Supplementary Information for

Loss of TET reprograms Wnt signaling through impaired demethylation to promote lung cancer development

Qin Xu†, Chao Wang†, Jia-Xin Zhou, Zhi-Mei Xu, Juan Gao, Peng-Fei Sui, Colum P. Walsh, Hong-Bin Ji*, Guo-Liang Xu*

† These authors contributed equally to this work.
*Corresponding authors: Hong-Bin Ji & Guo-Liang Xu
Email: hbji@sibcb.ac.cn (H.-B.J.); glxu@sibcb.ac.cn (G.-L.X.).

This PDF file includes:
SI Materials and Methods
Figures S1 to S12
Tables S1 to S4
SI References
Supplementary Information Text

SI Materials and Methods

Genotyping. Genomic DNA was extracted from the tail tips of mice. The tail tissues were lysed by incubation with 400 μl lysis buffer (10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mM EDTA, 0.1% SDS) containing 1:50 proteinase K (LifeFeng) overnight at 55 °C. DNA was precipitated with an equal volume of isopropanol, followed by centrifugation at maximum speed for 15 min to discard supernatant, and then dissolved in 200 μl TE buffer. Genomic PCR reactions were carried out using 2×Taq PCR Master (LifeFeng, PT102-01). Genotyping primers and protocols have been reported previously (1-3).

In vivo adenoviral or lentiviral delivery. To initiate lung tumors, mice with appropriate genotypes at 6-8 weeks of age were infected with 4×10^6 plaque-forming units (PFUs) of CMV-Cre expressing adenovirus (Ad5-CMV-Cre) or 1×10^6 transforming units (TFUs) of lentiviruses co-expressing Cre and cDNA of interest or 2×10^5 TFUs of pSECC lentiviruses co-expressing Cre, Cas9 and specific sgRNAs per mouse through intranasal inhalation as previously described (4, 5). Then, mice were monitored for signs of lung tumorigenesis and euthanized at the indicated times following adenoviral or lentiviral intranasal inhalation for histopathological assessment. All mice with the same genotype or administration were generally housed within the same cage. Mice of both sexes were randomly selected for inducing tumor initiation.

Mice were monitored weekly for signs of morbidity and were euthanized by CO2 asphyxiation until signs of illness necessitated euthanasia. Date of euthanasia relative to date of adenoviral intranasal inhalation was used for Kaplan-Meier survival analyses.

Lung preparation and whole-mount inspection. Mice were anesthetized with avertin, and then tumor-bearing lungs were perfused through the trachea with cold 4%
paraformaldehyde (PFA). Collected lungs dissected from mice were then subjected to fixation in 4% PFA overnight. After being washed with PBS, lung tissues were placed on plates to capture whole-mount bright-field images by using a Canon EOS 1100D digital camera or transferred to 70% ethanol for histopathological analysis.

**Histological quantification.** Adobe Photoshop CS6 was used to quantify the percentage of total lung area occupied by tumor area (tumor burden) on H&E-stained slides. ImageJ was used to determine the frequency of cells that were positive for specified antigens as a fraction of total tumor cells. Data points represent individual mice for tumor burden graphs and individual tumor regions for Ki-67 by immunohistochemistry. Individual mouse tumor lesions or grades were analyzed by using modified tumor-grading schemes as previously described (2, 4).

**Cloning of human TET2 wild-type or mutant constructs.** Full-length wild-type human TET2 cDNA was obtained from pCAG-TET2 (NCBI NM_001127208.3), and pCAG-TET2 was then engineered by site directed mutagenesis to introduce missense mutations occurring in LUAD patients using overlapping PCR method of two fragments. Finally, full-length fragments of either TET2 wild-type or mutant cDNA were cloned into the pCDH-CMV-3×FLAG-EF1α-copGFP lentiviral vector using seamless assembly cloning kit (CloneSmarter, C5891), according to the manufacturer’s protocol. All constructs were validated by Sanger sequencing, and these validated constructs were transfected into HEK293T cells using Lipofectamine 3000 Transfection Reagent (Invitrogen, L3000015).

**Lentiviral production and infection.** Lentivirus was generated by the co-transfection of lentiviral backbone constructs (5.4 μg) with packaging plasmid psPAX2 (Addgene, #12260, 5.4 μg) and envelope plasmid pMD2.G VSV-G (Addgene, #12259, 1.2 μg) into Lenti-X 293T cells (Takara, 632180) using Lipofectamine 2000 Transfection Reagent (Invitrogen, 11668-019). Lentivirus-containing supernatants were collected and cleared of cellular debris by 0.45-μm filtration 48-72 h after transfection. For in
in vivo infection, lentiviral supernatants were concentrated by ultracentrifugation at 50,000 g for 2 h, and resuspended overnight in an appropriate volume of Dulbecco’s Phosphate Buffered Saline (DPBS) or OptiMEM (Gibco, 31985-062).

Target cells were exposed to lentiviral supernatants supplemented with 8 μg/ml polybrene for 24h periods before being washed, grown for at least 48 h in fresh medium and then subjected to purification of copGFP+ cells by FACS. copGFP+ cells were cultured under fresh medium until they were used for experiments.

To determine lentiviral titer, the engineered 3T3 reporter cell line was generated. First, lentiviruses were diluted in a gradient. Upon infection with lentiviruses, reporter cells became GFP+ 2-3 days after infection, and then the frequency of GFP+ cells was analyzed by flow cytometry to calculate TFUs.

**Generation of tumor-derived KT cell lines.** KT tumors generated from individual KTR mice were dissected, and digested in protease and DNase solution. Dissociated cells were further processed for FACS sorting. Total DAPI−/CD45−/CD31−/tdTomato+ LUAD cells were sorted, and spun at 300 g for 3 min. Cell pellets were resuspended in fresh medium, and then plated in 100-mm plates to allow for attachment. Cell lines were genotyped for Kras, Tet1, Tet2 and Tet3 after at least 5 passages in culture. All KT cell lines were grown in DMEM supplemented with 10% FBS and 1% pen-strep, and tested negative for mycoplasma contamination. The KT cell lines used in this study were established from mouse LUAD over the course of the study.

**Cell proliferation assay.** In brief, 2×10⁴ cells were seeded in duplicate into 6-well plates per well. Cells were counted every single day with trypan blue staining to determine cell viability using a Countess Automated Cell Counter (Invitrogen, C10281).

**Subcutaneous allograft transplantation.** Parental KT cells or KT cells constitutively expressing TET2CD or TET2HD were resuspended in DPBS, and 1×10⁶ cells in a volume of 100 μl were subcutaneously injected into both flanks of randomized 6-
week-old male BALB/c athymic nude mice (six mice per group). Mice were monitored every 3 days for tumor formation until 2 weeks. After inoculation for 15 days, mice were euthanized, and allograft tumors were dissected and examined for tumor weight, then fixed in 4% PFA. Paraffin-embedded allograft tumor tissues were cut into 5-μm thick sections, and H&E staining was performed on these sections. Differences in the weight of allograft tumors were statistically assessed. The maximal tumor volumes were in accordance with the Institutional Animal Care and Use Committee (IACUC) of Shanghai Institute of Biochemistry and Cell Biology (SIBCB), Chinese Academy of Sciences.

**Analysis of human LUAD datasets.** The following human LUAD datasets were used for mutation, mutual exclusivity and co-occurrence and survival analysis: (TSP, Nature 2008); (Broad, Cell 2012); (TCGA, Nature 2014); (TCGA, Firehose Legacy); (TCGA, PanCancer Atlas); (MSKCC, Science 2015); (MSKCC, Cancer Discov 2017). Oncoprint, correlation and survival curves were obtained by analysis of the above human LUAD datasets using cBioportal for cancer genomics (http://www.cbioportal.org) (6, 7).

**Quantitative real-time PCR analysis.** Total RNA was treated with gDNA Eraser and reverse-transcribed into cDNA by PrimeScript RT reagent Kit (Takara, RR047A). Quantitative real-time PCR was performed using Bio-Rad CFX96 with SYBR Premix Ex Taq (Takara, RR820). Fold change was calculated based on $2^{-\Delta\Delta Ct}$ method after normalization to the transcript level of the housekeeping gene Actb.

**Western blot analysis.** Cells were collected and lysed by ProteinExt Mammalian Total Protein Extraction Kit (TransGen, DE101). The supernatants were loaded onto a 10% SDS-PAGE gel and the proteins were blotted onto a nitrocellulose membrane. Membranes were blocked for 1 h with 5% milk in TBST, incubated with anti-FLAG mouse monoclonal antibody (Sigma, F1804, 1:1,000) or anti-α-Tubulin mouse monoclonal antibody (Sigma, T5168, 1:10,000) overnight at 4 °C or anti-HRP-
conjugated GAPDH mouse monoclonal antibody (Proteintech, HRP-60004, 1:10,000) for 1 h at room temperature, and then rinsed 3 times each for 5 min before incubation with peroxidase-conjugated affinipure goat anti-mouse IgG (Jackson, 115-035-003, 1:10,000) or goat anti-rabbit IgG (Jackson, 111-035-003, 1:10,000) for 1 h. The blots were detected with Immobilon Western Kit (Millipore, WBKLS0100) and images were obtained using a myECL imager (Thermo).

**UHPLC–MS/MS analysis.** In brief, purified high-quality DNA was digested by nuclease P1 (Sigma, N8630) in the presence of 0.2 mM ZnSO₄ and 20 mM NaAc (pH 5.3) at 55 °C for at least 4 h and then was dephosphorylated with calf intestinal alkaline phosphatase (CIAP, Takara, 2250A) at 37 °C for at least 4 h. The samples were centrifuged at maximum speed for 5 min and the supernatants were then subjected to ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) analysis for quantitation of 5mC and 5hmC. The UHPLC–MS/MS analyses were performed using a UHPLC system (Agilent Technologies, 1290 series) coupled to a triple quadrupole mass spectrometer (Agilent Technologies, Agilent 6495 QQQ).
Fig. S1. Mutations of KRAS and TET family genes co-occur and associate with poor survival in human LUAD. (A) Barplot showing frequency of genetic alterations in TET1, TET2 and TET3 among all patients in human LUAD datasets. Mut, mutation; Amp, amplification; Del, deep deletion. (B) OncoPrint from cBioPortal showing mutation profiles of KRAS, TP53 and TET family genes in human...
LUAD datasets \((n = 2678 \text{ samples})\). The corresponding mutation frequencies, types and study of origin are shown. Connecting lines denote that the mutational status of \(TET\) genes is not profiled in these samples. (C) Lollipop plots depicting distribution of \(TET1, TET2\) and \(TET3\) mutations (missense, truncating and in-frame) found in LUAD patients along the protein sequences. (D) Pie charts representing frequency of \(TET\) family genes related to loss-of-function mutations in human LUAD datasets. (E) Contingency table analysis of mutation frequencies between \(KRAS\) and \(TET\) family genes or \(TP53\) in 7 human LUAD datasets. “+” and “−” indicate mutated, or not altered, respectively. Mutual exclusivity and co-occurrence analysis of four pairs between \(KRAS\) and \(TET1, TET2, TET3\) and all three \(TET\) genes was tested. \(P\)-value was derived from one-sided Fisher’s Exact Test. (F) Kaplan–Meier survival analysis of LUAD patients whose tumors contain mutations in \(KRAS, TET\) family genes or both \(KRAS\) and \(TET\) family genes. The number of samples is indicated. Survival data of patients were obtained from the clinical sequencing cohorts shown in (B). Significance was determined by Log-rank (Mantel-Cox) test. Source data can be found from cBioPortal (www.cbioportal.org). ns, not significant; *, \(P \leq 0.05\); **, \(P \leq 0.01\); ***, \(P \leq 0.001\); ****, \(P \leq 0.0001\).
Fig. S2. The mutually exclusive pattern of \textit{EGFR} and \textit{TET} mutations in human LUAD. (A) OncoPrint from cBioPortal showing mutation profiles of \textit{EGFR} and \textit{TET} family genes in human LUAD datasets (\(n = 2678\) samples). The corresponding mutation frequencies, types and study of origin are shown. Connecting lines denote that the mutational status of \textit{TET} genes is not profiled in these samples. (B) Contingency table analysis of mutation frequencies between \textit{EGFR} and \textit{TET} family genes in human LUAD datasets. “+” and “−” indicate mutated, or not altered, respectively. Mutual exclusivity and co-occurrence analysis of four pairs between \textit{KRAS} and \textit{TET1}, \textit{TET2}, \textit{TET3} and all three \textit{TET} genes was tested. \(P\)-values were derived from one-sided Fisher’s Exact Test. (C) Kaplan–Meier survival analysis of LUAD patients whose tumors contain mutations in \textit{EGFR}, \textit{TET} family genes or both \textit{EGFR} and \textit{TET} family genes. The number of samples is indicated. Survival data of
patients were obtained from the clinical sequencing cohorts shown in (A).
Significance was determined by Log-rank (Mantel-Cox) test. ns, not significant; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. 
Fig. S3. Impacts of human LUAD-derived TET mutations on gene expression and catalytic activity. (A and B) Downregulation of TET expression in TET-mutant LUAD versus TET wild-type LUAD in the setting of both wild-type and mutant KRAS (A) or mutant KRAS alone (B). Source data were derived from TCGA PanCancer Altas. (C) Global DNA methylation levels in TET-mutant LUAD versus TET wild-type LUAD. Source data were derived from TCGA Firehose Legacy. (D) Western blot analysis of FLAG-tagged human TET2 variants (mutations outside the catalytic domain: R1214L, G1235A, G1282V and F1287S; mutations inside the catalytic domain: F1300L, P1342S and H1382Y/D1384A) in transfected HEK293T cells. (E) Enzymatic activity assay of the indicated TET2 variants ectopically expressed in HEK293T cells. Relative 5hmC levels in genomic DNA were quantified by mass spectrometry from two independent experiments. (F) Downregulation of TET expression in TCGA LUAD compared with corresponding healthy tissues. Source data of tumors and normal tissues were derived from TCGA LUAD datasets and
Genotype-Tissue Expression Project (GTEx), respectively. (G) Downregulation of TET expression in TET wild-type tumors during LUAD progression based on the histological grades. Source data were derived from OncoSG (Nat Genet, 2020) (H and I) Kaplan–Meier plots showing survival probability of patients in TCGA PanCancer Atlas partitioned by the expression level of TET (high, ≥ 0.5 standard deviations (s.d.); low, ≤ −0.5 s.d. below the mean). The number of samples is indicated for each group (A-C, F-I), and P values were calculated using two-tailed unpaired Student’s t-test (A-C, F and G) and Log-rank Mantel–Cox test (H and I). *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001.
Fig. S4. Generation of GEMMs to simultaneously induce the activation of oncogenic Kras and the deletion of Tet family genes. (A) Schematic of Kras<sup>G12D</sup> point mutation activation and targeted disruption of Tet family genes. Conditional allele of Kras<sup>LSL-G12D</sup> modeled an oncogenic mutation by a glycine to aspartic acid transition at codon 12 in the endogenous locus. A LoxP-STOP-LoxP (LSL) cassette was engineered into the first intron of Kras gene. The LSL cassette consists of transcriptional and translational stop elements flanked by LoxP sites, and prevents the expression of the mutant allele until the stop elements are removed by Cre recombinase. For conditional alleles of Tet1, Tet2 and Tet3, coding exons are shown as black rectangles. Frt sites flanking the neo selection marker are shown as purple triangles and LoxP sites flanking the targeted region are shown as red triangles. Stars indicate the exon encoding the His-X-Asp motif of the iron binding site of TET enzymes. (B) Gel images show genotyping PCR results of the corresponding GEMMs with conditional alleles of Kras<sup>LSL-G12D</sup> and Tet family genes. (C) Kras<sup>LSL-G12D/+</sup> mice, Tett<sup>fl<sup>ox</sup>/fl</sup> mice, Tett<sup>2</sup><sup>fl<sup>ox</sup>/fl</sup> mice and Tett<sup>3</sup><sup>fl<sup>ox</sup>/fl</sup> mice were crossed to generate the
cohorts as below: K (\textit{Kras}^{\text{LSL-G12D/+}}), KT Het (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet1}^{\text{flox/+}}; \textit{Tet2}^{\text{flox/+}}; \textit{Tet3}^{\text{flox/+}}), KT (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet1}^{\text{flox/flox}}, \textit{Tet2}^{\text{flox/flox}}, \textit{Tet3}^{\text{flox/flox}}), KT1 (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet1}^{\text{flox/flox}}), KT2 (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet2}^{\text{flox/flox}}), KT3 (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet3}^{\text{flox/flox}}), KT23 (\textit{Kras}^{\text{LSL-G12D/+}}; \textit{Tet2}^{\text{flox/flox}}, \textit{Tet3}^{\text{flox/flox}}) and T (\textit{Tet1}^{\text{flox/flox}}, \textit{Tet2}^{\text{flox/flox}}, \textit{Tet3}^{\text{flox/flox}}).

Genetic lesions were induced by intranasal inhalation of adenovirus or lentivirus expressing Cre recombinase at 6-8 weeks of age, after which mice were monitored for signs of disease and euthanized from 4 to 12 weeks after viral infection for histopathological analysis.
Fig. S5. Triple knockout of *Tet* genes triggers a spectrum of rapidly-proliferating and poorly-differentiated malignant lesions. (A) Histopathological images of...
representative H&E-stained lung tumors within different tumor grades from KT mice at 12 weeks after Ad-CMV-Cre infection. AAH, atypical adenomatous hyperplasia; ADC, adenocarcinoma. Scale bars, top panel, G1 100 μm, G2 200 μm, G3 500 μm; bottom panel, 50 μm. (B and C) Immunohistochemistry images for Ki-67 staining (B) and quantification of Ki-67 positive tumor cells (C) in K and KT tumors at 8 or 12 weeks after Ad-CMV-Cre infection. Each circle represents an individual tumor region. 

\( n = 30 \) tumor regions for each genotype at various time points. (D and E) Representative images of H&E-stained tumor-bearing lung sections (D) and corresponding quantification of tumor burden (E) in K and KT mice at 8 or 12 weeks after Ad-CMV-Cre infection. Each dot represents an individual mouse. \( n = 8 \) mice for each group. Data are mean ± s.e.m. Statistical significance was assessed using two-tailed unpaired Student’s \( t \)-test with Welch’s correction (C, E). *, \( P \leq 0.05 \); **, \( P \leq 0.01 \); ***, \( P \leq 0.001 \); ****, \( P \leq 0.0001 \). Scale bars, left (8 weeks), 100 μm; right (12 weeks), 200 μm, and insets are 20 μm (B); 2 mm (D). (F) Representative whole-mount views of lung tissues and H&E-stained lung sections from T mice at 12 (left) or 54 (right) weeks after Ad-CMV-Cre infection. Scale bars, whole-mount lung tissues and sections, 2 mm; magnified H&E sections, 100 μm. (G) Immunofluorescence images for SPC (AT2 marker), PDPN (AT1 marker), SCGB1A1 (club cell marker) and FOXJ1 (ciliated cell marker) staining on lung sections from T mice at 12 weeks after Ad-CMV-Cre infection. Scale bars, 50 μm; magnified images, 10 μm.
Fig. S6. Single or double knockout of Tet genes induce tumors with varying degrees of differentiation. (A) Quantitative real-time PCR (qRT-PCR) analysis of Tet1, Tet2 and Tet3 transcripts in mouse adult lung tissues (n = 3 mice). (B) qRT-PCR analysis of Tet1, Tet2 and Tet3 transcripts in K tumors and adjacent tumor-free tissues collected at 22 weeks after Ad-CMV-Cre infection. The expression levels are normalized to Actb and presented as mean ± s.e.m. (A, B). Statistical significance was determined using two-tailed unpaired Student’s t-test with Welch’s correction (B). ns, not significant; ***, P ≤ 0.001. (C-F) Immunohistochemistry images for NKX2-1, SPC and SCGB1A1 staining in KT1 (C), KT2 (D), KT3 (E) and KT23 (F) tumors at 12 weeks after Ad-CMV-Cre infection. H&E staining is shown on top. Scale bars, 200 μm, and insets are magnified 10× (C, D, E, F).
Fig. S7. Tet2 attenuates proliferation of established KT LUAD cells dependent on its catalytic activity. (A) Representative FACS plots of tdTomato+ primary tumor cells.
cells sorted from dissociated tumor-bearing lung tissues of three KT mice with 
\textit{Rosa26LSL-AdTomato/+} reporter allele at 12-16 weeks after Ad-CMV-Cre infection. 
DAPI−/CD45−/CD31−/tdTomato+ cells (total cancer cell fraction) are gated in red rectangles. (B) Representative images of cell morphology in three independent cell lines sub-cultured from the sorted tdTomato+ tumor cells at least 5 passages. (C) Gel images show genotyping PCR results of conditional inactivation of \textit{Tet1}, \textit{Tet2} and \textit{Tet3} in three independent cell lines. (D) Gross (top) and H&E-stained (bottom) images of KT LUAD cells grown for 2 weeks as subcutaneous allografts. (E) Schematic of lentiviral constructs of the truncated or full-length mouse Tet2 (Tet2CD or Tet2FL) and the catalytically-inactive counterparts (Tet2HD or Tet2FL-HD) used for constitutive expression in KT cells. copGFP is used for purification of infected cells by FACS. CMV and EF1\(\alpha\) promoters drive transgene or fluorescence marker expression, respectively. A FLAG epitope added to the N terminus of Tet2CD, Tet2FL, Tet2HD or Tet2FL-HD allows detection of the fusion protein. (F and G) qRT-PCR (F) and western blot (G) analysis of truncated \textit{Tet2} mRNA or truncated FLAG-Tet2 fusion protein expression in KT cells complemented with Tet2CD or Tet2HD. Parental KT cells (Vec) served as negative control. (H and I) Quantification of 5hmC (H) and 5mC (I) nucleosides in genomic DNA extracted from KT cells complemented with Tet2CD or Tet2HD and parental cells (Vec) determined by quantitative mass spectrometry from two independent experiments. Downward arrows indicate not detected. (J) Growth curves of KT LUAD cells complemented with Tet2CD or Tet2HD. Cell number was determined at every passage by counting the trypan blue negative cells at indicated time points from two biological replicates. (K) KT LUAD cells complemented with Tet2CD or Tet2HD were subcutaneously injected into randomized athymic nude mice (six mice per group) and grown for 15 days. The weight of allograft tumors was statistically analyzed. (L) Western blot analysis of FLAG-tagged full-length mouse Tet2 protein in KT cells complemented with Tet2FL or Tet2FL-HD. Parental KT cells (Vec) served as negative control. (M and N) Quantification of 5hmC (M) and 5mC (N) nucleosides in genomic DNA extracted from KT cells complemented with Tet2FL or Tet2FL-HD and parental cells (Vec).
nucleoside contents were determined by quantitative mass spectrometry. Downward arrows indicate “not detected”. (O) Growth curves of KT cells complemented with Tet2FL or Tet2FL-HD. Cell number was determined by counting trypan blue negative cells at indicated time points. Data are presented as mean ± s.e.m. Statistical significance was calculated using two-tailed unpaired Student’s t-test (J, K, O). *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. Scale bars, 200 μm (B); 2 mm (top), 5 μm (bottom) (D); 5 mm (K). #1, #2 and #3 represent three biologically independent KT LUAD cell lines.
Fig. S8. Functional validation of lentiviruses of Cre, Cre-Tet2CD or Cre-Tet2HD in engineered 3T3 cells. (A) Schematic of engineered 3T3 fibroblast cells containing a LoxP-RFP-STOP-LoxP-GFP cassette. The 3T3 cells expressed GFP 2-3 days after infection with a lentivirus expressing Cre recombinase. (B) Representative images of GFP+ cells infected with lentiviruses of Cre, Cre-Tet2CD or Cre-Tet2HD, respectively. Total cells were shown under bright field (BF). Scale bars, 100 μm. (C and D) Confirmation of truncated Tet2 mRNA or truncated FLAG-TET2 fusion protein overexpressed in sorted GFP+ cells infected with indicated lentiviruses using qRT-PCR (C) and western blot (D). (E and F) 5hmC (E) and 5mC (F) dot blot analysis of genomic DNA extracted from sorted GFP+ cells infected with indicated lentiviruses. Total genomic DNA was measured by methylene blue staining.
**Fig. S9. Isolation of pre-malignant cells in early neoplasia used for low-input RNA-seq and WGBS analysis.** (A) An example of the FACS gating strategy used to sort live tdTomato+ lung epithelial cells. Dissociated lung cells were first gated as DAPI− and tdTomato+ cells. Epithelial cells were then identified as CD45−/CD31−/EPCAM+. (B) FACS analysis of pre-malignant cells defined as tdTomato+ lung epithelial cells sorted from KR and KTR mice at 3 weeks after Ad-CMV-Cre infection. Representative plots from three independent experiments. (C) FACS quantification for the total proportion of pre-malignant cells in KR and KTR mouse lung tissues. KR, n = 7 mice; KTR, n = 8 mice. Data are shown as mean ± s.e.m. Statistical significance was calculated using two-tailed unpaired Student’s t-test. ns,
not significant. (D) Log expression of lung epithelial cell marker genes from low-input RNA-seq data. (E) Global CpG methylation levels of K and KT pre-malignant cells across the different genomic elements. DNA methylation level was calculated by extracting CpG sites with at least 5× coverage. (F) Distribution of methylation levels of individual CpGs across the designated genomic elements.
Fig. S10. Transcriptional profiles of established KT cells complemented with Tet2CD or Tet2HD by bulk RNA-seq. (A) Principal component analysis of RNA-seq data from parental KT cells (black) and KT cells complemented with Tet2CD (red) or Tet2HD (grey). (B and C) Scatter plot showing the DEGs in KT cells complemented with Tet2CD versus parental cells (B) or cells complemented with Tet2HD (C). nCount stands for the normalized counts (mean of the counts divided by size factors). The dashed lines indicate the four-fold change threshold for defining DEGs. Red and blue dots denote significantly changed genes (log2(fold change) ≥ 2 or ≤ −2 and adjusted P value ≤ 0.05) and grey dots denote genes without significant changes. (D) Venn diagram showing the overlap of DEGs from (B) and (C). The number of upregulated or downregulated genes is indicated. (E and F) GO analysis of DEGs (log2(fold change) ≥ 2 or ≤ −2 and adjusted P value ≤ 0.01) either upregulated
(E) or downregulated (F) in KT cells complemented with Tet2CD. Related categories of Wnt signaling pathway are denoted in red (E), and related categories of cell proliferation and cell adhesion are denoted in blue (F). #1 and #2 represent two biologically independent KT LUAD cell lines.
Fig. S11. Methylation profiles of established KT cells complemented with Tet2CD or Tet2HD by TAPS. (A) Overall CpG methylation levels of parental KT cells and KT cells complemented with Tet2CD or Tet2HD across the various genomic elements. DNA methylation level was calculated by extracting CpG sites with at least 10× coverage. (B) Global methylome comparison between Tet2CD-expressing and Tet2HD-expressing KT cells. DMRs were filtered by length (≤1000 bp) and CpG number (at least 5 CpG sites). Those with at least 20% absolute methylation level difference were defined as hypermethylated or hypomethylated DMRs. Each dot represents the methylation level of each single DMR. Red or blue triangle contains hypermethylated or hypomethylated DMRs. (C) Distribution of hypomethylated DMRs in Tet2CD-expressing KT cells among various genomic elements. (D) GO analysis of all upregulated DEGs harboring hypomethylated DMRs in promoter regions in Tet2CD-expressing KT cells. The number of genes in each category is indicated. Wnt-related categories are denoted in red. (E) Representative methylation
tracks of the Wnt antagonizing genes Dact1 (top) and Tmem88 (bottom) in KT cells complemented with Tet2CD or Tet2HD. Vertical bars of methylation tracks indicate the methylation level at individual CpG dyads. The grey-shaded box indicates the hypomethylated DMRs in the promoter regions of Dact1 and Tmem88 in Tet2CD-expressing KT cells. (F and G) Differential methylation (F) or expression (G) levels of WNT antagonizing genes DACT1 and TMEM88 in TET-mutant LUAD compared with TET wild-type LUAD. Source data of methylation values (β-values) were derived from TCGA Firehose Legacy (F), and source data of gene expression were derived from TCGA PanCancer Atlas (G). In the box plots, the bottom and the top rectangles indicate the first quartile and third quartile. The horizontal lines in the middle indicate the median, and the vertical lines that extend from the top and the bottom of the plot indicate the maximum and minimum values. The number of samples is indicated for each group, and P values were calculated using a two-tailed unpaired Student’s t-test. *, P ≤ 0.05.
Fig. S12. *In vitro* validation of efficiency of sgRNAs targeting the β-catenin gene *Ctnnb1* in engineered 3T3 cells using pSECC lentiviruses. (A) CRISPR-Cas9-mediated *Ctnnb1* gene disruption. sgRNA targeting sites were chosen in *Ctnnb1* exon 3, and four serine/threonine residues are included (codons are underlined), which
could be phosphorylated, leading to β-catenin degradation. Three individual sgRNA-targeting sequences are underlined in blue, and the corresponding protospacer adjacent motif (PAM) sequences are marked in orange. The region for PCR genotyping is indicated as horizontal arrows. (B) Representative images of GFP+ 3T3 cells infected with pSECC lentiviruses of sgGFP, sgCtnnb1.1, sgCtnnb1.2 or sgCtnnb1.3, respectively. Total cells were shown under bright field. Scale bars, 100 μm. (C-E) Representative examples of indel (insertion/deletion) mutations detected in Ctnnb1 endogenous locus from the genomic DNA of GFP+ 3T3 cells infected with the lentiviruses of sgCtnnb1.1 (C), sgCtnnb1.2 (D) or sgCtnnb1.3 (E). Percentages of indels arising from non-homologous end joining (NHEJ) mediated by sgCtnnb1.1, sgCtnnb1.2 or sgCtnnb1.3 were summarized from total tested samples. sgRNA targeting sequences are labelled in blue. PAM sequences are marked in yellow. Cas9 cleavage sites are denoted by downward arrows in wild-type gene sequence.
## Table S1. PCR primers for mouse genotyping

| Mouse gene | Primer name | Primer sequence (5’-3’) |
|------------|-------------|-------------------------|
| **Tet1**   | C           | CAGTAGTATTTTGCTGCTGCAT  |
|            | F           | CATCCTAAATAACCAACCACCA  |
|            | R           | TTCCCTAAGGAGTTACTGCAACG |
| **Tet2**   | C           | ACACAGAGAAAAGGTACGTGAA  |
|            | F           | ACTCATTAGTGAATATGGAATG  |
|            | R           | CTGCTTAGTACATGCAACC    |
| **Tet3**   | C           | GCTCTTTTCCTGAGCAGCAAATGT|
|            | F           | TAGGTGTGGAACAAACATGGAG  |
|            | R           | ACAGCTTTACAGGACACCAGAGAT|
| **KrasLSL-G12D** | Forward | CTAAGCCACCATGGCTTGAGT |
|             | Reverse     | TCCGAATTCAGTGACACTACAGAT|
| **Rosa26LSL-tdTomato/+** | LSL-tdTomato-F | CTGTGCCTGTACGGCATGG |
|             | LSL-tdTomato-R | GCCATTAAGCAGCAGTATCC |
|             | Rosa26WT-F  | AAGGGAGCTGAGTGGAGTA |
|             | Rosa26WT-R  | CGAATACTCGTGGGAAGTC |
| Mouse gene | Primer name | Primer sequence (5’-3’) |
|------------|-------------|------------------------|
| Tet1       | Forward     | CATTCTCAACAGGACATTCCACAACA |
|            | Reverse     | AGTAAAAAGTACGGCTCTTCTCTTG |
| Tet2       | Forward     | GCTCAATATACAGAAGCCTTGCAC |
|            | Reverse     | TATGAGGGTACCCACTGTACT    |
| Tet3       | Forward     | CTATGCAGGGAGGTCAGAAATGAG |
|            | Reverse     | ACAGTGCACCCATGTAGAGGTAT  |
| Lrp4       | Forward     | GCTATGGGAGAGGCTAGAGAAA   |
|            | Reverse     | CGACCAGCGTTCAGATGG       |
| Ctnnbip1   | Forward     | GCCACAGCAGCTCAGCCAC      |
|            | Reverse     | GTCTCGATCTGGAAAACGC      |
| Dact1      | Forward     | GGGTGGCCCAATCTGGCAAG     |
|            | Reverse     | GTGCGTTCAGCTGGATCC       |
| Tmem88     | Forward     | ATCCCTAGATGCTGGG         |
|            | Reverse     | AGCGCAAAACTGGAGGTG       |
| Actb       | Forward     | GGCTGTATTTCCCTCCATCG     |
|            | Reverse     | CCAGTAAACATGGCATGT       |
Table S3. PCR primers for Bisulfite or APOBEC-coupled epigenetic (ACE) Sanger sequencing

| Promoter regions of mouse gene | Primer name | Primer sequence (5’-3’) |
|-------------------------------|-------------|-------------------------|
| Lrp4 DMR-1                    | BS-Forward  | TGGGAATATAGGAGGTAAGATG  |
|                               | BS-Reverse  | AACTCTACCAAAAACCTAAACC  |
| Lrp4 DMR-2                    | BS-Forward  | AGAATAGGGATTTTTGATAGTGAGA |
|                               | BS-Reverse  | ACTCCTACTCAATCTACCCTAACATTCC |
| Ctnnbip1 DMR                  | BS-Forward  | TTAGATGGTTGTAAGGTTGTTTG |
|                               | BS-Reverse  | ACCCAACAATTATCTAAACACAACCTTCC |
Table S4. sgRNA targeting sequences

| Target gene | Primer name | 20 nt sgRNA sequence (5'-3') | PAM | Strand |
|-------------|-------------|------------------------------|------|--------|
| Ctnnb1 (β-Catenin) | sgRNA-1-F | CTGTGGTGGTGCGACCAGAA | TGG | - |
|             | sgRNA-1-R | TTCTGGTGGCAACCAGAGGAGCAG |     | |
| Ctnnb1 (β-Catenin) | sgRNA-2-F | AGCTCCCTCCCTGAGTGCA | AGG | + |
|             | sgRNA-2-R | TGCCAATCTCAGGAAGAGAGCT |     | |
| Ctnnb1 (β-Catenin) | sgRNA-3-F | AGCTTTGCTCTTGAGTGA | AGG | - |
|             | sgRNA-3-R | TCACGCAAGAGAAAGTAGCT |     | |
| eGFP        | sgRNA-F   | GGGCGAGGAGCTGTACCAGC | GGG | + |
|             | sgRNA-R   | CGGTAACAGCTCCTCGCCC |     | |
| Ctnnb1      | Forward   | GCCAGACTGGCCTGTTTCTC |     | For targeting region analysis |
|             | Reverse   | TGTGGCTAGCTGTCACACA |     | |

33
SI References

1. H. Q. Dai et al., TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature* **538**, 528-532 (2016).
2. E. L. Jackson et al., Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243-3248 (2001).
3. L. Madisen et al., A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133-140 (2010).
4. M. DuPage, A. L. Dooley, T. Jacks, Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* **4**, 1064-1072 (2009).
5. F. J. Sanchez-Rivera et al., Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428-431 (2014).
6. E. Cerami et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).
7. J. Gao et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).