miRNA binding site variants, rs2240688 and rs3130, may be potential biomarkers for genetic susceptibility to GC and possible predictors for survival in GC patients but require further validation by larger studies.

Key words: gastric cancer; polymorphism; stem cell; survival; microRNA

INTRODUCTION

Gastric cancer (GC) is the fourth most common cancer worldwide and the second most frequent cause of all cancer deaths [1]. It is believed that both environmental and genetic factors, such as *Helicobacter pylori* [2], alcohol consumption [3], smoking [4], and family cancer history [5], may play roles in the etiology. In addition, patients diagnosed with late-stages of GC suffer from increased mortality and morbidity [6]. Hence, it is imperative to identify biomarkers for prevention, early diagnosis, monitor tumor progression and potential therapeutic targets of GC.

Cancer stem cells (CSCs) have been hypothesized as the origin of cancer. Like normal stem cells, CSCs are immortal and can self-renew and differentiate into all types of cells [7]. CD133 (also called AC133 or PROM1) is a trans-membrane cell-surface glycoprotein, a specific surface marker found in embryonic stem cells, normal tissue stem cells, stem cell niches, and circulating endothelial progenitors as well as CSCs [8].

Single-nucleotide polymorphisms (SNPs) have been believed to harbor information about genetic variation in functionality of the genome and susceptibility to cancer, representing a rich source for investigating functional biomarkers implicated in cancer etiology as well as patient treatment outcomes [9]. For example, a common missense SNP (Met1Thr, rs2294008) in exon 1 of the prostate stem cell antigen (PSCA) gene is not only associated with susceptibility to diffuse-type GC in Japanese [10], Korean [10], Chinese [11] and Caucasian [12] populations, but also with survival in Chinese patients with diffuse-type GC [13].

The present address of Qingyi Wei is Duke Cancer Institute, Duke University School of Medicine, 10 Bryan Searle Drive, Durham, NC 27710.

Abbreviations: GC, gastric cancer; CSC, cancer stem cell; microRNA, miRNA; SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; UTR, untranslated regions.

Grant sponsor: Texas 4000 Endowed Distinguished Professorship.

*Correspondence to: Department of Epidemiology, Box 189, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. E-mail: qwei@mdanderson.org

Received 11 January 2013; Revised 29 October 2013; Accepted 1 November 2013

DOI 10.1002/mc.22113

Published online in Wiley Online Library (wileyonlinelibrary.com).

Although its function still remains unclear, *CD133* is broadly found among normal tissue stem cells as well as putative CSC populations, thought to serve as a marker of asymmetric division, lineage plasticity, tumor cell dormancy, and inherent embryonic gene expression [8]. A recent case-control study in a Chinese population [14], in which 13 potentially functional SNPs were genotyped in the promoter, exon and 3′-UTR regions of *CD133*, showed that rs2240688C variant genotypes were associated with reduced risk of lung cancer and may also be associated with a protective effect on OS. Further, the rs2240688A-to-C transition was found to reduce *CD133* expression as measured by a new binding site of microRNA hsa-miR-135a/b in a functional assay. In another study, Pohl et al. genotyped three SNPs (rs2240688, rs3130, and rs2286455) located in the 3′- untranslated region of *CD133* in 91 patients with metastatic colorectal cancer who had been treated with bevacizumab-based chemotherapy, they found that patients with the favorable combination of alleles (either homozygous for C/C in both polymorphisms or the combination of C/T in rs2086455 with either C/T or T/T in rs3130) showed a significantly increased PFS, compared with a decreased PFS for patients with C/C in one polymorphism and C/T or T/T in the other polymorphism [15]. Therefore, we hypothesized that functional stem cell marker SNPs in *CD133* could modulate susceptibility to and prognosis of GC. In the present study, we tested this hypothesis by assessing the associations of functional SNPs in *CD133* with both GC risk and survival.

MATERIALS AND METHODS

Study Subjects

Study subjects were patients with newly diagnosed and histologically confirmed GC at The University of Texas MD Anderson Cancer Center (Houston, TX) between March 2002 and February 2012, who were recruited for an ongoing case-control study. Meanwhile, cancer-free control subjects were also recruited by using frequency matching on age, sex, and ethnicity from an ongoing molecular epidemiology study of the head and neck cancer in the nearly same period [16]. These cancer-free control subjects were hospital visitors who were neither seeking health care, nor genetically unrelated to the cases. Additional information about risk factors, such as smoke [current smoker, former smoker (quit >1 year), and never smoked] and alcohol use [current drinker, former drinker (quit >1 year), and never drinker], family history of any cancer in first-degree relatives, was collected from each eligible subject. A written informed consent was obtained, and a one-time blood sample was taken from each of the subjects. The study protocol was approved by The University of Texas MD Anderson Cancer Center institutional review board.

Outcome Data Collection

All cases included patients who had adenocarcinoma located at the gastroesophageal junction (GEJ, entailing tumors arising between 5 cm proximal to the *z*-line and 5 cm distal to the *z*-line into the stomach) or at the stomach. The overall survival (OS) time was calculated from the date of registration at M.D. Anderson to the date of last contact or death. Patients who were still alive at the last contact were considered as a censored event in the analysis. The age at diagnosis, sex, and type of treatments (i.e., surgery and chemotherapy) were used as covariates in the analysis. *H. pylori* infection status was confirmed in serum *H. pylori* antibody tested by ELISA.

SNP Selection and Genotyping

Genomic DNA was extracted from the buffy coat fraction of each blood sample by using a Blood Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. DNA purity and concentrations were determined by spectrophotometric measurement of absorbance at 260 and 280 nm by UV spectrophotometer.

All common [minor allele frequency (MAF) ≥ 0.05 in Caucasians] SNPs with potentially functional significance [that is, located in the splicing regulation region, the transcription factor binding site (TFBS), the coding regions with amino acid changes, or the 3′-untranslated regions (3′-UTRs), according to NCBI dbSNPs and SNPinfo Web Server (http://snpinfo.niehs.nih.gov/snpfunc.htm) (the last search date: May, 2012)] in *CD133* were identified (Supplementary Fig. S1). As a result, four *CD133* SNPs (rs2240688A>C, rs7686732C>G, rs10022537T>A, and rs3130 C>T) were selected and genotyped by using the TaqMan methodology in 384-well plates, and the outputs were read with the Sequence Detection Software on an ABI-Prism 7900 instrument, according to the manufacturer's instructions, Applied Biosystems (Foster City, CA), which had also supplied primers and probes. Each plate included four negative controls, duplicated positive controls and eight repeated samples. The conditions of amplification were as follows: 50°C for 2 min, 95°C for 10 min followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min.

Statistical Analysis

The χ^2 tests were performed to compare the distributions of demographic variables and selected risk factors, such as smoking status, alcohol use and family history of cancer, between cases and controls. The Hardy–Weinberg equilibrium was tested by a goodness-of-fit χ^2 test to compare the observed genotype frequencies with the expected ones in cancer-free controls. The associations of genotypes of *CD133* SNPs with risk of GC were estimated by

calculating the odds ratios (OR) and their 95% confidence intervals (CIs) from both univiariate and multivariate logistic regression models, followed by stratification analysis. All these analyses were performed with or without adjustment for demographic variables and selected risk factors. Kaplan–Meier plots and log-rank test were applied to assess differences in OS by genotypes from each SNP. Hazard ratios (HR) and their corresponding 95% CIs for OS were estimated by applying the Cox proportional hazards regression model with adjustment for ethnicity, alcohol, stage, location, and application of surgery, chemotherapy, and radiotherapy. Proc HAPLOTYPE in SAS/Genetics software using the expectation–maximization (EM) algorithm was conducted to generate maximum likelihood estimates of haplotype frequencies based on the observed genotypes. Haplotypes with frequencies <5% were combined into one group. All tests were two-sided, and a $P < 0.05$ was considered the cut-off for statistical significance. All of the statistical analyses were performed with Statistical Analysis System software (Version 9.2; SAS Institute, Cary, NC).

RESULTS

Demographic Characteristics and Risk Factors for GC

The present study included 371 GC patients and 313 cancer-free controls. Table S1 summarizes the distribution of demographic characteristics and selected risk factors for GC. Due to frequency matching used in the study design, there were neither significant difference in distributions of age (median age of 59 years), sex and ethnicity, nor were there differences in frequency distributions of smoking, drinking, and family cancer history status between cases and controls. However, these variables were further adjusted for in further multivariate logistic regression models to control for any residual confounding on the main effect of selected SNPs.

CD133 Genotypes and Risk of GC

Genotype and allele distributions of the selected four functional *CD133* SNPs between the cases and controls are listed in Table 1. The observed genotype frequencies of these four SNPs were all in agreement with those expected by the Hardy–Weinberg equilibrium in control subjects ($P = 0.360$ for rs2240688 A>C, $P = 0.721$ for rs7686732 C>G and $P = 0.886$ for rs10022537 T>A), except for rs3130 ($P = 1.833E\text{-}10$ for rs3130 C>T). Compared with the rs2240688 A allele, the C allele was associated with significantly increased GC risk in an allele dose–response manner ($P_{trend} = 0.037$); Furthermore, compared with the AA genotype, the AC + CC genotypes were associated with significantly increased GC risk under a dominant model (adjusted OR = 1.52, 95% CI = 1.09–2.13). For the *CD133* rs3130 C>T SNP, another *CD133* SNP in the miRNA binding sites, the TT genotype was associated with significantly reduced GC risk under a recessive model, compared with the CC + CT genotypes (adjusted OR = 0.68, 95% CI = 0.48–0.97; Table 1).

We further performed the stratified analysis for the *CD133* SNPs in the miRNA binding site, rs2240688 A>C, by age, sex, ethnicity, smoking status, drinking status, and family history of cancer. Compared with the AA genotype, the elevated risks associated with rs2240688 AC + CC variant genotypes did remain statistically significant in the subgroup of <59 years, males, current smokers, ever drinker, and with family history ($P < 0.05$), but a borderline elevated risk in the subgroup of ever smokers ($P = 0.051$; Table 2).

Characteristics of the GC Patients

Clinical and pathological characteristics of the 336 patients who had available follow-up data on outcome are shown in Table S2. There were 207 males (61.61%) and 129 females (38.39%), whose median age was 59 years. Using the Cox hazards regression analysis for OS by clinicopathological characteristics, we found that age, sex, smoking, family history, and *H. pylori* infection status were not statistically associated with OS ($P = 0.275$, 0.168, 0.297, 0.092, 0.906, and 0.507, respectively), but ethnicity and alcohol status were ($P = 0.045$ and 0.016, respectively).

Of the 260 patients with follow-up data, 188 (77.38%) had tumors located at the stomach and 76 (22.62%) had tumors located at the GEJ. Of these, 182 (54.17%) patients were intestinal 130 (38.69%), signet ring (8.34%), and others 24 (7.14%). We grouped the types of differentiation into the following two categories: poor and moderate-well, and the number and percentage of these two groups were 75 (22.32%) and 255 (75.89%; six cases with data missing), respectively. In all patients, clinicopathological characteristics, including histology and differentiation status, were not significantly associated with OS in the univariate analysis ($P = 0.691$ and 0.061, respectively), but tumor location was ($P < 0.0001$). Clinical tumor stages according to the International Union against Cancer (UICC) criteria were as follows: 97 (28.87%) had stage I + II and 215 (63.99%) had stage III + IV. Among the 167 patients, 54 (16.07%) received surgery, 205 (61.01%) received chemotherapy, and 65 (19.35%) received radiotherapy; at the end of the follow-up period, 178 (52.98%) patients had died. The median follow-up time was 18.0 months, and the median survival time for all patients was 29.00 ± 3.30 months. Advanced stage, surgery, chemotherapy, and radiotherapy were all associated with OS ($P = 0.000$, 0.000, 0.002, and 0.005, respectively; Table S2). Because ethnicity, alcohol, stage, location, application of surgery, chemotherapy, and radiotherapy may be confounding factors for the effect of the genotypes on OS, these variables were further adjusted for in the multivariable analysis for the main effect of the selected *CD133* SNPs.

Table 1. Logistic Regression Analysis of Associations between the Genotypes of *CD133* Variants and Gastric Cancer Susceptibility

Genotypes	Cases		Controls		*P*-value[a]	Adjusted OR[b] (95% CI)	*P*-value[c]
	n	%	*n*	%			
PROM1 rs2240688 A>C					0.124		
AA	224	60.38	213	68.05		1.00	
AC	126	33.96	87	27.80		1.49 (1.05–2.11)	**0.027**
CC	20	5.39	13	4.15		1.76 (0.84–3.72)	0.136
P_{trend}					**0.037**		
Dominant model		AC + CC vs. AA			**0.042**	**1.52 (1.09–2.13)**	**0.014**
Recessive model		CC vs. AA + AC			0.448	1.54 (0.74–3.21)	0.249
A	574	77.36	513	81.95	**0.045**	1.00	**0.013**
C	166	22.37	113	18.05		**1.43 (1.08–1.89)**	
PROM1 rs7686732 C>G					0.518		
CC	280	75.41	255	81.47		1.00	
CG	74	19.95	56	17.89		1.13(0.74–1.71)	0.572
GG	4	1.08	2	0.64		1.36 (0.19–9.91)	0.763
P_{trend}					0.262		
Dominant model		CG + GG vs. CC			0.295	1.14 (0.75–1.71)	0.545
Recessive model		GG vs. TT + GT			0.517	1.33 (0.18–9.70)	0.778
C	634	85.44	566	90.41	0.267	1.00	0.538
G	82	11.05	60	9.58		1.12 (0.77–1.64)	
PROM1 rs10022537 T>A					0.264		
TT	225	60.65	209	66.77		1.00	
TA	124	33.42	88	28.12		1.28 (0.57–2.87)	0.548
AA	20	5.39	15	4.79		1.04 (0.48–2.28)	0.922
P_{trend}					0.143		
Dominant model		AA + TA vs. TT			0.104	0.84 (0.60–1.17)	0.302
Recessive model		AA vs. TA + TT			0.719	0.89 (0.41–1.94)	0.777
A	574	77.36	506	80.83	0.133	1.00	0.444
T	164	22.10	118	18.85		1.12 (0.84–1.49)	
PROM1 rs3130 C>T					**0.050**		
CC	128	34.50	108	34.50		1.00	
CT	146	39.35	100	31.95		1.17 (0.80–1.72)	0.427
TT	96	25.88	105	33.55		0.73 (0.49–1.10)	0.135
P_{trend}					0.210		
Dominant model		CT + TT vs. CC			0.980	0.95 (0.68–1.33)	0.754
Recessive model		TT vs. CC + CT			**0.030**	**0.68 (0.48–0.97)**	**0.033**
C	402	54.18	316	50.48	0.157	1.00	0.111
T	338	45.55	310	49.52		0.83 (0.66–1.04)	

[a]Two-sided χ^2 test for either genotype distribution or allele frequency.
[b]Adjusted for age, sex, ethnicity, smoking status, drinking status and family history of cancer in a logistic regression model.
[c]Two-sided χ^2 test for difference in frequency distribution of genotype between cases and controls by adjusted for age, sex, ethnicity, smoking status, drinking status and family history of cancer.
The results are in bold, if statistically significant.

CD133 Genotypes and GC Survival by Clinicopathological Characteristics

The genotype distributions of the four SNPs in *CD133* and their associations with OS are summarized in Table 3. In all patients, variant genotypes of rs2240688 A>C and rs3130 C>T were statistically significantly associated with OS (log-rank $P = 0.047$ under an additive model for rs2240688, Figure 1A; log-rank $P = 0.033$ under a recessive model for rs2240688, Figure 1B; log-rank $P = 0.018$ under an additive model for rs3130, Figure 1C; log-rank $P = 0.017$ under a recessive model for rs3130, Figure 1D). After adjustment for ethnicity, alcohol, stage, tumor location, surgery, chemotherapy and radiotherapy, the rs2240688 C variant genotype were not significantly associated with OS, while the rs3130 T variant genotypes remained significantly associated with OS (adjusted $P_{trend} = 0.016$ under an additive model Figure 1C; adjusted HR = 0.77, 95% CI = 0.63–0.93, adjusted $P = 0.007$ under a recessive model Figure 1D), but this association was not observed for the other two SNPs investigated in this study. We further performed the stratified analysis for rs3130 C>T by age, sex, ethnicity, smoking status, drinking status, family history of cancer, application for surgery, chemotherapy and radiotherapy. Compared with the CC + CT genotype, the OS associated with

Table 2. Stratified Analysis for Associations between the *CD133* rs2240688 Polymorphism and Gastric Cancer Susceptibility

	CD133 rs2240688 A>C (cases/controls)				
Variables	AC + CC	AA	Crude OR (95% CI)	Adjusted OR[a] (95% CI)	*P*-value[b]
Age, year					
<59	69/45	97/102	**1.61 (1.01–2.56)**	**1.82 (1.11–2.94)**	**0.018**
≥59	77/54	127/111	1.25 (0.81–1.92)	1.28 (0.81–2.04)	0.286
Gender					
Males	95/67	135/140	**1.47 (0.99–2.17)**	**1.64 (1.08–2.50)**	**0.021**
Females	51/32	89/73	1.30 (0.75–2.22)	1.33 (0.74–2.38)	0.341
Ethnicity					
White	99/72	131/134	1.41 (0.95–2.08)	1.49 (0.99–2.22)	0.054
Non-white	47/27	93/79	1.47 (0.85–2.56)	1.56 (0.42–2.86)	0.155
Smoking					
Never	68/53	106/103	1.25 (0.79–1.96)	1.41 (0.88–2.72)	0.157
Former	56/36	83/70	1.32 (0.78–2.22)	1.37 (0.78–2.44)	0.266
Current	20/10	35/35	**2.00 (0.82–4.76)**	**2.94 (1.11–7.69)**	**0.029**
Ever (F + C)	76/46	118/105	**1.47 (0.93–2.33)**	**1.67 (1.03–2.70)**	**0.037**
Drinking					
Never	68/45	116/93	1.20 (0.76–1.92)	1.33 (0.81–2.17)	0.251
Former	22/18	43/46	1.30 (0.62–2.78)	1.37 (0.60–3.13)	0.446
Current	54/36	63/69	1.64 (0.95–2.86)	1.61 (0.89–2.86)	0.112
Ever (F+C)	76/54	106/115	**1.52 (0.99–2.38)**	**1.59(1.00–2.56)**	**0.051**
Family history					
Yes	78/58	103/115	**1.49 (0.97–1.37)**	**1.56 (1.01–2.44)**	**0.048**
No	46/41	69/93	1.52 (0.89–2.56)	1.52 (0.88–2.56)	0.134

[a]Adjusted for age, sex, ethnicity, smoking status, drinking status and family history of cancer in a logistic regression model where it was appropriate.
[b]Two-sided χ^2 test for the ORs obtained from the multivariate logistic regression with adjustment for age, sex, ethnicity, smoking status, drinking status and family history of cancer.
The results are in bold, if statistically significant.

the rs3130 TT variant genotype also remained statistically significant in the subgroup of <59 years, male, white, ever smokers, never drinker, without family history, and without surgery, without chemotherapy nor without radiotherapy ($P < 0.05$; Table 4). We found that rs2240688 C variant genotypes were significantly associated with tumor location under all three genetic models, with metastasis and clinical stage under both addictive and recessive models and that rs3130 T variant genotypes were only associated with *H. pylori* infection under a dominant model (Table 5).

CD133 Haplotypes and GC Risk and Survival

We also explored the haplotypes in determining their association with GC risk and survival. As shown in Figure S1, based on genotyping data of 371 cases and 313 controls, these four validated *CD133* SNPs are in low LD with the highest found for rs3130 and rs2240688 ($r^2 = 0.13$). Six haplotypes were shown to have frequencies >5% among all the cases and control, while other less common haplotypes (frequencies <5%) were combined into one group in the analysis. The six common haplotypes in the order of *CD133* rs2240688A>C, rs7686732C>G, rs10022537T>G and rs3130T>C (A-C-T-T, A-C-T-C, C-C-T-C, A-C-A-C, A-C-A-T, and A-G-T-T) accounted for 88.68% and 92.65% of the chromosomes of the cases and controls, respectively. Of these, A-C-T-T and C-C-T-C were shown to be significantly different between the cases and the controls (adjusted $P = 0.010$ and 0.001, respectively) and C-C-T-C was associated with increased GC risks (crude OR = 1.67, 95% CI = 1.20–2.31 and adjusted OR = 1.76, 95% CI = 1.25–2.49, respectively), compared with the A-C-T-T haplotype. A-C-T-C was statistically significantly associated with poor OS (crude HR = 1.14, 95% CI = 0.85–1.53 and adjusted HR = 1.46, 95% CI = 1.07–1.99, adjusted $P = 0.017$, respectively), compared with the A-C-T-T haplotype. Although the combined group of other uncommon haplotypes was also associated with statistically significantly increased GC risk and survival, the results might have been biased due to inadequate sample sizes (Table S3).

DISCUSSION

In this hospital-based case-control study, we found that the variant rs2240688 C allele in the miRNA binding site of the stem cell marker gene *CD133* was associated with significantly increased GC risk in an allele dose-response manner, whereas for rs3130C>T, another SNP in the miRNA binding site of *CD133*, the TT genotypes were associated with significantly

Table 3. Associations between the *CD133* Genotypes and Overall Survival of Gastric Cancer Patients

Genotypes	No. of patients 688 A>C	No. of deaths	MST (months)	Log-rank *P*	Adjusted HR (95% CI)[a]	Adjusted *P*[b]
PROM1 rs2240						
AA	198	108 (54.55)	28.00	**0.047**	1.00	0.748[b]
AC	121	59 (48.76)	31.00		0.94 (0.67–1.33)	0.741
CC	16	11 (68.75)	16.00		1.22 (0.64–2.33)	0.546
Dominant model	AC + CC vs. AA		31.00 vs. 28.00	0.467	0.98 (0.71–1.35)	0.901
Recessive model	CC vs. AC + AA		16.00 vs. 30.00	**0.033**	1.25 (0.66–2.35)	0.492
PROM1 rs7686732 C>G						
CC	259	139 (53.67)	—	—	1.00	0.633[b]
CG	61	31 (50.82)	—		1.22 (0.81–1.83)	0.340
GG	3	0 (0)	—		0.00 (0.00–0.00)	0.942
Dominant model	CG+GG vs. CC		40.00 vs. 28.00	0.779	0.99 (0.66–1.50)	0.977
Recessive model	GG vs. CT+CC		—	0.095	0.00(0.00–0.00)	0.942
PROM1 rs10022537 T>A						
TT	204	114 (55.88)	30.00	0.849	1.00	0.964[b]
TA	113	55 (48.67)	29.00		1.09 (0.51–2.33)	0.824
AA	17	8 (47.06)	40.00		1.05 (0.50–2.19)	0.896
Dominant model	TA + AA vs. TT		29.00 vs. 30.00	0.577	1.04 (0.75–1.44)	0.819
Recessive model	AA vs. TA + TT		40.00 vs. 29.00	0.775	0.97(0.47–2.01)	0.940
PROM1 rs3130 C>T						
CC	114	62 (54.39)	31.00	**0.018**	1.00	**0.016**[b]
CT	134	79 (58.96)	20.00		1.19 (0.84–1.68)	0.327
TT	87	37 (42.53)	43.00		0.64 (0.42–1.00)	**0.048**
Dominant model	CT + TT vs. CC		28.00 vs. 31.00	0.836	0.95 (0.69–1.31)	0.759
Recessive model	TT vs. CT + CC		43.00 vs. 26.00	**0.017**	0.77 (0.63–0.93)	**0.007**

MST, median survival time.
[‡]Mean survival time was provided when MST could not be calculated [b]*P* for trend.
[a]Adjusted for ethnicity, alcohol, stage, tumor location, application of surgery, chemotherapy and radiotherapy (Yes or No).
[b]*P* values were obtained from the Cox hazards model with adjustment for ethnicity, alcohol, stage, tumor location, application of surgery, chemotherapy and radiotherapy (Yes or No).
The results are in bold, if statistically significant.

reduced GC risk under a recessive model. In all patients, variant genotypes of rs2240688 and rs3130 were statistically significantly associated with OS; and variant genotypes of rs2240688 were also significantly associated with tumor location, metastasis and clinical stage, whereas rs3130 T variant genotypes were associated only with *H. pylori* infection. After adjustment for ethnicity, alcohol, stage, tumor location, surgery, chemotherapy and radiotherapy, the rs2240688 C variant genotypes were not significantly associated with OS, while the rs3130 TT genotype remained significantly associated with OS. Additionally in the stratified analysis, the rs2240688A>C variant genotypes also showed some effects in subgroups of smoking, drinking family history status and surgery, chemotherapy and radiotherapy. In further haplotype analysis, the C-C-T-C haplotype was associated with a significantly increased GC risk, and the A-C-T-C haplotype was statistically significantly associated with poor OS, both compared with the A-C-T-T haplotype. However, because these SNPs are potentially functional (Table S4) and are not in LD (Table SS), the haplotype effects may be attributed to single SNP effects.

The CSC theory proposes that cancers are maintained by subpopulations of tumor cells that possess stem cell and progenitor characteristics. These cells can initiate tumor formation, differentiate along multipotent pathways, and they are relatively resistant to conventional chemotherapy [17] and will eventually form metastasis as a "seed"[18]. Recent studies have defined a subset of cancer stem cells found in cancers of the breast [19], brain [20], blood [21], colon [22], pancreas [23], lung [24], liver [25], skin [26], ovaries [27], bladder [28], head and neck [29], and stomach [30,31].

CD133 (prominin-1) is a 120-kDa glycoprotein with an N-terminal extracellular domain and two large extracellular loops, which are strongly N-glycosylated, and an intracellular C-terminus [32]. In GC, nearly half of human tumor samples are positive for *CD133*, and the positivity seems to be correlated with tumor progression [33,34], but it is controversial whether CD133 is a surface marker that may enrich the putative gastric CSC fraction [35].

While tobacco use and alcoholic consumption have been regarded as the major risk factors for GC and CSC has been hypothesized as the origin of GC, few studies have investigated the relationship between tobacco use, alcoholic consumption and gastric stem cell. In the present study, SNPs in the miRNA binding sites of the stem cell marker gene *CD133*, rs2240688 A>C and

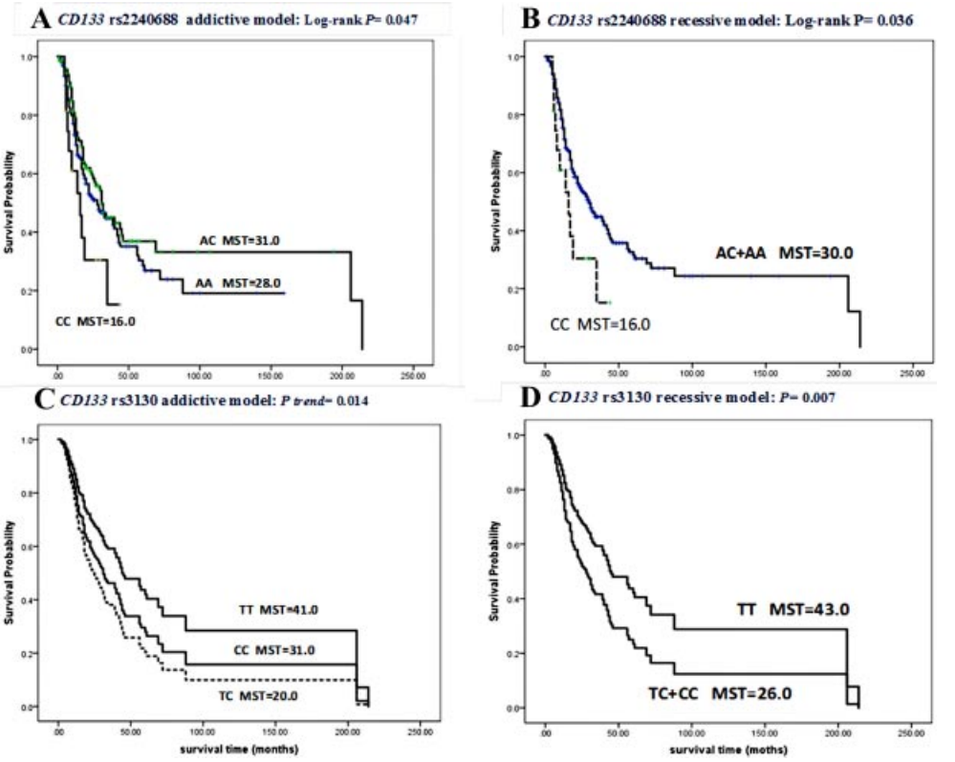

Figure 1. Survival curves of gastric patients by CD133 genotypes of rs2240688 (A and B) and rs3130 (C and D) under both additive and recessive genetic models.

rs3130 C>T, showed statistically significant associations with GC risk in subgroups of smokers, drinkers and those with a family history of cancer. These data inferred that the miRNA binding site SNPs in the 3′UTR of *CD133* may be functional by interacting with environmental risk factors and stem cells and thus may contribute to GC susceptibility, which are consistent with stem cell and CSC hypothesis.

Previous study has indicated that expression of the CD133 protein and mRNA is associated with the occurrence of lymph node metastasis, later TNM stage and poorer survival of GC patients [33,34,36]. Indeed, in the present study, variant genotypes of rs2240688 and rs3130 in the miRNA binding sites of *CD133* were also statistically significantly associated with OS, whereas variant genotypes of rs2240688 were significantly associated with metastasis and clinical stage. Bioinformatics analysis (SNPinfo Web server: http://snpinfo.niehs.nih.gov/) predicted that the rs2240688 A-to-C transition gained a new binding site of the microRNA hsa-miR-135a/b or has-miR-942, and hsa-miR-519e to target *CD133* rs3130 C or T allele. Recently, Cheng et al. validated that the rs2240688 A-to-C transition gained a new binding site of the microRNA hsa-miR-135a/b and decreased the CD133 expression by functional assays [14]. Since the sequence complementarity play an essential role in the interaction of miRNA with its targets, it is conceivable that SNPs in a target-binding site could alter the miRNA–mRNA interaction and thus affect the expression of miRNA targets. So, the present study suggests that the metastasis may be an intrinsic property of small cellular populations—CSC and that CD133 might modify metastasis competence of GC by the miRNA binding site polymorphisms, which could be a putative target for improved therapies for treating metastasis.

Previous reports indicated that GEJ may be different from distant GC with poorer prognosis [37,38]. In the present study, we showed that GEJ had poorer OS than distant GC (MST 17 vs. 34 months), that distributions of the variant genotypes of rs2240688 were significantly different between GEJ and distant GC, and that GC cases with the homozygous CC variant genotype also had a greater risk of GEJ and a poorer survival. Interestingly, as shown in the SNPinfo Web Server (http://snpinfo.niehs.nih.gov/snpfunc.htm), rs2240688 was the only common functional SNP predicted as having different effects between alleles in the *CD133* gene. Given that GEJ may be different from GC because of different cultural or epidemiologic/epigenetic factors involved [39,40], our results also suggested that CD133 might

Table 4. Stratified Analysis for Associations between the *CD133* rs3130 Polymorphism and Overall Survival of Gastric Cancer Patients

	PROM1 rs3130 C>T (deaths/patients)				
Variables	TT	CT + CC	MST (months)	Adjusted OR [a] (95% CI)	*P*-value[b]
Age					
<59	16/41	67/115	43.00 vs. 24.00	**0.56 (0.31–0.98)**	**0.043**
≥59	21/46	74/133	34.00 vs. 28.00	0.64 (0.36–1.13)	0.127
Gender					
Males	25/53	94/154	34.00 vs. 22.00	**0.55 (0.34–0.88)**	**0.013**
Females	12/34	47/94	90.30 vs. 68.80	0.77 (0.38–1.57)	0.470
Ethnicity					
White	25/52	103/159	34.00 vs. 22.00	**0.51 (0.31–0.83)**	**0.007**
Non-white	12/35	38/89	57.00 vs. 39.00	0.70 (0.36–1.40)	0.317
Smoking					
Never	19/46	57/110	57.00 vs. 32.00	0.59 (0.33–1.04)	0.069
Ever (F + C)	18/41	84/138	39.00 vs. 22.00	**0.54 (0.31–0.93)**	**0.028**
Drinking					
Never	15/44	65/120	214.00 vs. 29.00	**0.41 (0.21–0.80)**	**0.009**
Ever (F + C)	22/41	76/128	24.00 vs. 24.00	0.80 (0.49–1.31)	0.379
Family history					
Yes	23/49	78/126	43.00 vs. 20.00	0.65 (0.39–1.07)	0.090
No	11/23	55/87	39.00 vs. 23.00	**0.50 (0.25–1.00)**	**0.049**
Surgery					
Yes	2/16	19/48	80.67 vs. 106.18[‡]	0.38 (0.09–1.72)	0.211
No	35/71	122/200	28.00 vs. 19.00	**0.64 (0.43–0.96)**	**0.031**
Chemotherapy					
Yes	27/49	98/156	28.00 vs. 19.00	0.72 (0.46–1.13)	0.150
No	10/38	43/92	214.00 vs. 44.00	**0.38 (0.17–0.87)**	**0.022**
Radiotherapy					
Yes	3/13	17/40	122.22 vs. 54.76[‡]	0.58 (0.14–2.33)	0.442
No	33/73	124/208	34.00 vs. 22.00	**0.57 (0.38–0.87)**	**0.010**

[a]Adjusted for ethnicity, alcohol, stage, location, application of surgery, chemotherapy, and radiotherapy.
[b]*P* values were obtained from the Cox hazards model with adjustment for ethnicity, alcohol, stage, location, application of surgery, chemotherapy, and radiotherapy.
[‡]Mean survival time was provided when MST could not be calculated.
The results are in bold, if statistically significant.

genetically modify GC risk and prognosis differently in the GEJ and distant GC.

Our data suggested that patients carrying the rs3130TT genotype had a significantly decreased risk of GC, and that GC cases with the TT genotype had significantly better survival; in addition, patients carrying a rs2240688C allele had a significantly increased risk, and GC cases with the homozygous variant CC genotype also had a greater risk of GEJ, metastasis and advanced clinical stage; moreover, two haplotypes were significantly associated with susceptibility to and survival of GC, indicating that these protective/risk effects of the variant alleles did extend to disease severity, a finding consistent with stem cell and CSC hypothesis, suggesting *CD133* as a putative GC stem cell marker.

Because the present study is, to our knowledge, the first study on SNPs of the stem cell marker gene and GC risk or prognosis with a limited sample size, particularly for the subgroup analysis, our findings are best considered preliminary. As one of most common surface markers of stem cell and CSC [8], CD133 SNPs may not be a specific but common marker for susceptibility to and prognosis of cancer. In addition, the current study lacked of information on the *H. pylori* infection status of the control subjects, although we did find that rs3130 variant genotypes were associated with *H. pylori* infection in GC patients. Since *H. pylori* infection in GC patients was relatively uncommon in the US, compared with those in Asian countries [41], all controls and 66.07% (222/336) patients recruited in our study were not tested for the infection upon their visits to the hospital. Therefore, inclusion of the information about *H. pylori* infection should be considered in future larger studies to best illustrate the interaction between environmental and genetic factors in the GC etiology and prognosis. Lastly, current studies had made many comparisons, which might have led to possible false positive findings. Therefore, larger studies are warranted to further assess the role of these SNPs in the etiology and prognosis of various kinds of cancer, particularly in different ethnic groups.

Table 5. *CD133* rs2240688 and rs3130 Polymorphism and Clinico-Pathological Characteristics of Patients ($N = 336$)

Variables	*CD133* rs2240688										*CD133* rs3130									
	Addictive Model			*P*-value[a]	Dominant model		*P*-value[a]	Recessive model		*P*-value[a]	Addictive model			*P*-value[a]	Dominant model		*P*-value[a]	Recessive model		*P*-value[a]
	AA	AC	CC		AA	AC + CC		CC	AC + AA		CC	CT	TT		CC	CT + TT		TT	CT + CC	
H. pylori[‡]																				
No	59	34	2	0.102	59	36	0.162	2	93	0.058	24	47	24	0.092	24	71	**0.034**	71	24	0.784
Yes	8	8	2		8	10		2	16		9	5	4		9	9		14	4	
Histology																				
Intestinal	188	65	8	0.973	108	73	0.875	8	173	0.916	57	77	47	0.504	57	124	0.228	47	134	0.670
Signet ring	77	45	7		77	53		7	123		51	47	32		51	79		32	98	
Other	13	10	1		13	11		1	23		6	10	8		12	18		8	16	
Location																				
Stomach	162	88	9	**0.020**	162	97	**0.018**	9	150	**0.039**	93	10 2	64	0.370	93	166	0.332	64	195	0.181
GEJ	36	33	7		36	40		7	69		21	32	23		21	45		23	43	
Differentiation[‡]																				
Well-mid dle	38	33	3	0.182	38	36	0.102	3	71	0.713	21	34	19	0.405	21	53	0.243	19	55	0.918
Poor	158	84	13		158	97		13	242		91	97	67		91	164		67	188	
Tumor[‡]																				
T_{0-2}	49	38	2	0.561	49	37	0.398	2	84	0.643	34	32	20	0.483	34	52	0.242	20	66	0.832
T_{3-4}	93	50	5		92	55		5	142		47	64	36		47	100		36	111	
Node status[‡]																				
N_{0-1}	74	55	2	0.176	74	57	0.829	2	129	0.073	44	57	30	0.484	44	87	0.293	30	101	0.342
N_{2-3}	6	3	1		6	4		1	9		5	4	1		5	5		1	9	
Metastasis																				
No	100	61	2	**0.012**	100	63	0.416	2	161	**0.003**	53	64	46	0.645	53	110	0.569	46	117	0.360
Yes	98	60	14		98	74		14	158		61	70	41		61	111		41	131	
Stage[‡]																				
I–II	55	41	0	**0.014**	55	41	0.654	0	96	**0.008**	31	42	23	0.862	31	65	0.714	23	73	0.827
III–IV	129	71	15		129	86		15	200		74	87	54		74	141		54	141	

MST, median survival time.
[a]Two-sided χ^2 test.
[‡]Some cases data missed.
The *P* values <0.05 are in bold.

ACKNOWLEDGMENTS

We thank Margaret Lung and Jessica A Fiske for assistance in recruiting the subjects; Min Zhao, Kejing Xu and Jianzhong He for laboratory assistance.

REFERENCES

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
2. Helicobacter and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: A combined analysis of 12 case control studies nested within prospective cohorts. Gut 2001;49:347–353.
3. Tramacere I, Negri E, Pelucchi C, et al. A meta-analysis on alcohol drinking and gastric cancer risk. Ann Oncol 2012; 23:28–36.
4. Gonzalez CA, Agudo A. Carcinogenesis, prevention and early detection of gastric cancer: Where we are and where we should go. Int J Cancer 2012;130:745–753.
5. Yaghoobi M, Bijarchi R, Narod SA. Family history and the risk of gastric cancer. Br J Cancer 2010;102:237–242.
6. Higuchi K, Tanabe S, Azuma M, Sasaki T, Ishido K, Koizumi W. Future perspectives for the development of chemotherapy for advanced gastric cancer: Japanese and global status. Pathobiology 2011;78:334–342.
7. Kim CF, Dirks PB. Cancer and stem cell biology: How tightly intertwined? Cell Stem Cell 2008;3:147–150.
8. Yu X, Lin Y, Yan X, Tian Q, Li L, Lin EH. CD133, stem cells, and cancer stem cells: Myth or reality? Curr Colorectal Cancer Rep 2011;7:253–259.
9. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. Hum Genet 2011;130:59–78.
10. Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet 2008;40:730–740.
11. Wu C, Wang G, Yang M, et al. Two genetic variants in prostate stem cell antigen and gastric cancer susceptibility in a Chinese population. Mol Carcinog 2009;48:1131–1138.
12. Sala N, Munoz X, Travier N, et al. Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: Results from the EPIC-EURGAST study. Int J Cancer 2012;130:2417–2427.
13. Wang M, Bai J, Tan Y, et al. Genetic variant in PSCA predicts survival of diffuse-type gastric cancer in a Chinese population. Int J Cancer 2011;129:1207–1213.
14. Cheng M, Yang L, Yang R, et al. A microRNA-135a/b binding polymorphism in CD133 confers decreased risk and favorable prognosis of lung cancer in Chinese by reducing CD133 expression. Carcinogenesis 2013;34:2292–2299.
15. Pohl A, El-Khoueiry A, Yang D, et al. Pharmacogenetic profiling of CD133 is associated with response rate (RR) and progression-free survival (PFS) in patients with metastatic colorectal cancer (mCRC), treated with bevacizumab-based chemotherapy. Pharmacogenomics J 2013;13:173–180.
16. Liu Z, Wei S, Ma H, et al. A functional variant at the miR-184 binding site in TNFAIP2 and risk of squamous cell carcinoma of the head and neck. Carcinogenesis 2011;32:1668–1674.
17. Dick JE. Future prospects for animal models created by transplanting human haematopoietic cells into immune-deficient mice. Res Immunol 1994;145:380–384.
18. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science 2011;331:1559–1564.
19. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA 2003;100:3983–3988.
20. Singh SK, Clarke ID, Terasaki M, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res 2003;63:5821–5828.
21. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994;367:645–648.
22. Ricci-Vitiani L, Lombardi DG, Pilozzi E, et al. Identification and expansion of human colon-cancer-initiating cells. Nature 2007;445:111–115.
23. Li C, Heidt DG, Dalerba P, et al. Identification of pancreatic cancer stem cells. Cancer Res 2007;67:1030–1037.
24. Kim CF, Jackson EL, Woolfenden AE, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 2005;121:823–835.
25. Wu PC, Lai VC, Fang JW, Gerber MA, Lai CL, Lau JY. Hepatocellular carcinoma expressing both hepatocellular and biliary markers also expresses cytokeratin 14, a marker of bipotential progenitor cells. J Hepatol 1999;31:965–966.
26. Schatton T, Murphy GF, Frank NY, et al. Identification of cells initiating human melanomas. Nature 2008;451:345–349.
27. Szotek PP, Pieretti-Vanmarcke R, Masiakos PT, et al. Ovarian cancer side population defines cells with stem cell-like characteristics and Mullerian Inhibiting Substance responsiveness. Proc Natl Acad Sci USA 2006;103:11154–11159.
28. Chan KS, Espinosa I, Chao M, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. Proc Natl Acad Sci USA 2009;106:14016–14021.
29. Prince ME, Sivanandan R, Kaczorowski A, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proc Natl Acad Sci USA 2007;104:973–978.
30. Takaishi S, Okumura T, Tu S, et al. Identification of gastric cancer stem cells using the cell surface marker CD44. Stem Cells 2009;27:1006–1020.
31. Fukuda K, Saikawa Y, Ohashi M, et al. Tumor initiating potential of side population cells in human gastric cancer. Int J Oncol 2009;34:1201–1207.
32. Corbeil D, Roper K, Fargeas CA, Joester A, Huttner WB. Prominin: A story of cholesterol, plasma membrane protrusions and human pathology. Traffic 2001;2:82–91.
33. Zhao P, Li Y, Lu Y. Aberrant expression of CD133 protein correlates with Ki-67 expression and is a prognostic marker in gastric adenocarcinoma. BMC Cancer 2010;10:218.
34. Ishigami S, Ueno S, Arigami T, et al. Prognostic impact of CD133 expression in gastric carcinoma. Anticancer Res 2010;30:2453–2457.
35. Rocco A, Liguori E, Pirozzi G, et al. CD133 and CD44 cell surface markers do not identify cancer stem cells in primary human gastric tumors. J Cell Physiol 2012;227:2686–2693.
36. Yu JW, Zhang P, Wu JG, et al. Expressions and clinical significances of CD133 protein and CD133 mRNA in primary lesion of gastric adenocacinoma. J Exp Clin Cancer Res 2010; 29:141.
37. Siewert JR, Bottcher K, Stein HJ, Roder JD, Busch R. Problem of proximal third gastric carcinoma. World J Surg 1995;19:523–531.
38. Sakaguchi T, Watanabe A, Sawada H, et al. Characteristics and clinical outcome of proximal-third gastric cancer. J Am Coll Surg 1998;187:352–357.
39. Yao JC, Tseng JF, Worah S, et al. Clinicopathologic behavior of gastric adenocarcinoma in Hispanic patients: Analysis of a single institution's experience over 15 years. J Clin Oncol 2005; 23:3094–3103.
40. An C, Choi IS, Yao JC, et al. Prognostic significance of CpG island methylator phenotype and microsatellite instability in gastric carcinoma. Clin Cancer Res 2005;11:656–663.
41. Guan X, Zhao H, Niu J, Tang D, Ajani JA, Wei Q. The VEGF-634G>C promoter polymorphism is associated with risk of gastric cancer. BMC Gastroenterol 2009;9:77.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.