BMJ Open

Polymorphisms of the stem cell marker gene *CD133* are associated the clinical outcome in a cohort of Chinese non-small cell lung cancer patients

Qing-Feng Liu, Zhi-Fei Zhang, Guang-Jie Hou, Guang-Yu Yang, Yi He

To cite: Liu Q-F, Zhang Z-F, Hou G-J, *et al*. Polymorphisms of the stem cell marker gene *CD133* are associated the clinical outcome in a cohort of Chinese non-small cell lung cancer patients. *BMJ Open* 2017;**7**:e016913. doi:10.1136/bmjopen-2017-016913

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-016913).

Received 21 March 2017
Revised 18 July 2017
Accepted 24 July 2017

Department of Thoracic Surgery, Henan Provincial People's Hospital (Zhengzhou University People's Hospital), Zhengzhou, China

Correspondence to
Dr Yi He;
13903866310@163.com

ABSTRACT

Objectives To evaluate the prognostic relevance of four functional single nucleotide polymorphisms (SNPs) in *CD133* (*rs2240688*A>C, *rs10022537*T>A, *rs7686732*C>G, and *rs3130*C>T) on overall survival (OS) of non-small cell lung cancer (NSCLC) patients.

Design Retrospective cohort study.

Setting Department of General Surgery, in a general hospital, Henan Province, China.

Participants NSCLC patients aged ≥18 years, who were not receiving preoperative neoadjuvant therapies and had a blood sample available for genotyping, were eligible for inclusion. Those participants who were pregnant or breastfeeding, had a previous history of cancer, had other primary tumours, or who had had primary tumours of the skin and nasopharynx, were excluded from the study.

Outcome measures The primary endpoint was OS, which was calculated from the date of enrolment until the date of death or date of last follow-up.

Results There was a total of 1383 participants, with a median age of 63 years; 726 (52.5%) were male. Compared with the*rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (HR 1.27, 95% CI 1.12 to 1.45 for AC genotype; HR 2.32, 95% CI 1.91 to 2.80 for CC genotype). Higher hazard ratios for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy (HR 1.86, 95% CI 1.52 to 2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55 to 2.33).

Conclusions The study confirmed the significant association between the SNP *rs2240688 A>C* of CD133 and OS of NSCLC patients. Larger population-based studies in different ethnic groups are necessary to further validate the role and mechanisms of *CD133* in NSCLC.

Strengths and limitations of this study

- A large cohort (1383 participants) was studied to explore the association between functional single nucleotide polymorphisms (SNPs) in *CD133* and overall survival (OS) of lung cancer patients.
- Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (HR 1.27, 95% CI 1.12 to 1.45 for AC genotype; HR 2.32, 95% CI 1.91 to 2.80 for CC genotype).
- Higher hazard ratios for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy and radiotherapy for curative intent.
- All participants were recruited from a single hospital, which may limit the generalisability of the study's results.

INTRODUCTION

Lung cancer remains the most common cause of cancer-related deaths in China and worldwide.[1] The number of patients newly diagnosed with lung cancer was estimated at 1.8 million worldwide in 2012 and 500 000 in China. The number is expected to reach one million in China by 2025.[2] Approximately 83% of lung cancer patients have non-small cell lung cancer (NSCLC).[3] In addition, despite improvements in technologies and development of multiple treatments including surgery, radiotherapy, chemotherapy and utilisation of other biological agents, the prognosis of NSCLC is very poor due to recurrence and metastasis, with an overall 5-year survival rate <16%.[4–6] Hence, it is necessary to identify biomarkers for the prevention, early diagnosis, monitoring of progression and therapeutic effects of NSCLC.

It is widely conceived that cancer stem cells (CSCs) are able to be self-renew and to produce heterogeneous lineages of cancer cells.[7 8] CSCs have been hypothesised as the origin of cancer due to their potent tumour-driving capability on tumour initiation, growth, metastasis and relapse.[9 10] The cell surface marker *CD133*, which is located in cellular protrusions, is related to tumorigenesis and cancer progression.[11] The *CD133* antigen, also known as prominin-1, has been used extensively as a biomarker of CSCs among different types of tumours, including colon cancer,[12] liver cancer,[13] gastric cancer,[14]

brain tumour,[15] endometrial cancer[16] and ovarian cancer.[17]

The expression of *CD133* is significantly correlated with the development and prognosis of NSCLC.[18] As single nucleotide polymorphisms (SNPs) are considered to harbour information about genetic variation in functionality of the genome and susceptibility to tumour development,[19] we hypothesised that potential functional SNPs in *CD133* may influence the function of *CD133* and, consequentially, prognosis. A recent case–control study conducted by our team in a Chinese population showed that the *rs2240688* variant AC/CC genotypes were associated with a statistically increased risk of lung cancer.[19] However, another study found that *rs2240688C* variant genotypes were negatively associated with lung cancer and have a protective effect on overall survival (OS).[20] Considering the limited number of studies involving *CD133* genetic variants with NSCLC and their inconsistent results,[19 20] we investigated the associations between four functional SNPs in *CD133* and the prognosis of NSCLC in a Chinese population.

METHODS

Study populations

The sample in this retrospective cohort study included 1383 patients with histologically confirmed NSCLC, who were treated at the Department of General Surgery, Henan Provincial People's Hospital between January 2006 and December 2014. All participants with NSCLC in a case–control study which identified the relationship between polymorphisms of *CD133* and the risk of lung cancer were included in the previous retrospective cohort study.[19]

Eligibility for inclusion were the following criteria: (1) willing to participate in the study and sign an informed consent form; (2) aged ≥18 years with pathologically confirmed NSCLC; (3) had not received preoperative neoadjuvant therapies (including chemotherapy and radiotherapy); and (4) had a blood sample available for genotyping four selected SNPs in *CD133*. Those participants who were pregnant or breastfeeding, had previous cancer history, had other primary tumours, or had had primary tumours of the skin and nasopharynx, were excluded from this study. A personal identification number was assigned to every study subject at enrolment and specified on each case report form in order to maintain confidentiality.

The present study was performed in accordance with the Declaration of Helsinki, and the protocol and informed consent form have been reviewed and approved by the Institutional Review Boards of Henan Provincial People's Hospital. Written informed consent was provided by all the participants.

Initial screening, assessment and follow-up

In this retrospective cohort study, blood samples and medical data were obtained from the biobank for lung cancer patients at the Department of General Surgery, Henan Provincial People's Hospital. This biobank consists of an electronic database of demographic and clinicopathological data (including age, sex, smoking status, histological tumour type, tumour-node-metastasis stage, chemotherapy or radiotherapy treatment), as well as blood samples. At the time of inclusion, written informed consent was obtained, and each participant was interviewed to obtain a detailed medical history. Blood samples were separated by centrifugation within 2 hours of collection. Serum samples were stored in aliquots below −70°C until analysis.

Details on surgical treatment were recorded including dates, types of surgeries, and complications. From the electronic database, we identified patients treated with adjuvant chemotherapy and we classified patients as being treated with adjuvant chemotherapy if the patient received platinum-based chemotherapy within 3 months of surgery. We classified patients as having received radiotherapy if they received external beam radiation, radioactive implants, radioisotopes, brachytherapy or other types of radiotherapy within 6 months of cancer diagnosis. We classified patients as having received curative intent radiotherapy if the patients with early disease (stages I and II) were treated with surgical resection in addition to radiotherapy. We classified patients as having received palliative radiotherapy if the patients with advanced or distant disease (stages III and IV) were treated with radiotherapy or in combination with other treatments for supportive care. Family history of cancer was defined as any types of cancer present in first degree relatives of the participants (parents, siblings and children).

OS was evaluated for all the patients with regular follow-up at 3-month intervals for the first 2 years after surgery, at 6-month intervals for years 3–5, and yearly thereafter according to the hospital guidelines. The patients who failed to attend follow-up visits were telephoned or their family members were contacted. Follow-up of patients for the present study was performed until July 2016.

SNP selection and genotyping

Potential SNPs should be common (≥5% minor allele frequency) in Chinese populations. Candidate *CD133* gene SNPs with potentially functional significance (that is, located in the promoter, the transcription factor-binding site, exon and 3'-untranslated region (UTR), or the coding regions with amino acid changes) were selected based on NCBI dbSNPs (http://www.ncbi.nlm.nih.gov) and SNPinfo Web Server (http://www.snpinfo.niehs.nih.gov/snpfunc.htm). As a result, four *CD133* candidate SNPs (*rs2240688* A>C, *rs10022537* T>A, *rs7686732* C>G, *rs3130* C>T) were identified and chosen in our model analysis. Three SNPs (*rs2240688*, *rs7686732*, *rs3130*) were located in the 3'-UTR of CD133, and *rs10022537* was located within the intron of the *CD133* gene.

Table 1 Demographic and clinicopathological characteristics of non-small cell lung cancer patients recruited from Henan Provincial People's Hospital between January 2006 and December 2014

Characteristics	Lung adenocarcinoma (n=793)	Lung squamous cell cancer (n=331)	Others (n=259)	Total (n=1383)
Age (years)				
<65	476 (60.0)	151 (45.6)	154 (59.5)	781 (56.5)
≥65	317 (40.0)	180 (54.4)	105 (40.5)	602 (43.5)
Sex				
Male	372 (46.9)	206 (62.2)	148 (57.1)	726 (52.5)
Female	421 (53.1)	125 (37.8)	111 (42.9)	657 (47.5)
Smoking status				
Non-smoker	182 (23.0)	24 (7.3)	33 (12.7)	239 (17.3)
Former smoker	322 (40.6)	165 (49.8)	100 (38.6)	587 (42.4)
Current smoker	289 (36.4)	142 (42.9)	126 (48.6)	557 (40.3)
Pack-years				
≤25	158 (25.9)	23 (7.5)	28 (12.4)	209 (15.1)
26–50	233 (38.1)	121 (39.4)	96 (42.5)	450 (32.5)
>50	220 (36.0)	163 (53.1)	102 (45.1)	485 (35.1)
Stage				
I-II	234 (29.5)	110 (33.2)	50 (19.3)	394 (28.5)
III	271 (34.2)	155 (46.8)	102 (39.4)	528 (38.2)
IV	288 (36.3)	66 (19.9)	107 (41.3)	461 (33.3)
Family history				
Yes	121 (15.3)	45 (13.6)	39 (15.1)	205 (14.8)
No	672 (84.7)	286 (86.4)	220 (84.9)	1178 (85.2)
Surgery				
No	274 (34.5)	101 (30.5)	166 (64.1)	541 (39.1)
Lobectomy	271 (34.2)	67 (20.2)	79 (30.5)	417 (30.2)
Segmentectomy	130 (16.4)	62 (18.7)	3 (1.2)	195 (14.1)
Wedge resection	118 (14.9)	101 (30.5)	11 (4.3)	230 (16.6)
History of radiotherapy				
No	263 (33.2)	100 (30.2)	143 (55.2)	506 (36.6)
Palliative therapy	116 (14.6)	112 (33.8)	27 (10.4)	255 (18.4)
Curative intent	414 (52.2)	119 (36.0)	89 (34.4)	622 (45)
Adjuvant chemotherapy				
No	444 (56.0)	177 (53.5)	155 (59.9)	776 (56.1)
Yes	349 (44.0)	154 (46.5)	104 (40.1)	607 (43.9)

Genomic DNA was extracted from the buffy coat fraction of each blood sample with a DNA blood Mini Kit (Qiagen Inc, Valencia, California, USA) according to the manufacturer's instructions. The genotyping methods of the four *CD133* SNPs are described in detail elsewhere.[19]

Statistical analysis

We expect 3-year survival rates of 35% in patients with variant genotypes of *rs2240688* (AC/CC) and 27% in patients with *rs2240688* AA genotype. Based on a difference of 15% between groups on the primary outcome, assuming a 10% drop-out rate, a total of 1234 participants (at 1:1 ratio, 617 subjects in each group) are required to provide 80% power, with the use of a two-sided significance level of 0.05.

All statistical tests were performed using SAS 9.3 software (Cary, North Carolina, USA). Descriptive analysis results were presented as median and interquartile range (IQR) for continuous variables and frequencies (percentage) for categorical variables. Distributions of

Table 2 Associations between *CD133* genotypes and overall survival among non-small cell lung cancer patients recruited from Henan Provincial People's Hospital between January 2006 and December 2014

Genotypes	No. of patients N (%)	No. of deaths N (%)	MST (months)	HR (95% CI)	Adjusted HR (95% CI)*
rs2240688					
AA	652	463 (71.0)	20.3	1.0	1.0
AC	555	434 (78.2)	15.6	1.29 (1.13 to 1.47)	1.27 (1.12 to 1.45)
CC	172	143 (83.1)	8.2	2.22 (1.84 to 2.68)	2.32 (1.91 to 2.80)
Recessive					
AA/AC	1207	897 (74.3)	18.1	1.0	1.0
CC	172	143 (83.1)	8.2	1.98 (1.66 to 2.36)	2.07 (1.73 to 2.48)
Dominant					
AA	652	463 (71.0)	20.3	1.0	1.0
AC/CC	727	577 (79.4)	13.0	1.43 (1.27 to 1.62)	1.43 (1.26 to 1.61)
rs10022537					
TT	913	689 (75.5)	17.2	1.0	1.0
TA	413	311 (75.3)	14.5	1.06 (0.93 to 1.22)	1.10 (0.96 to 1.27)
AA	39	30 (76.9)	15.2	1.14 (0.79 to 1.64)	1.00 (0.69 to 1.44)
Dominant					
TT/TA	1326	1000 (75.4)	16.8	1.0	1.0
AA	39	30 (76.9)	15.2	1.12 (0.78 to 1.61)	0.96 (0.67 to 1.39)
Recessive					
TT	913	689 (75.5)	17.2	1.0	1.0
TA/AA	452	341 (75.4)	14.5	1.07 (0.94 to 1.22)	1.13 (0.99 to 1.29)
rs7686732†					
CC	398	286 (71.9)	17.1	1.0	1.0
CG	88	63 (71.6)	15.9	1.06 (0.80 to 1.39)	1.12 (0.85 to 1.49)
GG	5	4 (80.0)	20.1	1.33 (0.49 to 3.56)	1.26 (0.47 to 3.41)
Dominant					
CC/CG	486	349 (71.8)	17.0	1.0	1.0
GG	5	4 (80.0)	20.1	1.31 (0.49 to 3.52)	1.24 (0.46 to 3.34)
Recessive					
CC	398	286 (71.9)	17.1	1.0	1.0
CG/GG	93	67 (72.0)	15.9	1.07 (0.82 to 1.39)	1.13 (0.86 to 1.49)
rs3130†					
CC	134	92 (68.7)	18.0	1.0	1.0
CT	269	201 (74.7)	17.0	1.13 (0.88 to 1.44)	1.14 (0.89 to 1.47)
TT	92	64 (69.6)	14.4	0.98 (0.72 to 1.36)	0.97 (0.70 to 1.34)
Dominant					
CC/CT	403	293 (72.7)	17.3	1.0	1.0
TT	92	64 (69.6)	14.4	0.91 (0.69 to 1.19)	0.89 (0.68 to 1.18)
Recessive					
CC	134	92 (68.7)	18.0	1.0	1.0
CT/TT	361	265 (73.4)	16.8	1.09 (0.86 to 1.38)	1.09 (0.86 to 1.39)

*Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.
†Genotyping of the two SNPs *rs7686732 and rs3130* was only carried out for a portion of the participants.
MST, median survival time; SNPs, single nucleotide polymorphisms.

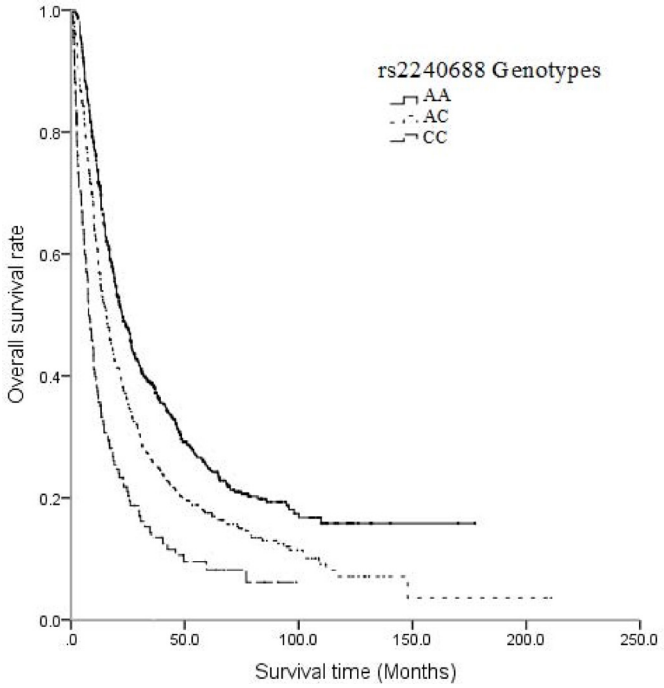

Figure 1 Kaplan-Meier curves for overall survival among non-small cell lung cancer patients stratified by the *rs2240688* genotypes.

categorical variables, including demographic variables, prognosis factors and clinicopathological characteristics, were compared using the χ^2 test/Fisher's exact test as appropriate. The primary endpoint was OS, which was calculated from the date of enrolment until the date of death or date of last follow-up. Survival curves of OS were estimated by the Kaplan-Meier method and compared by the log-rank test. The associations of *CD133* SNPs with OS were estimated by calculating hazard ratios (HR) and corresponding 95% confidence intervals (CI) from both univariate and multivariate Cox proportional hazards regression models, followed by stratification analysis by age, sex, smoking status, histopathology type, stage, family history, and application of chemotherapy and radiotherapy. In addition, the associations of *CD133* SNPs with OS were analysed under specific genetic models: genotypic, recessive and dominant models, but only the dominant model was used in the stratification analysis of *CD133 rs2240688* polymorphism as *rs2240688* (AC/CC) are variant genotypes. All these analyses were performed with or without adjustment for demographic variables and selected clinicopathological characteristics. All tests were two-sided and a value of p<0.05 was considered to be statistically significant for all analyses.

RESULTS

Baseline characteristics of the study population

There were a total of 1383 participants with histologically confirmed NSCLC included in this retrospective cohort, including 793 (57.3%) lung adenocarcinomas, 331 (23.9%) lung squamous cell cancers, and 259 (18.7%) other types of NSCLC. table 1 summarises the baseline characteristics of the study population by histopathology type. There were 726 (52.5%) males and 657 (47.5%) females, with ages ranging from 28 to 92 years (median 63 years; IQR 54–70 years). There were 394 (28.5%), 528 (38.2%) and 461 (33.3%) participants with stage I-II, III and IV NSCLC, respectively. There were a total of 842 (60.9%) participants who received surgical treatment, including 417 (30.2%) with lobectomy, 195 (14.1%) with segmentectomy and 230 (16.6%) with wedge resection. There were 607 (43.9%) participants who underwent adjuvant chemotherapy and 877 (63.4%) who underwent radiotherapy, including 622 (45.0%) for curative intent and 255 (18.4%) for palliative therapy.

Association of *CD133* genotypes with OS

The enrolled NSCLC patients who returned for at least one follow-up visit had been followed for a median of 14.4 months (IQR 24.4 months). At the end of the study, 339 (24.5%) patients were alive and 1044 (75.5%) patients had died of any cause during follow-up.

The genotype distributions of the selected four SNPs in *CD133* and their associations with OS of NSCLC patients are shown in table 2. In all patients, variant genotypes of *rs2240688* (AC/CC) were statistically significantly associated with OS (log-rank p<0.001 under a recessive model). Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were associated with a statistically poorer OS of NSCLC (HR 1.29, 95% CI 1.13 to 1.47 for AC genotype; HR 2.22, 95% CI 1.84 to 2.68 for CC genotype). As shown in multivariate survival analysis using Cox proportional hazards regression, *rs2240688* variant genotypes remained significantly associated with OS (HR 1.27, 95% CI 1.12 to 1.45 for AC genotype; HR 2.32, 95% CI 1.91 to 2.80 for CC genotype) after adjustment for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy. However, the association between genotype distribution of the other three SNPs (*rs10022537* T>A, *rs7686732* C>G, *rs3130* C>T) and OS of NSCLC patients was not observed.

Figure 1 shows the Kaplan-Meier curves for the OS among all NSCLC patients stratified by *rs2240688* genotypes. The median OS was 20.3 months (95% CI 18.7 to 22.7) for participants with *rs2240688* AA genotype, 15.6 months (95% CI 13.0 to 17.1) with *rs2240688* AC genotype, and 8.2 months (95% CI 7.1 to 9.9) with *rs2240688* CC genotype.

CD133 genotypes and OS of NSCLC by clinicopathological characteristics

Stratified analysis was further performed for *rs2240688* A>C by age, sex, smoking status, histopathology type, stage, family history of cancer, and application of chemotherapy and radiotherapy (table 3). Compared with the *rs2240688* AA genotype, the association between OS and *rs2240688* AC/CC variant genotypes also remained statistically significant in the subgroup of all ages, all sexes, former smoker, current smoker, lung adenocarcinoma,

Table 3 Stratified analysis for associations between *CD133 rs2240688* polymorphism (dominant for the C allele) and overal survival among non-small cell lung cancer patients recruited from Henan Provincial People's Hospital between January 2006 and December 2014

Variables	*rs2240688* (death/patients)		MST (Months)	HR (95% CI)	Adjusted HR (95% CI)*
	AA	AC/CC			
Age (years)					
<65	254/367	324/411	20.1 vs 14.1	1.36 (1.16 to 1.61)	1.44 (1.22 to 1.70)
≥65	209/285	253/316	21.2 vs 11.9	1.54 (1.28 to 1.85)	1.46 (1.21 to 1.76)
Sex					
Male	249/334	326/390	18.1 vs 11.8	1.53 (1.30 to 1.80)	1.45 (1.23 to 1.71)
Female	214/318	251/337	23.0 vs 16.8	1.32 (1.10 to 1.59)	1.45 (1.21 to 1.74)
Smoking status					
Non-smoker	82/116	89/120	18.3 vs 16.8	1.19 (0.88 to 1.60)	1.28 (0.94 to 1.74)
Former smoker	188/278	241/309	26.0 vs 13.4	1.54 (1.27 to 1.86)	1.52 (1.25 to 1.84)
Current smoker	193/258	247/298	19.0 vs 11.6	1.43 (1.18 to 1.72)	1.44 (1.19 to 1.74)
Histopathology type					
Lung adenocarcinoma	253/377	313/414	23.0 vs 16.0	1.40 (1.19 to 1.66)	1.44 (1.22 to 1.70)
Lung squamous cell cancer	118/161	132/170	21.0 vs 12.0	1.40 (1.09 to 1.79)	1.32 (1.03 to 1.71)
Others	92/114	132/143	14.6 vs 9.0	1.54 (1.18 to 2.01)	1.51 (1.14 to 2.00)
Stage					
I-II	78/169	140/224	53.9 vs 18.0	2.09 (1.58 to 2.77)	2.28 (1.72 to 3.03)
III	217/272	219/256	19.0 vs 15.5	1.15 (0.95 to 1.39)	1.17 (0.97 to 1.42)
IV	168/211	218/247	13.4 vs 11.2	1.42 (1.16 to 1.73)	1.43 (1.16 to 1.75)
Family history					
Yes	71/102	80/103	22.2 vs 11.6	1.52 (1.10 to 2.09)	1.56 (1.12 to 2.17)
No	392/550	497/624	20.0 vs 13.2	1.42 (1.24 to 1.62)	1.41 (1.23 to 1.61)
Surgery					
No	191/232	274/306	12.3 vs 9.6	1.36 (1.13 to 1.64)	1.38 (1.15 to 1.66)
Lobectomy	157/210	179/206	26.0 vs 17.0	1.38 (1.11 to 1.71)	1.40 (1.12 to 1.74)
Segmentectomy	61/95	61/100	25.0 vs 21.6	1.14 (0.80 to 1.62)	1.27 (0.87 to 1.84)
Wedge resection	54/115	63/115	48.9 vs 23.4	1.72 (1.19 to 2.48)	1.75 (1.20 to 2.54)
History of radiotherapy					
No	196/237	230/266	15.0 vs 12.0	1.17 (0.97 to 1.42)	1.24 (1.02 to 1.51)
Palliative therapy	103/125	113/130	19.0 vs 16.8	1.19 (0.91 to 1.56)	1.16 (0.88 to 1.53)
Curative intent	164/290	234/331	31.2 vs 12.8	1.90 (1.55 to 2.32)	1.90 (1.55 to 2.33)
Adjuvant chemotherapy					
No	294/384	326/391	19.3 vs 16.2	1.23 (1.05 to 1.44)	1.22 (1.04 to 1.43)
Yes	169/268	251/336	22.7 vs 9.9	1.78 (1.46 to 2.16)	1.86 (1.52 to 2.26)

*Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.
MST, median survival time.

lung squamous cell cancer, other types of NSCLC, stage I-II, stage IV, with or without family history of cancer, without surgical treatment, with lobectomy, with wedge resection, with radiotherapy for curative intent, without radiotherapy, and with or without adjuvant chemotherapy. Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy (HR 1.86, 95% CI 1.52 to 2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55 to 2.33), compared with patients without adjuvant chemotherapy (HR 1.22, 95% CI 1.04 to 1.43) and those without radiotherapy (HR 1.24, 95% CI 1.02 to 1.51). Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were also observed in patients with stage I-II and wedge resection surgery (table 3).

DISCUSSION

It is generally accepted that SNPs represent genetic variation in functionality of the genome and they are potential functional biomarkers for cancer aetiology. CSCs are responsible for tumour initiation, growth, migration, aggressiveness, metastasis, drug resistance and pluripotency.[8] In this study, information was collected to determine

the role of CSCs in the clinical outcomes of NSCLC. We focused on the *CD133* gene that has been used to isolate CSCs. Four potential functional SNPs in the *CD133* gene locus were selected from SNP websites and peer-reviewed literature by using the candidate gene approach.

The prognostic and clinicopathological values of *CD133* protein and mRNA expression have been indicated in other studies.[3 21] In this hospital-based cohort study, we found that the variant genotypes (AC/CC) of *rs2240688* A>C in the miRNA binding site of the stem cell marker gene *CD133* was associated with a significantly poorer prognosis for NSCLC patients. The association remained statistically significant (HR 1.27, 95% CI 1.12 to 1.45 for AC genotype; HR 2.32, 95% CI 1.91 to 2.80 for CC genotype) after adjustment for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy. Additionally in the stratified analysis, the poorer prognosis associated with *rs2240688* A>C variant genotypes did remain statistically significant in most subgroups. It was validated that *rs2240688* A-to-C transition gained a new binding site of the microRNA has-miR-135a/b, which may play a pivotal role in modulating the effect of the SNP on *CD133* expression.[20] The *rs2240688* A>C variant genotypes are located in the 3'-UTR of *CD133*. SNPs in the 3'-UTR have been shown to have functional effects on the control of mRNA stability and efficiency through the regulation of miRNA, including miR-34a, −101,–128, −137 and −138.[22] It is inferred that SNPs in a target-binding site could alter the miRNA–mRNA interaction and thus affect the expression of miRNA targets. Considering the tumour-driving capability of CSCs on tumour growth and metastasis, the present study suggests that *CD133* might modify their metastasis competence of NSCLC by the miRNA binding site polymorphisms, which could be a putative target for improved therapies for treatment. Our subgroup analysis results showed that *rs2240688* A>C variant genotypes had more effects on the prognosis of NSCLC among patients receiving adjuvant chemotherapy or radiotherapy. This may be due to its association with resistance to chemotherapy and radiotherapy.[23] Higher HRs for associations between CD133 rs2240688 polymorphism and OS were also observed in patients with stage I-II and wedge resection surgery. It may due to the effect of the *rs2240688* A>C variant genotypes on tumour growth and metastasis which would result in greater impact among early stage patients.

However, the prognostic value of CSCs marker *CD133* in NSCLC remains controversial. Another study in China found that the *rs2240688* variant genotypes were associated with a favourable survival. Several studies found no significant association between the expression level of *CD133* and OS of NSCLC patients.[24 25] The inconsistent results may be explained in part by the different ethnic population, in addition to the different sampling methods used to select the populations under study. A meta-analysis showed that NSCLC patients with higher *CD133* expression had poor OS only in Asian patients, but not in Caucasian patients.[3] Therefore, high quality and interethnic studies with large samples should be undertaken to confirm the prognostic and clinical value of *CD133*.

This study had several limitations that should be taken into account. First, selection bias cannot be excluded even though inclusion/exclusion criteria were determined to minimise the bias. Potential confounding factors, such as clinicopathological characteristics, may be associated with SNPs in the *CD133* gene and also exert an effect on the overall mortality in our cohort of NSCLC patients. However, the independent association between SNPs in the *CD133* gene and OS of NSCLC patients was determined by using multivariate Cox proportional hazards regression models. Moreover, the HRs were largely very similar in all subgroups and similar to the overall HR, which implied no confounding by these factors. Second, in the hospital-based cohort study, all participants were recruited from a single hospital in Henan Province. Therefore, our study setting may limit the generalisability of our results. Finally, our study made many statistical comparisons, which might increase type I error.

CONCLUSIONS

This study confirmed a significant association between the SNP *rs2240688* A>C of *CD133* and OS for NSCLC patients. Larger population-based studies in different ethnic groups are necessary to further validate the role and mechanisms of *CD133* in NSCLC.

Contributors Q-FL, Z-FZ and YH conceived and designed the experiments. G-JH and G-YY performed the experiments. Q-FL and G-YY analysed the data. Q-FL, Z-FZ and YH contributed to the writing of the manuscript. All authors contributed to and have approved the final manuscript.

Funding This study was funded by Henan Provincial Science and Technology Department Grant (201201023).

Competing interests None declared.

Patient consent Obtained.

Ethics approval Institutional Review Boards of Henan Provincial People's Hospital.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Jemal A, Center MM, DeSantis C, *et al*. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;19:1893–907.
2. Liang B, Shao Y, Long F, *et al*. Predicting diagnostic gene biomarkers for non-small-cell lung cancer. *Biomed Res Int* 2016;2016:1–8.
3. Chen E, Zeng Z, Bai B, *et al*. The prognostic value of CSCs biomarker CD133 in NSCLC: a meta-analysis. *Oncotarget* 2016;7:56526–39.

4. Siegel R, DeSantis C, Virgo K, *et al*. Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin* 2012;62:220–41.
5. Hachey KJ, Colson YL. Current innovations in sentinel lymph node mapping for the staging and treatment of resectable lung cancer. *Semin Thorac Cardiovasc Surg* 2014;26:201–9.
6. Matsuda A, Yamaoka K, Tango T. Quality of life in advanced non-small cell lung cancer patients receiving palliative chemotherapy: a meta-analysis of randomized controlled trials. *Exp Ther Med* 2012;3:134–40.
7. Kim CF, Dirks PB. Cancer and stem cell biology: how tightly intertwined? *Cell Stem Cell* 2008;3:147–50.
8. Sales KM, Winslet MC, Seifalian AM. Stem cells and cancer: an overview. *Stem Cell Rev* 2007;3:249–55.
9. Clarke MF. A self-renewal assay for cancer stem cells. *Cancer Chemother Pharmacol* 2005;56(Suppl 1):64–8.
10. Al-Hajj M, Wicha MS, Benito-Hernandez A, *et al*. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 2003;100:3983–8.
11. Li Z. CD133: a stem cell biomarker and beyond. *Exp Hematol Oncol* 2013;2:17.
12. Artells R, Moreno I, Díaz T, *et al*. Tumour CD133 mRNA expression and clinical outcome in surgically resected colorectal cancer patients. *Eur J Cancer* 2010;46:642–9.
13. Sasaki A, Kamiyama T, Yokoo H, *et al*. Cytoplasmic expression of CD133 is an important risk factor for overall survival in hepatocellular carcinoma. *Oncol Rep* 2010;24:537–46.
14. Wang Q, Liu H, Xiong H, *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.
15. Metellus P, Nanni-Metellus I, Delfino C, *et al*. Prognostic impact of CD133 mRNA expression in 48 glioblastoma patients treated with concomitant radiochemotherapy: a prospective patient cohort at a single institution. *Ann Surg Oncol* 2011;18:2937–45.
16. Nakamura M, Kyo S, Zhang B, *et al*. Prognostic impact of CD133 expression as a tumor-initiating cell marker in endometrial cancer. *Hum Pathol* 2010;41:1516–29.
17. Silva IA, Bai S, McLean K, *et al*. Aldehyde dehydrogenase in combination with CD133 defines angiogenic ovarian cancer stem cells that portend poor patient survival. *Cancer Res* 2011;71:3991–4001.
18. Woo T, Okudela K, Mitsui H, *et al*. Prognostic value of CD133 expression in stage I lung adenocarcinomas. *Int J Clin Exp Pathol* 2010;4:32–42.
19. Liu QF, Zhang ZF, Hou GJ, *et al*. Polymorphisms of the stem cell marker gene CD133 and the risk of lung cancer in Chinese population. *Lung* 2016;194:393–400.
20. 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;34:2292–9.
21. Su C, Xu Y, Li X, *et al*. Predictive and prognostic effect of CD133 and cancer-testis antigens in stage Ib-IIIA non-small cell lung cancer. *Int J Clin Exp Pathol* 2015;8:5509–18.
22. Wang X, Hu JF, Tan Y, *et al*. Cancer stem cell marker Musashi-1 rs2522137 genotype is associated with an increased risk of lung cancer. *PLoS One* 2014;9:e95915.
23. Keysar SB, Jimeno A. More than markers: biological significance of cancer stem cell-defining molecules. *Mol Cancer Ther* 2010;9:2450–7.
24. Gottschling S, Jensen K, Herth FJ, *et al*. Lack of prognostic significance of neuroendocrine differentiation and stem cell antigen co-expression in resected early-stage non-small cell lung cancer. *Anticancer Res* 2013;33:981–90.
25. Shien K, Toyooka S, Ichimura K, *et al*. Prognostic impact of cancer stem cell-related markers in non-small cell lung cancer patients treated with induction chemoradiotherapy. *Lung Cancer* 2012;77:162–7.