SPANDIDOS PUBLICATIONS

ONCOLOGY REPORTS 47: 59, 2022

rs9390123 and rs9399451 influence the DNA repair capacity of lung cancer by regulating *PEX3* and *PHACTR2-AS1* expression instead of *PHACTR2*

QIANG SHI, QIAO-NA SHI, JIANG-WEI XU, HONG-YAN WANG, YA-JIE LI, XIN-XIN ZHANG, YU-HANG FU, RU-HUI TIAN, RU JIANG, CHUN-CHUN LIU and CHANG SUN

College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi 710119, P.R. China

Received July 8, 2021; Accepted January 7, 2022

DOI: 10.3892/or.2022.8270

Abstract. Lung cancer is a common cancer type, and has the highest mortality rate in the world. A genome-wide association study suggests that the genetic marker rs9390123 is significantly associated with DNA repair capacity (DRC) in lung cancer. Analysis of the data derived from the 1000 Genomes Project indicated that there is another single nucleotide polymorphism (SNP), rs9399451, in strong linkage disequilibrium with rs9390123 in Caucasian individuals, thus suggesting that this SNP could be associated with DRC. However, the causal SNP and mechanism of DRC remain unclear. In the present study, dual luciferase assay results indicated that both SNPs are functional in lung cells. Through chromosome conformation capture, an enhancer containing the two functional SNPs was observed to bind the promoter of peroxisomal biogenesis factor 3 and phosphatase and actin regulator 2 antisense RNA 1 (*PHACTR2-AS1*). Knockdown of *PHACTR2-AS1* could significantly influence lung cell proliferation, colony formation, migration and wound healing, which verified that *PHACTR2-AS1* is a novel oncogene for lung cancer. Through chromatin immunoprecipitation, the transcription factor POU class 2 homeobox 1 (POU2F1) was identified to bind to the surrounding segments of these two SNPs, and their interaction was investigated. The present study identified the mechanism via which rs9390123 and rs9399451 could influence DRC.

Correspondence to: Dr Chang Sun, College of Life Sciences, Shaanxi Normal University, 620 West Chang'an Road, Xi'an, Shaanxi 710119, P.R. China
E-mail: sunchang@snnu.edu.cn

Abbreviations: FUCA2, alpha-L-fucosidase 2; BAC, bacterial artificial chromosome; CDS, coding DNA sequence; ChIP, chromatin immunoprecipitation; 3C, chromosome conformation capture; CHB, Han Chinese in Beijing; DRC, DNA repair capacity; EMSA, electrophoretic mobility shift assay; EXO1, exonuclease 1; eQTL, expression quantitative trait locus; FBS, fetal bovine serum; FPKM, fragments per kilobase of transcripts per million fragments mapped; GWAS, genome-wide association studies; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; LD, linkage disequilibrium; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; PED, pedigree; PEX3, peroxisomal biogenesis factor 3; PHACTR2-AS1, PHACTR2 antisense RNA 1; PHACTR2, phosphatase and actin regulator 2; POU2F1, POU class 2 homeobox 1; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas; TF, transcription factor; CEU, Utah Residents with Northern and Western European Ancestry; VCF, Variant Call Format; YRI, Yoruba in Ibadan, Nigeria

Key words: rs9390123, rs9399451, *PEX3*, *PHACTR2-AS1*, lung cancer, DNA repair capacity

Introduction

Lung cancer is a common malignant tumor type with a high morbidity and mortality in the world, particularly in males (1). The onset of lung cancer is usually due to the interaction of multiple factors, including genetic and environmental factors such as cigarette smoking and air pollution in particular. Cigarette smoking and air pollution could induce the inhalation of multiple carcinogens, particularly benzo[a]pyrene, which could lead to DNA damage and further tumor onset. Therefore, a lower DNA repair capacity (DRC) may be associated with lung cancer susceptibility (2). A significant difference in individual DRC has been observed, and it has been suggested to be associated with genetic variations in the human genome (3). To disclose the potential genetic contribution to DRC, genome-wide association studies (GWAS) have been performed, and one single nucleotide polymorphism (SNP) in chromosome 6q24.2, rs9390123, was identified to be associated with DRC in Caucasians and thus has been suggested to contribute to lung cancer (4). Since this SNP is located on the intron of phosphatase and actin regulator 2 (*PHACTR2*), it was proposed that the association was through the function of this gene (4). However, to the best of our knowledge, this mechanism has not been investigated to date. Due to the limited capacity of microarrays, only ~317K SNPs were selected as tags to represent the human genome in a previous study (4). Consequently, the causal SNP(s) for DRC and lung cancer may be not only the genetic marker rs9390123, but also the one(s) that were not present in the microarray but exhibited linkage disequilibrium (LD) with rs9390123, despite the fact

that there were no other SNPs in LD with rs9390123 according to the International HapMap Project (5). However, to the best of our knowledge, the LD pattern in this locus has not been investigated thus far.

In the present study, the genotype data for the rs9390123 surrounding region were downloaded from the 1000 Genomes Project (6), and the LD pattern was analyzed. The function of these SNPs and the mechanism were evaluated by functional genomics assay. Through chromosome conformation capture (3C), the regulatory target genes were identified, one of which is *PHACTR2*-antisense RNA 1 (*AS1*), a long non-coding RNA (lncRNA). The function of *PHACTR2-AS1* in lung cancer was further evaluated in the present study. The current findings identified a novel oncogene for lung cancer and illustrated the potential genetic mechanism for DRC.

Materials and methods

Genotype data download and analysis. The genotype of rs9390123 surrounding 1-Mb segment (chr6: 143443314-144443314 relative to human genome build 37) was downloaded from the 1000 Genomes Project for three representative populations in the world, namely CEU (Utah Residents with European Ancestry), CHB (Han Chinese in Beijing) and YRI (Yoruba in Ibadan), by using an online program, VCF (Variant Call Format) to PED (pedigree) converter (http://grch37.ensembl.org/Homo_sapiens/Tools/VcftoPed). The obtained genotype data file (PED format) and locus information file (INFO format) were loaded into Haploview 4.2 (Broad Institute) (7) and r^2 was calculated with default parameter. When $r^2 \geq 0.8$, the SNPs were supposed to be in strong LD.

Tissue culture. The human lung epithelial cell line Beas-2B (cat. no. KCB200922YJ), lung cancer cell line A549 (cat# KCB200434YJ) and 293TN cells (cat. no. KCB2013068YJ) were maintained in high-glucose DMEM (HyClone; Cytiva) supplemented with 10% fetal bovine serum (FBS, Biological Industries; Sartorius AG) in 5% CO_2 at 37˚C. All cell lines were purchased from Kunming Cell Bank of Type Culture Collection of Chinese Academy of Sciences (http://www.kmcellbank.com/). The cell bank performed short tandem repeat genotyping, karyotype and isoenzyme analyses to verify the cell identity in order to avoid potential contamination.

Luciferase plasmid construction and transfection. To include all the potential cis-regulatory elements and amplify the segment efficiently, rs9390123 and its nearby region (~1.5 kb; chr6: 143942517-143944088) were amplified via PCR with the forward primer 5'-GGCAAATCCTTCCCATAGTTCC-3' and the reverse one 5'-TAGGGCCAAATTTCACAGGTCTTA-3' containing *Mlu*I and *Bgl*II restriction sites, respectively. The DNA isolated from Beas-2B cells was utilized as a template. The thermocycling conditions are as follows: 98˚C for 30 sec; 35 cycles of 98˚C for 10 sec, 65˚C for 30 sec, 72˚C for 30 sec; 72˚C for 2 min. To avoid potential PCR errors, amplification was performed by utilizing Q5 High-Fidelity DNA Polymerase (New England BioLabs, Inc.). Upon digestion with the aforementioned restriction enzymes (New England BioLabs, Inc.), the PCR product was inserted into the cloning site of the pGL3-promoter plasmid (Promega Corp.). The plasmid containing another allele was produced for each SNP by using a Q5 Site-Directed Mutagenesis kit (New England BioLabs, Inc.) and the primers listed in Table SI.

After seeding (~10^5 cells) into a 24-well plate and culture for 24 h as described above, 475 ng plasmid with the aforementioned inserted segment was transfected into the Beas-2B cells by using Lipofectamine® 2000 (Thermo Fisher Scientific, Inc.). After additional 36 h of culture and cell harvest, the luciferase expression was determined by using a Dual Luciferase Reporter Assay System (Promega Corp.). A total of 25 ng pRL-TK plasmid (Promega Corp.) was transfected simultaneously as an internal control to normalize the transfection efficiency, and the luciferase ratio between firefly and *Renilla* was utilized to represent the activity of the enhancer containing rs9390123 and rs9399451. In total, six replicates were performed for each plasmid transfection.

3C. The spatial conformation of the enhancer and the related gene promoter was examined by using 3C as previously described (8). Briefly, after subjecting Beas-2B cells (~10^8) to crosslink by using formaldehyde and subsequently lysing the cells, the chromatin was directly digested with *Bsr*GI (New England BioLabs, Inc.) and further ligated with T4 DNA ligase (New England BioLabs, Inc.). The ligation product was purified by using the standard phenol-chloroform method (9). Simultaneously, a BAC (bacterial artificial chromosome) RP11-1012I24 (BACPAC Genomics; https://bacpacresources.org/) harboring the 6q24.2 segment was grown in LB (Luria-Bertani) medium and purified by Large-Construct Kit (Qiagen Corp.) according to the manufacturer's protocol. After digestion, ligation and purification as aforementioned, the product was utilized as a control to normalize the primer efficiency. The relative enrichment of the 3C product was assessed by quantitative PCR (qPCR) with the unidirectional primers shown in Table SII and iTaq Universal SYBR Green Supermix (Biorad Corp.). The thermocycling conditions are as follows: 95˚C for 10 min; 40 cycles of 95˚C for 5 sec, 60˚C for 30 sec. The experiment was repeated in triplicate. All 3C PCR products were verified using sequencing with constant primer (Table SII) and BigDye® Terminator v3.1 (Thermo Fisher Scientific, Inc.) according to the manufacturer's recommendation.

Chromatin immunoprecipitation (ChIP). Potential transcription factors (TFs) were predicted by an online program Match (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi). ChIP was performed with EZ-ChIP kit (MilliporeSigma). In brief, cells (~10^7) were crosslinked by formaldehyde, lysed by lysis buffer (MilliporeSigma) and fragmented into segments of 200-800 bp by using sonication. After diluting by dilution buffer and preclearing by Protein A beads (MilliporeSigma), the chromatin/protein complex was captured overnight at 4˚C by addition of 2 μg anti-mouse POU class 2 homeobox 1 (POU2F1) antibody (Santa Cruz Biotechnology, Inc.; cat. no. sc-53830) or the same amount of normal mouse IgG (Santa Cruz Biotechnology, Inc.; cat. no. sc-2025) and further immunoprecipitated by using Protein A Beads (MilliporeSigma). After washing, resuspending, crosslink reversing and protein

digestion with related buffer (MilliporeSigma), the DNA was recovered by using a column (MilliporeSigma), and quantified via qPCR as described above with the primers shown in Table SIII. The experiment was repeated in triplicate, and the PCR product was sequenced for validation by using forward PCR primer in Table SIII and BigDye® Terminator v3.1 as described above.

TF overexpression and gene expression analysis. The POU2F1 coding region was obtained by using nested PCR with Q5 High-Fidelity DNA Polymerase. The PCR primer sequences, annealing temperature and target region are displayed in Table SIV. After dual enzymatic cleavage with *Kpn*I and *Hind*III (New England BioLabs, Inc.), the POU2F1 coding region was inserted into the pEGFP-N1 overexpression vector (Clontech Laboratories; Takara Bio USA, Inc.). A total of 500 ng POU2F1 overexpression plasmid was transfected into Beas-2B cells with Lipofectamine® 2000 as aforementioned. pEGFP-N1 was utilized as a negative control. After 48 h of culture, the relative mRNA levels of *POU2F1* and target genes were evaluated by qPCR with primers in Table SV. In total, three independent repeats were carried out.

Electrophoretic mobility shift assay (EMSA). The probes for both alleles of rs9390123 and rs9399451 (Table SVI) were labeled with biotin. Nuclear extracts were isolated from Beas-2B cells by using a Nuclear and Cytoplasmic Protein Extraction kit (Beyotime Institute of Biotechnology) and incubated with biotin-labeled probes (Sangon Biotech; 10 fmol). The probe-protein complexes were separated via electrophoresis in 4.9% non-denatured polyacrylamide gel and transferred to nylon membranes with positive charge (Beyotime Institute of Biotechnology). For each SNP allele, electrophoresis of only biotin-labeled probes and probe-protein complexes incubated with non-labeled probes (competitor oligonucleotides) was performed as a control. After incubation with 7.5 μg streptavidin-HRP (horseradish peroxidase) conjugate (Beyotime Institute of Biotechnology; cat. no. A0303) at room temperature for 15 min with 150 rpm rotation, the membrane was visualized with enhanced chemiluminescence.

Nuclear and cytoplasmic RNA isolation. RNA in the nucleus and cytoplasm of Beas-2B cells was separated and isolated by using an RNA subcellular isolation kit (Active Motif, Inc.). Reverse transcription (RT) was performed with SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Inc.). *PHACTR2-AS1* expression was determined by using qPCR with primer pairs from the literature (10-12) (Table SV). *U6* and *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) are known to be expressed mainly in the nucleus (13) and cytoplasm (14), respectively. Therefore, their expression was also determined by qPCR with primer pairs shown in Table SV, and it was used as positive controls to verify the separation of nucleus and cytoplasm.

Short hairpin RNA (shRNA) design, plasmid construction, lentivirus packaging and transfection. shRNA (5'-GCC CTGCATACTGTGGATTCA-3') for *PHACTR2-AS1* was designed with the online software BLOCK-iT RNAi Designer (https://rnaidesigner.thermofisher.com/rnaiexpress/; Thermo Fisher Scientific, Inc.). The restrictive sites for *Eco*RI and *Bam*HI were added into shRNA sequence and both forward and reverse strands were synthesized (Sangon Biotech). After annealing, the two strands formed double-strand DNA with sticky end. After *Eco*RI and *Bam*HI (New England BioLabs, Inc.) digestion, the pGreenPuro vector (System Biosciences, LCC) was ligated with the DNA segment containing shRNA by T4 DNA ligase. The lentiviral pGreenPuro vector (3 μg) was transfected into 293TN cells by Lipofectamine® 2000 along with packaging (psPAX2; addgene# 12260; 2.5 μg) and envelope (pMD2.G; addgene# 12259; 2.5 μg) plasmids to produce viruses. 293TN is a genetically modified cell line to produce high titer virus. After 48 h of culture, the viruses were collected via centrifugation. The empty pGreenPuro vector (3 μg) was also transfected into 293TN cells along with packaging and envelope plasmids to produce viruses, which were further used as a negative control (sh-NC). Beas-2B and A549 cells were transfected with the aforementioned lentivirus and screened with puromycin (2 μg/ml final concentration) for 2 weeks in order to obtain a stable cell line.

Total RNA was purified from Beas-2B and A549 cells using TRIzol® (Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions, quantified by using Nanodrop 2000 (Thermo Fisher Scientific, Inc.) and RT was carried out as aforementioned. *PHACTR2-AS1* expression after knockdown was evaluated via PCR or qPCR with the primer pair listed in Table SV. *GAPDH* was also amplified as a positive control by using the primer pair listed in Table SV.

Cell proliferation assay. Beas-2B and A549 cells with lentiviral transfection were cultured for 4 days in 96-well plates (6,000 cells per well). Every 24 h, 20 μl MTT (5 mg/ml; Beijing Solarbio Science & Technology Co., Ltd.) was added to the cells and incubated for 4 h at 37˚C. Four time points were included in this assay. After removing the supernatant and adding 150 μl DMSO to each well, the plate was covered with foil and placed on an orbital shaker for 15 min at room temperature. Next, the optical density (OD) value was examined at a wavelength of 490 nm. A total of three independent repeats were carried out for each time point.

Colony formation assay. Beas-2B and A549 cells were adjusted to a density 500 cells per well and cultured for 14 days in a 6-well plate. After washing with PBS, the cells were fixed with paraformaldehyde and stained with 1% crystal violet (Beijing Solarbio Science & Technology Co., Ltd.) for 20 min at room temperature. The colony number was then determined visually using a DMi8 Automated Inverted Phase Contrast Fluorescence Microscope (Leica Microsystems). Three independent repeats were performed.

Cell migration assay. Beas-2B and A549 cells (10^4) were resuspended into DMEM without FBS and placed in the upper chamber of a Transwell cell culture plate (MilliporeSigma), while DMEM with 10% FBS was placed in the lower chamber. After 2 days of culture at 37˚C, paraformaldehyde fixing and crystal violet staining of Beas-2B and A549 cells were performed as aforementioned, and the number of cells was examined using a DMi8 Automated Inverted Phase Contrast

Fluorescence Microscope as described above. Three independent repeats were carried out.

Wound healing assay. Beas-2B and A549 cells were adjusted to a density of $7x10^5$ cells per well and maintained to near full (95%) confluence in DMEM with 10% FBS. After scratching the cell monolayer with a pipette tip, the cells were washed with PBS for 2 min at room temperature (twice), and fresh DMEM without FBS was added. The wound areas were imaged at 0 and 48 h using an inverted microscope, and were analyzed using ImageJ software v1.51 (https://imagej.net/). Three independent repeats were performed.

Statistical analysis. The luciferase activity for each plasmid, ChIP enrichment, gene expression after POU2F1 overexpression, and cell proliferation, colony formation, cell migration and wound healing abilities were presented as mean ± standard deviation (SD). Independent (unpaired) Student's t-test was utilized to evaluate difference in the above data between two groups. All statistical analyses were performed with SPSS 20.0 (IBM Corp.). P<0.05 was considered to indicate a statistically significant difference.

Results

SNPs and LD pattern in the rs9390123 surrounding region. To identify the potential SNP(s) associated with DRC, the genotype data surrounding rs9390123 in its nearby region were obtained from the 1000 Genomes Project for three representative populations in the world. For the 1M segment surrounding rs9390123, 3,927, 4,285 and 6,800 SNPs were found to exist in the CEU, CHB and YRI populations, respectively. However, only one SNP, rs9399451, presented high LD with rs9390123 in CEU (r^2=0.979) and YRI (r^2=0.851; Fig. S1), thus suggesting that this SNP may also be associated with DRC in these two populations. In the CHB population, the LD between these two SNPs was found to be moderate (r^2=0.645; Fig. S1). Except rs9399451, all other SNPs within this region showed relatively low LD with rs9390123 (all r^2≤0.668, 0.633 and 0.675 in CEU, CHB and YRI, respectively; Fig. S1). There was a distance of only 560 bp between these two SNPs, suggesting that they both could form part of the same functional element. For the entire human population, T is the minor allele for both rs9390123 and rs9399451, and their frequencies were found to be ~38% and 30% for CEU and YRI populations, respectively. In contrast, T frequencies in CHB were ~58 and ~68% for rs9390123 and rs9399451, respectively.

Function of rs9390123 and rs9399451. Since neither rs9390123 nor rs9399451 are within protein-coding regions, it was hypothesized that these two SNPs could influence target gene expression. To evaluate the function of rs9390123 and rs9399451 on gene expression, the segment surrounding rs9390123 and rs9399451 was inserted into the cloning site of the pGL3-promoter vector. The sequencing indicated that the nucleotide is T for both rs9390123 and rs9399451 in the plasmid construct (Fig. 1). The plasmids with corresponding allele, i.e., C at rs9390123 and G at rs9399451, were generated via mutagenesis, and subsequently all plasmids were transfected. Since most lung cancer events derive from normal epithelial cells, the lung epithelial cell line Beas-2B was used in following functional genomics work.

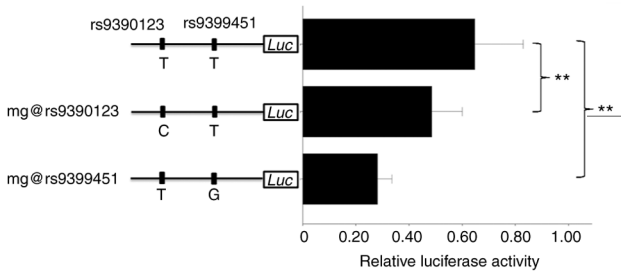

Figure 1. Relative enhancer activity for the rs9390123 and rs9399451 alleles in Beas-2B cells. Each bar represents one plasmid. The top plasmid is the original construct, while the bottom two are derived from mutagenesis. The x-axis denotes relative luciferase level. Data are normalized to the value of the pGL3-promoter plasmid (empty vector), and are displayed as the mean ± SD. Six independent transfections are performed for each plasmid in this assay. **P<0.01. Mg, mutagenesis.

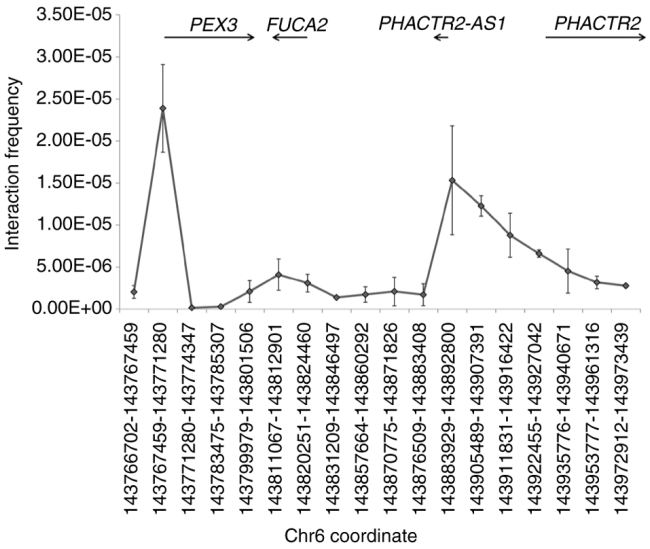

Figure 2. Interacting efficiency between the enhancer containing rs9390123 and rs9399451 and multiple genome elements within 6q24.2 in Beas-2B cells. The x-axis designates the position of the restrictive fragment in chromosome 6 (according to the coordinates of human genome build 37), while the y-axis represents the level of chromosome conformation capture-PCR normalized to the bacterial artificial chromosome (BAC) clone. The arrow indicates the location and transcript orientation of each gene in this region. Three independent repeats are performed. Data are shown as the mean ± SD. *PEX3*, peroxisomal biogenesis factor 3; *FUCA2*, alpha-L-fucosidase 2; *PHACTR2*, phosphatase and actin regulator 2; AS1, antisense RNA 1.

As shown in Fig. 1, the C allele of rs9390123 showed ~24.8% lower relative luciferase activity than the T allele (P=0.0041), while the G allele of rs9399451 presented ~56.7% lower luciferase activity than the T one ($P<10^{-5}$; Fig. 1), which indicated that both SNPs are functional in lung cells, and rs9399451 may have a more important effect than rs9390123 in regulating target gene expression.

Interaction between the enhancer containing rs9390123 and rs9399451, and the PEX3 and PHACTR2-AS1 promoter. Since rs9390123 and rs9399451 are not located at the promoter of any known genes, it was proposed that these two SNPs may be located within an enhancer region and could regulate the

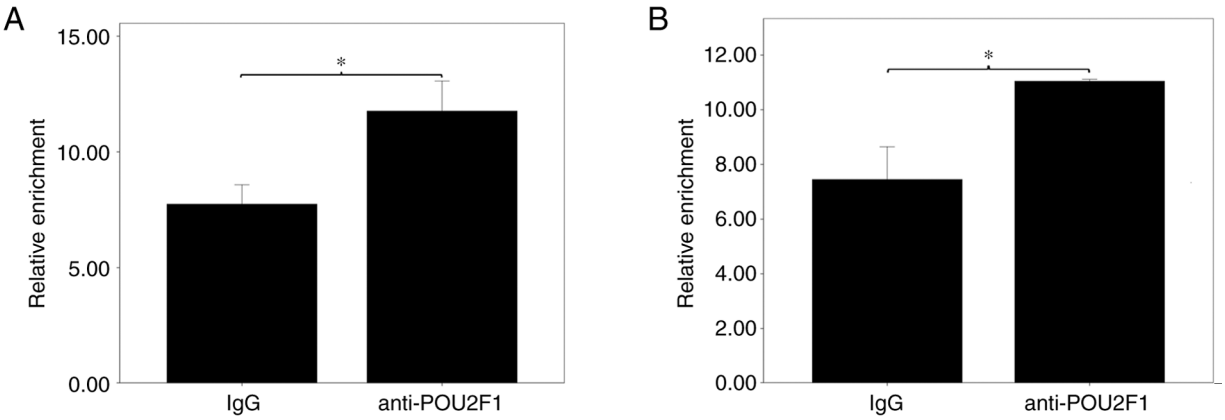

Figure 3. Chromatin enrichment of the region spanning the position of (A) rs9390123 and (B) rs9399451 in Beas-2B cells. The y-axis represents relative enrichment normalized by the input. Three independent repeats are performed. Data are shown as the mean ± SD. *P<0.05. POU2F1, POU class 2 homeobox 1.

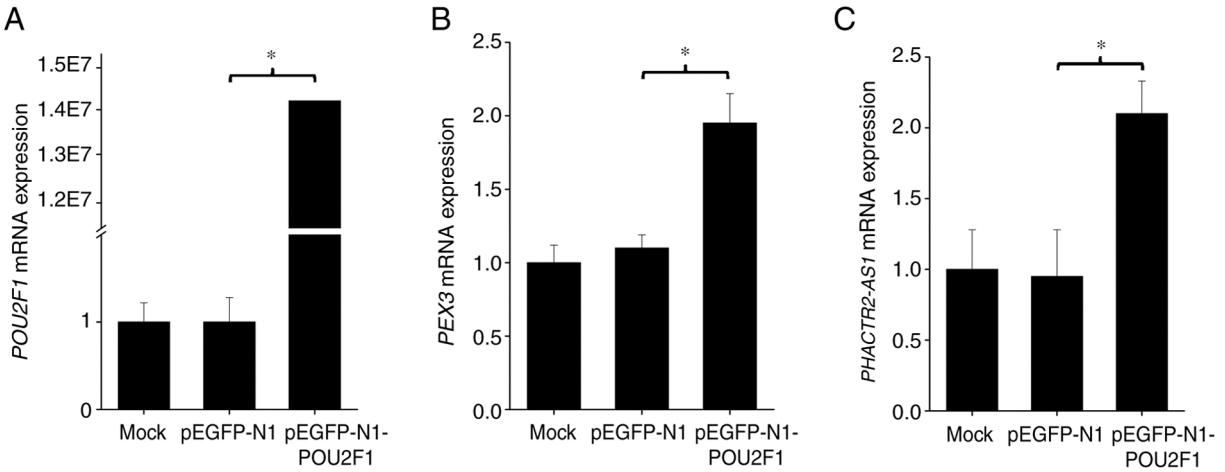

Figure 4. Effect of overexpression vector transfection on (A) *POU2F1*, (B) *PEX3* and (C) *PHACTR2-AS1* expression. Three independent transfections are performed. Data are shown as the mean ± SD. *P<0.05. *POU2F1*, POU class 2 homeobox 1; *PEX3*, peroxisomal biogenesis factor 3; *PHACTR2*, phosphatase and actin regulator 2; AS1, antisense RNA 1.

expression of target genes. This suggestion could be supported by the histone modification signal. Searching the data from the Encyclopedia of DNA Elements project (15) in the UCSC genome browser (http://genome.ucsc.edu) revealed that there are clear H3K27Ac (acetylation at the 27th lysine residue of the histone H3 protein) and H3K4me1 (mono-methylation at the 4th lysine residue of the histone H3 protein) peaks, which are two common histone modifications around an active enhancer (16), near this region in the human lung cancer cell line A549 (Fig. S2). However, the target gene remained unclear. To identify the potential target gene for this enhancer, 3C was used. The rationale of 3C is that the enhancer is far away from its target gene(s) in sequence but close in space. Thus, after 3C library construction, the target gene promoter can be ligated in higher efficiency with enhancer than other random genome regions, which can be disclosed by qPCR. To normalize the primer efficiency, the BAC RP11-1012I24 including nearby region was used as a control. There are only three protein-coding genes [*PEX3*, *FUCA2* (α-L-fucosidase 2) and *PHACTR2*] and one lncRNA *PHACTR2-AS1* in this BAC. Therefore, in our 3C assay, unidirectional primers were designed to anchor the promoter of these four genes, the

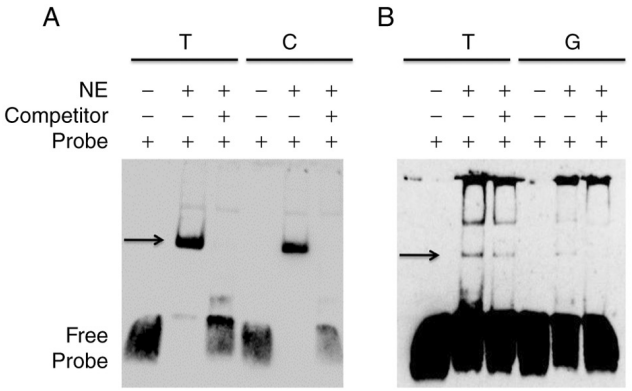

Figure 5. Difference in binding affinity between Beas-2B nuclear proteins and (A) rs9390123 and (B) rs9399451 alleles in EMSA. The top line indicates different alleles for each SNP. NE denotes nuclear extracts, and the arrow indicates the position of the protein-probe complex. EMSA, electrophoretic mobility shift assay; SNP, single nucleotide polymorphism.

aforementioned enhancer and several random regions. As shown in Fig. 2, no increase was observed in ligation frequency at the *PHACTR2* promoter (corresponding to the 15th point in

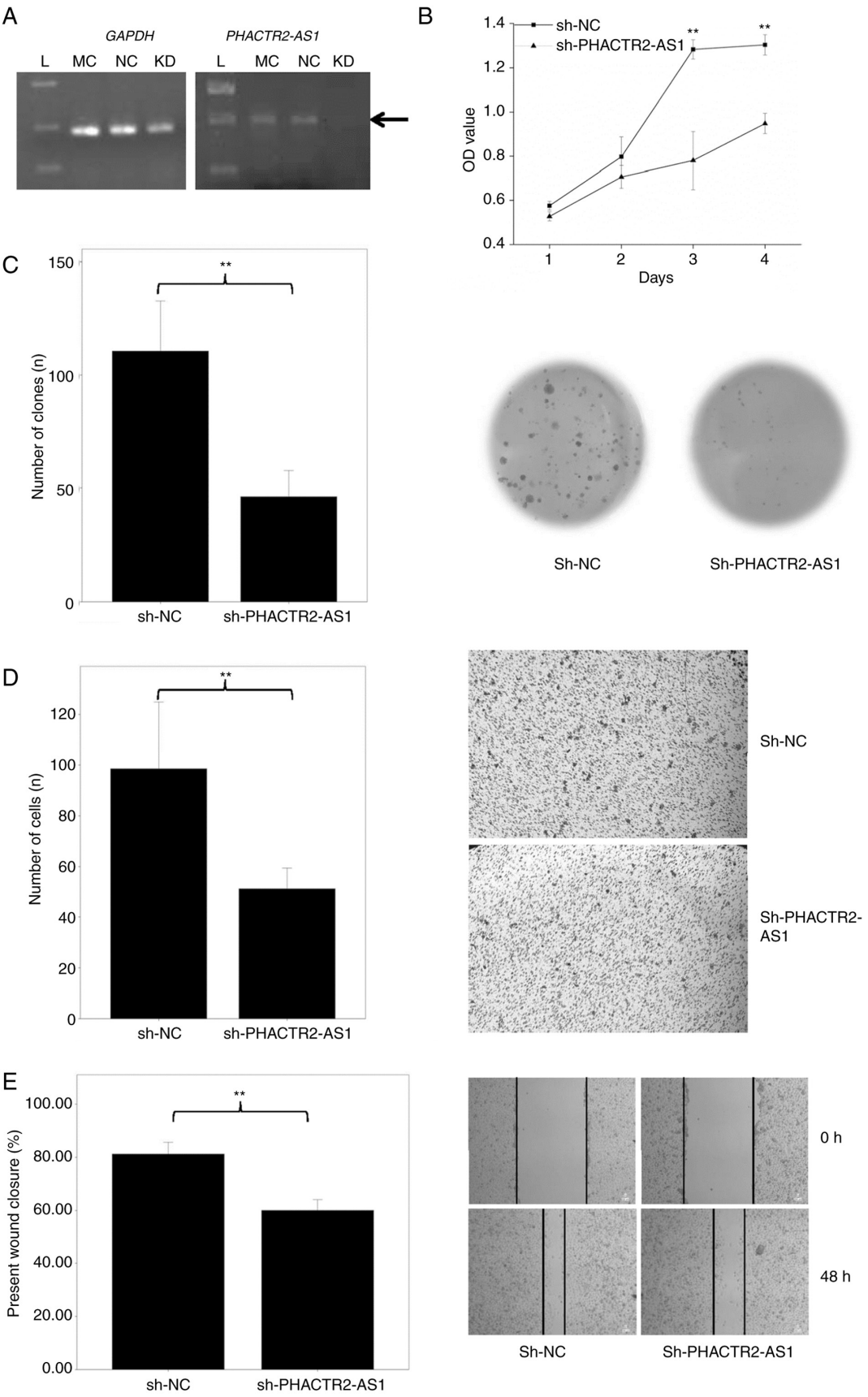

Figure 6. Results of (A) amplification and its effect on the (B) cell proliferation, (C) colony formation, (D) cell migration and (E) wound healing of *PHACTR2-AS1* stable knockdown Beas-2B cells (sh-*PHACTR2-AS1*). In panel A, the left side corresponds to *GAPDH*, while the right side corresponds to *PHACTR2-AS1*. For each side, the first lane is the DNA ladder, while the second and third lanes contain the cDNA from sh-NC and sh-*PHACTR2-AS1*, respectively. The arrow indicates the position of the *PHACTR2-AS1* PCR product, which was undetectable after knockdown. In panel B, the x-axis designates culture time, while the y-axis represents relative cell number. In panel D, the background for sh-*PHACTR2-AS1* is relatively lighter, which may induce an overlook of some cells in vision. Three independent repeats are performed in panel B-E. Data are shown as the mean ± SD. $^{**}P<0.01$. sh, small hairpin; *PHACTR2*, phosphatase and actin regulator 2; AS1, antisense RNA 1; MC, Mock; L, ladder; NC, Negative control; KD, knockdown.

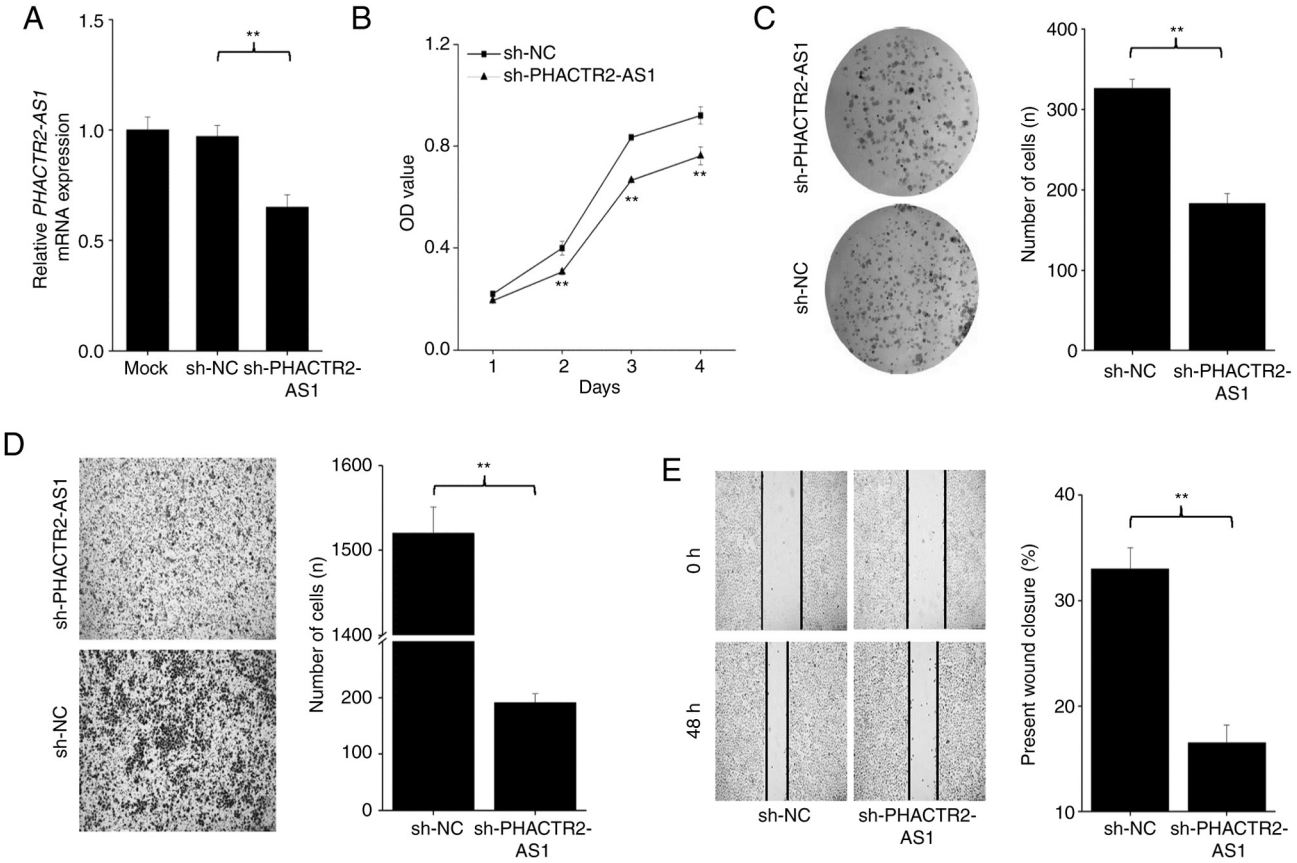

Figure 7. (A) qPCR result and the effect on (B) cell proliferation, (C) colony formation, (D) migration and (E) wound healing of *PHACTR2-AS1* stable knockdown in A549 cells (sh-*PHACTR2-AS1*). The data are shown as mean ± SD; **P<0.01. For part B, the x-axis designates culture time while the y-axis represents relative cell number. sh, small hairpin; *PHACTR2*, phosphatase and actin regulator 2; AS1, antisense RNA 1; NC, Negative control.

the x-axis, ~17.9 kb away from the enhancer), suggesting that the regulatory target of this enhancer is not *PHACTR2*. The *FUCA2* promoter region (7th point in the x-axis) exhibited a similar result. By contrast, a higher ligation efficiency was obtained in the *PEX3* promoter region (2nd point in the x-axis, ~173.6 kb away from the enhancer). In addition, the promoter of the lncRNA *PHACTR2-AS1* (12th point in the x-axis, ~50.5 kb away from the enhancer), also presented a high ligation efficiency, indicating that *PEX3* and *PHACTR2-AS1* are the regulatory targets of this enhancer.

TF binding to rs9390123 and rs9399451. Since rs9390123 and rs9399451 are located within a non-coding region of the genome, it could be hypothesized that these two SNPs were located at a TF binding region, and could influence the interaction between TF and DNA. Since TFs usually bind to DNA through recognizing specific motif, the potential TF can be predicted by bioinformatics approach. If one TF can interact with specific DNA segment, the TF antibody can immunoprecipitate more DNA-TF complex than IgG, since IgG presents random interaction with TF. The surrounding region containing T of rs9390123 and rs9399451 are matching with the motif of POU2F1 (see https://jaspar.genereg.net/ for detail). In contrast, both C of rs9390123 and G of rs9399451 can disrupt the recognizing site for POU2F1, thus resulting much lower scores for POU2F1 interaction in prediction (data not shown). Therefore, it was hypothesized that rs9390123 and rs9399451 effect through influencing POU2F1 binding. To confirm this suggestion, ChIP with POU2F1 antibodies was utilized to identify the surrounding region of the potential binding TF. As shown in Fig. 3, the anti-POU2F1 antibody significantly pulled down more chromatin for the region surrounding rs9390123 and rs9399451 compared with that pulled down by IgG (P=0.001 and 0.004, respectively), thus supporting that POU2F1 has the ability to bind these two regions in lung cells.

Effect of POU2F1 on gene expression. To further examine the function of POU2F1 on gene transcription, a POU2F1 overexpression vector was constructed and transfected into Beas-2B cells and mRNA levels were evaluated by RT-qPCR. After transfection of the overexpression plasmid, POU2F1 expression was increased ~$1.4x10^7$ fold ($P<10^{-7}$; Fig. 4A), which verified that the transfection efficiency was very high. This could further induce ~95.0 and ~110% increase in mRNA expression of *PEX3* (P=0.0002; Fig. 4B) and *PHACTR2-AS1* (P=0.012; Fig. 4C) in lung cells, respectively, suggesting that POU2F1 can influence the expression of target genes.

Difference in TF binding affinity between alleles at the rs9390123 and rs9399451 SNPs. Our prediction, ChIP and overexpression assay verify the effect of POU2F1 in the expression of these two genes. To further investigate the binding efficiency between two alleles at these two SNPs,

EMSA was performed based on biotin-labeled probes. As shown in Fig. 5, the two alleles for rs9390123 and rs9399451 showed an apparent different affinity to nuclear proteins from Beas-2B cells. Furthermore, these patterns could be abolished by adding competitor oligonucleotides (Fig. 5), which verified the hypothesis that these two SNPs could alter TF binding affinity.

Transcript location and effect of PHACTR2-AS1 on cell proliferation, colony formation, migration and wound healing. Since *PHACTR2-AS1* is the regulatory target of the enhancer containing rs9390123 and rs9399451, the function and transcript location of this gene in lung cells were investigated. To determine the location of the lncRNA *PHACTR2-AS1*, nuclear and cytoplasmic RNA was isolated and quantified separately. As shown in Fig. S3, this lncRNA is predominantly expressed in the nucleus. To investigate the function of *PHACTR2-AS1* in lung cells, a stable knockdown cell line was constructed using Beas-2B cells by transfecting a lentivirus containing shRNA, and the knockdown efficiency is displayed in Fig. 6A. In Fig. 6A, the left panel is the PCR result for *GAPDH* and the right panel is for *PHACTR2-AS1*. For each part, the four lanes, from left to right, are ladder (L), Mock (MC), negative control (NC) and knockdown (KD) cells, respectively. The arrow in Fig. 6A points out the position of the *PHACTR2-AS1* PCR product. As shown in Fig. 6A, *PHACTR2-AS1* expression is persistent in the NC but decreased to undetectable amount in the KD cells, which indicates that the transfection efficiency was high and the knockdown was successful. Cell proliferation was evaluated with an MTT assay. As shown in Fig. 6B, cell proliferation was significantly inhibited in the KD group compared with that of the control group (P=0.0017 and $<10^{-5}$ for days 3 and 4, respectively). The number of formed clones was also significantly decreased after *PHACTR2-AS1* knockdown (P=0.00085; Fig. 6C). Migration and wound healing abilities were also significantly attenuated after knockdown (P=0.00048 and 0.00024, respectively; Fig. 6D and E).

To substantiate the function of *PHACTR2-AS1* in tumorigenic lung cells, we further transfected lung cancer cell line A549 with the lentivirus. As shown in Fig. 7A, *PH2ACTR2-AS1* expression was similar in the mock and negative control (NC) (P>0.05). After RNA interference, *PH2ACTR2-AS1* expression was decreased ~40% (Fig. 7A) compared with the mock and NC, which indicates that the transfection efficiency was high. Consequently, the cell count was reduced significantly at day 2, 3 and 4 (all P=0.0012, $<10^{-5}$ and 0.0011, respectively; Fig. 7B). Meanwhile, the clone formation (P=0.00012; Fig. 7C), migration ($P<10^{-6}$; Fig. 7D) and wound healing (P=0.00062; Fig. 7E) abilities were also significantly inhibited. Overall, these results indicated that *PHACTR2-AS1* can promote lung cell proliferation and migration *in vitro*, and suggested that it could be a novel oncogene for lung cancer.

Discussion

The present study aimed to clarify the underlying mechanism between the genetic marker rs9390123 and DNA repair capacity (DRC). Data from the 1000 Genomes Project were utilized to analyze the linkage disequilibrium (LD) spectrum, and another single nucleotide polymorphism (SNP), rs9399451, was identified to be in strong LD with rs9390123, particularly in Caucasians. Dual luciferase assay results suggested that both SNPs could alter enhancer activity. Chromosome conformation capture (3C) revealed that the regulatory target genes may be *PEX3* and lncRNA *PHACTR2-AS1* instead of *PHACTR2*. By using ChIP and EMSA, the associated transcription factors (TFs) and molecular mechanism were identified. The function of *PHACTR2-AS1* in lung cancer was further evaluated. The present findings suggest an association between the genetic marker rs9390123 and DRC.

Phosphatase and actin regulator 2 (*PHACTR2*) was reported to be mainly expressed in the nervous system (17). Since rs9390123 is located within the intron of *PHACTR2*, this gene was proposed to be the regulatory target of this enhancer (4). However, the present 3C results rejected a potential spatial interaction between the enhancer containing rs9390123 and the *PHACTR2* promoter, which indicated that *PHACTR2* is not the regulatory target of this enhancer.

Instead, the results of 3C suggested a clear spatial contact between this enhancer and the promoter of *PEX3* and *PHACTR2-AS1*. The peroxisome is an organelle containing numerous enzymes (18), which plays a central role in the metabolism of multiple substrates, particularly reactive oxygen species (ROS) (19). Active ROS in cells can induce DNA damage and further cancer onset (20). Therefore, an association between the peroxisome and cancer has been proposed (21). Peroxisomal biogenesis factor 3 (*PEX3*) is a membrane import receptor that plays an important role in peroxisome biogenesis (22). Therefore, high *PEX3* expression could benefit DRC and further reduce the possibility of lung cancer onset. A search for *PEX3* expression in The Cancer Genome Atlas database by online software TANRIC (23) indicated that the expression of this gene is significantly suppressed in lung adenocarcinoma tissues compared with that in normal tissues (P=0.0059; Fig. S4). However, in lung squamous cell carcinoma, no significant difference was observed (P=0.17; data not shown), which may be due to the relatively low sample size of the normal group (n=17). Considering the function of *PEX3*, this gene should have a tumor-suppressor effect through a long-term pattern, and knockdown of this gene would not directly influence the cell cycle, particularly in culture condition.

PHACTR2-AS1 (previously known as lncIHS or NR027113) (10-12), another regulatory target of this enhancer containing rs9390123 and rs9399451 besides *PEX3*, is a novel lncRNA with various cellular locations and functions. In breast tissue, *PHACTR2-AS1* is suggested to be a tumor-suppressor gene for breast cancer, and injecting this lncRNA fragment into mice could inhibit tumor growth and metastasis (24). By contrast, this lncRNA was suggested to be an oncogene for hepatocellular, gastric and tongue squamous cell carcinoma (10-12,25). In tongue squamous carcinoma cells, *PHACTR2-AS1* is mainly expressed in the cytoplasm and functions by interacting with microRNA-137 (25). In hepatocellular and gastric carcinoma cells, this lncRNA was proposed to be present in the nucleus and is able to regulate the activity of the ERK and AKT signaling pathways (10-12). Considering the same location and function, and the wide distribution of members

of the ERK and AKT signaling pathways in human tissues, it could be hypothesized that *PHACTR2-AS1* could also activate the ERK and AKT signaling pathways in lung tissues. Furthermore, it has been proposed that the ERK and AKT signaling pathways could activate DNA repair (26-31), and inhibition of either pathway could impair DRC in several cancer types (32-35). Therefore, the association between the genetic marker rs9390123 and DRC in lung tissue (4) may also be interpreted by the capability of regulating *PHACTR2-AS1* expression.

Due to the important role of DRC in the onset of lung cancer, it could be concluded that rs9390123 may be associated with lung cancer by regulating *PEX3* and *PHACTR2-AS1* expression and further DRC. However, the association between rs9390123 and lung cancer risk failed to reach the genome-wide significance threshold (4). This inconsistency may be due to the function of the *PHACTR2-AS1* transcript. Indeed, higher *PHACTR2-AS1* expression would not only benefit DRC but also promote cell proliferation, as shown in the present knockdown results, which may attenuate the association between rs9390123 and lung cancer risk.

The present work has some limitations in the lack of evidence for an association between rs9399451 or the two target genes and DRC. Since measuring DRC is beyond our ability, we cannot provide direct evidence for this issue. However, it is notable that the r^2 between rs9399451 and rs9390123 was as high as 0.979 in the CEU population, which is near a complete LD and suggests the same result of these two SNPs in association study. Moreover, there has been no evidence for the involvement of *PEX3* and *PHACTR2-AS1* in DRC. However, considering the function of the peroxisome (18) and the role of ERK and AKT pathway in DNA repair activation (26-31), a lower expression of these two genes induced by genetic factors will definitely decrease individual DRC, which deserves further investigation.

Considering the function of rs9390123 and rs9399451 and the present 3C results, it could be proposed that these two SNPs may be an expression quantitative trait locus (eQTL) for *PEX3* and *PHACTR2-AS1*. To verify this hypothesis, a search was performed in GTEx (https://gtexportal.org/), a database including data on the regulation of gene expression in multiple tissues (36). However, no association was observed between this locus and the expression of the *PEX3* and *PHACTR2-AS1* genes. This discrepancy may be interpreted by the fact that the sensitivity and power of eQTL analysis could be influenced by environmental and physiological effects (37). This effect is more serious for genes involved in the cell cycle and response to exogenous treatment (38).

In conclusion, our research effort indicates that rs9390123 and rs9399451 influence DRC of lung cancer through regulating *PEX3* and *PHACTR2-AS1* expression, which illuminates the mechanism for the association in GWAS and guarantees the usage of the expression levels of these two genes to measure individual DRC.

Acknowledgements

We thank Professor Huanjie Shao (Shaanxi Normal University) for technical advice.

Funding

This research was supported by the Fundamental Research Funds for the Central Universities (nos. 2018CBLY005 and GK202001004), National Natural Science Foundation of China (no. 31370129) and Qinzhu Grant (no. KY2019YB017).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

QS and CS designed the project and applied for the funding. QS, QNS, HYW, YJL, XXZ, JWX, YHF, RHT, RJ and CCL performed the experiments. QS analyzed the data. CS composed the manuscript. All authors read and approved the manuscript and ensure the integrity of the collected data and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

Ethics approval was not applicable to the current study as no human subjects were involved. The data in Fig. S4 are from a public database.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' information

Dr Chang Sun: ORCID: 0000-0002-2425-9485.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. Wei Q, Cheng L, Amos CI, Wang LE, Guo Z, Hong WK and Spitz MR: Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: A molecular epidemiologic study. J Natl Cancer Inst 92: 1764-1772, 2000.
3. Wei Q and Spitz MR: The role of DNA repair capacity in susceptibility to lung cancer: A review. Cancer Metastasis Rev 16: 295-307, 1997.
4. Wang LE, Gorlova OY, Ying J, Qiao Y, Weng SF, Lee AT, Gregersen PK, Spitz MR, Amos CI and Wei Q: Genome-wide association study reveals novel genetic determinants of DNA repair capacity in lung cancer. Cancer Res 73: 256-264, 2013.
5. International HapMap Consortium: The International HapMap Project. Nature 426: 789-796, 2003.
6. The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR: A global reference for human genetic variation. Nature 526: 68-74, 2015.

7. Barrett JC, Fry B, Maller J and Daly MJ: Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265, 2005.
8. Yang YC, Fu WP, Zhang J, Zhong L, Cai SX and Sun C: rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in east Asian populations. Carcinogenesis 39: 1216-1221, 2018.
9. Kochl S, Niederstatter H and Parson W: DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. Methods Mol Biol 297: 13-30, 2005.
10. Chen Z, Zhou ZY, He CC, Zhang JL, Wang J and Xiao ZY: Down-regulation of LncRNA NR027113 inhibits cell proliferation and metastasis via PTEN/PI3K/AKT signaling pathway in hepatocellular carcinoma. Eur Rev Med Pharmacol Sci 22: 7222-7232, 2018.
11. Chen QF, Hu CF, Sun K and Lang YP: LncRNA NR027113 promotes malignant progression of gastric carcinoma via EMT signaling pathway. Eur Rev Med Pharmacol Sci 23: 4746-4755, 2019.
12. Chen Z, Yu W, Zhou Q, Zhang J, Jiang H, Hao D, Wang J, Zhou Z, He C and Xiao Z: A novel lncRNA IHS promotes tumor proliferation and metastasis in HCC by regulating the ERK- and AKT/GSK-3β-signaling pathways. Mol Ther Nucleic Acids 16: 707-720, 2019.
13. Pessa HK, Will CL, Meng X, Schneider C, Watkins NJ, Perälä N, Nymark M, Turunen JJ, Lührmann R and Frilander MJ: Minor spliceosome components are predominantly localized in the nucleus. Proc Natl Acad Sci USA 105: 8655-8660, 2008.
14. Tristan C, Shahani N, Sedlak TW and Sawa A: The diverse functions of GAPDH: Views from different subcellular compartments. Cell Signal 23: 317-323, 2011.
15. ENCODE Project Consortium: An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74, 2012.
16. Calo E and Wysocka J: Modification of enhancer chromatin: What, how, and why? Mol Cell 49: 825-837, 2013.
17. Allen PB, Greenfield AT, Svenningsson P, Haspeslagh DC and Greengard P: Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins. Proc Natl Acad Sci USA 101: 7187-7192, 2004.
18. de Duve C: The peroxisome: A new cytoplasmic organelle. Proc R Soc Lond B Biol Sci 173: 71-83, 1969.
19. Fransen M, Lismont C and Walton P: The peroxisome-mitochondria connection: How and why? Int J Mol Sci 18: 1126, 2017.
20. Prasad S, Gupta SC and Tyagi AK: Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals. Cancer Lett 387: 95-105, 2017.
21. Dahabieh MS, Di Pietro E, Jangal M, Goncalves C, Witcher M, Braverman NE and Del Rincón SV: Peroxisomes and cancer: The role of a metabolic specialist in a disease of aberrant metabolism. Biochim Biophys Acta Rev Cancer 1870: 103-121, 2018.
22. Sugiura A, Mattie S, Prudent J and McBride HM: Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes. Nature 542: 251-254, 2017.
23. Li J, Han L, Roebuck P, Diao L, Liu L, Yuan Y, Weinstein JN and Liang H: TANRIC: An interactive open platform to explore the function of lncRNAs in cancer. Cancer Res 75: 3728-3737, 2015.
24. Chu W, Zhang X, Qi L, Fu Y, Wang P, Zhao W, Du J, Zhang J, Zhan J, Wang Y, *et al*: The EZH2-PHACTR2-AS1-Ribosome axis induces genomic instability and promotes growth and metastasis in breast cancer. Cancer Res 80: 2737-2750, 2020.
25. Yuan F, Miao Z, Chen W, Wu F, Wei C, Yong J and Xiao C: Long non-coding RNA PHACTR2-AS1 promotes tongue squamous cell carcinoma metastasis by regulating Snail. J Biochem 168: 651-657, 2020.
26. Golding SE, Morgan RN, Adams BR, Hawkins AJ, Povirk LF and Valerie K: Pro-survival AKT and ERK signaling from EGFR and mutant EGFRvIII enhances DNA double-strand break repair in human glioma cells. Cancer Biol Ther 8: 730-738, 2009.
27. Sato A, Sunayama J, Matsuda K, Seino S, Suzuki K, Watanabe E, Tachibana K, Tomiyama A, Kayama T and Kitanaka C: MEK-ERK signaling dictates DNA-repair gene MGMT expression and temozolomide resistance of stem-like glioblastoma cells via the MDM2-p53 axis. Stem Cells 29: 1942-1951, 2011.
28. de Laval B, Pawlikowska P, Barbieri D, Besnard-Guerin C, Cico A, Kumar R, Gaudry M, Baud V and Porteu F: Thrombopoietin promotes NHEJ DNA repair in hematopoietic stem cells through specific activation of Erk and NF-κB pathways and their target, IEX-1. Blood 123: 509-519, 2014.
29. Liu Q, Turner KM, Alfred Yung WK, Chen K and Zhang W: Role of AKT signaling in DNA repair and clinical response to cancer therapy. Neuro Oncol 16: 1313-1323, 2014.
30. Sun Y, Zhai L, Ma S, Zhang C, Zhao L, Li N, Xu Y, Zhang T, Guo Z, Zhang H, *et al*: Down-regulation of RIP3 potentiates cisplatin chemoresistance by triggering HSP90-ERK pathway mediated DNA repair in esophageal squamous cell carcinoma. Cancer Lett 418: 97-108, 2018.
31. Rezatabar S, Karimian A, Rameshknia V, Parsian H, Majidinia M, Kopi TA, Bishayee A, Sadeghinia A, Yousefi M, Monirialamdari M and Yousefi B: RAS/MAPK signaling functions in oxidative stress, DNA damage response and cancer progression. J Cell Physiol: 2019 (Online Ahead of Print).
32. Toulany M, Kasten-Pisula U, Brammer I, Wang S, Chen J, Dittmann K, Baumann M, Dikomey E and Rodemann HP: Blockage of epidermal growth factor receptor-phosphatidylinositol 3-kinase-AKT signaling increases radiosensitivity of K-RAS mutated human tumor cells in vitro by affecting DNA repair. Clin Cancer Res 12: 4119-4126, 2006.
33. Kao GD, Jiang Z, Fernandes AM, Gupta AK and Maity A: Inhibition of phosphatidylinositol-3-OH kinase/Akt signaling impairs DNA repair in glioblastoma cells following ionizing radiation. J Biol Chem 282: 21206-21212, 2007.
34. Chen YR, Liu MT, Chang YT, Wu CC, Hu CY and Chen JY: Epstein-Barr virus latent membrane protein 1 represses DNA repair through the PI3K/Akt/FOXO3a pathway in human epithelial cells. J Virol 82: 8124-8137, 2008.
35. Marampon F, Gravina GL, Di Rocco A, Bonfili P, Di Staso M, Fardella C, Polidoro L, Ciccarelli C, Festuccia C, Popov VM, *et al*: MEK/ERK inhibitor U0126 increases the radiosensitivity of rhabdomyosarcoma cells in vitro and in vivo by downregulating growth and DNA repair signals. Mol Cancer Ther 10: 159-168, 2011.
36. Battle A, Brown CD, Engelhardt BE and Montgomery SB; GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)-Analysis Working Group; Statistical Methods groups-Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site-NDRI; Biospecimen Collection Source Site-RPCI, *et al*: Genetic effects on gene expression across human tissues. Nature 550: 204-213, 2017.
37. Gagneur J, Stegle O, Zhu C, Jakob P, Tekkedil MM, Aiyar RS, Schuon AK, Pe'er D and Steinmetz LM: Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. PLoS Genet 9: e1003803, 2013.
38. Umans BD, Battle A and Gilad Y: Where are the disease-associated eQTLs? Trends Genet 37: 109-124, 2021.