Supplementary Methods

Genetically engineered mouse models

Human EZH2 cDNA was cloned into transgenic targeting vectors and co-electroporated into v6.5 C57BL/6(F) × 129/sv(M) embryonic stem cells (Open Biosystems, #MES1402) with plasmid expressing FLP recombinase as described (1). Embryonic stem cells were screened for integration of the transgene by PCR. Correctly targeted embryonic stem cells were injected into Black 6 blastocysts, and the resulting chimeras were bred with BALB/c strain wildtype mice for germline transmission of the transgenes. EZH2 transgenic mice were crossed with Ubiquitin-Cre-ERT2 (2) or Actin-Cre mice (Jackson Laboratories). Mice were maintained in a mixed background strain, and bitransgenic mice were given 2 mg of 4-hydroxytamoxifen (4-OHT) (Sigma, #1054029-01) at 6 weeks of age for 4 consecutive days. All mice were housed in pathogen-free animal facilities, and all experiments were performed with the approval of the Animal Care and Use Committee at Harvard Medical School and Dana-Farber Cancer Institute.

Immunohistochemistry

Tissues were fixed overnight in 10% buffered formalin and paraffin-embedded (FFPE) overnight to two days. Staining was performed as previously described (3).

Histone extraction

Murine normal or tumor lung tissues were homogenized in TEB buffer (0.5% Triton X-100, 2 mM PMSF, 0.02% NaN_3 in PBS) at 200 mg/mL. Homogenates were centrifuged at 3000 rpm for 5 min at 4°C. Pellets were resuspended in 3 volumes of extraction buffer (0.5 N HCl, 10% glycerol) and incubated for 30 min on ice. Mixtures were centrifuged at 12,000 rpm for 5 min at 4°C, and supernatants were removed and mixed with 8 volumes of acetone and incubated overnight at -20°C. Samples were centrifuged at 12,000 rpm for 5 min, and resulting histone pellets were air-dried. Isolated histones were dissolved in distilled water, quantified, and subjected to Western blotting with H3 (Abcam, #AB1791), H3-3K4 (Millipore, #07-473), H3-3K27 (Millipore, #07-449), and H3-3K36 (Abcam, #AB9050) antibodies per manufacturers' recommendations. Relative quantifications were performed using ImageJ software.

Soft agar assay

Agar was made by heating 5% noble agar (Sigma A5431) in sterile water and diluting the dissolved solution to 0.5% in warm media. Two milliliters of agar was added to the bottom each well of a 6-well plate. After solidified, 40,000 cells in complete growth media were mixed with a final concentration of 0.35% agar and evenly seeded into each well. After solidified, plates were placed in an incubator, and cells were allowed to grow for 2–4 weeks.

Crystal violet staining

Cells were incubated with compounds for 7 days at which time the media was aspirated and the cells were fixed with formalin for 20min. The plates were washed with PBS and

stained with crystal violet (10% ethanol, 0.1% crystal violet) for 15-20 min. Plates were washed with water and dried. Crystal violet dye was extracted using 10% acetic acid and absorbance was determined at 590nm.

ChIP-Seq alignment

All ChIP-Seq datasets were aligned using Bowtie2 (version 2.2.1) to build version NCB37/MM9 of the mouse genome (4). Alignments were performed using default parameters that preserved reads mapping uniquely to the genome without mismatches. ChIP-Seq raw reads and alignment statistics reported in Supplementary Table S3.

Identifying ChIP-Seq enriched regions

We used the MACS version 1.4.2 (Model based analysis of ChIP-Seq) peak finding algorithm to identify regions of ChIP-Seq enrichment over background (5). A p-value threshold of enrichment of 1e-9 was used for all datasets.

Mapping enhancers and super enhancers

Enhancers and super enhancers were mapped using the ROSE software package available at (younglab.wi.mit.edu/super_enhancer_code.html) (6, 7). ChIP-Seq enriched regions for H3K27ac in WT Lung, EZH2 OE Normal Lung, EZH2 OE Tumor Lung, KRas Mutant Lung were used to map enhancers and super-enhancers. Super enhancer associated genes were designated as transcriptionally active genes within 50kb from a super enhancer. Active genes were defined as genes containing H3K27ac enrichment +/- 1kb from the TSS and detectable gene expression (> 1 FPKM).

Identifying pairwise gained/lost enhancer regions

Differential super enhancer regions between Mutant and Normal samples were determined by comparing background subtracted H3K27ac signal at the set of all enhancer regions considered ‘super’ in at least one condition. Gained/lost super enhancers were determined as those with a greater than 2 fold change in either direction.

Identifying regions with increased H3K27me3

To identify changes associated with increased H3K27me3, we used the set of H3K27ac super enhancer regions as indicative of active chromatin. The ratio of H3K27ac over H3K27me3 was calculated at super enhancer containing regions found in either WT or Tumor H3K27ac. Regions with a Log_2 ratio less than 0 were considered regions with strong H3K27me3 gain.

Identifying candidate transcription factor binding regions

Regions between peaks of H3K27ac have been shown to harbor transcription factor binding sites (8). To identify these regions, we first defined the valleys in each enhancer region (9). 10 bp bins were assigned a valley score, then adjacent bins were stitched and their sequence was extended 100 bp on either side to search for transcription factor binding motifs.

Building core regulatory circuits

To identify core regulatory elements within each condition, motif enrichment for active transcription factors was determined at condition specific super enhancer valleys (as described previously) using Find Original Motif Occurrences (FIMO) (10). Enrichment was determined against a collection of super enhancer valleys found in the *EZH2_OE*, *KRAS*, and *WT* datasets. Using active, FIMO enriched transcription factor motifs, edges were drawn between known binding partners. Each transcription factor was assigned a score based on inward (number of incoming edges) and outward binding (number of outward edges). Cliques were determined among all possible binding circuits and ranked by constituent TF inward and outward binding score. TFs were then compared between tumor samples and wild-type to determine unique changes in the Super enhancer driven transcription factor networks.

RNA-Seq

Total RNA was extracted using Trizol (Invitrogen) followed by RNA cleanup (Qiagen, #74204). mRNA-Seq libraries were constructed using 1µg of total RNA with the mRNA-Seq Kit (Cat# RS-122-2001). The samples were multiplexed per at 3 per lane and 50-cycle paired-end sequencing runs were generated with an Illumina HiSeq2500. QC-passed reads were aligned to the mouse reference genome using MapSplice (11). The alignment profile was determined by Picard Tools v1.64. Aligned reads were sorted and indexed using SAMtools and translated to transcriptome coordinates then filtered for indels, large inserts, and zero mapping quality using UBU v1.0. Transcript abundance estimates for each sample were performed using RSEM, an expectation-maximization algorithm using the UCSC known Gene transcript and gene definitions (12). Raw RSEM read counts for all RNAseq samples were normalized to the overall upper quartile (13).

RNA-Seq data analysis

Paired-end fastq files were aligned to the mm9 genome assembly using HiSat2 with default parameters. RefSeq annotated transcripts were assembled and FPKM values were generated using cuff quant and cuffnorm from the cufflinks pipeline (14). Active transcripts were defined as transcripts with a normalized FPKM value greater than 0.1.

Single sample gene set enrichment analysis (ssGSEA)

ssGSEA projection was performed according to the methods described by Barbie, DA *et al*., 2009. The GenePattern portal of the Broad Institute was used to perform the analysis. The following standard parameters were used: sample normalization method "Rank"; weighing exponent of 0.75; min gene set size of 10. The expression profiles of TCGA lung adenocarcinoma tumors and normal lung were used to perform ssGSEA. For each tumor, the driver mutation was determined based on TCGA mutation analysis. All ALK and NF1 mutant tumors were removed from the analysis. EZH2 high tumors were defined as the top 20% of tumors with highest EZH2 expression and lacking any mutation in KRAS, EGFR, ALK or NF1. The KRAS- and EGFR-mutant categories were defined as being mutant in KRAS or EGFR and having low EZH2 expression (tumors with 20% lowest EZH2 expression). Analysis was performed with the downstream RAS

pathway signatures (gene lists) present in the C6 oncogenic signatures collection (MSigDB) of the Broad Institute (MEK: MEK_UP.V1_UP and mTOR: mTOR_UP.N4.V1_DN) and our own gene list of H3K27Me3-bound genes.

Surface plasmon resonance

SPR studies were performed using a Biacore™ T100 instrument (GE Health Sciences Inc.). Biotinylated JQEZ5 (JQEZ6) was immobilized to SA chip (~8.3 RU) by injecting 800 nM probe dissolved in 1X PBS-P+ buffer (GE Health Sciences Inc.) at a 10 uL/min flow rate for 20s. Five-component PRC2 was diluted in buffer (25 mM TRIS pH 7.9, 50 mM NaCl, and 0.01% Tween-20) at various concentrations and kinetic analysis was performed with a 120 s contact time, and a wash time of 1200s. Curve fitting and *K*D determinations were performed with the Biacore T200 Evalutation software (GE Health Sciences Inc.). Four concentrations of PRC2 were tested in twofold serial dilution experiments in duplicate.

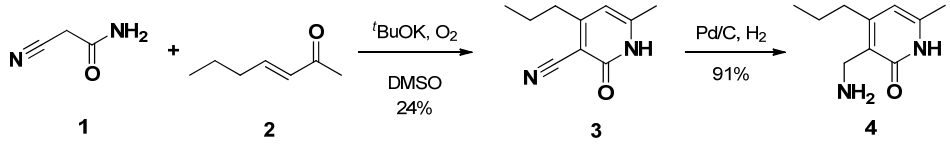
Docking and ligand-interaction determination

All structural modeling was conducted using the Schrodinger computational suite (update 8/2013) through the SBgrid. Ligands of interest were imported into Maestro and an exhaustive conformational search was preformed through Monte-Carlo simulation using the MM5 force field. The resulting conformations were clustered by 3D similarity and an exemplar structure was chosen from each structure for subsequent docking. The docking was performed on a previously reported published binding model with the ligand-binding site used to define the receptor grid (15). The protein was preprocessed with Maestro and grid generation and docking was preformed with Glide using standard input parameters. All docking was performed using the XP level of precision and results were ranked using the Glide Score. Ligand-interaction diagrams were generated with Maestro for the top docking poses.

Synthesis of small molecule inhibitors

Reactions were run as described in the individual procedures using standard double manifold and syringe techniques; glassware was dried by baking in an oven at 130 °C for 12h prior to use. The flasks were fitted with rubber septa and reactions were conducted under a positive pressure of nitrogen. Solvents for reactions were purchased anhydrous from Sigma-Aldrich and used as received; the only exception being EtOH, which was stored over 4 Å molecular sieves. HPLC grade solvents were used for aqueous work ups and chromatography. Reagents were used as received. Flash column chromatography was performed as described by Still *et al.* using silica gel (60 Å pore size, 40-63 μm, 4-6% H_2O content, Zeochem). Thin layer chromatography plates were visualized by exposure to iodine vapor. All the intermediates and final product were fully characterized with proton and carbon-13 nuclear magnetic resonance (^{1}H NMR and ^{13}C NMR) spectra and mass spectra (MS). Reactions were monitored by thin-layer chromatography using EMD silica gel 60 F254 (250-micron) glass-backed plates (visualized by UV fluorescence quenching and staining with KMnO4) and by LC-MS using a Waters Aquity BEH C18 2 x 50 mm 1.7 μm particle column (50 °C) eluting at 1 mL/min with H2O/acetonitrile [0.2% v/v added formic acid or concentrated NH_4OH(aq)

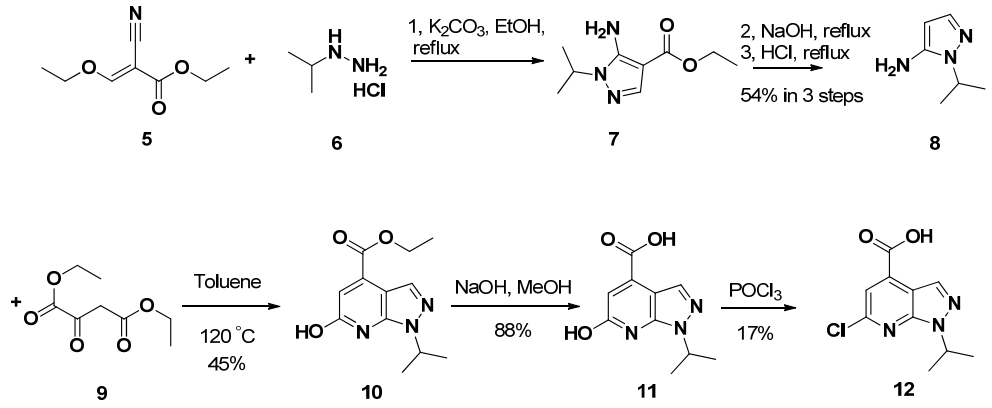
solution; 95:5(0min)→1:99(3.60min)→ 1:99(4.00min)] using alternating positive/negative electrospray ionization (125-1000 amu) and UV detection (210-350 nm). Flash column chromatography was carried out using Merck grade 9385 silica gel 60 Å pore size (230-400 mesh). Melting points were obtained using a capillary melting point apparatus and are uncorrected. 1H NMR spectra were recorded at 400 MHz on a Bruker spectrometer and are reported in ppm using the residual solvent signal (dimethylsulfoxide-d6 = 2.50 ppm; chloroform-d = 7.27 ppm; methanol-d4 = 3.31 ppm; dichloromethane-d2 = 5.32 ppm) as an internal standard. Data are reported as: {(δ shift), [(s = singlet, d = doublet, dd, doublet of doublets, ddd = doublet of a dd, t = triplet, quin = quintet, sept = septet, br = broad, ap = apparent), (J = coupling constant in Hz) and (integration)]}. Proton-decoupled 13C NMR specta were recorded at 100 MHz on a Bruker spectrometer and are reported in ppm using the residual solvent signal (chloroform-d = 77.0 ppm; dimethylsulfoxide-d6 = 39.51 ppm; methanol-d4 = 49.15 ppm) as an internal standard. Infrared spectra were recorded using an ATR-FTIR instrument. High resolution mass spectra were acquired by flow injection on a qTOF Premiere Mass Spectrometer operating in ES+ ionization with resolution ~15,000.



Scheme 1. Synthesis of intermediate **4**.

A mixture of potassium *t*-butyloxide (BuOK, 4 g, 35.7 mmol), 2-cyanoacetamide (**1**) (3.3 g, 39.2 mmol) and 3-hepten-2-one (**2**) (4 g, 35.7 mmol) in DMSO (60 mL) was stirred at room temperature for 30 min. Then additional *t*-BuOK (12 g, 107 mmol) was added and the reaction mixture was stirred under an atmosphere of oxygen for 1 h. The reaction mixture was purged with nitrogen, and diluted slowly with water (250 mL) and *aq.* HCl (4 N, 300 mL). The reaction mixture was filtered to collect the yellow precipitate, which was washed with water and dried to give 1.5 g of **3** (24% yield) as a yellow solid. MS: m/z 177 $(M+H)^+$.

A mixture of the above **3** (1.5 g) in THF (20 mL) was added Pd/C (10%, 1.5 g) and *conc.* HCl (1 mL). The mixture was stirred at room temperature overnight under an atmosphere of hydrogen. The mixture was filtered through Celite and the filtration was concentrated in vacuo. the residue was recrystallized from PE/EA (v/v = 10) to give 1.4 g of **4** (91% yield) as a tan-yellow solid. MS: m/z 181.1 $(M+H)^+$; 164.1 $(M-NH_3+H)^+$; ^{1}H NMR (500 MHz, $DMSO_{d6}$) δ 11.87 (br s, 1H), 8.12 (s, 3H), 6.49 (br s, 1H), 6.00 (s, 1H), 3.78 (m, 2H), 2.18 (s, 3H), 1.50 (m, 2H), 0.92 (t, *J* = 7.0 Hz, 3H) ppm.



Scheme 2. Synthesis of intermediate **12**

A mixture of ethyl (ethoxymethylene)cyanoacetate (**5**) (5 g, 29.6 mmol), isopropylhydrazine hydrochloride (**6**) (3.9 g, 35.5 mmol) and potassium carbonate (8.2 g, 59.2 mmol) in ethanol (100 mL) was refluxed for 16 h. The volatiles were removed in vacuo to give crude **7** as a yellow solid containing inorganic salt, which was used for the next step without further purification. MS: m/z 198.1 $(M+H)^+$.

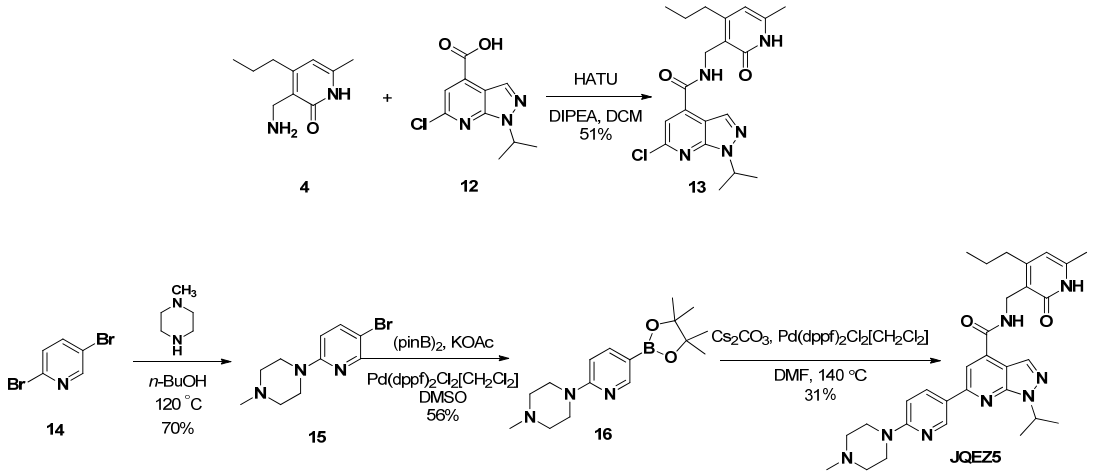
A suspension of the above crude **7** in *aq.* sodium hydroxide (4 N, 50 mL) was refluxed for 16 h. The mixture was cooled and acidified with *conc.* HCl to pH ~ 3.5. HCl in dioxane (4 N, 2 mL to pH < 1) was added to the reaction mixture and was refluxed for 16 h. The organic layer was separated off and the aqueous solution was neutralized with *aq.* sodium hydroxide (4 N, to pH > 10). Then the mixture was extracted with methylene chloride. The combined organic solution was washed with brine, dried over sodium sulfate and concentrated to give about 2 g of **8** as orange oil (54% yield in 3 steps), and was used for the next step without further purification. MS: m/z 126.1 $(M+H)^+$.

A mixture of **8** (1.00 g, 7.99 mmol) and diethyl oxalacetate (**9**) (2.26 g, 12.0 mmol) in toluene (20 mL) was refluxed for 16 h. The volatiles were removed *in vacuo* and the residue was dissolved in acetic acid (10 mL) and refluxed for 4 hours. Then the mixture was diluted with water (20 mL) and extracted with ethyl acetate. The combined organic solution was washed with water and brine, dried over sodium sulfate and concentrated. The crude product was recrystallized from methylene chloride to give 0.9 g of **10** (45% yield) as a white solid. MS: m/z 250 $(M+H)^+$.

A solution of **10** (0.90 g, 3.61 mmol) in THF (10 mL) was added *aq.* sodium hydroxide (4 N, 5 mL), and was stirred at room temperature overnight. The resulting mixture was then acidified with *conc.* HCl to pH ~ 2 and was extracted with methylene chloride. The organic solution was washed with brine, dried over sodium sulfate and concentrated in vacuo. The crude product was recrystallized from methylene chloride to give 0.7 g of **11** (88% yield) as a yellow solid. MS: m/z 222 $(M+H)^+$.

A mixture of **11** (600 mg, 2.71 mmol) in phosphorus oxychloride (15 mL) was stirred at 120 ^oC in a sealed tube overnight. Most of the phosphorus oxychloride was removed in vacuo and the residue was quenched with water at 0 ^oC. The mixture was extracted

with methylene chloride. The combined organic solution was washed with water, dried over sodium sulfate and concentrated in vacuo. The residue was purified by silica gel column chromatography (PE/EA, v/v = 1) to give 110 mg of 12 (17% yield) as a white solid. MS: m/z 240 $(M+H)^+$. MS: m/z 240.0 $(M+H)^+$; ^1H NMR (500 MHz, $CDCl_3$) δ 8.51 (s, 1H), 7.83 (s, 1H), 5.34 (m, 1H), 1.64 (d, *J* = 6.5 Hz, 6H) ppm.



Scheme 3. Synthesis of compound **JQEZ5**

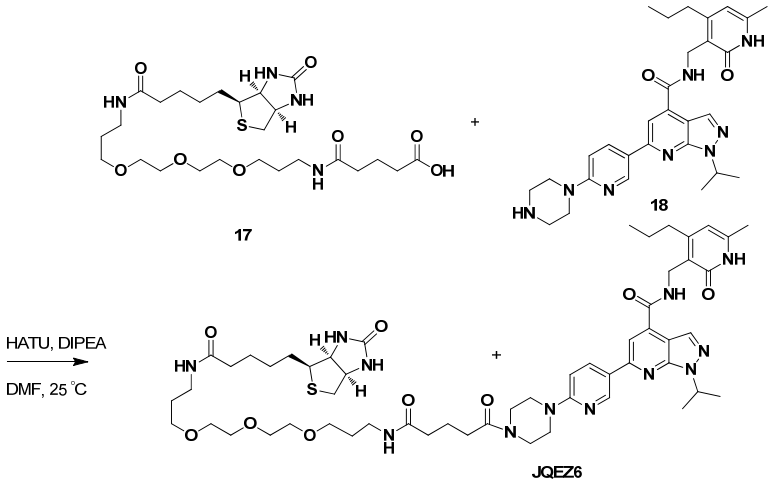
A mixture of **12** (70 mg, 0.29 mmol), HATU (277 mg, 0.73 mmol) and diethylpropyl ethyl amine (DIPEA) (1 mL) in methylene chloride (10 mL) was stirred for 10 min and then was added **4** (127 mg, 0.58 mmol). The reaction mixture was stirred at room temperature for 3 hours. The mixture was diluted with methylene chloride (50 mL) and washed with water and brine, dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by flash chromatography (50% EA in PE) to give 60 mg of **13** (51% yield) as a white solid. MS: m/z 402.1 $(M+H)^+$; ^1H NMR (500 MHz, $CDCl_3$) δ 11.91 (br s, 1H), 8.38 (s, 1H), 8.24 (t, *J* = 5.0 Hz, 1H), 7.43 (s, 1H), 6.00 (s, 1H), 5.26 (m, 1H), 4.65 (d, *J* = 5.5 Hz, 2H), 2.69 (t, *J* = 7.5 Hz, 2H), 2.28 (s, 3H), 1.64 (m, 2H), 1.55 (d, *J* = 6.5 Hz, 6H), 1.02 (t, *J* = 7.0 Hz, 3H) ppm.

A mixture of **14** (2.37 g, 10 mmol) and *N*-methylpiperazine (4 g, 40 mmol) in *n*-BuOH (25 mL) was refluxed for 96 h. The volatiles were removed *in vacuo* and the crude product was purified by silica gel column chromatography (PE/EA, v/v = 3 to EA) to give 1.8 g of **15** (70% yield) as a yellow semi-solid. MS: m/z 256.0 $(M+H)^+$, 258.0 $(M+H, Br)^+$.

A mixture of **15** (1.5 g, 5.9 mmol), bis(pinacolato)diboron (1.6 g, 6.5 mmol), potassium acetate (1.8 g, 18 mmol) and $Pd(dppf)_2Cl_2[CH_2Cl_2]$ (0.73 g, 0.89 mmol) in DMSO (20 mL) was protected with nitrogen and stirred at 80 ^oC overnight. The mixture was diluted with water and extracted with EA. The organic solution was concentrated *in vacuo* and the residue was purified by silica gel column chromatography (PE/EA, v/v = 3 to EA) to give 1.0 g of **16** (56% yield) as a brown solid. MS: m/z 304 $(M+H)^+$.

A mixture of **13** (133 mg, 0.33 mmol), **16** (200 mg, 0.66 mmol), $Pd(dppf)_2Cl_2[CH_2Cl]_2$ (41 mg, 0.050 mmol) and Cs_2CO_3 (215 mg, 0.66 mmol) in DMF (5 mL) was protected with argon and irradiated with microwave at 140 ^oC for 30 minutes. The mixture was diluted

with ethyl acetate and filtered through Celite. The filtrate was concentrated and the residue was purified by flash chromatography and prep-HPLC (HCOOH system) to give 55 mg of **JQEZ5** (31% yield) as a dark solid. MS: m/z 543.4 $(M+H)^+$; ^1H NMR (500 MHz, $CDCl_3$) δ 9.00 (d, *J* = 2.5 Hz, 1H), 8.33 (s, 1H), 8.31 (dd, *J* = 9.0 and 2.5 Hz, 1H), 8.14 (t, *J* = 6.0 Hz, 1H), 7.87 (s, 1H), 6.75 (d, *J* = 9.0 Hz, 1H), 6.02 (s, 1H), 5.38 (m, 1H), 4.69 (d, *J* = 6.0 Hz, 2H), 3.88 (m, 4H), 2.92 (m, 4H), 2.75 (m, 2H), 2.62 (s, 3H), 2.27 (s, 3H), 1.67 (m, 2H), 1.62 (d, *J* = 6.5 Hz, 6H), 1.05 (t, *J* = 7.0 Hz, 3H) ppm.



A mixture of **17** (11.2 mg, 0.02 mmol), **16** (10.6 mg, 0.02 mmol), HATU (10 mg, 0.022 mmol) and DIPEA (25 μL) in DMF (0.5 mL) was stirred at 25 ^oC for 2 h. The mixture was directly purified by prep-HPLC (TFA system) to give 5.5 mg of **JQEZ6** (26 % yield) as colorless oil. MS: m/z 1071.6 (M+H)

Supplementary References

1. Beard C, Hochedlinger K, Plath K, Wutz A, Jaenisch R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis. 2006;44(1):23-8.
2. Ruzankina Y, Pinzon-Guzman C, Asare A, Ong T, Pontano L, Cotsarelis G, et al. Deletion of the developmentally essential gene ATR in adult mice leads to age-related phenotypes and stem cell loss. Cell Stem Cell. 2007;1(1):113-26.
3. Xu C, Fillmore CM, Koyama S, Wu H, Zhao Y, Chen Z, et al. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. Cancer Cell. 2014;25(5):590-604.
4. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
5. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
6. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-34.

7. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-19.
8. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100.
9. Ramsey SA, Knijnenburg TA, Kennedy KA, Zak DE, Gilchrist M, Gold ES, et al. Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. Bioinformatics. 2010;26(17):2071-5.
10. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017-8.
11. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010;38(18):e178.
12. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
13. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11:94.
14. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5.
15. Kalinic M, Zloh M, Eric S. Structural insights into binding of small molecule inhibitors to Enhancer of Zeste Homolog 2. J Comput Aided Mol Des. 2014;28(11):1109-28.