Polymorphisms of \( \text{EpCAM} \) gene and prognosis for non-small-cell lung cancer in Han Chinese

Yuefan Yang,¹,² Fei Fei,²,³ Yang Song,³ Xiaofei Li,⁴ Zhipei Zhang,⁴ Zhou Fei,¹ Haichuan Su¹ and Shaogui Wan⁵

¹Department of Neurosurgery, Xijing Hospital, The Fourth Military Medical University, Xi’an; ²State Key Laboratory of Cancer Biology, Cell Engineering Research Center and Department of Cell Biology, The Fourth Military Medical University, Xi’an; ³Departments of Oncology; ⁴Thoracic Surgery, Tangdu Hospital, The Fourth Military Medical University, Xi’an; ⁵Pharmaceutical College of Henan University, Kaifeng, Henan, China

Key words
Chinese population, EpCAM, non-small-cell lung cancer, prognosis, single nucleotide polymorphism

Correspondence
Shaogui Wan, Pharmaceutical College of Henan University, Jinming District, Kaifeng, Henan 475001, China. Tel: +86-371-2269-1382; Fax: +86-371-2269-1381; E-mail: shaoguiwan@hotmail.com

Haichuan Su, Department of Oncology, Tangdu Hospital, The Fourth Military Medical University, 569 Xinsi Road, Xi’an 710038, China. Tel: +86-29-84778455; Fax: +86-29-84772286; E-mail: suhc@fmmu.edu.cn

These authors contributed equally to this work.

Funding information
Program for New Century Excellent Talents in University (81201583). National Key Scientific and Technological Project (2011ZX09307-001-04).

Received August 28, 2013; Revised October 28, 2013; Accepted October 30, 2013

Cancer Sci | 2014 | Vol. 105 | No. 1 | 89–96

doi: 10.1111/cas.12318

Lung cancer is the most common malignancy in the world. It is a main cause of cancer-associated death in both men and women. Non-small-cell lung cancer (NSCLC) accounts for approximately 80% of primary lung cancers. Tobacco use is known as the most significant risk factor leading to lung cancer. Other known risk factors for lung cancer include environmental and occupational exposures. However, the major causes underlying lung cancer development are genetic and epigenetic damage resulting from environmental carcinogens. In spite of advances achieved in diagnosis and treatment in recent years, the clinical outcome for patients with lung cancer remains disappointing. The 5-year survival rate of NSCLC patients who receive surgical resection is only approximately 15%. Although the key prognostic determinant in lung cancer is clinical stage at diagnosis, it is common to observe that patients at a similar stage have considerably different clinical outcomes. In fact, it is beneficial to improve survival rates after surgical operations by finding suitable biomarkers for prognosis prediction in lung cancer.

The \( \text{EpCAM} \) gene is located on chromosome 2q21, which encodes a carcinoma-associated type I membrane protein of 314 amino acids. It is expressed in a variety of human epithelial tissues and cancer tissues, as well as progenitor and stem cells. Previous studies have reported that epithelial cell adhesion molecule (EpCAM) plays diverse roles in different cancers. Overexpressed EpCAM has been confirmed to increase the expression of c-Myc and cyclins A and E, and then induce a direct impact on cell cycle control. Cells with high levels of EpCAM have a reduced requirement for growth factors; they grow in an anchorage-independent manner and proliferate more rapidly. However, previous studies have also reported that EpCAM has protective effects in some cancers such as renal cell, thyroid, oral squamous cell, and gastric carcinomas. Therefore, whether EpCAM functions as an oncogene or as a tumor suppressor gene mainly relies on the tumor microenvironment. With regard to lung cancer, cumulative experimental evidence has shown that knockdown of \( \text{EpCAM} \) inhibits tumor growth and results in massive apoptosis of cancer cells, which indicates a promising therapeutic target for lung cancer patients.

Single nucleotide polymorphisms located in encoded regions that alter the amino acid sequence of a protein are termed non-
synonymous SNPs (ns-SNPs). Recently, ns-SNPs have been investigated as biomarkers for risk assessment and prognosis prediction, as they are closely associated with a variety of diseases and their development.\(^{(20,21)}\) Previously published reports showed that an ns-SNP in the EpCAM gene (rs1126497) is associated with risk of breast cancer and cervical cancer.\(^{(22,23)}\) In addition, EpCAM plays an important role in hereditary diseases of non-polypsis colorectal cancer (Lynch syndrome) and congenital tufting enteropathy.\(^{(24)}\) Another study reported that a nonsense mutation in the EpCAM gene is significantly associated with congenital tufting enteropathy patients.\(^{(25)}\) All of these results show considerable functions of the EpCAM gene in human hereditary and tumor diseases.

In this study, we selected two SNPs in the EpCAM gene and assessed their effects on the clinical outcomes of NSCLC patients. To the best of our knowledge, this is the first study on the prognostic role of SNPs in the EpCAM gene in NSCLC.

Materials and Methods

**Study group.** A total of 577 patients with NSCLC who had received surgical resection treatment were initially recruited into an ongoing molecular epidemiological study at the Department of Thoracic Surgery at Tangdu Hospital (Fourth Military Medical University, Xi’an, China) between July 2009 and February 2012. None of the patients had been treated with surgery, chemotherapy, and/or radiotherapy before enrolment into the study. There was no restriction for age, gender, or disease stage at enrolment, but the study was restricted to patients with the NSCLC type of lung cancer. In this prognosis study, we excluded 71 patients, including 32 who had incomplete clinical information or failed follow-up, 13 who died within 2 months after surgery, 17 patients with metastatic diseases, and nine with recurrence within 1 month after surgery. Finally, 506 NSCLC patients were included in the present study and successfully genotyped. All patients were Han Chinese. This study was approved by the Institutional Review Boards of the Fourth Military Medical University, and signed informed consent was obtained from each participant.

**Demographic and clinical data.** Demographic data was collected through in-person interviews using a standard epidemiological questionnaire, including age, gender, ethnicity, residential region, smoking status, alcohol use, education status, and family history of cancer. Detailed clinical information was collected through medical chart review, including medical imaging data, pathological diagnosis, TNM stage, and treatment information. A standard follow-up was carried out by a trained clinical specialist through medical record review or telephone interview. The latest follow-up data in this analysis were obtained in November 2012, and 41 (8.1%) patients were lost during follow-up and censored for the analyses. All the patients donated 5 mL blood before treatment for genomic DNA extraction using the EZNA Blood DNA Midi Kit (Omega Bio-Tek, Norcross, GA, USA) in the laboratory.

**Single nucleotide polymorphism selection and genotyping.** Functional SNPs in the EpCAM gene were selected using a set of Web-based SNP selection tools. We found two functional SNPs, rs1421 and rs1126497. The former, s1421, is located in the 3′-UTR, which was predicted to be a potential microRNA binding site by the FuncPred tool (available from http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) and may influence the mRNA degradation or translation. The SNP rs1126497 is an ns-SNP located in exon 3, which changes the 115 Met to Thr and may affect the protein function. Genotyping was carried out using the iPLEX genotyping system (Sequenom, San Diego, CA, USA). Laboratory personnel who undertook the genotyping assays were blinded to patients’ information. The average call rate for the genotyping was 98.5%. Strict quality control measures were implemented during genotyping with more than 99% concordance in samples that were randomly selected to be genotyped in duplicate.

**Statistical analysis.** The srs version 19.0 statistical software (SPSS Inc., Chicago, IL, USA) was used for all statistical analyses. Three genetic models (additive, dominant, and recessive) were applied to assess the association between single SNP and clinical outcomes of NSCLC patients. Two major endpoints, overall survival (OS) and recurrence-free survival (RFS), were analyzed in this study. Overall survival was defined as the time from diagnosis to death from any cause; RFS was defined as the time from diagnosis to the first date when patients developed recurrence. Considering that there was a very small number (≤5) of patients with the rare homozygous variant genotype for these two SNPs of the EpCAM gene, we did not present the data from the recessive model in our analyses to avoid unstable estimations. Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated from a multivariate Cox proportional hazards model, adjusting for age, gender, smoking status, histology, TNM stage, differentiation, and adjuvant chemotherapy or radiotherapy. Kaplan–Meier survival curves and the log-rank test were used to assess the differences among patient groups in OS and RFS analyses. All P-values in this study were two-sided. P < 0.05 was considered statistically significant.

**Results**

**Demographic and clinical characteristics and prognosis analysis.** A total of 506 NSCLC patients were included in this study from Tangdu Hospital. They were an ethnically homogenous group of Han Chinese, the median age was 59 years (range, 27–86 years), and 392 (77.5%) were male patients. Only 160 (31.6%) patients were never smokers, and all others had a smoking history. Of the total 506 patients, 271 were diagnosed with squamous cell carcinoma, 152 with adenocarcinoma, 62 with adenosquamous carcinoma, 16 with carcinosarcoma, three with large-cell lung cancer, and two with mucoepidermoid carcinoma. There were 284 (56.1%) patients diagnosed with stage I and stage II disease, and 222 (43.9%) patients had stage III disease. There were 347 (68.6%) patients with well or moderately differentiated tumors, and 159 (31.4%) with poorly differentiated or undifferentiated tumors. According to the records, 375 patients received adjuvant chemotherapy or radiotherapy.

**Associations between patients’ characteristics and NSCLC outcomes.** We carried out a multivariate Cox regression analysis to assess the associations between clinical characteristics and outcomes in NSCLC patients. As shown in Table 1, patients with adenocarcinoma showed better OS (HR = 0.61; 95% CI, 0.39–0.93) than those with squamous cell carcinoma. Patients at advanced tumor stage had significantly increased risk for death and recurrence with HRs of 1.94 (95% CI, 1.41–2.68) and 2.01 (95% CI, 1.53–2.65), respectively. Patients with poorly differentiated or undifferentiated tumors showed significant association with worse RFS (HR for recurrence = 1.61; 95% CI, 1.14–2.28). In addition, adjuvant chemotherapy or radiotherapy significantly improved the OS of patients with NSCLC (HR for death = 0.67; 95% CI, 0.47–0.69).

**Associations between SNPs and clinical outcome of NSCLC.** The Multivariate Cox regression model was used to evaluate the effect of EpCAM SNPs on death and recurrence in NSCLC patients. To the best of our knowledge, this is the first study on the functional role of SNPs in the EpCAM gene in NSCLC. The average call rate for the genotyping was 98.5%, and the concordance in samples that were randomly selected to be genotyped in duplicate was 99%. Strict quality control measures were implemented during genotyping. The average call rate for the genotyping was 98.5%. The results show considerable functions of the EpCAM gene in human hereditary and tumor diseases.
patients. As shown in Table 2, patients carrying CT or TT genotypes in the SNP rs1126497 showed a significantly higher death risk when compared with those carrying the CC genotype (HR for additive model = 1.46; 95% CI, 1.09–1.97, \( P = 0.012 \); HR for dominant model = 1.41; 95% CI, 1.02–1.94, \( P = 0.039 \)). Similar results were observed in the RFS analysis (HR for additive model = 1.37; 95% CI, 1.06–1.76, \( P = 0.016 \); HR for dominant model = 1.34; 95% CI, 1.01–1.77, \( P = 0.039 \)). Additionally, there were no significant results observed for the SNP rs1421 in either the OS or RFS analyses. For rs1126497, Kaplan–Meier curves significantly distinguished the patient groups with different genotypes in the OS analysis (log–rank \( P \) for additive and dominant models = 0.007 and 0.021, respectively) and RFS analysis (log–rank \( P \) for additive and dominant models = 0.008 and 0.006, respectively) (Fig. 1).

**Associations between EpCAM polymorphisms and NSCLC progression.** We applied binary logistic regression to explore the association between EpCAM SNPs and cancer progression. As shown in Table 3, the AG+GG genotype in rs1421 was significantly associated with early T stages (T1/T2) when compared with the AA genotype (odds ratio [OR] for late stage = 0.65; 95% CI, 0.44–0.96, \( P = 0.029 \)). However, there was no significant association between the SNP rs1421 and N stage. Additionally, the SNP rs1126497 had no significant association with either the T or N stage of NSCLC.

### Table 1. Distribution of characteristics in Chinese patients with non-small-cell lung cancer (n = 506) and prognosis analysis

| Variables                      | No. of total patients | Overall survival | Recurrence-free survival |
|--------------------------------|-----------------------|------------------|-------------------------|
|                                | \( n = 506, \text{ no. (\%)} \) | \( n = 159, \text{ no. (\%)} \) | \( n = 220, \text{ no. (\%)} \) |
| Age                            |                       | \( \text{HR (95\% CI)\text{\( ^{\dagger} \)}} \) | \( \text{P-value} \) | \( \text{HR (95\% CI)\text{\( ^{\dagger} \)}} \) | \( \text{P-value} \) |
| \( \leq 59 \text{ years} \)    | 251 (49.6)            | 77 (48.4)        | Ref.                    | 117 (53.2) | Ref.                    | 0.340 |
| \( >59 \text{ years} \)       | 255 (50.4)            | 82 (51.6)        | 0.99 (0.72–1.38)        | 103 (46.8) | 0.87 (0.66–1.15)        | 0.340 |
| Gender                         |                       |                  |                         |           |                         |       |
| Female                         | 114 (22.5)            | 36 (22.6)        | Ref.                    | 54 (24.5) | Ref.                    | 0.622 |
| Male                           | 392 (77.5)            | 123 (77.4)       | 0.97 (0.57–1.66)        | 166 (75.5) | 1.12 (0.71–1.78)        | 0.231 |
| Smoking status                 |                       |                  |                         |           |                         |       |
| Never smoker                   | 160 (31.6)            | 50 (31.4)        | Ref.                    | 75 (34.1) | Ref.                    |       |
| Ever smoker                    | 346 (68.4)            | 109 (68.6)       | 0.97 (0.59–1.58)        | 145 (65.9) | 0.78 (0.51–1.17)        | 0.643 |
| Histology                      |                       |                  |                         |           |                         |       |
| Squamous cell carcinoma        | 271 (53.6)            | 82 (51.6)        | Ref.                    | 112 (50.9) | Ref.                    |       |
| Adenocarcinoma                 | 152 (30.0)            | 35 (22.0)        | 0.61 (0.39–0.93)        | 59 (26.8) | 0.77 (0.54–1.09)        | 0.135 |
| Other†                         | 83 (16.4)             | 42 (26.4)        | 1.35 (0.83–2.22)        | 49 (22.3) | 1.11 (0.72–1.71)        | 0.643 |
| TNM stage                      |                       |                  |                         |           |                         |       |
| I/II                           | 284 (56.1)            | 72 (45.3)        | Ref.                    | 98 (44.5) | Ref.                    | 0.001 |
| III                            | 222 (43.9)            | 87 (54.7)        | 1.94 (1.41–2.68)        | 122 (55.5) | 2.01 (1.53–2.65)        | <0.001 |
| Differentiation                |                       |                  |                         |           |                         |       |
| Well/moderate                  | 347 (68.6)            | 88 (55.3)        | Ref.                    | 129 (58.6) | Ref.                    | 0.007 |
| Poorly/undifferentiated        | 159 (31.4)            | 71 (44.7)        | 1.50 (0.99–2.26)        | 91 (41.4) | 1.61 (1.14–2.28)        |       |
| Adjuvant chemotherapy or radiotherapy | 1.31 (25.9) | 46 (28.9)        | Ref.                    | 54 (24.5) | Ref.                    |       |
| No                             | 375 (74.1)            | 113 (71.1)       | 0.67 (0.47–0.96)        | 166 (75.5) | 0.86 (0.62–1.19)        | 0.365 |

\( ^{\dagger}\text{Adjusted by age, gender, smoking status, histology, TNM stage, differentiation, and adjuvant chemotherapy or radiotherapy where appropriate.} \)

### Table 2. Association between single nucleotide polymorphisms (SNPs) in the EpCAM gene and clinical outcomes in Chinese patients with non-small-cell lung cancer

| SNP locus | Genotype | Overall survival | Recurrence-free survival |
|----------|----------|------------------|--------------------------|
|          |          | Death/total      | HR (95% CI)\text{\( ^{\dagger} \)} | P-value | Death/total | HR (95% CI)\text{\( ^{\dagger} \)} | P-value |
| rs1421 miRNA binding site | AA | 108/329 | Ref. | 144/329 | Ref. |
|        | AG | 44/147 | 1.14 (0.79–1.63) | 0.485 | 67/147 | 1.31 (0.97–1.77) | 0.078 |
|        | GG | 6/18 | 1.33 (0.57–3.07) | 0.508 | 7/18 | 0.81 (0.38–1.75) | 0.596 |
|        | Dominant | 50/165 | 1.16 (0.82–1.64) | 0.404 | 74/165 | 1.24 (0.93–1.65) | 0.151 |
|        | Additive | 1.14 (0.86–1.53) | 0.364 | 1.12 (0.88–1.41) | 0.360 |
| rs1126497 non- synonymous SNP | CC | 93/336 | Ref. | 131/336 | Ref. |
|        | CT | 59/157 | 1.34 (0.96–1.86) | 0.087 | 80/157 | 1.29 (0.97–1.71) | 0.076 |
|        | TT | 5/10 | 3.44 (1.37–8.65) | 0.006 | 6/10 | 2.61 (1.14–6.02) | 0.024 |
|        | Dominant | 64/167 | 1.40 (1.02–1.94) | 0.040 | 86/167 | 1.34 (1.02–1.72) | 0.039 |
|        | Additive | 1.46 (1.08–1.96) | 0.013 | 1.37 (1.06–1.76) | 0.016 |

\( ^{\dagger}\text{Adjusted by age, gender, smoking status, histology, TNM stage, differentiation, and adjuvant chemotherapy or radiotherapy. CI, confidence interval; HR, hazard ratio; Ref., reference.} \)
Associations between EpCAM polymorphisms and prognosis of NSCLC stratified by patient characteristics. To investigate the modulator effects of clinical characteristics on the association between EpCAM polymorphisms of rs1126497 and prognosis, we carried out multivariate Cox proportional hazard analyses in a dominant model stratified by age, gender, smoking status, histology, TNM stage, differentiation, and adjuvant chemotherapy or radiotherapy. As shown in Table 4, the CT + TT genotype was significantly associated with both worse OS and RFS in NSCLC patients with older age, male gender, smoking history, squamous cell type tumor, well or moderately differentiated tumor, and adjuvant chemotherapy or radiotherapy. We observed that patients with the CT + TT genotype had significantly increased risk for death (HR = 1.60; 95% CI, 1.02–2.53) and recurrence (HR = 1.67; 95% CI, 1.11–2.51) in patients with older age. For ever smoker patients, the HRs for death and recurrence were 1.52 (95% CI, 1.04–2.23) and 1.73 (95% CI, 1.24–2.42), respectively. For patients with squamous cell carcinoma, the HRs for death and recurrence were 1.94 (95% CI, 1.24–3.03) and 1.54 (95% CI, 1.04–2.28), respectively. For patients with well or moderately differentiated tumors, the HRs for death and recurrence were 1.84 (95% CI, 1.20–2.81) and 1.58 (95% CI, 1.10–2.26), respectively. For patients who received adjuvant chemotherapy or radiotherapy, the HRs for death and recurrence were 1.52 (95% CI, 1.04–2.21) and 1.57 (95% CI, 1.15–2.15), respectively. Tumor stage, differentiation, and adjuvant treatments are common and important clinical variables in cancer survival analysis. Therefore, we assessed the prognostic values of rs1126497 on OS analysis stratified by these factors. As the results show in Figure 2, compared to patients carrying wild-type (CC) rs1126497, patients with variant genotypes showed significantly shorter survival in the strata of patients

Table 3. Association between single nucleotide polymorphisms (SNPs) in the EpCAM gene and disease progression in Chinese patients with non-small-cell lung cancer

| SNP      | Genotype | T stage | N stage |
|----------|----------|---------|---------|
|          |          | T1/T2 (%) | T3/T4 (%) | OR (95% CI) | P-value | N0 (%) | N1/N2/N3 (%) | OR (95% CI) | P-value |
| rs1421   | AA       | 192 (63.0) | 137 (72.5) | Ref. | 148 (65.5) | 181 (67.5) | Ref. |
|          | AG       | 101 (33.1) | 46 (24.3) | 0.64 (0.42–0.96) | 0.033 | 70 (31.0) | 77 (28.7) | 0.89 (0.61–1.33) | 0.594 |
|          | GG       | 12 (3.9) | 6 (3.2) | 0.70 (0.26–1.91) | 0.488 | 8 (3.5) | 10 (3.7) | 1.02 (0.39–2.66) | 0.964 |
|          | AG+GG    | 113 (37.0) | 52 (27.5) | 0.65 (0.44–0.96) | 0.029 | 78 (34.5) | 87 (32.5) | 0.91 (0.63–1.33) | 0.630 |
| rs1126497| CC       | 215 (69.4) | 121 (62.7) | Ref. | 157 (68.0) | 179 (65.8) | Ref. |
|          | CT       | 90 (29.0) | 67 (34.7) | 1.32 (0.89–1.95) | 0.156 | 72 (31.2) | 85 (31.3) | 1.02 (0.71–1.51) | 0.857 |
|          | TT       | 5 (1.6) | 5 (2.6) | 1.78 (0.50–6.26) | 0.371 | 2 (0.9) | 8 (2.9) | 3.51 (0.73–16.77) | 0.116 |
|          | CT+TT    | 95 (30.6) | 72 (37.3) | 1.35 (0.92–1.97) | 0.123 | 74 (32.0) | 93 (34.2) | 1.10 (0.76–1.60) | 0.609 |

N stage, regional lymph node involvement; Ref., reference; T stage, size and tissue invasion of the primary tumor.

© 2013 The Authors. Cancer Science published by Wiley Publishing Asia Pty Ltd on behalf of Japanese Cancer Association.

Cancer Sci | January 2014 | vol. 105 | no. 1 | 92

www.wileyonlinelibrary.com/journal/cas
with early-stage disease ($P = 0.038$), better differentiated tumors ($P = 0.002$), and those who received adjuvant chemotherapy or radiotherapy ($P = 0.014$).

**Discussion**

In this study, we investigated the effects of two SNPs in the *EpCAM* gene on the survival of NSCLC patients. We found that a variant genotype (CT+TT) of rs1126497 in the *EpCAM* gene was significantly associated with poor prognosis (both OS and RFS) of NSCLC, whereas there was no significant association observed between the SNP rs1421 and the clinical outcomes of NSCLC patients. Further stratified analysis indicated that the effect of rs1126497 had more prominence in older patients, smokers, patients with squamous cell carcinoma, patients with well or moderately differentiated tumor, and patients who received adjuvant chemotherapy or radiotherapy treatment. Moreover, we examined the associations between these two SNPs and NSCLC progression by the logistic regression model and found that a variant genotype of rs1421 had significant association with advanced T-stage disease in NSCLC patients. To the best of our knowledge, this is the first study to investigate the association between *EpCAM* gene polymorphisms and NSCLC prognosis.

Previous studies have indicated that a functional polymorphism (rs1126497) in the *EpCAM* gene was associated with the risk of breast and cervical cancer. One of these studies showed that subjects carrying the CT+TT genotype had a 1.40-fold increased risk for breast cancer than patients with the CC genotype. Another study found that there was a dose–response of cancer risk for patients with the CT genotype (adjusted OR = 1.72) and TT genotype (adjusted OR = 1.96) for cervical cancer. Nonetheless, there is no report about the association between functional polymorphisms in the *EpCAM* gene and cancer prognosis. Our study found that variant genotypes of rs1126497 had a significant association with worse prognosis of NSCLC patients in both OS and RFS analyses, which is in concordance with those previous reports.

As a cell adhesion molecule, EpCAM is involved in cell adhesion and proliferation, which plays a regulative role in human cancer. It has been indicated that EpCAM is overexpressed in various types of cancers, including cancers of the colon and rectum, breast, prostate, bladder, and also frequently in lung cancer. Downregulation of EpCAM by RNAi decreases the proliferation, migration, and invasive capacity of breast cancer cells. In addition, EpCAM knockdown leads to substantial apoptosis in lung cancer cell lines rather than in a bronchial epithelial cell line. Moreover, recent studies have revealed that EpCAM plays a versatile role in carcinogenesis, in that it not only affects cell–cell adhesion but is also involved in many processes including cell migration, differentiation, and metabolism. All these studies suggest that EpCAM play an essential role in the development of cancer, including lung cancer.

The EpCAM molecule consists of three domains: an extracellular domain with epidermal growth factor-like repeat and thyroglobulin repeat-like (TY) domain, a single transmembrane domain, and an intracellular domain (EplCD). The TY domain is capable of selectively discriminating cathepsin L and S, two similar cysteine proteases. It has the potential...
The role of inhibiting the activity of cysteine proteases, and this activity depends on the specific interaction with target protein. The SNP rs1126497 is an ns polymorphism (M115T) located in the TY domain of EpCAM. Thus, the rs1126497 polymorphism could alter the primary structure of the TY domain, which may decrease the interaction between the TY domain and cathepsin L, weaken the inhibiting ability of cathepsin L, and thus eventually affect NSCLC patients’ prognosis. More recently, many studies have focused on the EpCAM intracellular signaling pathway, in which EpCAM is cleaved by tumor necrosis factor-α converting enzyme (TACE/ADAM17) and γ-secretase and the ectodomain is released into the medium and EpICD into the cytoplasm. The released EpICD as a signal transducer is transported into the nucleus and forms a nuclear complex with FHL2, β-catenin, and Lef-1 to stimulate expression of target gene c-myc. This result supports the proliferative signaling activity of overexpressed EpCAM.

In the stratified analysis, we observed that the significant associations between SNP rs1126497 and OS were only evident in NSCLC patients with better clinical parameters, such as early stage disease and well or moderately differentiated tumors, and in those receiving adjuvant chemotherapy or radiotherapy (Fig. 2). To date, the detailed mechanism underlying the effect of SNP rs1126497 on cancer progression is largely unknown. However, this SNP has been reported to be significantly associated with risk of breast and cervical cancers in Chinese populations. These data suggest that this ns-SNP may play an important role in cancers at early stage. Moreover, it is generally accepted that most SNPs only have subtle effects on complex traits, such as cancers. This may partially explain why rs1126497 did not show a significant effect on prognosis in patients with worse clinical parameters, as the worse the
cancer, the more complex and varied the disease. Additionally, a previous in vitro study suggested that EpCAM is a promising target to sensitize tumor cells to chemotherapy, which is in concordance with our finding that the effect conferred by this rsNSNP on prognosis was only evident in patients receiving adjuvant treatment. Taken together, it is reasonable to find that the effect of SNP rs1126497 on prognosis is more prominent in those patients with better clinical parameters and receiving adjuvant treatments.

Additionally, although our data did not show any association between rs1421 and NSCLC prognosis, we found that the rs1421 AG+GG genotypes were significantly associated with early T stage disease. This could be due to the location of rs1421 in mRNA binding sites of the EpCAM 3'-UTR region that may affect expression of EpCAM. Further study is needed to understand the effects of the rs1421 polymorphism on tumor progression.

The biggest advantage of our study is that all subjects’ characteristics were homogeneous, and they were treated using similar procedures, which minimized the confounding effects. Nevertheless, there are limitations to our study. First, the follow-up time was relatively short. However, this limitation can be partially compensated by the high recurrence and death rate in our study population. Second, the relatively small sample size limited the analysis on the recessive model and caused an unreliable estimation for the stratified analyses. A larger number of patients will be needed to assess the association between the SNP rs1126497 at EpCAM and NSCLC survival in future studies. Finally, functional assays are warranted to establish the underlying mechanism.

In summary, our study indicates for the first time that the SNP rs1126497 in the EpCAM gene is significantly associated with the prognosis of NSCLC. Our findings contribute to current understanding of the effects of EpCAM polymorphisms on NSCLC.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011; 61: 69–90.
2 Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005; 55: 74–108.
3 Sekido Y, Fong KM, Minna JD. Molecular genetics of lung cancer. Annu Rev Med 2003; 54: 73–87.
4 Cho J. The International Association for the Study of Lung Cancer—the Lung Cancer Staging Project: better data, better decisions, better outcomes. Hawaii Med J 2006; 67: 220–2.
5 Mulshine JL, Sullivan DC. Clinical practice. Lung cancer screening. N Engl J Med 2005; 352: 2714–20.
6 Chong JM, Speicher DW. Determination of disulfide bond assignments and N-glycosylation sites of the human gastrointestinal carcinoma antigen GA733-2 (CO17-1A, EGP, KSI-4, KSA, and Ep-CAM). J Biol Chem 2001; 276: 5804–13.
7 Munz M, Baeuerle PA, Gires O. The emerging role of EpCAM in cancer and stem cell signaling. Cancer Res 2009; 69: 5627–9.
8 van der Graaf MT, Musters RJ, Rutiers MH, de Leij LF, McLaughlin PM, Rots MG. EpCAM in carcinogenesis: the good, the bad or the ugly. Carcinogenesis 2010; 31: 1913–21.
9 Kimura H, Kato H, Faried A. Epithelial cell adhesion molecule (EpCam) in renal epithelial tumors. Gan To Kagaku Ryoho 2007; 34: 171–9.
10 Went P, Dinnhof S, Salvisberg T et al. Expression of epithelial cell adhesion molecule (EpCam) in renal epithelial tumors. Am J Surg Pathol 2005; 29: 83–8.
11 Brunner A, Prelog M, Verdorfer I, Tzankov A, Mikuz G, Ensinger C. EpCAM is predominantly expressed in high grade and advanced stage urothelial carcinoma of the bladder. J Clin Pathol 2008; 61: 307–10.
12 Pak MG, Shin DH, Lee CH, Lee MK. Significance of EpCAM and TROP2 expression in non-small cell lung cancer. World J Oncol 2012; 10: 753–9.
13 Went P, Vasei M, Bubendorf L et al. Frequent high-level expression of the immunotherapeutic target EpCam in colon, stomach, prostate and lung cancers. Br J Cancer 2006; 94: 128–35.
14 Ni J, Cozzi PJ, Duan W et al. Role of the EpCam (CD326) in prostate cancer metastasis and progression. Cancer Metastasis Rev 2012; 31: 799–91.
15 Munz M, Kieu C, Mack B, Schmitt B, Zeidler R, Gires O. The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. Oncogene 2004; 23: 5748–58.
16 Seligson DB, Pantuck AJ, Liu X et al. Epithelial cell adhesion molecule (KSA) expression: pathobiology and its role as an independent predictor of survival in renal cell carcinoma. Clin Cancer Res 2004; 10: 2659–69.
33 Guncar G, Pungercic G, Klemencic I, Turk V, Turk D. Crystal structure of MHC class II-associated p41 II fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J 1999; 18: 793–803.

34 Maetzel D, Denzel S, Mack B et al. Nuclear signalling by tumour-associated antigen EpCAM. Nat Cell Biol 2009; 11: 162–71.

35 Carpenter G, Red Brewer M. EpCAM: another surface-to-nucleus missile. Cancer Cell 2009; 15: 165–6.

36 Patriarca C, Macchi RM, Marschner AK, Mellstedt H. Epithelial cell adhesion molecule expression (CD326) in cancer: a short review. Cancer Treat Rev 2011; 38: 68–75.

37 Hussain S, Pluckthun A, Allen TM, Zangemeister-Wittke U. Chemosensitization of carcinoma cells using epithelial cell adhesion molecule-targeted liposomal antisense against bcl-2/bcl-xL. Mol Cancer Ther 2006; 5: 3170–80.