

Association of cancer stem cell markers genetic variants with gallbladder cancer susceptibility, prognosis, and survival

Anu Yadav¹ · Annapurna Gupta¹ · Neeraj Rastogi² · Sushma Agrawal² · Ashok Kumar³ · Vijay Kumar⁴ · Balraj Mittal¹

Received: 29 June 2015 / Accepted: 12 August 2015 / Published online: 30 August 2015
© International Society of Oncology and BioMarkers (ISOBM) 2015

Abstract Genes important to stem cell progression have been involved in the genetics and clinical outcome of cancers. We investigated germ line variants in cancer stem cell (CSC) genes to predict susceptibility and efficacy of chemoradiotherapy treatment in gallbladder cancer (GBC) patients. In this study, we assessed the effect of SNPs in CSC genes (surface markers CD44, ALCAM, EpCAM, CD133) and (molecular markers NANOG, SOX-2, LIN-28A, ALDH1A1, OCT-4) with GBC susceptibility and prognosis. Total 610 GBC patients and 250 controls were genotyped by using PCR-RFLP, ARMS-PCR, and TaqMan allelic discrimination assays. Chemotoxicity graded 2–4 in 200 patients and tumor response was recorded in 140 patients undergoing neoadjuvant chemotherapy (NACT). Differences in genotype and haplotype frequency distributions were calculated by binary logistic regression. Gene-gene interaction model was analyzed by generalized multifactor dimensionality reduction (GMDR). Overall survival was assessed by Kaplan-Meier survival curve and multivariate Cox-proportional methods. ALCAM A_{rs1157}C_{rs10511244} ($P=0.0035$) haplotype was significantly associated with GBC susceptibility.

Electronic supplementary material The online version of this article (doi:10.1007/s13277-015-3929-6) contains supplementary material, which is available to authorized users.

✉ Balraj Mittal
bml_pgi@yahoo.com

- ¹ Department of Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, UP, India
- ² Department of Radiotherapy, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, UP, India
- ³ Department of Gastroenterology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, UP, India
- ⁴ Surgical Oncology, KGMU, Lucknow, UP, India

In GMDR analysis, ALCAM rs1157G>A, EpCAM rs1126497T>C emerged as best significant interaction model with GBC susceptibility and ALDH1A1 rs13959T>G with increased risk of grade 3–4 hematological toxicity. SOX-2 rs11915160A>C, OCT-4 rs3130932T>G, and NANOG rs11055786T>C were found best gene-gene interaction model for predicting response to NACT. In both Cox-proportional and recursive partitioning ALCAM rs1157GA+AA genotype showed higher mortality and hazard ratio. ALCAM gene polymorphisms associated with GBC susceptibility and survival while OCT-4, SOX-2, and NANOG variants showed an interactive role with treatment response.

Keywords Cancer stem cell (CSC) · Gallbladder cancer (GBC) · Treatment outcomes · Survival recursive partitioning (Rpart)

Introduction

Cancer is known as malignant neoplasia related to uncontrolled proliferation. Until now, there are over 200 different kinds of cancers that can affect the human body [1]. Among them, gallbladder cancer (GBC) is the most common biliary tract cancer in North Indian population, characterized by its high lethality, aggressive nature, and dismal prognosis [2]. For all advances in chemoradiation, the outcome of treatment with advanced GBC is still poor, because of chemoresistance and recurrence due to heterogeneity of tumor cells. Cancer stem cell (CSC) presents in a heterogeneous tumor with properties of tumor progression, chemoresistance, and recurrence [3].

Cancer stem cells are identified by expression of surface markers CD44, ALCAM, EpCAM, CD133, and molecular markers NANOG, SOX-2, LIN-28A, ALDH1A1, and OCT-4. Surface markers are transmembrane glycoprotein in nature

and are involved in cell proliferation, invasion, and metastasis of cancer stem cell [4, 5]. CD44+CD133+ cells showed chemoresistance and CSCs like characteristics in gallbladder cancer cell lines [6, 7]. Our earlier reports have shown the important role of CD44 gene variants in GBC susceptibility [8]. ALCAM-ALCAM interconnections have a role in the development and maintenance of tissue architecture and tumor progression [9]. ALCAM considered as a breast cancer prognostic marker [10]. EpCAM+ cells have been possessed a tumor-initiating role [11]. EpCAM is overexpressed in different tumors, including colon, lung, pancreas, breast, and ovary [12].

Molecular marker ALDH1A1 enzyme belongs to dehydrogenases family of proteins that participates in cellular detoxification and differentiation through the oxidation of intracellular aldehydes [13] while Oct-4, Sox-2, and Nanog transcription factors responsible for pluripotency of stem cell. Oct-4, Sox-2, and Nanog induce expression in cooperative manner for maintaining pluripotency of cancer stem cells [14]. LIN28 encodes a microRNA-binding protein that binds to let-7 microRNA and inhibits production of the mature let-7 microRNA in stem cells [15].

Many studies have reported significant associations of CSC gene polymorphisms with susceptibility and treatment response in cancers [16–18]. Till now; there are a limited number of studies involving CSC genetic variants with GBC [8]. Therefore, in this study, we investigated 15 CSCs (surface and molecular marker) germline polymorphisms that have been previously associated with different cancers, to evaluate their influence on the risk and treatment outcomes of GBC in the North Indian population. In future, some cancer stem cell genetic variants may act as potential predictive and prognostic markers for GBC.

Materials and methods

Recruitment of subjects

In this study, 610 histologically confirmed gallbladder cancer patients and 250 controls were recruited after taking their informed consent. All healthy controls were age-, gender-, and ethnicity-matched and without gastrointestinal disorders and belonged to same geographical region. The study protocol was approved by the Ethics Committee of SGPGIMS. Two hundred GBC patients receiving adjuvant or neoadjuvant chemoradiotherapy, in a sequential manner for three cycles, were followed for treatment response. Radiotherapy was given by IMRT and 3D-CRT techniques. Chemotherapeutic drugs gemcitabine, 5-fluorouracil, and cisplatin were given in combination. Two hundred patients were followed-up for drug toxicity according to the National Cancer Institute (NCI) Common Toxicity Criteria for Adverse Events (CTCAE), version 3.0 (<http://ctep.cancer.gov>). Tumor response was assessed in

140 NACT-treated patients, according to the Response Evaluation Criteria in Solid Tumors (RECIST1.1 criteria). Details of patient enrollment and treatment plan are shown in Fig. 1.

Patients with static and progressive disease were considered as nonresponders while with complete and partial pathological response were considered as responders. In chemotoxicity, grade 0–1 was considered as mild toxicity and grade 2–4 as moderate to severe. Hematological toxicity was in terms of anemia, leucopenia, and thrombocytopenia. We evaluated overall survival on the basis of patient's histopathological and clinical stages in 199 patients. Treatment modality and clinical stage affect the survival pattern, so we divided 199 patients in three groups' postoperative, metastatic, and locally advanced for survival analysis. Median follow-up survival period was 18 months.

Basic characteristics of study population and clinical data of follow up patients are given in Tables 1 and 2, respectively.

SNP selection

Most Informative tagger SNPs (Haploview software 4.2) were selected with >5 % MAF in GIH and CEU population by using the HapMap project database. The sample size was calculated by using QUANTO1.1. This study achieved 80 % power.

Genotyping

Venous blood (4 ml) was collected from patients, and the genomic DNA was extracted from peripheral blood using the standard salting out method [19]. Genotyping of the SNPs was carried out using the TaqMan allelic discrimination assay, ARMS-PCR, and PCR-RFLP. Details of genotyping methods are summarized in Table S1 (supplementary data).

Statistical evaluation

Descriptive statistics of patients were presented as mean and standard deviations for continuous measures and frequencies and percentages for categorical measures. Correlations between various genotypes and treatment outcomes were examined using binary logistic regression. The association was expressed with odds ratios (OR) or risk estimates with 95 % confidence intervals (CI). Statistical analysis was done using SPSS statistical analysis software, version 20.0. Haplotype analysis was done by using the SNPstat Software (bioinfo.iconcologia.net/SNPstats). Global *P* value represents a difference between the means associated with pair of haplotypes. To reduce chances of obtaining false-positive results (type I errors) Bonferroni correction was applied in subgroup analysis. The Bonferroni correction is an adjustment made to *P* values when several dependent or independent statistical tests are being performed simultaneously on a single dataset.

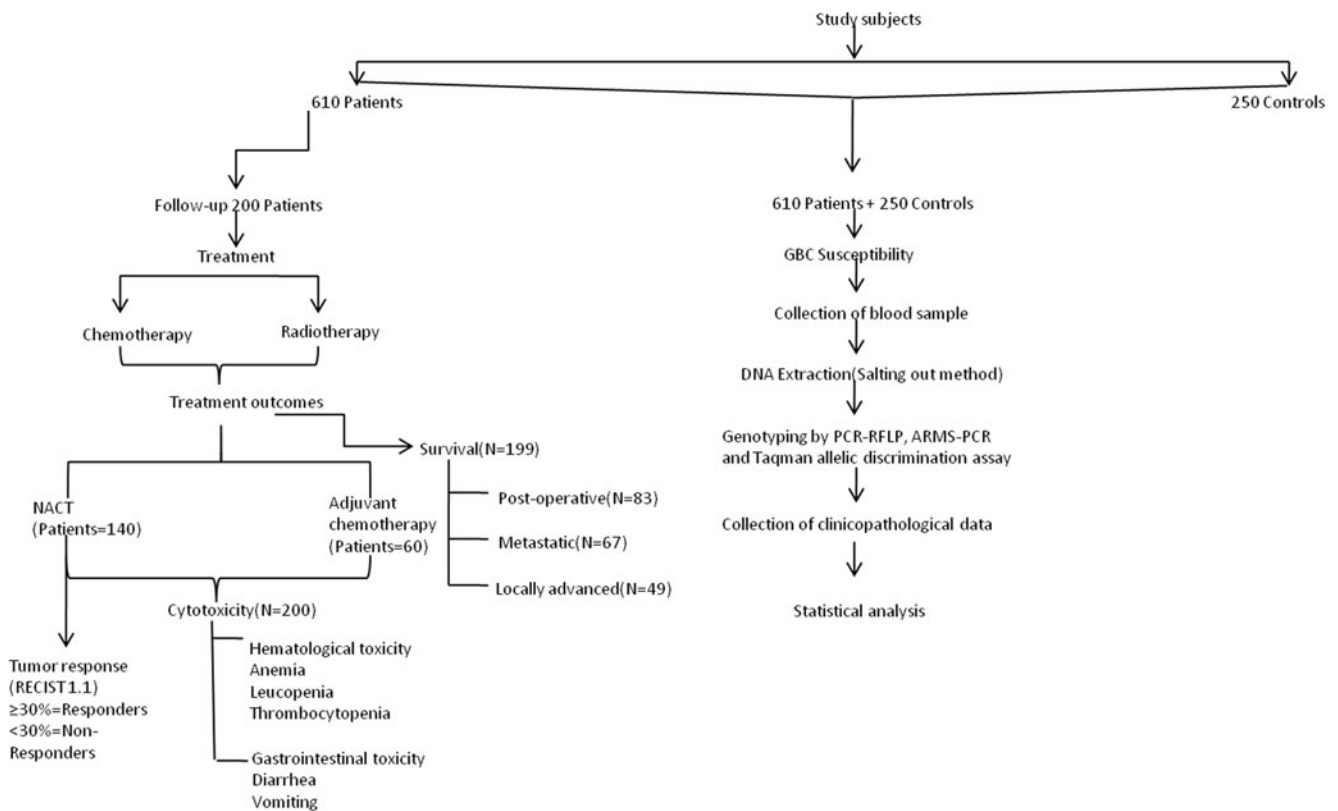


Fig. 1 Study plan flow diagram. GBC susceptibility group and follow-up group with treatment outcomes

To perform a Bonferroni correction, we divided the critical *P* value (0.05) by the number of comparisons being made ($0.05/2=0.025$). The statistical power of the study is then calculated based on this modified *P* value (0.025).

The functional effects of polymorphisms were determined by online F-SNP (<http://compbio.Cs.Queens.CA/F-SNP/>) database. F-SNP database provides integrated information about the functional effects of SNPs obtained from 16 bioinformatics tools and databases and provides output in terms of FS score. The functional effects are predicted and indicated at

Table 1 Characteristics of the subjects

	Cases, <i>N</i> (%)	Controls, <i>N</i> (%)
Subjects	610 (100)	250 (100)
Female	435 (71.3)	145 (58.0)
Male	175 (28.7)	105 (42.0)
Age±SD	52.19±10.4	43.24±11.5
Gallstone present	346 (56.8 %)	None
Gallstone absent	264 (43.2 %)	250 (100)
Tobacco users	146 (25.3 %)	NA
Tobacco nonusers	423 (73.3 %)	NA
Unknown tobacco status	8 (1.4 %)	NA
Early age of onset (<50)	371 (60.8 %)	NA
Late age of onset (≥50)	239 (39.2 %)	NA

NA not applicable

Table 2 Characteristics of followed up patients

GBC200[%]	
Gender	
Male	55 (27.5 %)
Female	145 (72.5 %)
TNM status	
Stage I (T1 N0 M0)	5 (2.5 %)
Stage II (T2 N0 M0)	15 (7.5 %)
Stage IIIA (T3 N0 M0)	21 (10.5 %)
Stage IIIB (T1–3 N1 M0)	27 (13.5 %)
Stage IVA (T4 N0–1 M0)	35 (17.5 %)
Stage IVB (T4 N0-1 M1)	97 (48.5 %)
Radiotherapy	49 (24.5 %)
Chemotherapy	
Adjuvant	60 (30 %)
Neoadjuvant	140 (70 %)
Treatment response	
Responders=CR+PR	48 (34.2 %)
Nonresponders=PD+SD	92 (65.7 %)
Extent of tumor	
Postoperative	84 (42.0 %)
Metastasis	67 (33.5 %)
Locally advanced	49 (24.5 %)

Table 3 Frequency distribution of CSCs surface markers polymorphisms in GBC and controls (age and gender adjusted)

Genotypes	Controls, <i>N</i> (%)	GBC, <i>N</i> (%)	OR (95 % CI)	<i>P</i> value
CD44rs13347C>T				
CC	162 (64.8)	378 (62.0)	1 (reference)	–
CT	80 (32.0)	201 (33.0)	0.87 (0.59–2.88)	0.48
TT	8 (3.2)	31 (5.1)	1.99 (0.80–4.96)	0.14
C	404 (80.8)	957 (78.44)	1 (reference)	–
T	96 (19.2)	263 (21.56)	1.17 (0.7–2.9)	0.12
CD44 rs353639A>C				
AA	167 (66.8)	388 (63.6)	1 (reference)	–
AC	68 (27.2)	174 (28.5)	1.05 (0.70–1.57)	0.81
CC	15 (6.0)	48 (7.9)	1.71 (.795–3.67)	0.17
A	402 (80.4)	950 (77.87)	1 (reference)	–
C	98 (19.6)	270 (22.13)	1.79 (0.27–2.53)	0.13
CD44 rs187116G>A				
GG	99 (39.6)	226 (37.0)	1 (reference)	–
GA	89 (35.6)	230 (37.7)	0.97 (0.64–1.47)	0.89
AA	62 (24.8)	154 (25.2)	0.82 (0.52–1.29)	0.39
G	287 (57.4)	682 (55.9)	1 (reference)	–
A	213 (42.6)	538 (44.1)	0.85 (0.66–1.10)	0.21
CD44 rs187115 T>C				
TT	150 (60.0)	353 (57.9)	1 (reference)	–
TC	70 (28.0)	178 (29.2)	1.13 (0.76–1.70)	0.54
CC	30 (12.0)	79 (13.0)	0.71 (0.41–1.26)	0.24
T	370 (74.0)	884 (72.46)	1 (reference)	–
C	130 (26.0)	336 (27.54)	0.81 (0.61–1.08)	0.15
ALCAM rs1157 G>A				
GG	134 (53.6)	302 (49.5)	1 (reference)	–
GA	105 (42.0)	259 (42.5)	1.23 (0.85–1.77)	0.28
AA	11 (4.4)	49 (8.0)	2.40 (1.09–5.29)	0.03
G	373 (58.65)	863 (70.74)	1 (reference)	–
A	263 (41.35)	357 (29.26)	1.42 (1.07–4.89)	0.016
ALCAM rs10511244 T>C				
TT	174 (69.6)	403 (66.1)	1 (reference)	–
TC	64 (25.6)	160 (26.2)	1.02 (0.68–1.54)	0.92
CC	12 (4.8)	47 (7.7)	2.10 (0.97–4.55)	0.06
T	412 (82.4)	966 (79.18)	1 (reference)	–
C	88 (17.6)	254 (20.82)	1.42 (1.02–4.96)	0.036
EpCAM rs1126497 T>C				
TT	67 (26.8)	161 (26.4)	1 (reference)	–
TC	120 (48.0)	293 (48.0)	0.76 (0.50–1.16)	0.20
CC	63 (25.2)	156 (25.6)	0.87 (0.53–1.42)	0.57
T	254 (50.8)	615 (50.5)	1 (reference)	–
C	246 (49.2)	605 (49.5)	0.99 (0.77–1.27)	0.94
EpCAM rs1421 T>C				
TT	205 (82.0)	484 (79.3)	1 (reference)	–
TC	32 (12.8)	90 (14.8)	1.09 (0.64–1.84)	0.75
CC	13 (5.2)	36 (5.9)	1.27 (0.59–2.74)	0.54
T	442 (88.4)	1058 (86.7)	1 (reference)	–
C	58 (11.6)	162 (13.3)	0.90 (0.62–1.31)	0.58
CD133 rs3130 C>T				
CC	157 (62.8)	371 (60.8)	1 (reference)	–
CT	77 (30.8)	192 (31.5)	1.04 (0.70–1.55)	0.84
TT	16 (6.4)	47 (7.7)	0.99 (0.48–2.04)	0.98
C	391 (78.2)	934 (76.56)	1 (reference)	–
T	109 (21.8)	286 (23.44)	1.09 (0.80–1.48)	0.59
CD133 rs2240688 T>G				
TT	148 (59.2)	353 (57.9)	1 (reference)	–
TG	81 (32.4)	206 (33.8)	1.02 (0.69–1.49)	0.94
GG	21 (8.4)	51 (8.4)	1.45 (0.75–2.82)	0.28
T	377 (75.4)	912 (74.7)	1 (reference)	–
G	123 (24.6)	308 (25.25)	0.98 (0.73–1.30)	0.86

Significant *P* values shown in bold

Table 4 GMDR: gene-gene interaction

GBC susceptibility	Best interaction model	Testing accuracy	CVC	<i>P</i> value	OR (95 % CI)
Molecular marker	OCT-4 rs3130932, LIN-28A rs4274112	0.568	10/10	0.007	1.84 (2.18–5.88)
Surface marker	ALCAM rs1157, EpCAM rs1126497	0.574	10/10	0.005	1.8 (2.17–6.78)
Treatment outcomes	Molecular markers				
Gastrointestinal toxicity	LIN-28A rs4274112	0.594	8/10	0.179	2.97 (1.583–15.18)
Hematological toxicity	ALDH1A1 rs13959	0.613	9/10	0.016	3.0 (1.25–7.47)
Treatment response	NANOGrs11055786, OCT-4rs3130932, SOX-2rs11915160	0.695	7/10	0.003	5.6 (1.71–18.3)
Surface markers					
Gastrointestinal toxicity	CD44 rs187116	0.631	10/10	0.056	5.1 (0.85–31.3)
Hematological toxicity	ALCAM rs1157	0.592	8/10	0.072	2.1 (0.9317–4.77)
Treatment response	CD44 rs187116, CD44 rs187115, CD133 rs2240688	0.747	10/10	0.30	1.16 (2.79–7.0)

Significant *P* values shown in bold

the splicing, transcriptional, translational, and posttranslational level. As such, the F-SNP database helps identify and focus on SNPs with potential pathological effect to human health.

Generalized multifactor dimensionality reduction (GMDR) analysis was done for the adjustment of quantitative and discrete covariants with gene-gene interactions [20]. GMDR replication of permutation determines accuracy of the *P* value assessed by permutation. We used 1000 replication of permutation to get significant *P* values.

The best gene-gene interaction model was obtained by reducing multi-locus genotypes into low-risk and high-risk groups on the basis of highest testing accuracy and permutation results, and cross-validation consistency (CVC) was considered to be statistically significant at the 0.05 level.

Survival curve was constructed using Kaplan-Meier method, and differences between the groups were tested by the log-rank method. The hazard ratio was calculated by Cox-

proportional method at 95 % confidence interval. More intuitive survival model was generated by recursive partitioning (rpart version 3.1). Protein direct and functional interactions were determined by STRING database.

Results

Association of genetic variants with GBC susceptibility

In this study, we evaluated the frequency distribution of genetic polymorphisms in cancer stem cell surface markers CD44, ALCAM, EpCAM, and CD133 and molecular markers ALDH1A1, OCT-4, SOX-2, LIN-28A, and NANOG polymorphisms between healthy controls and GBC patients. All studied polymorphisms were in Hardy-Weinberg equilibrium in controls ($P > 0.05$).

Fig. 2 Kaplan-Meier survival curve with ALCAM rs1157 in metastatic GBC patients

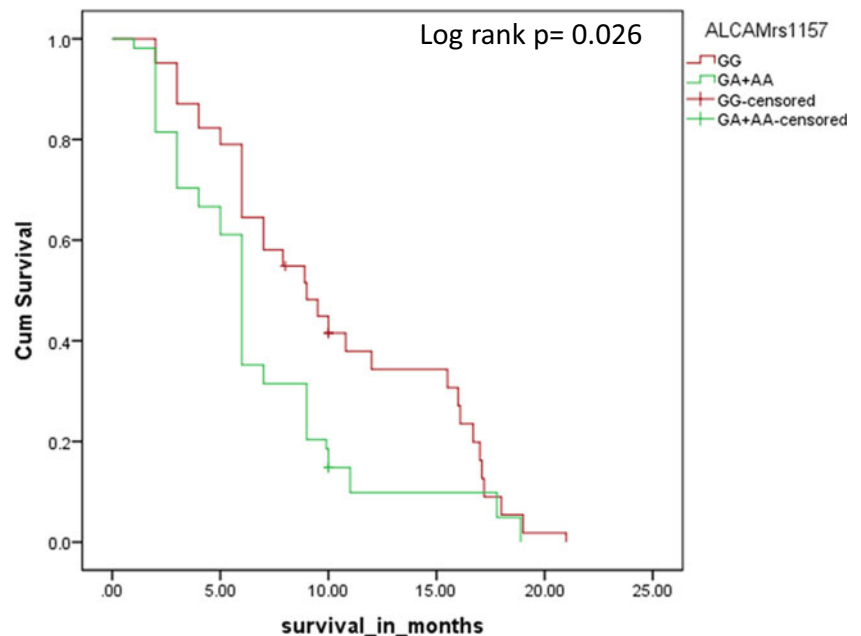


Table 5 Cox-proportional hazard model

Variables	HR	95 % CI	P value
GG and GA+AA genotype for ALCAM rs1157	1.714	1.032–2.844	0.037
Gender	1.154	0.643–2.073	0.631
Gallstone	0.837	0.500–1.399	0.497
Gastrointestinal toxicity	0.598	0.322–1.111	0.104
Hematological toxicity	1.061	0.632–1.780	0.823

The frequencies of ALCAM rs1157G>A and rs10511244T>C polymorphisms were found to be significantly associated with GBC risk, at genotypic [rs1157(AA), OR (CI)=2.4 (1.09–5.29), *P* value=0.03] and allelic levels [rs1157(A), OR (CI)=1.42 (1.07–4.89), *P* value=0.016; rs10511244 (C), OR (CI)=1.42 (1.02–4.96), *P* value=0.036], respectively (Tables 3 and S2).

To determine the effect of multiple single nucleotide polymorphisms within a gene, haplotype analysis was done. Those haplotypes, which did not cross rare haplotype frequency (<0.05 %), were excluded from analysis. Bonferroni correction was applied in subgroup analysis. The A_{rs1157}C_{rs10511244} haplotype of ALCAM was significantly associated with increased risk in GBC patients [Table S3 OR (CI)=2.40 (1.34–4.32), *P* value=0.0035], but the significance was lost after Bonferroni correction. ALCAM A_{rs1157}C_{rs10511244} haplotype in GBC patients harboring gallstones [Table S4 OR (CI)=3.71 (1.90–7.26), *P* value=1e–04*] and GBC females [Table S5 OR (CI)=3.62 (1.53–8.57), *P* value=0.0035*] remained significant after Bonferroni correction (*P*<0.025).

Individual CD44 gene polymorphisms had no significant differences in frequency distribution between GBC cases and controls, but haplotype of CD44 C_{rs13347}A_{rs353639}A_{rs187116}C_{rs187115} [OR (CI)=0.51 (0.27–0.95)], *P* value=0.033] was significantly associated with lower risk of GBC. Here also, the significance was lost after Bonferroni correction (Table S3).

There were no significant associations of EpCAM rs1126497T>C, EpCAM rs1421T>C, CD133 rs3130C>T, CD133 rs2240688T>G, ALDH1A1 rs13959A>G, OCT-4 rs3130932T>G, SOX-2 rs11915160A>C, LIN-28A rs4274112T>C, and NANOG rs11055786T>C polymorphisms with GBC risk in terms of overall case control frequency distribution, gender stratification, and gallstone status (Tables 3, S2, S3, S4, and S5).

Gene-gene interaction for GBC susceptibility GMDR analysis was used to evaluate gene-gene interaction with GBC susceptibility by adjusting age, gender, tobacco, and gallstone status covariants.

We performed separate gene-gene interaction analysis for surface and molecular markers. In surface markers, ALCAM rs1157 and EpCAM rs1126497 [Table 4, OR (CI)=1.8 (1.17–

6.78), *P* value=0.005] was best significant interaction model for GBC susceptibility. In molecular markers OCT-4 rs3130932, LIN-28A rs4274112 [Table 4, OR (CI)=1.84 (1.18–5.88), *P* value=0.007] was best significant interaction model for GBC susceptibility.

Association of CSC genetic variants with treatment response and toxicity

Univariable analysis did not show significant association of variants in surface and molecular marker between responders and nonresponders of chemoradiotherapy treatment outcomes at allele, genotype, and haplotype levels (haplotypes, Table S6). Results were consistent after multivariable analysis. Hematological and gastrointestinal toxicity of treatment in GBC patients also did not associate with any particular CSC genetic variants (Tables S7 and S8).

The GMDR analysis was used to evaluate gene-gene interaction with clinical outcome by adjustment of the covariants tumor stage, age, gender, and treatment modality. The best combination model for clinical outcome is given in Table 4.

For surface markers, we did not get significant interaction model for treatment response. However, in molecular markers, the best significant model was NANOG rs11055786, OCT-4 rs3130932, and SOX-2 rs11915160 (Table 4, OR (CI)=5.6 (1.71–18.3), *P* value=0.003] for treatment response. ALDH1A1 rs13959 (Table 4, OR (CI)=3.0 (1.25–7.47), *P* value=0.016) emerged as the best interaction model for hematological toxicity.

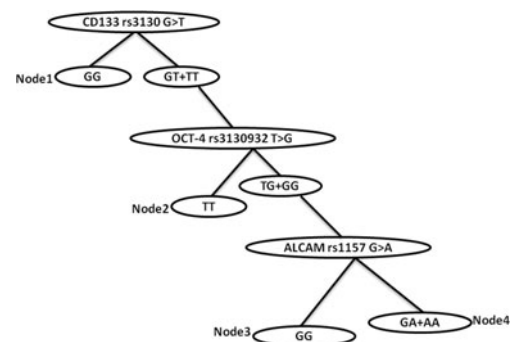
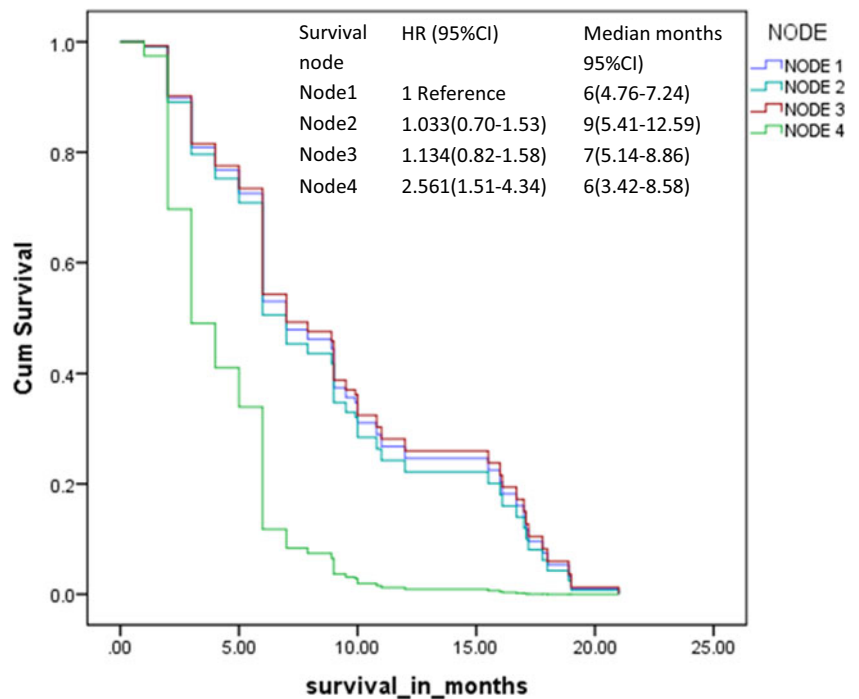
**Fig. 3** Decision tree constructed by recursive partitioning analysis

Fig. 4 Kaplan-Meier survival curve of the risk groups based on a specific gene variant profile including CD133 rs3130G>T, OCT-4 rs3130932, and ALCAM rs1157 G>A



Associations between the CSCs genetic variants and GBC survival Kaplan-Meier survival curves were used to assess the associations between the CSCs polymorphisms and survival time. Survival was calculated separately in the three patient groups’ postoperative, metastatic, and locally advanced. The median survival period for metastatic was 9.5 months. Metastatic GBC cases with ALCAM rs1157GG genotype showed a higher survival rate than the individuals with GA+AA genotype (log-rank $p=0.026$) (Fig. 2). In Cox-proportional hazard model ALCAM rs1157GA+AA genotype in GBC cases, associated with higher hazard ratio (HR=1.7, CI=1.03–2.84, Table 5) as compared with GG genotype. Kaplan-Meier survival curve did not show any significant differences in other CSC polymorphisms’ survival rate between wild and variant genotypes. Model-based classical recursive partitioning is a statistical method that creates a decision tree with the split criteria of log-rank test that strives to correctly classify according survival times and the Kaplan-

Meier graphs. Tree-structured survival analysis of our data represented that ALCAM (GA+AA) end node had highest hazard ratio and lowest survival period as compared to CD133rs3130C>T and OCT-4 rs3130932 (Figs. 3 and 4).

Discussion

Cancer stem cells are responsible for tumor growth, migration, invasion, aggressiveness, resistance, and pluripotency of a tumor. This study was conducted to determine the role of cancer stem cell gene variants in gallbladder cancer susceptibility and treatment outcomes.

CSC genetic variants and GBC susceptibility Our results obtained by analyzing 610 GBC patients and 250 controls demonstrated that ALCAM rs1157G>A and rs10511244T>C polymorphisms were significantly associated with increased

Table 6 In silico analysis

Result of F SNP				
Genetic variation	Functional category	Prediction tool	Prediction result	FS score
CD44 rs13347 C>T	Transcriptional regulation	TFSearch	Changed	0.176
CD44 rs353639 A>C	Transcriptional regulation	TFSearch	Changed	0.176
CD44 rs187116 G>A	Transcriptional regulation	TFSearch	Not changed	0
CD44 rs187115 T>C	Transcriptional regulation	TFSearch	Changed	0.176
ALCAM rs10511244 T>C	Transcriptional regulation	GoldenPath	Exist	0.101
ALCAM rs1157G>A	Transcriptional regulation	GoldenPath	Exist	0.101
ALDH1A1 rs13959 A>G	Splicing regulation	ESRsearch	Changed	0.237

risk in GBC patients at genotype and haplotype analysis. On the other hand, CD44 $C_{rs13347}A_{rs353639}A_{rs187116}C_{rs187115}$ haplotype was significantly associated with a protective effect. The ALCAM and CD44 are cell surface adhesion molecules which mediate adhesion interactions between cell-cell and cell-substrates [9, 21]. Previous studies also showed ALCAM rs1157 homozygous variant to be associated with breast cancer susceptibility [22]. The ALCAM rs1157G>A and ALCAM rs10511244T>C are present in 3'UTR and intron1 region of the gene, respectively. By F-SNP, we hypothesized that the variant alleles of ALCAMrs1157 G>A and ALCAMrs10511244T>C may be involved in transcription regulation (Table 6). Variant allele may result in increased ALCAM expression which may disrupt adhesion interactions that disturb tissue architecture between cells [23].

In our study, CD44 $C_{rs13347}A_{rs353639}A_{rs187116}C_{rs187115}$ haplotype conferred lower risk of GBC which reconfirms our earlier observation in smaller number of samples [8]. However, CD44 rs187116–rs187115 T-A haplotype has been reported to confer higher risk of gastric adenocarcinoma [24]. CD44rs187115 variant genotype was also correlated with risk of oral cancer development [25]. CD44 has been reported to be highly expressed in subserous gallbladder carcinoma [26] and CD44 protein variant overexpression was related to histologic dedifferentiation of gallbladder carcinoma [27]. Several mechanistic studies have shown dualistic nature of CD44, after its ligand hyaluronic acid interaction. In silico analysis using F-SNP predicted change in transcriptional regulation for rs13347, rs353639, and rs187115 SNPs (Table 6), supporting the influence of selected CD44 genetic variants on gene transcription and splicing mechanisms.

On applying logistic regression and haplotype analysis, we did not find significant association of EpCAM rs1126497T>C, EpCAM rs1421T>C, CD133 rs3130C>T, CD133 rs2240688T>G, ALDH1A1 rs13959A>G, OCT-4 rs3130932T>G, SOX-2 rs11915160A>C, LIN-28A rs4274112 T>C, and NANOG rs11055786T>C polymorphisms with overall frequency distribution, gender, and gallstone status. Some other studies showed that EpCAM rs1126497 CT+TT and OCT-4 rs3130932 T>G were associated with susceptibility to breast cancer risk, whereas no association was observed with EpCAM rs1421 [28, 29].

In genetic association studies involving low penetrance genes, effect of individual genes in disease risk is rather limited and gene-gene interaction have been proposed as important contributor in risk assessments. For gene-gene interaction analysis by GMDR with adjusting covariants, we found that SNPs ALCAM rs1157 and EpCAM rs1126497 surface marker as best significant interaction model with increased risk of GBC susceptibility. F-SNP predicted EpCAM rs1126497 functional role in splicing regulation. STRING 10 also suggested ALCAM and EpCAM protein-protein interaction

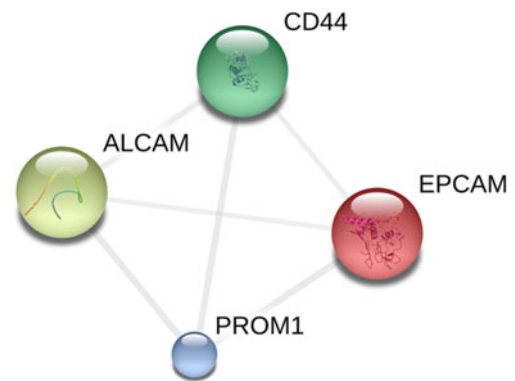


Fig. 5 STRING 10 protein-protein interactions for cancer stem cell surface marker

(Fig. 5). So, there must be an interacting role of ALCAM and EpCAM in GBC susceptibility.

Likewise, GMDR analysis predicted interactive role of OCT-4 rs3130932 and LIN-28A rs4274112 molecular marker with GBC susceptibility. Lin-28a is a mRNA binding protein and facilitates posttranscriptional regulation of OCT-4 protein [30]. Our previous study also reported LIN-28A rs4274112 (AG+GG) association with lymph node metastases and OCT-4rs3130932 with hormone receptor positive tumors in breast cancer [31]. These findings uncover a new level of gene-gene interaction in gallbladder cancer susceptibility.

CSCs genetic variants and GBC treatment response At both univariable and multivariable logistic regression and haplotype analyses, none of the CSC polymorphisms were significantly associated with GBC treatment response. Even after GMDR analysis, CSCs genetic variants for surface markers still did not significantly associate with treatment response and toxicity. In a previous study also, ALCAM rs1157 had no statistical significance with 5-FU-treated colon cancer recurrence [32] while in other study, it was shown to be associated with colon cancer recurrence [17]. CD44rs187116 and rs187115 T-A haplotype had lower risk to develop gastric adenocarcinoma recurrence [24]. The expression of CD44 with tumor prognosis is controversial. Certain studies have indicated that the overexpression of CD44 is correlated with elevated chemoradiotherapy resistance and increased risk of recurrence [33] whereas others have related poor prognosis to CD44 downregulation in tumor cells [34]. For EpCAM

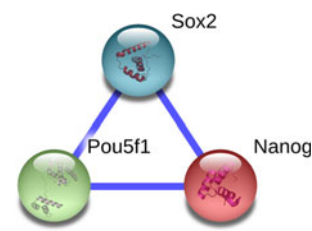


Fig. 6 STRING 10 protein-protein interactions for cancer stem cell molecular markers

rs1126497T>C, rs1421T>C, our results with treatment response and toxicity are in tune with previous studies which also did not show any significant effect of these variants on prognosis in nonsmall cell lung cancer [35].

Variants for molecular marker genes showed an interactive association with treatment outcomes. In GMDR analysis, SNPs in SOX2 rs11915160, OCT4 rs3130932, and NANOG rs11055786 turned out to be the best interaction models for predicting poor response to NACT in GBC. The cooperative interaction sox-oct cis regulatory element is essential for nanog pluripotency. STRING 10 analysis showed interaction among SOX-2, OCT-4, and NANOG proteins (Fig. 6). Our earlier report also suggested SOX-2 rs11915160, OCT-4 rs3130932, and NANOG rs11055786 interactive role with treatment outcomes in breast cancer [31].

ALDH1A1 rs13959 elicited as the best model for higher grade 3–4 hematological toxicity. ALDH1A1 belongs to dehydrogenase family and plays a role in the detoxification of active cyclophosphamide metabolites. Our findings suggest that ALDH1A1 rs13959A>G genetic variations may affect the risk of severe hematological toxicity caused by myelosuppressive chemotherapy drugs. ALDH1A1 rs13959A>G is a synonymous polymorphism present in exon 3. F-SNP predicted its role in change in splicing regulation (Table 6). This variant allele could thus affect ALDH enzyme activity, thus resulting in an increased risk of chemotherapeutic drug toxicity. ALDH1A1 polymorphisms (rs3764435 C>A-rs63319C>A) A-A were also reported to be associated with grade 3–4 hematological toxicity in breast cancer [36].

CSC genetic variants and GBC survival Survival analysis results showed that ALCAM rs1157 GA+AA genotype is associated with unfavorable survival in GBC metastasis cases. The Cox-proportional hazard regression analysis demonstrated that GA+AA genotype had higher hazard ratio. Swedish population study also showed similar results; ALCAM rs1157 homozygous minor allele carriers had worse survival in breast cancer [37]. CD44, EpCAM, CD133, NANOG, SOX-2, LIN28A, OCT-4, and ALDH1A1 polymorphisms did not significantly associate with overall survival. But previously reported, CD44 rs187116G>A–rs187115T>C T-A haplotype and CD44rs13347 variant genotype associated with lowered survival rate in breast cancer patients [24, 38]. Survival decision tree terminal node determined ALCAM GA+AA as worse survival subgroup with higher hazard ratio.

These data collectively suggest important biological role of the CSC genetic variants in the susceptibility and prognosis of gallbladder cancer. However, functional studies are needed to elucidate the effects of the CSC genetic variants in various cancers. In future, cancer stem cells may prove to be important in defeating gallbladder cancer.

In conclusion, our findings suggest important role of CSC genetic variants in GBC susceptibility, prognosis, and survival outcomes.

Acknowledgments The authors would like to thank the Department of Biotechnology, Indian Council of Medical Research, and the Department of Science and Technology, Government of India, for financial support.

Conflicts of interest None

References

1. Rebecca Siegel MJM, Zou Z, Jemal A. Cancer statistics. *CA Cancer J Clin.* 2014;64:29.
2. Misra S, Chaturvedi A, Misra NC, Sharma ID. Carcinoma of the gallbladder. *Lancet Oncol.* 2003;4:167–76.
3. Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea—a paradigm shift. *Cancer Res.* 2006;66:1883–90. **discussion 1895–1886.**
4. Bourguignon LY, Singleton PA, Zhu H, Diedrich F. Hyaluronan-mediated CD44 interaction with RhoGEF and Rho kinase promotes Grb2-associated binder-1 phosphorylation and phosphatidylinositol 3-kinase signaling leading to cytokine (macrophage-colony stimulating factor) production and breast tumor progression. *J Biol Chem.* 2003;278:29420–34.
5. Bourguignon LY. CD44-mediated oncogenic signaling and cytoskeleton activation during mammary tumor progression. *J Mammary Gland Biol Neoplasia.* 2001;6:287–97.
6. Baumann M, Krause M. CD44: a cancer stem cell-related biomarker with predictive potential for radiotherapy. *Clin Cancer Res.* 2010;16:5091–3.
7. Shi C, Tian R, Wang M, Wang X, Jiang J, Zhang Z, et al. CD44+ CD133+ population exhibits cancer stem cell-like characteristics in human gallbladder carcinoma. *Cancer Biol Ther.* 2010;10:1182–90.
8. Sharma KL, Yadav A, Gupta A, Tulsayan S, Kumar V, Misra S, et al. Association of genetic variants of cancer stem cell gene CD44 haplotypes with gallbladder cancer susceptibility in North Indian population. *Tumour Biol.* 2014;35:2583–9.
9. Bowen MA, Patel DD, Li X, Modrell B, Malacko AR, Wang WC, et al. Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *J Exp Med.* 1995;181:2213–20.
10. Burandt E, Bari Noubat T, Lebeau A, Minner S, Burdelski C, Janicke F, et al. Loss of ALCAM expression is linked to adverse phenotype and poor prognosis in breast cancer: a TMA-based immunohistochemical study on 2,197 breast cancer patients. *Oncol Rep.* 2014;32:2628–34.
11. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. *Gastroenterology.* 2009;136:1012–24.
12. Imrich S, Hachmeister M, Gires O. EpCAM and its potential role in tumor-initiating cells. *Cell Adhes Migr.* 2012;6:30–8.
13. Pereira F, Rosenmann E, Nylen E, Kaufman M, Pinsky L, Wroegemann K. The 56 kDa androgen binding protein is an aldehyde dehydrogenase. *Biochem Biophys Res Commun.* 1991;175: 831–8.
14. Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem.* 2005;280:24731–7.
15. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science.* 2008;320:97–100.

16. Chou YE, Hsieh MJ, Chiou HL, Lee HL, Yang SF, Chen TY. CD44 gene polymorphisms on hepatocellular carcinoma susceptibility and clinicopathologic features. *Biomed Res Int.* 2014;2014:231474.
17. Gerger A, Zhang W, Yang D, Bohanes P, Ning Y, Winder T, et al. Common cancer stem cell gene variants predict colon cancer recurrence. *Clin Cancer Res.* 2011;17:6934–43.
18. Wang Q, Liu H, Xiong H, Liu Z, Wang LE, Qian J, et al. Polymorphisms at the microRNA binding-site of the stem cell marker gene CD133 modify susceptibility to and survival of gastric cancer. *Mol Carcinog.* 2015;54:449–58.
19. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215.
20. Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet.* 2007;80:1125–37.
21. Goodison S, Urquidi V, Tarin D. CD44 cell adhesion molecules. *Mol Pathol.* 1999;52:189–96.
22. Jiang Y, Qin Z, Hu Z, Guan X, Wang Y, He Y, et al. Genetic variation in a hsa-let-7 binding site in RAD52 is associated with breast cancer susceptibility. *Carcinogenesis.* 2013;34:689–93.
23. Benchimol S, Fuks A, Jothy S, Beauchemin N, Shirota K, Stanners CP. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell.* 1989;57:327–34.
24. 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. *Int J Cancer.* 2011;129:1096–104.
25. Chou YE, Hsieh MJ, Hsin CH, Chiang WL, Lai YC, Lee YH, et al. CD44 gene polymorphisms and environmental factors on oral cancer susceptibility in taiwan. *PLoS One.* 2014;9:e93692.
26. Roa I, Villaseca M, Araya J, Roa J, de Aretxabala X, Ibacache G, et al. CD44 (HCAM) expression in subserous gallbladder carcinoma. *Rev Med Chil.* 2001;129:727–34.
27. Yanagisawa N, Mikami T, Mitomi H, Saegusa M, Koike M, Okayasu I. CD44 variant overexpression in gallbladder carcinoma associated with tumor dedifferentiation. *Cancer.* 2001;91:408–16.
28. Jiang L, Zhang C, Li Y, Yu 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.
29. Katafigiotis S, Papamichos SI, Katopodi R, Papazisis K, Mylonaki E, Repana D, et al. A case-control study on the rs3130932 single nucleotide polymorphism in the OCT4B translation initiation codon in association with cancer state. *Eur J Cancer Prev.* 2011;20:248–51.
30. Qiu C, Ma Y, Wang J, Peng S, Huang Y. Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.* 2010;38:1240–8.
31. Tulsyan S, Agarwal G, Lal P, Mittal B. Significant association of combination of OCT4, NANOG, and SOX2 gene polymorphisms in susceptibility and response to treatment in North Indian breast cancer patients. *Cancer Chemother Pharmacol.* 2014;74:1065–78.
32. Szkandera J, Herzog S, Pichler M, Stiegelbauer V, Stotz M, Schaberl-Moser R, et al. Lgr5 rs17109924 is a predictive genetic biomarker for time to recurrence in patients with colon cancer treated with 5-fluorouracil-based adjuvant chemotherapy. *Pharmacogenomics J.* 2015.
33. Situ D, Long H, Lin P, Zhu Z, Wang J, Zhang X, et al. Expression and prognostic relevance of CD44v6 in stage I non-small cell lung carcinoma. *J Cancer Res Clin Oncol.* 2010;136:1213–9.
34. Stoll C, Baretton G, Soost F, Terpe HJ, Domide P, Lohrs U. Prognostic importance of the expression of CD44 splice variants in oral squamous cell carcinomas. *Oral Oncol.* 1999;35:484–9.
35. Yang Y, Fei F, Song Y, Li X, Zhang Z, Fei Z, et al. Polymorphisms of EpCAM gene and prognosis for non-small-cell lung cancer in Han Chinese. *Cancer Sci.* 2014;105:89–96.
36. Yao S, Sucheston LE, Zhao H, Barlow WE, Zirpoli G, Liu S, et al. Germline genetic variants in ABCB1, ABCC1 and ALDH1A1, and risk of hematological and gastrointestinal toxicities in a SWOG phase III trial S0221 for breast cancer. *Pharmacogenomics J.* 2014;14:241–7.
37. Varadi V, Bevier M, Grzybowska E, Johansson R, Enquist-Olsson K, Henriksson R, et al. Genetic variation in ALCAM and other chromosomal instability genes in breast cancer survival. *Breast Cancer Res Treat.* 2012;131:311–9.
38. Jiang L, Deng J, Zhu X, Zheng J, You Y, Li N, et al. Cd44 rs13347 c>t polymorphism predicts breast cancer risk and prognosis in Chinese populations. *Breast Cancer Res.* 2012;14:R105.