

# 论文 1 及检索证明



www.nature.com/cddis

ARTICLE OPEN



## PHF13 epigenetically activates TGF $\beta$ driven epithelial to mesenchymal transition

Yating Sun<sup>1,2,7</sup>, Dan Li<sup>1,7</sup>, Hongmei Liu<sup>3,4,7</sup>, Yongye Huang<sup>5</sup>, Fanyu Meng<sup>1</sup>, Jiahao Tang<sup>1</sup>, Zhanjun Li<sup>6,8</sup> and Wanhua Xie<sup>1,8</sup>

© The Author(s) 2022

Epigenetic alteration is a pivotal factor in tumor metastasis. PHD finger protein 13 (PHF13) is a recently identified epigenetic reader of H3K4me2/3 that functions as a transcriptional co-regulator. In this study, we demonstrate that PHF13 is required for pancreatic-cancer-cell growth and metastasis. Integrative analysis of transcriptome and epigenetic profiles provide further mechanistic insights into the epigenetic regulation of genes associated with cell metastasis during the epithelial-to-mesenchymal transition (EMT) induced by transforming growth factor  $\beta$  (TGF $\beta$ ). Our data suggest PHF13 depletion impairs activation of TGF $\beta$  stimulated genes and correlates with a loss of active epigenetic marks (H3K4me3 and H3K27ac) at these genomic regions. These observations argue for a dependency of TGF $\beta$  target activation on PHF13. Furthermore, PHF13-dependent chromatin regions are enriched in broad H3K4me3 domains and super-enhancers, which control genes critical to cancer-cell migration and invasion, such as SNAIL1 and SOX9. Overall, our data indicate a functional and mechanistic correlation between PHF13 and EMT.

*Cell Death and Disease* (2022)13:487; <https://doi.org/10.1038/s41419-022-04940-4>

### INTRODUCTION

Epithelial-to-mesenchymal transition (EMT) is a cellular process that plays a central role in embryonic development, wound healing, and malignant tumor progression [1]. Abnormal activation of EMT confers multiple characteristics on carcinoma cells during cancer progression, involving stemness, invasiveness, and therapeutic resistance associated with high-grade malignancy [2–4].

The TGF $\beta$  pathway is one of the most well-characterized pathways that is known to induce EMT. In response, mesenchymal genes (N-cadherin, vimentin, fibronectin, etc.) and EMT-induced transcription factors (SNAIL1, SNAIL2, ZEB1, and TWIST) are activated. These EMT-TFs, in turn, repress epithelial genes and elevate the expression of mesenchymal genes. These widespread transcriptional changes are accompanied by complex epigenetic regulatory mechanisms, notably alterations in the post-transcriptional modification of histones. Interestingly, TGF $\beta$  treatment induces widespread alteration of histone modifications via increasing chromatin accessibility [5]. Furthermore, epigenetic-profiling studies have identified novel chromatin signatures, notably broad H3K4me3 domains and super-enhancers, that contribute to the initiation and development of cancers and provide potential targets for therapeutic intervention [6–8].

The broad H3K4me3 domains regulate embryonic development and human disease [8–12]. For example, the broad H3K4me3 domains are associated with the high transcriptional activity of tumor-suppressor genes in normal cells [6–8]. Interestingly, some studies have found newly formed broad H3K4me3 on oncogenes

in cancer development [6, 13]. In addition, WDR5 and MLL4, which function as epigenetic writers to facilitate broad H3K4me3, are required to activate mesenchymal genes and EMT-TFs [8, 10, 14, 15]. However, the importance of the newly formed broad H3K4me3 in cancer progression (e.g., in the EMT process) remains unclear.

Enhancers, signed by active histone markers, including H3K27ac and H3K4me1, interact with promoters to form chromatin loops, resulting in transcriptional activation. The epigenetic landscape of transcription activators has identified a novel epigenetic signature termed “super-enhancers” (SEs), which strongly activate the expression of genes that control cell identity [16–18]. Oncogenic SEs, which are formed in cancers rather than normal tissues, play critical roles in various kinds of tumors [18]. Interestingly, oncogenic SEs are more sensitive to perturbation than typical enhancers and show potential therapeutic epigenetic targets for many malignant tumors [19–26]. In addition, widespread chromatin opening of enhancers has been documented in TGF $\beta$ -induced EMT [5, 27]. However, whether TGF $\beta$ -induced EMT coincides with changes in SE activity is still unknown.

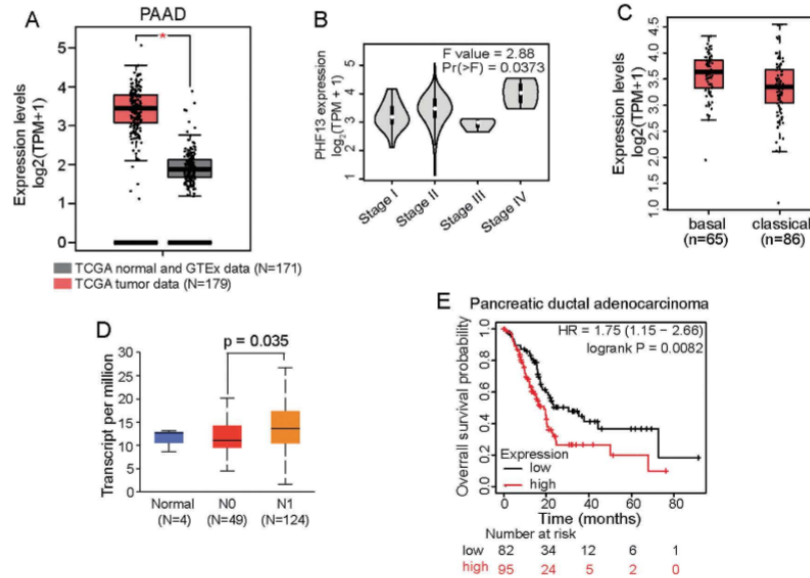
PHF13 is a plant homeodomain (PHD) finger-containing protein that plays an essential role in modulating the cell cycle [28], DNA damage repair [29], and chromatin structure [28]. A recent study found that PHF13 specifically recognizes H3K4me2/3 and functions as a transcription coactivator to facilitate gene transcription [30–32]. Interestingly, PHF13 also interacts with EZH2, a core subunit of the polycomb repressive complex 2 (PRC2), on a

<sup>1</sup>The Precise Medicine Center, Department of Basic Medical College, Shenyang Medical College, Shenyang 110034, China. <sup>2</sup>School of Pharmacy, Shenyang Medical College, Shenyang 110034, China. <sup>3</sup>CAS Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China. <sup>4</sup>Research Unit of Generation of Large Animal Disease Models, Chinese Academy of Medical Sciences (2019RU015), Guangzhou 510530, China. <sup>5</sup>College of Life and Health Sciences, Northeastern University, Shenyang 110169, China. <sup>6</sup>Key Laboratory of Zoonosis Research, Ministry of Education, College of Animal Science, Jilin University, Changchun 130062, China. <sup>7</sup>These authors contributed equally: Yating Sun, Dan Li, Hongmei Liu. <sup>8</sup>email: [liizj\\_1998@jlu.edu.cn](mailto:liizj_1998@jlu.edu.cn); [wanhua\\_xie@icloud.com](mailto:wanhua_xie@icloud.com)  
Edited by Dr. Anna Woloszyńska

Received: 21 October 2021 Revised: 4 May 2022 Accepted: 12 May 2022  
Published online: 21 May 2022

Official journal of CDDpress

SPRINGER NATURE  
CDDpress



**Fig. 1** High expression of *PHF13* is associated with human cancer. **A** Boxplot showing TPM-normalized expression of *PHF13* in normal tissues and pancreatic adenocarcinoma (PAAD). The results were generated by the online tool: GEPIA [73]. Given that there are only 4 normal pancreatic samples in TCGA data, we combined them with the pancreatic samples in GTEx data. **B** Violin plot showing the expression of *PHF13* at different clinical stages in patients with PAAD. The data was generated using the GEPIA tool [73]. **C** Boxplot comparing the TPM-normalized expression of *PHF13* between basal-like and classical subtypes. The data was generated using the GEPIA2 tool [73]. **D** Boxplot showing normalized expression of *PHF13* in normal pancreatic tissues (Normal), pancreatic tumors with no regional lymph node metastasis (N0), and pancreatic tumors with metastases in 1 to 3 axillary lymph nodes (N1) in the TCGA cohort. N number of patients. The result was generated from the online tool UALCAN [74]. **E** Kaplan–Meier plot showing the overall survival of patients based on *PHF13* expression. Data is generated from the webserver (<http://kmplot.com/analysis/>) [75]. HR hazard ratio, CI confidence interval.

particular chromatin region termed “bivalent,” which is co-occupied by the active histone marker H3K4me3 and repressive histone marker H3K27me3 [31]. Bivalent domains keep master regulators of cell fate transitions silent yet “poised” for rapid activation by extracellular signals, such as TGF $\beta$  [27, 33, 34]. However, the function of PHF13 on bivalent domains remains unclear. Highly expressed PHF13 has been observed in unresectable ovarian cancer and is associated with a lower survival rate [35]. However, its role and therapeutic targetability in human carcinomas remain unknown.

In this study, we demonstrate that PHF13-depleted pancreas ductal adenocarcinoma cells Panc-1 show a significantly decreased cell proliferation in vitro and in vivo. Moreover, cell culture-based assays reveal that PHF13 is required for TGF $\beta$ -induced EMT. Using the integrated transcriptomic and epigenetic analysis, we provide the first evidence that a subset of TGF $\beta$ -activated genes requires PHF13 to increase the activity of poised regions, broad H3K4me3 domains, and super-enhancers. Together, these findings suggest a novel epigenetic mechanism of PHF13 in promoting cancer-cell migration and identify PHF13 as a potential therapeutic target for human pancreatic cancer.

## RESULTS

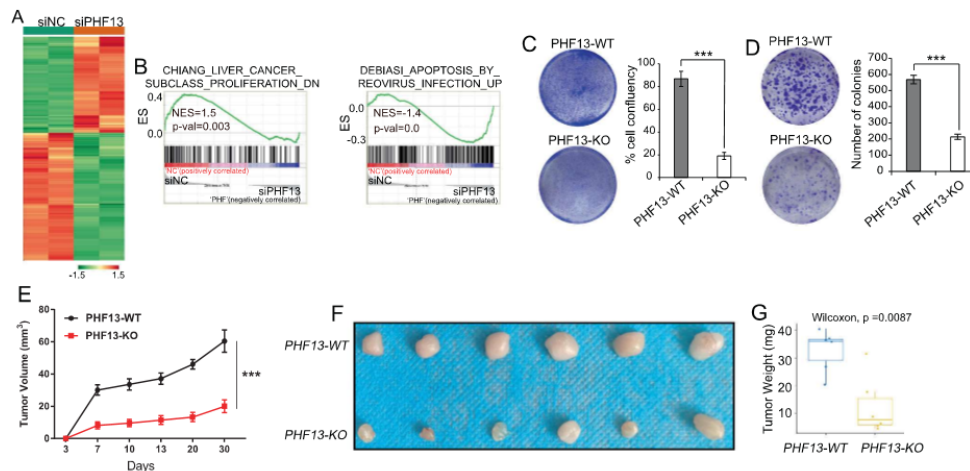
### High expression of PHF13 is associated with pancreatic adenocarcinoma

To understand the role of PHF13 in pancreatic cancer, we investigated its expression pattern in human tumors, compared

with normal tissues, and found an elevated expression of *PHF13* in pancreatic adenocarcinoma (Fig. 1A). Moreover, highly expressed *PHF13* was significantly correlated with high-grade pancreatic cancer (Fig. 1B). In addition, *PHF13* showed much higher expression in the basal-like subtype of pancreatic cancer that is characterized by aggressive activity, undifferentiated histopathology, and worse prognosis, compared to the classical subtype (Fig. 1C). The expression of *PHF13* was also significantly higher in pancreatic cancer with metastases in axillary lymph nodes (Fig. 1D). Importantly, Kaplan–Meier analysis demonstrated that high *PHF13* expression significantly worsened survival rates in patients with pancreatic ductal adenocarcinoma (Fig. 1E). Together, these results demonstrate the correlation between PHF13 expression and patient outcomes, indicating a potential oncogenic role for PHF13 in pancreatic cancer.

### Perturbation of PHF13 suppresses pancreatic cancer cell proliferation in vitro and in vivo

To determine the role of PHF13 in human pancreatic cancers, we performed mRNA-seq upon siRNA-mediated depletion of PHF13 in Panc-1 (Supplementary Fig. 1A). In support of the role of PHF13 in regulating mitosis [28], GSEA analysis of the differential expressed genes upon PHF13 depletion (Fig. 2A, Supplementary Table 3) revealed a significantly negative enrichment of the cell proliferation-related pathway (Fig. 2B, left panel) and over-representation of apoptosis-associated gene set in PHF13-depleted cells (Fig. 2B, right panel). The in vitro proliferation assay further verified the negative effect of PHF13 depletion on Panc-1 cell growth (Supplementary Fig. 1B).



**Fig. 2** Perturbation of *PHF13* suppresses pancreatic cancer cell proliferation in vitro and in vivo. **A** Heatmap showing the differential expressed genes following *PHF13* depletion in pancreatic adenocarcinoma cells Panc-1. The color key was shown at the bottom of the graph. The significantly regulated genes were selected based on: the absolute value of log2-fold change > 0.58, the  $p$ -adj < 0.05 **B** GSEA showing the significant enrichment of gene sets associated with cell proliferation (left panel), and apoptosis (right panel). NES, normalized enrichment score. **C** Cell proliferation assay for *PHF13*-WT and *PHF13*-KO Panc1. Image of crystal violet staining (left panel). The confluency of stained cells in each well was calculated by ImageJ (right panel). The data are shown as the mean  $\pm$  SD. \*\*\* $p$  < 0.001, calculated by two-side  $t$  test. all experiments were performed in triplicate. **D** Colony-forming assay for *PHF13*-WT and *PHF13*-KO Panc1 cells. The left panel indicates the image of crystal violet staining for the colonies. ImageJ was utilized to calculate the numbers of stained colonies in each well (right panel). Biological triplicate experiments were performed. The data are represented as the mean  $\pm$  SD. \*\*\* $p$  < 0.001, calculated by two-side  $t$  test. **E** Representative graphs showing the size of tumors formed by transplanting *PHF13*-WT and *PHF13*-KO Panc1 cells into the right groin of the mouse to establish an orthotopic model. Error bars represent the standard error of the mean between the six biological replicates. \*\*\* $p$  < 0.001, calculated by two-side  $t$  test. **F** The pictures showed the tumors formed by transplanting *PHF13*-WT and *PHF13*-KO Panc1 cells. The xenografts were harvested at 30 days. **G** boxplot showing the weight of tumors implanting *PHF13*-WT and *PHF13*-KO Panc1 cells for 30 days.  $P$  value was calculated by the unpaired Wilcoxon–Mann–Whitney Test.

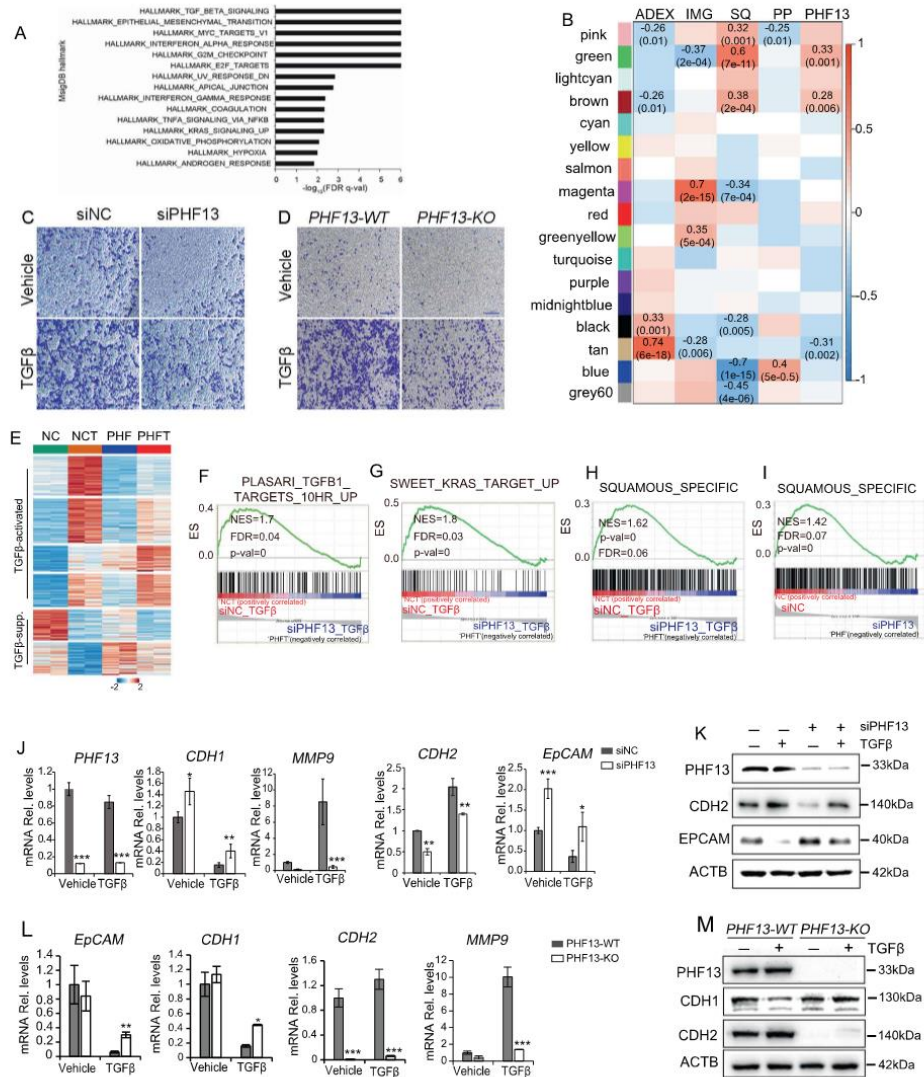
We next sought to substantiate these findings in an in vivo system. For this purpose, we utilized the CRISPR/Cas9 gene-editing tool to induce *PHF13* mutation in Panc-1 cells. Two guide RNAs were designed to specifically target exon 3 of the *PHF13* gene (Supplementary Fig. 1C). Allele mutation of *PHF13* in single-cell clone was screened by genotyping PCR and Sanger sequencing (Supplementary Fig. 1D, E). *PHF13* knockout consistently and significantly suppressed the proliferation and colony formation of Panc-1 cells (Fig. 2C, D and Supplementary Fig. 1F). Notably, *PHF13* deletion significantly reduced the size and weight of tumors formed by Panc-1 cells in vivo (Fig. 2E–G and Supplementary Table 4). Thus, in vitro and in vivo data provide further support for the role of *PHF13* in controlling pancreatic-cancer-cell proliferation.

#### Depletion of *PHF13* impairs TGF $\beta$ -induced EMT

Interestingly, GSEA further identified significant enrichment of cell migration-associated pathways involved in TGF $\beta$  signaling and EMT via comparing the transcriptomic profiles of *PHF13*-depleted and control Panc-1 (Fig. 3A). To confirm the findings in vitro, we analyzed the public available RNA-seq data of human pancreatic cancers to identify the differential expressed genes between *PHF13* highly and lowly expressed pancreatic cancers (Supplementary Fig. 2A–C and Supplementary Table 5). Significantly, GO enrichment analysis of the differentially expressed genes identified the extracellular matrix organization pathway, associated with cell migration (Supplementary Fig. 2D–F). Moreover, weighted gene co-expression network analysis (WGCNA) was performed to determine the correlations between the expression of *PHF13* and

the subtypes of pancreatic cancers. We demonstrated that the SQ trait-associated “green” and “brown” gene modules, which displayed significant enrichment of pathways involved in the extracellular matrix, tissue development, and cell junctions (Supplementary Fig. 2H, I), were significantly related to the high expression level of *PHF13* (Fig. 3B). Thus, those findings uncovered *PHF13* as a potential regulator of cell migration-related gene programs in human pancreatic cancer.

To confirm the potential role of *PHF13* in controlling tumor metastasis, we investigated the influence of *PHF13* depletion on TGF $\beta$ -induced EMT. Upon TGF $\beta$  stimulation, Panc-1 cells acquired a more pronounced mesenchymal phenotype than the *PHF13*-deleted cells (Supplementary Fig. 2J). In addition, the transwell migration assay revealed a significantly decreased migration potential in *PHF13*-depleted and -knockout Panc-1 (Fig. 3C, D). Furthermore, we performed a transcriptome-wide analysis of the effects of *PHF13* depletion on TGF $\beta$ -regulated genes. siRNA-mediated depletion of *PHF13* was confirmed by RT-qPCR and western blot analysis (Fig. 3J, K). K-means clustering coupled with a genome-wide pairwise comparison of heatmaps further revealed that the depletion of *PHF13* negatively or positively regulated a large group of TGF $\beta$ -controlled genes (Fig. 3E and Supplementary Table 3). GSEA confirmed significantly opposite effects of *PHF13* depletion on TGF $\beta$ -regulated genes (Fig. 3F; Supplementary Fig. 2K, L). Interestingly, *PHF13* depletion significantly downregulated the targets of *KRAS* (3G), which is frequently activated in most pancreatic adenocarcinomas and is known to be an essential driver of human cancer [36–38]. Moreover, the preferentially



expressed genes in the most aggressive SQ subtype of pancreatic cancers [39] were negatively enriched in PHF13-depleted cells (Fig. 3H, I). Furthermore, using RT-qPCR and western blot assay, we confirmed that PHF13 perturbation significantly upregulated epithelial markers (*CDH1* and *EpCAM*) and downregulated mesenchymal markers (*CDH2* and *MMP9*) (Fig. 3J–M). In contrast, re-expressing PHF13 in PHF13-KO Panc-1 resulted in a significant reduction of *CDH1* and gain of *CDH2* (Supplementary Fig. 2M).

These results suggest that PHF13 plays an essential role in controlling pancreatic-cancer-cell metastasis through maintaining and directing EMT-associated gene expression.

#### PHF13 is required for TGFβ-mediated activation of enhancers and promoters

To determine the epigenetic function of PHF13 in gene expression, we performed ChIP-seq analysis of PHF13. We further overlapped the

**Fig. 3 PHF13 is required for TGF $\beta$ -induced EMT.** **A** The top 15 significantly enriched MSigDB Hallmark gene sets via GSEA for the transcriptomic profiles of PHF13-depleted and control cells. FDR false discovery rate. **B** Heatmap showing the statistical significance of correlations between each subtype of pancreatic cancers, PHF13 expression levels, and gene modules identified by WGCNA. Bailey et al. determined the four molecular subtypes of pancreatic cancers: pancreatic progenitor (PP), immunogenic (IMG), aberrantly differentiated endocrine exocrine (ADEX), and squamous (SQ) [39]. The significantly correlated cells contain Student's asymptotic *P* values and Pearson correlations. Transwell migration assay showing a significant decrease in migration potential of TGF $\beta$ -treated cells in response to PHF13 depletion (**C**) or CRISPR/Cas9-mediated knockout (**D**). The migrated cells were stained with 0.1% (w/v) crystal violet (scale bar 100  $\mu$ m). The experiment was performed in duplicate. **E** Heatmaps showing the significantly regulated genes upon TGF $\beta$  treatment in Panc-1 cells. K-means clustering (*K* = 6) was performed based on the normalized expression level using the option scale = "row" during the following conditions: Control (NC), Control with TGF $\beta$  treatment (NCT), PHF13 knockdown (PHF), and TGF $\beta$ -treated cells depleted for PHF13 (PHFT). The color scale bar shows the row-wise normalized value. GSEA showing a significant decrease in the enrichment of TGF $\beta$  targets (**F**), KRAS targets (**G**), and squamous-specific genes (**H**, **I**) in PHF13-depleted Panc-1 cells. Squamous-specific genes were generated from the published data [39]. **J** RT-qPCR analysis of PHF13 knockdown efficiency, epithelial (*CDH1* and *EpCAM*), and mesenchymal (*CDH2* and *MMP9*) markers in Panc-1 cells in response to TGF $\beta$  treatment and PHF13 depletion. The data represent the mean  $\pm$  SD. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001, and calculated by two-side *t* test. **K** Western blots showing the protein level of PHF13, CDH2, and EpCAM in Panc-1 following TGF $\beta$  induction and PHF13 knockdown. The level of ACTB was set as a loading control. **L** RT-qPCR analysis of epithelial (*CDH1* and *EpCAM*) and mesenchymal (*CDH2* and *MMP9*) markers in PHF13-WT and PHF13-KO Panc1 following TGF $\beta$  treatment. **M** Western blot analysis of the protein level of PHF13, CDH1, and CDH2 in the given samples. ACTB was set as a loading control.

accessible sites (ATAC-seq) with the PHF13-bound sites to preferentially select the PHF13-occupied regions (Supplementary Fig. 3A). Interestingly, the PHF13-occupied regions were mainly distributed on TSS-proximal promoters and distal enhancers (Fig. 4A), in support of previous findings [31], heatmap analysis of the occupancy of H3K4me3, H3K27me3, and H3K27ac surrounding PHF13-bound regions revealed that PHF13 was enriched in the active and poised chromatin regions (Fig. 4B). Next, we analyzed ChIP-seq profiles of H3K4me3, H3K27me3, and H3K27ac to uncover the role of PHF13 in the TGF $\beta$ -triggered alterations in chromatin activity. Upon stimulation of TGF $\beta$ , the regions displaying differential occupancy of H3K4me3, H3K27me3, and H3K27ac in Panc-1 were identified via DiffBind analysis (Supplementary Fig. 3B). The pairwise comparison analysis of heatmaps further revealed that the occupancy of H3K4me3 and H3K27ac were decreased, whereas the occupancy of H3K27me3 was increased in a subset of TGF $\beta$ -activated regions in response to PHF13 depletion in TGF $\beta$ -treated cells compared to cells with TGF $\beta$  treatment alone (Fig. 4C).

To determine which chromatin states are highly dependent on PHF13, chromatin segmentation was performed, and 13 specific chromatin states were identified via ChromHMM [40] analysis of the H3K27me3, H3K36me3 [41], H2Bub1, H3K4me1 [42], H3K4me3, and H3K27ac occupancy profiles in Panc-1 cells. Based on the enrichment of each chromatin state on the CpG island, TSS, and TSS-proximal promoter (TSS  $\pm$  3 kb), the 13 chromatin states fit into nine functional classes: active promoter, weak promoter, poised promoter, strong enhancer, poised enhancer, strong transcribed, no/weak transcribed, no/weak signal, and polycomb-repressed regions (Fig. 4D). Boxplot further compared the occupancy of H3K4me3, H3K27ac, and H3K27me3 at each chromatin state. In general, TGF $\beta$  treatment in Panc-1 resulted in increased activity of each chromatin state except for poised enhancers and polycomb-repressed regions, showing significantly increased occupancy of H3K4me3 and H3K27ac and decreased occupancy of H3K27me3 compared to the control cells (Fig. 4E–I). In comparison to TGF $\beta$  treatment alone, PHF13 depletion in TGF $\beta$ -treated Panc-1 caused opposite effects (Fig. 4E–I). However, there were no visible changes in poised enhancers (Fig. 4J). In the polycomb-repressed regions, the occupancy of H3K27me3 was significantly increased in TGF $\beta$ -induced cells compared to the control cells, while slightly decreased in PHF13-depleted TGF $\beta$ -treated cells compared to cells with TGF $\beta$  treatment alone (Fig. 4K). Thus, we propose that depletion of PHF13 impairs TGF $\beta$ -triggered epigenetic activation.

#### The expression of TGF $\beta$ -stimulated poised genes is highly dependent on PHF13

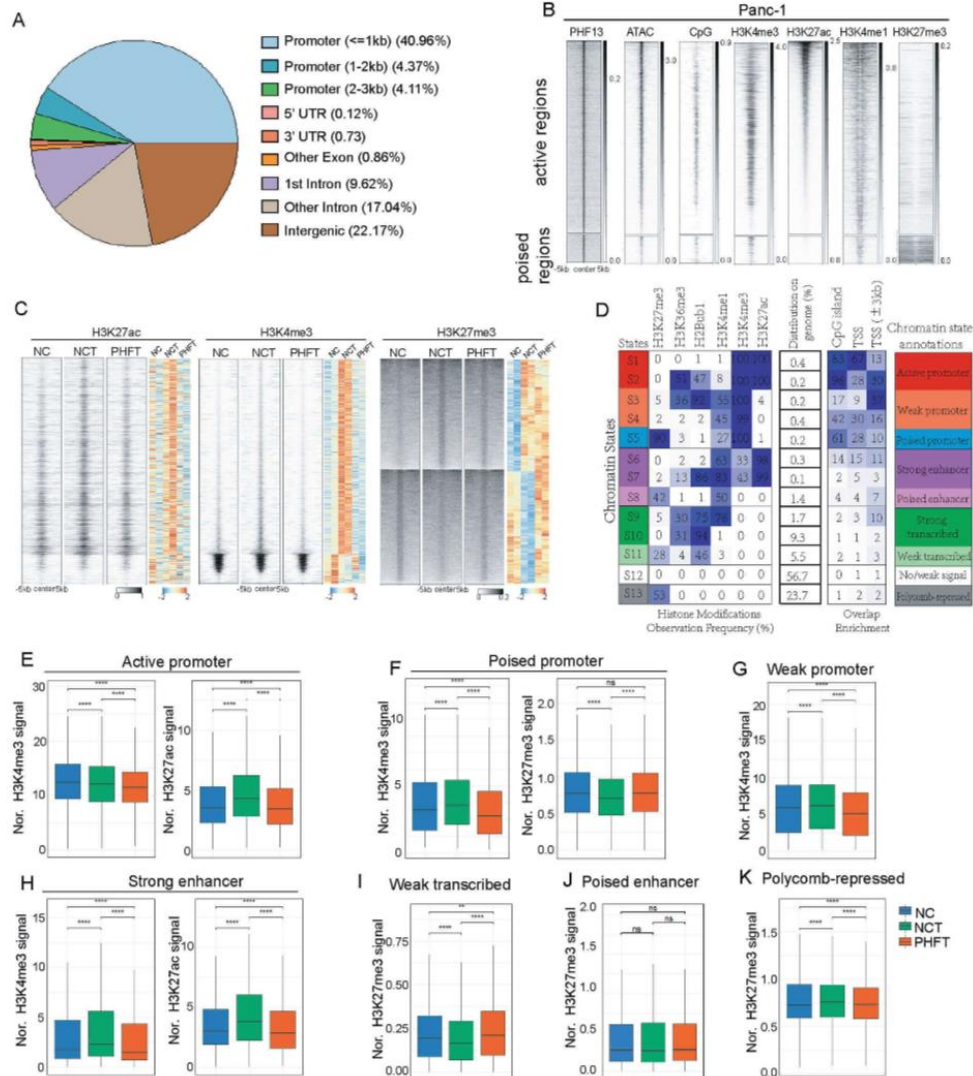
Given the essential role of the poised-chromatin state in cell fate determination [43], we further investigated the function of PHF13

on poised-chromatin regions in TGF $\beta$ -induced EMT. Aggregate profile analysis of H3K4me3, H3K27me3, and H3K27ac surrounding PHF13-occupied poised regions suggested a significantly increased occupancy of H3K27ac and decreased occupancy of H3K27me3 in TGF $\beta$ -treated cells relative to the control cells, whereas PHF13 depletion in TGF $\beta$ -treated cells significantly decreased H3K4me3 and H3K27ac occupancy and increased H3K27me3 occupancy compared to cells with TGF $\beta$  treatment alone (Fig. 5A). Pairwise comparison of heatmaps showed that a large fragment of TGF $\beta$ -stimulated poised genes was downregulated upon PHF13 depletion (Fig. 5B, Supplementary Table 3), supporting the alterations in the occupancy of histone modifications (Supplemental Fig. 4A). GO analysis of those genes revealed significant enrichment of pathways controlling the extracellular matrix and axonogenesis (Fig. 5C). Consistent with previous findings [31], endogenous co-immunoprecipitation (IP) assays showed that immunoprecipitation of PHF13 caused a co-IP of EZH2 (Fig. 5D) and vice versa (Fig. 5E). Interestingly, western blot analysis in Panc-1 showed a visibly increased level of H3K27me3 upon PHF13 depletion (Fig. 5F). This observation was mirrored by depleting PHF13 in HCT116 and Bel-7404 cells (Supplementary Fig. 4B). Consistently, boxplot analysis confirmed the significantly increased occupancy of H3K27me3 in PHF13-depleted and TGF $\beta$ -treated cells (Supplementary Fig. 4C). The effect of PHF13 depletion on H3K27me3 occupancy was confirmed at individual PHF13-dependent and TGF $\beta$ -stimulated poised genes (*PCDH8* and *PAPLN*) via the genomic browser and ChIP-qPCR analysis (Fig. 5G, H). Thus, the data suggest that PHF13 is required for TGF $\beta$ -induced resolution of the poised states to active states.

#### TGF $\beta$ -activated broad H3K4me3 domains are sensitive to PHF13 depletion

To investigate the epigenetic function of PHF13 on TSS-proximal regions, we divided TSS-proximal promoters into five clusters based on histone modifications as previously reported [44]: broad H3K4me3 occupied (broad), typical H3K4me3 occupied (active), H3K4me3/H3K27me3 co-occupied (poised), only H3K27me3 occupied (H3K27me3 only), and no histone markers (neither) (Fig. 6A). Interestingly, the chromatin sites, which displayed differential occupancy of H3K4me3 in each condition, were significantly enriched in the broad H3K4me3 and poised chromatin regions (Fig. 6B). In support of this observation, the upregulated genes upon TGF $\beta$  stimulation were enriched in "poised" and "broad" clusters (Supplementary Fig. 5A), whereas most of the two cluster genes were downregulated upon PHF13 depletion (Supplementary Fig. 5B, C).

Given the importance of broad H3K4me3 domains in cell fate determination and human disease [8, 10, 11, 44], we further examined the role of PHF13 in controlling the broad H3K4me3



domain in TGF $\beta$ -induced EMT. Therefore, we selected the broad-H3K4me3-occupied genes in untreated, TGF $\beta$ -induced, as well as PHF13-depleted TGF $\beta$ -treated Panc-1 cells, as in the previous report [44] (Supplementary Fig. 5D, Supplementary Table 3). Interestingly, aggregate profile and boxplot analysis revealed that the H3K4me3 and H3K27ac peaks surrounding TSS of broad H3K4me3-linked genes were broadened in TGF $\beta$ -treated cells compared to the control cells, whereas narrowed in PHF13-depleted TGF $\beta$ -stimulated cells relative to cells treated with TGF $\beta$  alone (Fig. 6C–E), which corresponded to the downregulation of

those genes following PHF13 depletion (Fig. 6F). A similar effect was observed in the typical active regions, but not stronger than that in the broad H3K4me3 regions (Supplementary Fig. 5E). Furthermore, Venn diagram was performed to select the TGF $\beta$ -activated and broad H3K4me3-linked genes (231 + 28 genes showing broader H3K4me3 peaks in TGF $\beta$ -treated cells compared to the control) in Panc-1 (Fig. 6G). GO analysis further revealed that those selected genes were significantly enriched in pathways involved in NF-kappaB signaling, cell migration, and the mRNA catabolic process (Fig. 6H). Notably, we confirmed this finding in

**Fig. 4 PHF13 is required for the TGF $\beta$ -mediated activation of enhancers and promoters.** **A** Distribution of PHF13 occupancy on distinct chromatin elements. **B** Heatmaps show the occupancy of chromatin assembly (ATAC-seq), CpG island, H3K4me3, H3K27ac, H3K4me1, and H3K27me3 surrounding the center of PHF13 peaks ( $\pm 5$  kb). The color key of each heatmap is shown on the right side. **C** Pairwise heatmaps show the normalized density and distribution of the differentially changed H3K27ac, H3K4me3, and H3K27me3 in siNC (NC), siNC + TGF $\beta$  (NCT), and siPHF13 + TGF $\beta$  (PHFT) Panc1 cells. The differentially changed chromatin sites upon TGF $\beta$  treatment were identified by DiffBind (Supplementary Fig. 3B). The region flanking  $\pm 5$  kb of each peak center is shown. The right color heatmap shows the row-wise normalized signal of H3K27ac, H3K4me3, and H3K27me3. **D** ChromHMM was used to learn and characterize chromatin states in Panc1 cells. The first table from left to right shows the emission parameters determined de novo with ChromHMM from the given chromatin markers over the entire genome. The second table indicates the distribution of each characterized chromatin state on the genome. The third table displays the relative fold enrichment of each chromatin state on several chromatin elements, including CpG islands, TSS, and promoters (surrounding TSS  $\pm 3$  kb). Chromatin annotations for each state are shown on the rightmost table. Boxplots compare the occupancy of H3K4me3, H3K27ac, and H3K27me3 at the genomic regions corresponding to the identified states, including active promoter (**E**), poised promoter (**F**), weak promoter (**G**), strong enhancer (**H**), weak transcribed (**I**), poised enhancer (**J**), and polycomb-repressed (**K**) chromatin regions, in siNC (NC), siNC + TGF $\beta$  (NCT), and siPHF13 + TGF $\beta$  (PHFT) Panc1 cells. *p* value was calculated using the unpaired Wilcoxon–Mann–Whitney Test. ns nonsignificant difference; \*\**p* value greater than 0.001 but less than 0.01; \*\*\*\**p* value less than 0.0001.

the individual example of PHF13-bound and TGF $\beta$ -activated gene *SNAI1*, whose expression was dependent on PHF13 and highly associated with cancer metastasis (Fig. 6I–K). Furthermore, we compared the transcriptomic profiles of PHF13-depleted and control Panc-1 using the SNAI1-regulated gene set collection in LNCaP [45] and LS174T [46] cells. GSEA revealed a significant overrepresentation of the SNAI1-suppressed gene set and negative enrichment of the SNAI1-dependent gene set in PHF13-depleted cells (Supplementary Fig. 5F). ChIP-seq profiles showed a visibly decreased occupancy of H3K4me3 and H3K27ac at the promoter and enhancer of the SNAI1 target *MMP2*, which corresponded to the downregulation of *MMP2* in PHF13-depleted TGF $\beta$ -treated cells compared to cells treated with TGF $\beta$  alone (Supplementary Fig. 5G). Thus, the data suggest that PHF13 is essential in facilitating the transcription of TGF $\beta$ -activated oncogenes because it maintains the establishment of broad H3K4me3 domains.

Given that our data provide the first evidence that TGF $\beta$  selectively elevates the expression of oncogenes by controlling the broad H3K4me3 domains, we re-analyzed the public data in normal mouse mammary gland NMuMG cells to avoid any discrepancies [5]. Similarly, the TGF $\beta$ -induced and broad H3K4me3-linked genes in NMuMG cells were selected (Supplementary Fig. 5H, I), which were significantly enriched in pathways involved in cell adhesion, differentiation, and the TGF $\beta$  signaling (Supplementary Fig. 5J). Consistently, aggregate profiling analysis confirmed that TGF $\beta$  treatment resulted in a widespread increase of H3K4me3, H3K27ac, and H3K4me1 occupancy surrounding TSS of the TGF $\beta$ -triggered broad H3K4me3-linked genes (Supplementary Fig. 5K) but not on the random controls (Supplementary Fig. 5L). Thus, the role of TGF $\beta$  in controlling broad H3K4me3 domains is highly conserved across species.

#### PHF13 is required for the increased activity of a subset of super-enhancers in TGF $\beta$ -induced EMT

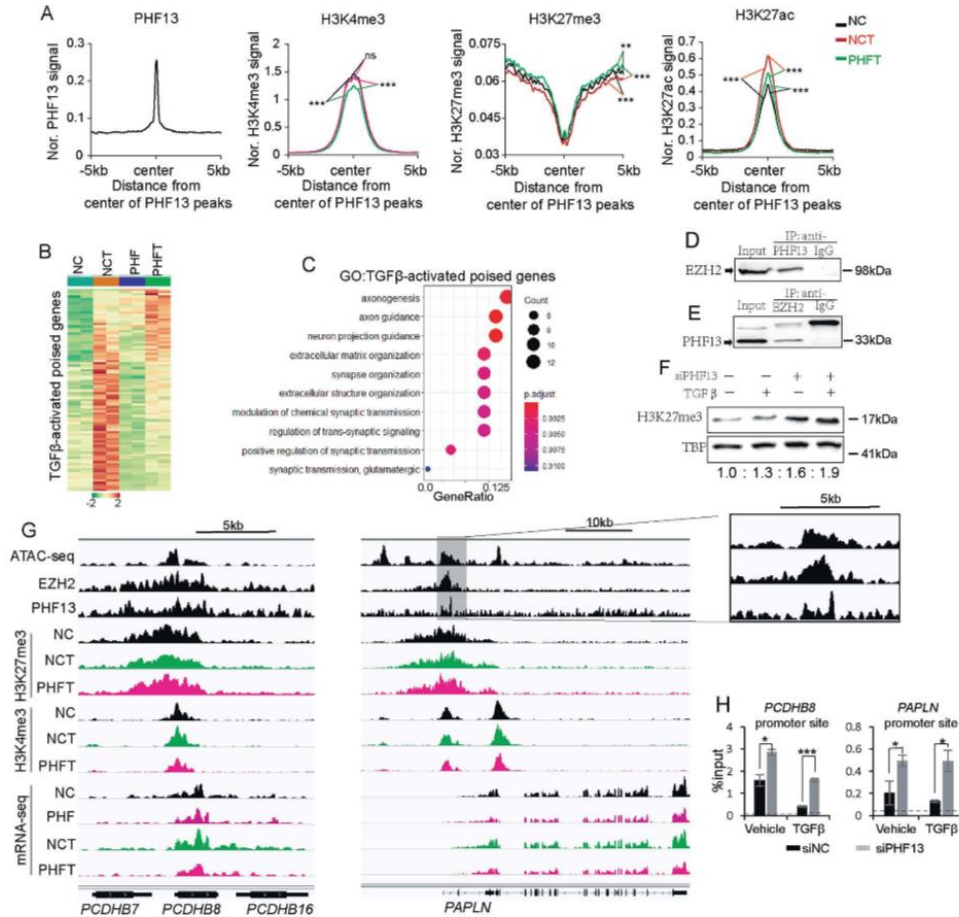
To further explain the changes in gene expression, we investigated alterations in enhancer activity through DiffBind analysis of H3K27ac occupancy (Supplementary Fig. 6A). The genome-wide pairwise comparison further revealed a decreased occupancy of H3K27ac at a large fragment of TGF $\beta$ -activated enhancers in PHF13-depleted TGF $\beta$ -treated cells compared to cells treated with TGF $\beta$  alone (Fig. 7A). Furthermore, Venn diagram analysis was utilized to select the PHF13-dependent enhancers (Supplementary Fig. 6A, B). The Genomic Regions Enrichment of Annotations Tool (GREAT) analysis revealed that the PHF13-dependent enhancers were significantly enriched in pathways controlling the extracellular matrix, cell migration, and EMT (Supplementary Fig. 6C). In addition, a de novo motif analysis was performed on the PHF13-dependent enhancers and identified motif enrichment of AP-1 members FOSL2, JUN, and FRA2, among others (Fig. 7B and Supplementary Fig. 6D). This finding was further confirmed by

mapping the given regions to the publicly available transcription factor binding source [47] (Fig. 7C). Interestingly, consistent with recent findings that TGF $\beta$  broadly activates AP1-dependent enhancers via upregulating the expression of AP1 family members [5], we observed that the expression of AP1 members (*FOSL2*, *JUNB*, *ATF3*, and *ATF6*) were increased by TGF $\beta$  treatment in Panc-1 cells (Supplementary Fig. 6E). Significantly, the mRNA levels of most AP1 members (*JUNB*, *ATF3*, *FOSL1*, *ATF5*, *JUND*, *JDP2*, *ATF1*, *ATF6*, and *BATF3*) were reduced upon PHF13 depletion (Supplementary Fig. 6E). Together, these data indicate that PHF13 controls the AP1-dependent enhancers via facilitating AP1 family gene expression, thereby suggesting that PHF13 plays a role in regulating AP1-dependent enhancer-mediated genes in TGF $\beta$ -mediated EMT.

Given the essential role of super-enhancers (SEs) in cancer metastasis, we next sought to examine the role of PHF13 in controlling SEs during TGF $\beta$ -induced EMT. Specifically, 594 SEs were distinguished from typical enhancers via ROSE analysis of H3K27ac density [17] (Supplementary Table 3). Upon TGF $\beta$  stimulation, most SEs showed increased H3K27ac occupancy in Panc-1, whereas the occupancy of H3K27ac was decreased in most SEs in PHF13-depleted and TGF $\beta$ -treated cells compared to cells with TGF $\beta$  treatment alone (Supplementary Fig. 6F). We further found that enhancers showing differential occupancy of H3K27ac were enriched at SEs in siNC\_TGF $\beta$  versus siNC and siPHF13\_TGF $\beta$  versus siNC\_TGF $\beta$  conditions, but not in siPHF13\_TGF $\beta$  versus siNC condition (Fig. 7D). In addition, the changes in H3K27ac occupancy were much stronger in SEs than in TE (Fig. 7E). K-means clustering and genome-wide pairwise comparison further revealed a decreased occupancy of H3K27ac at a large number of TGF $\beta$ -activated SEs in PHF13-depleted TGF $\beta$ -treated cells compared to cells with TGF $\beta$  treatment alone (Fig. 7F). Consistent with the alteration in chromatin activity, these SE-linked genes were significantly downregulated following PHF13 depletion (Supplementary Fig. 6G). Interestingly, the PHF13-dependent SEs were enriched in gene sets associated with cell metastasis, such as the extracellular matrix (Fig. 7G). Ranking the levels of the H3K27ac signal revealed that TGF $\beta$ -activated SEs were linked to the EMT-associated genes *MMP17*, *ITGA2*, *COL1A1*, and *SOX9* (Fig. 7H). This effect was confirmed at individual PHF13-dependent TGF $\beta$ -activated gene *SOX9* (Fig. 7I) and provides further support that the activation of a subset of SEs in TGF $\beta$ -induced EMT is highly dependent on PHF13.

#### DISCUSSION

Whole-genome sequencing of primary and metastatic tumors in the same patients has revealed limited heterogeneity of known driver mutations among all subclones [48]. Therefore, epigenetic alterations have been suggested as a pivotal factor in tumor metastasis [21, 49, 50]. Here, we showed that PHF13, a newly identified epigenetic reader of H3K4me2/3 [31, 32], is highly expressed in human pancreatic tumors characterized by

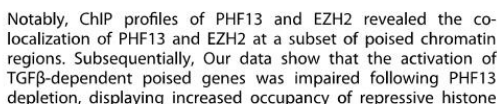


**Fig. 5** The expression of TGF $\beta$ -stimulated poised genes is highly dependent on PHF13. **A** Aggregate profiles show the occupancy of H3K4me3, H3K27me3, and H3K27ac surrounding the center of PHF13-occupied poised chromatin regions ( $\pm 5$  kb) in siNC, siNC + TGF $\beta$  (NCT), and siPHF13 + TGF $\beta$  (PHFT) Panc-1 cells. *P* values were calculated by Wilcoxon–Mann–Whitney Test. ns nonsignificant difference;  $^{**}0.01 < p < 0.001$ ;  $^{***}p < 0.001$ . **B** Heatmaps show gene expression of TGF $\beta$ -activated poised genes in siNC (NC), siPHF13 (PHF), siNC + TGF $\beta$  (NCT), and siPHF13 + TGF $\beta$  (PHFT) Panc-1 cells. The color scale bar shows the row-wise normalized expression signal. **C** GO analysis of the TGF $\beta$ -activated poised genes. The top 10 significantly enriched gene terms were shown. Western blot analysis of nuclear extract inputs, anti-PHF13 IP (**D**), anti-EZH2 IP (**E**), and anti-IgG IP from Panc1 nuclear extract. **F** Western blot showing the level of H3K27me3 in the given samples. The protein TBP was set as a loading control. The down numbers indicated the normalized signal of H3K27me3. The signal of each band was measured using ImageJ. The signal of H3K27me3 was normalized to TBP and represented as fold change relative to the control cells. **G** Genome browser tracks the normalized signal of chromatin assembly (ATAC-seq), EZH2, PHF13, H3K27me3, H3K4me3 occupancy, and mRNA-seq data at *PCDHB8* and *PAPLN* locus from siNC (NC), siPHF13 (PHF), siNC + TGF $\beta$  (NCT), siPHF13 + TGF $\beta$  (PHFT) Panc-1 cells. **H** ChIP-qPCR analysis of H3K27me3 occupancy near the TSS of the *PCDHB8* and *PAPLN* genes in siControl treated, PHF13-depleted, TGF $\beta$ -treated, and PHF13-depleted TGF $\beta$ -treated Panc-1 cells. *p* value was calculated using a two-side *t* test.  $^{*}p < 0.05$ ;  $^{***}p < 0.001$ . ChIP-qPCR for IgG was set as a negative control. The dotted line indicated the average signal of IgG.

metastasis. Utilizing a cell culture-based approach, we provide evidence that depletion of PHF13 impaired TGF $\beta$ -induced EMT in pancreatic ductal adenocarcinoma cell Panc-1. Mechanistically, integrated analysis of transcriptome and epigenetic data suggest that PHF13 acts at a subset of poised promoters, broad H3K4me3

domains, and SEs and is required for the expression of a subset of TGF $\beta$ -activated genes.

EZH2, a subunit of PRC2, shows context-dependent oncogenic and tumor suppressor functions [51–55]. For example, EZH2 could promote tumor progression by interacting with SNAIL1 and being



**Fig. 6** The broad H3K4me3 domain in TGF $\beta$ -induced EMT is highly dependent on PHF13. **A** The heatmaps show H3K4me3, H3K27me3, and H3K27ac on the regions flanking  $\sim$ 5 kb to 10 kb around the TSS of all genes in Panc1 cells. All TSS were clustered based on the occupancy of the given histone modifications: the broad H3K4me3 occupied TSS (broad, defined based on Supplementary Fig. 5A), the typical H3K4me3 occupied TSS (active), H3K4me3 and H3K27me3 co-occupied TSS (poised), only H3K27me3-occupied TSS (H3K27me3 only), and non-histone modifications marked TSS (neither). The color key at the right side of each heatmap indicates the occupied density. **B** Observed/Expected enrichment of regulated and unchanged H3K4me3 occupied sites calculated from siNC\_TGF $\beta$  versus siNC (up panel), siPHF13\_TGF $\beta$  versus siNC\_TGF $\beta$  (middle panel), and siPHF13\_TGF $\beta$  versus siNC (down panel) at "broad", "active", and "poised" TSS regions. Aggregate profiles compare H3K4me3 (**C**) and H3K27ac (**D**) occupancy in siNC, siNC\_TGF $\beta$ , and siPHF13\_TGF $\beta$  Panc-1 cells at the regions from upstream 5 kb to downstream 10 kb of the broad H3K4me3-occupied TSS. **E** Boxplot compares the length of the broad H3K4me3 and H3K27ac peaks in siNC, siNC\_TGF $\beta$ , and siPHF13\_TGF $\beta$  Panc-1 cells. The unpaired Wilcoxon–Mann–Whitney Test was used to calculate *p* values. **F** Boxplot compares the log<sub>2</sub>-fold changes in the expression of broad H3K4me3-occupied genes in the following conditions: siPHF13 versus siNC (PHF/NC), siNC + TGF $\beta$  versus siNC (NCT/NC), and siPHF13 + TGF $\beta$  versus siNC + TGF $\beta$  (PHFT/NCT). **G** Venn diagram displaying the numbers of overlap and unique broad H3K4me3-occupied genes between siNC, siNC\_TGF $\beta$ , and siPHF13\_TGF $\beta$  Panc-1 cells. The broad H3K4me3-occupied genes in various conditions were obtained according to Supplementary Fig. 5A. **H** GO analysis of the TGF $\beta$ -gained broad H3K4me3-linked genes (correlated to Fig. 6G, the "231" and "28" genes). The enrichment map displays the connection of the top 25 significantly enriched gene sets. Four functional biological modules have been identified, including NF-kappaB signaling, cell migration, mRNA process, and endoderm formation. **I** Genome browser tracks the normalized signal of chromatin assembly (ATAC-seq), PHF13, H3K27ac, H3K4me3 occupancy, and mRNA-seq data at *SNAI1* locus from siNC (NC), siPHF13 (PHF), siNC + TGF $\beta$  (NCT), siPHF13 + TGF $\beta$  (PHFT) Panc-1 cells. **J** Western blot analysis shows the expression of *SNAI1* in Panc-1 cells following TGF $\beta$  treatment and PHF13 depletion. **K** Western blots analysis shows the expression of *SNAI1* in PHF13-overexpressed Panc-1 cells.

modification H3K27me3 and decreased occupancy of active histone modifications H3K4me3 and H3K27ac in comparison to cells treated with TGF $\beta$  alone. These findings support the hypothesis that PHF13 antagonizes the catalytic activity of EZH2 on H3K27me3 to activate TGF $\beta$ -stimulated poised genes.

Broad H3K4me3 domains are tightly linked to genes controlling cell identity [10, 11, 44] and mark oncogenes in tumor cells [6, 7] or tumor suppressor genes [6] in normal cells [8]. Moreover, alterations in the width of H3K4me3 peaks on oncogenes in different tumor cells are associated with dysregulated transcription [6]. We further determined that TGF $\beta$  treatment in two different cell lines results in the widespread broadening of H3K4me3 peaks on metastasis-associated genes. Interestingly, the stabilization of the broad H3K4me3 domain stimulated by TGF $\beta$  was much more dependent on PHF13 than the typical H3K4me3-occupied regions. Following the depletion of PHF13, the length of TGF $\beta$ -triggered broad H3K4me3 domains was significantly narrowed. Thus, we suggest the epigenetic reader of H3K4me3 PHF13 recognizes and recruits to H3K4me3 to maintain TGF $\beta$ -induced broad H3K4me3 domains. Notably, this effect has been confirmed in the central EMT-TF *SNAI1*. In addition, depletion of PHF13 downregulated *SNAI1*-dependent genes, such as *MMP2* and *MMP9*. Thus, a novel epigenetic regulator pathway has been hypothesized to explain PHF13-mediated cancer-cell invasion, in which the H3K4me3 reader PHF13 stabilizes the TGF $\beta$ -activated broad H3K4me3 domain to maintain the highly transcribed level of *SNAI1*, thereby triggers a transcriptional program leading to EMT.

In conclusion, we have provided mechanistic insights to explain the role of PHF13 in cancer metastasis by employing genome-wide analysis to understand its effects on chromatin activity and gene expression during TGF $\beta$ -induced EMT. Furthermore, our data indicate that PHF13 is required for TGF $\beta$  to endorse EMT-related oncogene expression by stabilizing the broad H3K4me3 domains and SEs. Interestingly, PHF13 may have diverse molecular functions in gene transcription. In addition to cooperating with H3K4me2/3 to increase chromatin activity [31], PHF13 interacts with H3K9 KMTs (G9A, GLP, and SetDB1) to mediate H3K9me3-dependent chromatin condensation [29]. Notably, G9a, which shows elevated expression in multiple human cancers, plays an essential role in tumor metastasis [58]. Moreover, ChIP-seq for PHF13 in our study showed a large fragment of PHF13-bound regions not localized in chromatin assembly sites, indicating a potential role of PHF13 in the H3K9me3-marked highly compacted chromatin regions. Altogether, these findings reveal PHF13 as a novel therapeutic target for human cancers and provide a basis for

further investigation of its oncogenic role in various types of human cancer.

## MATERIALS AND METHODS

### Cell culture, plasmids, and siRNA transfection

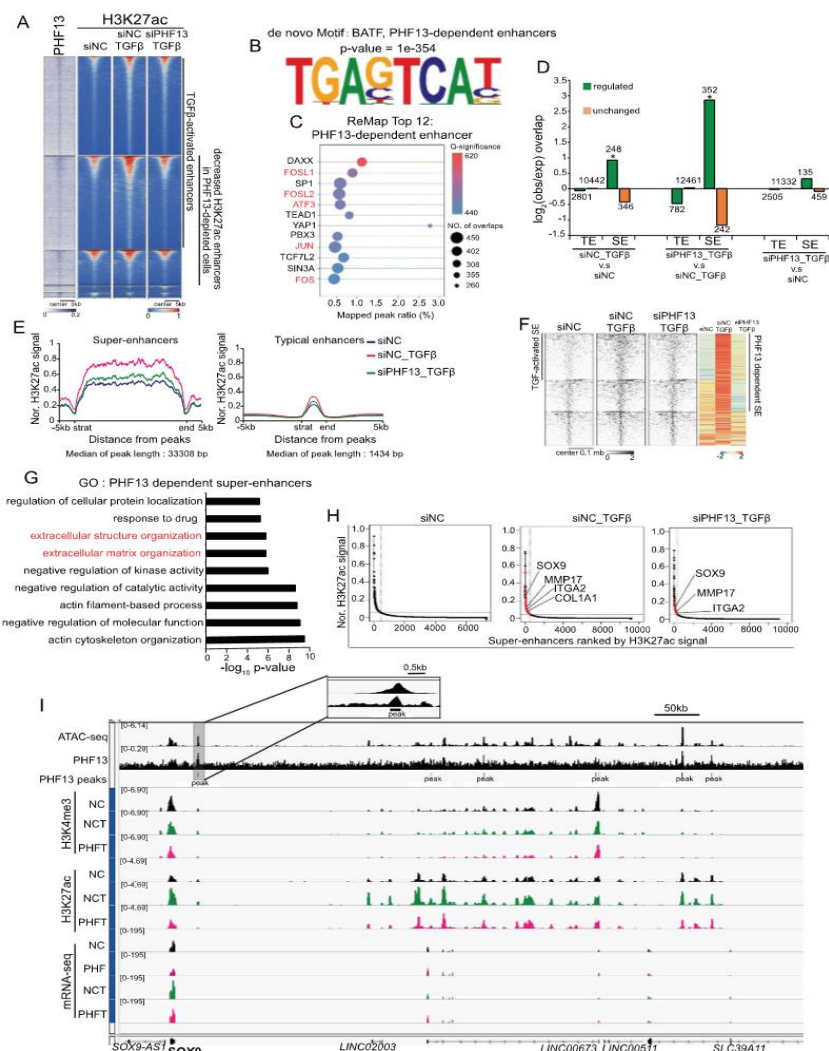
Panc-1 cells were kindly provided by Stem Cell Bank, Chinese Academy of Sciences. Panc-1 cells were authenticated by DNA profiling using twenty different and highly polymorphic short tandem repeat (STR) loci. Mycoplasma was also measured by PCR. Panc-1 cells were cultured in high-glucose DMEM/F12 (Thermo Fisher Scientific, New York, USA) supplemented with 10% FBS (10099, Thermo Fisher Scientific, Australia) and 1% penicillin-streptomycin (Thermo Fisher Scientific, New York, USA) at 37°C, 5% CO<sub>2</sub>. For plasmids transfection, cells were transfected with pcDNA3.1(+)-PHF13 overexpressed vector using FuGENE<sup>®</sup>HD reagent (E2311, Promega, Wisconsin, USA) according to the manufacturer's instructions. pcDNA3.1(+) empty vector was set as a negative control. siRNA transfections were performed using RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. Targeted human PHF13 siRNAs (Shanghai GenePharma Co., Ltd, China) containing the following sequences: 5'-CCCUUUAUCCGAAGGAGGAATT-3', 5'-UUCUCCUUCGGUAUAGG GTT-3', 5'-CCUGAUCAGGUCAAAGAAATT-3', 5'-UUUCUUUGACCUAGCAGGTT-3'. Non-targeting siRNA was used as a negative control (Shanghai GenePharma Co., Ltd, China). Afterward, Panc-1 cells were exposed to 10 ng/ml of TGF $\beta$  (P01137, Peprotech, USA) supplemented growth medium for 48 h to induce epithelial-mesenchymal transition.

### CRISPR/Cas9-mediated mutation of PHF13 in Panc-1 cells

PHF13 knockout Panc-1 cells were generated using the CRISPR/Cas9 system as described before [59]. In brief, guide RNAs targeting exon 3 of the PHF13 genes were designed using the MIT CRISPR design software and further cloned into a pSpCas9(BB)-2A-GFP (PX458) plasmids from Feng Zhang lab (Addgene plasmid #48138). According to the manufacturer's instructions, the constructed plasmids were transfected in Panc1 cells using the X-tremeGENE HP DNA transfection reagent (XTGHP-RO, Roche, USA). GFP+ cells were obtained by sorting 48 h after plasmids transfection. Then, the single-cell clones were expanded in 96-well plates for two weeks. Finally, the positive clones were selected by genotyping and sequencing. The additional information about gRNAs and genotyping primers are listed in Supplementary Table 1.

### The pancreatic cancer xenograft mouse model

Animal experiments were performed in Zhanjun Li's lab (Jilin University) and approved by the Institutional Animal Care and Use Committee of Jilin University. Female nude BALB/c mice (4 weeks old) were obtained from Liaoning Changsheng Biotechnology Co., Ltd. (Liaoning, China) and were randomly assigned to experimental groups, no blinding was performed. 200  $\mu$ l of  $2 \times 10^6$  Panc1 cells were injected into the right groin of each mouse. Tumor sizes were measured and calculated as  $\frac{1}{2}$  (length  $\times$  width<sup>2</sup>). After 30 days, the mice were sacrificed, and the formed tumors were weighed. Each cell line was injected into six mice.



### Western blotting, immunoprecipitation, and gene expression analysis

Protein extraction and western blot analysis were performed as previously described [60]. Cells were lysate in RIPA buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% NP-40, 0.1% SDS] with protease inhibitors. Following sonication, protein extraction was incubated with SDS loading dye at 95 °C

for 10 min. Proteins were further separated by PAGE and analyzed by western blotting using the specific antibodies that were listed in Supplementary Table 1.

Gene expression analysis was performed by quantitative real-time PCR (qRT-PCR) as described previously [60]. The primers for qRT-PCR were listed in Supplementary Table 1. In addition, the expression level of each gene was

**Fig. 7 PHF13 is required for the increased activity of a subset of super-enhancers in TGF $\beta$ -induced EMT.** **A** Pairwise heatmaps show the occupancy and distribution of PHF13 and H3K27ac over differential changed enhancers in siNC, siNC + TGF $\beta$ , and siPHF13 + TGF $\beta$  Panc-1 cells. Enhancers clustered according to the DiffBind analysis results (Supplementary Fig. 6A). TGF $\beta$ -activated enhancers, increased in siNC + TGF $\beta$  compared to siNC; PHF13-dependent enhancers, decreased in siPHF13 + TGF $\beta$  compared to siNC + TGF $\beta$ . The region flanking  $\pm 5$  kb of each enhancer center is shown. The color scale bar shows the row-wise normalized signal of H3K27ac. **B** The de novo motif sequence of PHF13-dependent enhancers. DNA sequence-based motif analysis was performed on the PHF13-dependent enhancers (the 1201 PHF13 peaks identified in Supplementary Fig. 6B). The most significantly enriched motif is shown here. **C** ReMap analysis shows the top 12 transcription factors that hit PHF13-dependent enhancers. **D** Observed/Expected enrichment of regulated and unchanged H3K27ac occupied sites calculated from siNC\_TGF $\beta$  versus siNC (left panel), siPHF13\_TGF $\beta$  versus siNC\_TGF $\beta$  (middle panel), and siPHF13\_TGF $\beta$  versus siNC (right panel) at TEs and SEs. **E** Metagene profiles show the normalized signal of H3K27ac at typical enhancers (TE) and super-enhancers (SE) in siNC, siNC + TGF $\beta$ , and siPHF13 + TGF $\beta$  Panc-1 cells. The length of enhancers (from "start" to "end") was calculated relative to the median length (TE, 1.4 kb; SE, 33 kb). **F** Pairwise heatmaps show the normalized density and distribution of H3K27ac over SEs in siNC, siNC + TGF $\beta$ , and siPHF13 + TGF $\beta$  Panc-1 cells. SEs were clustered using K-means ( $K = 3$ ) based on the normalized occupancy of H3K27ac. The regions flanking  $\pm 100$  kb of each SEs center are shown. The right color heatmap shows the row-wise normalized signal of H3K27ac and highlights the PHF13-dependent and TGF $\beta$ -activated SEs identified by K-means clustering. **G** GO analysis of PHF13-dependent SEs associated genes. The GREAT tool was used to predict SE annotated genes. The top 10 significantly enriched gene terms are shown. **H** Ranking of enhancers based on the normalized signal of H3K27ac in siNC, siNC + TGF $\beta$ , and siPHF13 + TGF $\beta$  Panc-1 cells. SEs are annotated by a GREAT online tool. The positions of SEs associated with *SOX9*, *MMP17*, *ITGA2*, and *COL1A1*, as well as the unique SEs in siNC + TGF $\beta$  or siPHF13 + TGF $\beta$  Panc-1 cells, are highlighted (red points). **I** Genome browser tracks the normalized signal of chromatin assembly (ATAC-seq), PHF13, H3K27ac, H3K4me3 occupancy, and mRNA-seq data at *SOX9* locus from siNC (NC), siPHF13 (PHF), siNC + TGF $\beta$  (NCT), siPHF13 + TGF $\beta$  (PHFT) Panc1 cells. One of the enhancers has been zoom out.

normalized to the expression of *RPLPO* and represented as fold change relative to the control cells. All experiments were performed in biological triplicate.

#### Cell migration assay

Cell migratory potential was analyzed in duplicate for Panc-1 cells by transwell migration assay as described previously [61]. In addition, 20,000 cells 24 h after being treated with 10 ng/ml of TGF $\beta$  were seeded into 8.0  $\mu$ m transparent PET membrane cell culture inserts (Corning, USA) and were cultured for another 48 h with TGF $\beta$  treatment. After fixation with methanol for 20 min, the migrated cells were stained with 0.1% (w/v) crystal violet diluted in PBS for 10 min. All experiments were performed in biological triplicate.

#### Colony formation assay

One thousand cells were seeded in each well of a 6-well plate. The cells were cultured for 12 days and then stained with 0.1% (w/v) crystal violet diluted in PBS. The numbers of colonies in each well were counted using ImageJ software. The experiments were performed in triplicate.

#### RNA sequencing (RNA-seq) and data analysis

48 h after siRNA transfection and TGF $\beta$  treatment, total RNA in Panc-1 cells was isolated using QIAzol reagent (79306, Qiagen, Germany) and quantitated using a Bioanalyzer 2100 (Agilent). The library was prepared using VAHTS mRNA-seq V3 Library Prep Kit according to the manufacturer's protocol (NR611, Vazyme, Nanjing, China) and sequenced using the paired-end 150-bp method on the Illumina HiSeq X platform. The experiments were performed in duplicate.

The paired-end raw reads were trimmed for adaptor sequence and low-quality sequence ( $-q 30$ ) using Trim Galore 0.4.1. The cleaned data were mapped on the human reference transcriptome (hg19) by STAR 2.7.0 [62]. Htseq-count 0.12.4 was used to count the reads mapped on the exon of each gene. Normalization and differential analysis of gene counts were further performed by the DESeq2 package [63]. Significant differential expressed genes were selected based on  $q$  value  $< 0.05$  and the absolute value of  $\log_2$  fold change  $> 0.58$ .

#### Gene ontology (GO) and gene set enrichment (GSEA) analysis

GO enrichment analysis for selected genes was performed with the R package "clusterprofile" [64]. The significantly enriched gene terms were determined based on "p-adjust  $< 0.05$ ". Similarly, GSEA software was used to identify a set of genes in the Molecular Signatures Database (MSigDB) whose expression shows coordinately changed in response to PHF13 depletion [65].

#### Chromatin immunoprecipitation (ChIP), ChIP-sequencing (ChIP-seq), and ChIP-qPCR

ChIP of histone modifications was performed as described previously [44]. 48 h after siRNA transfection and TGF $\beta$  treatment,  $10^5$  Panc-1 cells in 10 cm plates were crosslinked in 1% formaldehyde for 10 min and quenched using 125 mM of Glycine for 5 min. Cells were washed using ice-cold PBS

twice and lysed in nelson buffer [150 mM NaCl, 20 mM EDTA (pH 8), 50 mM Tris (pH 7.5), 0.5% NP-40, 1% Triton-X-100, 20 mM NaF]. Nuclei were pelleted and suspended in sonication buffer [150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 50 mM Tris-HCl, 20 mM EDTA, 20 mM NaF, 0.5% SDS]. The chromosome was fragmented in the Bioruptor<sup>®</sup> Pico via sonicating 5 cycles. The fragmented chromosome samples were diluted in two volumes of dilution buffer [150 mM NaCl, 20 mM EDTA (pH 8), 50 mM Tris (pH 8.0), 1% NP-40, 1% Triton-X-100, 20 mM NaF, and 0.5% sodium deoxycholate]. The unspecific binding sites were blocked and pulled down by incubating with 100  $\mu$ l Sepharose 50% slurry for 1 h. The supernatants were incubated with each primary antibody (H3K4me3, H3K27me3, and H3K27ac) overnight at 4 °C. The DNA-protein complex was pulled down by incubating with protein A agarose beads for 1.5 h at 4 °C. The immunoprecipitated beads were washed twice with each ice-cold buffer: IP buffer [150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 50 mM Tris-HCl (pH 8.0), 20 mM EDTA, 20 mM NaF, 0.1% SDS], LiCl buffer [100 mM Tris-HCl (pH 8.5), 500 mM LiCl, 1% v/v NP-40, 1% w/v sodium deoxycholate, 20 mM EDTA, and 20 mM NaF], IP buffer again, TE buffer [10 mM Tris-HCl (pH 8) and 1 mM EDTA (pH 8.0)]. RNAs were degraded in 10 mM Tris-HCl (pH 8.0) buffer supplemented with 0.2  $\mu$ g/ $\mu$ l RNase A. Immunoprecipitated DNA was eluted in high SDS buffer [50 mM Tris-HCl (pH 8), 10 mM EDTA (pH 8.0), 1% SDS] supplemented with 0.2  $\mu$ g/ $\mu$ l proteinase-K overnight at 65 °C. DNA was purified using phenol: chloroform: Isoamyl alcohol (25:24:1) and quantitated using Qubit 3.0 according to the manufacturer's protocol (Life technologies). The experiment was performed in duplicate.

ChIP of EZH2 and PHF13 was performed as previous descriptions [66, 67]. Briefly, Panc-1 cells in one 15 cm plate were crosslinked sequentially as the followings: for 40 min in 5 mM EGS (21565, Thermo Fisher Scientific, Michigan, USA) at room temperature, for 40 min in 1% paraformaldehyde (V900894, Sigma, Missouri, USA) at 4 °C, then for 10 min in 1% formaldehyde (252549, Sigma, Missouri, USA). Then, according to the manufacturer's protocol, ChIP was further processed using the ChIP-IT High Sensitivity Chromatin Immunoprecipitation kit (53040, Active Motif, California, USA). The rabbit polyclonal antibodies against human PHF13 for ChIP were newly designed and prepared by Wuhan ABclonal Technology company.

The sequencing library was prepared using the VAHTS Universal DNA Library Prep Kit according to the manufacturer's protocol (ND607, Vazyme, Nanjing, China) and sequenced on the HiSeq X (Illumina) platforms.

For ChIP-qPCR, Input DNA (10% of the amount utilized for immunoprecipitation) was prepared from chromatin extracts and used for the normalization of ChIP samples. The occupancy of histone modifications was determined by RT-qPCR using the indicated primers (Supplementary Table 1), normalized to input DNA, and displayed as "% of input".

#### ChIP-seq data analysis

The quality of sequencing reads was examined with the FastQC package and ChIPQC analysis (Supplementary Table 2). Trim Galore (Version 0.4.1) was used to filter out the adaptor and low-quality reads ( $-q 30$ ). The cleaned reads were aligned to the reference hg19 genome using Bowtie 2 (Version 2.4.1 [68]), with default parameters. Samtools were used to save the mapped

paired reads. PCR duplicates in bam files were moved by using the Picard tool. MACS2 (Version 2.2.7.1) [69] was used to call narrow peaks for PHF13, H3K4me3, and H3K27ac with the parameters (-q 0.001) or broad peaks for EZH2 and H3K27me3 with the parameters (-q 0.1 -broad). The input sequencing data was set as a control. The height of H3K4me3 peaks was calculated using "refinepeak" in MACS2. Furthermore, the bedtools closest tool was used to calculate the distance between each peak of H3K4me3 or H3K27ac and the closest transcription start site (TSS). If the distance was less than 5000 bp, this corresponding histone modification was considered to be enriched on the genes, otherwise on a potential enhancer region.

Differential binding analysis of ChIP-seq data peaks called by MACS2 was performed using the R package "DiffBind" (Version 2.16.2 [70]). Significant differential binding peaks were determined using the DESeq2 method with the parameter ( $p < 0.05$ ).

The bamCoverage tool was used to generate a coverage track normalized by counts per million (CPM) (Version 3.5.0 [71]). Furthermore, the normalized score on each genomic region was called using computeMatrix. Finally, the average scores over the given genomic regions were calculated by plotProfile.

The normalized H3K27ac ChIP-seq signal was used to call super-enhancer with Rank Ordering of Super-Enhancers (ROSE) as described previously [17]. The GREAT (version 4.04) online tool was used to annotate the given enhancers on the basis of association rules: 5 kb upstream, 5 kb downstream, and 1000 kb max extension.

PHF13-bound enhancers were performed in the motif enrichment analysis. Then, the DNA sequences at the selected regions were further loaded to find motifs using HOMER (version 4.8.0) [72]. Similarly, the ReMap online tool (version 1.2) was used to determine whether a transcriptional regulator significantly co-localized with the prearranged regions [47].

## DATA AVAILABILITY

ATAC-seq data in Panc-1 cells were obtained from ENCODE project. H3K4me3, H3K4me1, and H3K27ac ChIP-seq data in mouse normal mammary gland NMuMG cells are available in NCBI GEO with the accession code "GSE140552". The deep sequencing data in this study have been submitted to the NCBI GEO database with the accession numbers: GSE164824 and GSE180064.

## REFERENCES

- Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*. 2019;20:69–84.
- Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. *Cell*. 2012;148:349–61.
- Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*. 2008;133:704–15.
- Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene*. 2010;29:4741–51.
- Guerrero-Martinez JA, Ceballos-Chavez M, Koehler F, Peiro S, Reyes JC. TGFbeta promotes widespread enhancer chromatin opening and operates on genomic regulatory domains. *Nat Commun*. 2020;11:6196.
- Gopi LK, Kidder BL. Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. *Nat Commun*. 2021;12:1419.
- Park S, Kim GW, Kwon SH, Lee JS. Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *FEBS J*. 2020;287:2891–902.
- Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet*. 2015;47:1149–57.
- Sankar A, Lerdup M, Manaf A, Johansen JV, Gonzalez JM, Borup R, et al. KDM4A regulates the maternal-to-zygotic transition by protecting broad H3K4me3 domains from H3K9me3 invasion in oocytes. *Nat Cell Biol*. 2020;22:380–8.
- Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdup M, et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*. 2016;537:548–52.
- Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell*. 2015;163:1281–6.
- Liu X, Wang C, Liu W, Li J, Li C, Kou X, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature*. 2016;537:558–62.
- Terranova C, Tang M, Maitiuheti M, Raman AT, Schulz J, Amin SB, et al. Bivalent and broad chromatin domains regulate pro-metastatic drivers in melanoma. Preprint at <https://www.biorxiv.org/content/10.1101/721480v1> (2019).
- Dhar SS, Zhao D, Lin T, Gu B, Pal K, Wu SJ, et al. MLL4 is required to maintain broad H3K4me3 peaks and super-enhancers at tumor suppressor genes. *Mol Cell*. 2018;70:825–41.e826.
- Wu MZ, Tsai YP, Yang MH, Huang CH, Chang SY, Chang CC, et al. Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. *Mol Cell*. 2011;43:811–22.
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–47.
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307–19.
- Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153:320–34.
- Huang CS, You X, Dai C, Xu QC, Li F, Wang L, et al. Targeting super-enhancers via nanoparticle-facilitated BRD4 and CDK7 inhibitors synergistically suppresses pancreatic ductal adenocarcinoma. *Adv Sci*. 2020;7:1902926.
- Nguyen TTT, Zhang Y, Shang E, Shu C, Torrini C, Zhao J, et al. HDAC inhibitors elicit metabolic reprogramming by targeting super-enhancers in glioblastoma models. *J Clin Invest*. 2020;130:3699–716.
- Andricovich J, Perkail S, Kai Y, Casasanta N, Peng W, Tzatsos A. Loss of KDM6A activates super-enhancers to induce gender-specific squamous-like pancreatic cancer and confers sensitivity to BET inhibitors. *Cancer Cell*. 2018;33:512–26. e518.
- Hamdan FH, Johnsen SA. Super enhancers - new analyses and perspectives on the low hanging fruit. *Transcription*. 2018;9:123–30.
- Wang X, Kutsch AP, Yamada M, Prokakis E, Botcher P, Tanaka K, et al. Bromodomain protein BRDT directs DeltaNp63 function and super-enhancer activity in a subset of esophageal squamous cell carcinomas. *Cell Death Differ*. 2021;28:2207–2220.
- Han J, Meng J, Chen S, Wang X, Yin S, Zhang Q, et al. YY1 complex promotes quaking expression via super-enhancer binding during EMT of hepatocellular carcinoma. *Cancer Res*. 2019;79:1451–64.
- Peng L, Jiang B, Yuan X, Qiu Y, Peng J, Huang Y, et al. Super-enhancer-associated long noncoding RNA HCLLS is activated by ZEB1 and promotes the malignancy of hepatocellular carcinoma. *Cancer Res*. 2019;79:572–84.
- Zhang C, Wei S, Sun WP, Teng K, Dai MM, Wang FW, et al. Super-enhancer-driven AJUBA is activated by TCF4 and involved in epithelial-mesenchymal transition in the progression of Hepatocellular Carcinoma. *Theranostics*. 2020;10:9066–82.
- Fuoye R, Iacobucci S, Pappa S, Estaras C, Lois S, Vicioso-Mantis M, et al. Lineage specific transcription factors and epigenetic regulators mediate TGFbeta-dependent enhancer activation. *Nucleic Acids Res*. 2018;46:3351–65.
- Kinkley S, Staeger H, Mohrmann G, Rohaly G, Schaub T, Kremmer E, et al. SPOC1: a novel PHD-containing protein modulating chromatin structure and mitotic chromosome condensation. *J Cell Sci*. 2009;122:2946–56.
- Mund A, Schubert T, Staeger H, Kinkley S, Reumann K, Kriegs M, et al. SPOC1 modulates DNA repair by regulating key determinants of chromatin compaction and DNA damage response. *Nucleic Acids Res*. 2020;48:10013–4.
- Fuchs A, Torroba M, Kinkley S. PHF13: a new player involved in RNA polymerase II transcriptional regulation and co-transcriptional splicing. *Transcription*. 2017;8:106–12.
- Chung HR, Xu C, Fuchs A, Mund A, Lange M, Staeger H, et al. PHF13 is a molecular reader and transcriptional co-regulator of H3K4me2/3. *Elife*. 2016;5:e10607.
- Zhao S, Chuh KN, Zhang B, Dul BE, Thompson RE, Farrelly LA, et al. Histone H3Q5 serotonylation stabilizes H3K4 methylation and potentiates its readout. *Proc Natl Acad Sci USA*. 2021;118:e2016742118.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006;125:315–26.
- Xi Q, Wang Z, Zaromytidou AI, Zhang XH, Chow-Tsang LF, Liu JX, et al. A poised chromatin platform for TGF-beta access to master regulators. *Cell*. 2011;147:1511–24.
- Mohrmann G, Hengstler JG, Hofmann TG, Ende S, Lee B, Stelzer C, et al. SPOC1, a novel PHD-finger protein: association with residual disease and survival in ovarian cancer. *Int J Cancer*. 2005;116:547–54.
- Forrester K, Almoguera C, Han K, Grizzle WE, Perucho M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature*. 1987;327:298–303.
- Hong SM, Vincent A, Kanda M, Leclerc J, Omura N, Borges M, et al. Genome-wide somatic copy number alterations in low-grade PanINs and IPMNs from individuals with a family history of pancreatic cancer. *Clin Cancer Res*. 2012;18:4303–12.
- Kanda M, Matthaei H, Wu J, Hong SM, Yu J, Borges M, et al. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. *Gastroenterology*. 2012;142:730–3. e739.
- Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531:47–52.

40. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9:215–6.
41. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
42. Zhang J, Lee D, Dhiman V, Jiang P, Xu J, McGillivray P, et al. An integrative ENCODE resource for cancer genomics. *Nat Commun*. 2020;11:3696.
43. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. *Genes Dev*. 2013;27:1318–38.
44. Xie W, Nagarajan S, Baumgart SJ, Kosinsky RL, Najafova Z, Kari V, et al. RNF40 regulates gene expression in an epigenetic context-dependent manner. *Genome Biol*. 2017;18:32.
45. Stylianou N, Lehman ML, Wang C, Fard AT, Rockstroh A, Fazli L, et al. Correction: a molecular portrait of epithelial-mesenchymal plasticity in prostate cancer associated with clinical outcome. *Oncogene*. 2019;38:2436.
46. Freihe V, Ronsch K, Mastroianni J, Frey P, Rose K, Boerries M, et al. SNAIL1 employs beta-Catenin-LEF1 complexes to control colorectal cancer cell invasion and proliferation. *Int J Cancer*. 2020;146:2229–42.
47. Cheneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res*. 2018;46:D267–D275.
48. Makohon-Moore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet*. 2017;49:358–66.
49. McDonald OG, Li X, Saunders T, Tryggvadottir R, Mentch SJ, Warmoes MO, et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat Genet*. 2017;49:367–76.
50. Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, et al. Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell*. 2017;170:875–88.e820.
51. Pasini D, Di Croce L. Emerging roles for Polycomb proteins in cancer. *Curr Opin Genet Dev*. 2016;36:50–8.
52. Kim KH, Roberts CW. Targeting EZH2 in cancer. *Nat Med*. 2016;22:128–34.
53. Ntziachristos P, Tsirogas A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med*. 2012;18:298–301.
54. Mallen-St Clair J, Soydaner-Azeloglu R, Lee KE, Taylor L, Livanos A, Pylayeva-Gupta Y, et al. EZH2 couples pancreatic regeneration to neoplastic progression. *Genes Dev*. 2012;26:439–44.
55. Koppens M, van Lohuizen M. Context-dependent actions of Polycomb repressors in cancer. *Oncogene*. 2016;35:1341–52.
56. Cao Q, Yu J, Dhanasekaran SM, Kim JH, Mani RS, Tomlins SA, et al. Repression of E-cadherin by the polycomb group protein EZH2 in cancer. *Oncogene*. 2008;27:7274–84.
57. Vanharanta S, Shu W, Brenet F, Hakimi AA, Heguy A, Viale A, et al. Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nat Med*. 2013;19:50–56.
58. Chen MW, Hua KT, Kao HJ, Chi CC, Wei LH, Johansson G, et al. H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer Res*. 2010;70:7830–40.
59. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281–308.
60. Bedi U, Scheel AH, Hennion M, Begus-Nahrman Y, Ruschhoff J, Johnsen SA. SUPT6H controls estrogen receptor activity and cellular differentiation by multiple epigenomic mechanisms. *Oncogene*. 2015;34:465–73.
61. Mishra VK, Subramaniam M, Kari V, Pitel KS, Baumgart SJ, Naylor RM, et al. Kruppel-like Transcription Factor KLF10 Suppresses TGFbeta-Induced Epithelial-to-Mesenchymal Transition via a Negative Feedback Mechanism. *Cancer Res*. 2017;77:2387–400.
62. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
63. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
64. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
65. Subramaniam A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.
66. Sen M, Wang X, Hamdan FH, Rapp J, Eggert J, Kosinsky RL, et al. ARID1A facilitates KRAS signaling-regulated enhancer activity in an AP1-dependent manner in colorectal cancer cells. *Clin Epigenetics*. 2019;11:92.
67. Zirkel A, Nikolic M, Sofiadis K, Mallm JP, Brackley CA, Gothe H, et al. HMG82 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. *Mol Cell*. 2018;70:730–44. e736.
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
69. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
70. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481:389–93.
71. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42:W187–191.
72. 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;38:576–89.
73. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45:W98–W102.
74. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19:649–58.
75. Nagy A, Munkacsy G, Györfy B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep*. 2021;11:6047.

## ACKNOWLEDGEMENTS

We acknowledge Shanghai Personal Biotechnology Co., Ltd for the help in performing high-throughput sequencing. We thank H. Liu in the Z. Li lab for performing the in vivo study. We thank Y. Huang, X. Wang, and Z. Li for the helpful discussion.

## AUTHOR CONTRIBUTIONS

The experiments were designed by W.X., Y.S., and D.L.; W.X. performed the bioinformatic analysis; Y.S., D.L., H.L., F.M., and J.T. performed the experiments; W.X., Y.H., and Z.L. wrote the paper. All authors read and approved the final manuscript.

## FUNDING

This work was funded by the National Natural Science Foundation of China (No. 81902873) and the Department of Science and Technology of Liaoning Province (No. 20180530084).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41419-022-04940-4>.

**Correspondence** and requests for materials should be addressed to Zhanjun Li or Wanhua Xie.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022