Identification of a Functional SNP rs17079281 at 6q22.2 Locus That Is Associated with Lung Cancer Risk

Yu Wang[1], Ben Liu[1], Jinyu Kong[1], Jingxin Li[1], Rongna Ma[2], Ming Gao[1], Herbert Yu[3], Biyun Qian[2*]

Affiliations

1. Department of Cancer Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Tianjin 300060, China.

2. Hongqiao International Institute of Medicine, Shanghai Tongren Hospital and Faculty of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China.

3. Epidemiology Program, University of Hawaii Cancer Center, 701 Ilalo Street, Honolulu, HI 96813, USA.

*** Corresponding author:**

Biyun Qian, MD, PhD. Professor.

Hongqiao International Institute of Medicine, Shanghai Tongren Hospital/Faculty of Public Health, Shanghai Jiao Tong University School of Medicine, 227 South Chongqing Road, Shanghai 200025. Email: qianbiyun@shsmu.edu.cn

Abstract

Genome-wide association studies (GWAS) have identified numerous genetic polymorphisms that are associated with cancer risk, but their biological relevance to the disease is not known for many of them. The study performed linkage disequilibrium (LD) analysis on a GWAS-discovered SNP rs9387478 and identified four potentially functional SNPs (rs17079281, rs6911915, rs9320604 and rs4946259) in *DCBLD1* with high LD. Associations of these SNPs with lung cancer were examined in 766 Chinese cases and 773 matched controls. The results suggested that lung cancer risk was associated with SNP rs17079281. Genotype C/T and T/T had lower risk compared with genotype C/C (OR=0.78, 95% CI=0.63-0.98). Luciferase assays demonstrated that YY1 had higher binding–affinity to the T alleles of rs17079281 than C alleles, and the binding was associated with reduced transcription of *DCBLD1*. We further observed a trend showing a decrease in *DCBLD1* expression in the C/T and T/T carriers compared to C/C carriers. DCBLD1 knockdown inhibited lung cancer cell migration and invasion. Our findings suggest that rs17079281 (C>T) may affect the binding affinity of YY1 to *DCBLD1* and influence the risk of lung cancer.

Introduction

Lung cancer is one of the most common malignant tumors worldwide. In 2012, there were 1.8 million new cases and 1.6 million deaths[1]. In China, around 600,000 patients die from lung cancer each year, and lung cancer mortality has been more than quadrupled over the past 3 decades[2]. The development of lung cancer is known to be multifactorial. Tobacco smoking and indoor/outdoor air pollutions are major risk factors of the disease[3,4]. Genetic factors may also play a role in disease susceptibility as well as in the susceptibility to lung cancer risk factors, such as addiction to tobacco use. Although enormous efforts have been made to understand the etiology of lung cancer, the exact mechanisms of lung carcinogenesis remain to be elucidated.

The advent of GWAS (genome-wide association study) provides new avenues to investigate the genetic basis of disease susceptibility[5,6]. GWAS examines common SNPs known as "tag SNPs" to identify their associations with disease risk[7]. Recently, several GWAS reports have been published suggesting that a number of SNPs in chromosome 8q24, 3q28, 5q15.33, 3q12.12 , 13q12.12 and 22q12.2 are associated with lung cancer risk in Han Chinese[8-10]. Since majority of the SNPs (~93%) reported lie within the non-protein coding regions[11,12], we face a huge challenge to understand the biological relevance of these genetic variants to cancer risk. Thus, it is critical for us to identify causal variants from those tag SNPs and characterize their biological implications in carcinogenesis. One of the most frequently used strategies is to seek the variants which have high linkage disequilibrium (LD) with the SNPs discovered by GWAS[13].

In a previous GWAS [14], SNP rs9387478 in 6q22.2 was found to be associated with lung cancer risk. SNP rs9387478 is situated between two genes, *DCBLD1* and *ROS1*, which encode the discoidin, CUB and LCCL domain containing 1 protein and ROS proto-oncogene receptor tyrosine kinase, respectively. Both proteins are involved in regulation of cell proliferation and possibly invasion [15,16]. To search for potential hidden causal SNPs in the 6q22.2 region, we performed LD analysis to look for functional SNPs having high LD with rs9387478. For the SNPs identified to be in high LD with rs9387478, we analyzed their genotypes in 766 lung cancer patients and 773 matched control subjects to determine the associations of genotypes with lung cancer risk. We also conducted in vitro experiments to assess their potential biological actions.

Results

Characteristics of study subjects

Demographic features and risk factors of lung cancer patients and their matched controls in our study were presented in Table 1. Age and gender were not substantially different between patients and controls, suggesting adequate matching on these factors. No significant differences were found between cases and controls in their history of lung diseases (p=0.710) and family history of cancer (p=0.072). As expected, smoking status was associated with the risk of lung cancer (OR=1.66; 95%*CI*: 1.15-1.84; p<0.001); a dose-response relationship was also noticed between the disease risk and pack-year of smoking. In addition, cases had lower BMI than controls (p<0.001).

DCBLD1 genotypes and lung cancer risk

A total of 14 SNPs were found in high LD with rs9387478, and of them four SNPs were in the *DCBLD1* gene (Table 2). Table 3 shows the genotypes of the *DCBLD1* SNPs in cases and controls and their associations with lung cancer (ORs and 95%CI) after adjustment for age, gender, smoking status, BMI and family history of cancer. Two SNPs (rs17079281, rs6911915) were in Hardy-Weinberg equilibrium (p>0.05), and two (rs4946259, rs9320604) were not.

We found that the distribution of rs17079281 genotype was different between cases and controls, and controls had more C/T heterozygous than patients (p=0.039).

Compared to individuals carrying the wild C/C genotype, those with C/T heterozygote had lower risk for lung cancer (adjusted OR=0.74; 95%CI: 0.58-0.94). Under the dominant model, those with the T allele had an adjusted OR=0.78 (95%CI: 0.63-0.98).

For rs4946259 (G>A), we found no statistically significant associations of genotypes or genetic models with lung cancer risk. Similar results were observed for rs9320604 (G>A). An association, however, was noticed for SNP rs6911915 (T>C). Individuals with the T/C genotype had lower risk of lung cancer compared to those with the T/T genotype.

Effects of rs17079281 on transcription activity

SNP rs17079281 is located in the promoter of *DCBLC1*, and the SNP is situated on a transcription factor YY1 binding site. Two luciferase reporter plasmids containing rs17079281 C or T allele were constructed (Figure 1A). To investigate the effect of transcription factor YY1 on transactivation of different rs17079281 alleles and determine whether YY1 was responsible for the T allele-related decrease in the *DCBLD1* promoter activity, A549, H1299 and HEK293T were cotransfected with a YY1-expression vector (GV144-YY1) and either pGL3-CC or pGL3-TT plasmid. As predicted, YY1-mediated transactivation was significantly lower in A549 and H1299 cells co-transfected with pGL3-TT plasmid than with pGL3-CC plasmid (Figure 1C&1D). In addition, as shown in Figure 1B, when either pGL3-CC or pGL3-TT

plasmid was cotransfected with control vector (GV144), pGL3-TT plasmid had significantly lower luciferase expression than pGL3-CC plasmid in HEK293T cell, but not in A549 and H1299 cells (Figure 1C&1D). These results support the hypothesis that YY1 exerts allele-specific influences on *DCBLD1* expression through its higher binding-affinity to the T-allele promoter, resulting in reduced transcription activity of *DCBLD1*.

DCBLD1 expression by rs17079281 genotypes

To test whether this differential binding could modulate *DCBLD1* expression, we measured *DCBLD1* mRNA in cancer tissue from 198 patients using quantitative PCR. We found that patients with homozygous T/T genotype had lower *DCBLD1* expression than those with C/C genotype. In addition, individuals with T allele also had lower *DCBLD1* mRNA levels than C/C carriers. However, these differences did not reach statistical significance (Figure 2A&B).

DCBLD1 knockdown and tumor cell migration and invasion

We performed wound healing and transwell assays to assess whether DCBLD1 downregulation influences the migration and invasion capacities of lung cancer cells. To further confirm the effectiveness of *DCBLD1* knockdown by siRNA, we analyzed protein level of DCBLD1 by western blot in A549 and H1299. The results showed that DCBLD1 expression reduced by 80% (Figure 3A). Figure 3B showed the results of wound closure experiments at 48 hours. Compared with the cells transfected with

control siRNA where wound closure reached by 90%, wound closure in *DCBLD1* knockdown cells was only 50%. Consistently, *DCBLD1* knockdown suppressed the number of migrating cells compared with control cells (Figure 3C). As shown in Figure 3D, transwell assay indicated more than 2-fold decreases in cell invasion ($p<0.0001$) after downregulation of *DCBLD1* expression. These data suggested that silencing DCBLD1 inhibited migration and invasion of lung cancer cells.

Discussion

In the study, we analyzed four putative functional SNPs which were in high LD with rs9387478, a SNP discovered by GWAS to be associated with lung cancer risk. We found that one of the four SNPs, rs17079281 in the *DCBLD1* promoter, was associated with lung cancer. We also demonstrated that SNP rs17079281 had an impact on DCBLD1 expression, and low expression of DCBLD1 was associated with less aggressive tumor behaviors.

In our genotype analysis, individuals with rs17079281 heterozygous C/T genotype had a significantly lower risk for lung cancer compared to those with the C/C genotype, but no risk difference was observed between individuals with homozygous T/T and C/C genotypes. A possible explanation for this observation could be heterozygous advantage, in which heterozygous genotype had better selection advantage than either homozygous genotype. This phenomenon was also observed for a polymorphism in a sickle cell gene which affects the morbidity and mortality of malaria[17].

SNP rs17079281 is located in the promoter region of the *DCBLD1* gene, 682 bp upstream of the transcription start site (TSS). Bioinformatic analysis suggests that the C-to-T polymorphism is situated in a transcription factor binding site to which transcription factor YY1 binds. In addition, studies have indicated that CCAT and ACAT are two types of core sequences that possess high binding-affinity to YY1 in gene promoters [18]. Interestingly, the sequence surrounding the T allele of rs17079281 matches to the sequence of CCA**T**. Our *in vitro* experiment showed that when control

vector (no YY1) was cotransfected into tumor cells with *DCBLD1* promoter, there was no significant difference in luciferase expression between PGL3-TT and PGL3-CC plasmid. This may be explained by low expression of YY1 in A549 and H1299. However, when cells were cotransfected with YY1 and *DCBLD1* promoter, luciferase activities were significantly different by the rs17079281 genotype. T alleles had much lower expression than C alleles, suggesting that transcription factor YY1 could depress *DCBLD1* expression by interact with the rs17079281 T alleles more strongly than the C alleles in the *DCBLD1* promoter. YY1 has been implicated in the transcriptional repression of gene expression because it regulates histone deacetylation and H3-K27 methylation[19]. YY1 possesses a strong repression domain at C-term which consists of four GLI-Krüppel type zinc fingers [20]. Interestingly, D-Limonene, listed as the food additives by Code of Federal Regulation, mediated chemoprevention of hepatocarcinogenesis in AKR mice by increasing YY1 protein level[21]. Transcription factor YY1 had different effect on the alleles of SNP rs17079281, which provided clues to the study of individualized prevention.

Expression quantitative trait loci (eQTLs) are polymorphisms within the regulatory regions that influence gene expression. Multiple studies have shown that SNPs in the regulatory regions alter gene expression in a tissue-specific manner [22,23]. To further elucidate the effect of SNP rs17079281 on transcription activity of its target gene *DCBLD1*, we analyzed *DCBLD1* genotype and phenotype in lung tumor samples to perform eQTLs analysis. However, no association was found between SNP rs17079281 and *DCBLD1* mRNA levels. This finding did not exclude the possibility

that there were other mechanisms which might affect gene expression in tumor samples, such as copy number variation and methylation changes[24]. Whether these mechanisms exist in the *DCBLD1* gene is currently unknown. Furthermore, the effect of a single SNP on gene expression tends to be small, and other subtle influences on gene expression may mask the SNP's impact. Lastly, *DCBLD1* expression was measured at one time point, which may not reflect the overall situation of gene expression. It is known that gene expression varies by time and location, and an eQTL-target gene relationship may only exist during a particular time point or location. A tendency between low *DCBLD1* mRNA and T alleles was observed in our study, and it was in accordance with the results of luciferase assay.

The *DCBLD1* gene is located at 6q22 with 151 kp in length. There are 214 SNPs in the region (based on dbSNP database), of which only one SNP rs17574269 located in intron 1 was reported to be associated with overall survival (OS) of small-cell lung cancer (SCLC)[25]. Few studies have explored the biological function of DCBLD1 in lung cancer. Our study indicated that downregulation of DCBLD1 expression suppressed lung cancer cell migration and invasion. Interestingly, studies on a related gene *DCBLD2* (encoding discoidin, CUB and LCCL domain containing 2, also named *CLCP1*) showed that the expression of this molecule was up-regulated in highly metastatic lung cancer cell lines NCI-H460-LNM35 and lung cancer specimens. Down-regulation of DCBLD2 suppressed cell motility [16,26]. In addition, overexpression of YY1 in lung cancer cells inhibited DCBLD1 expression. Wang et al. reported that transcription factor YY1 upregulated invasion suppressor HLJ1

expression and inhibited cancer cell invasion in lung cancer cells [27,28].

In summary, in search for potential functional SNPs in high LD with rs9387478 identified by lung cancer GWAS[14], we found SNP rs17079281, located in the promoter region of DCBLD1, to be associated with lung cancer risk. This SNP appears to affect the promoter activity of DCBLD1 by influencing the binding affinity of transcription factor YY1. Our experiments indicated that downregulation of DCBLD1 suppressed lung cancer cell migration and invasion. More studies are needed to understand the biological mechanisms of this SNP in connection to lung cancer. The associations of rs17079281 with lung cancer both individually and in combination with other risk alleles are also required to be further confirmed in large epidemiological studies.

Materials and methods

Study subjects

The study included 766 cases and 773 controls. The patients in the study were recruited from Tianjin Medical University Cancer Hospital (TMUCH) between January 2006 and January 2013. All patients were diagnosed with histologically confirmed primary non-small cell lung cancer (NSCLC). The exclusion criteria included a previous history of cancer or radio/chemotherapy. Patient information on tumor histology, disease stage, treatment, and body weight and height was obtained from their medical charts and pathology reports. Healthy individuals who underwent regular health checkup during the same time as the patients being recruited were enrolled in the study as controls. The control subjects were matched to cases on gender and age (±5 years).

All study participants were interviewed by trained research staff using a structured questionnaire which elicits information on demographic features, medical history of lung diseases, family history of cancer, status of tobacco smoking, years of smoking, and number of cigarettes smoked per day. Individuals who had smoked at least one cigarette per day for more than 12 months in lifetime were defined as smokers. Pack-years was calculated by multiplying the number of packs smoked daily with number of years smoking. Family history of cancer was defined as self-reported cancer in first-degree relatives.

The study was approved by the medical ethics committee at TMUCH and carried out in accordance to the approved guidelines. Each study subject signed an informed

consent before being enrolled in the study and providing 20 ml blood samples for research. Tumor specimens were collected from 198 patients. The study subjects were unrelated Han ethnic Chinese.

SNP selection and genotyping

Our search for candidate causal SNPs was centered on SNP rs9387478 which was found by GWAS to be associated with lung cancer risk. We conducted a LD-based search of the HapMap database for SNPs that are located in 100 kb up and downstream of rs9387478 in Han Chinese using the Haploview softer 4.2. The SNPs were selected on the basis of following criteria: (i) located in the same Han Chinese haplotype block with SNP rs9387478; (ii) in high LD with rs9387478 (D'>0.8); (iii) located in the coding or regulatory regions of the two nearby gene, *DCBLD1* and *ROS1*; and (iv) had MAF>0.05. Four SNPs in *DCBLD1* met these selection criteria, including rs17079281 in the promoter region, and rs6911915, rs9320604 and rs4946259 in the first intron.

Genomic DNA was extracted from peripheral blood using the TKM method and genotyped with the TaqMan assay using the 7900HT Fast Real-time PCR System. The TaqMan assays targeting on the 4 SNPs were designed and ordered from ABI (Applied Biosystems Inc., US), and the PCR assays were performed according to the manufacturer's instruction. Five percent of the samples were randomly selected for retesting, and the results were in complete concordance.

Constructions of luciferase reporter gene plasmids

Our bioinformatic analysis using TRANSFAC® (http://www.gene-regulation.com/pub/databases.html) and ALGGEN PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi/DirDB=TF_8.3) indicated that SNP rs17079281, a C/T polymorphism, might affect the binding affinity of the transcription factor YY1 to the *DCBLD1* promoter (Figure 1A）. To determine whether this polymorphism has an impact on the transcriptional activity of *DCBLD1*, we constructed two luciferase reporter plasmids containing either C or T allele of rs17079281. DNA fragments with 1070 bp bearing the *DCBLD1* promoter region (nucleotides -959 to +110，relative to the transcription start site) were amplified by PCR using the genomic DNA isolated from control subjects who carry either homozygous C/C or T/T genotype. The PCR primers were designed as follows: forward with *KPN I* site, 5'-GGGGTACCCCTCCCCAAACCCTTCTCCGC-3' and reverse with *Bgl II* site, 5'-GAAGATCTCAGCTTGGCAAGCTCGGCCT-3'. PCR conditions included initial denaturing at 95^{o}C for 10 min and 35 cycles at 94^{o}C for 40 sec, 64.9^{o}C for 30 sec and 72^{o}C for 1 min and 20 sec, followed by elongation at 72^{o}C for 10 min. The PCR products between the *KPN I* and *Bgl II* sites were cloned into the pGL3-Basic vector (Promega, US) containing the firefly luciferase gene as a reporter. The recombined vector containing the expected C or T allele of rs17079281 were verified by direct sequencing. The resultant plasmid containing C allele of rs17079281 was designated as pGL3-CC; the resultant plasmid, named pGL3-TT, was confirmed to contain T allele at rs17079281. To elucidate the effect of transcription

factor YY1 on transactivation, an overexpression vector (GV144-YY1) and its negative control were purchased from Genechem (Shanghai, China).

Cell lines culture

Human NSCLC cell lines A549 and H1299 and embryonic kidney cell line HEK293T were selected for *in vitro* experiments. These cells were cultured in RPMI1640 medium, containing 10% of fetal bovine serum (FBS). All cell lines were incubated at 37^{o}C with 5% CO_2.

Luciferase reporter assays

A549, H1299 and HEK293T were seeded at 5×10^4 per well in 24-well plates, and allowed to attach for 24 hours. The cells were transfected with 0.4ug of each reporter constructs (pGL3-Basic, pGL3-CC, pGL3-TT) and 0.4 □g of either GV144-YY1 expression plasmid or empty GV144 vector using Lipofectamine 3000 (Invitrogen, US) according to the manufacturer's instructions. PRL-TK (8 ng) (Promega, US) containing *Renilla* luciferase gene was co-transfected to standardize transfection efficiency. The relative luciferase activity was determined at 24 hours after transfection using Dual Luciferase Reporter Assay System (Promega, US). The experiments were performed with triplicate.

siRNA transfections

SMART pool siRNAs targeting the *DCBLD1* gene and non-targeting siRNA

pools (Dharmacon, US) were transfected into lung cancer cell A549 and H1299 according to the RNAiMAX protocol (Invitrogen, US). Briefly, 5 □l RNAiMAX transfection reagent was mixed with final siRNA concentration of 50 □M. The mixture was incubated for 5 minute at room temperature and then was dispersed into each well of a six-well plate. Transfected cells were collected to perform the following assay.

Wound-healing assay

A549 and H1299 cells were seeded into 6-well plates to complete confluence overnight. The monolayer cells were scratched with a sterile 10 □l pipette tip. Detached cells were removed by washing with PBS. Then, the medium was replaced with 2 ml of fresh medium. The migrated distance of cells was monitored under a microscope. Photographs were taken at 0 and 48 hours after transfection. Three independent experiments were performed.

Transwell assay

Cell migration and invasion were performed using an 8mm pore size insert (BD, US). Lung cancer cells were seeded in serum-free medium at a density of 2×10^4 cells in normal upper chamber for migration assay. For invasion assay, 4×10^4 cells were seeded in matrigel-coat (BD, US) chamber containing serum-free medium. The bottom chamber was filled with medium containing 10% FBS. After incubating for 24 hours at 37^oC, cells were washed with PBS. Cells in the upper chamber were removed

with a cotton swab and cells on the bottom surface of the membrane were fixed with 4% Paraformaldehyde and stained with Rapid Wright-Giemsa Staining Solution (Sangon Biotech, China). Numbers of migrated cells were counted in at least three random fields (×200) per filter.

Western blot

Cells were lysed for 30 minutes on ice in 1×RIPA Lysis Buffer supplemented with a protease inhibitor cocktail (Millipore, US). Denatured proteins (40 □g) were electrophoresed on 10% SDS-PAGE gel and were electro transferred from the gel to PDVF membranes (Millipore, US). After blocking with 5% no-fat milk in Tris Buffered saline with 0.05% Tween-20 for 1 hour at room temperature, the membranes were incubated overnight at 4℃ with primary antibodies against *DCBLD1* (Abcam, US) at a dilution of 1:500, β-actin 1:1000 (Sata, US). Secondary antibodies were added at concentrations of 1:4000. Enhance chemiluminesence (ECL) (Millipore, US) was used to visualize protein expression.

Detection of DCBLD1 by quantitative RT-PCR

Total RNA was extracted from lung tumor samples using the TRizol reagents. RNA (5ug) was then reverse transcribed to complementary DNA using oligo (dT) and MLV (Invitrogen, US). The cDNA level was relatively quantified in the 7900HT Fast Real-time PCR system using the TaqMan gene expression assay according to the manufacturer's protocol. Levels of β-actin expression were measured as internal

reference, and each sample was tested in triplicate. Log($2^{-\Delta\Delta Ct}$) was calculated as the level of mRNA expression.

Statistical analysis

Chi-square test was used to analyze the differences in distributions of demographic variables, risk factors and genotypes between cases and controls. Hardy-Weinberg equilibrium was calculated in the control subjects using the goodness-of-fit chi-square test to compare the expected genotypes with the observed ones. To estimate the associations between SNPs and lung cancer risk in additive, dominate and recessive models, unconditional logistic regression models with adjustment for age, gender, smoking status, BMI and family history of cancer was used to compute odds ratios (ORs) and their 95% confidence intervals (CIs). The P_{trend} was calculated by defining genotypes in the model as a continuous ordinal variable [29]. Student's t test was employed to assess the differences in levels of luciferase reporter gene expression. DCBLD1 mRNA expression differences among rs17079281 genotype were assessed by one-way ANOVA. All statistical tests were two-sided, and SAS software version 9.1 (SAS, US) was used for statistical analysis. P value less than or equal to 0.05 was selected as the significant level.

References

1 Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer. Journal international du cancer* **136**, E359-386, doi:10.1002/ijc.29210 (2015).

2 She, J., Yang, P., Hong, Q. & Bai, C. Lung cancer in China: challenges and interventions. *Chest* **143**, 1117-1126, doi:10.1378/chest.11-2948 (2013).

3 Mu, L. *et al.* Indoor air pollution and risk of lung cancer among Chinese female non-smokers. *Cancer causes & control : CCC* **24**, 439-450, doi:10.1007/s10552-012-0130-8 (2013).

4 Loomis, D., Huang, W. & Chen, G. The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: focus on China. *Chinese journal of cancer* **33**, 189-196, doi:10.5732/cjc.014.10028 (2014).

5 Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728-731, doi:10.1038/nature07631 (2008).

6 McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews. Genetics* **9**, 356-369, doi:10.1038/nrg2344 (2008).

7 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature reviews. Genetics* **6**, 95-108, doi:10.1038/nrg1521 (2005).

8 Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nature genetics* **43**, 792-796, doi:10.1038/ng.875 (2011).

9 Liu, S. G. *et al.* Association of genetic polymorphisms in TERT-CLPTM1L with lung cancer in a Chinese population. *Genetics and molecular research : GMR* **14**, 4469-4476, doi:10.4238/2015.May.4.4 (2015).

10 Zhang, X. *et al.* Polymorphisms on 8q24 are associated with lung cancer risk and survival in Han Chinese. *PloS one* **7**, e41930, doi:10.1371/journal.pone.0041930 (2012).

11 Wang, K. *et al.* Interpretation of association signals and identification of causal variants from genome-wide association studies. *American journal of human genetics* **86**, 730-742, doi:10.1016/j.ajhg.2010.04.003 (2010).

12 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195, doi:10.1126/science.1222794 (2012).

13 Zhu, Q. *et al.* Prioritizing genetic variants for causality on the basis of preferential linkage disequilibrium. *American journal of human genetics* **91**, 422-434, doi:10.1016/j.ajhg.2012.07.010 (2012).

14 Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nature genetics* **44**, 1330-1335, doi:10.1038/ng.2456 (2012).

15 Rimkunas, V. M. *et al.* Analysis of receptor tyrosine kinase ROS1-positive tumors in non-small cell lung cancer: identification of a FIG-ROS1 fusion. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 4449-4457, doi:10.1158/1078-0432.CCR-11-3351 (2012).

16 Koshikawa, K. *et al.* Significant up-regulation of a novel gene, CLCP1, in a highly

metastatic lung cancer subline as well as in lung cancers in vivo. *Oncogene* **21**, 2822-2828, doi:10.1038/sj.onc.1205405 (2002).

17 Aidoo, M. *et al.* Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet* **359**, 1311-1312, doi:10.1016/S0140-6736(02)08273-9 (2002).

18 Yant, S. R. *et al.* High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. *Nucleic acids research* **23**, 4353-4362 (1995).

19 Zhang, Q., Stovall, D. B., Inoue, K. & Sui, G. The oncogenic role of Yin Yang 1. *Critical reviews in oncogenesis* **16**, 163-197 (2011).

20 Galvin, K. M. & Shi, Y. Multiple mechanisms of transcriptional repression by YY1. *Molecular and cellular biology* **17**, 3723-3732 (1997).

21 Parija, T. & Das, B. R. Involvement of YY1 and its correlation with c-myc in NDEA induced hepatocarcinogenesis, its prevention by d-limonene. *Molecular biology reports* **30**, 41-46 (2003).

22 Yang, L. *et al.* A functional polymorphism at microRNA-629-binding site in the 3'-untranslated region of NBS1 gene confers an increased risk of lung cancer in Southern and Eastern Chinese population. *Carcinogenesis* **33**, 338-347, doi:10.1093/carcin/bgr272 (2012).

23 Fu, J. *et al.* Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS genetics* **8**, e1002431, doi:10.1371/journal.pgen.1002431 (2012).

24 Monteiro, A. N. & Freedman, M. L. Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. *Journal of internal medicine* **274**, 414-424, doi:10.1111/joim.12085 (2013).

25 Han, J. Y. *et al.* A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy. *The pharmacogenomics journal* **14**, 20-27, doi:10.1038/tpj.2013.7 (2014).

26 Nagai, H. *et al.* CLCP1 interacts with semaphorin 4B and regulates motility of lung cancer cells. *Oncogene* **26**, 4025-4031, doi:10.1038/sj.onc.1210183 (2007).

27 Wang, C. C. *et al.* The transcriptional factor YY1 upregulates the novel invasion suppressor HLJ1 expression and inhibits cancer cell invasion. *Oncogene* **24**, 4081-4093, doi:10.1038/sj.onc.1208573 (2005).

28 Wang, C. C. *et al.* Synergistic activation of the tumor suppressor, HLJ1, by the transcription factors YY1 and activator protein 1. *Cancer research* **67**, 4816-4826, doi:10.1158/0008-5472.CAN-07-0504 (2007).

29 Ryan, B. M. *et al.* Identification of a functional SNP in the 3'UTR of CXCR2 that is associated with reduced risk of lung cancer. *Cancer research* **75**, 566-575, doi:10.1158/0008-5472.CAN-14-2101 (2015).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.81573231), Natural Science Foundation of Shanghai (Grant No.15ZR1424400) and Translational Medicine Foundation of Shanghai Jiao Tong University School of Medicine (Grant No.15ZH1001)

Author Contributions

B.Q. conceived and designed the study; Y.W and B.L. performed the experiments; J.K., J.L. and R.M. analyzed the data and participated in the discussion; M.G collected subjects and clinical data; B.Q., Y.W., and H.Y. wrote and revised the manuscript. All authors reviewed the manuscript.

Additional Information

Competing financial interest: The authors indicate no potential conflict of interest

Figure1. Luciferase expression assays with constructs containing *DCBLD1* promoter in different cell lines. pRL-TK was cotransfected to standardize transfection efficiency. (A) Schematic of the rs17079281 position relative to the transcription start site (TSS) and illustration of reporter constructs containing rs17079281 C or T allele. Bioinformatic analysis showed that C-to-T mutation created a binding site for a transcription factor. Luciferase reporter plasmids were cotransfected with control vector or YY1 overexpression plasmids in HEK293T (B), A549(C) and H1299 (D) cells. The values represent fold changes of luciferase activities relative to those of cells cotransfected with pGL3-basic. The experiments were carried out in triplicates in three independent transfection experiments. *P<0.05；**P<0.01, in comparison to pGL3-CC construct.

Figure2. Analysis of *DCBLD1* expression in lung tumor samples by *DCBLD1* genotype at rs17079281. (A) No statistically significant differences in *DCBLD1* expression between rs17079281 CC, CT and TT genotypes. (B) *DCBLD1* expression was lower in subjects either with rs17079281 CT or TT genotypes compared to those with the CC genotypes, but the differences were not statistically significant. All values were normalized to *GAPDH* and expressed as means±SD of three independent experiments.

Figure3. Down-regulation of *DCBLD1* expression in lung cancer cell lines suppressed cell migration and invasion. (A) Western blot showed that protein levels of DCBLD1

were decreased in lung cancer cell lines after transfected with *DCBLD1* siRNA. β-actin was used as an internal control. (B) In the wound-healing assay, A549 and H1299 transfected with negative control siRNA or *DCBLD1* siRNA were seeded in 6-well plates, and a scratched wound was applied after 24 hours of confluence. Transfection of *DCBLD1* siRNA could inhibit cell migration. A549 and H1299 transfected with control siRNA or *DCBLD1* siRNA were seeded in the upper chambers without(C) or with (D)matrigel. After 24 hours of incubation, migration/invasion cells were stained with Giemsa and counted from three independent experiments. Cells transfected with *DCBLD1* siRNA had reduced ability to invade compared to those transfected with control siRNA. *$P<0.05$.

Table 1. Characteristics of lung cancer patients and healthy controls

Variables	**N（%）**		***P*[a]**	***OR*（95%*CI*）**
	Control（773）	**Case（766）**		
Age at diagnosis			0.350	
<60	404(52.26)	382(49.87)		1.00
≥60	369(47.74)	384(50.13)		1.05（0.94-1.16）
Gender			0.510	
Male	466（60.28)	468(61.90)		1.00
Female	307(39.72)	288(38.10)		0.96(0.87-1.07)
Lung disease history			0.710	
No	550(90.02)	685(90.61)		1.00
Yes	61(9.98)	71（9.39）		0.90（0.67-1.30）
Family history of cancer			0.072	
No	671(87.83)	634(84.65)		1.00
Yes	93(12.17)	115(15.35)		1.26(0.97-1.62)
Smoking status			**<0.001**	
No	475(61.45)	269(35.77)		1.00
Yes	298(38.55)	483(64.23)		**1.66(1.15-1.84)**
Pack-years			**<0.001**	
0	475(66.43)	269(36.30)		1.00
0.50-35	145(20.28)	203(27.40)		**2.40(1.90-3.20)**
≥35	95(13.29)	269(36.30)		**5.00(3.70-6.60)**
BMI			**<0.001**	
>24	227(34.87)	348(47.22)		1.00
24-28	264(40.55)	295(40.03)		**0.73(0.57-0.92)**
>28	160(24.58)	94(12.75)		**0.38(0.28-0.52)**

[a] *P* values for a two-sidedχ^2test

Table 2. SNPs in 6q22.2 region that are in high LD with rs9387478, based on the genotype information in the CHB population in the HapMap

SNPs	**location**	**LD SNP**	**D'**	**r^2**	**MAF**
rs7749229	NA	rs9387478	0.87	0.42	0.35
rs13205986	NA	rs9387478	0.87	0.46	0.37
rs2104064	NA	rs9387478	0.86	0.40	0.34
rs7763979	NA	rs9387478	0.87	0.46	0.37
rs7746536	NA	rs9387478	0.87	0.46	0.37
rs9320604	**DCBLD1 intron1**	**rs9387478**	**0.88**	**0.78**	**0.45**
rs6942067	NA	rs9387478	0.88	0.48	0.33
rs717969	NA	rs9387478	0.88	0.48	0.38
rs6930292	NA	rs9387478	0.88	0.50	0.40
rs9489193	NA	rs9387478	0.86	0.41	0.30
rs17079281	**DCBLD1 5'flanking**	**rs9387478**	**0.86**	**0.40**	**0.34**
rs6911915	**DCBLD1 intron1**	**rs9387478**	**0.93**	**0.83**	**0.50**
rs4946259	**DCBLD1 intron1**	**rs9387478**	**0.86**	**0.68**	**0.40**
rs1555401	NA	rs9387478	0.81	0.49	0.35

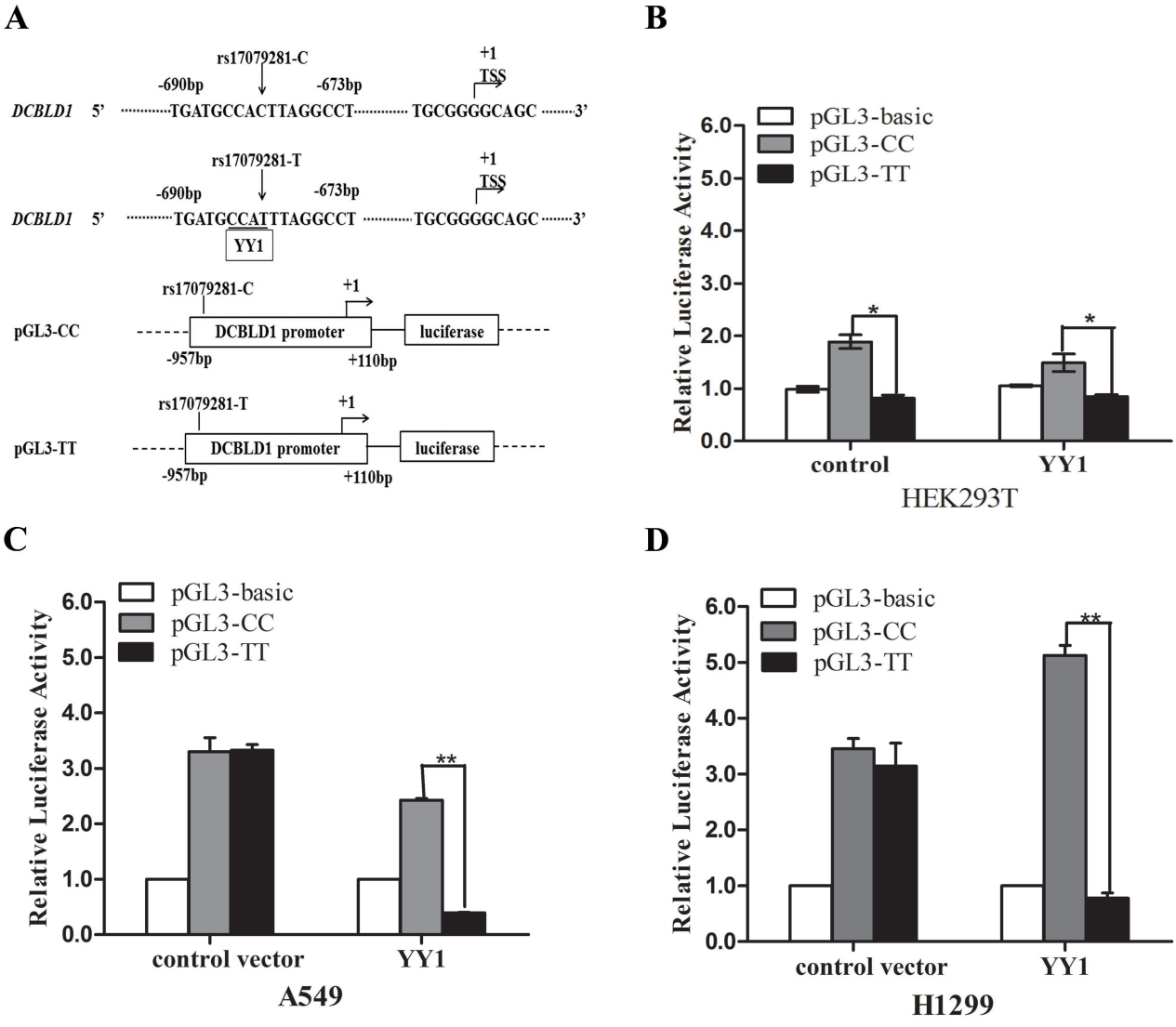
Table 3. Associations of lung cancer risk and SNPs that are in high LD with rs9387478 in 6q22.2 region

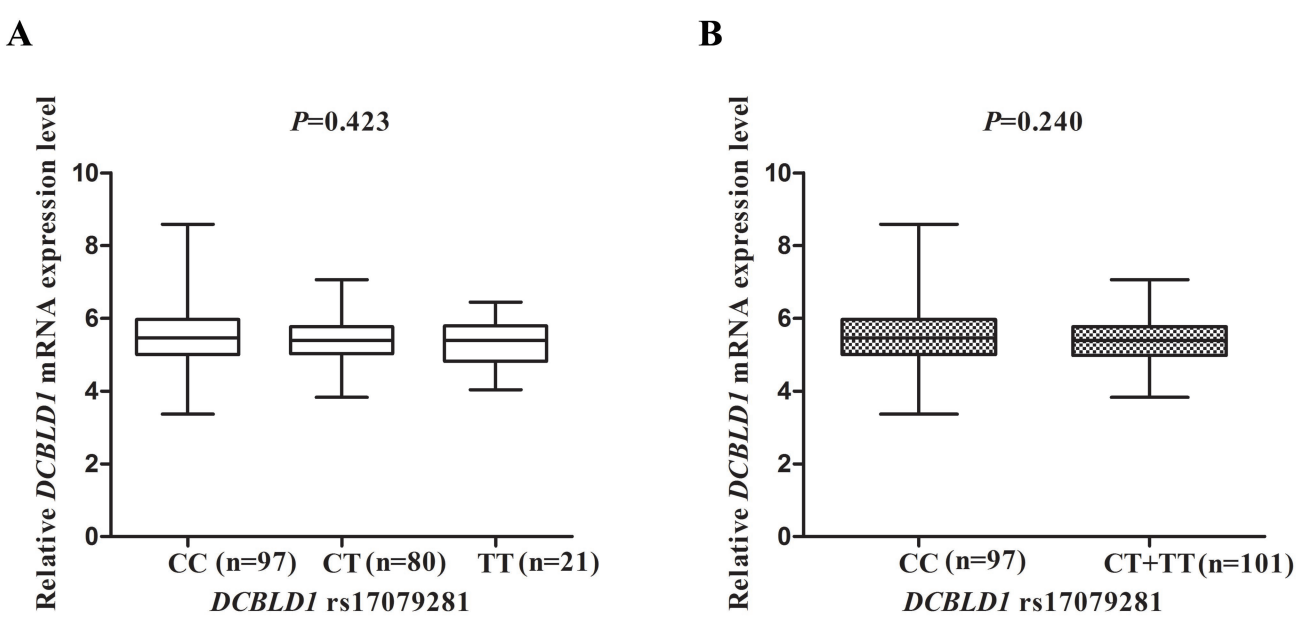
Genotype	N (%)		P^a	$OR(95\%CI)^b$
	Control(773)	Case(766)		
DCBLD1 rs17079281(C>T)			**0.039**	
CC	338(43.78)	375(48.96)		1.00
CT	361(46.76)	309(40.34)		**0.74(0.58-0.94)**
TT	73(9.46)	82(10.70)		1.02(0.69-1.53)
P trend			0.246	
Dominate model				
CC	338(43.78)	375(48.96)		1.00
CT+TT	434(56.23)	391(51.04)		**0.78(0.63-0.98)**
Recessive model				
CC+CT	699(90.54)	684(89.30)		1.00
TT	73(9.46)	82(10.70)		1.18(0.81-1.73)
DCBLD1 rs4946259(G>A)			0.130	
GG	246(31.82)	263(34.33)		1.00
AG	408(52.87)	366(47.78)		0.79(0.62-1.02)
AA	119(15.39)	137(17.89)		1.18(0.84-1.66)
P trend			0.805	
Dominate model				
GG	246(31.82)	263(34.33)		1.00
AG+GG	527(68.18)	503(65.67)		0.88(0.69-1.11)
Recessive model				
GG+AG	654(84.60)	629(82.11)		1.00
AA	119(15.39)	137(17.89)		1.35(0.99-1.84)
DCBLD1 rs6911915(T>C)			0.246	
TT	219(29.40)	244(32.88)		1.00
CT	380(51.01)	348(46.90)		**0.77(0.59-0.99)**
CC	146(19.60)	150(20.22)		0.92(0.66-1.28)
P trend			0.447	
Dominate model				
TT	219(29.40)	244(32.88)		1.00
CT+CC	426(57.18)	498(67.12)		0.81(0.63-1.03)
Recessive model				
TT+CT	599(80.40)	592(79.78)		1.00
CC	146(19.60)	150(20.22)		1.09(0.82-1.44)
DCBLD1 rs9320604(G>A)			0.285	
GG	186(28.18)	209(32.20)		1.00
AG	410(62.12)	381(58.71)		0.79(0.61-1.02)
AA	64(9.70)	59(9.09)		0.75(0.48-1.15)
P trend			0.089	
Dominate model				

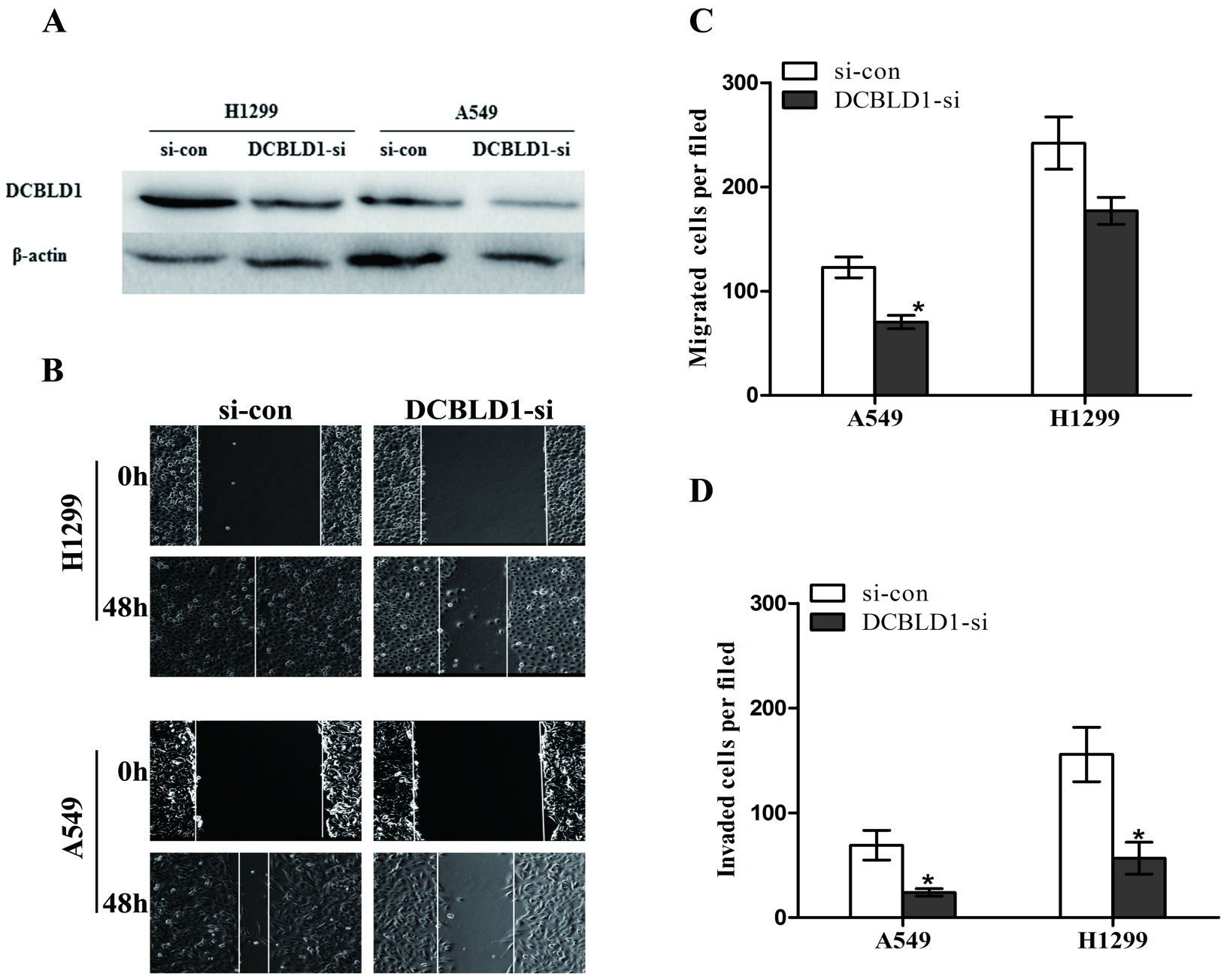
GG	186(28.18)	209(32.20)	1.00
AG+AA	474(71.82)	440(67.80)	0.78(0.61-1.01)
Recessive model			
GG+AG	596(90.30)	590(90.91)	1.00
AA	64(9.70)	59(9.09)	0.88(0.59-1.30)

[a] *P* values for a two-sided χ^2 test

[b] Adjusted by age, gender ,smoke ,BMI and family history of cancer


A
rs17079281-C
+1
TSS
-690bp
-673bp
DCBLD1 5' TGATGCCACTTAGGCCT TGCGGGGCAGC 3'
rs17079281-T
+1
TSS
-690bp
-673bp
DCBLD1 5' TGATGCCATTTAGGCCT TGCGGGGCAGC 3'
YY1
rs17079281-C
+1
pGL3-CC
DCBLD1 promoter
luciferase
-957bp
+110bp
rs17079281-T
+1
pGL3-TT
DCBLD1 promoter
luciferase
-957bp
+110bp
B
pGL3-basic
pGL3-CC
pGL3-TT
Relative Luciferase Activity
6.0
5.0
4.0
3.0
2.0
1.0
0.0
*
*
control
YY1
HEK293T
C
pGL3-basic
pGL3-CC
pGL3-TT
Relative Luciferase Activity
6.0
5.0
4.0
3.0
2.0
1.0
0.0
**
control vector
YY1
A549
D
pGL3-basic
pGL3-CC
pGL3-TT
Relative Luciferase Activity
6.0
5.0
4.0
3.0
2.0
1.0
0.0
**
control vector
YY1
H1299


A
P=0.423
Relative DCBLD1 mRNA expression level
10
8
6
4
2
0
CC (n=97)
CT (n=80)
TT (n=21)
DCBLD1 rs17079281
B
P=0.240
Relative DCBLD1 mRNA expression level
10
8
6
4
2
0
CC (n=97)
CT+TT (n=101)
DCBLD1 rs17079281


A
H1299
A549
si-con
DCBLD1-si
si-con
DCBLD1-si
DCBLD1
β-actin
B
si-con
DCBLD1-si
H1299
0h
48h
A549
0h
48h
C
si-con
DCBLD1-si
Migrated cells per filed
300
200
100
0
*
A549
H1299
D
si-con
DCBLD1-si
Invaded cells per filed
300
200
100
0
*
*
A549
H1299