Deficiency of the Polycomb Protein RYBP and TET Methylcytosine Oxidases Promotes Extensive CpG Island Hypermethylation and Malignant Transformation

Wei Cui[1], Zhijun Huang[1], Seung-Gi Jin[1], Jennifer Johnson[1], Kin H. Lau[2], Galen Hostetter[3], and Gerd P. Pfeifer[1]

ABSTRACT

Hypermethylation of CpG islands (CGI) is a common feature of cancer cells and predominantly affects Polycomb-associated genomic regions. Elucidating the underlying mechanisms leading to DNA hypermethylation in human cancer could help identify chemoprevention strategies. Here, we evaluated the role of Polycomb complexes and 5-methylcytosine (5mC) oxidases in protecting CGIs from DNA methylation and observed that four genes coding for components of Polycomb repressive complex 1 (PRC1) are downregulated in tumors. Inactivation of RYBP, a key activator of variant PRC1 complexes, in combination with all three 5mC oxidases (TET proteins) in non-tumorigenic bronchial epithelial cells led to widespread hypermethylation of Polycomb-marked CGIs affecting almost 4,000 target genes, which closely resembled the DNA hypermethylation landscape observed in human squamous cell lung tumors. The RYBP- and TET-deficient cells showed methylation-associated aberrant regulation of cancer-relevant pathways, including defects in the Hippo tumor suppressor network. Notably, the quadruple knockout cells acquired a transformed phenotype, including anchorage-independent growth and formation of squamous cell carcinomas in mice. This work provides a mechanism promoting hypermethylation of CGIs and shows that such hypermethylation can lead to cell transformation. The breakdown of a two-pronged protection mechanism can be a route towards genome-wide hypermethylation of CGIs in tumors.

Significance: Dysfunction of the Polycomb component RYBP in combination with loss of 5-methylcytosine oxidases promotes widespread hypermethylation of CpG islands in bronchial cells and induces tumorigenesis, resembling changes seen in human lung tumors.
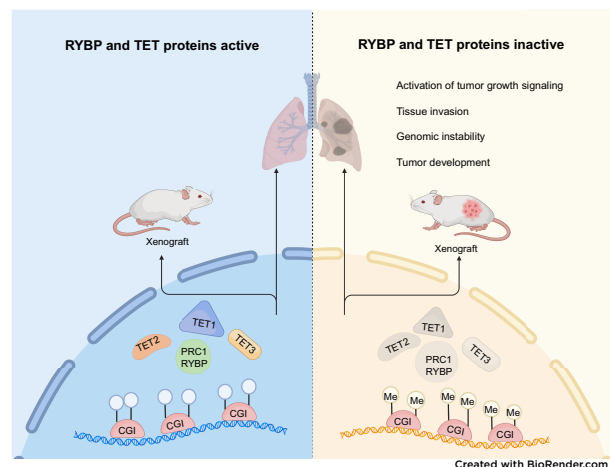
Introduction

Epigenetic reprogramming is a common feature of cancer cells (1). Changes of the epigenome in tumors include altered DNA methylation patterns and aberrant nucleosome modifications. Cancer cells have long been known to harbor genome-wide DNA hypomethylation relative to the corresponding normal cells and show focal DNA hypermethylation at many CpG islands (CGI; refs. 2–4). DNA hypermethylation correlates with silencing of several tumor suppressor genes, for example genes involved in cell cycle control, anti-growth signaling pathways, or DNA repair (5). DNA hypermethylation of CGIs is widespread, affecting a thousand or more target genes in individual tumors and is seen across most cancer types. This event does not occur randomly but is strongly targeted to genes encoding developmental transcription factors and signaling molecules that are regulated by the Polycomb complex (6–10). Because DNA methylation changes are reversible, at least in principle, major efforts are made to use DNA methylation inhibitors for cancer therapy (11).

DNA methylation patterns in normal and malignant cells are established by DNA methyltransferases (DNMT), which operate preferentially at 5'CpG sequences and produce 5-methylcytosine (5mC). DNMT1 maintains methylation patterns owing to its preferential activity on hemimethylated CpG sites formed shortly after DNA replication, and DNMT3A and DNMT3B can methylate unmethylated sites *de novo* (12). DNA methylation patterns can be reversed to an unmethylated state by enzymatic activities that covert the methyl group of 5mC to oxidized forms creating 5-hydroxymethyl-, 5-formyl-, and 5-carboxyl-cytosine followed by base excision repair of the latter two modified bases. These reactions are carried out by a small family of 5mC oxidases, the TET proteins (TET1, TET2, and TET3; refs. 13–15). DNMT and TET proteins are found mutated in some human tumors,

[1]Department of Epigenetics, Van Andel Institute, Grand Rapids, Michigan. [2]Bioinformatics and Biostatistics Core, Van Andel Institute, Grand Rapids, Michigan. [3]Pathology and Biorepository Core, Van Andel Institute, Grand Rapids, Michigan.

W. Cui and Z. Huang contributed equally to this article.

Corresponding Author: Gerd P. Pfeifer, Department of Epigenetics, Van Andel Institute, Grand Rapids, MI 49503. Phone: 616-234-5398; E-mail: gerd.pfeifer@vai.org

but this process primarily affects DNMT3A and TET2 in hematological malignancies (16, 17). On a more global scale, the TET-mediated 5mC oxidation process is defective in most human solid tumors, independent of specific genomic mutations in genes encoding TET proteins or genes critical in the biochemical pathways that support 5mC oxidation (18).

Although DNA methylation patterns in human cancer have been extensively characterized, we still have no good understanding of how these methylation changes arise mechanistically. We hypothesized that the specificity of DNA hypermethylation for Polycomb target genes is based on a diminished function of Polycomb and is accompanied by defective 5mC oxidation. In a model system using human bronchial epithelial cells, we have inactivated a critical component of Polycomb repressive complexes (PRC1), RYBP, and the TET proteins and analyzed the consequences of these deficiencies on genome-wide DNA methylation and cellular phenotype.

Materials and Methods

Cell culture

Human bronchial epithelial cells (HBEC-3KT, CRL-4051; RRID: CVCL_X491) and 293T (HEK293T, CRL-3216) were obtained from ATCC in April 2015 (HEK293T) and in November 2017 (HBEC3-KT). Cell lines were authenticated by ATCC by short tandem repeat profiling. HBEC-3KT cells were cultured in keratinocyte serum-free medium (KSFM; Thermo, 17005042) containing 50 μg/mL of bovine pituitary extract and 5 ng/mL of human recombinant epidermal growth factor. 293T cells were maintained in DMEM (Life Technologies, 11965–092) containing 10% FBS (Life Technologies, 26140–079). Cells were incubated at 37°C with 5% CO_2. The media was changed every 3 to 4 days. All cell lines including the generated knockout (KO) cells have been regularly verified as *Mycoplasma*-free using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza, LT07–710), last tested on November 18, 2021. Cell lines were used within five passages from thawing.

Generation of HBEC3 KO cell lines using CRISPR/Cas9

CRISPR gRNAs were designed to target the genomic sequence, and to introduce frameshift mutations into exon 2 of RYBP and YAF2, or the N-terminal region of the catalytic domains for TET1, TET2, and TET3. All CRISPR gRNAs used in this study are listed in Supplementary Table S1. We used the software CCTop, a CRISPR-Cas9 target online predictor (https://cctop.cos.uni-heidelberg.de; ref. 19), to identify potential off-target sites. The guide RNAs showed a few potential off-targets with three or four mismatches and none at two mismatches or less. Four potential off-targets were in exons of other genes. Using DNA sequencing results, we confirmed that these potential off-target sites did not contain any insertion or deletion mutations.

Lentivirus production

The pLentiCRISPR E plasmid (Addgene, 78852, a gift from Phillip Abbosh) was modified with different selectable marker genes, blasticidin, hygromycin or with EGFP, then cotransfected into HEK293T cells with lentiviral packaging plasmids psPAX2 and pMD2.G (Addgene, 12260 and 12259, gifts from Didier Trono). HEK293T cells were cultured in DMEM plus 10% FBS and seeded in 6-well cell plates the day before transfection in such a way that they would be 80% to 90% confluent at the time of transfection. For transfection of one well, 20 μL of FuGene HD Transfection Reagent (Promega, E2311) was diluted into 500 μL of Opti-MEM and then the following DNAs were added: 1 μg gRNA vector, 0.75 μg psPAX2, and 0.25 μg pMD2.G. The solution was briefly vortexed and incubated at room temperature for 20 minutes. The mixture was then gently added to 6-well plates with 1.5 mL DMEM. All media was aspirated after 10 hours and replaced with fresh DMEM containing 10% FBS. Viral particles were harvested 48 hours after this media change and frozen at −80°C.

Identification of mutant clones

Two days after virus infection of HBEC3-KT cells, blasticidin (4 μg/mL, ThermoFisher, A1113903) or hygromycin B (40 μg/mL, ThermoFisher, 10687010) were added into the medium for selection for 1 week. For EGFP selection, the cells were cultured for 2 weeks until cells with visible green color appeared. For cells with no selection marker, the cells were cultured for 2 weeks. HBEC cells were dissociated into single cells and replated at 50 to 100 cells per 10-cm dish. Cells were allowed to grow until colonies from single cells became visible (10–14 days), then the single colonies were manually picked under a microscope and distributed into 24-well plates. The cells were kept under antibiotics. Colonies were expanded, and 30 to 40 plasmid sequences for each colony were analyzed by Sanger sequencing at gRNA-targeted positions. Sequencing primers are shown in Supplementary Table S1. For each genotype of HBECs, we derived three independent cell clones. The genotype of each clone is listed in Supplementary Table S2.

RNA isolation, mRNA sequencing, and data analysis

Total RNA was isolated with PureLink RNA Mini Kit (Invitrogen, 12183020) from HBEC controls and various HBEC targeted clones (three different clones for each knockout). RNA quality was assessed with the Agilent Bioanalyzer 2100, and RIN scores of all samples were above 9.90. We performed RT-PCR as described (20). Briefly, 2 μg of RNA for each sample was converted to cDNA with gene-specific primers using SuperScript III Reverse Transcriptase Kit (Invitrogen, 18088). mRNA level was determined with pfuUltra II Fusion HS DNA polymerase (Agilent, 600670). PCR primers are listed in Supplementary Table S1. For mRNA sequencing (mRNA-seq), the libraries were prepared from total RNA with the KAPA RNA HyperPrep kit with RiboErase (KAPA Biosystems). Library size distributions were validated on the Bioanalyzer (Agilent Technologies), then the libraries were sequenced using an Illumina NextSeq500 system with 75 bp single read runs at the Van Andel Institute Genomics Core.

For mRNA-seq data analysis, 75-bp single-end reads were trimmed with Trim Galore (version 0.4.0; RRID:SCR_011847). Reads were aligned to the human genome hg19 with STAR (version 2.5.1; RRID:SCR_004463), then gene count was performed with STAR. Differential gene expression was determined with the Limma (version 3.38.2) statistical package as descried previously (21). Differential expression *P* values were adjusted for multiple testing correction using the Benjamini–Hochberg method in the stats package. Statistical significance for differentially expressed genes (DEG) was fold change > 2 with $q < 0.05$. Heat maps were generated with pheatmap (version 1.0.12).

Chromatin immunoprecipitation sequencing and data analysis

For RYBP chromatin immunoprecipitation (ChIP; antibody from Millipore, AB3637), cells were fixed with 2.5 mmol/L ethylene glycol-bis-succinimidyl-succinate (Thermo Scientific, 21565) for 45 minutes, then washed twice with PBS and fixed again with 1% formaldehyde in 50 mmol/L HEPES-KOH, pH 7.5, 100 mmol/L NaCl, 1 mmol/L EDTA, pH 8.0, 0.5 mmol/L EGTA, pH 8.0, for 8 minutes. For H2AK119ub1 ChIP (Cell Signaling Technology, antibody 8240) and H3K27me3 ChIP (Cell Signaling Technology, antibody 9733), cells were cross-linked with 1% formaldehyde only, for 8 minutes, then fixation was stopped by adding glycine at 2.5 mol/L concentration for
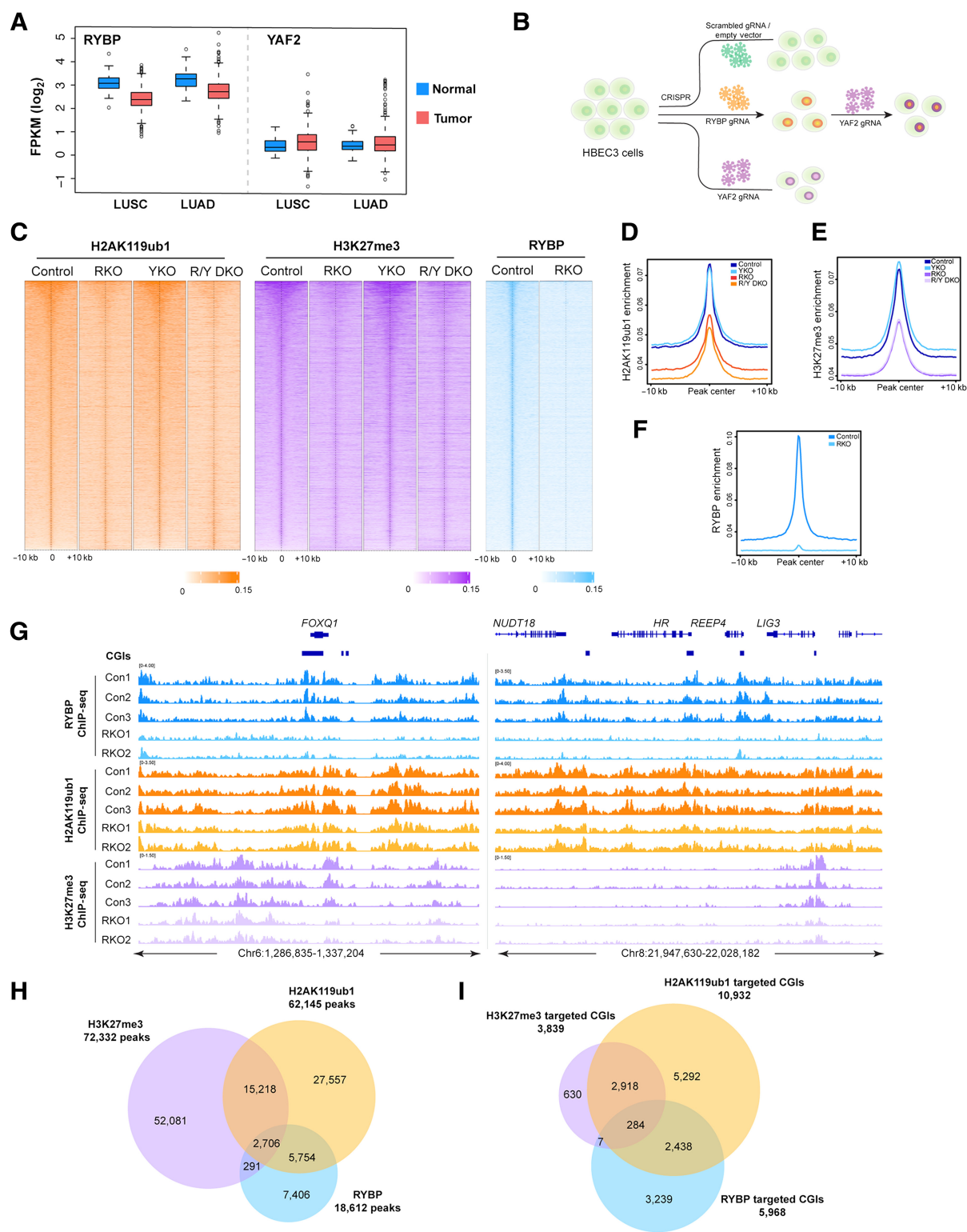
Figure 1.
Polycomb marks in control bronchial cells and in cells lacking RYBP. **A,** mRNA expression of *RYBP* and *YAF2* in human lung adenocarcinomas (LUAD; $n = 537$; normal solid tissue, $n = 59$) and LUSC ($n = 502$; normal solid tissue, $n = 51$) in the TCGA dataset. *RYBP* is downregulated in LUAD ($P < 1.2e\text{-}11$) and LUSC ($P < 2.2e\text{-}16$). **B,** Strategic overview to generate KO clones that were infected with gRNA lentivirus. **C,** Heat map of H2AK119ub1 and H3K27me3 peaks (±10 kb) in control, RKO, YKO, and R/Y DKO and heat map of RYBP peaks (±10 kb) in control and RKO cells. (*Continued on the following page.*)

5 minutes. Fixed cells were suspended in lysis buffer (50 mmol/L HEPES-KOH, pH 7.9, 140 mmol/L NaCl, 1 mmol/L EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100, protease inhibitors) for 10 minutes on ice. Then the cells were pelleted and re-suspended with wash buffer (10 mmol/L Tris-Cl, pH 8.1, 200 mmol/L NaCl, 1 mmol/L EDTA, pH 8.0, 0.5 mmol/L EGTA, pH 8.0, protease inhibitors) and shearing buffer (0.1% SDS, 1 mmol/L EDTA, 10 mmol/L Tris-Cl, pH 8.1, protease inhibitors). The cell pellet was resuspended with 1 mL of shearing buffer, and the DNA was sheared to 300 to 500 bp size fragments with a Covaris E220 Evo sonicator. Twenty micrograms of chromatin, antibody, and protein-G bead complexes were incubated overnight on a rotator at 4°C. Then, the beads were washed with low salt buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 20 mmol/L HEPES-KOH, pH 7.9, 150 mmol/L NaCl), high salt buffer (0.1% SDS, 1% Triton X-100, 2 mmol/L EDTA, 20 mmol/L HEPES-KOH, pH7.9, 500 mmol/L NaCl), and LiCl buffer (100 mmol/L Tris-Cl, pH 7.5, 500 mmol/L LiCl, 1% NP40, 1% sodium deoxycholate) twice and with TE buffer (10 mmol/L Tris-Cl, pH 8.1, 1 mmol/L EDTA) once. Purified DNA was quantified for library preparation.

TruSeq ChIP Sample Preparation Kit (Illumina, IP-202–1012, IP-202–1024) was used according to the manufacturer's instructions to perform chromatin immunoprecipitation sequencing (ChIP-seq) library preparation. Briefly, 5 ng of DNA was used as starting material for input and IP samples. Libraries were amplified using 14 cycles on the thermocycler. Libraries were validated using the Agilent High Sensitivity DNA Kit. Then, libraries were submitted to the Van Andel Institute Genomics Core for sequencing with an Illumina NextSeq500 sequencer with 75 bp single read runs.

The 75-bp single-end reads were trimmed with Trim Galore (version 0.4.0) with default parameters, then the trimmed reads were mapped to human genome hg19 with Bowtie2 (version 2.3.5; RRID: SCR_016368). After deduplication with the software picard (version 2.19.0), the deduplicated reads were used for peak calling by HOMER (version 4.10; ref. 22) with the following settings: -B -region -size 1000 -minDist 2500 -P 0.01 -F 2.0 -L 2.0 -LP 0.01. Peak motifs were identified with HOMER based on the peak calling information. Output bedgraph files from HOMER were processed into bigwig files with UCSC Exe Utilities bedGraphToBigWig (version 1.04.00; ref. 23), then bigwig files were loaded into Integrative Genomics Viewer (IGV) for genome browser snapshots of ChIP-seq enrichment (24). Density plots of ChIP-seq enrichment at peak center regions were made with R package genomation (version 1.28.0; ref. 25). Heat maps for comparing the enrichment pattern between control and KO samples were plotted with R package Enrichedheatmap (version 1.22.0; ref. 26), which are presented in **Fig. 1C**. Peaks were sorted by the peaks in control. For all overlap analysis of ChIP peaks, we used bedtools (RRID:SCR_006646) intersection function, as well as for all other region overlap analysis in this study. Significant difference between two groups in metaplots was calculated by Welch two-sample t test with R (version 4.2.0).

Whole-genome bisulfite sequencing and data analysis

Genomic DNA was extracted from HBEC controls and various HBEC derivative cells (three different clones for each group) using a Quick-DNA Miniprep Plus kit (Zymo Research, D4070). DNA samples were submitted to the Van Andel Institute's Genomics Core [Controls, RYBP KO (RKO), TET2 KO, TET2/3 DKO, TET123 (T123) triple KO (TKO), and RYBP/T123 quadruple KO (QKO)] or Fulgent [YAF2 KO (YKO) and TET3 KO] for library preparation. The Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences, Ann Arbor, MI, 30024) was used according to the manufacturer's instructions to perform bisulfite conversion and library preparation. Sequencing was performed with an Illumina NovaSeq 6000 or HiSeq 2500 system with 150 bp paired end read runs. Generally, all libraries displayed high Q-scores (>30) in both read pairs. We obtained data corresponding to approximately 24x genome coverage on average. Paired-end sequencing reads were aligned to the hg19 human genome using Bismark (27). Adaptors and low-quality reads were trimmed using the parameters as described previously (21).

We used DMRseq (version 0.99.0; ref. 28) to identify differentially methylated regions (DMR). CpG methylation values were called by the Bismark methylation extractor script provided with Bismark, using the parameters as described previously (21). Sequencing depth for CpGs with at least three covering reads for each sample were considered for DMR calling, as well as for further analysis of DNA methylation. A single CpG coefficient cutoff of 0.05 was used for candidate regions. Significant DMRs between the control and modified cells were identified using $q < 0.05$.

To analyze the DMR motifs, HOMER was used for identifying putative motifs within DMRs. We used the following parameters on bed files of DMRs: findMotifsGenome.pl -size given -mask -cpg. Known motifs (as opposed to *de novo* motifs) from HOMER were scored. We identified the functional enrichments of DMRs and plotted them with ReactomePA (version 1.40.0; ref. 29), ChIPseeker (version 1.32.0; ref. 30) and the clusterProfiler (version 4.4.4) package (31). To identify the DMRs common among different genotype groups, upset plots and pairwise plots were plotted by Python tool Intervene (32), with intervene upset and intervene pairwise function. IGV was used to view the data tracks. We plotted the heat maps of DMRs and CGIs among different genotype groups with R package pheatmap (version 1.0.12).

The CpG densities of methylation distribution between samples were calculated and plotted by the R package genomation (version 1.28.0). We plotted the Circos plot of DMRs of different genotype groups with circlize (version 0.4.10; ref. 33).

Significant difference between two groups in metaplots was calculated by Welch two-sample t test with R (version 4.2.0).

The sequencing copy number on chromosome 12 was visualized with the GenVisR (version 1.28.0.; ref. 34).

Gene ontology and pathway enrichment analysis

We used various gene sets for gene ontology and pathway enrichment analysis of differential expressed genes using DAVID functional annotation analysis (35). Functional enrichment of ChIP-peaks was performed with R package ReactomePA (version 1.40.0; ref. 29), ChIPseeker (version 1.32.0; ref. 30), and clusterProfiler (version 4.4.4) package (31).

(*Continued.*) The ChIP-seq signal is ranked from highest to lowest for each mark in control cells. **D,** ChIP-seq cumulative enrichment deposition centered at peak summit of H2AK119ub1 in control, RKO, YKO, and R/Y DKO cells. **E,** ChIP-seq cumulative enrichment deposition centered at peak summit of H3K27me3 in control, RKO, YKO, and R/Y DKO cells. **F,** ChIP-seq cumulative enrichment deposition centered at peak summit of RYBP in control and RKO cells. **G,** IGV browser view of the *FOXQ1* gene and a cluster of genes, *NUDT18, HR, REEP4, LIG3* for RYBP, H2AK119ub1, and HEK27me3 ChIP-seq data acquired from control and RKO cells. Three independent clones in control cells and two independent clones in RKO cells are presented. **H,** Venn diagram showing overlap of H2AK119ub1, H3K27me3, and RYBP peaks in HBEC3 control cells. **I,** Venn diagram showing overlap of CGIs targeted by H2AK119ub1, H3K27me3, and RYBP peaks in HBEC3 control cells.

Nuclear protein and histone extraction and Western blotting

We isolated nuclear proteins with NE-PER Nuclear and Cytoplasmic Extraction Reagents kit (Thermo, 78835) according to the manufacturer's instructions.

Histone acid extraction was performed by washing the cell pellet from one 10 cm dish with PBS twice. The pellet was resuspended and incubated with 1 mL hypotonic lysis buffer including additionally 50 μmol/L of the deubiquitinase inhibitor PR-619 (Sigma Aldrich) for 30 minutes on a rotator at 4°C. We resuspended the nuclei in 0.4 N H_2SO_4 and incubated on a rotator overnight at 4°C. We added trichloroacetic acid to the supernatant containing histones and incubated the solution on ice for 30 minutes. We collected precipitates by centrifugation. Before dissolving the histone pellet in water, we performed two steps of washing with ice-cold acetone.

The antibodies used for Western blotting are listed in Supplementary Table S1. We used Image J (RRID:SCR_003070) for quantitative analysis of Western blot signals of H2AK119ub1, H3K27me3 and histone H3.

Immunofluorescence and microscopy

Cells were seeded in a chamber and fixed with ice-cold 100% methanol for 15 minutes at −20°C, then rinsed 3 times in PBS for 5 minutes each. Then, we blocked the samples with 5% BSA in PBST (PBS containing 0.3% Triton X-100) for 1 hour, and subsequently incubated them with the diluted primary antibodies in 1% BSA in PBST for overnight at 4°C. Cells were washed 3 times with PBS, incubated with fluorescent secondary antibody in 1% BSA for 1 hour at room temperature in the dark, and finally decanted and washed in the dark prior to taking images. We captured fluorescent images on a Leica inverted microscope.

Cell proliferation and soft agar growth assays

We monitored cell proliferation using MTT assay (Sigma, M5655). One thousand viable cells were seeded into 96-well plates, and cells were incubated with MTT solution in a CO_2 incubator for 2 hours. We added 150 microliters of DMSO to each well until all crystals were dissolved. We measured the absorbance at 570 nm on a plate reader (Biotech Instruments).

We determined the clonogenic capacity of HBEC3 and modified cells by soft agar assays. We suspended 5,000 viable cells and seeded them in 0.35% agar in KFSM medium in 6-well plates and overlaid them with a 0.5% agar base in the same medium. We stained the colonies with 0.02% iodonitrotetrazolium chloride (0.02% INT, Sigma, I10406) and counted them using a microscope.

In vivo tumorigenicity and histology analysis

We used female NSG mice (4–5 weeks old; Jackson Laboratories) for these studies. Mice were housed at the Van Andel Institute animal care facility. All animal experiments were approved by the Van Andel Institute Animal Care and Use Committee. All animal care and use protocols followed were in accordance with guidelines of the Institutional Animal Care and Use Committee. We injected mice subcutaneously in the left and right flanks with 6×10^6 viable control cells or RYBP/TET123 QKO cells in 0.2 mL of KSFM with Cultrex BME (R&D Systems, 3632–010–02). Mice were monitored every week for tumor formation for up to two months. Formalin-fixed, paraffin-embedded xenograft tumor tissue was sectioned and stained with hematoxylin and eosin (H&E) for histologic analysis. We performed IHC staining for CK5, p40, SNAIL and VIM at the Van Andel Institute Pathology and Biorepository Core with antibodies specified in Supplementary Table S1. IHC staining was performed on an automated immunostainer (Dako) using standard protocols and stained slides were scanned to digital files (Aperio Scanscope).

The Cancer Genome Atlas data analysis

For analysis of Infinium 450K methylation data from the Cancer Genome Atlas (TCGA), we downloaded the raw IDAT files by R Bioconductor package TCGAbiolinks (version 2.24.3; ref. 36). We used R Bioconductor package minfi (version 1.42.0; ref. 37) to process the raw IDAT files. The function preprocessNoob was applied for background correction based on the normal-exponential out-of-band (NOOB) background subtraction method (38), and then preprocessFunnorm function was used for further processing by functional normalization method (FunNorm; ref. 39). We used the object containing Beta-values returned by the preprocessFunnorm function for DMR calling by bumphunter (version 1.38.0; ref. 37) with bump hunting algorithm (40), with the cutoff 0.1 for the Beta-values and a large number of permutations, B = 1000. For the downstream analysis, only the DMRs with at least two probes in a minimal region of 300 bp were considered. Hypergeometric probability distribution was used for statistical analysis of overlap between cell model and TCGA data. For RNA expression analysis of TCGA data, we downloaded the expression data by FirebrowseR (version 1.1.35), then generated box plots for comparing normal solid tissue and primary tumor with R (version 4.2.0).

For the TCGA data and CGI methylator phenotype (CIMP) analysis, RNA-seq read counts for TCGA lung squamous cell carcinoma (LUSC) samples based on GENCODE v29 gene annotations were obtained using the recount3 v1.6.0 R package (41). Per-gene raw counts were calculated using the 'compute_read_counts' function. Samples with less than or equal to 15 million reads were removed. Samples with the same sample ID were deduplicated by retaining only the sample with the highest number of uniquely mapping reads. CIMP subclasses (42) were merged with the sample meta data based on the TCGA sample ID and used to group the samples as Normal, CIMP.L (low), CIMPi (intermediate), or CIMP.H (high). Tumor samples with no matching CIMP labels were removed. To remove lowly expressed genes, only genes with more than 10 read counts in at least 51 samples (sample size of the smallest group, which was the Normal samples) were retained. Data from all groups, as defined above, were fit using a design of '∼ group' using the default workflow encoded in the 'DESeq' function in DESeq2 v1.36.0 (43). Pairwise contrasts were tested using the 'results' function with the parameter, 'alpha = 0.01'. Variance stabilizing transformed (VST) counts were obtained using the 'vst' function with the parameter, 'blind = FALSE'. Violin plots were produced using the scater v1.24.0 (44) and ggpubr v0.6.0 packages, after storing the sample meta data, gene meta data and VST counts in a SingleCellExperiment object (45).

Statistical analysis

Data are presented as mean ± SD (unless otherwise noted) and were derived from at least three independent experiments. Data on replicates is given in the Figure legends. Statistical analysis was performed using the two-tailed *t* test or one-way comparison ANOVA. Statistical analyses for DNA methyl-seq, RNA-seq, and ChIP-seq data are described in those respective methods sections.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All sequencing datasets are available under the Gene Expression Omnibus accession number GSE208689. The data analyzed in this study were obtained

from TCGA in datasets TCGA-LUSC (https://portal.gdc.cancer.gov). All other raw data are available upon request from the corresponding author.

Results

Inactivation of RYBP in bronchial cells

We first analyzed TCGA database and examined the expression of genes coding for subunits of Polycomb repression complexes (PRC1 and PRC2). Among over 50 genes analyzed for PRC1 and PRC2, we found that most Polycomb genes were upregulated in tumors across a wide spectrum of malignancies. The only downregulated genes in lung cancer and other tumors were *EZH1* for PRC2 and *RYBP*, *PCGF5*, *CBX6* (in lung adenocarcinomas only), and *CBX7* for PRC1 (**Fig. 1A**; Supplementary Fig. S1A and S1B). *EZH1* downregulation is accompanied by a strong upregulation of *EZH2* as seen in many cancer types (TCGA database). We decided to focus on RYBP because this protein is present in most if not all variant (also called noncanonical) PRC1 complexes where it is mutually exclusive with CBX proteins (46). Variant PRC1 complexes are responsible for the largest fraction of total histone H2A K119 monoubiquitylation (H2AK119ub1) in cells (47). RYBP is a strong activator of the RING1B ubiquitin ligase activity of PRC1 that produces H2AK119ub1 and provides a positive feedback loop for focused PRC1 activities through its binding to H2AK119ub1 (47). *RYBP* is downregulated in lung adeno- and squamous cell carcinomas (**Fig. 1A**) but also across many other types of solid tumors (Supplementary Fig. S1A). Because the focus of our research has been on DNA methylation in lung cancers (8, 48), we proceeded to inactivate RYBP in human bronchial epithelial cells. We used the telomerase- and CDK4-immortalized bronchial cell line HBEC3-KT (49, 50). HBEC3-KT cells are nontumorigenic and could not even be transformed by simultaneous introduction of a *KRAS* mutation, *EGFR* mutation and by inhibiting the TP53 tumor suppressor (50). We employed CRISPR-Cas9 technology to target the RYBP gene. Creating frameshift mutations (Supplementary Fig. S1C; Supplementary Table S2), we generated several RYBP-deficient clones, along with control clones (**Fig. 1B**; Supplementary Fig. S1C and S1D). Because RYBP has a paralogue in the human genome, YAF2, we also inactivated YAF2 (Supplementary Fig. S1E), alone or in combination with RYBP, creating a double KO (Supplementary Fig. S1F). *YAF2* was not downregulated in lung tumors relative to normal lung (**Fig. 1A**). We next assessed the levels of the PRC1 Polycomb mark, H2AK119Ub1, and the PRC2 Polycomb mark, histone H3 K27me3 (H3K27me3), considering the extensive crosstalk between PRC1 and PRC2 activities (46, 47). Western blotting to assess global levels of these marks showed a reduction of H2AK119ub1 and to a lesser extent of H3K27me3 in the RYBP inactivated cells or the RYBP/YAF2 double KOs (R/Y DKO) but not in the single YKO (Supplementary Fig. S1G–S1I).

We then performed ChIP-seq of the two Polycomb marks and of RYBP protein in control and RYBP/YAF2 targeted cells (**Fig. 1C–I**) and determined peak distributions (Supplementary Table S3, Excel file). This data shows an almost 50% reduction of the H2AK119ub1 signal in cells lacking RYBP (**Fig. 1C** and **D**) and a nearly complete loss of RYBP peaks in RYBP-deleted cells (**Fig. 1C** and **F**). Loss of H3K27me3 was also substantial (**Fig. 1C** and **E**). Examples of ChIP-seq data for H2AK119ub1 across two gene loci is shown in **Fig. 1G**. To analyze the overlap between H2AK119ub1, H3K27me3, and RYBP peaks, we created Venn diagrams (**Fig. 1H** and **I**). There was a partial overlap between the two histone marks but less than 10% of all peaks were common to all three mapped parameters. This data shows

that PRC1 and PRC2 activities do not always coincide in these somatic cells. However, considering CGIs, about 40% of the RYBP-targeted CGIs also carried the H2AK119ub1 mark (**Fig. 1I**). Of note, comparison with data for H3K4me3 mapping in our bronchial epithelial cells revealed that most H3K27me3-associated CGIs did not carry H3K4me3 in somatic lung epithelial cells, in other words, they were not traditional 'bivalent' CGIs. Focusing on CGIs marked by H2AK119ub1 alone or in combination with RYBP, we performed gene ontology analysis. These CGIs were associated with genes involved in developmental and differentiation processes (Supplementary Fig. S2A and S2B). In addition, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed that the RYBP/H2AK119ub1 co-occupied CGIs are linked to genes involved in several cancer-relevant pathways including, for example AKT signaling and the Hippo pathway (Supplementary Fig. S2C).

Next, we performed RNA-seq in the RYBP, YAF2, and R/Y DKO cells (Supplementary Fig. S2D–S2H). Somewhat surprisingly, YAF2 single KO cells did not show any DEGs that reached genome-wide statistical significance. RYBP single and R/Y DKOs had about 1,000 upregulated genes, which was > 3-times more than the number of downregulated genes, consistent with a mostly repressive function of RYBP (Supplementary Fig. S2D–S2F; Supplementary Table S4, Excel file). Upregulated genes showed a loss of the H2AK119ub1 and H3K27me3 marks near the transcription start site (TSS), and downregulated genes showed a small increase of these repressive marks (Supplementary Fig. S2E and S2F). Gene ontology analysis revealed that the DEGs after loss of RYBP fell into several cancer-relevant pathways including cell migration, proliferation, epithelial-to-mesenchymal transition (EMT), and signal transduction (Supplementary Fig. S2G). Upregulated genes included several transcription factors, for example *ERG1*, *ERG2*, *NR4A1*, and *ATF3* (Supplementary Fig. S2H).

We then determined if the loss of RYBP and PRC1/2 histone marks leads to a change in DNA methylation patterns. We analyzed three control clones and three independently derived RKO or YKO clones by whole-genome bisulfite sequencing (WGBS), a technique that comprehensively interrogates all CpG sites in the human genome that have sufficient read coverage. DMR-seq analysis (*see* Methods) identified about 16,000 DMRs, and more than 80% of these DMRs were hypermethylated in the RYBP-deficient cells (**Fig. 2A**). The genomic distribution of hypermethylation DMRs is shown in **Fig. 2B**; they were characterized by a loss of H2AK119ub1 in the RKO and the R/Y DKO (**Fig. 2C** and **D**). YKO showed no significant DMRs. The DNA hypermethylation after loss of RYBP often affected larger genomic regions that lost the PRC1 histone mark H2AK119ub1, as shown for example for the *HOXB* locus, a well-known Polycomb target (**Fig. 2D** and **E**). Some of the hypermethylation events were confirmed by manual, locus-specific bisulfite sequencing (**Fig. 2F**). However, when considering CGIs, the predominant targets of DNA hypermethylation in cancer, we observed that only 298 CGIs contained hypermethylated DMRs, which were abundant instead in intergenic and intragenic regions (**Fig. 2B**). Thus, RYBP inactivation alone was insufficient to cause extensive and specific CGI methylation as seen in tumors.

DNA hypermethylation after inactivation of RYBP and TET proteins

We next considered that 5mC oxidases, the TET proteins, can revert inappropriately incorporated 5mC back to the unmethylated state. TET1 and TET3 have been mapped to CGIs, whereas TET2 has more commonly been associated with enhancer regions (51–53).
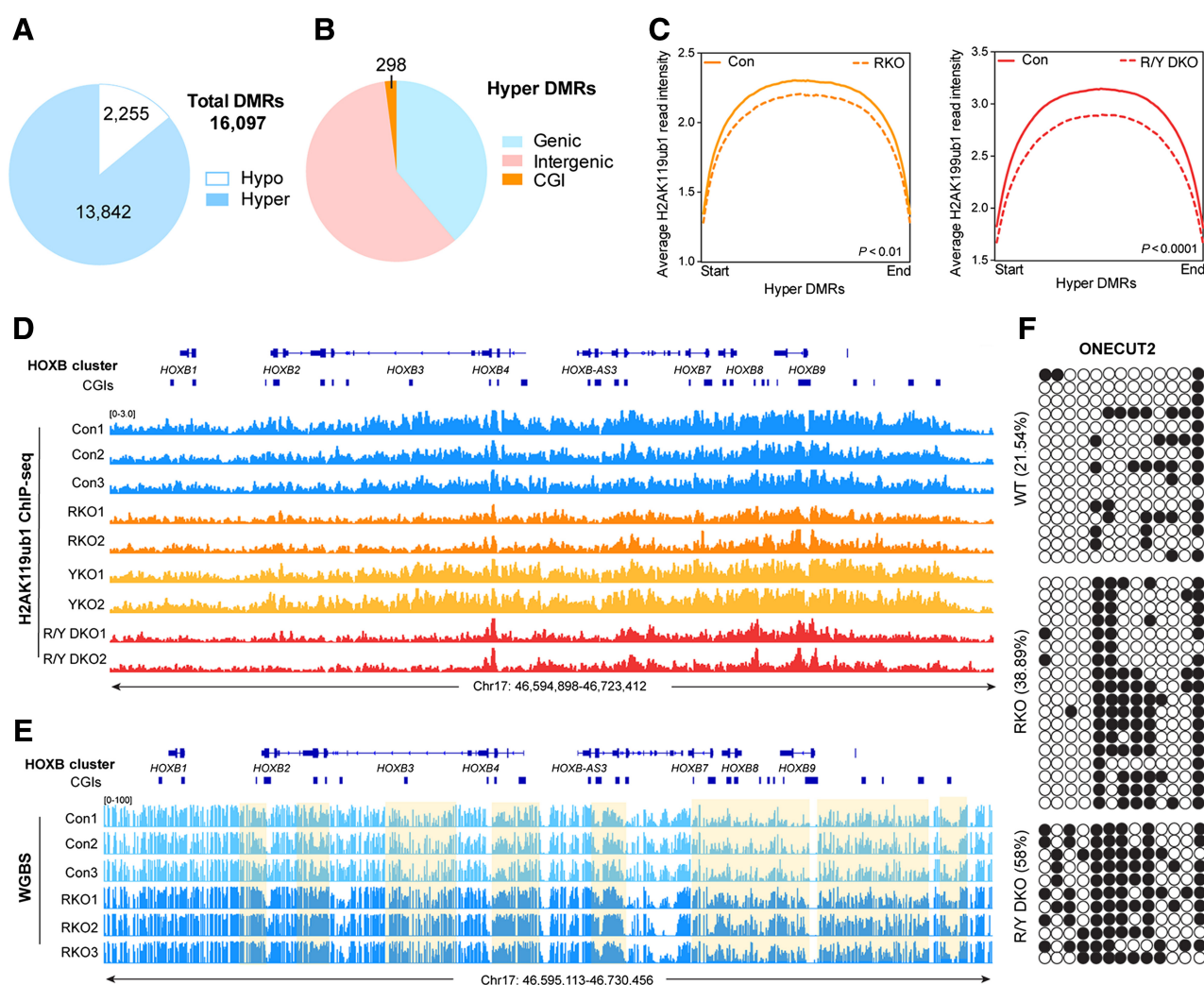
Figure 2.
DNA hypermethylation in RYBP-deficient cells. **A,** Quantities of hypermethylated DMRs (Hyper DMRs) and hypomethylated DMRs (Hypo-DMRs) in RKO cells compared with controls. **B,** Quantities of hypermethylated DMRs that overlapped with the indicated genomic features in RKO cells compared with controls. **C,** Intensity plots illustrating H2AK119ub1 occupancy around hypermethylated DMRs of RKO and RYBP/YAF2 DKO (R/Y DKO), respectively. The combined analysis of three independent clones is presented. *P* value is indicated. **D,** IGV browser view of the *HOXB* cluster for H2AK119ub1 ChIP-seq data acquired from controls, RKO, YKO, and R/Y DKO cells. **E,** IGV browser view of the *HOXB* cluster for WGBS DNA methylation data acquired from control and RKO cells. The hypermethylated regions are shaded. **F,** Bisulfite sequencing analysis of a *ONECUT2* CGI in WT, RKO, and R/Y DKO cells. Solid black circles, methylated CpG sites; open circles, unmodified CpG sites. Total percentage of methylated cytosines (%Me) is indicated.

However, there is likely a considerable overlap between the three TET proteins in their functionality and genomic specificity. TET genes are generally upregulated in human cancer according to the TCGA database, including LUSCs. This contrasts with their inability to oxidize 5mC bases in human solid tumors, as first reported about a decade ago for LUSCs and many other tumor types (18). This phenomenon attests to a dysfunctional TET axis in human cancers.

We first assessed the expression levels of the three *TET* genes in the parental bronchial cells and in the RYBP-inactivated cells using RNA-seq data. Expression of *TET2* and *TET3* was readily detectable but *TET1* levels in the control cells were extremely low (Supplementary Fig. S3A and S3B). After loss of RYBP, expression of *TET1* and *TET2* was significantly increased although *TET1* levels remained relatively low. Coinciding with their activation, we observed a moderate DNA

hypomethylation at the *TET1* and *TET2* promoters in RYBP-deficient cells. On the basis of this data and our hypothesis that combined inactivation of RYBP and TET proteins could have a more pronounced phenotype, we proceeded to inactivate all three TET proteins in bronchial cells. We created single KOs for TET2 and TET3, and various combinations of double and triple KOs as illustrated by the clone tree (Supplementary Fig. S3D). The effort culminated in creation of a QKO that lacked RYBP, TET1, TET2, and TET3. For each targeting, three independent clones with biallelic frameshift mutations were created as confirmed by Western blotting for TET2 and RYBP (Supplementary Fig. S3C), with no suitable antibodies available for human TET1 and TET3, and by extensive DNA sequencing (Supplementary Table S2; Supplementary Fig. S3D).

We performed WGBS on most of the derived KO clones and determined DMRs (Supplementary Table S5, Excel file). A CIRCOS

plot of DMRs for TET2 single KO, TET2/TET3 double KO (DKO), RKO, TET123 TKO, and RYBP/TET123 QKO is shown in **Fig. 3A**. Heat map analysis is shown in **Fig. 3B** for total DMRs and in **Fig. 3C** for CGI DMRs. TET2 (and TET3) single KO and TET2/TET3 double KO cells showed no DMRs that reached genome-wide statistical significance (**Fig. 3A** and **D**). DMRs in the RKO, the TET123 triple, and the RYBP/TET123 QKOs were distributed uniformly along all chromosomes (**Fig. 3A**). Removing all three TET proteins led to more extensive DNA methylation changes and resulted in over 89,000 DMRs (**Fig. 3D**). However, the majority of DMRs in the TET TKO clones were hypomethylated, which may seem *a priori* unexpected for inactivation of 5mC oxidases. There is precedence however that TET protein loss leads to DNA hypomethylation, which was postulated to be caused by a major redistribution of DNMT proteins upon loss of TET (54). Global DNA hypomethylation is observed in many cancer types and primarily involves heterochromatic regions. In the RYBP/TET123 QKO cells, we observed 71,000 DMRs, but now the direction of the DMRs was shifted again towards hypermethylation (**Fig. 3D**).

Focusing on CGIs, the TET triple KOs and most strikingly, the QKOs had predominantly hypermethylated regions, which affected over 5,000 CGIs in the RYBP/TET123 QKO (**Fig. 3E**). This finding is different from the RKO, which had more hypomethylated DMRs than hypermethylated DMRs at CGIs (**Fig. 3E**). The extent of DNA hypermethylation was most dramatic in the RYBP/TET123 QKO cells (Supplementary Table S6, Excel file). The RYBP/TET123 QKO had 3,176 hypermethylated CGIs not found in either the RKO or the TET123 TKO (Supplementary Fig. S4A). We show examples for the gene *HAGLR*, which is embedded in the *HOXD* locus (**Fig. 3F**) and for the gene *ACTG1* (**Fig. 3G**). Making various pair-wise comparisons between cells of different genotype, we determined that several DMRs are shared as shown in Supplementary Figs. S4B and S4C. For example, comparing the DMRs found between QKO versus control, QKO versus TET TKO, and QKO versus RKO reveals that 1,018 CGIs are in common for these comparisons.

Hypermethylation of Polycomb-targeted CGIs

We intersected our Polycomb mapping data from our ChIP-seq experiments (**Fig. 1**; Supplementary Table S3) with the DNA methylation data obtained by WGBS. This analysis revealed that CGIs marked by H2AK119ub1 frequently become methylated in the RYBP/TET123 QKO cells (**Fig. 4**). **Figure 4A** shows several genome browser examples of DNA methylation at CGIs associated with homeobox and other transcription factor genes. These CGIs are all marked by H2AK119ub1. We centered all CGIs of the genome in composite profile plots. Although a substantial fraction of all genomic CGIs is methylated in control cells, we observed an overall increase of DNA methylation levels in the QKO samples versus controls (**Fig. 4B**). We saw a similar increase when we separately analyzed CGIs occupied by H2AK119ub1 (**Fig. 4C**), H3K27me3 (**Fig. 4D**), or RYBP (**Fig. 4E**). Of note, Polycomb marked CGIs—but less so RYBP-occupied regions—are already partially methylated in controls and then acquire higher methylation levels in the QKO. This is in a way similar to normal human tissues, in which Polycomb-targeted CpG-rich regions acquire partial methylation during noncancerous processes including inflammation and aging (55–61). The increase of methylation affected the entire length of the CGIs globally rather than being targeted to certain regions of them such as CGI centers or CGI shores (**Fig. 4B–E**).

However, the numbers of CGIs that became hypermethylated were different depending on the associated mark (**Fig. 4F**). The largest number of CGIs subjected to hypermethylation in the QKO were those that carried the H2AK119ub1 mark ($n = 3,995$), followed by H3K27me3 ($n = 2,024$) and RYBP ($n = 810$). Because the total number of CGIs with hypermethylated DMRs in the QKO was around 5,200 (**Fig. 4G**), we calculated from these numbers that 76% of all CGI-specific hypermethylated DMRs were targeted to H2AK119ub1-marked CGIs. We also determined the percentage of CGIs associated with a particular mark/protein that did undergo hypermethylation (**Fig. 4H**). This analysis showed that even though H3K27me3-marked targets are fewer in number (**Fig. 4F**), over 60% of them were methylated in the QKO. Forty-one percent of H2AK119ub1-marked CGIs became hypermethylated in the QKO (**Fig. 4H**). This number was much smaller in the RKO (5.6%) and in the TET-TKO (25%).

Because DNA methylation and Polycomb are often mutually exclusive, it was of interest to determine if the acquisition of DNA methylation at Polycomb-marked CGIs in the QKO was linked to a loss of the Polycomb marks. Indeed, the QKO cells showed a profound loss of H2AK119ub1 (57%) and to a lesser extent of H3K27me3 (15%) over CGIs (Supplementary Fig. S5A and S5B). When we sorted the H2AK119ub1 ChIP-seq data according to the extent of acquired DNA methylation, we observed that those CGIs with the highest level of H2AK119ub1, showed the greatest methylation gain (Supplementary Fig S5C).

We wondered how the hypermethylation of CGIs seen in the QKO bronchial epithelial cells relates to DNA methylation changes seen in human lung cancers, in particular LUSCs. We extracted the DNA methylation values at CGIs from the TCGA database and established a list of LUSC hypermethylated CGIs for those samples for which normal lung controls were available ($n = 40$). This list had 3,666 hypermethylated CGIs (**Fig. 4I**). Almost half of them (1,743 = 47.5%) overlapped with the hypermethylated CGIs found in the RYBP/TET123 QKO bronchial cells, which was statistically highly significant (**Fig. 4I**; Supplementary Table S7, Excel file). A substantial fraction (56.6%) of the TCGA hypermethylated CGIs carried the H2AK119ub1 mark in bronchial cells. More than 1,300 CGIs (36.1%) carried the PRC1 mark and became hypermethylated in both the QKO and in human lung SCC (**Fig. 4I**).

According to the extent of CGI hypermethylation, LUSCs have been previously classified in a CIMP-like manner (42). We have used this data, which is based on TCGA, to categorize samples according to the extent of CGI methylation. Following this classification, we analyzed the expression levels of *RYBP* and of the *TET* genes in relation to the "CIMP-low," "CIMP-intermediate," and "CIMP-high" assigned status (Supplementary Fig. S6). This data shows that the CIMP-high samples have the lowest expression of *RYBP*, but *TET* genes do not follow this correlation.

We asked if the DNA hypermethylation events at CGIs in the RYBP/TET123 QKO and TET TKO cells were associated with specific DNA sequences. Motif analysis revealed the enrichment of transcription factor motifs in the hypermethylated regions (Supplementary Fig. S7). Most striking was the specific enrichment of homeobox transcription factor motifs, which occurred in both KOs and were also seen when we subsampled the hypermethylated regions to focus on those marked by H2AK119ub1 (Supplementary Fig. S7A–S7C). These homeobox binding motifs bear same similarity, are generally AT-rich, and surprisingly lack any CpG dinucleotides (Supplementary Fig. S7A–S7C). However, there is a smaller set of CGIs that become hypermethylated but are not marked by H2AK119ub1 (about 1,000). Motif analysis of this subset showed the enrichment of transcription factors with motifs containing CG sequences, including, for example members of the E-twenty-six (ETS) family of transcription factors, YY1, and E2F (Supplementary Fig. S7D). We do not currently know the reasons why such different motifs are found depending on
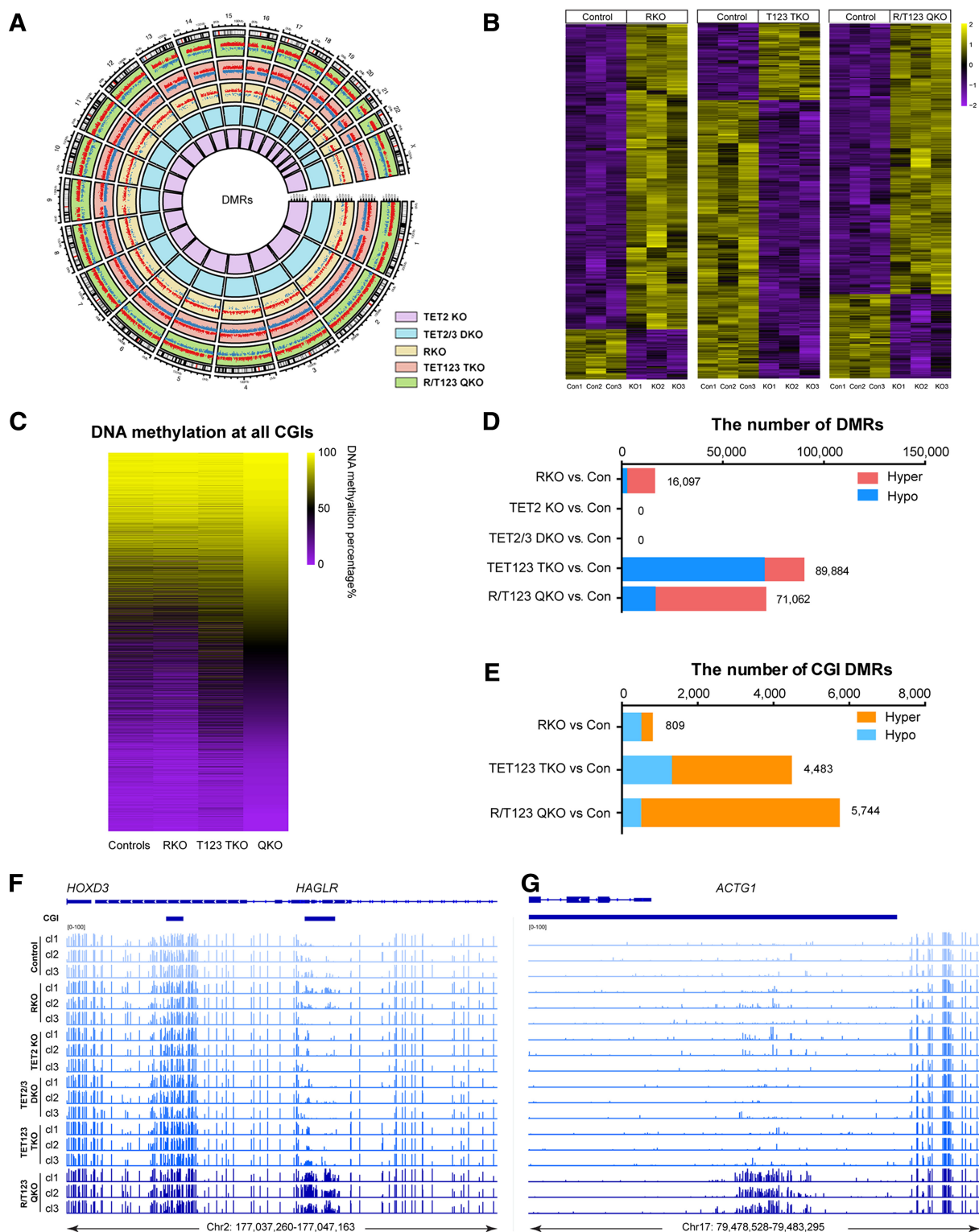
Figure 3.
DNA methylation changes in cells deficient in RYBP and TET proteins. **A,** CIRCOS plot showing distribution of hypermethylation-DMRs and hypomethylation-DMRs on all chromosomes in RKO, TET2 KO, TET2/3 DKO, TET123 TKO, and RYBP/TET123 QKO cells. Red dots, hypermethylation-DMRs; blue dots, hypomethylation-DMRs. The coordinates represent the methylation percentage; positive values represent hypermethylation; negative values represent hypomethylation. Different background colors are used for the different KO cells. (*Continued on the following page.*)

Polycomb marking. One possibility is that the non-CG motif-binding homeobox transcription factors are involved in recruiting PRC1 complexes at an early stage of development. However, the hypermethylation of CGIs that are not marked by PRC1 should proceed by a mechanism that does not involve loss of PRC1. One possibility is that the CGIs not marked by H2AK119ub1 are poorly protected from DNA methylation by TETs given the known association of TET proteins with Polycomb complexes (53). However, other scenarios are possible. For example, ETS-like factors may recruit DNMTs, or perhaps methylation of these CGIs occurs because of selection of a tumor-promoting phenotype (we consider this less likely).

Cell transformation and tumor growth of RYBP- and TET-deficient cells

We performed cell growth assays in soft agar for the control cells, TET2 single KO, RKO, TET123 TKO and the RYBP/TET123 QKO cells (**Fig. 5A–C**). Confirming earlier data showing that HBEC3-KT cells are nontumorigenic (49, 62), the parental control cells did not grow in soft agar and neither did TET2 and RKOs cells (**Fig. 5A–C**). TET123 TKO cells showed few, very small colonies. However, the RYBP/TET123 QKO cells showed high colony forming efficiency and generated large size colonies in soft agar, indicating that they had acquired properties of the transformed phenotype (**Fig. 5A–C**). The modified cells did not display an enhanced cell division rate because all TET- or RYBP-deficient cells grew more slowly than control cells (**Fig. 5D**).

To determine if growth in soft agar is related to tumor growth in mouse xenotransplantation experiments, we transplanted control and QKO cells subcutaneously into the flanks of immune-deficient female NSG mice. Two controls, wildtype cells and empty vector transformed wildtype cells, did not form tumors. Tumor formation was observed at 29 of 31 injection sites with the three QKO cell clones (QKO7, QKO32, and QKO39; **Fig. 5E**). This result was repeated in an independent transplantation experiment. Histopathology review of QKO cell line xenografts showed varied amounts to tumor cells in nodular configuration and epithelioid clusters of tumor cells with small cystic structures, single cell necrosis, compressed and angulated cellular contours with invasive features and single cell migration suggestive of EMT (most prominent for clone Q32). The cytologic features of tumors were most remarkable as the supporting matrix used in cell culturing (Matrigel-like material) was moderately well retained in the xenograft with minimal host (mouse) lymphoid or macrophage contaminant allowing for clean immunostaining. The clones QKO7 and QKO39 formed clusters of tumor cells with differentiation and cysts of variable size with single lining of tumor cells as determined by H&E staining (**Fig. 5F**), whereas QKO32 tumors presented with more isolated and solid clusters of cells and in streaming pattern at the xenograft periphery (Supplementary Fig. S8D). We used IHC methods to stain all QKO tumor samples with antibodies against the squamous cell carcinoma markers p40 (recognizing the short isoform of TP63 expressed from the *TP63* gene), and with antibodies against cytokeratin 5 (CK5). This staining showed high positivity, suggesting that the tumors formed were squamous cell carcinomas (**Fig. 5F**). We also stained the sections with anti-vimentin (VIM) and SNAIL markers to determine if cells had undergone an EMT. The staining results suggested that clone Q32 has properties of EMT (Supplementary Fig. S8D). In Q32, although expression of *CDH1* (E-cadherin) was not downregulated by RNA-seq, several EMT marker genes were upregulated including *VIM*, *CDH2* (N-cadherin), *SNAI1/2, TWIST1/2, and TGFB2* and this was confirmed at the protein level and by immunofluorescence for VIM (Supplementary Fig. S8A–S8C).
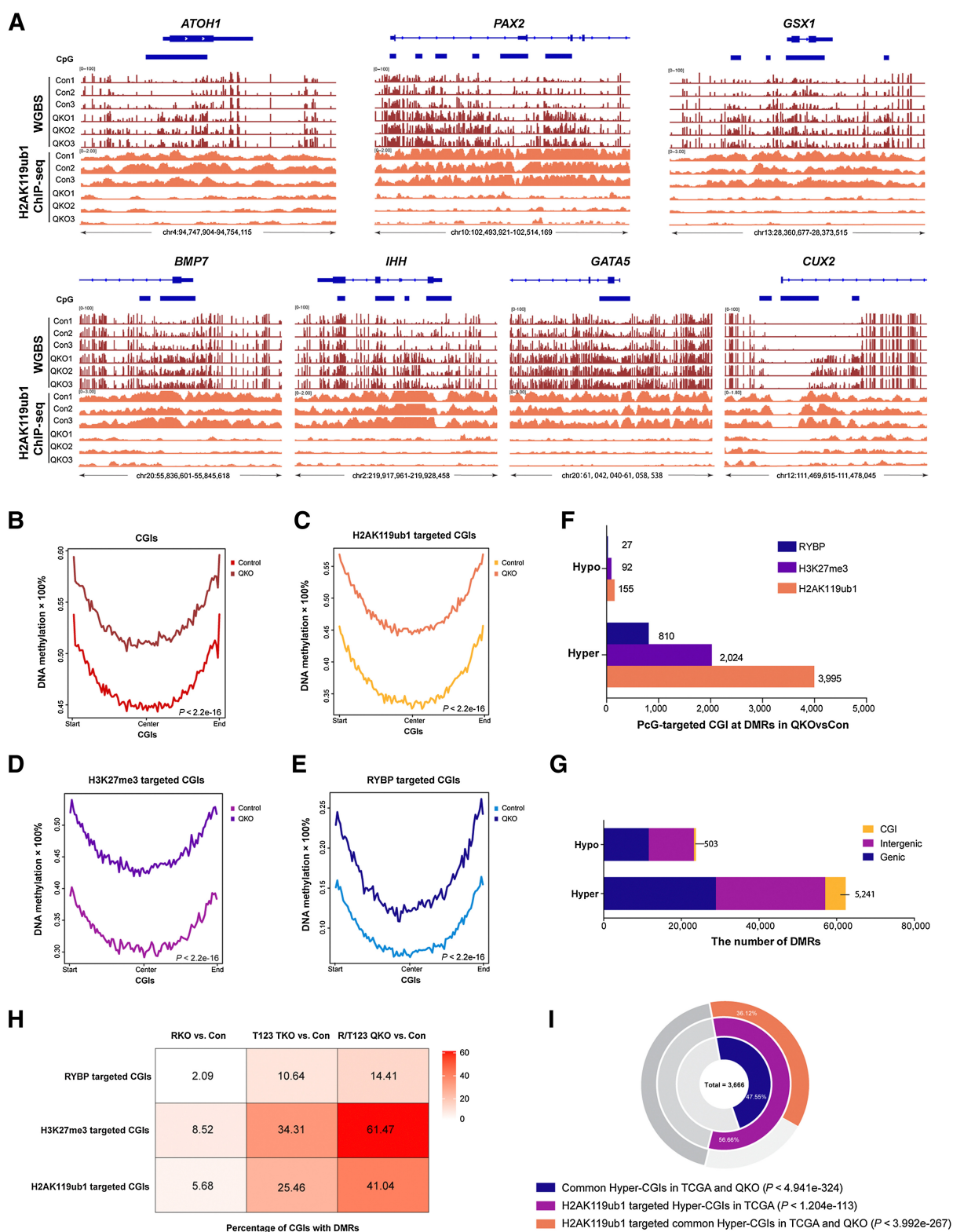
Activation of tumor-promoting signaling pathways

Because we observed efficient cell transformation for the R/TET123 QKO cells, we investigated if cancer-related pathways had become aberrant. We used RNA-seq to analyze gene expression patterns in the RKO cells, TET123 TKOs and the RYBP/TET123 QKOs (Supplementary Table S4). The TET-TKO had the smallest number of DEGs, only about 400 (**Fig. 5G**). Both the RKO and the RYBP/TET123 QKO cells had over 2,000 DEGs. Many cancer-relevant genes were differentially expressed in the QKO versus control comparison. Some of these genes are listed in **Fig. 5G** and genome browser views are shown in **Figs. 5I** and **J**. There was a strong upregulation of genes of the AP-1 transcription factor family (*FOS, JUN, ATF3, FOSL1*, and others; **Fig. 5I**). KEGG pathway analysis (**Fig. 5H**) showed the enrichment of cancer pathway-related terms, including Hippo signaling and various other signal transduction pathways. **Figure 5J** shows the substantial upregulation of downstream growth-promoting genes negatively regulated by the Hippo pathway [*CCN2*, also called connective tissue growth factor (*CTGF*), *ANKRD1*, and *TGFB2*]. To understand why these genes may be upregulated, perhaps in connection with the extensive DNA hypermethylation we had induced in the QKO cells, we looked for DNA methylation-associated gene silencing of potential regulators of these growth-enhancing pathways.

First, we analyzed the correlation between methylation at promoter-associated CGIs and gene expression using global analysis (Supplementary Fig. S9A). After applying a cutoff, we found 350 genes that were hypermethylated and downregulated (R = 0.27; *P* = 3.4e-07; Supplementary Fig. S9A). For the downregulated genes, which in many cases carried the unusual 'bivalent' combination of both the H3K4me3 and H2AK119ub1 marks in control bronchial cells, there is increased methylation at the TSS (Supplementary Fig. S9B). We show an example for the *GJB2* gene encoding a gap junction protein (Supplementary Fig. S9C), where methylation at the promoter-associated CGI coincides with strong downregulation of the gene.

Looking at known cancer-relevant genes, we did not find DNA hypermethylation of the CGI associated with the *CDKN2A* gene. There are several negative regulators of the WNT pathway. However, we could not identify methylation-linked downregulation of any of them. For example, the Secreted Frizzled Related Protein genes (*SFRPs*) or *DKK* genes were either not expressed in bronchial cells (*DKK2, DKK4, DKKL1, SFRP2–5*) or not downregulated in the QKO (*DKK1, DKK3, SFRP1*). We next analyzed upstream regulators of the Hippo pathway. FAT protocadherins are known positive regulators of Hippo anti-growth signaling (63). *FAT2* was strikingly

---

(*Continued.*) **B,** Heat map of DNA methylation changes (all DMRs in KOs compared with controls) for RKO, TET123 TKO, and RYBP/TET123 QKO. Each row of the heat maps represents a DMR. **C,** Heat map showing DNA methylation at all CGIs in control, RKO, TET123 TKO, and RYBP/T123 QKO. Each row represents one CGI ranked by the methylation level in QKO. **D,** The numbers of all hypermethylated DMRs and hypomethylated DMRs in the RKO, TET2 KO, TET2/3 DKO, TET123 TKO, and R/TET123 QKO cells (KOs compared with controls). **E,** The numbers of hypermethylated DMRs and hypomethylated DMRs that overlap with CGIs in the RKO, TET123 TKO, and R/TET123 QKO (KOs compared with controls). The combined analysis of three independent clones is presented in **A–E. F** and **G**. IGV browser views of the *HAGLR* and *ACTG1* loci including CGIs for WGBS data acquired from controls, TET2 KO, TET2/3 DKO, TET123 TKO, and RYBP/TET123 QKO cells. Three independent clones in the different KO cell lines are presented.

Figure 4.
Hypermethylation of Polycomb target genes in the RYBP/TET123 KO cells. **A,** IGV browser views of the genes, *ATOH1, PAX2, GSX1, BMP7, IHH, GATA5,* and *CUX2,* showing WGBS data acquired from controls and RYBP/T123 QKO cells and corresponding H2AK119ub1 ChIP-seq tracks. **B,** Metaplot illustrating DNA methylation levels at all CGIs in controls and R/T123 QKO cells. **C,** Metaplot illustrating DNA methylation levels at H2AK119ub1-targeted CGIs in controls and RYBP/T123 QKO cells. **D,** Metaplot illustrating DNA methylation levels at H3K27me3-targeted CGIs in controls and RYBP/T123 QKO cells. (*Continued on the following page.*)

downregulated in the QKO cells (Supplementary Fig. S10A). According to ENCODE data, this gene is controlled by a proximal enhancer/promoter region and an upstream enhancer. Both regions were occupied by RYBP and by the histone mark H2AK119ub1. The two regions became hypermethylated in the QKO (Supplementary Fig. S10A). In addition, we observed methylation and downregulation of several members of the *RASSF* gene family (*RASSF1A*, *RASSF4*, and *RASSF6*), which encode proteins that activate the mammalian MST1/2 Hippo kinases. The RASSF1A isoform is a known lung tumor suppressor (20). We see methylation of its H2AK119ub1-marked promoter CGI in the QKO and downregulation of the *RASSF1A* isoform (Supplementary Fig. S10B).

The RYBP/TET123 QKO cells showed an upregulation of the *KRAS* gene (Supplementary Fig. S11A). There were no *KRAS* mutations in the QKO as determined by extensive DNA sequencing. The short arm of chromosome 12, where *KRAS* is located, showed a copy number gain in the QKO cells, leading to upregulation of most genes on 12p (Supplementary Fig. S11B and S11C). We hypothesized that the 12p chromosomal instability in these cells may have been induced by the extensive DNA hypermethylation in these cells and by downregulation of genes that control genetic stability. Indeed, we found two genes that fulfill these criteria. The gene *HLTF* coding for helicase-like transcription factor, a RAD5-related translocase, is involved in genome stability and DNA repair (64). This gene is downregulated in the QKO, its 5′ CGIs is associated with RYBP and H2AK119ub1 in bronchial cells, and the CGI becomes hypermethylated in the QKO (Supplementary Fig. S11D). Another candidate gene for maintaining genome stability is *KLHDC8B*, which encodes a protein involved in centrosome function and chromosome stability (65). This gene is also strongly downregulated and becomes hypermethylated at the promoter CGI in the QKO cells (Supplementary Fig. S11E).

Discussion

One of the most pervasive epigenetic aberrations in human tumors is CGI methylation, which occurs preferentially at Polycomb target genes. CGI methylation in cancer is widespread, is seen in every malignancy, and is observed at a lower intensity already in some premalignant tissues (66, 67), and even in tissues undergoing inflammatory reactions, and during aging (55, 56, 58–61). We propose that Polycomb-marked CGIs are inherently susceptible to hypermethylation owing to long-term instability of two protection mechanisms. The methylation increase is dramatically accelerated during malignant progression, eventually resulting in several thousand of such CGIs becoming methylated in tumors.

Various mechanisms for the CGI hypermethylation process have been proposed. They include, for example overexpression of DNMT proteins in cancer, but this hypothesis has not been fully substantiated. DNMT targeting might be involved, for example via the recently described interaction of the unstructured N-terminus of DNMT3A1 and H2AK119ub1 (68).

We focused on protection mechanisms that may break down over time and during malignant transformation. These protective shields against methylation are two-pronged, consisting of the Polycomb

complexes themselves and of the 5mC oxidases; both are known to localize to CGIs. Here we examined the role of the RYBP protein because it is a critical activator of variant PRC1 complexes, and unlike most other Polycomb genes, *RYBP* is downregulated in many tumor types (**Fig. 1A**; Supplementary Fig. S1A, Supplementary Fig. S6). However, this choice does not mean that RYBP dysfunction is the only possible defect of the Polycomb system in cancer. Other members of PRC1, or even PRC2, may also become dysfunctional through various mechanisms, such as downregulation of expression (e.g., *PCGF5*, *CBX7*), increased protein instability, or altered posttranslational modifications. Furthermore, simultaneous dysfunction of several members of these large complexes may be additive.

Cells contain a powerful backup system to correct inappropriately introduced 5mCs by demethylation mediated by the TET proteins. Loss of all three TET proteins in mouse embryos or embryonic stem cells leads to hypermethylation of CGIs (69). However, the effects of TET protein loss on CGI methylation patterns in somatic cells have not been explored in much detail. TET1 and TET3 are known to bind to CGIs, likely through their CXXC domains, which have affinity to unmethylated CpG-rich DNA sequences. TET1 has been found in association with Polycomb complexes and binds to Polycomb target genes (13, 53). Also, TET2, which is often mapped to enhancers, may be recruited to CGIs by its associated binding partner CXXC4/IDAX, which is another CXXC domain family protein (70). This considerable redundancy helps to protect CGIs from DNA methylation. Importantly, TET protein function appears to be substantially reduced in tumors. First, TET2 is mutated in hematological malignancies. Mutations in isocitrate dehydrogenase (IDH1, IDH2) create a neomorphic enzyme that generates 2-hydroxyglutarate, a competitive inhibitor of TET proteins (71). However, IDH mutations are limited to certain tumor types such as gliomas or cholangiocarcinomas. On a much broader scale, TET protein activity appears to be strongly diminished in all human solid tumors, as shown by vanishing levels of 5-hydroxymethylcytosine in the malignant tissue, even in nondividing cells of the tumor mass (18). It has been puzzling to understand the cause of this 5hmC loss, but this finding is not simply related to downregulation of *TET* genes in tumors (Supplementary Fig. S6). TET protein function may be impaired by other cancer-associated disturbances such as changes in metabolism that affect the availability of the cofactors α-ketoglutarate and/or ascorbic acid, and/or a hypoxic tumor environment (72) or by other unknown mechanisms. Regardless of the exact mechanisms for loss of TET functionality in cancer, our data shows that inactivation of TET proteins in lung epithelial cells in combination with a dysfunctional PRC1 complex leads to extensive DNA hypermethylation of Polycomb target genes, mimicking a situation commonly observed in most human malignancies.

We detected the conversion of nontumorigenic bronchial cells to transformed cells that show anchorage-independent growth and form tumors in immune-deficient mice. In previous studies, these HBEC3-KT cells could not even be transformed by simultaneous introduction of a *KRAS* mutation, *EGFR* mutation and by inhibiting the TP53 tumor suppressor (50). *TET* genes are mutated in a small percentage (<10%) of human lung tumors. A recent study observed tumor formation in a mouse model with *Kras* mutation and loss of

(*Continued.*) **E,** Metaplot illustrating DNA methylation levels at RYBP-targeted CGIs in controls and RYBP/T123 QKO cells. *P* value is indicated in **A–E**. **F,** Number of RYBP-, H3K27me3-, and H2AK119ub1-targeted CGIs that contain Hyper-DMRs and Hypo-DMRs in RYBP/T123 QKO cells relative to controls. **G,** Number of Hyper-DMRs that overlap with the indicated genomic features in RYBP/T123 QKO compared with controls. **H,** Heat map showing the percentage of all RYBP-, H3K27me3-, and H2AK119ub1-targeted CGIs that contain DMRs in RKO, TET123 TKO, and RYBP/T123 QKO cells, respectively. **I,** A total of 3666 hypermethylated CGIs from LUSC tumor samples are found in the TCGA dataset. The ring chart indicates the percentage of H2AK119ub1-targeted hypermethylated CGIs (Hyper-CGIs) in the TCGA dataset and the percentage of them that overlap with Hyper-CGIs in the RYBP/T123 QKO cells. *P* values (hypergeometric probability) are indicated.
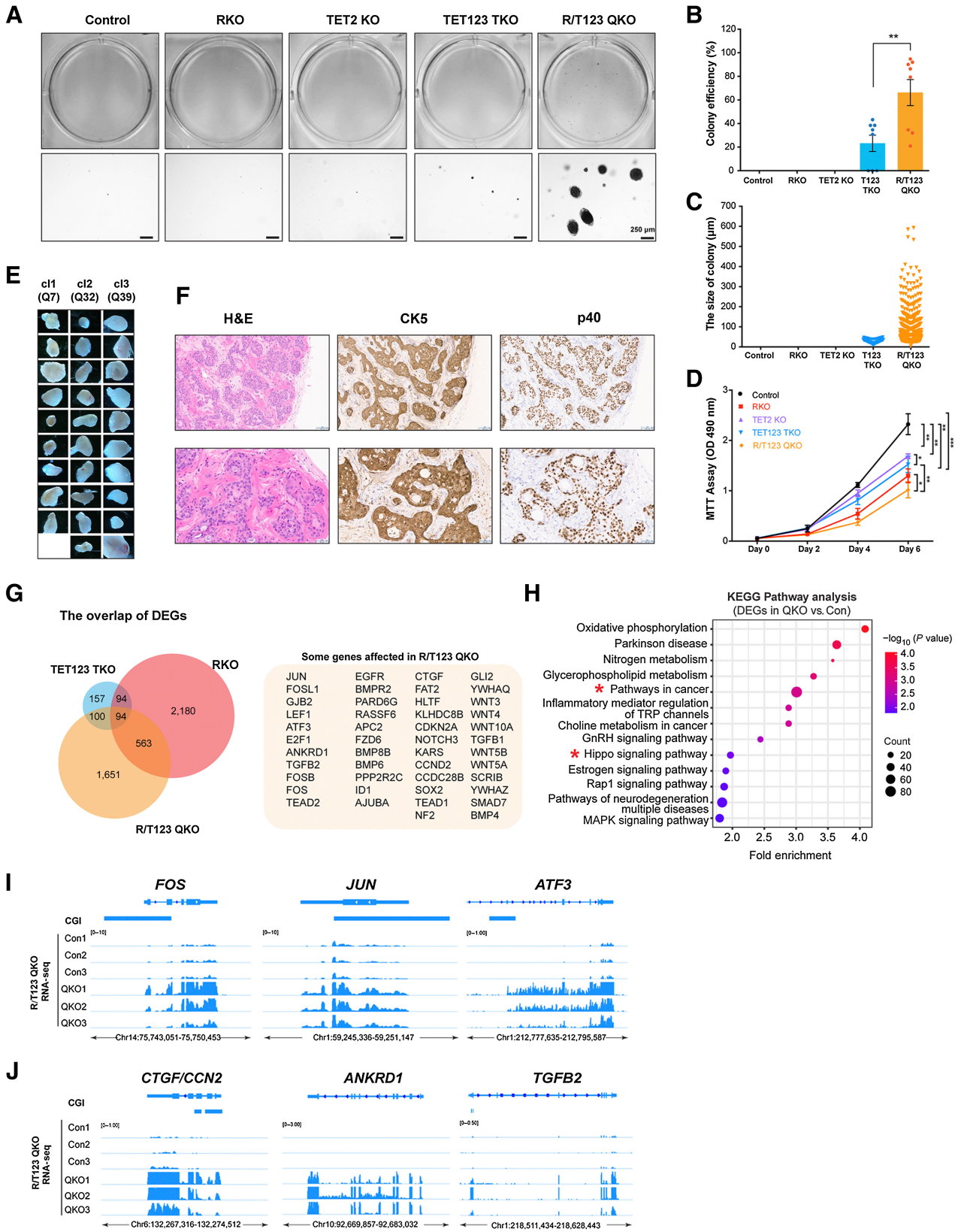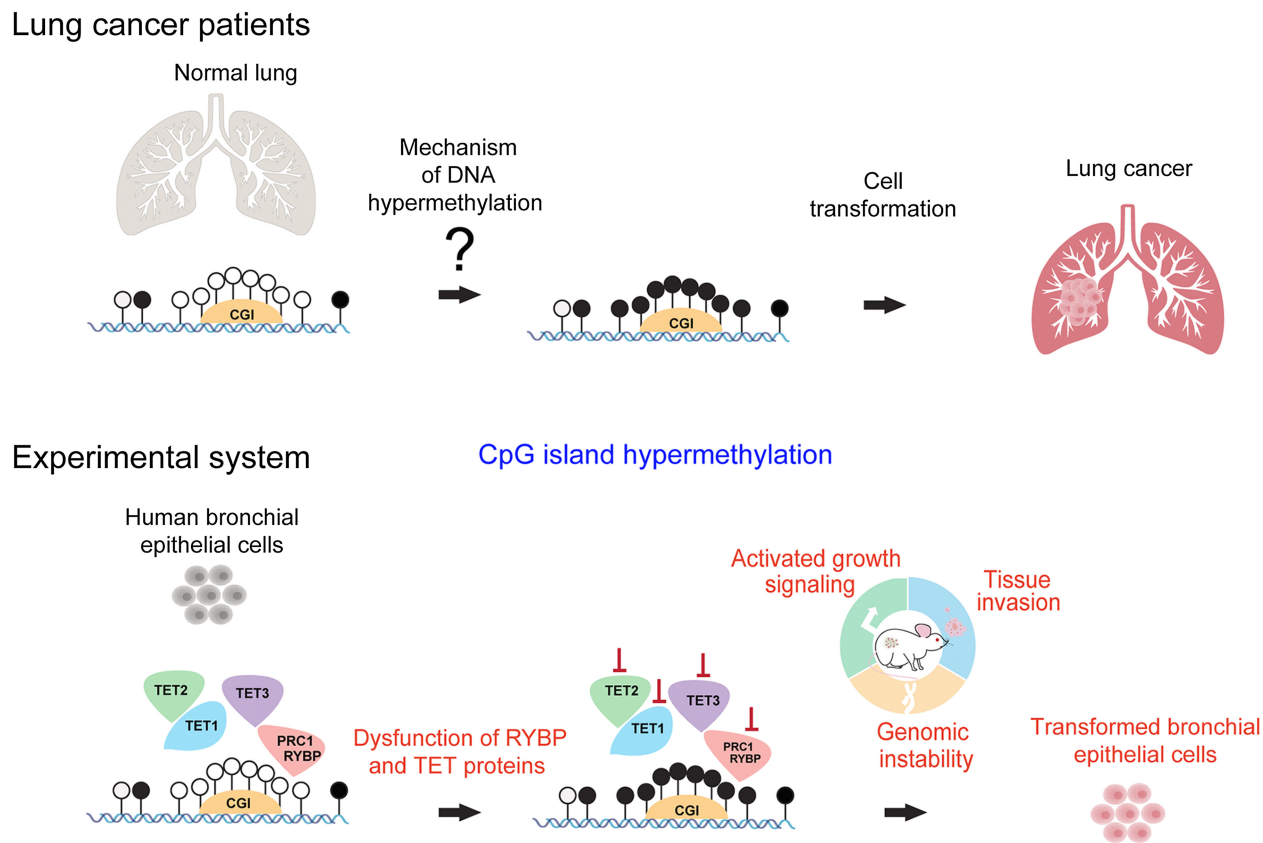
Figure 6.
Model of possible DNA hypermethylation mechanisms in cancer and lung tumor formation. The top part of the diagram shows the DNA hypermethylation events observed in patients with LUSCs and the open question of how DNA hypermethylation is initiated. The bottom part displays our experimental system of bronchial epithelial cells with dysfunction of RYBP and TET proteins leading to CGI hypermethylation and cell transformation.

Tet genes, but the induced hypermethylation events were not specifically targeted to CGIs (73). Although we observed a 1.9-fold increase of *KRAS* expression due to a copy number gain of chr12p, there were no *RAS* gene mutations, and we favor the model that the combined disturbance of several other cancer-relevant signaling pathways was due to epigenetic perturbations. We observed a reduced output of the Hippo tumor suppressor pathway as shown by strong upregulation of canonical Hippo target genes such as *CTGF* (**Fig. 5J**). Another characteristic finding in our study was the increased expression of AP-1 genes, which may be a critical event in certain types of lung cancer (74). The TEAD transcription factors

that are negatively controlled by the Hippo pathway often colocalize and collaborate with AP-1 proteins, and this combined activity drives cell proliferation (75). We found several induced CGI hypermethylation events that have the potential to be cancer drivers. These methylation target genes were positive regulators of the Hippo anti-growth pathway (*RASSF* genes and *FAT2*) that became downregulated in association with promoter methylation and several genes controlling genomic stability. The *FAT2* gene is also mutated in close to 10% of non–small cell lung cancers (COSMIC database) and this gene is highly and specifically expressed in human basal respiratory cells of the bronchi (proteinatlas.org). In lung cancers,

Figure 5.
Growth properties of bronchial cells that lack RYBP and the TET proteins. **A,** Top, soft agar assays to assess the colony forming potential of controls, RKO, TET2 KO, TET123 TKO, and RYBP/TET123 (R/T123) QKO cells. Bottom, representative images of soft agar colonies are shown. Three replicates, three independent experiments, and three independent cell clones were used. Scale bars, 250 μm. **B** and **C,** Efficiency and size of colony formation as obtained in **A**. Data represent mean $\pm$ SE. **$P < 0.01$. **D,** Cell proliferation curves of controls, RKO, TET2 KO, TET123 TKO, and R/TET123 QKO cells. Three independent clones were tested in five replicates and two independent experiments. Data represent mean $\pm$ SD. *, $P < 10e^{-3}$–$10e^{-10}$; **, $P < 10e^{-10}$–$10e^{-15}$; ***, $P < 0.10e^{-15}$–$10e^{-25}$. **E,** Images of xenografts from mice injected with RYBP/T123 QKO cells. Ten or eleven injection sites were used for each cell clone. Magnification, ×1. **F,** Representative histology of xenografts obtained from mice injected with RYBP/T123 QKO cells (top row, clone Q39; bottom row, clone Q7). Parallel sections were stained from each xenograft. Representative sections were stained with H&E, anti-CK5, and anti-p40. Scale bars, 100 μm, top row; 50 μm, bottom row. **G,** Overlap of DEGs in RKO, TET123 TKO, and RYBP/TET123 (R/T123) QKO cells. The combined analysis of three independent clones for each genotype is presented. Some examples of DEGs in the QKO vs. controls are listed. **H,** KEGG pathway analysis of all DEGs in R/T123 QKO cells. **I,** IGV browser views of the *FOS*, *JUN*, and *ATF3* genes for RNA-seq data acquired from controls and R/T123 QKO cells. **J,** IGV browser views of the Hippo pathway-controlled genes *CTGF/CCN2*, *ANKRD1*, and *TGFB2* for RNA-seq data acquired from controls and R/T123 QKO cells.

expression of *FAT2* is very heterogenous between patients with many having close to zero expression. It would be difficult to dissect which one of the multiple DNA hypermethylation events is the most consequential one for tumor formation because they co-occur upon misregulation of the two epigenetic pathways.

In summary, based on our findings in lung epithelial cells, a model can be proposed for a mechanism leading to DNA hypermethylation in lung cancer, as presented in **Fig. 6**. Further studies will test whether these mechanisms can explain the initiation or progression of other human cancers based on epigenetic alterations.

Authors' Disclosures

G.P. Pfeifer reports grants from NIH/NCI during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

W. Cui: Data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **Z. Huang:** Data curation, formal analysis, writing–review and editing. **S.-G. Jin:** Resources, investigation. **J. Johnson:** Resources, investigation. **K.H. Lau:** Formal analysis. **G. Hostetter:** Data curation, formal analysis, investigation. **G.P. Pfeifer:** Conceptualization, formal analysis, supervision, funding acquisition, visualization, writing–original draft, project administration, writing–review and editing.

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

References

1. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov 2022;12:31–46.
2. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301:89–92.
3. Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res 1986;46:2917–22.
4. Issa JP. CpG island methylator phenotype in cancer. Nat Rev Cancer 2004;4: 988–93.
5. Baylin SB, Jones PA. Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol 2016;8:a019505.
6. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 2007;39:237–42.
7. Rauch T, Li H, Wu X, Pfeifer GP. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res 2006;66:7939–47.
8. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc Natl Acad Sci USA 2007;104: 5527–32.
9. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 2007;39:232–6.
10. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, et al. Epigenetic stem cell signature in cancer. Nat Genet 2007;39:157–8.
11. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. Nat Rev Genet 2016;17:630–41.
12. Edwards JR, Yarychkivska O, Boulard M, Bestor TH. DNA methylation and DNA methyltransferases. Epig Chrom 2017;10:23.
13. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 2017;18:517–34.
14. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009;324:930–5.
15. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 2011; 333:1303–7.
16. Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 2009;114:144–7.
17. Yang L, Rau R, Goodell MA. DNMT3A in haematological malignancies. Nat Rev Cancer 2015;15:152–65.
18. Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, et al. 5-Hydroxy-methylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res 2011;71:7360–5.
19. Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PLoS One 2015;10:e0124633.
20. Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. Epigenetic inactivation of a RAS association domain family protein from the lung tumor suppressor locus 3p21.3. Nat Genet 2000;25:315–9.
21. Huang Z, Yu J, Cui W, Johnson BK, Kim K, Pfeifer GP. The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. Sci Adv 2021;7: eabb9149.
22. 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.
23. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 2010;26: 2204–7.
24. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178–92.
25. Akalin A, Franke V, Vlahovicek K, Mason CE, Schubeler D. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics 2015;31:1127–9.
26. Gu Z, Eils R, Schlesner M, Ishaque N. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics 2018;19:234.
27. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics 2011;27:1571–2.
28. Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA. Detection and accurate false discovery rate control of differentially methylated regions from whole-genome bisulfite sequencing. Biostatistics 2019;20:367–83.
29. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst 2016;12:477–9.
30. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison, and visualization. Bioinformatics 2015;31: 2382–3.
31. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2021;2:100141.
32. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinf 2017;18:287.

33. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics 2014;30:2811–2.

34. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: genomic visualizations in R. Bioinformatics 2016;32:3012–4.

35. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13.

36. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.

37. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363–9.

38. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation Beadarrays. Nucleic Acids Res 2013;41:e90.

39. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 2014;15:503.

40. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol 2012;41:200–9.

41. Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biol 2021;22:323.

42. Sanchez-Vega F, Gotea V, Margolin G, Elnitski L. Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. Epigenetics Chromatin 2015;8:14.

43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

44. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 2017;33:1179–86.

45. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with bioconductor. Nat Methods 2020;17:137–45.

46. Piunti A, Shilatifard A. The roles of Polycomb repressive complexes in mammalian development and cancer. Nat Rev Mol Cell Biol 2021;22:326–45.

47. Blackledge NP, Klose RJ. The molecular principles of gene regulation by Polycomb repressive complexes. Nat Rev Mol Cell Biol 2021;22:815–33.

48. Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci USA 2008;105:252–7.

49. Ramirez RD, Sheridan S, Girard L, Sato M, Kim Y, Pollack J, et al. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. Cancer Res 2004;64:9027–34.

50. Sato M, Vaughan MB, Girard L, Peyton M, Lee W, Shames DS, et al. Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. Cancer Res 2006;66:2116–28.

51. Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell 2014;56:286–97.

52. Jin SG, Zhang ZM, Dunwell TL, Harter MR, Wu X, Johnson J, et al. Tet3 reads 5-Carboxylcytosine through Its CXXC domain and is a potential guardian against neurodegeneration. Cell Rep 2016;14:493–505.

53. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 2011;473:343–8.

54. Lopez-Moyado IF, Tsagaratou A, Yuita H, Seo H, Delatte B, Heinz S, et al. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. Proc Natl Acad Sci USA 2019;116:16933–42.

55. Hahn MA, Hahn T, Lee DH, Esworthy RS, Kim BW, Riggs AD, et al. Methylation of Polycomb target genes in intestinal cancer is mediated by inflammation. Cancer Res 2008;68:10280–9.

56. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, et al. Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 2010; 20:332–40.

57. Nejman D, Straussman R, Steinfeld I, Ruvolo M, Roberts D, Yakhini Z, et al. Molecular rules governing de novo methylation in cancer. Cancer Res 2014;74: 1475–83.

58. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, et al. Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol 2016;17:205.

59. Klutstein M, Moss J, Kaplan T, Cedar H. Contribution of epigenetic mechanisms to variation in cancer risk among tissues. Proc Natl Acad Sci USA 2017;114: 2230–4.

60. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, Maekita T, et al. Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. Cancer Res 2010;70: 1430–40.

61. Takeshima H, Ushijima T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. NPJ Precis Oncol 2019;3:7.

62. Vaz M, Hwang SY, Kagiampakis I, Phallen J, Patil A, O'Hagan HM, et al. Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations. Cancer Cell 2017;32:360–76.

63. Gridnev A, Misra JR. Emerging mechanisms of growth and patterning regulation by dachsous and Fat protocadherins. Front Cell Dev Biol 2022; 10:842593.

64. van Toorn M, Turkyilmaz Y, Han S, Zhou D, Kim HS, Salas-Armenteros I, et al. Active DNA damage eviction by HLTF stimulates nucleotide excision repair. Mol Cell 2022;82:1343–58.

65. Krem MM, Luo P, Ing BI, Horwitz MS. The kelch protein KLHDC8B guards against mitotic errors, centrosomal amplification, and chromosomal instability. J Biol Chem 2012;287:39083–93.

66. Hanley MP, Hahn MA, Li AX, Wu X, Lin J, Wang J, et al. Genome-wide DNA methylation profiling reveals cancer-associated changes within early colonic neoplasia. Oncogene 2017;36:5035–44.

67. Hu X, Estecio MR, Chen R, Reuben A, Wang L, Fujimoto J, et al. Evolution of DNA methylome from precancerous lesions to invasive lung adenocarcinomas. Nat Commun 2021;12:687.

68. Gu T, Hao D, Woo J, Huang TW, Guo L, Lin X, et al. The disordered N-terminal domain of DNMT3A recognizes H2AK119ub and is required for postnatal development. Nat Genet 2022;54:625–36.

69. Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell 2014;29:102–11.

70. Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 2013;497:122–6.

71. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 2009; 462:739–44.

72. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, et al. Tumor hypoxia causes DNA hypermethylation by reducing TET activity. Nature 2016;537:63–8.

73. Xu Q, Wang C, Zhou JX, Xu ZM, Gao J, Sui P, et al. Loss of TET reprograms Wnt signaling through impaired demethylation to promote lung cancer development. Proc Natl Acad Sci USA 2022;119:e2107599119.

74. Kim J, Minna JD. AP-1 leads the way in lung cancer transformation. Dev Cell 2022;57:292–4.

75. Tang F, Xu D, Wang S, Wong CK, Martinez-Fundichely A, Lee CJ, et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. Science 2022;376:eabe1505.