



# Synergistic alterations in the multilevel chromatin structure anchor dysregulated genes in small cell lung cancer



Dan Guo<sup>a,1</sup>, Qiu Xie<sup>a,1</sup>, Shuai Jiang<sup>c,1</sup>, Ting Xie<sup>a</sup>, Yaru Li<sup>b</sup>, Xin Huang<sup>b</sup>, Fangyuan Li<sup>a</sup>, Tingting Wang<sup>a</sup>, Jian Sun<sup>d</sup>, Anqi Wang<sup>a</sup>, Zixin Zhang<sup>a</sup>, Hao Li<sup>b</sup>, Xiaochen Bo<sup>b</sup>, Hebing Chen<sup>b,\*</sup>, Zhiyong Liang<sup>d,\*</sup>

<sup>a</sup> Medical Science Research Centre, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

<sup>b</sup> Beijing Institute of Radiation Medicine, Beijing 100850, China

<sup>c</sup> State Key Laboratory of Proteomics, Institute of Basic Medical Sciences, National Center of Biomedical Analysis, Beijing 100850, China

<sup>d</sup> Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

## ARTICLE INFO

### Article history:

Received 12 August 2021

Received in revised form 2 November 2021

Accepted 2 November 2021

Available online 01 November 2021

November 2021

### Keywords:

SCLC  
TARGET  
Hi-C  
TAD  
NanoString  
Tumorigenesis

## ABSTRACT

Small cell lung cancer (SCLC) is an aggressive form of lung cancer that uniquely changes the chromosomal structure, although the basis of aberrant gene expression in SCLC remains largely unclear. Topologically associated domains (TADs) are structural and functional units of the human genome. Genetic and epigenetic alterations in the cancer genome can lead to the disruption of TAD boundaries and may cause gene dysregulation. To understand the potential regulatory role of this process in SCLC, we developed the TAD boundary alteration-related gene identification in tumors (TARGET) computational framework, which enables the systematic identification of candidate dysregulated genes associated with altered TAD boundaries. Using TARGET to compare gene expression profiles between SCLC and normal human lung fibroblast cell lines, we identified >100 genes in this category, of which 24 were further verified in samples from patients with SCLC using NanoString. The analysis revealed synergistic chromatin structure alteration at the A/B compartment and TAD boundary levels that underlies aberrant gene expression in SCLC. TARGET is a novel and powerful tool that can be used to explore the relationship of chromatin structure alteration to gene dysregulation related to SCLC tumorigenesis, progression, and prognosis.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Lung cancer is the leading cause of cancer incidence and mortality worldwide [1]. Small cell lung cancer (SCLC), an aggressive and chemosensitive malignancy with a high relapse rate and poor prognosis, accounts for approximately 14% of all lung cancers [2–4]. It causes unique changes in biological function and chromosomal structure. The dysregulation of tumor suppressor genes, oncogenes, and signaling pathways; the upregulation of receptor tyrosine kinases, growth factors, and cellular markers; and the activation of early development pathways have been reported [5]. Chromosomal rearrangements, such as TP53 and RB1, occur constitutionally in the general population and somatically in patients with SCLC, as for the majority of cancers [6]. In addition, four reciprocal translocations [t(1;17)(p10;p10), t(3;6)(q24;q21), t(12;17)

(p10;p10), and complex t(2;6)] have been identified in the NCI-H82, NCI-H2009, and NCI-H1437 SCLC cell lines [7]. High-throughput chromosome conformation capture (Hi-C) was used for the precise detection and characterization of chromosomal rearrangements and copy number variations in human tumors [8]. The molecular mechanisms responsible for tumor development and clinical behavior and the tumor landscape have been explored, but results have been generally inconclusive [9]. The detailed biology of SCLC remains poorly understood, and adequate study of this disease continues to be challenging.

Investigation of the regulatory basis of key SCLC-driving genes will shed light on the mechanism underlying SCLC tumorigenesis, which requires decoding of the function of the non-coding genome, the “dark matter” [10] of DNA (accounting for > 97% of the entire human genome). Although the regulatory function of the majority of this genome remains elusive, the recent development of genome-wide chromatin conformation capture technology provides an opportunity to link alterations in non-coding DNA to functional impacts on gene transcriptional regulation through the higher-order chromatin architecture.

\* Corresponding authors.

E-mail addresses: [chb-1012@163.com](mailto:chb-1012@163.com) (H. Chen), [liangzhiyong1220@yahoo.com](mailto:liangzhiyong1220@yahoo.com) (Z. Liang).

<sup>1</sup> These authors contributed equally to the paper as first authors.

The three-dimensional chromatin state underpins the structural and functional basis of the genome by bringing regulatory elements and genes into close spatial proximity to ensure proper cell type-specific gene expression profiles. Changes in domain structure are accompanied by novel cancer-specific chromatin interactions in topologically associated domains (TADs) that are enriched in regulatory elements, such as enhancers, promoters, and insulators, and associated with alterations in gene expression [11]. However, a computational tool for the direct identification of aberrantly expressed genes associated with the disruption of the higher-order chromatin architecture in cancer is lacking. Here, we report on our development of the TAD boundary alteration-related gene identification in tumors (TARGET) computational framework for the systematic identification of candidate dysregulated genes associated with altered TAD boundary landscapes based on Hi-C and RNA sequencing (RNA-seq) data from the NCI-H209 and DMS153 SCLC cell lines and the MRC-5 normal cell line. We also aim to characterize the interplay between alterations in chromatin architecture at different scales and unravel their functional role in the dysregulation of gene expression in SCLC.

## 2. Results

### 2.1. Identification of candidate genes affected by altered TAD boundaries in SCLC cell lines

Hi-C and RNA-seq analysis of the MRC-5 human embryonic lung fibroblast cell line and the NCI-H209 and DMS153 SCLC cell lines was performed, and the TARGET pipeline was applied to the Hi-C and RNA-seq datasets to facilitate the identification of candidate genes that were dysregulated near altered TADs in SCLC (Fig. 1). We identified 201 and 152 candidate boundary alteration-associated genes in NCI-H209 and DMS153, respectively. Briefly, Hi-C data were binned to generate Hi-C contact maps at 20-kb resolution, with each bin representing a 20-kb locus, and biases in Hi-C contact maps were removed. TAD boundaries that showed considerable alteration of their ability to insulate chromatin contacts in tumor cells compared with normal cells were identified. Subsequently, genes with significant differential expression near these altered boundaries were identified as candidate boundary alteration-associated genes, using pre-set cutoffs for expression fold changes and effective boundary alteration ranges. Based on the RNA-seq data, we identified 2,988 and 3,329 differentially expressed genes in NCI-H209 and DMS153, respectively, compared with expression in MRC5 (Supplemental Fig. S1, Supplemental Table S1). Gene ontology analysis revealed that genes transcribed more actively in SCLC tissues corresponded to functional enrichment in the cell cycle and cell division, whereas those with significantly lesser expression in SCLC tissues corresponded to functional enrichment in the regulation of apoptotic processes (Supplemental Fig. S1). To assess the similarity of the NCI-H209 and DMS153 transcriptomes, we calculated Pearson's correlation coefficient for normalized expression levels of all genes in these cell lines. The coefficient was 0.91 ( $p < 2.2 \times 10^{-16}$ ), indicating a high degree of transcriptome similarity (Supplemental Fig. S2A). In addition, most differentially expressed genes were identified in both cell lines (Supplemental Fig. S2B). Thus, the two SCLC cell lines were highly similar.

To assess the potential influence of the alignment software used, we performed RNA-seq alignment using Tophat2 and HISAT2 respectively. We first assessed transcriptome similarity between gene expression levels that were calculated based on these two different sequence alignment tools. The Pearson's coefficients of correlation between log-normalized gene expression levels  $\{\log_2[\text{fragments per kilobase of transcript per million mapped reads}$

(FPKM) + 1 $\}$  were  $> 0.96$  for all RNA-seq datasets generated in this study, indicating that highly similar gene expression profiles were generated with Tophat2 and HISAT2 (Supplemental Fig. S3). To assess the impact of the application of different sequence alignment tools, we investigated the similarity between differentially expressed genes that were identified based on either Tophat2 and HISAT2, and found that the differentially expressed genes identified using either Tophat2 or HISAT2 overlapped strongly (Supplemental Fig. S4). Thus, the use of HISAT2 only had a weak influence on transcriptome quantification and differentially expressed gene identification.

Based on Hi-C data at 20-kb resolution, we identified 3,701, 3,987, and 5,023 TADs in the MRC5, NCI-H209, and DMS153 cell lines, respectively. The identification of increased numbers of TADs in the two SCLC cell lines is consistent with previous findings for cancer cells [11,12]. We identified 15,916 and 16,510 loci with altered insulation scores (ISs) in NCI-H209 and DMS153, respectively (Supplemental Fig. S5, Supplemental Table S2), including 1,297 and 1,628 respective loci corresponding to altered TAD boundaries. Two hundred one and 152 candidate boundary alteration-associated genes were identified in NCI-H209 and DMS153, respectively.

### 2.2. Characterization of altered TAD boundaries in SCLC cell lines

To confirm the identification of altered TAD boundaries, we first determined whether the loci identified as having altered ISs were more likely to be boundary loci (those corresponding to identified TAD boundaries in normal or tumor cells) than non-boundary loci (those not corresponding to TAD boundaries in either cell type). As expected, boundary loci in both SCLC cell lines were prone to have altered ISs. In the NCI-H209 cell line, 12.80% of boundary loci and 10.07% of non-boundary loci had altered ISs ( $P = 6.8 \times 10^{-19}$ , Fisher's exact test) (Fig. 2A). In DMS153, these percentages were 13.92% and 10.37%, respectively ( $P = 1.6 \times 10^{-33}$ , Fisher's exact test). The enrichment of loci with altered ISs also peaked near boundary loci in both SCLC cell lines (Fig. 2B). TAD boundaries with increased ISs comprised 7.75% (609/7863) and 12.07% (1195/9899) of all identified TAD boundaries in NCI-H209 and DMS153, respectively; those with decreased ISs comprised 8.75% (688/7863) and 4.37% (433/9899) of all identified TAD boundaries in these respective cell lines. In total, 295 altered TAD boundaries had increased ISs and 242 had decreased ISs in both SCLC cell lines (Fig. 2C). Most TAD boundaries with altered ISs were not conserved between normal and tumor cell lines, but corresponded to "gains" or "losses" in the SCLC cell lines compared with the normal cell line (Fig. 2D).

We next generated a differential Hi-C contact matrix to determine whether the identified TAD boundaries with altered ISs were indeed boundaries with differential insulation power between the normal and tumor cell lines. We analyzed conserved and unconserved boundaries separately. We examined the TAD boundaries with decreased ISs in NCI-H209. The averaged Hi-C contact map centered on unconserved TAD boundaries showed a TAD boundary-like pattern, indicating strong insulation of these loci, in MRC5; this pattern was barely discernible, supporting the loss of these boundaries, in NCI-H209 (Fig. 2E). The averaged Hi-C contact map centered on conserved TAD boundaries with decreased ISs also showed the TAD boundary-like pattern in MRC5. This pattern was weak, but clear, in NCI-H209, supporting the presence of these TAD boundaries in both cell lines (Fig. 2E). The differential Hi-C contact map showed that the insulation power at these conserved boundaries was weakened, with increased Hi-C contact frequency between downstream and upstream loci, in NCI-H209 relative to MRC5 (Fig. 2E). The differential contact map for MRC5-specific TAD boundaries was similar to that of conserved TAD boundaries with decreased ISs, reflecting similar degrees of

## TARGET computational framework

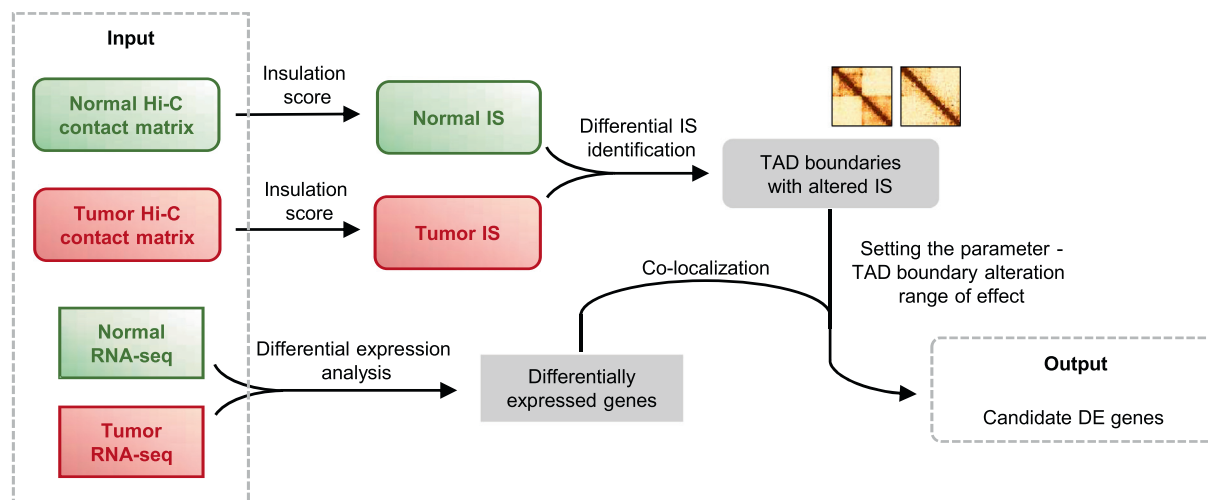


Fig. 1. Work flow for identification of candidate boundary alteration affected genes.

IS alteration (Fig. 2E). However, as the insulation of conserved TAD boundaries in MRC5 was stronger than that of MRC5-specific TAD boundaries, the IS decrease reflected the loss of the weaker MRC5-specific TAD boundaries in the NCI-H209 cell line, while the similar degree of IS decrease reflected only weakened TAD boundaries for the stronger conserved TAD boundaries. Subsequently, we performed these analyses using TAD boundaries with increased ISs. As expected, TAD boundaries that were gained in NCI-H209 corresponded to loci with no discernible boundary-like pattern in MRC5, and conserved boundaries with increased ISs were weaker in MRC5 than in NCI-H209 (Supplemental Fig. S6A). These analyses were also performed for DMS153 (Supplemental Fig. S6B, C); the patterns observed were similar to those observed for NCI-H209. Thus, the altered TAD boundaries identified with TARGET indeed represent TAD boundaries with altered insulation power in SCLC cell lines.

### 2.3. Enrichment of differentially expressed genes in regions near altered TAD boundaries in SCLC cell lines

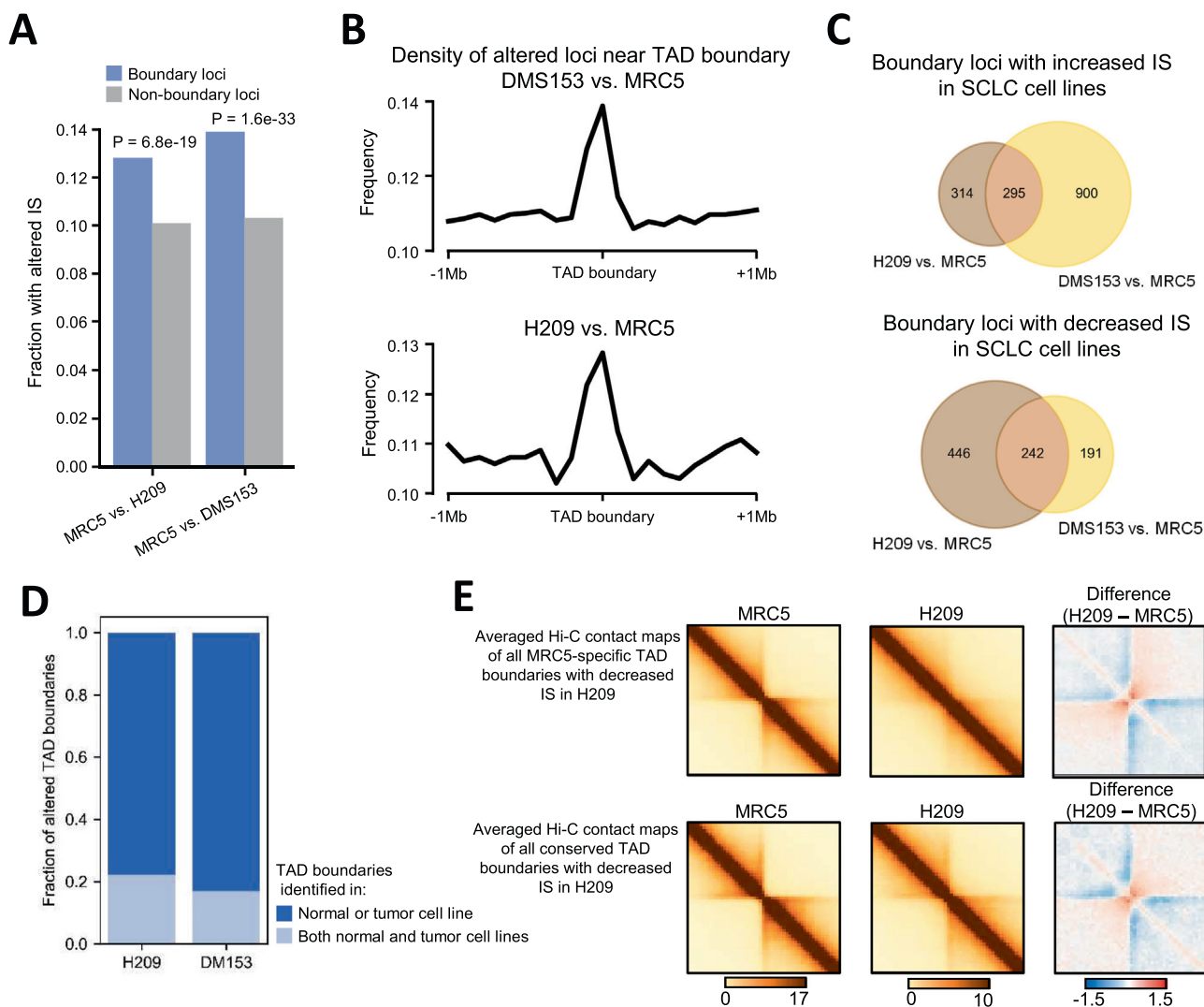
To understand the potential functional impact of the altered TAD boundaries in the SCLC cell lines, we investigated gene expression alteration in these cell lines based on RNA-seq data. Enrichment in differentially expressed genes was observed at altered TAD boundaries relative to other TAD boundaries in the NCI-H209 and DMS153 cell lines (Fig. 3A). Thus, genes at altered TAD boundaries are prone to transcriptional activity changes, suggesting an impact of gene transcriptional dysregulation at altered TAD boundaries in SCLC. Our findings are consistent with previous reports that genes located near altered TAD boundaries are affected by altered transcriptional regulatory circuits that are mediated by the three-dimensional (3D) chromatin structure [13]. As genes located near altered TAD boundaries also may be affected by disruption of the chromatin structure, we next investigated the potential impact of TAD boundary alteration on the expression levels of nearby genes to estimate the range of distance of the observed effects. We found significant enrichment of differentially expressed genes in genomic regions up to 400 kb upstream and downstream of altered boundaries (Fig. 3B). The enrichment fold change increased with decreasing distance, suggesting that genes in closer proximity to altered boundaries were more likely to be differentially expressed (Fig. 3B). We also used a series of increasing expression-fold-change cutoffs to identify six groups of genes

with increasing degrees of expression alteration, enabling investigation of the enrichment of these genes near altered TAD boundaries. The observed enrichment of genes near altered TAD boundaries with high degrees of expression alteration in SCLC cell lines was greater when we used a large cutoff (expression log<sub>2</sub> fold change  $\geq 6$ ; Fig. 3B) than when we used a lower cutoff. We also observed a general trend for each distance range of increasing enrichment of selected genes with the use of higher cutoff values for gene selection (Fig. 3B). These results indicate that TAD boundary alteration in SCLC may be more relevant to prominent expression alteration than to moderate expression changes, and that the use of an effective range  $< 400$  kb for the selection of candidate genes affected by altered TAD boundaries is plausible.

We demonstrated the utility of TARGET framework application to other datasets using publicly available Hi-C and RNA-seq datasets for the 22Rv1 and C4-2B prostate cancer cell lines and the corresponding RWPE1 normal cell line [14]. The analysis showed that differentially expressed genes were enriched near altered TAD boundaries identified in both prostate cancer cell lines. Using TARGET, we identified 96 and 90 candidate genes in 22Rv1 and C4-2B, respectively (Supplemental Table S4).

### 2.4. Coordinated TAD boundary alteration and A/B compartment flipping in SCLC cell lines

The human genome is segregated spatially into A/B compartments that participate in TAD formation [15], and the establishment and flipping of A/B compartments have been shown to play important roles in early embryonic development, cell differentiation, and cell reprogramming [16–18]. However, compartment alteration in the cancer genome and its relationship to TAD alteration have not been studied extensively. Based on the altered TAD boundaries and candidate genes identified in both SCLC cell lines examined in this study, we further explored the relationship between TAD boundary alteration and A/B compartment flipping, and the potential regulatory impacts of these processes, in SCLC. First, we attempted to identify the A/B compartment in all three cell lines at 200-kb resolution; we found that most genomic loci corresponded to unchanged compartment composition in all cell lines (Fig. 4A). We also identified loci with A-to-B and B-to-A compartment flipping in the SCLC cell lines relative to the MRC5 cell line (Supplemental Table S3). In NCI-H209 cells, 12.36% (1,509/12,206) of loci showed A-to-B compartment flips and 14.60% (1,782/12,206) showed B-to-A compartment flips. These



**Fig. 2.** Properties of altered TAD boundaries in SCLC. (A) Barplot showing the enrichment of altered IS in TAD boundaries than in other genomic regions. (B) Density of altered IS from upstream 1 Mb to downstream 1 Mb of TAD boundaries. (C) Venn diagram showing the overlap between altered TAD boundaries in H209 and DM153. The “cell type A vs. cell type B” is defined as “the value in cell type A minus the value in cell type B” when quantitative comparison is involved. For example, in Fig. 2C, boundary loci with increased IS for “H209 vs. MRC5” indicate higher IS in H209 than MRC5. (D) Stacked barplot showing the fraction of altered TAD boundaries that are identified in both normal and tumor cell lines or in only one cell line. Most altered TAD boundaries are gained or lost in tumor. (E) Averaged Hi-C contact map showing boundaries with decreased IS that are lost (top panels) or weakened (bottom panels) in H209 SCLC cell line. Differential contact maps are calculated based on observed/expected contact maps of MRC5 and H209 cell lines.

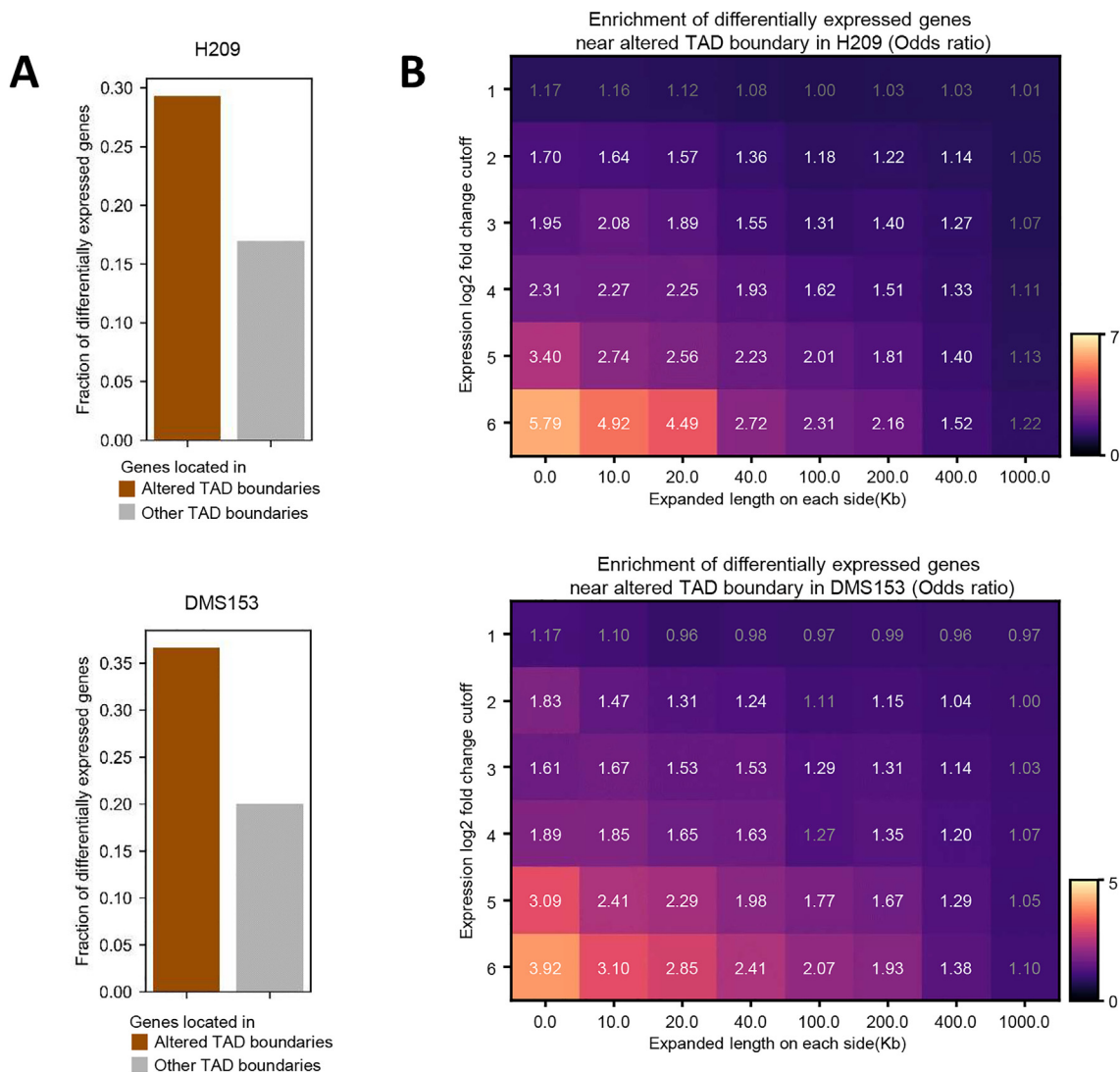
percentages in DMS153 cells were 14.56% (1,777/12,206) and 14.62% (1,785/12,206), respectively (Fig. 4A). These results showed that a subset of genomic regions may undergo compartment flipping during SCLC tumorigenesis.

The observed overall impacts of compartment flipping on gene expression in both SCLC cell lines are in agreement with previous findings [12]. Genes at A-to-B compartment-flipped loci were more likely to be down-regulated than were those at B-to-A compartment-flipped loci, and the latter were more likely to be up-regulated than the former (Fig. 4B). Small fractions of up-regulated genes in A-to-B compartment-flipped regions and down-regulated genes in B-to-A compartment-flipped regions were also present, indicating the complexity of transcriptional regulation and the potential contribution of other factors to the altered expression of these genes. We also found that altered TAD boundaries were enriched in compartment-flipped regions (Fig. 4C) in both SCLC cell lines, suggesting that local TAD boundary alteration is linked to the global reorganization of chromosome structure represented by the A/B compartment. To further explore

this connection, we tested the hypothesis that the direction of A/B compartment flipping depends to some extent on the direction of IS alteration at TAD boundaries. In both SCLC cell lines, TAD boundaries with increased ISs were more enriched in B-to-A compartment-flipped loci and boundaries with decreased ISs were more enriched in A-to-B compartment-flipped loci. These results indicate synergistic alteration of the 3D chromatin structure at the TAD and compartment levels, further supporting the connection between local TAD boundary disruption and global chromatin rearrangement (Fig. 4D).

### 2.5. Effects of the synergistic alteration of TAD boundary disruption and A/B compartment flipping on gene expression

We hypothesized that the similarity in the directions of A/B compartment flipping and TAD boundary strength alteration would correspond to a similar regulatory impact on gene expression. To test this hypothesis, we further explored the relationship between the directions of IS and gene expression alteration. In both

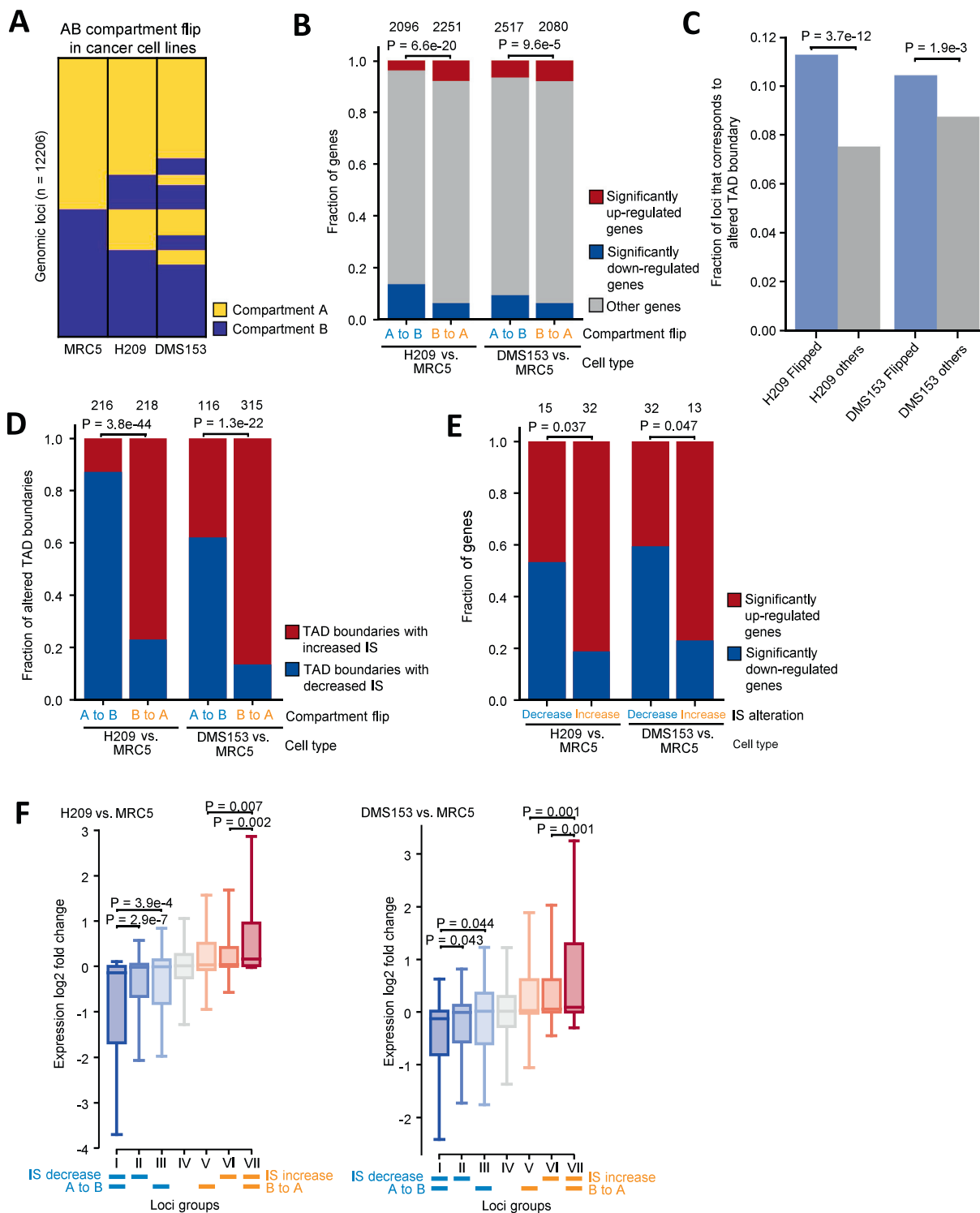


**Fig. 3.** Differentially expressed genes are enriched in regions near altered TAD boundaries in SCLC. (A) Barplot showing enrichment of differentially expressed genes (expression log<sub>2</sub> fold change > 2, q value < 0.05) in altered TAD boundaries compared to other genomic regions in SCLC cell lines. (B) Heatmap showing odds ratio of differentially expressed genes in regions near altered TAD boundaries than in other region. Differentially expressed genes are selected based on different expression log<sub>2</sub> fold change cutoffs (y-axis), and regions near altered TAD boundaries are selected by including regions up- and down- stream of altered TAD boundaries by a specific length (x-axis). Odds ratio in white text represent significant enrichment (p-value < 0.05), gray text represent no statistical significance.

SCLC cell lines, up-regulated and down-regulated gene enrichment was observed at altered TAD boundaries with increased and decreased insulation, respectively, similar to the A/B compartment flipping pattern (Fig. 4E). These results confirmed our hypothesis. We further hypothesized that this synergistic alteration of TAD boundaries and A/B compartments collaboratively strengthened the dysregulation of abnormally expressed genes in SCLC. We categorized all genomic loci, assigning them to groups I–VII, corresponding to combinations of presence/absence and direction of TAD boundary alteration and A/B compartment flipping [group I, transcription start sites (TSSs) at altered TAD boundaries with decreased ISs in A-to-B compartment-flipped regions; group II, TSSs at altered TAD boundaries with decreased ISs in unchanged compartment regions; group III, TSSs in compartment A-to-B-flipped regions with no TAD boundary alteration; group IV TSS with no TAD boundary alteration and in unchanged compartment regions; group V, TSSs in B-to-A compartment-flipped regions with no TAD boundary alteration; group VI, TSSs at altered TAD boundaries with increased ISs and in unchanged compartment regions; group VII, TSSs at altered TAD boundaries with increased

ISs in B-to-A compartment-flipped regions]. We investigated alteration in the expression of these gene categories. In both SCLC cell lines, the expression of group I genes was repressed more significantly than that of groups II and III genes (H2O9:  $P = 2.9 \times 10^{-7}$  for group I vs group II,  $P = 3.9 \times 10^{-4}$  for group I vs group III, two-tailed *t*-test; DMS153:  $P = 0.043$  for group I vs group II,  $P = 0.044$  for group I vs group III, two-tailed *t*-test), whereas expression fold changes of group VII genes were significantly greater than those of groups V and VI genes (H2O9:  $P = 0.002$  for group VI vs group VII,  $P = 0.007$  for group V vs group VII, two-tailed *t*-test; DMS153:  $P = 0.001$  for group VI vs group VII,  $P = 0.001$  for group V vs group VII, two-tailed *t*-test). Thus, genes with TSSs at loci with synergistic alteration of TAD boundaries and A/B compartments showed the greatest degree of transcriptional activity alteration (Fig. 4F). These data support our hypothesis that synergistic alteration plays a collaborative regulatory role in SCLC.

Chromatin loops have been shown to be enriched in regulatory elements, such as enhancers and promoters, which enables promotion of the cell type-specific transcriptional activity of genes and



**Fig. 4.** Relationship between TAD boundary alteration and AB compartment flip in SCLC cell lines. (A) Overview of AB compartment fraction in normal and SCLC cell lines. (B) Stacked barplot showing the fraction of differentially expressed genes in compartment A to B flip and in B to A flip in SCLC cell lines. The total number of genes are shown on the top of each stacked bar. (C) Barplot showing the enrichment of altered TAD boundaries in compartment flipped loci. Blue bars represent the fraction of altered TAD boundaries in all TAD boundaries that located in flipped compartment. Gray bars represent the fraction of altered TAD boundaries in all TAD boundaries that located in unchanged compartment. (D) Stacked barplot showing the enrichment of altered TAD boundaries with increased IS in compartment B to A flipped loci, and the enrichment of altered TAD boundaries with decreased IS in compartment A to B flipped loci. The total number of loci are shown on the top of each stacked bar. (E) Stacked barplot showing the enrichment of up-regulated genes in altered TAD boundaries with increased IS, and the enrichment of down-regulated genes in altered TAD boundaries with decreased IS. The total number of genes are shown on the top of each stacked bar. (F) Boxplot showing levels of expression alteration of genes located in different loci groups (group I-VII) of genomic regions categorized by the direction of alteration in both TAD boundary strength and A/B compartment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

underlies TAD formation [19,20]. Thus, gains or losses of chromatin loops in SCLC may serve as important structural bases for TAD boundary alteration and help to explain the correlation between such alteration and the dysregulation of candidate genes identified by TARGET in SCLC cell lines. To explore this hypothesis, we identified chromatin loops that were gained or lost in both SCLC cell lines relative to the MRC5 cell line. Anchors of altered chromatin loops showed more enrichment in candidate gene TSSs than in TSSs of other genes in both SCLC cell lines (Supplemental Fig. S7A). To further explore whether gained chromatin loops in SCLC promoted the transcription of candidate genes and *vice versa*, we examined candidate genes with TSSs 0–20 kb downstream or upstream of gained or lost chromatin loops in both SCLC cell lines (52/201 candidate genes in NCI-H209 cells, 53/152 candidate genes in DMS153 cells). We found that candidate genes with TSSs at or near gained chromatin loops were significantly more likely to be up-regulated than were those located at or near lost chromatin loops, and that candidate genes with TSSs located at or near lost chromatin loops were significantly more likely to be down-regulated than were those located at or near gained chromatin loops, in the SCLC cell lines (NCI-H209, odds ratio = 12.8; DMS153, odds ratio = 5.2; Supplemental Fig. S7B).

As an example, we identified *SOX2* as a candidate boundary alteration-associated gene. The Cancer Genome Atlas expression profile for *SOX2* [21,22] shows that it is activated in lung adenocarcinoma and lung squamous cell carcinoma (Supplemental Fig. S8). *SOX2* expression has been shown to be greater in SCLC than in non-SCLC tissues [23], and *SOX2* overexpression in SCLC is associated with poor prognosis [24]. The targeted suppression of *SOX2* expression blocked the proliferation of *SOX2*-amplified SCLC cell lines [25]. We found that *SOX2* was activated and was located near a TAD boundary with an increased IS in NCI-H209 and DMS153 cells (Fig. 5A, B). *SOX2* was also found at loci with negative compartment score (CS), corresponding to compartment B in MRC5 and positive CS, corresponding to compartment A, in both SCLC cell lines, indicating the occurrence of B-to-A compartment flipping in these genomic regions (Fig. 5A). In addition, the *SOX2* TSS was located in an anchor of a gained chromatin loop in NCI-H209 cells and in an anchor shared by two gained chromatin loops in DMS153 cells. These observations are in agreement with the reported roles of chromatin loops in TAD formation and the promotion of cell type-specific gene expression [26]. Thus, synergistic alteration of the 3D chromatin structure may underlie the activation of the *SOX2* oncogene in SCLC (Fig. 5C).

### 2.6. Verification of the identified candidate genes in patients with SCLC

For further verification, we selected 24 of the dysregulated candidate genes that we identified based on RNA-seq data from SCLC cell lines and that have been reported in most patients with SCLC (*CCND1*, *SFRP1*, *PSG5*, *GJA1*, *CHRM2*, *OAF*, *SLC38A5*, *WFS1*, *STC2*, *TNIP2*, *CACNA2D3*, *SOX2*, *PAX5*, *NELL1*, *CRTAC1*, *C5orf38*, *C14orf23*, *AUTS2*, *ANKRD6*, *CNTNAP2*, *BEX2*, *IRX2*, *TAGLN3*, and *EPCAM*). Selection was based mainly on marked differential expression, TAD boundary insulation strength, and proximity to altered TAD boundaries. Many of these genes showed synergistic alteration of the multilevel chromatin structures similar to that observed for *SOX2* (Supplemental Fig. S9). Of the 24 selected genes, 14 were up-regulated and 10 were down-regulated in both H209 and DMS153 tumor cell lines compared to MRC5 (Fig. 6). We performed NanoString analysis of these genes in clinical samples from 18 patients with histologically confirmed SCLC (Fig. 6, Supplemental Fig. S10). The patients' clinical characteristics are summarized in Table 1. Their median age at diagnosis was 63.4 (range, 41–81) years, and 72.2% ( $n = 15$ ) of the patients were male. Bronchial and pleural invasion were detected in 44.4% and 11.1% of cases,

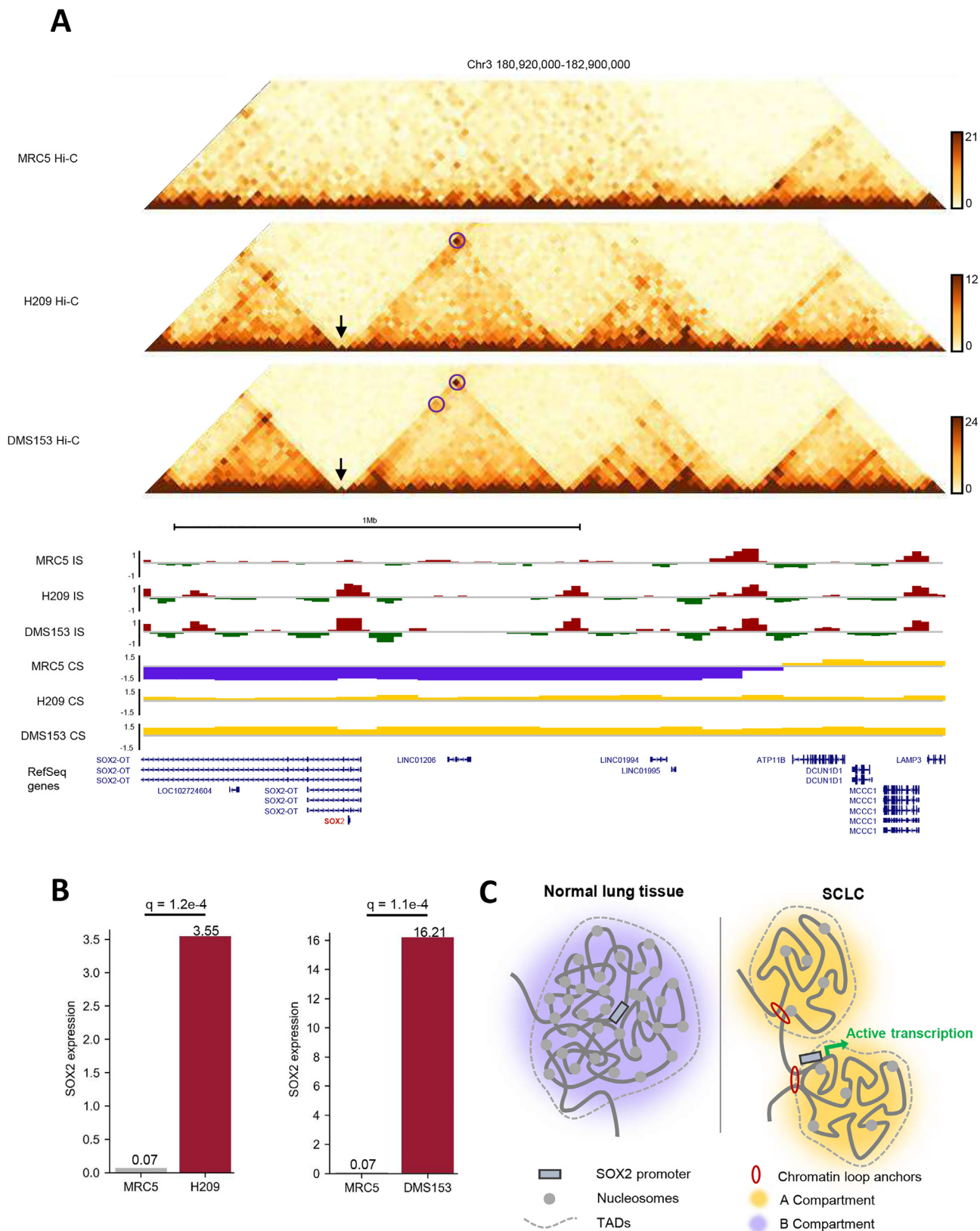
respectively, and >66.7% of the patients had regional lymph node metastasis. The Ki67 index (a cellular marker of proliferation) was extremely strong ( $\geq 50\%$ ) in 83.3% of cases [27,28], and necrosis was observed in three (16.7%) cases. In addition, more than three-quarters of the patients showed positivity for the biomarkers synaptophysin, chromogranin A, thyroid transcription factor-1 (TTF-1), CD56, and AE1/AE3, and only four (22.2%) cases were CK5/6 negative. The disease was limited (pTNM stages I–III) in 17 (94.4%) cases and extensive (stage IV) in 1 (5.6%) case [29].

Expression alteration trends similar to those observed in the SCLC cell lines were observed for most of the candidate genes in the patient samples. The trends of three genes (*CRTAC1*, *C5orf38*, and *IRX2*) were not consistent with those observed in SCLC cell lines in more than half of the patients; trends for 17 genes (*CCND1*, *PSG5*, *GJA1*, *CHRM2*, *OAF*, *SLC38A5*, *WFS1*, *TNIP2*, *SOX2*, *PAX5*, *NELL1*, *C14orf23*, *AUTS2*, *CNTNAP2*, *BEX2*, *TAGLN3*, and *EPCAM*) were consistent in  $\geq 16$  patient samples. Nineteen genes were differentially expressed between normal and tumor tissues (Supplemental Fig. S11); 10 (*EPCAM*, *TAGLN3*, *CNTNAP2*, *PAX5*, *BEX2*, *NELL1*, *SOX2*, *C14orf23*, *ANKRD6*, and *AUTS2*) were up-regulated and 9 (*OAF*, *SLC38A5*, *TNIP2*, *GJA1*, *CCND1*, *CHRM2*, *STC2*, *PSG5*, and *WFS1*) were down-regulated in tumor tissue and tumor cell lines. *CRTAC1* was up-regulated in the two tumor cell lines but down-regulated in SCLC tissue (Supplemental Fig. S11). Potential correlations of these expression alterations with the patients' clinical characteristics were examined; *EPCAM* was found to be up-regulated in SCLC tumor tissues without necrosis compared with those with necrosis (Supplemental Fig. S12). Thus, most of the candidate genes that we identified in the SCLC cell lines were also dysregulated in the tumor cells of patients with SCLC, and the candidate genes identified by TARGET based on RNA-seq and Hi-C data from SCLC cell lines may be representative of *in vivo* gene dysregulation associated with altered TAD boundaries in most patients with SCLC.

### 3. Discussion

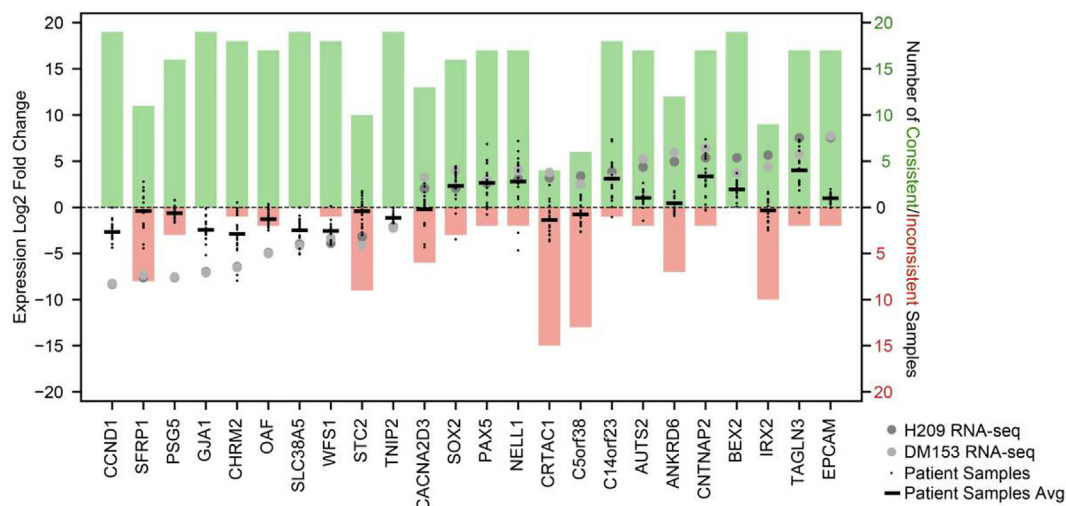
SCLC, a neuroendocrine tumor according to the 2015 World Health Organization revision, is associated with significantly higher rates of chromosomal aberrations, alterations, and remodeling [30,31]. The inactivation of chromatin remodeling genes may drive the transformation of neuroendocrine tumors, and variants in chromatin interaction regions may contribute to tumorigenesis via regulation of the expression of target genes [32–34].

In this study, we developed TARGET, a novel method for the systematic identification of candidate dysregulated genes associated with altered TAD boundary landscapes based on Hi-C and RNA-seq data, to facilitate the identification of genes with multilevel chromatin structure anchor dysregulation in SCLC, and to explore relationships between chromatin structure alteration and gene dysregulation related to SCLC tumorigenesis, progression, and prognosis. The TARGET framework involves the use of loess regression to estimate the expected variation in ISs between datasets at a certain IS level and to identify altered TAD boundaries based on the O/E IS variation threshold. Thus, altered TAD boundaries were identified not by using a fixed IS variation cutoff, but by using more accurate IS variations that were estimated by taking the IS variations of all loci at similar IS levels into account. This approach better reflects the characteristic of IS variation between datasets and enables the identification of altered TAD boundaries between Hi-C datasets with unreplicated Hi-C data. In addition, the TARGET method involves consideration not only of TAD boundaries that are gained and lost in tumor cells, but also those with altered boundary strength, which is implied by functional roles. In addition, TARGET outputs a list of candidate genes whose transcriptional activity alterations may be related to the alteration of



**Fig. 5.** Sox2 is significantly up-regulated in SCLC cell lines accompanied with coordinated increase in both IS and CS. (A) Genome browser-like plot showing Hi-C contact frequency, genes, IS and CS in the genomic region around SOX2. TAD boundaries with increased IS near SOX2 gene are marked by black arrows. Chromatin loops that anchored at SOX2 gene are marked by purple circles. (B) Barplot showing SOX2 expression in MRC5 and the two SCLC cell lines. (C) Cartoon showing the synergistic alteration in multi-level chromatin architectures underlying SOX2 activation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)





**Fig. 6.** Differential expression of candidate genes between normal and tumor tissue of 19 patients. Green bars showing the number of patients with consistent differential expression in cell lines for each candidate genes. Red bars showing the number of patients with inconsistent differential expression. Expression log<sub>2</sub> fold change of each patient are marked with black dots, and mean expression log<sub>2</sub> fold change are marked with black lines. Expression log<sub>2</sub> fold change in cell lines are marked with gray dots. Most candidate genes shows consistent differential expression in most patients and in cell lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nearby TAD boundaries. For downstream analysis of TARGET-identified candidate genes, we recommend that users use the GEPIA2 web-based platform [35] (<http://gepia2.cancer-pku.cn/>) as it provides comprehensive and user-friendly tools for tumor-associated analysis, such as differential expression analysis, survival analysis, and similar gene identification for user interested genes.

The application of the TARGET framework to datasets for other cancer types will expand our understanding of the extent to which aberrant chromatin architecture affects gene expression in cancer. Our study revealed the manner in which different types of TAD boundary alteration may preferentially activate and repress gene expression in tumors, and established the strengthened functional impact of synergistic alterations in the multilevel chromatin structure.

### 3.1. Synergistic alterations in the multilevel chromatin structure led to gene transcriptional dysregulation in SCLC

Disruption of the TAD boundary leads to aberrant chromatin folding and altered gene expression in different cancer types. The activation of proto-oncogenes was associated with the disruption of chromatin domains in T-cell acute lymphoblastic leukaemia [36]. Somatic copy number alterations resulting in the formation of new TADs were associated with insulin-like growth factor-2 dysregulation in colorectal cancer [37]. A reduction in CTCF binding was found to result in the loss of insulation between adjacent TADs due to hypermethylation of the CTCF DNA binding site, and to cause aberrant gene activation, in isocitrate dehydrogenase mutant gliomas [38]. In this study, we used the TARGET framework to identify dysregulated genes associated with altered TAD boundaries in SCLC cell lines on a genome-wide scale.

Importantly, we showed that differentially expressed genes in SCLC are significantly enriched within close range of TAD boundaries with altered insulation. This result suggests that the aberrant expression of these genes is associated with TAD boundary alteration in SCLC. In particular, we showed that genes with the most expression alteration are the most strongly enriched near altered TAD boundaries, further supporting the potential functional role of the altered TAD boundaries identified using Hi-C data.

By comparing A/B compartment distribution between normal and SCLC cell lines, we identified many genomic regions with flipped compartments. These results indicate that pervasive alteration of the compartment landscape occurs during SCLC tumorigenesis. We also found that these altered TAD boundaries were significantly enriched in compartment-flipped regions. TAD boundaries with increased insulation tended to be more enriched in B-to-A compartment-flipped regions, whereas those with decreased insulation were more likely to be enriched in A-to-B compartment-flipped regions. We proved that this similarity in the direction of 3D chromatin structural alteration corresponds to a similarity in functional impact by showing that, similar to genes located in B-to-A compartment-flipped regions, those near gained or strengthened TAD boundaries tended to be overexpressed, whereas genes near lost or weakened TAD boundaries tended to be repressed, resembling the tendency of transcriptional repression observed in genes located in A-to-B compartment-flipped regions. We further showed that the synergistic alteration of TAD boundaries and A/B compartments had a stronger impact on gene expression than did unilateral changes. The significant enrichment of altered chromatin loops near TSSs of TARGET-identified candidate genes suggests that the formation and disruption of chromatin loops in SCLC explains the relationship between the alteration of chromatin architecture and gene transcriptional activity, as loop extrusion forms TADs and promotes gene expression [26,39]. Thus, these synergistic changes in chromatin structure features at different scales may collaboratively strengthen the regulatory basis of abnormally expressed genes that underlies the genesis of SCLC; the formation of novel chromatin loops or disruption of existing loops may underlie the relationship between chromatin architecture and gene transcriptional activity alteration. TARGET identified SOX2, a known oncogene and Yamanaka factor [40], as a dysregulated gene associated with a nearby altered TAD boundary in both SCLC cell lines. The synergistic increase in the nearby boundary insulation and the local B-to-A compartment flip represents a novel chromatin architectural basis underling SOX2 activation in SCLC. Future studies of changes in the binding profiles of architectural proteins, such as CTCF and YY1, at anchors of identified chromatin loops and TAD boundaries may confirm the casual effect of specific TAD boundary alteration on gene dysregulation. Enhancers are key regulators of gene expression whose

**Table 1**  
Clinical characteristics of 18 patients with SCLC whose samples were analyzed for NanoString verification.

Clinicopathologic parameters	No. Cases (N = 18)
<b>Age</b>	
<65	10 (55.6%)
≥65	8 (44.4%)
<b>Gender</b>	
Female	5 (27.8%)
Male	13 (72.2%)
<b>Bronchia invasion</b>	
Yes	8 (44.4%)
No	10 (55.6%)
<b>Pleura invasion</b>	
Yes	2 (11.1%)
No	16 (88.9%)
<b>Lymph node metastasis</b>	
Yes	12 (66.7%)
No	6 (33.3%)
<b>Necrosis</b>	
Yes	3 (16.7%)
No	15(83.3%)
<b>Ki67</b>	
< 50%	1 (5.6%)
≥ 50%	15 (83.3%)
<b>TTF-1</b>	
Yes	15(83.3%)
No	1(5.6%)
<b>CD56</b>	
Yes	16 (88.9%)
No	0 (0%)
<b>AE1/AE3</b>	
Yes	14 (77.8%)
No	0
<b>CK5/6</b>	
Yes	0 (0%)
No	4 (22.2%)
<b>pT-Status</b>	
Tx/T1	8 (44.4%)
T2	9 (50.0%)
T3	1 (5.6%)
<b>pN Status</b>	
Nx/N0	7 (38.9%)
N1	8 (44.4%)
N2	3 (16.7%)
<b>Distant metastasis</b>	
Mx/M0	17 (80.0%)
M1	1 (5.6%)
<b>Clinical Stage</b>	
I	6 (33.3%)
II	7 (38.9%)
III	3 (16.7%)
IV	1 (5.6%)

function depends largely on physical contact with their target genes, mediated by the 3D chromatin architecture. Further studies could involve joint investigation of tumor-specific enhancer landscapes and CTCF binding via H3K27ac and CTCF chromatin immunoprecipitation sequencing experiments to further characterize the interplay with aberrant chromatin structure alteration in SCLC.

### 3.2. Functional roles and potential implications of TARGET-identified abnormally expressed genes in SCLC

The molecular alterations caused by SCLC include the inactivation of oncogenes and tumor suppressor genes [41], as well as enzymes involved in chromatin remodeling, receptor tyrosine

kinases, and their downstream effectors [5]. In this study, we comprehensively analyzed somatic genome alterations in SCLC cell lines and patient samples, and identified 24 differentially expressed and novel candidate genes, some of which were related to SCLC tumorigenesis, progression, and prognosis. Most of the genes that were differentially expressed between normal and tumor tissues from the 18 patients play roles in the cell cycle and transcription (e.g., *EPCAM*, *CNTNAP2*, *GJA1*, and *CRTAC1*, which are involved in cell adhesion and interactions); genes including *CACNA2D3*, *SLC38A5*, and *CHRM2* encode membrane proteins responsible for nutrient transport or receptors. In addition, genes such as *EPCAM* and *PAX5* participate in immune regulation.

Although our analytical filters support the involvement of these candidate genes in SCLC pathogenesis, functional experiments are required to clarify the biological roles of most genes. The stem cell transcription factor *SOX2* plays a crucial role in the regulation of embryonic development and is associated strongly with the inhibition of neuronal differentiation [42]. *SOX2* overexpression has been described in all types of lung cancer tissue, and the greater expression in SCLC than in non-SCLC cell lines and tissues indicates that *SOX2* can serve as a universal marker for the diagnosis of human lung cancer. The suppression of *SOX2* using short-hairpin RNAs blocked the proliferation of *SOX2*-amplified SCLC lines [23,43]. *SOX2* has been reported to be expressed in synovial sarcoma as part of the switch/sucrose non-fermentable chromatin remodeling complex, which affects histone methylation [25], but the role of *SOX2* in SCLC remains poorly characterized; further research is required.

Furthermore, our study showed that *PAX5* and *SOX2* had similar trends in SCLC tumor tissues and cell lines, with greater expression in tissues. The enhancement of *PAX5* protein expression has been described in lung tumors of neuroendocrine origin, such as carcinoid tumors, large cell neuroendocrine carcinoma, and SCLC, but not in general non-SCLC [44–46]. *PAX5* has been demonstrated to directly and positively regulate the transcription of c-Met, which is highly expressed in SCLC primary tumors and cell lines and plays significant roles in cell motility and tumor metastasis [47]. In other research, *PAX5* transcripts were found to be up-regulated in several SCLC cell lines, but not in non-SCLC cell lines [48,49], which is inconsistent with our results for tumor cells and tissues.

The epithelial cell adhesion molecule signaling pathway is associated with the proliferation, differentiation, and adhesion of epithelial cancer cells; gene expression has been observed in 21.6% of patients with SCLC [50], inconsistent with our finding of significant *EPCAM* upregulation at the RNA level in tumor cells and tissues. More specifically, we found that *EPCAM* was up-regulated in SCLC tumor tissues without necrosis compared with those with necrosis. The literature contains little information on the relationship between *EPCAM* and tumor necrosis; in our future research, we will examine the functions and mechanisms of dys-regulated genes in relation to SCLC tumorigenesis, progression, and prognosis in greater detail.

Recently, Rhie et al. [14] investigated the alteration of TAD structures comprehensively by comparing Hi-C data from normal and prostate cancer cells. They found that prostate cancer-specific TADs were smaller and more transcriptionally active than were normal cell-specific TADs, which were frequently in transcriptionally repressed chromatin states. Like Rhie et al. [14], we identified potential functional impacts of TAD alteration in tumor cells compared with normal cells. Our finding of transcriptionally activated gene enrichment near TAD boundaries with increased ISs in SCLC may reflect a similar relationship between TAD alteration and altered gene expression. The increased ISs at TAD boundaries in our study reflect new or strengthened TAD boundaries, similar to the splitting of normal cell-specific TADs into smaller tumor-specific TADs, and the enrichment of transcriptionally activated

genes is consistent with the greater transcriptional activity of split TADs in cancer cells reported by Rhie et al. [14]. However, there are several differences between our study and that of Rhie et al. [14]. We identified TAD boundaries with decreased ISs in tumor cells, synergistic alteration of TAD boundaries and A/B compartments, and greater enrichment of genes with more transcriptional alteration near altered TAD boundaries; Rhie et al. [14] did not report similar findings. In addition, we identified SOX2 as a candidate gene whose activation was mediated by synergistic alteration at multiple levels of the chromatin architecture, whereas Rhie et al. [14] revealed the activation of AR at a different locus. Overall, our results and those of Rhie et al. [14] indicate an important functional role of the establishment of tumor-specific chromatin architecture, particularly tumor-specific TAD boundaries in aberrant gene activation, in different tumor types. Whereas Rhie et al. [14] investigated histone marks, CTCF binding, and enhancer activity and their relationships to functional TAD alterations in prostate cancer, we revealed a regulatory role of synergistic alteration of the multilevel chromatin structure in SCLC.

The limitations of the present study include the generation of one Hi-C data replicate for each cell line. Although the TARGET framework can identify altered TAD boundaries and candidate genes based on unreplicated paired Hi-C datasets, potential technical variation between Hi-C datasets is difficult to identify or rule out. The addition of more replicates to experiments could help to minimize such variation and improve the sensitivity of the identification of biological differences. In addition, the Hi-C data resolution that were used in this study was 20 kb. Although it is adequate to identify chromatin compartment, TADs and many chromatin loops, there remains a subset of chromatin loops that can be identified at higher resolution of Hi-C data [51], and may provide additional insights in the interplay between multi-level chromatin architectures.

In conclusion, this study revealed synergistic chromatin structure alteration of A/B compartments and TAD boundaries that led to aberrant expression of nearby genes in SCLC. Tumor-specific chromatin loops may facilitate the construction of this structural basis underlying the transcriptional dysregulation associated with this disease. By enabling the identification of candidate dysregulated genes associated with altered chromatin architecture, the TARGET framework is a novel and powerful tool that can be used to explore the relationship of chromatin structure alteration with gene dysregulation related to SCLC tumorigenesis, progression, and prognosis, thereby broadening our understanding of the functional impact of aberrant 3D chromatin architecture in cancer.

## 4. Methods

### 4.1. Cell culture

The NCI-H209 and DMS153 human lung carcinoma cell lines and the MRC-5 noncancerous human lung fibroblast cell line were provided by the China Infrastructure of Cell Line Resource. The cells were cultured in RPMI 1640 medium (Life Technologies, Inc.) with 20% fetal bovine serum (HyClone) containing 0.1 mM nonessential amino acids (Invitrogen) and 0.1 mM GlutaMAX (Invitrogen). The cells were maintained at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub>, and passaged once a week. A split ratio of 1:3 was used.

### 4.2. Cell crosslinking and Hi-C library preparation

Approximately  $2.5 \times 10^7$  cells were collected gently and resuspended thoroughly in fresh culture medium without serum. Cell clumps were broken up by pipetting up and down, and crosslinking

was achieved with the addition of 2.5 ml 37% formaldehyde in one shot (2% final concentration). The tubes were inverted quickly several times for thorough sample mixing, and the samples were incubated at room temperature (RT) for exactly 10 min with gentle inversion every 1–2 min. To quench the formaldehyde reaction, 5 ml 2.5 M glycine was added in one shot and mixed well, followed by sample incubation for 5 min at RT and then on ice for at least 15 min to stop crosslinking completely. The crosslinked cell suspension was then split and centrifuged at 800×g for 10 min. The supernatant was discarded by aspiration. The fixed cells were resuspended in 1 ml lysis buffer [10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% Igepal CA-630, 1/10 vol proteinase inhibitor cocktail (Sigma)] and then incubated on ice for 20 min. The nuclei were pelleted by centrifugation at 4 °C and 600×g for 5 min and then washed with 1 ml lysis buffer, followed by centrifugation under similar conditions. After washing twice with restriction enzyme buffer, the nuclei were resuspended in 400 µl restriction enzyme buffer and transferred to safe-lock tubes. Next, the chromatin was solubilized with dilute sodium dodecyl sulfate (SDS) and incubated at 65 °C for 10 min. After quenching of the SDS with Triton X-100 overnight, digestion was performed with four cutter restriction enzymes (400 units MboI) at 37 °C on a rocking platform.

The next steps were Hi-C specific. The DNA ends were marked with biotin-14-deoxycytidine triphosphate, and blunt-end ligation of crosslinked fragments was performed. The proximal chromatin DNA was religated with the ligation enzyme. The nuclear complex crosslinks were reversed by incubation with proteinase K at 65 °C. The DNA was then purified by phenol-chloroform extraction. Biotin-C was removed from nonligated fragment ends using T4 DNA polymerase. The fragments were then sheared to 200–600-bp lengths by sonication. The fragment ends were repaired using a mixture of T4 DNA polymerase, T4 polynucleotide kinase, and Klenow DNA polymerase. The biotin-labeled Hi-C samples were specifically enriched using streptavidin C1 magnetic beads. The fragment ends were added by A-tailing with Klenow(exo-), followed by the addition of Illumina paired-end sequencing adaptor by ligation mix. Finally, the Hi-C libraries were amplified using 12–14 PCR cycles and sequenced on the Illumina HiSeq platform. Interacting partners were identified with an Illumina HiSeq instrument and  $2 \times 150$ -bp reads.

### 4.3. RNA-seq data processing

The sequences were aligned to the hg19 reference genome using TopHat v2.1.0 with the default parameters. Gene expression was normalized using Cufflinks v2.2.0 [52], and differential expression analysis was performed using Cuffdiff [53]. Log<sub>2</sub>-fold changes in expression were calculated based on the FPKM. A pseudocount that equals to 1 was added to each FPKM value to avoid log transformation issues.

### 4.4. Hi-C data processing

The sequences were aligned to the hg19 reference genome using Bowtie2 with the default parameters. A tag directory was created for each cell line using the “MakeTagDirectory” command in HOMER v4.8 [54]. Normalized Hi-C contact maps were created using the “analysis HiC” command in HOMER and the “-coverage Norm” parameter to remove bias. Compartment score (CS)s were calculated as described previously under 200-kb resolution [55]. TADs were identified using the “findTADsAndLoops.pl” script in HOMER, and their boundaries were defined as the first and last loci. ISs were calculated using the “findTADsAndLoops.pl” script in HOMER and normalized by z-score transformation for each cell line [54,56].

#### 4.5. Differential Hi-C contact map creation

The genomic distance of each intrachromosomal Hi-C contact map cell was measured between the centers of the two loci corresponding to that cell and used to further normalize the Hi-C contact maps. Specifically, the normalized Hi-C contact frequencies of all cells with the same genomic distance in all intrachromosomal Hi-C contact maps of a cell line were z-score normalized. To create differential Hi-C contact maps for a group of altered TAD boundaries, the averaged Hi-C contact map of the normal cell line was then calculated using the z-score-normalized Hi-C contact maps:

$$M_{ij} = \frac{\sum_{k=1}^n m_{ij}}{n} \quad 1 \leq i, j \leq 51,$$

where  $M_{ij}$  represents the cell corresponding to the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of the averaged Hi-C contact map, and  $m_{ij}$  represents the cell corresponding to the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of the 51 bins  $\times$  51 bins local z-score-normalized Hi-C contact map centered at the  $k^{\text{th}}$  TAD boundary.  $n$  represents the total number of TAD boundaries in the group. An averaged Hi-C contact map of each SCLC cell line was created in the same way. Differential Hi-C contact maps were then created as follows:

$$M_{ij}^D = M_{ij}^N - M_{ij}^T \quad 1 \leq i, j \leq 51,$$

where  $M^N$  and  $M^T$  represent the averaged Hi-C contact maps of the normal and SCLC cell lines, respectively.

#### 4.6. Identification of candidate boundary alteration-associated genes

To quantify IS variation at specific loci between cell lines, we calculated the observed/expected (O/E) differential IS for each locus. We obtained expected differential ISs based on the assumptions that different loci with different ISs would differ in the expected degree of IS variation across Hi-C datasets, whereas loci with similar ISs would have similar IS variation. We applied Loess regression to regress the IS variation between two cell lines against the mean IS for all loci to obtain the expected IS variation at any IS level.

First, the differential IS (absolute IS difference) and the mean IS for each locus of the two cell lines being compared were calculated. Then, Loess regression was performed using the “stats.models.api.nonparametric.lowess” function in the stats.models.python library. We used the parameter “frac = 0.01” to prevent oversmoothing of the fitted results. Then, the O/E differential IS was calculated as the ratio of the differential IS and the fitted IS variation value. We used the cutoff of O/E differential IS = 2 to identify loci with high degrees of IS variation (“altered” ISs). Altered boundaries were defined as TAD boundary loci with altered ISs after the removal of boundaries gained in SCLC cell lines with decreased ISs and those lost in SCLC cell lines with increased ISs. Candidate boundary alteration-associated genes were identified as genes with TSSs within the effective range of 40 kb (unless specified otherwise) from any altered TAD boundary with significant differential expression between the normal cell line and the SCLC cell line being examined. We excluded genes with absolute FPKM log<sub>2</sub>-fold changes < 2, unless specified otherwise.

#### 4.7. Functional enrichment analysis

Gene functional enrichment analysis was performed using a web-based tool provided by The Gene Ontology Resource (<http://geneontology.org/>) [26].

#### 4.8. Patients and clinical samples

Eighteen patients who underwent surgical treatment with a final diagnosis of SCLC at Peking Union Medical College Hospital (Beijing, China) between 2012 and 2015 were enrolled in this study. Their clinicopathological characteristics are summarized in Table 1.

Fresh specimens were fixed in formalin at RT for 48 h and then embedded in paraffin. Then, 4- $\mu$ m-thick sections were cut from each block and stained with hematoxylin and eosin (H&E). Two experienced pathologists reviewed the morphology of the stained sections to confirm the final diagnosis [57].

This study was approved by the Institutional Review Board of Peking Union Medical College Hospital. All patients provide written informed consent to the use of their tissues for research purposes.

#### 4.9. Immunohistochemistry

Immunohistochemical staining of 4- $\mu$ m-thick sections was performed using the Autostainer Link 48 (Dako; Agilent Technologies, Inc.). The tissue epitopes were repaired using the automated water-bath heating process with a PT Link device (Dako; Agilent Technologies, Inc.). The sections were incubated in TRIS-EDTA retrieval solution [10 mM Tris, 1 mM EDTA (pH 9.0)] at 98 °C for 20 min, and then incubated for 20 min at RT with a primary antibody (in ready-to-use form or diluted according to the manufacturer's instructions), followed by anti-rabbit immunoperoxidase polymer (Envision FLEX/HRP) for 20 min at RT according to the manufacturer's instructions. They were then developed with freshly prepared 0.05% 3,3'-diaminobenzidine tetrahydrochloride, counterstained with hematoxylin, dehydrated, and mounted.

#### 4.10. NanoString nCounter assay

Sections were cut from each paraffin block, and the tumor and normal tissues (identified by examination of H&E staining) were collected with a scratcher. Then, total RNA was extracted using the RNeasy FFPE kit (cat. no. 73504; Qiagen, Germany) according to the manufacturer's instructions. The total RNA concentrations and quality were determined with a NanoDrop One spectrophotometer (Thermo Fisher Scientific, Madison, WI, USA). RNA integrity was assessed with a 4200TapeStation system (#G2991AA; Agilent, Denmark) according to the manufacturer's instructions (NanoString Technologies).

Hybridization was performed according to the nCounter Element 48-plex Assay Manual (NanoString Technologies). Gene expression data were filtered using quality control (QC) criteria according to the manufacturer's recommendations. Raw counts of QC-passed samples were normalized using three reference genes (ACTB, GAPDH, CLTC, GUSB, and TBP) as internal controls. All QC and normalization procedures were performed with nSolver Analysis Software v2.0; all data were log<sub>2</sub> transformed before further analysis. Student's *t* test was used to compare normalized expression values between SCLC and normal tissues.

#### 4.11. Antibodies

The following antibodies were used: FLEX monoclonal mouse anti-human synaptophysin (DAK-SYNAP clone, ready to use), monoclonal mouse anti-human chromogranin A (DAK-A3 clone), FLEX monoclonal mouse anti-TTF-1 (8G7G3/1 clone, ready to use), FLEX monoclonal mouse anti-human CD56 (123C3 clone, ready to use), FLEX monoclonal mouse anti-human Ki-67 antigen (MIB-1 clone, ready to use), FLEX monoclonal mouse anti-human cytokeratin (AE1/AE3 clone, ready to use), and FLEX monoclonal mouse anti-

human cytokeratin 5/6 (D5/16 B4 clone, ready to use; all from DAKO; Agilent Technologies, Inc.).

#### 4.12. Statistical analysis

Statistical analysis was performed using the SciPy v1.3.0 library in the Python v3.7.2 environment. Fisher's exact test was used to assess enrichment of one set of genomic regions within another set. The two-tailed *t* test was used to assess the significance of differences between two groups. *P* values < 0.05 were considered to be significant.

#### Availability of data and materials

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE154094.

#### Ethical approval and consent to participate

This study was approved by the ethics committees of the Peking Union Medical College Hospital.

#### Consent for publication

All authors have agreed to the publication of this manuscript.

#### Funding

This work was supported by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (nos. 2017-I2M-1-001, 2017-I2M-2-001), the National Natural Science Foundation of China (no. 31801112), and the Beijing Nova Program of Science and Technology (no. Z191100001119064).

#### CRediT authorship contribution statement

**Dan Guo:** Conceptualization, Resources, Writing – review & editing. **Qiu Xie:** Data curation, Writing – original draft, Formal analysis, Visualization. **Shuai Jiang:** Resources, Methodology, Data curation, Writing – original draft, Software, Formal analysis, Visualization. **Ting Xie:** Writing – original draft. **Yaru Li:** Formal analysis, Writing – review & editing. **Xin Huang:** Writing – review & editing. **Fangyuan Li:** Data curation, Formal analysis. **Tingting Wang:** Data curation, Formal analysis. **Jian Sun:** Formal analysis. **Anqi Wang:** Resources. **Zixin Zhang:** Resources. **Hao Li:** Writing – original draft. **Xiaochen Bo:** Conceptualization, Writing – original draft. **Hebing Chen:** Conceptualization, Methodology, Writing – review & editing, Software, Funding acquisition. **Zhiyong Liang:** Conceptualization, Writing – review & editing, Funding acquisition.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

We thank all of the patients who participated in the study for their cooperation.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.11.003>.

#### References

- [1] Tian Y, Zhai X, Han A, Zhu H, Yu J. Potential immune escape mechanisms underlying the distinct clinical outcome of immune checkpoint blockades in small cell lung cancer. *J Hematol Oncol* 2019;12:67.
- [2] Armstrong SA, Liu SV. Immune checkpoint inhibitors in small cell lung cancer: a partially realized potential. *Adv Ther* 2019;36(8):1826–32.
- [3] Regzedmaa O, Zhang H, Liu H, Chen J. Immune checkpoint inhibitors for small cell lung cancer: opportunities and challenges. *Onco Targets Ther* 2019;12:4605–20.
- [4] Schulze AB, Evers G, Kerkhoff A, Mohr M, Schliemann C, Berdel WE, et al. Future options of molecular-targeted therapy in small cell lung cancer. *Cancers* 2019;11(5):690. <https://doi.org/10.3390/cancers11050690>.
- [5] Teicher BA. Targets in small cell lung cancer. *Biochem Pharmacol* 2014;87(2):211–9.
- [6] Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. *Nat Rev Dis Primers* 2021;7:1–20.
- [7] Grigorova M, Lyman RC, Caldas C, Edwards PAW. Chromosome abnormalities in 10 lung cancer cell lines of the NCI-H series analyzed with spectral karyotyping. *Cancer Genet Cytogenet* 2005;162(1):1–9.
- [8] Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol* 2017;18(1). <https://doi.org/10.1186/s13059-017-1253-8>.
- [9] Pezzuto F, Fortarezza F, Lunardi F, Calabrese F. Are there any theranostic biomarkers in small cell lung carcinoma? *J Thor Dis* 2019;11(S1):S102–12.
- [10] Chi KR. The dark side of the human genome. *Nature* 2016;538(7624):275–7.
- [11] Taberlay PC, Achinger-Kawecka J, Lun ATL, Buske FA, Sabir K, Gould CM, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* 2016;26(6):719–31.
- [12] Wu P, Li T, Li R, Jia L, Zhu P, Liu Y, et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat Commun* 2017;8(1). <https://doi.org/10.1038/s41467-017-01793-w>.
- [13] Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* 2016;17(12):771–82.
- [14] Rhie SK, Perez AA, Lay FD, Schreiner S, Shi J, Polin J, et al. A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nat Commun* 2019;10(1). <https://doi.org/10.1038/s41467-019-12079-8>.
- [15] Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol Cell* 2017;67(5):837–852.e7.
- [16] Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* 2018;50(2):238–49.
- [17] Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* 2017;170(2):367–381.e20.
- [18] Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, et al. Transformation of accessible chromatin and 3D nucleosome underlies lineage commitment of early T cells. *Immunity* 2018;48(2):227–242.e8.
- [19] Johnston MJ, Nikolic A, Ninkovic N, Guilhamon P, Cavalli FMG, Seaman S, et al. High-resolution structural genomics reveals new therapeutic vulnerabilities in glioblastoma. *Genome Res* 2019;29(8):1211–22.
- [20] Zhang L, He A, Chen B, Bi J, Chen J, Guo D, et al. A HOTAIR regulatory element modulates glioma cell sensitivity to temozolomide through long-range regulation of multiple target genes. *Genome Res* 2020;30(2):155–63.
- [21] Tang Z, Li C, Kang B, Gao G, Li C, et al. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 45: W98–W102.
- [22] Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;45(10):1113–20.
- [23] Chen Si, Xu Y, Chen Y, Li X, Mou W, Wang L, et al. SOX2 gene regulates the transcriptional network of oncogenes and affects tumorigenesis of human lung cancer cells. *PLoS ONE* 2012;7(5):e36326.
- [24] Yang F, Gao Y, Geng J, Qu D, Han Q, et al. Elevated expression of SOX2 and FGFR1 in correlation with poor prognosis in patients with small cell lung cancer. *Int J Clin Exp Pathol* 2013;6:2846–54.
- [25] Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat Genet* 2012;44(10):1111–6.
- [26] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 2000;25(1):25–9.

- [27] Schnabel PA, Junker K. Pulmonary neuroendocrine tumors in the new WHO 2015 classification: start of breaking new grounds? *Der Pathologe* 2015;36:283.
- [28] Wang HY, Li ZW, Sun W, Yang X, Zhou LX, et al. (2019) Automated quantification of Ki-67 index associates with pathologic grade of pulmonary neuroendocrine tumors. *Chin Med J (Engl)* 132: 551–561.
- [29] Bernhardt EB, Jalal SI. Small cell lung cancer. In: Reckamp KL, editor. *Lung Cancer: Treatment and Research*. Cham: Springer International Publishing; 2016. p. 301–22.
- [30] Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thor Oncol* 2015;10(9):1243–60.
- [31] Rossi G, Bertero L, Marchiò C, Papotti M. Molecular alterations of neuroendocrine tumours of the lung. *Histopathology* 2018;72(1):142–52.
- [32] Ji P, Ding D, Qin Na, Wang C, Zhu M, Li Y, et al. Systematic analyses of genetic variants in chromatin interaction regions identified four novel lung cancer susceptibility loci. *J Cancer* 2020;11(5):1075–81.
- [33] Fernandez-Cuesta L, Peifer M, Lu X, Sun R, Ozretić L, Seidel D, et al. Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids. *Nat Commun* 2014;5(1). <https://doi.org/10.1038/ncomms4518>.
- [34] Simbolo M, Mafficini A, Sikora KO, Fassan M, Barbi S, Corbo V, et al. Lung neuroendocrine tumours: deep sequencing of the four World Health Organization histotypes reveals chromatin-remodelling genes as major players and a prognostic role for TERT, RB1, MEN1 and KMT2D. *J Pathol* 2017;241(4):488–500.
- [35] Tang Z, Kang B, Li C, Chen T, Zhang Z (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 47: W556–w560.
- [36] Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016;351(6280):1454–8.
- [37] Weischenfeldt J, Dubash T, Drains AP, Mardin BR, Chen Y, Stütz AM, et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017;49(1):65–74.
- [38] Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 2016;529(7584):110–4.
- [39] Agaimy A, Hartmann A. Head and neck neoplasms: news from the WHO classification of 2017. *Pathologe* 2018;39:1–2.
- [40] Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell* 126: 663–676.
- [41] Sabari JK, Lok BH, Laird JH, Poirier JT, Rudin CM. Unravelling the biology of SCLC: implications for therapy. *Nat Rev Clin Oncol* 2017;14(9):549–61.
- [42] Bylund M, Andersson E, Novitch BG, Muhr J. Vertebrate neurogenesis is counteracted by Sox1–3 activity. *Nat Neurosci* 2003;6(11):1162–8.
- [43] Lu Y, Futtner C, Rock JR, Xu X, Whitworth W, Hogan BLM, et al. Evidence that SOX2 overexpression is oncogenic in the lung. *PLoS ONE* 2010;5(6):e11022.
- [44] Zhou X, Meng Q, Xu X, Zhi T, Shi Q, Wang Y, et al. Bex2 regulates cell proliferation and apoptosis in malignant glioma cells via the c-Jun NH2-terminal kinase pathway. *Biochem Biophys Res Commun* 2012;427(3):574–80.
- [45] Naderi A, Teschendorff AE, Beigel J, Cariati M, Ellis IO, et al. BEX2 is overexpressed in a subset of primary breast cancers and mediates nerve growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines. *Cancer Res* 2007;67:6725.
- [46] Yeting H, Qian X, Haiyan C, Jinjie H, YINUO T, et al. BEX2 promotes tumor proliferation in colorectal cancer. *Int J Biol Sci* 2017;13:286–94.
- [47] Taguchi A, Taylor AD, Rodriguez J, Çeliktaş M, Liu H, Ma X, et al. A search for novel cancer/testis antigens in lung cancer identifies VCX/Y genes, expanding the repertoire of potential immunotherapeutic targets. *Cancer Res* 2014;74(17):4694–705.
- [48] Kanteti R, Nallasura V, Loganathan S, Tretiakova M, Kroll T, et al. PAX5 is expressed in small-cell lung cancer and positively regulates MET transcription. *Labor Investig A J Tech Methods Pathol* 2009;89:301–14.
- [49] Baumann Kubetzko FB, di Paolo C, Maag C, Meier R, Schäfer BW, et al. The PAX5 oncogene is expressed in N-type neuroblastoma cells and increases tumorigenicity of a S-type cell line. *Carcinogenesis* 2004;25:1839–46.
- [50] Obermayr E, Agreiter C, Schuster E, Fabikan H, Weinlinger C, Baluchova K, et al. Molecular characterization of circulating tumor cells enriched by a microfluidic platform in patients with small-cell lung cancer. *Cells* 2019;8(8):880. <https://doi.org/10.3390/cells8080880>.
- [51] Rao SP, Huntley M, Durand N, Stamenova E, Bochkov I, Robinson J, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159(7):1665–80.
- [52] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28(5):511–5.
- [53] Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013;31(1):46–53.
- [54] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38(4):576–89.
- [55] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, NY)* 2009;326(5950):289–93.
- [56] Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 2015;523(7559):240–4.
- [57] El-Naggar AK, Chan J, Takata T, Grandis J, Sloodweg P (2017) The 4th Edition of the Head and Neck WHO Blue Book: Editor's Perspectives. *Human pathology: S0046817717301715*.