**Article**

# Remote regulation of rs80245547 and rs72673891 mediated by transcription factors C-Jun and CREB1 affect *GSTCD* expression



A

**Regulation mechanism of rs80245547 C and rs72673891 A to *GSTCD***

B

**Regulation mechanism of rs80245547 T and rs72673891 G to *GSTCD***

Jin-Xiu Li, Xue-Zhen Huang, Wei-ping Fu, ..., Li Zhong, Li Yu, Ya-ping Zhang

lizhong@snnu.edu.cn (L.Z.)
yuli@ynu.edu.cn (L.Y.)
zhangyp@mail.kiz.ac.cn (Y.-p.Z.)

## Highlights

The low expression of *GSTCD* mRNA is associated with COPD

Rs80245547 and rs72673891 regulate *GSTCD* through remote enhancer-promoter interactions

C-Jun and CREB1 are necessary for rs80245547 and rs72673891 to regulate the *GSTCD* gene

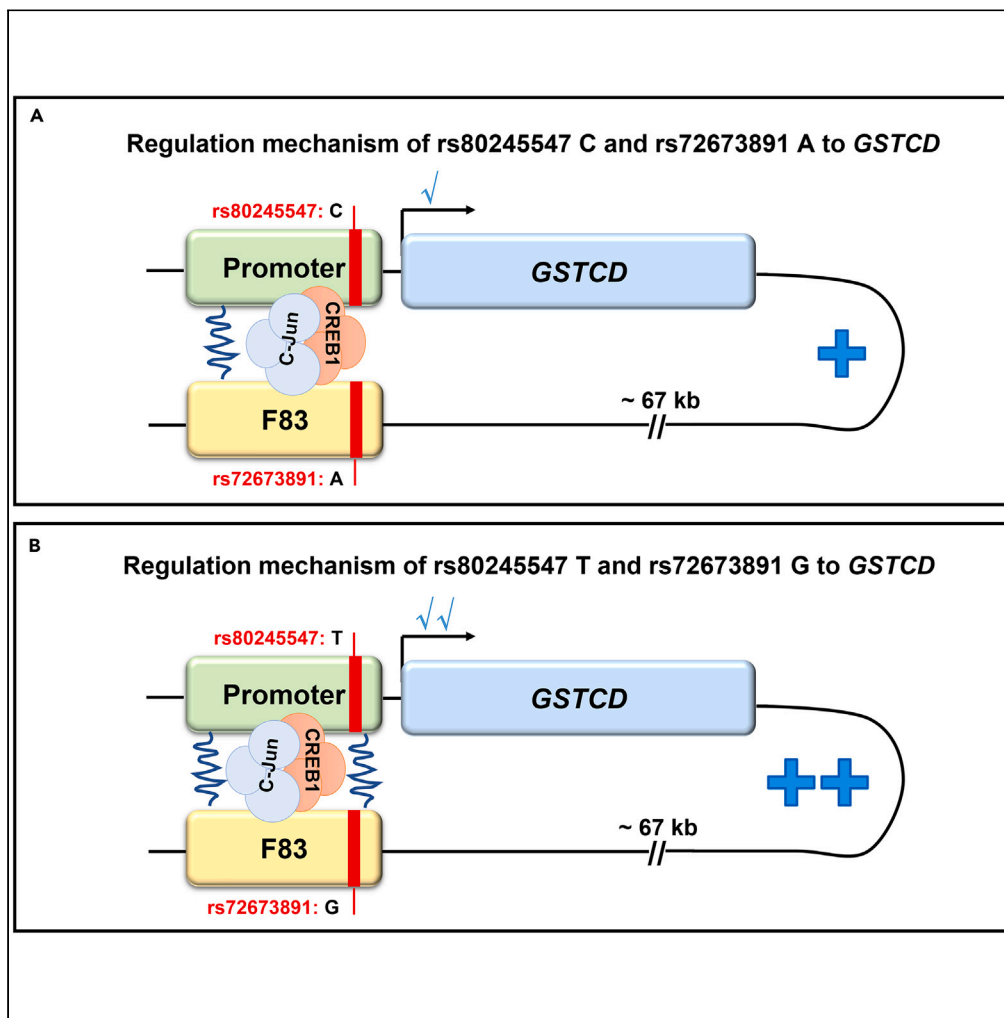Rs80245547T and rs72673891G had a stronger binding ability to C-Jun and CREB1

## Article

# Remote regulation of rs80245547 and rs72673891 mediated by transcription factors C-Jun and CREB1 affect *GSTCD* expression

Jin-Xiu Li,[1,6,7] Xue-Zhen Huang,[1,7] Wei-ping Fu,[2,7] Xiao-hua Zhang,[1] David H. Mauki,[3] Jing Zhang,[2] Chang Sun,[1,4] Lu-Ming Dai,[2] Li Zhong,[1,4,5,*] Li Yu,[1,*] and Ya-ping Zhang[1,6,8,*]

## SUMMARY

**Chronic obstructive pulmonary disease (COPD), the third leading cause of death worldwide, is influenced by genetic factors. The genetic signal rs10516526 in the glutathione S-transferase C-terminal domain containing (*GSTCD*) gene is a highly significant and reproducible signal associated with lung function and COPD on chromosome 4q24. In this study, comprehensive bioinformatics analyses and experimental verifications were detailly implemented to explore the regulation mechanism of rs10516526 and *GSTCD* in COPD. The results suggested that low expression of *GSTCD* was associated with COPD (p = 0.010). And C-Jun and CREB1 transcription factors were found to be essential for the regulation of *GSTCD* by rs80245547 and rs72673891. Moreover, rs80245547T and rs72673 891G had a stronger binding ability to these transcription factors, which may promote the allele-specific long-range enhancer-promoter interactions on *GSTCD*, thus making COPD less susceptible. Our study provides a new insight into the relationship between rs10516526, *GSTCD*, and COPD.**

## INTRODUCTION

As the third leading cause of death worldwide,[1] chronic obstructive pulmonary disease (COPD) kills more than 3 million people each year.[2] COVID-19, which poses a serious threat to human life and health, has also been found to have a higher infection rate in COPD patients,[3,4] suggesting that research on the pathophysiological mechanism of COPD should be strengthened. The discovery of genetic risks associated with diseases could shed light on the pathogenesis and identify possible molecular targets.[5] A genome-wide association study (GWAS) in European populations identified a highly significant and repeatable genetic signal locus, rs10516526 or its linkage loci rs11097901 and/or rs11728716, located in the intron region of *GSTCD* are strongly associated with lung function and COPD,[6–8] suggesting that *GSTCD* might be a new candidate gene for COPD.

*GSTCD* is expressed in structural cells in the respiratory system, including airway smooth muscle cells and bronchial epithelium.[9–11] It has been shown that high expression of *GSTCD* protein and mRNA levels during the pseudoglandular phase promotes the initiation of airway smooth muscle cells, mast cells, T lymphocytes, and dendritic cells in the lung parenchyma,[12] and that low expression of *GSTCD* can affect genes of airway fibrosis and inflammatory pathways.[9] Additionally, it has proven a role in mediating inflammatory signals that affect airway responses in *GSTCD* knockout mice.[13] All these indicate that *GSTCD* is a lung function related gene, whereas the expression of this gene in COPD remains unclear.

Moreover, it has been suggested that intron region mutations play an important role in gene expression regulation,[14–16] and these character-related mutations may regulate multiple genes.[11] It is suggested that since the SNP rs10516526 is in the intron of *GSTCD*, it may be reasonable to hypothesize that it is associated with COPD by the alteration of the *GSTCD* expression. The association between rs10516526 linkage sites and COPD should further be investigated, especially considering the potential regulatory sites in the intron region that were screened by GWAS, as they may be a prospective tag for unreported functional variations.[7]

Herein, a comprehensive analysis of rs10516526 and its linkage sites with *GSTCD* in COPD was performed. The mRNA transcription level of *GSTCD* was detected in lung tissues of the COPD case and control groups

[1]State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan, School of Life Sciences, Yunnan University, Kunming 650000, China

[2]Department of Respiratory Critical Care Medicine, The First Affiliated Hospital of Kunming Medical University, Kunming 650000, China

[3]Faculty of Pharmaceutical Sciences, Institute of Biomedicine and Biotechnology, Center for Cancer Immunology, Chinese Academy of Sciences, Shenzhen Institute of Advanced Technology, Shenzhen 518000, Guangdong China

[4]College of Life Sciences, Shaanxi Normal University, Xi'an 710000, China

[5]Provincial Demonstration Center for Experimental Biology Education, Shaanxi Normal University, Xi'an 710000, China

[6]State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650000, China

[7]These authors contributed equally

[8]Lead contact

*Correspondence: lizhong@snnu.edu.cn (L.Z.), yuli@ynu.edu.cn (L.Y.), zhangyp@mail.kiz.ac.cn (Y.-p.Z.)

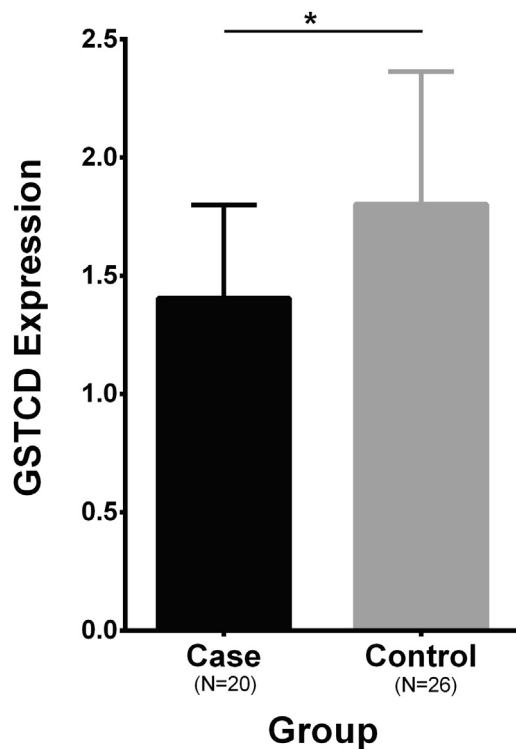https://doi.org/10.1016/j.isci. 2023.107383

**Figure 1. Gene mRNA expression of *GSTCD* in lung tissues of the COPD case and control groups**
Error bars denote SD, N is the number of samples. *p* values were determined by independent t test (*p < 0.05).

in the Han population of South China. The potential functional candidate SNPs were discriminated by exploring the rs10516526 and LD SNPs in the 500 kb region on chromosome 4q24. Furthermore, chromatin conformation capture assay (3C), dual-luciferase reporter assay, chromatin immunoprecipitation (ChIP), RNAi, and electrophoretic mobility shift assay (EMSA) were undertaken to identify the possible regulatory mechanism of candidate SNPs with *GSTCD*. In summary, a clear understanding of how rs10516526 (or LD SNPs) on *GSTCD* affects COPD was investigated.

## RESULTS

### *GSTCD* is associated with COPD at the mRNA level

The mRNA level of *GSTCD* was significantly reduced in COPD subjects compared to control subjects (p = 0.010) using *GAPDH* as reference after adjusting for gender, age, and smoking index, suggesting that *GSTCD* is a new candidate gene for COPD (Figure 1).

### Potential regulatory regions with rs10516526 are discovered

As shown in Figure 2A, there are four genes in this 500 kb region, including *ARHGEF38*, *INTS12*, *GSTCD*, and *NPNT*, and rs10516526 was in the intron of *GSTCD*. And seven PREs with strong regulatory signals were identified by histone marks (H3K4me1, H3K4me3, H3K9ac) and *DNase I* hypersensitive sites in the ENCODE project (Figure 2B). A total of 256 SNPs ($r^2 \geq 0.8$) were found on chromosome 4q24 (Figure 2C), multiple SNPs in LD with rs10516526 were found to overlap in seven PREs. Thus, total 42 overlapping LD SNPs were identified as potential functional sites for further analysis (Table S3). There is an SNP (rs80245547, C>T) in the PRE3 region, which is in the genomic region ~1 kb upstream of the transcription start site for *GSTCD* (Figure 2A).

### Promoter rs80245547 upregulates *GSTCD* gene expression

To verify the regulation of rs80245547 in the promoter region of *GSTCD* gene, promoter-driven luciferase reporter constructs were generated by the insertion of PCR-amplified fragments containing the *GSTCD* promoter into the *MluI* and *BglII* sites of pGL3-basic. As shown in Figure 3, the cloned region showed ~100-fold higher luciferase expression than pGL3-basic and demonstrated the strong promoter activity of this region. Moreover, rs80245547T showed an ~19% increase in promoter activity compared to rs 80245547C (p = 0.007 in the HEL1 cell line, Figures 3A and P = 0.005 in the HLF cell line; Figure 3B).
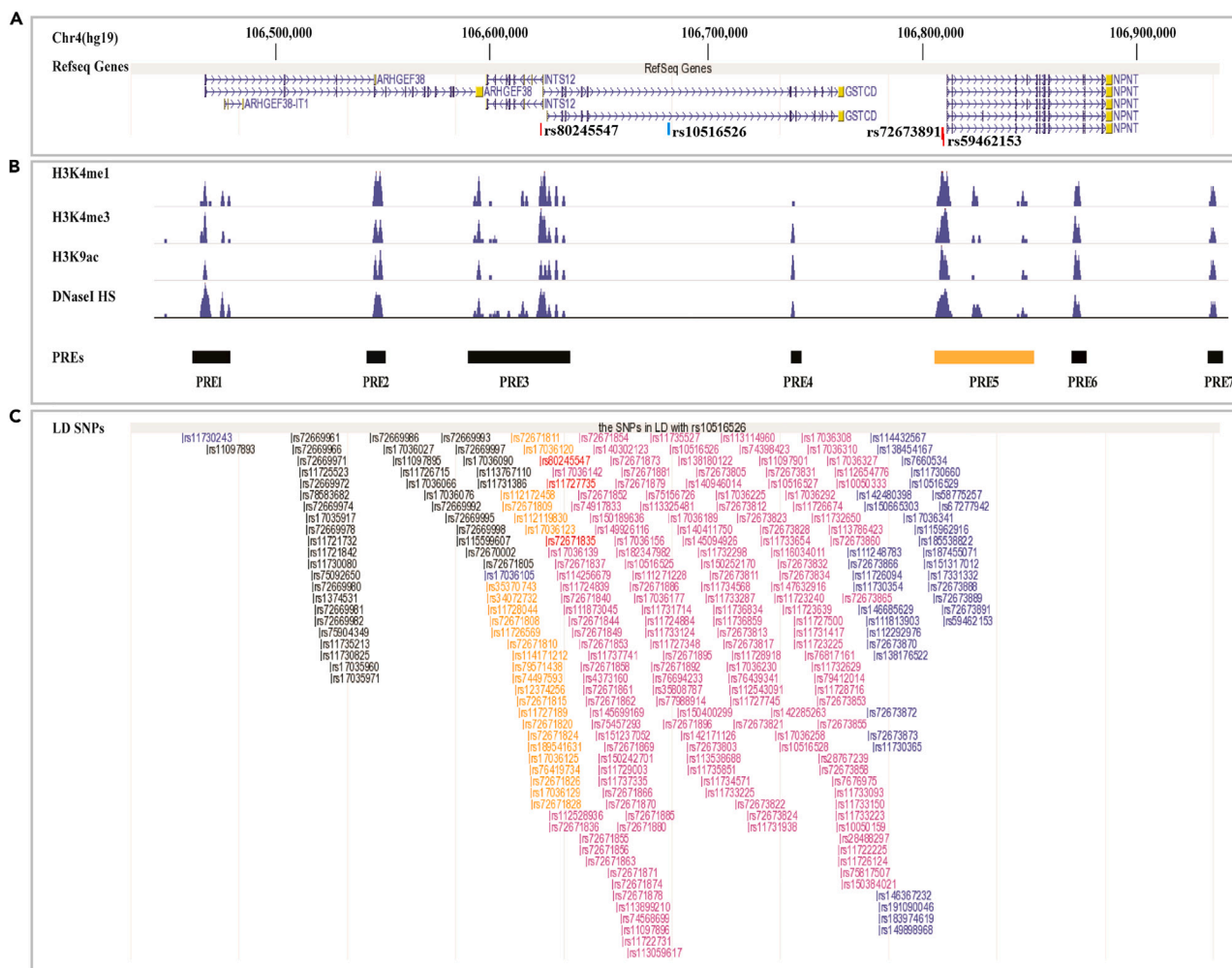
**Figure 2. Risk SNPs LD with rs10516526 and overlapping putative regulatory elements (PREs)**

(A) Genes located within 500 kb of the GWAS sentinel SNP rs10516526, the location of rs10516526 is marked with blue lines, rs80245547, rs72673891, and rs59462153 marked with red lines.

(B) Location of seven candidate PREs with strong regulatory signals were identified by histone marks and DNase I hypersensitive sites by the ENCODE project. H3K4me1 modification is an enhancer biomarker, H3K4me3 modification is a promoter biomarker, and H3K9ac modification represents active gene transcription. DNase I hypersensitive site maps all types of active elements including promoters and enhancers. The orange square represents the PRE region detected in the subsequent experiments, and the black square represents the other six PREs to be detected.

(C) LD SNPs ($r^2 \geq 0.8$) with rs10516526 within the 500 kb region. The blue labeled sites are intergenic SNPs, the black ones are SNPs within the range of the *ARHGEF38* gene, orange ones are SNPs within the range of *the INTS12* gene, peach ones are SNPs within the range of the *GSTCD* gene, and red ones are SNPs in the region where the *INTS12* and *GSTCD* genes overlap.

## Enhancer rs72673891 physically interacts with *GSTCD* promoters

Except for rs80245547, the other 41 SNPs had a considerable linear physical distance with *GSTCD*. To identify potential functional SNPs that interact with *GSTCD* at a distance, 3C assay was performed in HEL1 and HLF cells. In this study, a constant fragment containing the *GSTCD* promoter was interrogated across 29 fragments (Chr4:106,450,000–160,950,000, hg19) of COPD-associated GWAS sites (primer sequences are listed in Table S3). The interaction frequencies of any given two fragments were calculated by semiquantitative PCR with three independent repeated verification, and ligation of distal chromatin segments was verified by sequencing 3C-PCR products (data not shown). The results showed that there was an F83 fragment (~3 kb) with high signal at ~67 kb downstream of *GSTCD*, could stably interact with the *GSTCD* promoter (Figures 4A and S1), which may be a potential enhancer regulatory element of *GSTCD* gene. Notably, the F83 fragment contains two SNPs linked to rs10516526: rs72673891 (A>G) and rs59462153 (T>C) (Figure 4B).
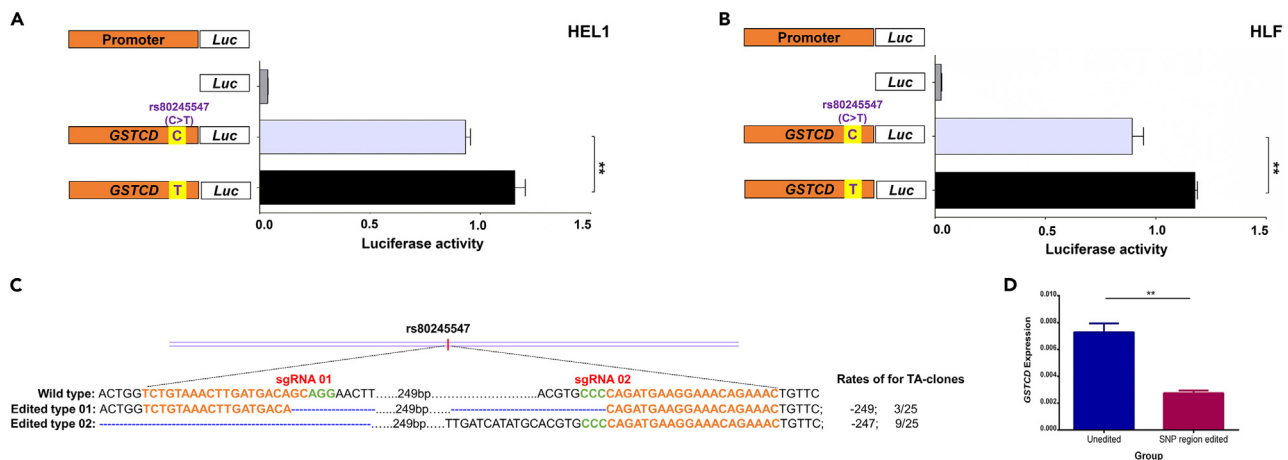
**Figure 3. Regulation of promoter rs80245547 and *GSTCD***

Luciferase reporter assays following transient transfection of HEL1 (A) and HLF (B) cell lines. The *GSTCD* promoter (orange stripe) containing rs80245547 (yellow stripe) was cloned into a PGL3-basic vector. The different alleles of rs80245547 are indicated in purple font. All shown are ratios of firefly luciferase expression to Renilla luciferase expression.

(C) Two sgRNA targeting rs80245547 region and TA clones Sanger sequencing results of modified cells with rates. Protospacers are in red text, PAMs are in green text. Deletion mutations (−) are shown in blue.

(D) Detection of *GSTCD* expression after editing rs80245547 region. Unedited represents the expression of *GSTCD* without editing, and SNP region edited represents the expression of *GSTCD* after editing in rs80245547. Geometric means and standard errors were calculated from six independent repeats. Error bars denote SD. *p* values were determined by independent t test (**$p < 0.01$).

Furthermore, to verify the regulatory effect of rs72673891 and rs59462153 in the F83 fragment on the *GSTCD* gene, a 1.1 kb PRE fragment containing rs72673891 and rs59462153 was cloned into the 5′ end of the *GSTCD* promoter with single- or double-mutation construct haplotypes in forward and reverse orientations. In the analysis of forwarding insertion of this 1.1 kb PRE fragment, rs72673891G showed significant enhancer activity compared with the *GSTCD* promoter ($p < 0.01$) in HEL1 cells, while rs59462153 (T or C) showed quantitatively modest enhancer activity ($p < 0.05$), as shown in Figure 5A. Moreover, double mutations at both SNPs (rs72673861G and rs59462153C) consistently showed an approximately significant increase in enhancer activity compared with the *GSTCD* promoter ($p = 0.024$), and this increase was mainly influenced by rs72673891G. Similar results were obtained in the analysis of reverse insertion of the segment (Figure 5C), and a consistent trend was observed in HLF cells (Figures 5B and 5D).

### SNPs region plays an important role in the expression of *GSTCD*

As shown in Figure 3C, the 249 bp region around rs80245547 in HEL1 cells was designed as the knockout region, and RT-PCR results showed that when the knockout efficiency of this region was 48.00% (Figure 3C), the expression of *GSTCD* gene could be significantly downregulated ($p = 1.578E-05$, Figure 3D). Similarly, 121 bp near rs72673891 was used as the knockout region (Figure 5E). When 15.38% of this region was knocked out (Figure 5E), the expression of *GSTCD* gene was numerically different with no significant ($p = 0.055$, Figure 5F). The results showed that rs80245547 and rs72673891 regions play a role in the expression of *GSTCD* gene.

### Transcription factors C-Jun and CREB1 is essential for regulating *GSTCD* expression

Interestingly, rs80245547 could bind both C-Jun and CREB1 transcription factors based on ENCODE ChIP-seq and JASPAR data, as could rs72673891, but no transcription factor binding including C-Jun and CREB1 was found near rs59462153 (Figure S2). Therefore, we first performed ChIP analysis on rs80245547 and rs72673891, and temporarily ignored the research on rs59462153. As shown in Figure 6, the antibody for C-Jun and CREB1 immunoprecipitated chromatin samples was significantly enriched for the region surrounding rs80245547 (Figure 6A, $P_{CREB1} = 0.014$, $P_{C-Jun} = 0.002$) and rs72673891 (Figure 6B, $P_{CREB1} = 0.032$, $P_{C-Jun} = 0.005$) compared with IgG in HEL1 cells.

Subsequent siRNA co-transfer analysis showed that after knockdown of transcription factors C-Jun and CREB1 in HEL1 cells (Figure 7A, $P_{CREB1} = 2.53E-05$, $P_{C-Jun} = 0.015$), and the expression level of *GSTCD*
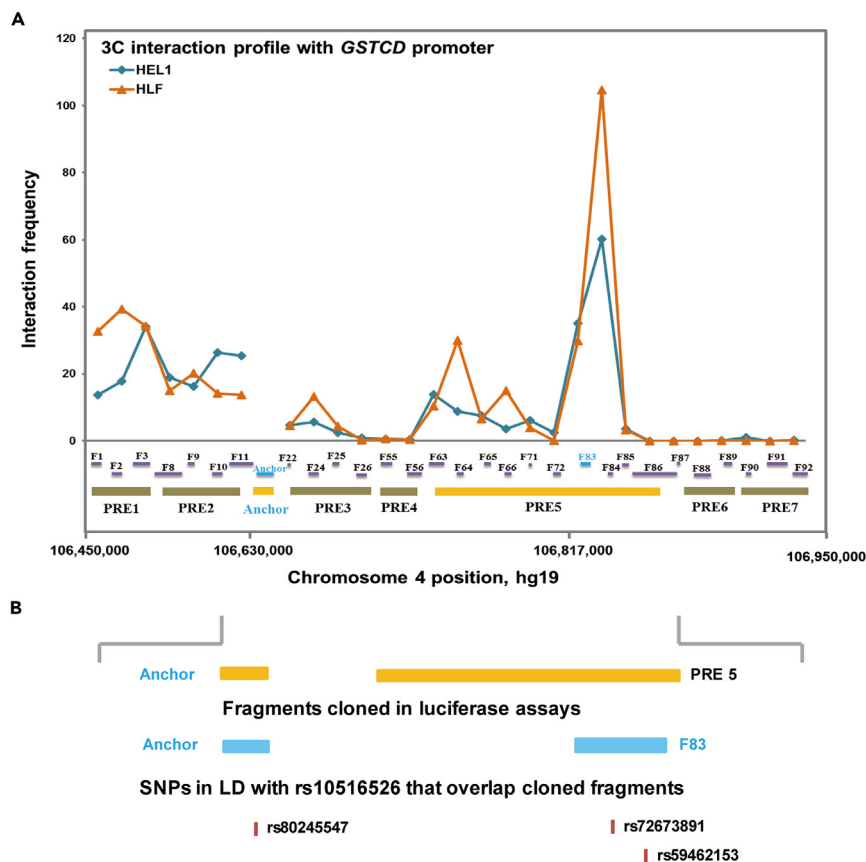
**Figure 4. Physical interaction of chromatin at the 4q24 risk region with the *GSTCD* promoter**

(A) 3C interaction profiles indicate frequent interactions between PRE regions and the *GSTCD* promoter region (Anchor). 3C libraries were generated with *HindIII*, with the anchor point set at the *GSTCD* promoter. A physical map of the region interrogated by 3C is shown below, the abscissa represents the physical map of the chromosome segments in the 3C region, and the vertical axis represents the regional interaction frequency. The orange stripe shows the detected interaction in the PRE region, the grass green stripe shows the remaining PRE regions to be detected, the blue line represents the interacting *Hind III* enzyme fragment within the detected PRE region, and the purple line represents other *Hind III* enzyme fragments to be tested.

(B) Location of the 3C interacting fragments, the two fragments (anchor and F83 fragment, shown in blue) cloned in the luciferase assays, and the SNPs that are in LD ($r^2 \geq 0.8$) with rs10516526 overlap with the cloned fragments.

gene was significantly decreased (Figure 7B, $P_{CREB1}$ = 4.09E-03, $P_{C-Jun}$ = 1.48E-03). When siRNA and expression vector co-transfected, there was no significant difference in the regulation of alleles of rs80245547 in the promoter region of *GSTCD* (Figures 7C and 7D, p > 0.05). Enhancer at rs72673891 was also disappeared after knockdown of transcription factors C-Jun (Figures 7E and 7F) and CREB1 (Figures 7G and 7H), there was no significant difference between enhancer fragment (F83) and *GSTCD* promoter (p > 0.05), or the expression level was significantly lower than *GSTCD* promoter (p < 0.01). These transcription factors are important for the regulation of *GSTCD* expression by rs80245547 and rs72673891.

Then EMSA was performed, and it was found that in rs80245547, when annealed probes of the two alleles were mixed with nuclear extract from HEL1 cells, a nucleoprotein complex was observed (arrow, Figure 8A), and there was an increasing intensity of rs80245547T probe allele than rs80245547C (Figure 8A, lane3, and lane4). To verify the specificity of the probe-protein bands, specific competitive probes (NC, NT, and MT) were added, and observed that the two bands of the nucleoprotein complex were specific to rs80245547 (Figure 8A, lanes 5–8). However, when the C-Jun antibody was added, we noted a reduction in the nucleoprotein complex, even though the supershifted band of the nucleoprotein-antibody complex was not clearly discerned (Figure 8A, lane9, and lane10). Of notice, similar finding was observed with rs80245547 and CREB1 (Figure 8B).
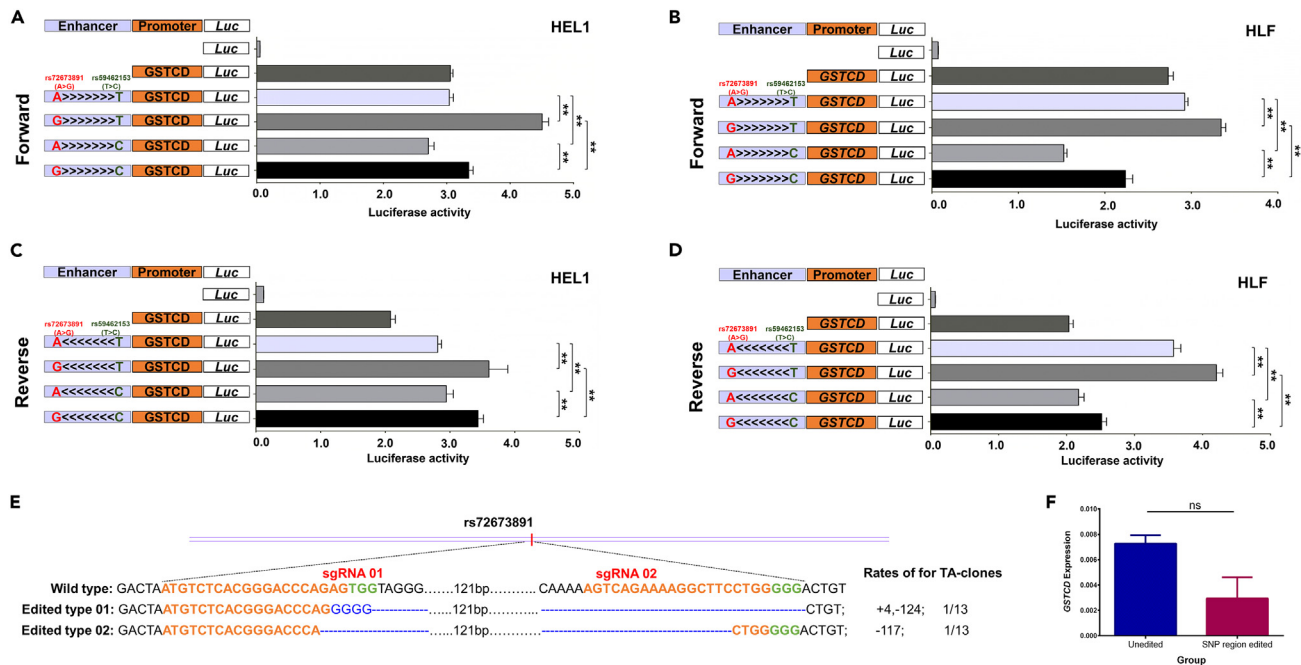
**Figure 5. Regulation of enhancer rs72673891 and *GSTCD***

Luciferase reporter assays following transient transfection of HEL1 (A and C) and HLF (B and D) cell lines. Fragment 83 (F83, ~1.1 kb around two key SNPs (rs72673891 and rs59462153), light purple stripe) was cloned in the forwarding orientation (≫>) and reverse orientation (≪<) and inserted upstream of the *GSTCD* promoter ("control," orange). Single-nucleotide alterations were introduced individually or combined into minimal enhancer constructs at forwarding (A and B) and reverse (C and D) orientations. The different alleles of rs72673891 are indicated in red font, and the different alleles for rs59462153 are indicated in green font. All shown are ratios of firefly luciferase expression to Renilla luciferase expression.

(E) Two sgRNA targeting rs72673891 region and TA clones Sanger sequencing results of modified cells with rates. Protospacers are in red text, PAMs are in green text. Mutations are in blue, and insertions with blue text, deletions (−).

(F) Detection of *GSTCD* expression after editing rs72673891 region. Unedited represents the expression of *GSTCD* without editing, and SNP region edited represents the expression of *GSTCD* after editing in rs72673891. Geometric means and standard errors were calculated from six independent repeats. Error bars denote SD. $p$ values were determined by independent t test (**$p < 0.01$).

For rs72673891 with C-Jun, nuclear protein complexes from HEL1 cells could bind to rs72673891, and the rs72673891G allele had better binding ability than rs72673891A (as shown in Figure 8C, lane 3, and lane 4). Moreover, the two bands were specific, as shown by complete elimination with an identical competitor (Figure 8C, lanes 5 and 6) or retention of the binding complex with the addition of a nonidentical competitor (Figure 8C, lanes 7 and 8). Furthermore, by adding a specific antibody of C-Jun against the probe-nucleoprotein binding complex, a supershift band was observed (Figure 8C, lanes 9 and 10), indicating that C-Jun was in the nucleoprotein complex, and the binding ability to C-Jun of the rs72673891G allele was stronger than that of rs72673891A. Similar result was also obtained in rs72673891 with CREB1, as shown in Figure 8D.

## DISCUSSION

As the third leading cause of death in the world,[1] COPD is mostly under investigation, and its genetic effects have been extensively explored. Recent studies reported that rs10516526 was significantly associated with lung function and COPD.[6–8,17] However, the relationship between rs10516526 and *GSTCD* is not fully understood. In this study, a comprehensive evaluation, including RT-PCR, 3C, ChIP, RNAi, and EMSA, was performed. The association of *GSTCD* with COPD was first demonstrated in the Han population of China, and the regulatory mechanisms of rs10516526 and *GSTCD* were elucidated for the first time. Two LD SNPs (rs80245547T and rs72673891G) with rs10516526 upregulated the expression of *GSTCD* by remotely co-binding the transcription factors C-Jun and CREB1, resulting in the protection of the population against COPD.

Although studies have indicated that *GSTCD* is expressed in structural cells in the respiratory system,[10,11] and several studies have analyzed the function of *GSTCD*,[9,12] which revealing the relationship between *GSTCD* and COPD is the principle of direct functional analysis. In the present study , we reported for
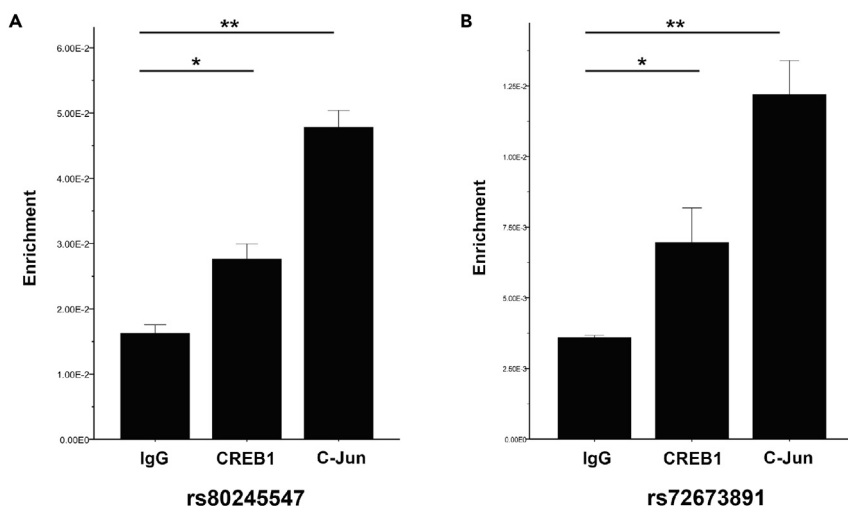
**Figure 6. Transcription factors and the SNPs rs80245547 and rs72673891**

The plot displays the binding of the transcriptions factors on rs80245547 (A) and rs72673891 (B) in HEL1 cell lines. ChIP assays were performed with CREB1 or C-Jun antibody, and nonimmune IgG served as a negative control (control). The abscissa represents the target protein detected, and the ordinate represents the relative amount of target DNA fragments enriched by the ChIP experiment. Graphs represent three biological replicates. Error bars denote SD. $p$ values were determined with a two-tailed $t$ test (**$p < 0.01$).

the first time that reduced *GSTCD* gene transcription is associated with COPD (Figure 1). However, there was a modest sample size in our study, and we therefore recommend for a much larger sample size and diversified group of populations for deeper investigation regarding this hypothesis.

Currently, it is convenient and powerful to discriminate candidate SNPs with COPD by leveraging the already publicly datasets from 1000 Genomes, ENCODE, and GTEx projects.[18,19] We attempted to identify the potential functional SNPs of COPD, by combining the public datasets and the findings emanated through the 3C assay. SNPs rs80245547 and rs72673891 that high LD with rs10516526 were discriminated successfully in our study (Figures 4 and S1). Rs80245547 was in the promoter region of *GSTCD*, and rs72673891 were located ~67 kb downstream of *GSTCD*. We also found some other small-frequency interaction fragments including F3 and F63 in the 3C experiment, which may be of research interest to others, and it is necessary to investigate the function of these fragments in the future.

Functional experiments and GTEx database analysis further confirmed that rs80245547 and rs72673891 were involved in regulating the expression of *GSTCD* gene (Figures 3 and 5). It is worth noting that rs80245547T also upregulated *GSTCD* gene expression in nerve-tibial tissues (Figure S3A) and skin sun-exposed tissues (Figure S3B) that were accessed in the Genotype-Tissue Expression (GTEx) Project[20] (https://www.gtexportal.org/home/) We also observed that, rs72673891G increased the expression of *GSTCD* in adipose-subcutaneous tissues (Figure S3C) and skin-sun-exposed tissues (Figure S3D) obtained from the GTEx project.

Recent studies have shown that transcription factors often have a significant role in the regulation and stabilization of chromatin loops and gene transcription.[21–23] It is reasonable to speculate that C-Jun and CREB1 transcription factors binding to rs80245547 and rs72673891 could stabilize the enhancer-promoter loop and the regulation of *GSTCD* gene transcription. When we knocked down C-Jun and CREB1 transcription factors, the expression of *GSTCD* was indeed affected (Figure 7). Moreover, rs80245547T and rs72673891G could upregulate the gene expression of *GSTCD* by binding C-Jun and CREB1 (Figure 8). It is particular concern that C-Jun[24,25] or CREB1[26–28] could also enhance the activity of gene promoters, combine with transcription factors, and mediate long-range enhancer-promoter interactions.[21] Interestingly, C-Jun and CREB1 are also the key transcription factors involved in innate immunity and inflammation,[29,30] which is an important pathophysiological process in COPD.

In conclusion, our findings first confirmed that *GSTCD* was a candidate gene for COPD and further revealed that the dual SNP (rs80245547T and rs72673891G) could remotely upregulate the expression of *GSTCD* by
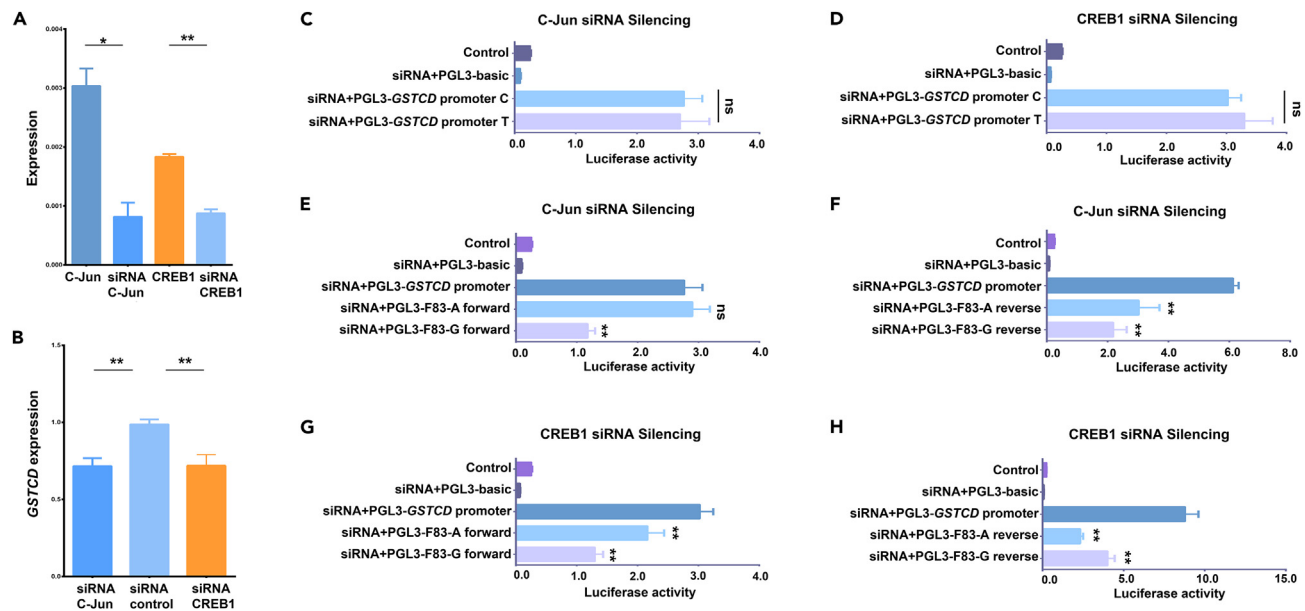
**Figure 7. Transcription factors C-Jun and CREB1 affect the regulation of rs80245547 and rs72673891 on *GSTCD***

(A) The expression of transcription factors C-Jun and CREB1 after siRNA silencing. C-Jun and CREB1 represent the expression of transcription factors in HEL1 cells interfered with siRNA control (siRNA contains a scrambled sequence conjugated), siRNA C-Jun and siRNA CREB1 represent the expression of C-Jun and CREB1 in HEL1 cells interfered with silencer siRNA (siRNAs against C-Jun or CREB1).

(B) The expression of *GSTCD* gene after siRNA silencing. siRNA control represents the expression of *GSTCD* gene in HEL1 cells after siRNA control interference, and siRNA C-Jun and siRNA CREB1 represents the expression of *GSTCD* gene in HEL1 cells interfered with silencer siRNA (siRNAs against C-Jun or CREB1).

(C–H) Dual-luciferase reporter analysis of the effect of C-Jun and/or CREB1 on regulation of rs80245547 in *GSTCD* promoter (C–D) and rs72673891 in enhancer fragment (F83) on *GSTCD* (E–H). Control represents HEL1 cells transfected with siRNA control. siRNA+PGL3-basic indicates the co-transfection of silencer siRNA and PGL3-basic. siRNA+PGL3-*GSTCD* promoter C represents the co-transfection of silencer siRNA and PGL3-*GSTCD* promoter containing rs80245547 C alleles. siRNA+PGL3-*GSTCD* promoter T represents the co-transfection of silencer siRNA with PGL3-*GSTCD* promoter containing rs80245547 T alleles. siRNA+PGL3-*GSTCD* promoter represents the co-transfection of silencer siRNA and PGL3-*GSTCD* promoter. siRNA+PGL3-F83-A indicates co-transfection of the silencer siRNA and an enhancer fragment (F83) containing the A allele of rs72673891, siRNA+PGL3-F83-G indicates co-transfection of the silencer siRNA and an enhancer fragment containing the G allele of rs72673891, forward or reverse represents the insertion orientation of the enhancer fragment (F83) containing rs72673891 in the upstream of the *GSTCD* promoter. Geometric means and standard errors were calculated from six independent repeats. Error bars denote SD, *p* values were determined by independent t test (**p < 0.01, ns for no significant difference).

binding C-Jun and CREB1, thus reducing susceptibility to COPD. These findings will help to reveal the genotype-phonotype associations on COPD function in the post-GWAS stage.

## Limitations of the study

If more extensive studies of *GSTCD* in structural cells of respiratory system can be carried out, it will be helpful for us to further understand the function of *GSTCD* gene in COPD.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human tissues and cell culture
- METHOD DETAILS
  - RNA isolation and RT-RCR
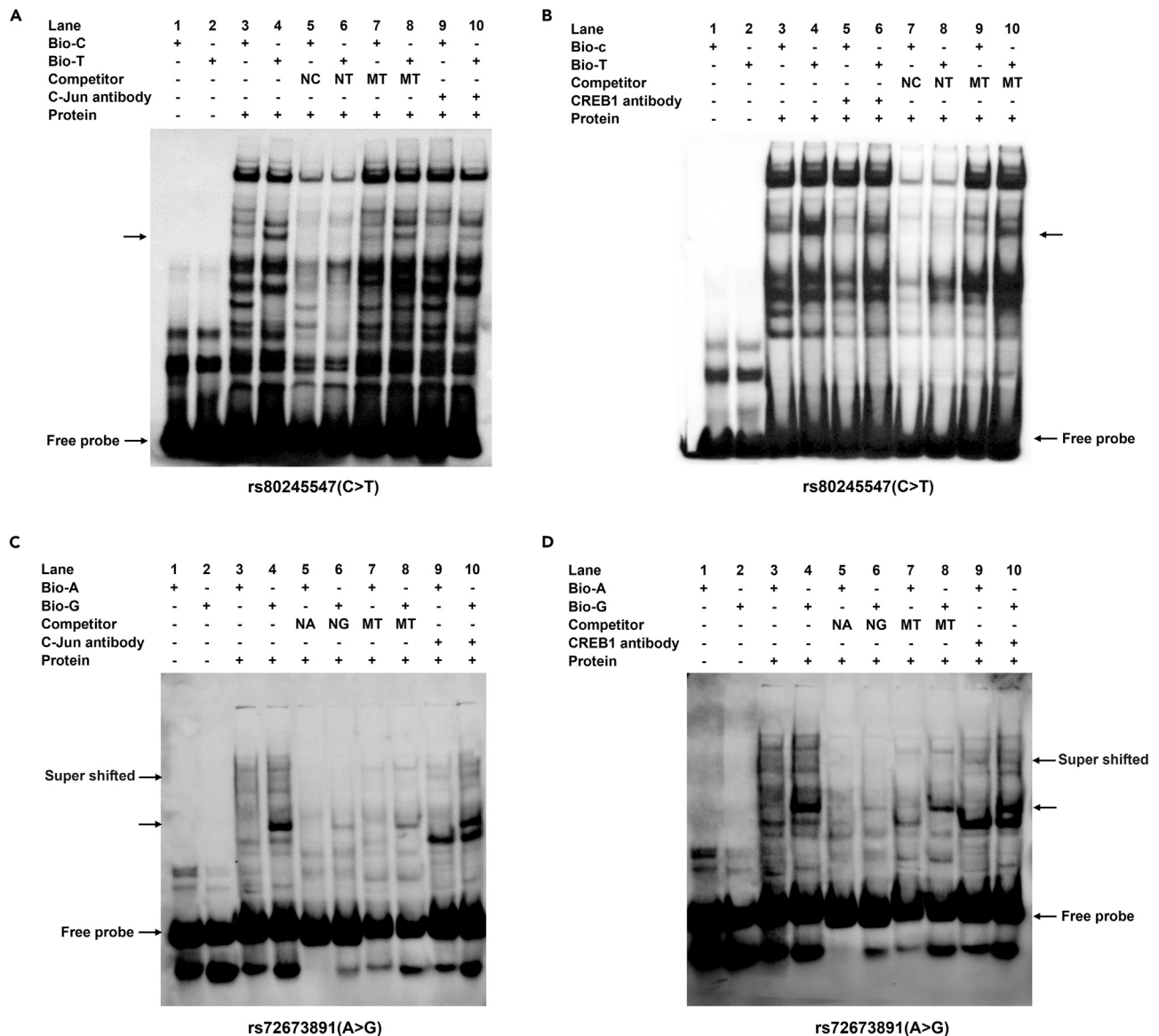  - Bioinformatic analysis

**Figure 8. Allele-specific binding of C-Jun and CREB1**

(A) and (B) Biotin-labeled oligonucleotides containing the sequence spanning rs80245547 with either C or T were incubated with nuclear extract from HEL1 cells. In Lanes 1 and 2, only probe C or T was added. In lanes 3 and 4, nuclear extract and probe C or T were added. In Lanes 5–10, based on the reaction of lanes 3 and 4, other samples to be tested were added. (A) rs80245547 alters a C-Jun binding site. Probes were incubated with identical competitor probe NC in lane 5, the NT probe in lane 6, and the 8 bp mutant probe (MT) in lanes 7 and 8. C-Jun antibodies were added to lanes 9 and 10. Arrows indicate specific nuclear proteins bound by the probe. (B) rs80245547 alters a CREB1 binding site. Probes were incubated with identical competitor probe NC in lane 7, the NT probe in lane 8, and the 8 bp mutant probe (MT) in lanes 9 and 10. CREB1 antibodies were added to lanes 5 and 6. Arrows indicate specific nuclear proteins bound by the probe.

(C) and (D) Biotin-labeled oligonucleotides containing the sequence spanning rs72673891 with either A or G were incubated with nuclear extracts from HEL1 cells. In Lanes 1 and 2, only probe A or G was added. In lanes 3 and 4, nuclear extract and probes A or G were added. In Lanes 5–10, based on the reaction of lanes 3 and 4, other samples to be tested were added. (C) rs72673891 alters a C-Jun binding site. Probes were incubated with identical competitor probe NC in lane 5, the NT probe in lane 6, and the 8 bp mutant probe (MT) in lanes 7 and 8. C-Jun antibodies were added to lanes 9 and 10. The specific nuclear proteins bound by the probe and supershifted C-Jun band in lanes 9 and 10 are indicated by arrows.

(D) rs72673891 alters a CREB1 binding site. Probes were incubated with identical competitor probe NC in lane 5, the NT probe in lane 6, and the 8 bp mutant probe (MT) in lanes 7 and 8. CREB1 antibodies were added to lanes 9 and 10. The specific nuclear proteins bound by the probe and supershifted CREB1 band in lanes 9 and 10 are indicated by arrows.

- ○ Chromatin conformation capture assay
- ○ Molecular cloning and dual-luciferase reporter assays
- ○ CRISPR Cas9 - Mediated deletion of SNP region in HEL1 cells
- ○ Transcription factor binding site prediction
- ○ Chromatin immunoprecipitation assays
- ○ RNAi for C-Jun and CREB1
- ○ Electrophoretic mobility shift assays
- ● QUANTIFICATION AND STATISTICAL ANALYSIS
- ○ Statistical analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107383.

## AUTHOR CONTRIBUTIONS

L.Z., L.Y., and Y-P.Z. conceived the study; J-X.L., X-Z.H., and X-H.Z. carried out the experimental operation in this study, J-X. L. and W-P. F. analyzed the experimental data, J.Z. completed the collection of tissue samples and their information, C.S. and L-M.D. provided suggestions for this study, J-X.L. and W-P.F. prepared and wrote the manuscript, D.H.M. revised the manuscript, and all authors read and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

1. Rennard, S.I., and Drummond, M.B. (2015). Early chronic obstructive pulmonary disease: definition, assessment, and prevention. Lancet *385*, 1778–1788. https://doi.org/10.1016/S0140-6736(15)60647-X.

2. Rabe, K.F., and Watz, H. (2017). Chronic obstructive pulmonary disease. Lancet *389*, 1931–1940. https://doi.org/10.1016/S0140-6736(17)31222-9.

3. Grasselli, G., Greco, M., Zanella, A., Albano, G., Antonelli, M., Bellani, G., Bonanomi, E., Cabrini, L., Carlesso, E., Castelli, G., et al. (2020). Risk Factors Associated With Mortality Among Patients With COVID-19 in Intensive Care Units in Lombardy, Italy. JAMA Intern. Med. *180*, 1345–1355. https://doi.org/10.1001/jamainternmed.2020.3539.

4. Lippi, G., and Henry, B.M. (2020). Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). Respir. Med. *167*, 105941. https://doi.org/10.1016/j.rmed.2020.105941.

5. Zhao, X., Qiao, D., Yang, C., Kasela, S., Kim, W., Ma, Y., Shrine, N., Batini, C., Sofer, T., Taliun, S.A.G., et al. (2020). Whole genome sequence analysis of pulmonary function and COPD in 19,996 multi-ethnic participants. Nat. Commun. *11*, 5182. https://doi.org/10.1038/s41467-020-18334-7.

6. Castaldi, P.J., Cho, M.H., Litonjua, A.A., Bakke, P., Gulsvik, A., Lomas, D.A., Anderson, W., Beaty, T.H., Hokanson, J.E., Crapo, J.D., et al. (2011). The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. Am. J. Respir. Cell Mol. Biol. *45*, 1147–1153. https://doi.org/10.1165/rcmb.2011-0055OC.

7. Hancock, D.B., Eijgelsheim, M., Wilk, J.B., Gharib, S.A., Loehr, L.R., Marciante, K.D., Franceschini, N., van Durme, Y.M.T.A., Chen, T.H., Barr, R.G., et al. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat. Genet. *42*, 45–52. https://doi.org/10.1038/ng.500.

8. Repapi, E., Sayers, I., Wain, L.V., Burton, P.R., Johnson, T., Obeidat, M., Zhao, J.H., Ramasamy, A., Zhai, G., Vitart, V., et al. (2010). Genome-wide association study identifies five loci associated with lung function. Nat. Genet. *42*, 36–44. https://doi.org/10.1038/ng.501.

9. Henry, A.P., Probert, K., Stewart, C.E., Thakker, D., Bhaker, S., Azimi, S., Hall, I.P., and Sayers, I. (2019). Defining a role for lung function associated gene GSTCD in cell homeostasis. Respir. Res. *20*, 172. https://doi.org/10.1186/s12931-019-1146-3.

10. Kheirallah, A.K., de Moor, C.H., Faiz, A., Sayers, I., and Hall, I.P. (2017). Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways. BMC Genom. *18*, 248. https://doi.org/10.1186/s12864-017-3628-3.

11. Obeidat, M., Miller, S., Probert, K., Billington, C.K., Henry, A.P., Hodge, E., Nelson, C.P., Stewart, C.E., Swan, C., Wain, L.V., et al.

(2013). GSTCD and INTS12 regulation and expression in the human lung. PLoS One *8*, e74630. https://doi.org/10.1371/journal.pone.0074630.

12. Melén, E., Kho, A.T., Sharma, S., Gaedigk, R., Leeder, J.S., Mariani, T.J., Carey, V.J., Weiss, S.T., and Tantisira, K.G. (2011). Expression analysis of asthma candidate genes during human and murine lung development. Respir. Res. *12*, 86. https://doi.org/10.1186/1465-9921-12-86.

13. Liu, B., Henry, A.P., Azimi, S., Miller, S., Lee, F.K., Lee, J.C., Probert, K., Kotlikoff, M.I., Sayers, I., and Hall, I.P. (2019). Exposure to lipopolysaccharide (LPS) reduces contractile response of small airways from GSTCD(-/-) mice. PLoS One *14*, e0221899. https://doi.org/10.1371/journal.pone.0221899.

14. Jacob, A.G., and Smith, C.W.J. (2017). Intron retention as a component of regulated gene expression programs. Hum. Genet. *136*, 1043–1057. https://doi.org/10.1007/s00439-017-1791-x.

15. Shaul, O. (2017). How introns enhance gene expression. Int. J. Biochem. Cell Biol. *91*, 145–155. https://doi.org/10.1016/j.biocel.2017.06.016.

16. Vaz-Drago, R., Custódio, N., and Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. Hum. Genet. *136*, 1093–1111. https://doi.org/10.1007/s00439-017-1809-4.

17. Wain, L.V., Shrine, N., Miller, S., Jackson, V.E., Ntalla, I., Soler Artigas, M., Billington, C.K., Kheirallah, A.K., Allen, R., Cook, J.P., et al. (2015). Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir. Med. *3*, 769–781. https://doi.org/10.1016/S2213-2600(15)00283-0.

18. Soler Artigas, M., Wain, L.V., Miller, S., Kheirallah, A.K., Huffman, J.E., Ntalla, I., Shrine, N., Obeidat, M., Trochet, H., McArdle, W.L., et al. (2015). Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nat. Commun. *6*, 8658. https://doi.org/10.1038/ncomms9658.

19. Obeidat, M., Hao, K., Bossé, Y., Nickle, D.C., Nie, Y., Postma, D.S., Laviolette, M., Sandford, A.J., Daley, D.D., Hogg, J.C., et al. (2015). Molecular mechanisms underlying variations in lung function: a systems genetics analysis. Lancet Respir. Med. *3*, 782–795.

20. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660. https://doi.org/10.1126/science.1262110.

21. Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. Proc. Natl. Acad. Sci. USA *106*, 20222–20227. https://doi.org/10.1073/pnas.0902454106.

22. Visser, M., Palstra, R.J., and Kayser, M. (2015). Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. Hum. Mol. Genet. *24*, 2649–2661. https://doi.org/10.1093/hmg/ddv029.

23. Zhou, X., Baron, R.M., Hardin, M., Cho, M.H., Zielinski, J., Hawrylkiewicz, I., Sliwinski, P., Hersh, C.P., Mancini, J.D., Lu, K., et al. (2012). Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Hum. Mol. Genet. *21*, 1325–1335. https://doi.org/10.1093/hmg/ddr569.

24. Qian, X.X., Mata-Greenwood, E., Liao, W.X., Zhang, H., Zheng, J., and Chen, D.B. (2007). Transcriptional regulation of endothelial nitric oxide synthase expression in uterine artery endothelial cells by c-Jun/AP-1. Mol. Cell. Endocrinol. *279*, 39–51. https://doi.org/10.1016/j.mce.2007.08.017.

25. Zhong, J.T., Wang, H.J., Yu, J., Zhang, J.H., Wang, S.F., Yang, X., and Su, W. (2017). Correlations of the expressions of c-Jun and Egr-1 proteins with clinicopathological features and prognosis of patients with nasopharyngeal carcinoma. Cancer Biomarkers *19*, 213–220. https://doi.org/10.3233/Cbm-161710.

26. Lv, X., Wang, D., Ma, Y., and Long, Z. (2018). Analysis of the oncogene BRAF mutation and the correlation of the expression of wild-type BRAF and CREB1 in endometriosis. Int. J. Mol. Med. *41*, 1349–1356. https://doi.org/10.3892/ijmm.2017.3342.

27. Popov, D.V., Lysenko, E.A., Vepkhvadze, T.F., Kurochkina, N.S., Maknovskii, P.A., and Vinogradova, O.L. (2015). Promoter-specific regulation of PPARGC1A gene expression in human skeletal muscle. J. Mol. Endocrinol. *55*, 159–168. https://doi.org/10.1530/JME-15-0150.

28. Zhang, M., Xu, J.J., Zhou, R.L., and Zhang, Q.Y. (2013). cAMP Responsive Element Binding Protein-1 Is a Transcription Factor of Lysosomal-Associated Protein Transmembrane-4 beta in Human Breast Cancer Cells. PLoS One *8*, e57520. https://doi.org/10.1371/journal.pone.0057520.

29. Hu, X., Paik, P.K., Chen, J., Yarilina, A., Kockeritz, L., Lu, T.T., Woodgett, J.R., and Ivashkiv, L.B. (2006). IFN-gamma suppresses IL-10 production and synergizes with TLR2 by regulating GSK3 and CREB/AP-1 proteins. Immunity *24*, 563–574. https://doi.org/10.1016/j.immuni.2006.02.014.

30. Patil, R.H., Naveen Kumar, M., Kiran Kumar, K.M., Nagesh, R., Kavya, K., Babu, R.L., Ramesh, G.T., and Chidananda Sharma, S. (2018). Dexamethasone inhibits inflammatory response via down regulation of AP-1 transcription factor in human lung epithelial cells. Gene *645*, 85–94. https://doi.org/10.1016/j.gene.2017.12.024.

31. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073. https://doi.org/10.1038/nature09534.

32. ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

33. Pazin, M.J. (2015). Using the ENCODE Resource for Functional Annotation of Genetic Variants. Cold Spring Harb. Protoc. *2015*, 522–536. https://doi.org/10.1101/pdb.top084988.

34. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910. https://doi.org/10.1016/j.cell.2015.07.038.

35. Li, J.X., Fu, W.P., Zhang, J., Zhang, X.H., Sun, C., Dai, L.M., Zhong, L., Yu, L., and Zhang, Y.P. (2018). A functional SNP upstream of the ADRB2 gene is associated with COPD. Int J Chronic Obstr *13*, 917–925. https://doi.org/10.2147/Copd.S151153.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Antibodies** | | |
| Mouse monoclonal anti-c-Jun Antibody (G-4)X | Santa Cruz | Cat#sc-74543 X; RRID: AB_1121646 |
| Mouse monoclonal anti-CREB-1 Antibody (X-12)X | Santa Cruz | Cat#sc-240 X; RRID: AB_627302 |
| **Bacterial and virus strains** | | |
| BAC clones 525F1 | Thermo Scientific | RPCI11.C |
| BAC clones 1041L12 | Thermo Scientific | RPCI11.C |
| BAC clones 277H20 | Thermo Scientific | RPCI11.C |
| pGL3-Basic Vector | Promega Corporation | E1751 |
| pRL-TK Vector | Promega Corporation | E2241 |
| pSpCas9(BB)-2A-GFP (PX458) | Addgene | 48138 |
| **Biological samples** | | |
| Human lung tissue | The First Affiliated Hospital of Kunming Medical College | N/A |
| **Chemicals, peptides, and recombinant proteins** | | |
| BglII | NEB | R0144V |
| HindIII | NEB | R0104L |
| KpnI | NEB | R3142V |
| SacI | NEB | R3156V |
| Halt Protease Inhibitor Cocktail | Thermo Scientific | 78430 |
| T4 DNA ligase | NEB | M0202L |
| Proteinase K | Takara | 9034 |
| RNase A | Takara | 2158 |
| c-Jun siRNA (h) | Santa cruz | sc-29223 |
| CREB-1 siRNA (h) | Santa cruz | sc-29281 |
| Control siRNA (FITC Conjugate)-A | Santa cruz | sc-36869 |
| **Critical commercial assays** | | |
| RNeasy Mini Kit (50) | Qiagen | Cat#74104 |
| QIAGEN Plasmid Plus Midi Sample Kit (5) | Qiagen | Cat#12941 |
| FastStart Universal SYBR Green Master (Rox) | Sigma-Aldrich | ROCHE 4913850001 |
| EZ-ChIP Assay Kit | Sigma-Aldrich | Cat#17-371 |
| LightShift™ Chemiluminescent EMSA Kit | Thermo Scientific | Cat#20148 |
| **Experimental models: Cell lines** | | |
| Human lung fibroblasts | The typical culture preservation committee Kunming Cell Bank of the Chinese Academy of Sciences | KCB 200695 |
| Human embryonic lung fibroblasts HEL1 | The typical culture preservation committee Kunming Cell Bank of the Chinese Academy of Sciences | KCB 86024 |
| **Oligonucleotides** | | |
| Primers for RT-PCR, see Table S2 | This paper | N/A |
| Primers for 3C experiments, see Table S3 | This paper | N/A |

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Primers for luciferase assay, see Table S4 | This paper | N/A |
| CRISPR Cas9 sgRNA and primer pairs for PCR amplifying, see Table S5 | | |
| Primers for ChIP experiments, see Table S6 | This paper | N/A |
| Oligonucleotide probe for EMSA, see Table S7 | This paper | N/A |
| **Recombinant DNA** | | |
| pGL3-GSTCD promoter | This paper | N/A |
| Forward AT | This paper | N/A |
| Forward GT | This paper | N/A |
| Forward AC | This paper | N/A |
| Forward GC | This paper | N/A |
| Reverse AT | This paper | N/A |
| Reverse GT | This paper | N/A |
| Reverse AC | This paper | N/A |
| Reverse GC | This paper | N/A |
| **Software and algorithms** | | |
| SPSS 18 | IBM | https://www.ibm.com |
| Lasergene v7.1 | DNASTAR | https://www.dnastar.com/software/lasergene |
| Prism 6 | GraphPad | https://www.graphpad.com |
| **Other** | | |
| LD SNPs with rs10516526 from the CEU population of 1000 genomes | This paper | http://www.1000genomes.org |
| Putative regulatory element (PRE) regions from Encyclopedia of DNA Elements (ENCODE) | This paper | https://www.encodeproject.org |
| Resource website for rs72673891 and rs80245547 transcription factor prediction | This paper | http://jaspar.binf.ku.dk |
| Resource website for rs72673891 and rs80245547 eQTL analysis | This paper | https://www.gtexportal.org/home |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ya-ping Zhang (zhangyp@mail.kiz.ac.cn).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- Data reported in this paper will be shared on scienceDB https://doi.org/10.57760/sciencedb.08839.

- This paper does not report original code.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Human tissues and cell culture

Forty-six human lung tissue samples from Asian population were obtained from the First Affiliated Hospital of Kunming Medical College (Kunming, China), 20 COPD cases ($FEV_1 < 80\%$) and 26 controls with normal lung function, all participants were between 42 and 70 years old, and informed consent was obtained from all participants. Detailed information on the cases and controls is listed in Table S1, age, smoking index, and pulmonary function data are displayed as the mean $\pm$ SD, and this study was approved by the institutional ethics committee of First Affiliated Hospital of Kunming Medical University.

Briefly, human lung fibroblasts (HLF, KCB 200695) and human embryonic lung fibroblasts (HEL1, KCB 86024) were purchased from the typical culture preservation committee Kunming Cell Bank of the Chinese Academy of Sciences, and both cell lines were isolated from the lung tissue of male donors. Then cells were grown in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% FBS, 100 U penicillin per mL, and 10 mg streptomycin per mL. The cells were incubated at 37°C in a 5% $CO_2$ condition.

## METHOD DETAILS

### RNA isolation and RT-RCR

Total RNA was extracted by the QIAGEN RNeasy Mini Kit (QIAGEN, Germantown, MD, USA). Then, cDNA was isolated by the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol.

The mRNA levels of *GSTCD* were determined by RT–PCR. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used as a reference gene as an internal control. The primers used for each gene and conditions for RT–PCR are detailed in Table S2. Relative expression levels in lung tissues were determined using the $2^{-\triangle \triangle Ct}$ method normalized to the endogenous control. Each experiment was conducted for at least 40 cycles, and the melting curve was obtained. Data represent at least three independent experiments. Statistical analysis was performed using Student's *t* test.

### Bioinformatic analysis

To discriminate the potential functional candidate SNPs with rs10516526, the SNP data in the ∼500 kb area around this SNP in the 4q24 genomic region were downloaded from the CEU population of 1000 genomes[31] (http://www.1000genomes.org/) for LD analysis with rs10516526 ($r^2 \geq 0.8$). Putative regulatory element (PRE) regions with strong regulatory signals were analyzed and distinguished by using publicly available datasets from Encyclopedia of DNA Elements (ENCODE)[32,33] (https://www.encodeproject.org/), which contains a comprehensive list of functional elements in the human genome, including elements that act at histone modification (H3K4me1, H3K4me3, H3K9ac), DNase I hypersensitivity regions.

### Chromatin conformation capture assay

We used the 3C protocol adopted from Guo et al.[34] with some modifications. In this study, *HindIII* (New England Biolabs, Ipswich, MA, USA) was used as restriction enzyme. Briefly, cells were cross-linked with 1% formaldehyde at room temperature for 10 min and stopped by adding glycine to 0.125 M. The cells were harvested by centrifugation for 10 min at 2,500 rpm and 4°C, and washed twice with 10 mL of cold PBS, repeatedly. Then, cells were lysed in a 5 mL of cold RIPA lysis buffer (Thermo Scientific, USA) with freshly added protease inhibitors (Thermo Scientific, USA) and rotated for 90 min at 4°C. The nuclei were suspended in 1.2 × restriction buffer with 0.3% SDS and incubated at 65°C for 15 min with shaking. The 10% SDS was then sequestered by adding Triton X-100 to 1.8%. The cells were collected by centrifugation for 10 min at 2,500 rpm and 4°C. The nuclei were resuspended in 0.5 mL of 1.2 × restriction buffer, and then hundred units of the restriction enzyme *HindIII* (NEB, USA) were added for digestion at 37°C. The reaction was stopped by incubation at 65°C for 25 min. The digested chromatin was diluted in a 7 mL 1 × ligation buffer (NEB, USA). The reaction was then cooled to 16°C, and 2000 U of T4 DNA ligase (NEB, USA) was added. The sample was incubated at 16°C overnight and then at room temperature for 30 min. After ligation, the chromatin mixture was incubated with 300 μg proteinase K (Takara Bio Inc., Kusatsu, Shiga, Japan) at 65°C overnight to reverse crosslinks. RNA was removed by adding 300 μg RNase A (Takara, Japan) treatment for 30 min at 37°C. The 3C sample was purified by phenol–chloroform extraction and precipitated DNA in isopropanol.

Control PCR templates were generated by digestion and random ligations of bacterial artificial chromosomes that together encompass rs10516526 within the 500 kb flanking regions (BAC clones 525F1, 1041L12, and 277H20, Thermo Scientific, USA). BAC DNA was isolated by using the QIAGEN Plasmid Midi Kit (QIAGEN, USA). Equimolar amounts of the three BAC clones were digested with *HindIII* and ligated.

Quantitative real-time PCR was performed using FastStart Universal SYBR Green Master Mix (ROX) (Roche, Basel, Switzerland) with appropriate primers from purified DNA as well as the control template (Primer sequences are listed in Table S3). The PCR cycling parameters were as follows: 2 min at 50°C, 10 min at 95°C, 40 cycles of 15 s at 95°C and 30 s at 60°C. Three independent repeated experiments were carried out. All PCR products were sequenced and verified for their quality.

## Molecular cloning and dual-luciferase reporter assays

According to Obeidat M et al.'s study on the structure of *GSTCD* gene,[11] the Transcription start site (TSS) region of the *GSTCD* gene was identified, and then selected about 1.7 kb (Chr4: 106,628,257–106,629,940, hg19) upstream of the TSS was selected as the promoter region and cloned into the pGL3 basic vector (Promega, Madison, WI, USA) at the *BglII* and *HindIII* (NEB, USA) digestion sites. Constructs including the candidate SNPs (~1.1 kb, containing rs72673891 and rs59462153) were assembled to the *KpnI* and *SacI* sites at the forward orientation (Forward AT/GT/AC/GC) and reverse orientation (Reverse AT/GT/AC/GC) of the pGL3-*GSTCD* promoter. Mutations in the *GSTCD* promoter (Rs80245547) and candidate region (Rs72673891 and rs59462153) were made using the Q5 Site-Directed Mutagenesis Kit (NEB, USA). All plasmids used for reporter assays were confirmed by sequencing. Primer sequences used in this study can be accessed in the Table S4.

For reporter assays, constructs were transfected into HEL1 or HLF cells using ViaFect Transfection Reagent (Promega, USA), and luciferase expression was normalized to Renilla luciferase expression. Twenty-four hours posttransfection, cells were collected to measure promoter activity using the Dual-Luciferase Reporter Assay System (Promega, USA) according to the manufacturer's protocol. Luminescence signals were captured in a GloMax 96 Microplate Luminometer (Promega, USA). Data represent at least six independent experiments. Student's two-tailed *t* test was used to analyze the statistical significance of the data.

## CRISPR Cas9 - Mediated deletion of SNP region in HEL1 cells

To further verify the effect of the region near SNP on the expression of *GSTCD*, A group of sgRNAs with moderately efficient (Table S5) were designed in the upstream and downstream of rs80245547 and rs7267391 to knock out 249 bp and 121 bp, respectively. Subsequently, sgRNA was inserted into pSpCas9(BB)-2A-GFP (PX458) (48138, Addgene), and 14ug plasmid was transfected into HEL1 cells with Lipofectamine 3000 in 100 mm Cell Culture Dishes (Thermo Scientific, USA). After 48 h of editing, the enriched cells were collected by flow cytometry (Becton Dickinson, USA) for SNP region knockout detection. The knockout efficiency was determined by TA cloning assay, and then the expression level of *GSTCD* gene was detected by RT-PCR. Data represent at least three independent experiments. Statistical analysis was performed using Student's *t* test.

## Transcription factor binding site prediction

To evaluate whether rs80245547 and rs72673891 would alter the binding affinity of the transcription factor, the transcription factor-binding site was predicted with 101 base pair DNA sequences centered on rs72673891 and rs80245547 by using ENCODE ChIP-seq data and the web-based JASPAR database (http://jaspar.binf.ku.dk/).

## Chromatin immunoprecipitation assays

ChIP was performed using the EZ-ChIP Assay Kit (Millipore, Billerica, Massachusetts, USA) essentially as previously described.[35] Briefly, approximately $1 \times 10^7$ HEL1 cells were cross-linked with 1% formaldehyde at 37°C for 10 min. The lysate was immunoprecipitated with antibodies against C-Jun or CREB1 (Santa Cruz Biotechnology, Dallas, TX, USA). The input genomic DNA and the immunoprecipitated DNA were then amplified by qPCR using the specific primers listed in Table S6. Data represent at least three independent experiments. Statistical analysis was performed using Student's *t* test.

### RNAi for C-Jun and CREB1

Two silencer siRNAs against C-Jun and CREB1, and a silencer select for control were purchased from Santa cruz (Santa Cruz Biotechnology, Dallas, TX, USA). For silencing, HEL1 cells were co-transfected with the relevant luciferase reporter plasmids and siRNAs using Lipofectamine 3000 (Thermo Scientific, USA) according to the manufacturer's protocol. Luciferase reporter assays were performed as described above after 48 h, and the efficiency of siRNA-based interference and *GSTCD* gene expression in HEL1 cells were detected by RT-PCR, the RT-PCR primers listed in Table S2. Data represent at least three independent experiments. Student's two-tailed *t* test was used to analyze the statistical significance of the data.

### Electrophoretic mobility shift assays

Nuclear extracts were prepared using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Scientific, USA) following the manufacturer's instructions. In brief, HEL1 cells were harvested and washed with cold PBS. The cell pellet was suspended in CRE I buffer and incubated on ice for 10 min. Following incubation with CRE II solution, the preparation was centrifuged at 12,000 rpm for 5 min. The pellet was treated with nuclear extraction reagent (NER) with vortexing for 15 s every 10 min for a total of 40 min. After centrifugation at 12,000 rpm for 10 min, the remaining supernatant was the nuclear extract. The protein concentrations were measured using a Pierce BCA Protein Assay Kit (Thermo Scientific, USA).

EMSA was performed using a LightShift Chemiluminescent EMSA Kit (Thermo Scientific, USA) following the manufacturer's instructions. Oligonucleotide probes were labeled with biotin (Thermo Scientific, USA) and annealed into a double strand for detection. A total of 4 μg of HEL1 nuclear extracts was incubated with 0.05 pmol of labeled probes in gel shift binding buffer for 20 min at room temperature. The samples were separated on a 6% polyacrylamide gel. Signals were recorded on an Amersham imager 600 (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA). The probe sequences are listed in Table S7. For competition experiments, either specific or nonspecific oligonucleotide competitors were added to the binding mixture 10 min before the addition of the labeled probe. For supershift experiments, anti-CREB1 and anti-C-Jun (Santa Cruz, USA) were preincubated with the nuclear extracts in binding reaction buffer for 40 min at room temperature.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Statistical analysis

All statistical tests were performed in SPSS 18.0 (SPSS Inc., Chicago, IL, USA) or GraphPad Prism software (GraphPad Software, San Diego, CA, USA). In this study, the significance level was accepted when the *P* (probability) value was <0.05.