BMJ Open

Polymorphisms of the stem cell marker gene CD133 are associated the clinical outcome of non-small cell lung cancer in Chinese population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016913
Article Type:	Research
Date Submitted by the Author:	21-Mar-2017
Complete List of Authors:	Liu, Qing-Feng Zhang, Zhi-Fei Hou, Guang-Jie Yang, Guang-Yu He, Yi; Henan Provincial People's Hospital, Department of thoracic surgery
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Genetics and genomics
Keywords:	Non-small cell lung cancer, CD133 polymorphism, Overall survival, China

Polymorphisms of the stem cell marker gene CD133 are associated the clinical outcome of non-small cell lung cancer in Chinese population

Qing-Feng Liu, M.S [1]; Zhi-Fei Zhang, M.S [1]; Guang-Jie Hou, M.D[1]; Guang-Yu Yang, M.D[1]; Ph.D; Yi He, M.S[1];

[1] Department of thoracic surgery, Henan Provincial People's Hospital (Zhengzhou University People's Hospital), 7 Weiwu Street, Zhengzhou city, Henan Province, 450003, PR China

Corresponding Author:

Yi He, Ph.D
Department of thoracic surgery
Henan Provincial People's Hospital
Zhengzhou University People's Hospital
7 Weiwu Street, Zhengzhou city
Henan Province, 450003, PR China
Phone: +86(371)86567371
Fax: +86(371)86567371
Email: hbwang2005@126.com

Running title: Polymorphisms of gene CD133 and the survival of NSCLC

Word Count:

Abstract: 254

Text: 2,441

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, without receiving preoperative neoadjuvant therapies, had available blood sample for genotyping, were eligible for inclusion. Those participants who were pregnant or breastfeeding, had previous cancer history, had second primary tumors, or who had had primary tumors of the skin and nasopharynx, were excluded from this study. Among 1,383 participants, median age was 63 years; 726 (52.5%) males.

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

Results: Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (hazard ration (HR) 1.27, 95% confidence interval (CI) 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype). Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy (HR 1.86, 95% CI 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33).

Conclusions: For the NSCLC patients with *rs2240688* A>C variant genotypes who receive chemotherapy and/or radiotherapy, the therapeutic effect may be enhanced by controlling the expression of *CD133* level. Larger population-based studies in

different ethnic groups are necessary to further validate the role and mechanisms of *CD133* in NSCLC.

Keywords: Non-small cell lung cancer; CD133 polymorphism; Overall survival; China

Article summary

Strengths and limitations of this study

1. A large cohort (1,383 participants) conducted to explore the association between functional SNPs in *CD133 and* OS of lung cancer patients.
2. Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype).
3. Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy and radiotherapy for curative intent.

Introduction

Lung cancer remains the most common cause of cancer-related deaths in China and worldwide [1]. The number of patients who were newly diagnosed with lung cancer was estimated at 1.8 million worldwide in 2012 and 500 thousand 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 other biological agent therapies, the prognosis of NSCLC is very poor due to recurrence and metastasis, with an overall 5-year survival rate less than 16% [4-6]. Hence, it is necessary to identify biomarkers for prevention, early diagnosis, monitoring 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 hypothesized as the origin of cancer due to their potent tumor-driving capability on tumor 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 tumors, including colon cancer [12], liver cancer [13], gastric cancer [14], brain tumor [15], endometrial cancer[16] and ovarian cancer [17].

The expression of *CD133* is significantly correlated with development and prognosis of NSCLC [18]. As single nucleotide polymorphisms (SNPs) are considered to harbor information about genetic variation in functionality of the genome and susceptibility to tumor development[19], we hypothesized 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 included 1,383 consecutive 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 in this retrospective cohort study.

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

The present study was performed in accordance with the Declaration of Helsinki, 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

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 tumor type, tumor-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 two hours of collection. Serum samples were stored in aliquots below −70 °C until analysis.

Details on surgical treatment are 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 three months of surgery. We classified patients as having received radiotherapy if they received external beam radiation, radioactive implants, radioisotopes, brachytherapy or other radiations within six 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 immediate blood relatives of the participants.

OS was evaluated for all the patients with regular follow-up at 3-months interval for the first 2 years after surgery, at 6-months 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.

SNPs 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, and *rs3130* C>T) were identified and chosen in our model analysis.

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

All statistical tests were performed using SAS™ 9.3 software (Cary, NC, USA). Descriptive analysis results were presented as median and inter-quartile range (IQR)

for continuous variables and frequencies (percentage) for categorical variables. Distributions of categorical variables, including demographic variables, prognosis factors and clinicopathological characteristics, were compared using the χ^2 test. The primary endpoint was OS, which was calculated from the date of enrollment until the date of death or date of last follow-up. Survival curves of OS were estimated by Kaplan-Meier method and compared by 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, application of chemotherapy and radiotherapy. All these analyses were performed with or without adjustment for demographic variables and selected clinicopathological characteristics. All tests were two-sided and a *P*-value <0.05 was considered to be statistically significant for all analyses.

Results

Baseline characteristics of the study population

There were a total of 1,383 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 summarizes 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%) 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,

there were 339 (24.5%) patients alive and 1,044 (75.5%) patients who 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*<.001 under an additive model; log-rank *P*<.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 1.13-1.47 for AC genotype; HR 2.22, 95% CI 1.84-2.68 for CC genotype). As shown in multivariate survival analysis using Cox proportional hazards regression model, *rs2240688* variant genotypes remained significantly associated with OS (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 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 and *rs3130* C>T) and OS of NSCLC patients was not observed in the study.

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, 18.7-22.7) for participants with *rs2240688* AA genotype, 15.6 months (95% CI, 13.0-17.1) with *rs2240688* AC genotype and 8.2 months (95% CI, 7.1-9.9) with *rs2240688* CC genotype.

CD133 Genotypes and OS of NSCLC by clinicopathological characteristics

The stratified analysis was further performed for *rs2240688* A>C by age, sex, smoking status, histopathology type, stage, family history of cancer, application for 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 <65 years, ≥65 years, male, female, former smoker, current smoker, lung adenocarcinoma, 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 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33), compared with patients without adjuvant chemotherapy (HR 1.22, 95% CI 1.04-1.43) and those without radiotherapy (HR 1.24, 95% CI 1.02-1.51).

Discussion

It is generally accepted that SNPs represent genetic variation in functionality of the genome and they are potential functional biomarkers for cancer etiology. CSCs are responsible for tumor initiation, growth, migration, aggressiveness, metastasis, drug resistance and pluripotency[8]. In this study, information was collected to determine the role of CSCs in clinical outcomes of NSCLC. We focused on *CD133* genes that have been used to isolate CSCs. Four potential functional SNPs in *CD133* gene loci 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 significantly poorer prognosis of NSCLC patients. The association remained statistically significant (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 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 locates in the 3'-UTRs of *CD133*. SNPs in the 3'-UTRs have been shown to have functional effects on control of mRNA stability and efficiency through the regulation of miRNA[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 tumor-driving capability of CSCs on tumor 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

prognosis of NSCLC among patients receiving adjuvant chemotherapy or radiotherapy. This may be due to its association with resistance to chemotherapy and radiotherapy [23]. For the NSCLC patients with *rs2240688* A>C variant genotypes who receive chemotherapy and/or radiotherapy, the therapeutic effect may be enhanced by controlling the expression of *CD133*.

However, the prognostic value of CSCs marker *CD133* in NSCLC remains controversial. Another study in China found that *rs2240688* variant genotypes were associated with a favorable 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 because of the different ethnic population, in addition to different sampling methods 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-scale sample should be done to confirm the prognostic and clinical value of *CD133*.

There are some limitations that should be taken into account in this study. First, selection bias cannot be excluded even though inclusion/exclusion criteria were determined to minimize the bias. Potential confounding factors, such as clinicopathological characteristics, may be associated with SNPs in *CD133*gene and also exert an effect on the overall mortality in our cohort of NSCLC patients. However, the independent association between SNPs in *CD133*gene and OS of NSCLC patients was determined by using multivariate Cox proportional hazards regression models. Secondly, in the hospital-based cohort study, all participants were recruited from a single hospital in Henan Province. Therefore, our study setting may limit the generalizability of our results. Finally, our study had made many comparisons in statistical analyses, which might increase type I error.

Conclusions

The study confirmed the significant association between the SNP *rs2240688* A>C of *CD133* and OS of NSCLC patients. For the NSCLC patients with *rs2240688* A>C variant genotypes who receive chemotherapy and/or radiotherapy, the therapeutic effect may be enhanced by controlling the expression of the *CD133* level. Larger population-based studies in different ethnic groups are necessary to further validate the role and mechanisms of *CD133* in NSCLC.

Figure Legends

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the *rs2240688* genotypes

OS: overall survival.
NSCLC: non-small cell lung cancer.

Author Contributions

Authors Qing-Feng Liu, Zhi-Fei Zhang and Yi He conceived and designed the experiments. Guang-Jie Hou and Guang-Yu Yang performed the experiments. Qing-Feng Liu and Guang-Yu Yang analyzed the data. Qing-Feng Liu, Zhi-Fei Zhang, and Yi He contributed to the writing of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest statement

All the authors declare no conflict of interest.

Funding

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

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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;**19**(8):1893-907 doi: 10.1158/1055-9965.EPI-10-0437[published Online First: Epub Date]|.
2. Liang B, Shao Y, Long F, et al. Predicting Diagnostic Gene Biomarkers for Non-Small-Cell Lung Cancer. BioMed research international 2016;**2016**:3952494 doi: 10.1155/2016/3952494[published Online First: Epub Date]|.
3. Chen E, Zeng Z, Bai B, et al. The prognostic value of CSCs biomarker CD133 in NSCLC: a meta-analysis. Oncotarget 2016 doi: 10.18632/oncotarget.10964[published Online First: Epub Date]|.
4. Siegel R, DeSantis C, Virgo K, et al. Cancer treatment and survivorship statistics, 2012. CA: a cancer journal for clinicians 2012;**62**(4):220-41 doi: 10.3322/caac.21149[published Online First: Epub Date]|.
5. Hachey KJ, Colson YL. Current innovations in sentinel lymph node mapping for the staging and treatment of resectable lung cancer. Seminars in thoracic and cardiovascular surgery 2014;**26**(3):201-9 doi: 10.1053/j.semtcvs.2014.09.001[published Online First: Epub Date]|.
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. Experimental and therapeutic medicine 2012;**3**(1):134-40 doi: 10.3892/etm.2011.368[published Online First: Epub Date]|.
7. Kim CF, Dirks PB. Cancer and stem cell biology: how tightly intertwined? Cell stem cell 2008;**3**(2):147-50 doi: 10.1016/j.stem.2008.07.012[published Online First: Epub Date]|.
8. Sales KM, Winslet MC, Seifalian AM. Stem cells and cancer: an overview. Stem cell reviews 2007;**3**(4):249-55 doi: 10.1007/s12015-007-9002-0[published Online First: Epub Date]|.
9. Clarke MF. A self-renewal assay for cancer stem cells. Cancer chemotherapy and pharmacology 2005;**56 Suppl 1**:64-8 doi: 10.1007/s00280-005-0097-1[published Online First: Epub Date]|.
10. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 2003;**100**(7):3983-8 doi: 10.1073/pnas.0530291100[published Online First: Epub Date]|.
11. Li Z. CD133: a stem cell biomarker and beyond. Experimental hematology & oncology 2013;**2**(1):17 doi: 10.1186/2162-3619-2-17[published Online First: Epub Date]|.
12. Artells R, Moreno I, Diaz T, et al. Tumour CD133 mRNA expression and clinical outcome in surgically resected colorectal cancer patients. European journal of cancer 2010;**46**(3):642-9 doi: 10.1016/j.ejca.2009.11.003[published Online First: Epub Date]|.
13. Sasaki A, Kamiyama T, Yokoo H, et al. Cytoplasmic expression of CD133 is an important risk factor for overall survival in hepatocellular carcinoma. Oncology reports 2010;**24**(2):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. Molecular carcinogenesis 2015;**54**(6):449-58 doi: 10.1002/mc.22113[published Online First: Epub Date]|.
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. Annals of surgical oncology 2011;**18**(10):2937-45 doi: 10.1245/s10434-011-1703-6[published Online First: Epub Date]|.
16. Nakamura M, Kyo S, Zhang B, et al. Prognostic impact of CD133 expression as a tumor-initiating cell marker in endometrial cancer. Human pathology 2010;**41**(11):1516-29 doi: 10.1016/j.humpath.2010.05.006[published Online First: Epub Date]|.
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 research 2011;**71**(11):3991-4001 doi: 10.1158/0008-5472.CAN-10-3175[published Online First: Epub Date]|.
18. Woo T, Okudela K, Mitsui H, et al. Prognostic value of CD133 expression in stage I lung adenocarcinomas. International journal of clinical and experimental pathology 2010;**4**(1):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**(3):393-400 doi: 10.1007/s00408-016-9876-1[published Online First: Epub Date]|.
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**(10):2292-9 doi: 10.1093/carcin/bgt181[published Online First: Epub Date]|.
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. International journal of clinical and experimental pathology 2015;**8**(5):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**(5):e95915 doi: 10.1371/journal.pone.0095915[published Online First: Epub Date]|.
23. Keysar SB, Jimeno A. More than markers: biological significance of cancer stem cell-defining molecules. Molecular cancer therapeutics 2010;**9**(9):2450-7 doi: 10.1158/1535-7163.MCT-10-0530[published Online First: Epub Date]|.
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 research 2013;**33**(3):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**(1):162-7 doi: 10.1016/j.lungcan.2012.02.006[published Online First: Epub Date]|.

Table 1. Demographic and clinicopathological characteristics of NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 2014

Characteristics		Lung adenocarcinoma (N=793)	Lung squamous cell cancer (N=331)	Others (N=259)
Age (years)	<65	476 (60.0)	151 (45.6)	154 (59.5)
	≥65	317 (40.0)	180 (54.4)	105 (40.5)
Sex	Male	372 (46.9)	206 (62.2)	148 (57.1)
	Female	421 (53.1)	125 (37.8)	111 (42.9)
Smoking Status	Non-smoker	182 (23.0)	24 (7.3)	33 (12.7)
	Former smoker	322 (40.6)	165 (49.8)	100 (38.6)
	Current smoker	289 (36.4)	142 (42.9)	126 (48.6)
Pack-years	≤25	158 (25.9)	23 (7.5)	28 (12.4)
	26-50	233 (38.1)	121 (39.4)	96 (42.5)
	>50	220 (36.0)	163 (53.1)	102 (45.1)
Stage	I-II	234 (29.5)	110 (33.2)	50 (19.3)
	III	271 (34.2)	155 (46.8)	102 (39.4)
	IV	288 (36.3)	66 (19.9)	107 (41.3)
Family history	Yes	121 (15.3)	45 (13.6)	39 (15.1)
	No	672 (84.7)	286 (86.4)	220 (84.9)
	No	274 (34.5)	101 (30.5)	166 (64.1)
Surgery	Lobectomy	271 (34.2)	67 (20.2)	79 (30.5)
	Segmentectomy	130 (16.4)	62 (18.7)	3 (1.2)
	Wedge Resection	118 (14.9)	101 (30.5)	11 (4.3)
History of radiotherapy	No	263 (33.2)	100 (30.2)	143 (55.2)
	Palliative therapy	116 (14.6)	112 (33.8)	27 (10.4)

	Curative intent	414 (52.2)	119 (36.0)	89 (34.4)
Adjuvant chemotherapy	No	444 (56.0)	177 (53.5)	155 (59.9)
	Yes	349 (44.0)	154 (46.5)	104 (40.1)

Table 2. Associations between *CD133* genotypes and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.13-1.47)	1.27 (1.12-1.45)
	CC	172	143 (83.1)	8.2	2.22 (1.84-2.68)	2.32 (1.91-2.80)
Dominant	AA/AC	1207	897 (74.3)	18.1	1.0	1.0
	CC	172	143 (83.1)	8.2	1.98 (1.66-2.36)	2.07 (1.73-2.48)
Recessive	AA	652	463 (71.0)	20.3	1.0	1.0
	AC/CC	727	577 (79.4)	13.0	1.43 (1.27-1.62)	1.43 (1.26-1.61)
rs10022537	TT	913	689 (75.5)	17.2	1.0	1.0
	TA	413	311 (75.3)	14.5	1.06 (0.93- 1.22))	1.10 (0.96-1.27)
	AA	39	30 (76.9)	15.2	1.14 (0.79- 1.64))	1.00 (0.69-1.44)
Dominant	TT/TA	1326	1000 (75.4)	16.8	1.0	1.0
	AA	39	30 (76.9)	15.2	1.12 (0.78-1.61)	0.96 (0.67-1.39)
Recessive	TT	913	689 (75.5)	17.2	1.0	1.0
	TA/AA	452	341 (75.4)	14.5	1.07 (0.94-1.22)	1.13 (0.99-1.29)
rs7686732	CC	398	286 (71.9)	17.1	1.0	1.0
	CG	88	63 (71.6)	15.9	1.06 (0.80-1.39)	1.12 (0.85-1.49)
	GG	5	4 (80.0)	20.1	1.33 (0.49-3.56)	1.26 (0.47-3.41)
Dominant	CC/CG	486	349 (71.8)	17.0	1.0	1.0
	GG	5	4 (80.0)	20.1	1.31 (0.49-3.52)	1.24 (0.46-3.34)
Recessive	CC	398	286 (71.9)	17.1	1.0	1.0
	CG/GG	93	67 (72.0)	15.9	1.07 (0.82-1.39)	1.13 (0.86-1.49)

rs3130	CC	134	92 (68.7)	18.0	1.0	1.0
	CT	269	201 (74.7)	17.0	1.13 (0.88-1.44)	1.14 (0.89-1.47)
	TT	92	64 (69.6)	14.4	0.98 (0.72-1.36)	0.97 (0.70-1.34)
Dominant	CC/CT	403	293 (72.7)	17.3	1.0	1.0
	TT	92	64 (69.6)	14.4	0.91 (0.69-1.19)	0.89 (0.68-1.18)
Recessive	CC	134	92 (68.7)	18.0	1.0	1.0
	CT/TT	361	265 (73.4)	16.8	1.09 (0.86-1.38)	1.09 (0.86-1.39)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

[§] MST, median survival time.

[†] Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.

Table 3. Stratified analysis for associations between *CD133 rs2240688* polymorphism and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.16-1.61)	1.44 (1.22-1.70)
	≥65	209/285	253/316	21.2	vs.	11.9	1.54 (1.28-1.85)	1.46 (1.21-1.76)
Sex	Male	249/334	326/390	18.1	vs.	11.8	1.53 (1.30-1.80)	1.45 (1.23-1.71)
	Female	214/318	251/337	23.0	vs.	16.8	1.32 (1.10-1.59)	1.45 (1.21-1.74)
Smoking Status	Non-smoker	82/116	89/120	18.3	vs.	16.8	1.19 (0.88-1.60)	1.28 (0.94-1.74)
	Former smoker	188/278	241/309	26.0	vs.	13.4	1.54 (1.27-1.86)	1.52 (1.25-1.84)
	Current smoker	193/258	247/298	19.0	vs.	11.6	1.43 (1.18-1.72)	1.44 (1.19-1.74)
Histopathology type	Lung adenocarcinoma	253/377	313/414	23.0	vs.	16.0	1.40 (1.19-1.66)	1.44 (1.22-1.70)
	Lung squamous cell cancer	118/161	132/170	21.0	vs.	12.0	1.40 (1.09-1.79)	1.32 (1.03-1.71)
	Others	92/114	132/143	14.6	vs.	9.0	1.54 (1.18-2.01)	1.51 (1.14-2.00)
Stage	I-II	78/169	140/224	53.9	vs.	18.0	2.09 (1.58-2.77)	2.28 (1.72-3.03)
	III	217/272	219/256	19.0	vs.	15.5	1.15 (0.95-1.39)	1.17 (0.97-1.42)
	IV	168/211	218/247	13.4	vs.	11.2	1.42 (1.16-1.73)	1.43 (1.16-1.75)
Family history	Yes	71/102	80/103	22.2	vs.	11.6	1.52 (1.10-2.09)	1.56 (1.12-2.17)
	No	392/550	497/624	20.0	vs.	13.2	1.42 (1.24-1.62)	1.41 (1.23-1.61)
Surgery	No	191/232	274/306	12.3	vs.	9.6	1.36 (1.13-1.64)	1.38 (1.15- 1.66)
	Lobectomy	157/210	179/206	26.0	vs.	17.0	1.38 (1.11-1.71)	1.40 (1.12- 1.74)
	Segmentectomy	61/95	61/100	25.0	vs.	21.6	1.14 (0.80-1.62)	1.27 (0.87- 1.84)
	Wedge Resection	54/115	63/115	48.9	vs.	23.4	1.72 (1.19-2.48)	1.75 (1.20- 2.54)

History of radiotherapy	No	196/237	230/266	15.0 12.0	vs.	1.17 (0.97-1.42)	1.24 (1.02-1.51)
	Palliative therapy	103/125	113/130	19.0 16.8	vs.	1.19 (0.91-1.56)	1.16 (0.88-1.53)
	Curative intent	164/290	234/331	31.2 12.8	vs.	1.90 (1.55-2.32)	1.90 (1.55-2.33)
Adjuvant chemotherapy	No	294/384	326/391	19.3 16.2	vs.	1.23 (1.05-1.44)	1.22 (1.04-1.43)
	Yes	169/268	251/336	22.7 9.9	vs.	1.78 (1.46-2.16)	1.86 (1.52-2.26)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

[§] MST, median survival time.

[†] Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.

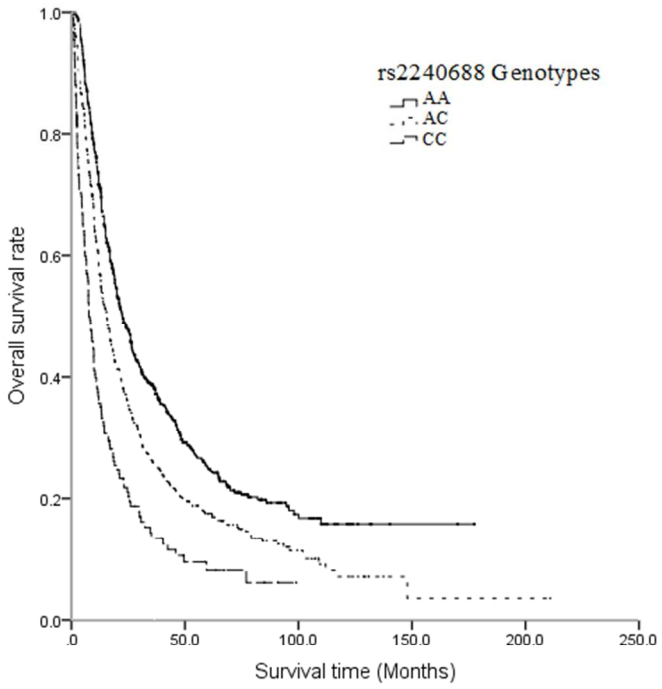

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the rs2240688 genotypes

171x128mm (106 x 106 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	4-5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4-5
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	NA
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, explain how loss to follow-up was addressed	NA
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	7
Outcome data	15*	Report numbers of outcome events or summary measures over time	7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7
		(*b*) Report category boundaries when continuous variables were categorized	7
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	7
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	8
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016913.R1
Article Type:	Research
Date Submitted by the Author:	13-Jun-2017
Complete List of Authors:	Liu, Qing-Feng Zhang, Zhi-Fei Hou, Guang-Jie Yang, Guang-Yu He, Yi; Henan Provincial People's Hospital, Department of thoracic surgery
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Genetics and genomics
Keywords:	Non-small cell lung cancer, CD133 polymorphism, Overall survival, China

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, M.S [1]; Zhi-Fei Zhang, M.S [1]; Guang-Jie Hou, M.D[1]; Guang-Yu Yang, M.D[1]; Ph.D; Yi He, M.S[1];

[1] Department of thoracic surgery, Henan Provincial People's Hospital (Zhengzhou University People's Hospital), 7 Weiwu Street, Zhengzhou city, Henan Province, 450003, PR China

Corresponding Author:

Yi He, Ph.D
Department of thoracic surgery
Henan Provincial People's Hospital
Zhengzhou University People's Hospital
7 Weiwu Street, Zhengzhou city
Henan Province, 450003, PR China
Phone: +86(371)86567371
Fax: +86(371)86567371
Email: hbwang2005@126.com

Running title: Polymorphisms of gene CD133 and the survival of NSCLC

Word Count:

Abstract: 245

Text: 2,640

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, without receiving preoperative neoadjuvant therapies, had available blood sample for genotyping, were eligible for inclusion. Those participants who were pregnant or breastfeeding, had previous cancer history, had second primary tumors, or who had had primary tumors of the skin and nasopharynx, were excluded from this study. Among 1,383 participants, median age was 63 years; 726 (52.5%) males.

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

Results: Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (hazard ratio (HR) 1.27, 95% confidence interval (CI) 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype). Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy (HR 1.86, 95% CI 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33).

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.

Keywords: Non-small cell lung cancer; CD133 polymorphism; Overall survival; China

Article summary

Strengths and limitations of this study

1. A large cohort (1,383 participants) conducted to explore the association between functional SNPs in *CD133 and* OS of lung cancer patients.
2. Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype).
3. Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy and radiotherapy for curative intent.
4. All participants were recruited from a single hospital, which may limit the generalizability of study results.

Introduction

Lung cancer remains the most common cause of cancer-related deaths in China and worldwide [1]. The number of patients who were newly diagnosed with lung cancer was estimated at 1.8 million worldwide in 2012 and 500 thousand 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 other biological agent therapies, the prognosis of NSCLC is very poor due to recurrence and metastasis, with an overall 5-year survival rate less than 16% [4-6]. Hence, it is necessary to identify biomarkers for prevention, early diagnosis, monitoring 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 hypothesized as the origin of cancer due to their potent tumor-driving capability on tumor 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 tumors, including colon cancer [12], liver cancer [13], gastric cancer [14], brain tumor [15], endometrial cancer[16] and ovarian cancer [17].

The expression of *CD133* is significantly correlated with development and prognosis of NSCLC [18]. As single nucleotide polymorphisms (SNPs) are considered to harbor information about genetic variation in functionality of the genome and susceptibility to tumor development[19], we hypothesized 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 included 1,383 consecutive 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 in this retrospective cohort study. 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 retrospective cohort study [19].

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

The present study was performed in accordance with the Declaration of Helsinki, 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 tumor type, tumor-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 two hours of collection. Serum samples were stored in aliquots below −70 °C until analysis.

Details on surgical treatment are 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 three months of surgery. We classified patients as having received radiotherapy if they received external beam radiation, radioactive implants, radioisotopes, brachytherapy or other radiations within six 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-months interval for the first 2 years after surgery, at 6-months 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.

SNPs 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, and *rs3130* C>T) were identified and chosen in our model analysis. Three SNPs (*rs2240688, rs7686732, rs3130*) located in the 3'-UTR of CD133, and *rs10022537* located within intron of the CD133 gene.

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 years survival rate 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 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 two-sided significance level of 0.05.

All statistical tests were performed using SAS™ 9.3 software (Cary, NC, USA). Descriptive analysis results were presented as median and inter-quartile range (IQR) for continuous variables and frequencies (percentage) for categorical variables. Distributions of 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 enrollment until the date of death or date of last follow-up. Survival curves of OS were estimated by Kaplan-Meier method and compared by 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, application of chemotherapy and radiotherapy. In addition, the associations of *CD133* SNPs with OS were analyzed under specific genetic models: genotypic, recessive and dominant model. All these analyses were performed with or without adjustment for demographic variables and selected clinicopathological characteristics. All tests were two-sided and a *P*-value <0.05 was considered to be statistically significant for all analyses.

Results

Baseline characteristics of the study population

There were a total of 1,383 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 summarizes 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%) 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, there were 339 (24.5%) patients alive and 1,044 (75.5%) patients who 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<.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 1.13-1.47 for AC genotype; HR 2.22, 95% CI 1.84-2.68 for CC genotype). As shown in multivariate survival analysis using Cox proportional hazards regression model, *rs2240688* variant genotypes remained significantly associated with OS (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 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 and *rs3130* C>T) and OS of NSCLC patients was not observed in the study.

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, 18.7-22.7) for participants with *rs2240688* AA genotype, 15.6 months (95% CI, 13.0-17.1) with *rs2240688* AC genotype and 8.2 months (95% CI, 7.1-9.9) with *rs2240688* CC genotype.

CD133 Genotypes and OS of NSCLC by clinicopathological characteristics

The stratified analysis was further performed for *rs2240688* A>C by age, sex, smoking status, histopathology type, stage, family history of cancer, application for 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 <65 years, ≥65 years, male, female, former smoker, current smoker, lung adenocarcinoma, 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 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33), compared with patients without adjuvant chemotherapy (HR 1.22, 95% CI 1.04-1.43) and those without radiotherapy (HR 1.24, 95% CI 1.02-1.51). 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 etiology. CSCs are responsible for tumor initiation, growth, migration, aggressiveness, metastasis, drug resistance and pluripotency[8]. In this study, information was collected to determine the role of CSCs in clinical outcomes of NSCLC. We focused on *CD133* gene that has been used to isolate CSCs. Four potential functional SNPs in *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 significantly poorer prognosis of NSCLC patients. The association remained statistically significant (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80

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 locates in the 3'-UTR of *CD133*. SNPs in the 3'-UTR have been shown to have functional effects on 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 tumor-driving capability of CSCs on tumor 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 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 *rs2240688* A>C variant genotypes on tumor 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 *rs2240688* variant genotypes were associated with a favorable 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 because of the different ethnic population, in addition to different sampling methods 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-scale sample should be done to confirm the prognostic and clinical value of *CD133*.

There are some limitations that should be taken into account in this study. First, selection bias cannot be excluded even though inclusion/exclusion criteria were

determined to minimize the bias. Potential confounding factors, such as clinicopathological characteristics, may be associated with SNPs in *CD133*gene and also exert an effect on the overall mortality in our cohort of NSCLC patients. However, the independent association between SNPs in *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. Secondly, in the hospital-based cohort study, all participants were recruited from a single hospital in Henan Province. Therefore, our study setting may limit the generalizability of our results. Finally, our study had made many comparisons in statistical analyses, which might increase type I error.

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.

Figure Legends

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the *rs2240688* genotypes

OS: overall survival.
NSCLC: non-small cell lung cancer.

Author Contributions

Authors Qing-Feng Liu, Zhi-Fei Zhang and Yi He conceived and designed the experiments. Guang-Jie Hou and Guang-Yu Yang performed the experiments. Qing-Feng Liu and Guang-Yu Yang analyzed the data. Qing-Feng Liu, Zhi-Fei Zhang, and Yi He contributed to the writing of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest statement

All the authors declare no conflict of interest.

Funding

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

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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;**19**(8):1893-907 doi: 10.1158/1055-9965.EPI-10-0437[published Online First: Epub Date]|.
2. Liang B, Shao Y, Long F, et al. Predicting Diagnostic Gene Biomarkers for Non-Small-Cell Lung Cancer. BioMed research international 2016;**2016**:3952494 doi: 10.1155/2016/3952494[published Online First: Epub Date]|.
3. Chen E, Zeng Z, Bai B, et al. The prognostic value of CSCs biomarker CD133 in NSCLC: a meta-analysis. Oncotarget 2016 doi: 10.18632/oncotarget.10964[published Online First: Epub Date]|.
4. Siegel R, DeSantis C, Virgo K, et al. Cancer treatment and survivorship statistics, 2012. CA: a cancer journal for clinicians 2012;**62**(4):220-41 doi: 10.3322/caac.21149[published Online First: Epub Date]|.
5. Hachey KJ, Colson YL. Current innovations in sentinel lymph node mapping for the staging and treatment of resectable lung cancer. Seminars in thoracic and cardiovascular surgery 2014;**26**(3):201-9 doi: 10.1053/j.semtcvs.2014.09.001[published Online First: Epub Date]|.
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. Experimental and therapeutic medicine 2012;**3**(1):134-40 doi: 10.3892/etm.2011.368[published Online First: Epub Date]|.
7. Kim CF, Dirks PB. Cancer and stem cell biology: how tightly intertwined? Cell stem cell 2008;**3**(2):147-50 doi: 10.1016/j.stem.2008.07.012[published Online First: Epub Date]|.
8. Sales KM, Winslet MC, Seifalian AM. Stem cells and cancer: an overview. Stem cell reviews 2007;**3**(4):249-55 doi: 10.1007/s12015-007-9002-0[published Online First: Epub Date]|.
9. Clarke MF. A self-renewal assay for cancer stem cells. Cancer chemotherapy and pharmacology 2005;**56 Suppl 1**:64-8 doi: 10.1007/s00280-005-0097-1[published Online First: Epub Date]|.
10. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 2003;**100**(7):3983-8 doi: 10.1073/pnas.0530291100[published Online First: Epub Date]|.
11. Li Z. CD133: a stem cell biomarker and beyond. Experimental hematology & oncology 2013;**2**(1):17 doi: 10.1186/2162-3619-2-17[published Online First: Epub Date]|.
12. Artells R, Moreno I, Diaz T, et al. Tumour CD133 mRNA expression and clinical outcome in surgically resected colorectal cancer patients. European journal of cancer 2010;**46**(3):642-9 doi: 10.1016/j.ejca.2009.11.003[published Online First: Epub Date]|.
13. Sasaki A, Kamiyama T, Yokoo H, et al. Cytoplasmic expression of CD133 is an important risk factor for overall survival in hepatocellular carcinoma. Oncology reports 2010;**24**(2):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. Molecular carcinogenesis 2015;**54**(6):449-58 doi: 10.1002/mc.22113[published Online First: Epub Date]|.
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. Annals of surgical oncology 2011;**18**(10):2937-45 doi: 10.1245/s10434-011-1703-6[published Online First: Epub Date]|.
16. Nakamura M, Kyo S, Zhang B, et al. Prognostic impact of CD133 expression as a tumor-initiating cell marker in endometrial cancer. Human pathology 2010;**41**(11):1516-29 doi: 10.1016/j.humpath.2010.05.006[published Online First: Epub Date]|.
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 research 2011;**71**(11):3991-4001 doi: 10.1158/0008-5472.CAN-10-3175[published Online First: Epub Date]|.
18. Woo T, Okudela K, Mitsui H, et al. Prognostic value of CD133 expression in stage I lung adenocarcinomas. International journal of clinical and experimental pathology 2010;**4**(1):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**(3):393-400 doi: 10.1007/s00408-016-9876-1[published Online First: Epub Date]|.
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**(10):2292-9 doi: 10.1093/carcin/bgt181[published Online First: Epub Date]|.
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. International journal of clinical and experimental pathology 2015;**8**(5):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**(5):e95915 doi: 10.1371/journal.pone.0095915[published Online First: Epub Date]|.
23. Keysar SB, Jimeno A. More than markers: biological significance of cancer stem cell-defining molecules. Molecular cancer therapeutics 2010;**9**(9):2450-7 doi: 10.1158/1535-7163.MCT-10-0530[published Online First: Epub Date]|.
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 research 2013;**33**(3):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**(1):162-7 doi: 10.1016/j.lungcan.2012.02.006[published Online First: Epub Date]|.

Table 1. Demographic and clinicopathological characteristics of NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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)
	No	274 (34.5)	101 (30.5)	166 (64.1)	541 (39.1)
Surgery	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)

Table 2. Associations between *CD133* genotypes and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.13-1.47)	1.27 (1.12-1.45)
	CC	172	143 (83.1)	8.2	2.22 (1.84-2.68)	2.32 (1.91-2.80)
Recessive	AA/AC	1207	897 (74.3)	18.1	1.0	1.0
	CC	172	143 (83.1)	8.2	1.98 (1.66-2.36)	2.07 (1.73-2.48)
Dominant	AA	652	463 (71.0)	20.3	1.0	1.0
	AC/CC	727	577 (79.4)	13.0	1.43 (1.27-1.62)	1.43 (1.26-1.61)
rs10022537	TT	913	689 (75.5)	17.2	1.0	1.0
	TA	413	311 (75.3)	14.5	1.06 (0.93- 1.22))	1.10 (0.96-1.27)
	AA	39	30 (76.9)	15.2	1.14 (0.79- 1.64))	1.00 (0.69-1.44)
Dominant	TT/TA	1326	1000 (75.4)	16.8	1.0	1.0
	AA	39	30 (76.9)	15.2	1.12 (0.78-1.61)	0.96 (0.67-1.39)
Recessive	TT	913	689 (75.5)	17.2	1.0	1.0
	TA/AA	452	341 (75.4)	14.5	1.07 (0.94-1.22)	1.13 (0.99-1.29)
rs7686732[¶]	CC	398	286 (71.9)	17.1	1.0	1.0
	CG	88	63 (71.6)	15.9	1.06 (0.80-1.39)	1.12 (0.85-1.49)
	GG	5	4 (80.0)	20.1	1.33 (0.49-3.56)	1.26 (0.47-3.41)
Dominant	CC/CG	486	349 (71.8)	17.0	1.0	1.0
	GG	5	4 (80.0)	20.1	1.31 (0.49-3.52)	1.24 (0.46-3.34)
Recessive	CC	398	286 (71.9)	17.1	1.0	1.0
	CG/GG	93	67 (72.0)	15.9	1.07 (0.82-1.39)	1.13 (0.86-1.49)

rs3130 ¶	CC	134	92 (68.7)	18.0	1.0	1.0
	CT	269	201 (74.7)	17.0	1.13 (0.88-1.44)	1.14 (0.89-1.47)
	TT	92	64 (69.6)	14.4	0.98 (0.72-1.36)	0.97 (0.70-1.34)
Dominant	CC/CT	403	293 (72.7)	17.3	1.0	1.0
	TT	92	64 (69.6)	14.4	0.91 (0.69-1.19)	0.89 (0.68-1.18)
Recessive	CC	134	92 (68.7)	18.0	1.0	1.0
	CT/TT	361	265 (73.4)	16.8	1.09 (0.86-1.38)	1.09 (0.86-1.39)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

§ MST, median survival time.

† Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.

¶ Genotyping of the two SNPs (*rs7686732 and rs3130*) was only carried out in part of the participants.

Table 3. Stratified analysis for associations between *CD133 rs2240688* polymorphism (dominant for the C allele) and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.16-1.61)	1.44 (1.22-1.70)
	≥65	209/285	253/316	21.2	vs.	11.9	1.54 (1.28-1.85)	1.46 (1.21-1.76)
Sex	Male	249/334	326/390	18.1	vs.	11.8	1.53 (1.30-1.80)	1.45 (1.23-1.71)
	Female	214/318	251/337	23.0	vs.	16.8	1.32 (1.10-1.59)	1.45 (1.21-1.74)
Smoking Status	Non-smoker	82/116	89/120	18.3	vs.	16.8	1.19 (0.88-1.60)	1.28 (0.94-1.74)
	Former smoker	188/278	241/309	26.0	vs.	13.4	1.54 (1.27-1.86)	1.52 (1.25-1.84)
	Current smoker	193/258	247/298	19.0	vs.	11.6	1.43 (1.18-1.72)	1.44 (1.19-1.74)
Histopathology type	Lung adenocarcinoma	253/377	313/414	23.0	vs.	16.0	1.40 (1.19-1.66)	1.44 (1.22-1.70)
	Lung squamous cell cancer	118/161	132/170	21.0	vs.	12.0	1.40 (1.09-1.79)	1.32 (1.03-1.71)
	Others	92/114	132/143	14.6	vs.	9.0	1.54 (1.18-2.01)	1.51 (1.14-2.00)
Stage	I-II	78/169	140/224	53.9	vs.	18.0	2.09 (1.58-2.77)	2.28 (1.72-3.03)
	III	217/272	219/256	19.0	vs.	15.5	1.15 (0.95-1.39)	1.17 (0.97-1.42)
	IV	168/211	218/247	13.4	vs.	11.2	1.42 (1.16-1.73)	1.43 (1.16-1.75)
Family history	Yes	71/102	80/103	22.2	vs.	11.6	1.52 (1.10-2.09)	1.56 (1.12-2.17)
	No	392/550	497/624	20.0	vs.	13.2	1.42 (1.24-1.62)	1.41 (1.23-1.61)
Surgery	No	191/232	274/306	12.3	vs.	9.6	1.36 (1.13-1.64)	1.38 (1.15- 1.66)
	Lobectomy	157/210	179/206	26.0	vs.	17.0	1.38 (1.11-1.71)	1.40 (1.12- 1.74)
	Segmentectomy	61/95	61/100	25.0	vs.	21.6	1.14 (0.80-1.62)	1.27 (0.87- 1.84)
	Wedge Resection	54/115	63/115	48.9	vs.	23.4	1.72 (1.19-2.48)	1.75 (1.20- 2.54)

History of radiotherapy	No	196/237	230/266	15.0 12.0	vs.	1.17 (0.97-1.42)	1.24 (1.02-1.51)
	Palliative therapy	103/125	113/130	19.0 16.8	vs.	1.19 (0.91-1.56)	1.16 (0.88-1.53)
	Curative intent	164/290	234/331	31.2 12.8	vs.	1.90 (1.55-2.32)	1.90 (1.55-2.33)
Adjuvant chemotherapy	No	294/384	326/391	19.3 16.2	vs.	1.23 (1.05-1.44)	1.22 (1.04-1.43)
	Yes	169/268	251/336	22.7 9.9	vs.	1.78 (1.46-2.16)	1.86 (1.52-2.26)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

[§] MST, median survival time.

[†] Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.

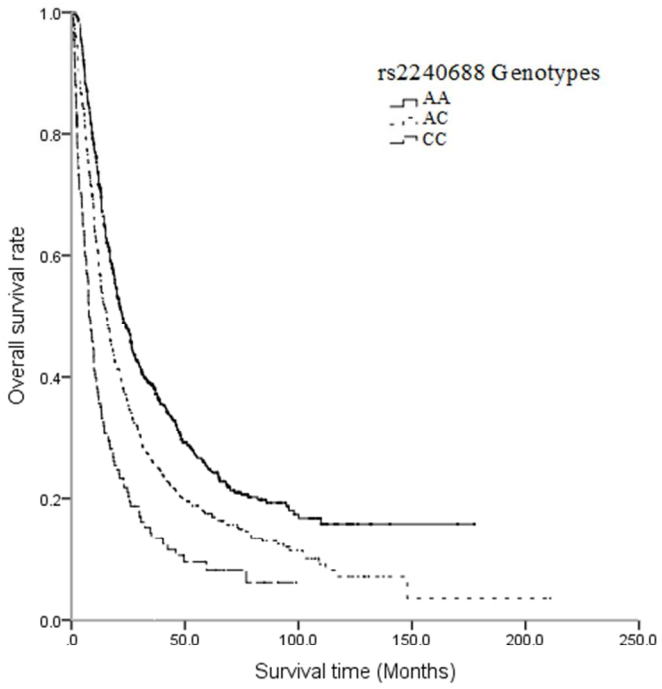

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the rs2240688 genotypes

171x128mm (106 x 106 DPI)

STROBE (STREGA) 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
		State if the study is the first report of a genetic association, a replication effort, or both.	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	4-5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	NA
		Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
		Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin)	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4-5
		Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches	6
Bias	9	Describe any efforts to address potential sources of bias	6
		For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was	NA

		undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.	
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
		If applicable, describe how effects of treatment were dealt with.	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		State software version used and options (or settings) chosen.	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, explain how loss to follow-up was addressed	7
		(*e*) Describe any sensitivity analyses	NA
		State whether Hardy–Weinberg equilibrium was considered and, if so, how	NA
		Describe any methods used for inferring genotypes or haplotypes	NA
		Describe any methods used to assess or address population stratification	7
		Describe any methods used to address multiple comparisons or to control risk of false positive findings	11
		Describe any methods used to address and correct for relatedness among subjects.	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		Report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	7
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		Consider giving information by genotype.	8
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	8
Outcome data	15*	Report numbers of outcome events or summary measures over time	8

		Report outcomes (phenotypes) for each genotype category over time.	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9
		(*b*) Report category boundaries when continuous variables were categorized	8-9
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8-9
		(d) Report results of any adjustments for multiple comparisons.	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	20 (Table 3)
		If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken	8-9
		If detailed results are available elsewhere, state how they can be accessed	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016913.R2
Article Type:	Research
Date Submitted by the Author:	18-Jul-2017
Complete List of Authors:	Liu, Qing-Feng Zhang, Zhi-Fei Hou, Guang-Jie Yang, Guang-Yu He, Yi; Henan Provincial People's Hospital, Department of thoracic surgery
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Genetics and genomics
Keywords:	Non-small cell lung cancer, CD133 polymorphism, Overall survival, China

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, M.S [1]; Zhi-Fei Zhang, M.S [1]; Guang-Jie Hou, M.D[1]; Guang-Yu Yang, M.D[1]; Ph.D; Yi He, M.S[1];

[1] Department of thoracic surgery, Henan Provincial People's Hospital (Zhengzhou University People's Hospital), 7 Weiwu Street, Zhengzhou city, Henan Province, 450003, PR China

Corresponding Author:

Yi He, Ph.D
Department of thoracic surgery
Henan Provincial People's Hospital
Zhengzhou University People's Hospital
7 Weiwu Street, Zhengzhou city
Henan Province, 450003, PR China
Phone: +86(371)86567371
Fax: +86(371)86567371
Email: hbwang2005@126.com

Running title: Polymorphisms of gene CD133 and the survival of NSCLC

Word Count:

Abstract: 245

Text: 2,640

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, were not receiving preoperative neoadjuvant therapies, and had an available blood sample for genotyping were eligible for inclusion. Those participants who were pregnant or breastfeeding, had previous cancer history, had other primary tumors, or who had had primary tumors of the skin and nasopharynx, were excluded from this study. Among 1,383 participants, median age was 63 years; 726 (52.5%) males.

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

Results: Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (hazard ratio (HR) 1.27, 95% confidence interval (CI) 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype). Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy (HR 1.86, 95% CI 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33).

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.

Keywords: Non-small cell lung cancer; CD133 polymorphism; overall survival; China

Article summary

Strengths and limitations of this study

1. A large cohort (1,383 participants) was studied to explore the association between functional SNPs in *CD133 and* OS of lung cancer patients.
2. Compared with *rs2240688* AA genotype, the variant AC/CC genotypes were independently associated with OS (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 for CC genotype).
3. Higher HRs for associations between *CD133 rs2240688* polymorphism and OS were observed in patients with adjuvant chemotherapy and radiotherapy for curative intent.
4. All participants were recruited from a single hospital, which may limit the generalizability of study results.

Introduction

Lung cancer remains the most common cause of cancer-related deaths in China and worldwide [1]. The number of patients who were newly diagnosed with lung cancer was estimated at 1.8 million worldwide in 2012 and 500 thousand 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 other biological agent therapies, the prognosis of NSCLC is very poor due to recurrence and metastasis, with an overall 5-year survival rate less than 16% [4-6]. Hence, it is necessary to identify biomarkers for prevention, early diagnosis, monitoring 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 hypothesized as the origin of cancer due to their potent tumor-driving capability on tumor 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 tumors, including colon cancer [12], liver cancer [13], gastric cancer [14], brain tumor [15], endometrial cancer[16] and ovarian cancer [17].

The expression of *CD133* is significantly correlated with development and prognosis of NSCLC [18]. As single nucleotide polymorphisms (SNPs) are considered to harbor information about genetic variation in functionality of the genome and susceptibility to tumor development[19], we hypothesized 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 included 1,383 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 in this retrospective cohort study. 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 retrospective cohort study [19].

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

The present study was performed in accordance with the Declaration of Helsinki, 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 tumor type, tumor-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 two 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 three months of surgery. We classified patients as having received radiotherapy if they received external beam radiation, radioactive implants, radioisotopes, brachytherapy or other radiations within six 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-months intervals for the first 2 years after surgery, at 6-months 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.

SNPs 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, and *rs3130* C>T) were identified and chosen in our model analysis. Three SNPs (*rs2240688, rs7686732, rs3130*) located in the 3'-UTR of CD133 and *rs10022537* located within intron of the CD133 gene.

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 years survival rate 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 1,234 participants (at 1:1 ratio, 617 subjects in each group) are required to provide 80% power, with the use of two-sided significance level of 0.05.

All statistical tests were performed using SAS 9.3 software (Cary, NC, USA). Descriptive analysis results were presented as median and inter-quartile range (IQR) for continuous variables and frequencies (percentage) for categorical variables. Distributions of 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 enrollment until the date of death or date of last follow-up. Survival curves of OS were estimated by Kaplan-Meier method and compared by 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 analyzed 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 *P*-value <0.05 was considered to be statistically significant for all analyses.

Results

Baseline characteristics of the study population

There were a total of 1,383 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 summarizes 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%) 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, there were 339 (24.5%) patients alive and 1,044 (75.5%) patients who 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<.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 1.13-1.47 for AC genotype; HR 2.22, 95% CI 1.84-2.68 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 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 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 and *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,

18.7-22.7) for participants with *rs2240688* AA genotype, 15.6 months (95% CI, 13.0-17.1) with *rs2240688* AC genotype and 8.2 months (95% CI, 7.1-9.9) with *rs2240688* CC genotype.

CD133 Genotypes and OS of NSCLC by clinicopathological characteristics

The 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, 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 1.52-2.26) and radiotherapy for curative intent (HR 1.90, 95% CI 1.55-2.33), compared with patients without adjuvant chemotherapy (HR 1.22, 95% CI 1.04-1.43) and those without radiotherapy (HR 1.24, 95% CI 1.02-1.51). 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 etiology. CSCs are responsible for tumor initiation, growth, migration, aggressiveness, metastasis, drug resistance and pluripotency[8]. In this study, information was collected to determine the role of CSCs in clinical outcomes of NSCLC. We focused on *CD133* gene that has been used to isolate CSCs. Four potential functional SNPs in *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 significantly poorer prognosis for NSCLC patients. The association remained statistically significant (HR 1.27, 95% CI 1.12-1.45 for AC genotype; HR 2.32, 95% CI 1.91-2.80 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 locates in the 3'-UTR of *CD133*. SNPs in the 3'-UTR have been shown to have functional effects on 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 tumor-driving capability of CSCs on tumor 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 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 *rs2240688* A>C variant genotypes on tumor 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 *rs2240688* variant genotypes were associated with a favorable 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 because of the different ethnic population, in addition to different sampling methods 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 done 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 minimize the bias. Potential confounding factors, such as clinicopathological characteristics, may be associated with SNPs in *CD133*gene and also exert an effect on the overall mortality in our cohort of NSCLC patients. However, the independent association between SNPs in *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. Secondly, in the hospital-based cohort study, all participants were recruited from a single hospital in Henan Province. Therefore, our study setting may limit the generalizability 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.

Figure Legends

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the *rs2240688* genotypes

OS: overall survival.
NSCLC: non-small cell lung cancer.

Author Contributions

Authors Qing-Feng Liu, Zhi-Fei Zhang and Yi He conceived and designed the experiments. Guang-Jie Hou and Guang-Yu Yang performed the experiments. Qing-Feng Liu and Guang-Yu Yang analyzed the data. Qing-Feng Liu, Zhi-Fei Zhang and Yi He contributed to the writing of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest statement

All the authors declare no conflict of interest.

Funding

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

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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;**19**(8):1893-907 doi: 10.1158/1055-9965.EPI-10-0437[published Online First: Epub Date]|.
2. Liang B, Shao Y, Long F, et al. Predicting Diagnostic Gene Biomarkers for Non-Small-Cell Lung Cancer. BioMed research international 2016;**2016**:3952494 doi: 10.1155/2016/3952494[published Online First: Epub Date]|.
3. Chen E, Zeng Z, Bai B, et al. The prognostic value of CSCs biomarker CD133 in NSCLC: a meta-analysis. Oncotarget 2016 doi: 10.18632/oncotarget.10964[published Online First: Epub Date]|.
4. Siegel R, DeSantis C, Virgo K, et al. Cancer treatment and survivorship statistics, 2012. CA: a cancer journal for clinicians 2012;**62**(4):220-41 doi: 10.3322/caac.21149[published Online First: Epub Date]|.
5. Hachey KJ, Colson YL. Current innovations in sentinel lymph node mapping for the staging and treatment of resectable lung cancer. Seminars in thoracic and cardiovascular surgery 2014;**26**(3):201-9 doi: 10.1053/j.semtcvs.2014.09.001[published Online First: Epub Date]|.
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. Experimental and therapeutic medicine 2012;**3**(1):134-40 doi: 10.3892/etm.2011.368[published Online First: Epub Date]|.
7. Kim CF, Dirks PB. Cancer and stem cell biology: how tightly intertwined? Cell stem cell 2008;**3**(2):147-50 doi: 10.1016/j.stem.2008.07.012[published Online First: Epub Date]|.
8. Sales KM, Winslet MC, Seifalian AM. Stem cells and cancer: an overview. Stem cell reviews 2007;**3**(4):249-55 doi: 10.1007/s12015-007-9002-0[published Online First: Epub Date]|.
9. Clarke MF. A self-renewal assay for cancer stem cells. Cancer chemotherapy and pharmacology 2005;**56 Suppl 1**:64-8 doi: 10.1007/s00280-005-0097-1[published Online First: Epub Date]|.
10. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 2003;**100**(7):3983-8 doi: 10.1073/pnas.0530291100[published Online First: Epub Date]|.
11. Li Z. CD133: a stem cell biomarker and beyond. Experimental hematology & oncology 2013;**2**(1):17 doi: 10.1186/2162-3619-2-17[published Online First: Epub Date]|.
12. Artells R, Moreno I, Diaz T, et al. Tumour CD133 mRNA expression and clinical outcome in surgically resected colorectal cancer patients. European journal of cancer 2010;**46**(3):642-9 doi: 10.1016/j.ejca.2009.11.003[published Online First: Epub Date]|.
13. Sasaki A, Kamiyama T, Yokoo H, et al. Cytoplasmic expression of CD133 is an important risk factor for overall survival in hepatocellular carcinoma. Oncology reports 2010;**24**(2):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. Molecular carcinogenesis 2015;**54**(6):449-58 doi: 10.1002/mc.22113[published Online First: Epub Date]|.
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. Annals of surgical oncology 2011;**18**(10):2937-45 doi: 10.1245/s10434-011-1703-6[published Online First: Epub Date]|.
16. Nakamura M, Kyo S, Zhang B, et al. Prognostic impact of CD133 expression as a tumor-initiating cell marker in endometrial cancer. Human pathology 2010;**41**(11):1516-29 doi: 10.1016/j.humpath.2010.05.006[published Online First: Epub Date]|.
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 research 2011;**71**(11):3991-4001 doi: 10.1158/0008-5472.CAN-10-3175[published Online First: Epub Date]|.
18. Woo T, Okudela K, Mitsui H, et al. Prognostic value of CD133 expression in stage I lung adenocarcinomas. International journal of clinical and experimental pathology 2010;**4**(1):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**(3):393-400 doi: 10.1007/s00408-016-9876-1[published Online First: Epub Date]|.
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**(10):2292-9 doi: 10.1093/carcin/bgt181[published Online First: Epub Date]|.
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. International journal of clinical and experimental pathology 2015;**8**(5):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**(5):e95915 doi: 10.1371/journal.pone.0095915[published Online First: Epub Date]|.
23. Keysar SB, Jimeno A. More than markers: biological significance of cancer stem cell-defining molecules. Molecular cancer therapeutics 2010;**9**(9):2450-7 doi: 10.1158/1535-7163.MCT-10-0530[published Online First: Epub Date]|.
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 research 2013;**33**(3):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**(1):162-7 doi: 10.1016/j.lungcan.2012.02.006[published Online First: Epub Date]|.

Table 1. Demographic and clinicopathological characteristics of NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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)

Table 2. Associations between *CD133* genotypes and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.13-1.47)	1.27 (1.12-1.45)
	CC	172	143 (83.1)	8.2	2.22 (1.84-2.68)	2.32 (1.91-2.80)
Recessive	AA/AC	1207	897 (74.3)	18.1	1.0	1.0
	CC	172	143 (83.1)	8.2	1.98 (1.66-2.36)	2.07 (1.73-2.48)
Dominant	AA	652	463 (71.0)	20.3	1.0	1.0
	AC/CC	727	577 (79.4)	13.0	1.43 (1.27-1.62)	1.43 (1.26-1.61)
rs10022537	TT	913	689 (75.5)	17.2	1.0	1.0
	TA	413	311 (75.3)	14.5	1.06 (0.93- 1.22))	1.10 (0.96-1.27)
	AA	39	30 (76.9)	15.2	1.14 (0.79- 1.64))	1.00 (0.69-1.44)
Dominant	TT/TA	1326	1000 (75.4)	16.8	1.0	1.0
	AA	39	30 (76.9)	15.2	1.12 (0.78-1.61)	0.96 (0.67-1.39)
Recessive	TT	913	689 (75.5)	17.2	1.0	1.0
	TA/AA	452	341 (75.4)	14.5	1.07 (0.94-1.22)	1.13 (0.99-1.29)
rs7686732[¶]	CC	398	286 (71.9)	17.1	1.0	1.0
	CG	88	63 (71.6)	15.9	1.06 (0.80-1.39)	1.12 (0.85-1.49)
	GG	5	4 (80.0)	20.1	1.33 (0.49-3.56)	1.26 (0.47-3.41)
Dominant	CC/CG	486	349 (71.8)	17.0	1.0	1.0
	GG	5	4 (80.0)	20.1	1.31 (0.49-3.52)	1.24 (0.46-3.34)
Recessive	CC	398	286 (71.9)	17.1	1.0	1.0
	CG/GG	93	67 (72.0)	15.9	1.07 (0.82-1.39)	1.13 (0.86-1.49)

rs3130 ¶	CC	134	92 (68.7)	18.0	1.0	1.0
	CT	269	201 (74.7)	17.0	1.13 (0.88-1.44)	1.14 (0.89-1.47)
	TT	92	64 (69.6)	14.4	0.98 (0.72-1.36)	0.97 (0.70-1.34)
Dominant	CC/CT	403	293 (72.7)	17.3	1.0	1.0
	TT	92	64 (69.6)	14.4	0.91 (0.69-1.19)	0.89 (0.68-1.18)
Recessive	CC	134	92 (68.7)	18.0	1.0	1.0
	CT/TT	361	265 (73.4)	16.8	1.09 (0.86-1.38)	1.09 (0.86-1.39)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

§ MST, median survival time.

† 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.

Table 3. Stratified analysis for associations between *CD133 rs2240688* polymorphism (dominant for the C allele) and OS among NSCLC patients recruited from Henan Provincial People's Hospital between Jan 2006 and Dec 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 (1.16-1.61)	1.44 (1.22-1.70)
	≥65	209/285	253/316	21.2	vs.	11.9	1.54 (1.28-1.85)	1.46 (1.21-1.76)
Sex	Male	249/334	326/390	18.1	vs.	11.8	1.53 (1.30-1.80)	1.45 (1.23-1.71)
	Female	214/318	251/337	23.0	vs.	16.8	1.32 (1.10-1.59)	1.45 (1.21-1.74)
Smoking Status	Non-smoker	82/116	89/120	18.3	vs.	16.8	1.19 (0.88-1.60)	1.28 (0.94-1.74)
	Former smoker	188/278	241/309	26.0	vs.	13.4	1.54 (1.27-1.86)	1.52 (1.25-1.84)
	Current smoker	193/258	247/298	19.0	vs.	11.6	1.43 (1.18-1.72)	1.44 (1.19-1.74)
Histopathology type	Lung adenocarcinoma	253/377	313/414	23.0	vs.	16.0	1.40 (1.19-1.66)	1.44 (1.22-1.70)
	Lung squamous cell cancer	118/161	132/170	21.0	vs.	12.0	1.40 (1.09-1.79)	1.32 (1.03-1.71)
	Others	92/114	132/143	14.6	vs.	9.0	1.54 (1.18-2.01)	1.51 (1.14-2.00)
Stage	I-II	78/169	140/224	53.9	vs.	18.0	2.09 (1.58-2.77)	2.28 (1.72-3.03)
	III	217/272	219/256	19.0	vs.	15.5	1.15 (0.95-1.39)	1.17 (0.97-1.42)
	IV	168/211	218/247	13.4	vs.	11.2	1.42 (1.16-1.73)	1.43 (1.16-1.75)
Family history	Yes	71/102	80/103	22.2	vs.	11.6	1.52 (1.10-2.09)	1.56 (1.12-2.17)
	No	392/550	497/624	20.0	vs.	13.2	1.42 (1.24-1.62)	1.41 (1.23-1.61)
Surgery	No	191/232	274/306	12.3	vs.	9.6	1.36 (1.13-1.64)	1.38 (1.15- 1.66)
	Lobectomy	157/210	179/206	26.0	vs.	17.0	1.38 (1.11-1.71)	1.40 (1.12- 1.74)
	Segmentectomy	61/95	61/100	25.0	vs.	21.6	1.14 (0.80-1.62)	1.27 (0.87- 1.84)
	Wedge Resection	54/115	63/115	48.9	vs.	23.4	1.72 (1.19-2.48)	1.75 (1.20- 2.54)

History of radiotherapy	No	196/237	230/266	15.0 12.0	vs.	1.17 (0.97-1.42)	1.24 (1.02-1.51)
	Palliative therapy	103/125	113/130	19.0 16.8	vs.	1.19 (0.91-1.56)	1.16 (0.88-1.53)
	Curative intent	164/290	234/331	31.2 12.8	vs.	1.90 (1.55-2.32)	1.90 (1.55-2.33)
Adjuvant chemotherapy	No	294/384	326/391	19.3 16.2	vs.	1.23 (1.05-1.44)	1.22 (1.04-1.43)
	Yes	169/268	251/336	22.7 9.9	vs.	1.78 (1.46-2.16)	1.86 (1.52-2.26)

CI, confidence interval; HR, hazard ratio; OS, overall survival.

[§] MST, median survival time.

[†] Adjusted for age, sex, smoking status, histopathology type, stage, chemotherapy and radiotherapy.

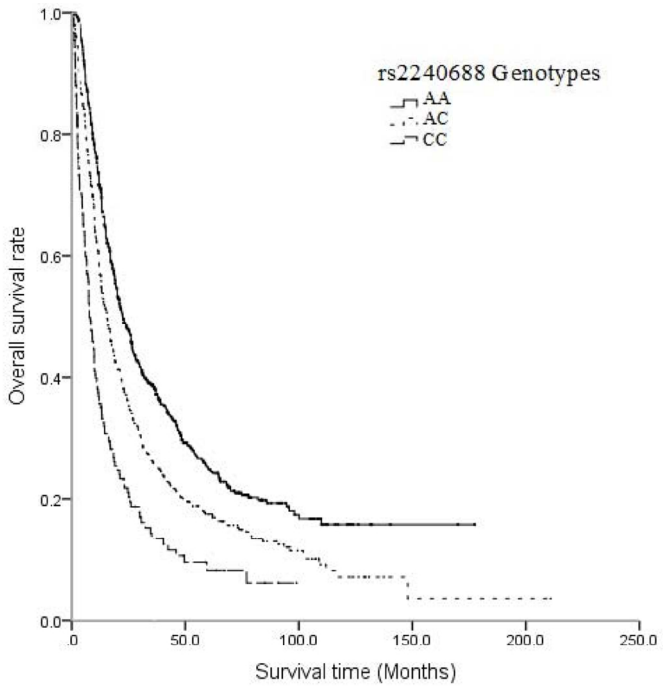

Figure 1. Kaplan-Meier curves for OS among NSCLC patients stratified by the rs2240688 genotypes

128x95mm (300 x 300 DPI)

STROBE (STREGA) 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
		State if the study is the first report of a genetic association, a replication effort, or both.	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	4-5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	NA
		Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
		Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin)	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4-5
		Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches	6
Bias	9	Describe any efforts to address potential sources of bias	6
		For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was	NA

		undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.	
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
		If applicable, describe how effects of treatment were dealt with.	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		State software version used and options (or settings) chosen.	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, explain how loss to follow-up was addressed	7
		(*e*) Describe any sensitivity analyses	NA
		State whether Hardy–Weinberg equilibrium was considered and, if so, how	NA
		Describe any methods used for inferring genotypes or haplotypes	NA
		Describe any methods used to assess or address population stratification	7
		Describe any methods used to address multiple comparisons or to control risk of false positive findings	11
		Describe any methods used to address and correct for relatedness among subjects.	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		Report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	7
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		Consider giving information by genotype.	8
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	8
Outcome data	15*	Report numbers of outcome events or summary measures over time	8

		Report outcomes (phenotypes) for each genotype category over time.	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9
		(*b*) Report category boundaries when continuous variables were categorized	8-9
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8-9
		(d) Report results of any adjustments for multiple comparisons.	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	20 (Table 3)
		If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken	8-9
		If detailed results are available elsewhere, state how they can be accessed	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.