**OMTN, Volume *27***

## Supplemental information

# CBX2 and EZH2 cooperatively promote the growth and metastasis of lung adenocarcinoma

**Fei-Fei Hu, Hao Chen, Yang Duan, Bei Lan, Chun-Jie Liu, Hui Hu, Xu Dong, Qiong Zhang, Yi-Ming Cheng, Min Liu, An-Yuan Guo, and Chenghao Xuan**

# Supplementary Information

## Supplementary Methods

### Regulator Collection

We collected 129 protein-coding genes that function as histone methylation writers, erasers and readers from the HIStome database[1] and literature[2–5] (Table S1).

### Cell Viability Assays

Examination of cell viability was performed using Cell Counting Kit-8 (CCK-8) (DOJINDO, Tokyo, Japan) assay or the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay. For CCK-8 assays, cells transfected with control or specific siRNAs were cultured in a 96-well culture plate, and then 10 μl of CCK-8 was added to each cell at the indicated time after transfection for 1-4 hours. Absorbance was photometrically measured at 450 nm. For MTT assays, 0.5 mg/ml MTT was added to each well, and cells were incubated for 4 hours at 37 °C. Cells were washed with phosphate-buffered saline (PBS) and lysed for 30 minutes at room temperature with DMSO (Solarbio, Beijing, China). Absorbance was photometrically measured at 490 nm.

### Apoptosis Assay

Apoptosis was measured using a FITC annexin V apoptosis detection kit (SUNGENE BIOTECH, Tianjin, China) according to the manufacturer's instructions. Briefly, cells were digested with trypsin into a single-cell suspension and collected by centrifugation at 1,800 rpm for three minutes to remove the supernatant. The cells were re-suspended in 100 μl of binding solution and 5 μl of annexin V-FITC. After incubation at room temperature for 10

minutes in the dark, 400 μl of binding solution and 5 μl of propidium iodide were added to the cell suspension, which was then subjected to flow cytometry.

**EdU-incorporation Assay**

EdU (5-ethynyl-2'-deoxyuridine) assays were performed according to the manufacturer's instructions (C10310, RiboBio, Guangzhou, China). Briefly, cells were cultured in 96-well plate forty-eight hours after siRNA transfection. Twenty-four hours later, the culture medium was replaced with medium containing EdU for 2 hours. Cells were fixed in 4% paraformaldehyde and processed for immunofluorescence.

**Migration Assay**

A total of $5 \times 10^4$ A549 and H1299 cells were re-suspended in serum-free DMEM and placed in the upper chamber (BD Biosciences, San Jose, CA, USA) coated with 50 μl of matrigel (BD Biosciences). Then, the chamber was placed in a 24-well culture dish (BD Biosciences) containing 500 μl of DMEM with 10% FBS. After incubation for 8-48 hours at 37 °C, transmigrated cells were fixed with 4% paraformaldehyde for 15 minutes at RT, stained with 0.1% crystal violet for 5 minutes, and counted.

**Statistical analysis**

All statistical analyses were performed using R language and all tests were two-sided. Spearman's rho statistic was utilized to estimate a rank-based measure of association between two variables. This test was carried out if the data did not come from a bivariate normal distribution and without ties, such as mRNA expression of two genes in TCGA tumor samples. On the other hand, to deal with data with ties such as immunohistochemical data, a random disturbance that obeys a normal distribution with a standard deviation of 0.01 was added. Correlation coefficient ($|r| \geq 0.3$) and FDR (<0.05, corrected p value) were treated as

standards to measure the association. Finally, scatters were fitted through linear models method.

Nonparametric tests (Wilcoxon rank-sum test for independent groups and Wilcoxon signed-rank test for paired groups) were used to compare the median values of two sets of continuous variables. Wilcoxon signed-rank test was performed for a comparison between tumor and normal samples from the same patient. For unpaired samples, such as the difference in immunohistochemical results between tumor and normal samples or between samples with or without distant metastasis, Wilcoxon rank-sum test was performed. Student's t test was performed to compare the difference between different experiment groups (n from 3 to 6).

**Pathway Enrichment**

KEGG was performed by KEGG function in R package clusterProfiler[6] with FDR control for 8,806 sorted (by fold change) genes, which were filtered by NOISeq with threshold CPM > 1 and FDR < 0.1 (prob > 0.9). This process was carried out to avoid losing information regarding genes with important functions that did not meet the differential expression filter conditions and to determine whether the gene set in KEGG pathways showed significant differences between the up- and downregulated gene sets. The enrichment score represents the degree of enrichment of pathway genes between two ends of differential expression (up and down). KEGG pathways were classified into up- and downregulated pathways using the enrichment score, and the downregulated pathways was focused. Moreover, to obtain the downregulated EZH2 and CBX2 targets involved KEGG pathways and due to the small size of this gene set, enrichKEGG function of R package clusterProfiler was executed without FDR control but with p value filtered ($p < 0.05$).

## Tumor Suppressor Gene and Oncogene Collection

Through the efforts of other studies, many cancer-related genes, including tumor-suppressor genes (TSGs) and oncogenes, have been revealed. Thanks to their efforts, 693 TSGs and 1,942 oncogenes were collected from 6 resources. And 525 lung cancer-related TSGs were obtained from the TSGene database[7]; 2,027 oncogenes were obtained from the Brushman laboratory (http://www.bushmanlab.org/links/genelists); 803 oncogenes were obtained from the oncogene database (http://ongene.bioinfo-minzhao.org/download.html); 574 cancer-related genes were obtained from the Cancer Gene Census (CGC) module from COSMIC database (https://cancer.sanger.ac.uk/cosmic/census?tier=1); 125 cancer-related genes were obtained from previous research[8]; 394 oncogenes and 1,031 TSGs were obtained from the UniProt database by searching "Proto-oncogene" and "Tumor Suppressor", and from the literature. Next, gene ID was normalized by R packages clusterProfiler and org.Hs.eg.db. Then, conflict screening, which eliminated genes with different functions recorded by different resources, was performed.

## Code Availability

The R script code we used to generate the figures and for basic data analysis is under github: https://github.com/Huffyphenix/LUAD_pic_code.

## References

1. Khare, S.P., Habib, F., Sharma, R., Gadewal, N., Gupta, S., and Galande, S. (2012). HIstome—a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res *40*, D337–D342.
2. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381–395.
3. Kouzarides, T. (2007). SnapShot: Histone-modifying enzymes. Cell *131*, 822.
4. Mohan, M., Herz, H.-M., and Shilatifard, A. (2012). Snapshot: Chromatin Lysine Methylase Complexes: Transcriptional Regulation and Epigenetics. Cell *149*, 498-498.e1.
5. Kutateladze, T.G. (2011). SnapShot: Histone Readers. Cell *146*, 842-842.e1.

6. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS : a Journal of Integrative Biology *16*, 284.
7. Zhao, M., Kim, P., Mitra, R., Zhao, J., and Zhao, Z. (2016). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res *44*, D1023–D1031.
8. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science *339*, 1546–1558.

## Supplementary Figures Legends

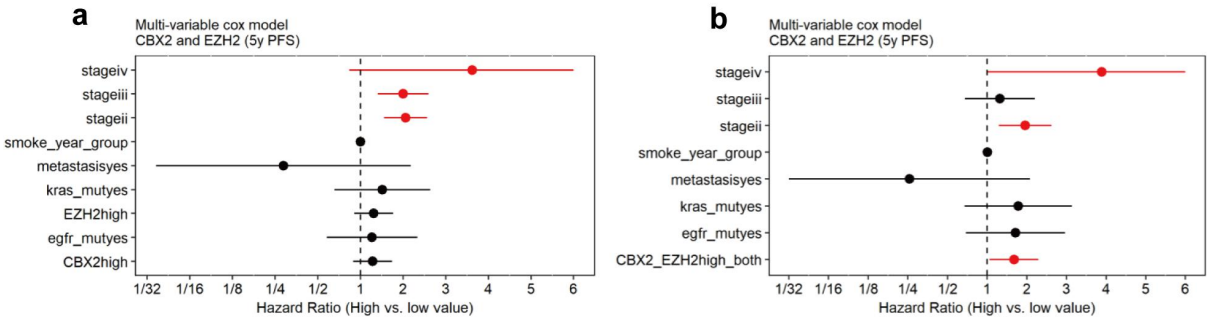


**Figure S1. Related to Figure 1. Multivariate survival analysis for CBX2 and EZH2; covariates are TNM stage, smoke years (continuous), metastasis, and mutation of KRAS and EGFR. (a)** Hazard ratio got from multi-variable cox model, comparing the risk of death for CBX2 or EZH2 expression, respectively. **(b)** Hazard ratio got from multi-variable cox model, comparing the risk of death between groups ($CBX2^{high}$-$EZH2^{high}$ and $CBX2^{low}$-$EZH2^{low}$).

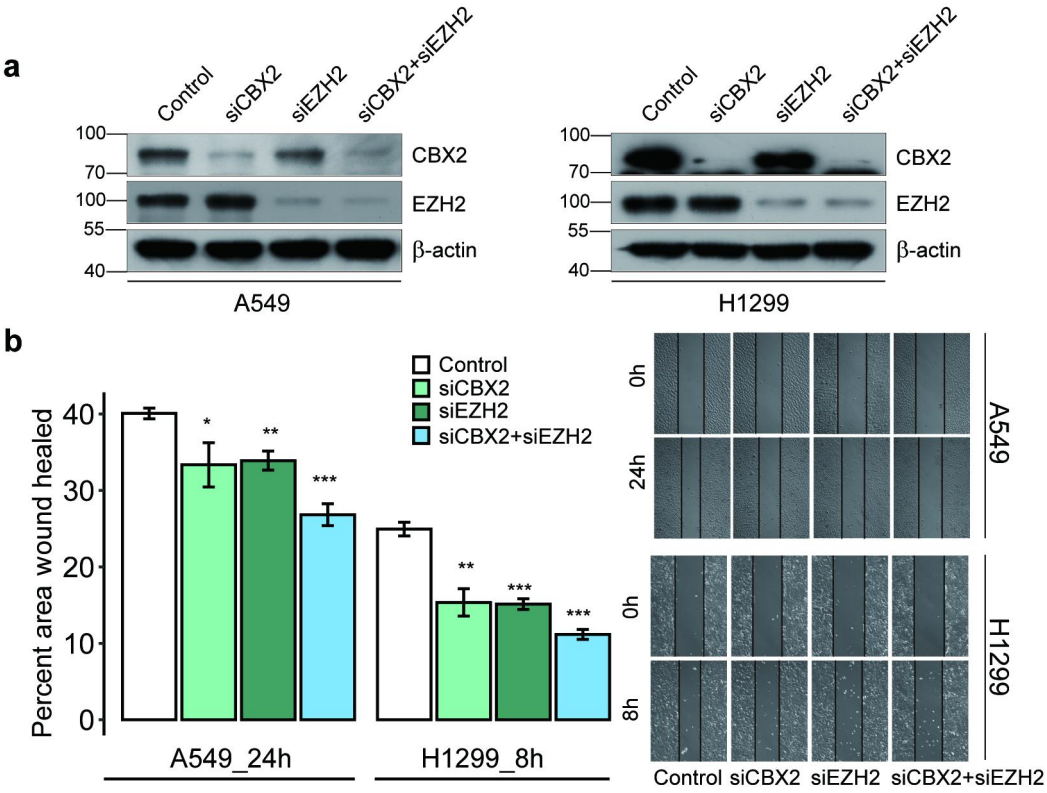


**Figure S2. Related to Figure 2. Wound-healing assays demonstrating the inhibition of cell migration by knockdown of CBX2, EZH2 and both. (a)** Western blot validating the inhibition effect of the protein expression by CBX2, EZH2 and double gene siRNAs in A549 and H1299 cells. The molecular weights (bp) are indicated on the left side of the plot. (b) A549 or H1299 cells transfected with control siRNAs, CBX2 siRNAs, EZH2 siRNAs or both CBX2 and EZH2 siRNAs were subjected to wound-healing assays, and representative images are shown. Quantitative measurement of cell migration was performed at 24 hours for A549 cells and 8 hours for H1299 cells. Each bar represents the mean ± S.D. of three independent biological replicates. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$ (Student's t test).

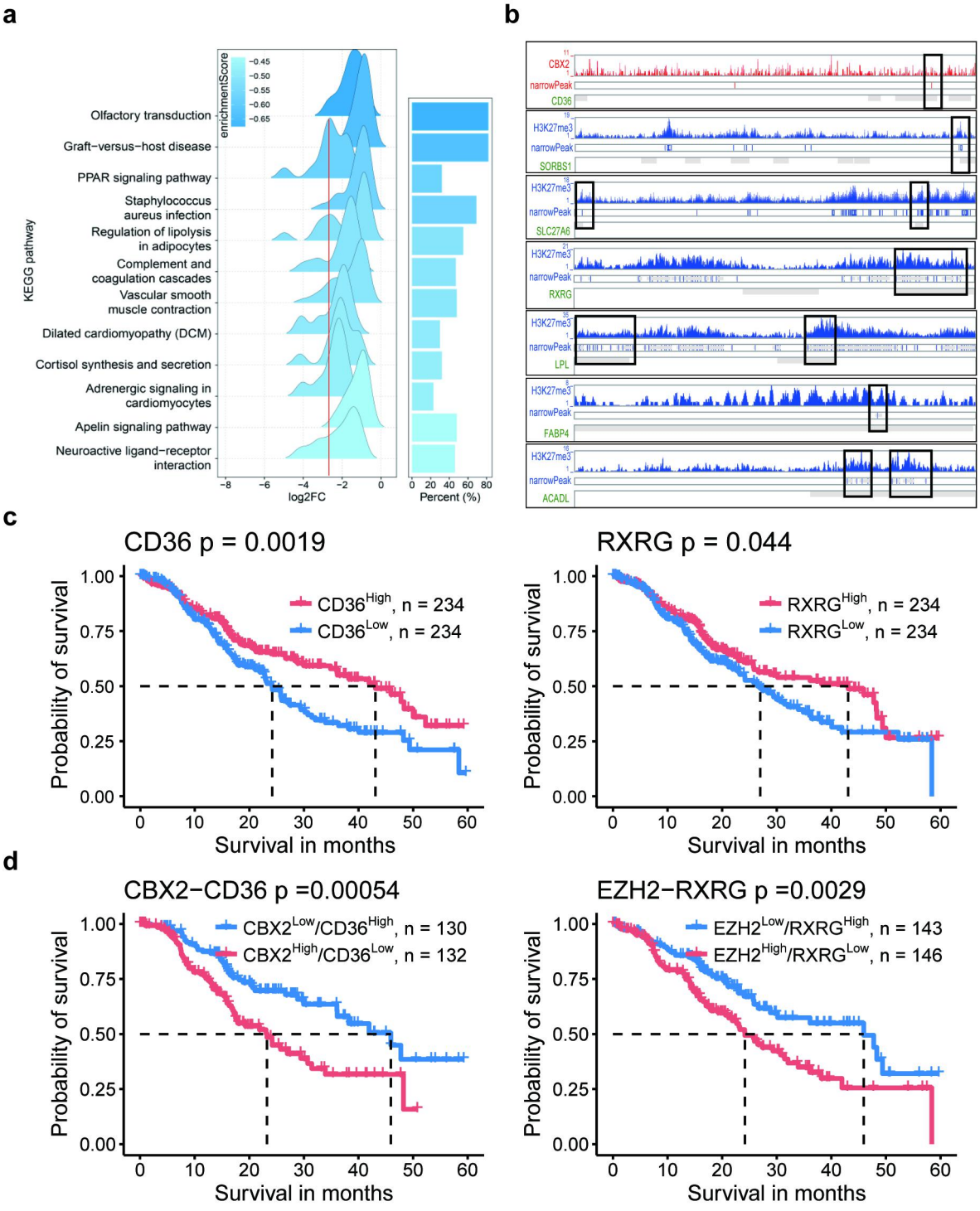


**Figure S3. Related to Figure 5 and 7. ChIP signal on PPAR signalling pathway genes targeted by CBX2 and EZH2, and survival analysis of PPAR signaling genes. (a)** KEGG pathway enrichment analysis for downregulated genes in tumor samples. Enrichment score: reflects the degree to which a gene set (here represents downregulated gene set) is overrepresented at the top or bottom of a ranked list of genes (here represents all genes ranked

by fold change). Percent (%): 100%×(the number of genes enriched in a pathway)/(number of genes in a pathway). **(b)** UCSC genome browser showing the binding peaks of CBX2 and EZH2 on PPAR signaling pathway genes, including CD36, FABP4, ACADL, SLC27A6, RXRG, LPL and SORBS1. Red and blue represent peaks of CBX2 and EZH2 respectively. The grey bar indicates the TSS regions (TSS-8kb ~ TSS+2kb) of different transcripts of one gene (Gencode V21). NarrowPeak was obtained by MACS2. We used black frame to highlight the peaks locate on the TSS region. **(c)** Kaplan-Meier survival curve for patients with high (red) and low (blue) expression of PPAR signalling genes, including CD36 and RXRG. (**d**) Kaplan-Meier survival curve for the combined expression of CBX2/EZH2 and PPAR genes.

Supplementary Table

Table S1. List of histone methylation-related genes

| | | | | |
|---|---|---|---|---|
| AEBP2 | DEFA1 | KDM4A | NCOA6 | RBBP7 |
| ASH1L | DEFA1B | KDM4B | NSD1 | RCBTB1 |
| ASH2L | DOT1L | KDM4C | NSD2 | SCML2 |
| ATF7IP | DPY30 | KDM4D | UHRF1 | SETD1A |
| BPTF | EED | KDM5C | WDR5 | SETD1B |
| BRD1 | EHMT1 | KDM6A | WDR82 | SETD2 |
| BRPF1 | EHMT2 | KDM7A | WHSC1 | SETD7 |
| CARM1 | EZH1 | KMT2A | NSD3 | SETD8 |
| CBX1 | EZH2 | KMT2B | PAGR1 | SETDB1 |
| CBX2 | FXR1 | KMT2C | PAXIP1 | SETDB2 |
| CBX3 | FXR2 | KMT2D | PDP1 | SETMAR |
| CBX4 | GAS7 | KMT2E | PHF1 | SFMBT1 |
| CBX5 | TDRD3 | KMT5A | PHF19 | SMYD2 |
| CBX6 | TFF1 | KMT5B | PHF2 | SMYD3 |
| CBX7 | TP53BP1 | KMT5C | PHF8 | SUV39H1 |
| CBX8 | TRRAP | L3MBTL1 | PRDM2 | SUV39H2 |
| CDY1 | GOLGA6A | L3MBTL2 | PRDM9 | SUV420H1 |
| CDY1B | HCFC1 | MBTD1 | PRMT1 | SUV420H2 |
| CDY2A | HP1BP3 | MDC1 | PRMT2 | SUZ12 |
| CDY2B | ING1 | MEN1 | PRMT5 | TAF3 |
| CHD1 | ING2 | MLLT10 | PRMT6 | WHSC1L1 |
| CHD4 | ING3 | MLLT3 | PRMT7 | XCL1 |
| CHP1 | ING4 | MLLT6 | PRMT8 | XCL2 |
| CNOT8 | ING5 | MORF4L1 | RAG2 | ZCWPW1 |

| | | | | |
|---|---|---|---|---|
| CRB2 | KDM3A | MSL3 | RBBP4 | ZGLP1 |
| CXXC1 | KDM3B | MTF2 | RBBP5 | |