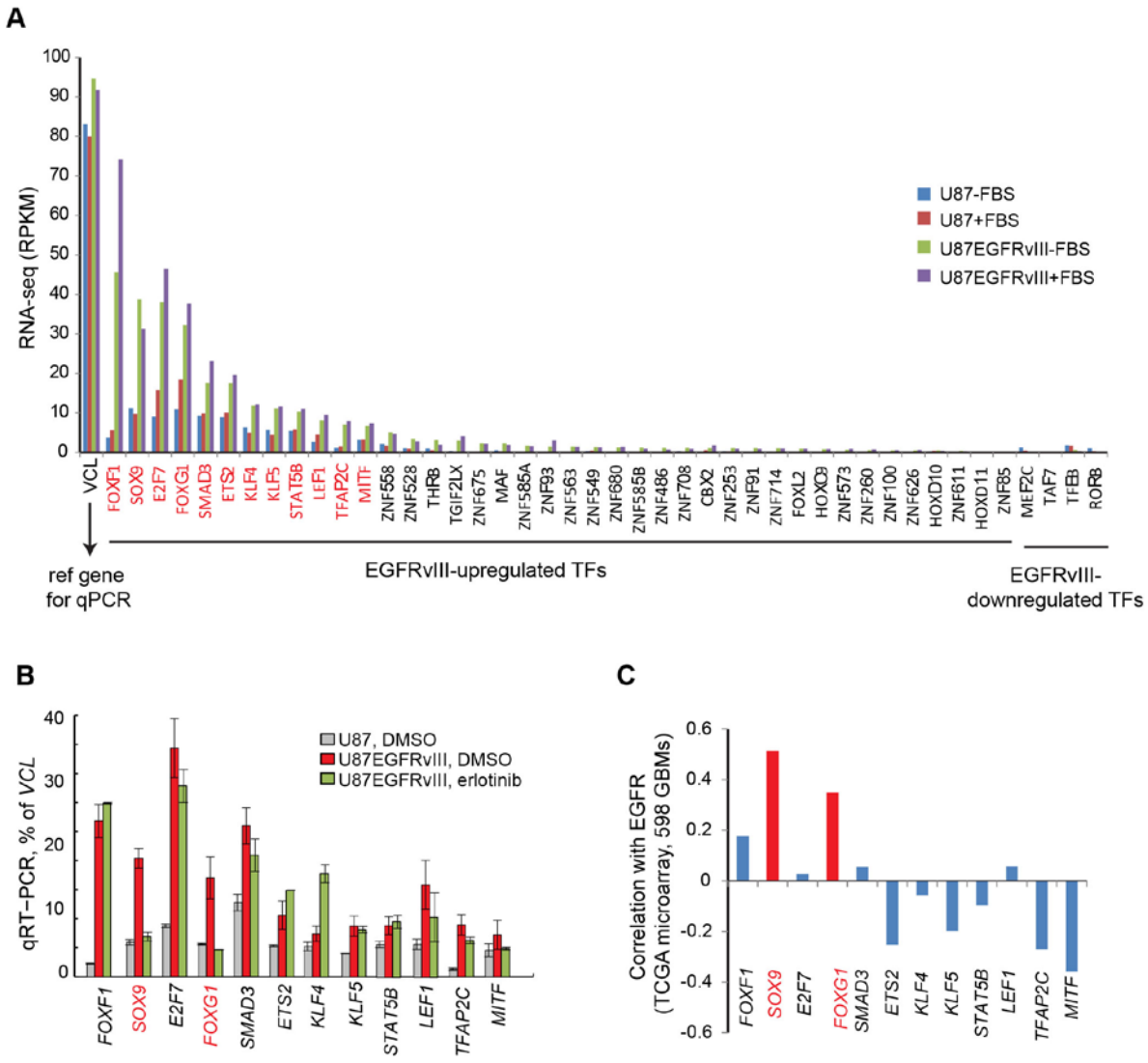

Figure S1, related to Figure 1. EGFRvIII-regulated transcription factors (TFs). **(A)** RNA-seq data of genes encoding DNA-sequence specific binding TFs. The y-axis shows reads per kilo bases per million reads (RPKM). EGFRvIII up- and down-regulated TFs are indicated below. *VINCULIN (VCL)* does not show significant differences between U87 and U87EGFRvIII cells, and is included as negative control. **(B)** Reponse of 12 EGFRvIII-activated TFs to 24 hr of 10 uM erlotinib treatment. Only the levels of *SOX9* and *FOXG1* decreased to the base levels in U87. **(C)** *SOX9* and *FOXG1* transcript levels are also most positively correlated with *EGFR* in 598 GBM tumors in the TCGA database.

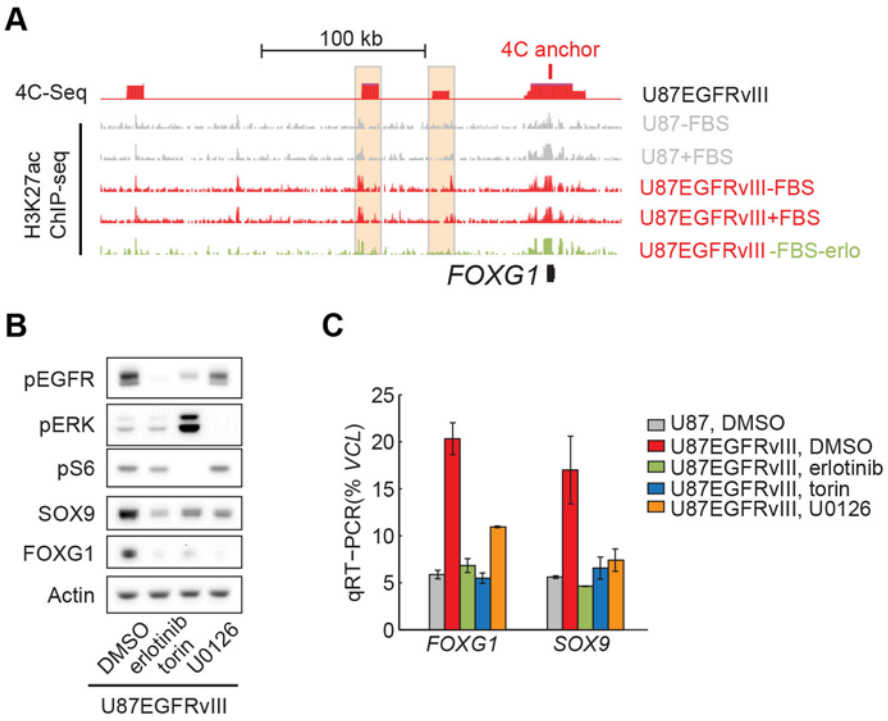

Figure S2, related to Figure 2. (A) Two EGFRvIII-responsive enhancers near the *FOXG1* locus interact with the promoter of *FOXG1*, as shown by the circularized chromosome confirmation capture assay followed by high-throughput sequencing (4C-Seq). **(B)** Western blots of cells treated by EGFR pathway inhibitors for 24 hours. Erlotinib (10 μm) is a tyrosine kinase inhibitor of EGFR/EGFRvIII. Torin (0.25 μm) is an mTOR (mTOR1/2) inhibitor. U0126 (10 μm) is a MEK inhibitor. pEGFR, pERK, and pS6 were used to monitor the effectiveness of the compounds at the doses used in the experiments. **(C)** qRT-PCR experiments indicate that these kinase inhibitors suppress SOX9 and FOXG1 at the level of transcription.

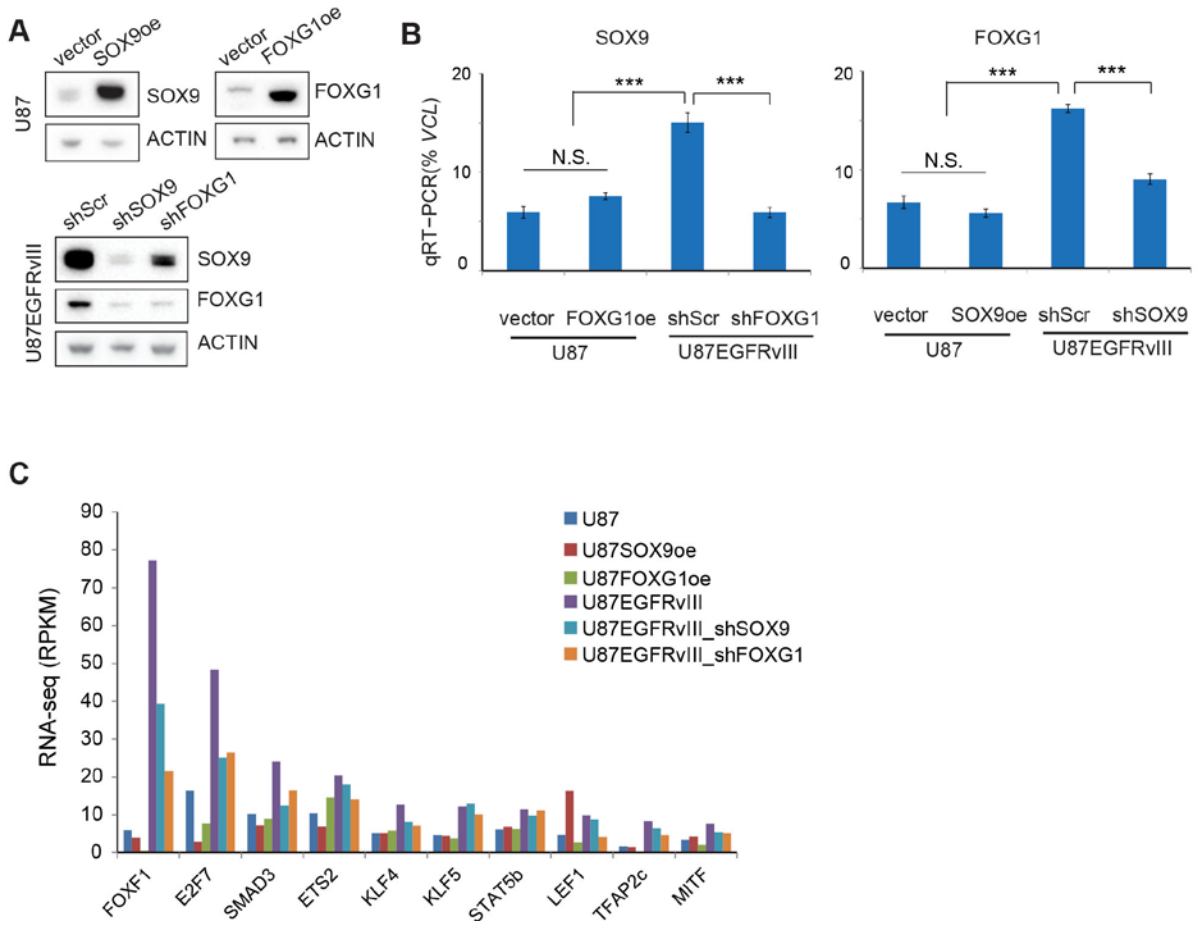

Figure S3, related to Figure 3. (A) Western blots of U87 cells over-expressing *SOX9* or *FOXG1*, or U87EGFRvIII cells expression short hairpins targeting *SOX9* and *FOXG1*. **(B)** qRT-PCR experiments indicate that overexpression of SOX9 or FOXG1 is insufficient to activate each other in U87 cells. However, in U87EGFRvIII cells, SOX9 is required for the expression of FOXG1, and vice versa. ***: $p<0.001$, *t*-test. N.S.: not significant. **(C)** RNA-seq data of the top 12 EGFRvIII-activated TFs in U87 cells overexpressing *SOX9* or *FOXG1*, as well as in U87EGFRvIII cells with *SOX9* and *FOXG1* knockdown.

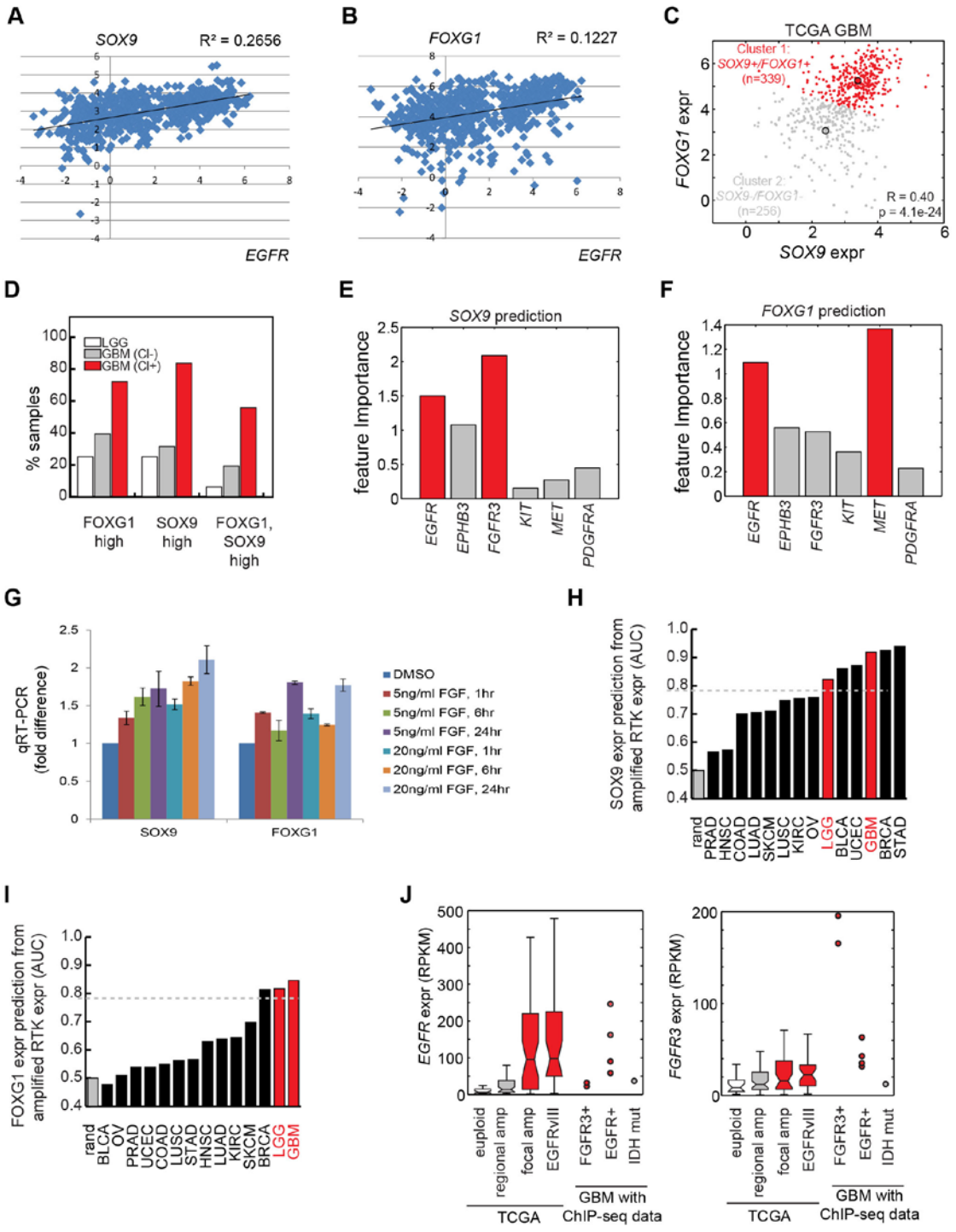

Figure S4, related to Figure 4. Correlation of *EGFR* with *SOX9* and *FOXG1* in clinical GBM samples. (**A, B**) Expression levels of *SOX9*, *FOXG1* and *EGFR* in TCGA microarray data of 598 GBM tumors. Each dot represents one tumor. Linear trend line and correlation coefficient (R^2) are shown. (**C**) *FOXG1* and *SOX9* are co-expressed at high levels in GBM. (**D**) *FOXG1* and

SOX9 are expressed at high levels more frequently in the classical subtype (Cl+) of GBM—which are characterized by EGFR amplification and/or mutation—than in non-classical (Cl-) subtype GBMs (the proneural, the neuronal, and the mesenchymal subtypes) and low grade glioma (LGG). (**E, F**) Feature importance of frequently amplified RTKs in predicting the levels of SOX9 and FOXG1 in TCGA GBM samples using a random forest classifier. (**G**) Stimulating a U87 cell line overexpressing *FGFR3* with FGF increases the level of SOX9, which is consistent with an association between FGFR3 and SOX9 in clinical GBM samples. (**H, I**) Summary of experiments using the expression levels of commonly amplified/mutated RTKs (Table S1) to predict the levels of *SOX9* and *FOXG1* in cancers by a random forest classifier. (**J**) Clinical GBM sample information. Tumors in the TCGA GBM database are grouped by the amplification/mutation status of *EGFR*. The genotypes of 7 GBM tumors used in this study for both RNA-seq and ChIP-seq (Figure 4D) are determined by PCR, FISH and ChIP-input DNA sequencing (see Table S2).

See also Table S1.

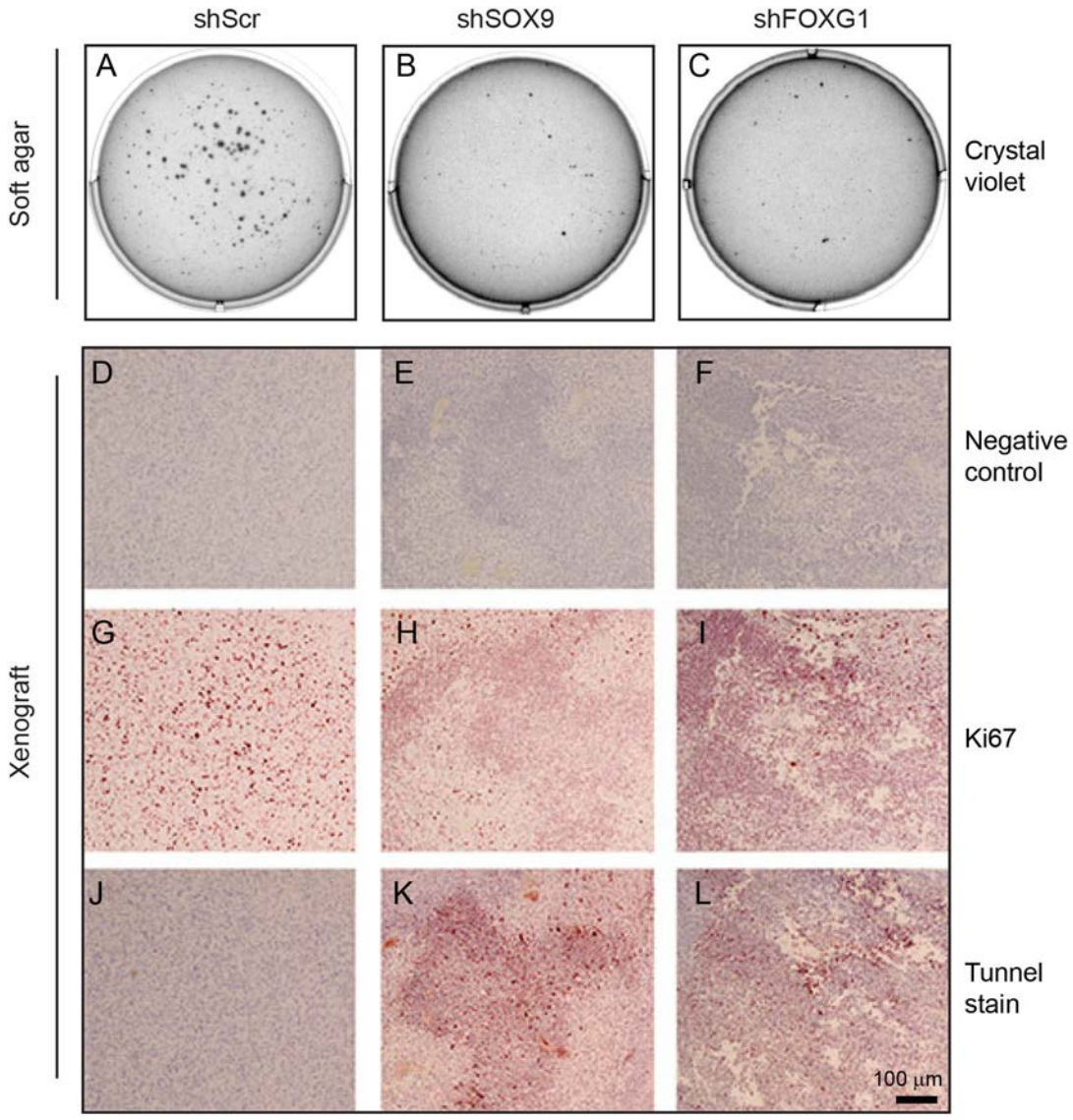

Figure S5, related to Figure 5. (A-C) Crystal violet stain of U87EGFRvIII cells cultured in suspension in soft agar for 17 days in 6-well culture plates. Knockdown of SOX9 or FOXG1 severely impairs the ability of U87EGFRvIII cells to form large colonies. **(D-L)** Immunohistochemical stain of adjacent sections of the mouse brains that contain intracranially injected U87EGFRvIII cells expressing short hairpin RNAs targeting *SOX9*, *FOXG1*, or a scrambled short hairpin RNA sequence. *SOX9* and *FOXG1* knockdowns both lead to dramatic decrease of Ki67 stain and increase of Tunnel stain in the tumors formed by U87EGFRvIII cells. Scale bar: 100 μm.

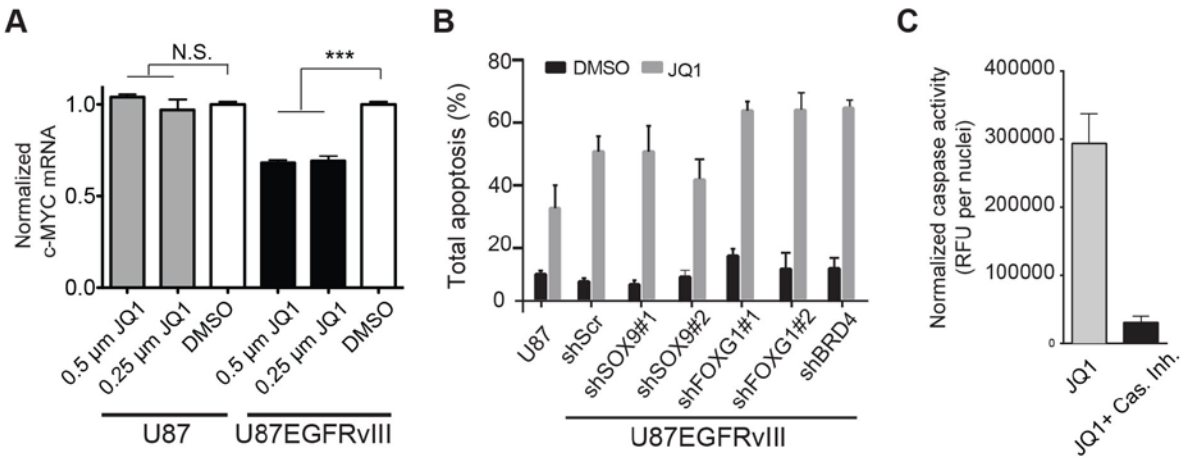

Figure S6, related to Figure 6. (A) JQ1 decreases *c-MYC* transcription specifically in U87EGFRvIII cells. qPCR was used to measure *c-MYC* mRNA levels in U87 and U87EGFRvIII cells treated with JQ1 or DMSO for 48 hours. *c-MYC* levels were normalized to both *GAPDH* and *ACTB*, and relative expression was compared to DMSO controls. **(B)** Total apoptosis of DMSO and JQ1 treated U87 and U87EGFRvIII cells, as determined by Annexin V/Propidium Iodide stain. **(C)** Quantitation of live-imaging assay of activated caspases. Incubation of U87 EGFRvIII cells (treated with 1 µM JQ1) with the pan-caspase inhibitor Z-DEVD-FMK (30 µM) immediately prior to staining significantly attenuates the fluorescence signal from the CellEvent reagent (p=0.0145).

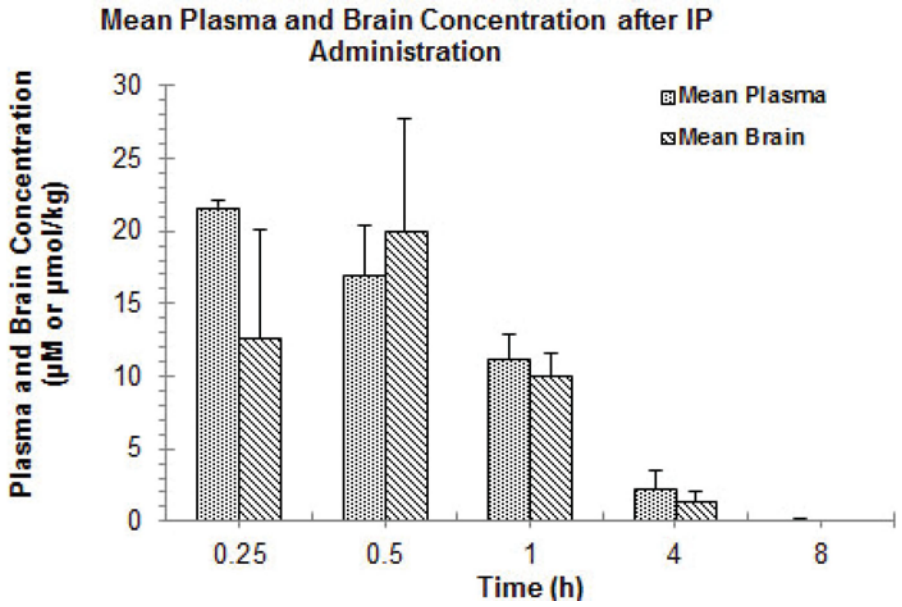

Figure S7, related to Figure 7. JQ1 pharmacokinetic profile. Following intraperitoneal injection of JQ1 in female nu/nu mice at 50 mg per kg body weight, drug concentrations in the plasma and the brain were measured by LC/MS at the indicated times. JQ1 exhibited high brain penetration with a maximum brain concentration of ~20 µM at 0.5 hr and a half-life of ~1 hr in both plasma and brain. Experiment was performed by WuXi AppTec, Shanghai, China.

Supplementary Tables

Table S1. Related to Figure S4.

Cancer type	Abbreviation	Frequently amplified receptor tyrosine kinases (RTKs)
Adrenocortical carcinoma	ACC	*FGFR3, FGFR4, FLT4*
Bladder Urothelial Carcinoma	BLCA	*DDR2, EGFR, ERBB2, FGFR1, FGFR3, NTRK1*
Breast invasive carcinoma	BRCA	*AATK, DDR2, EGFR, EPHB3, ERBB2, FGFR1, FGFR2, IGF1R, NTRK1, NTRK3*
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	*EGFR, EPHB3, ERBB2, IGF1R*
Colon adenocarcinoma	COAD	*EGFR, ERBB2, FGFR1, FLT1, FLT3, PTK7*
Esophageal carcinoma	ESCA	*EGFR, EPHB3, EPHB4, ERBB2, FGFR1, IGF1R, LMTK2, NTRK1, PTK7*
Glioblastoma	GBM	*EGFR, EPHB3, FGFR3, KDR, KIT, MET, PDGFRA*
Head and Neck squamous cell carcinoma	HNSC	*EGFR, EPHB3, EPHB4, ERBB2, FGFR1, IGF1R, LMTK2*
Kidney renal clear cell carcinoma	KIRC	*CSF1R, FGFR4, FLT4, PDGFRB*
Brain low grade glioma	LGG	*EGFR, EPHA1, EPHB3, IGF1R, KDR, KIT, PDGFRA, STYK1*
Liver hepatocellular carcinoma	LIHC	*AATK, DDR2, IGF1R, LMTK2, MET, NTRK1, PTK7*
Lung adenocarcinoma	LUAD	*AATK, DDR2, EGFR, EPHB3, ERBB2, FGFR1, FGFR4, KDR, KIT, MET, NTRK1, PDGFRA, PTK7, TIE1*
Lung adenocarcinoma	LUSC	*AATK, DDR2, EGFR, EPHA5, EPHA6, EPHB3, EPHB4, ERBB2, FGFR1, IGF1R, KDR, KIT, LMTK2, NTRK1, NTRK3, PDGFRA, TEK*
Ovarian serous cystadenocarcinoma	OV	*AATK, DDR2, EPHA1, EPHA10, EPHB3, EPHB6, ERBB3, FGFR3, IGF1R, NTRK1, NTRK3, PTK7, TIE1*
Pancreatic adenocarcinoma	PAAD	*LMTK2*

Prostate adenocarcinoma	PRAD	*FGFR1, FGFR3*
Rectum adenocarcinoma	READ	*DDR2, ERBB2, FGFR1, FLT1, FLT3, PTK7*
Sarcoma	SARC	*DDR2, NTRK1*
Skin Cutaneous Melanoma	SKCM	*AATK, DDR2, EPHA1, EPHB6FLT4, IGF1R, KIT, NTRK1, NTRK3, PDGFRA*
Stomach adenocarcinoma	STAD	*DDR2, EGFR, EPHB3, EPHB4, ERBB2, ERBB3, FGFR2, IGF1R, LMTK2, MET, NTRK1, NTRK3, PTK7*
Uterine Corpus Endometrial Carcinoma	UCEC	*AATK, DDR2, EPHA2, EPHB3, ERBB2, ERBB3, FGFR1, FGFR3, NTRK1, TIE1*
Uterine Carcinosarcoma	UCS	*AATK, EPHB3, ERBB3, FGFR1, FGFR3, NTRK1*

*Frequently amplified RTKs in different cancer types were according to the Broad Institute TCGA Copy Number Portal: http://www.broadinstitute.org/tcga/home

Table S2. Related to Figure 4.

Tumor ID	Tumor Nuclei	Pathology	EGFRviii PCR	EGFR Amplification	IDH1 mutation
B39	70%	GBM	positive	Yes	Negative
R28	90%	GBM	positive	Yes	Negative
S08	75%	GBM	positive	Yes	Negative
P69	80%	GBM	positive	Yes	Negative
2493	70%	GBM	negative	No	Negative
2585	70%	GBM	negative	No	Negative
IDHmut 5661	70%	Oligoastrocytoma III	negative	No	IDH1 R132H

Table S3 and Table S4 are Excel spreadsheets attached to this manuscript.

Table S3: PCR primer sets, related to Figures 2 and 3.

Table S4: ChIP-seq motif analysis, related to Figures 1 and 4.

Supplemental Experimental Procedures

Cell Culture

Cells were cultured at 37 °C in 5% CO_2, in Dulbecco's Modified Eagle Media (DMEM, Cellgro) supplemented with 10%FBS (Hyclone) and 1% penicillin/streptomycin/glutamine (Invitrogen). Serum-starved cells were obtained by culturing cells in DMEM for 24 hours. For drug treatment *in vitro*, cells were seeded in DMEM+0.5% FBS. DMSO or small compound stocks were added to culture media the next day. Drug used include: erlotinib hydrochloride (Selleck), U0126 (Millipore), torin1 (Selleck). JQ1 (Santai Labs).

High Throughput Sequencing Read Mapping

ChIP-seq experiments were mapped to human genome build hg18 using Bowtie v0.12.5 (Langmead et al., 2009) with the following options: "-v 3 -m 1 --best –strata". H3K27ac and input ChIP-seq experiments in normal tissue were obtained from the Roadmap Epigenomics Mapping Consortium (GSE17312) and lifted over from hg19 to hg18. PCR duplicates were removed with the Picard tool (http://broadinstitute.github.io/picard). Paired-end RNA-seq experiments were mapped using Tophat v1.4.0(Trapnell et al., 2009) with the following parameters: "-g 1 --library-type=fr-firststrand". 4C-Seq data were mapped using 4Cseqpipe (van de Werken et al., 2012) with the first cutter set as *NlaIII* (CATG) and the second cutter set as *DpnII* (GATC).

Enhancer Predictions

To identify putative enhancers in cell lines with H3K4me1, H3K4me3 and H3K27ac ChIP-seq data, we used RFECS (Rajagopal et al., 2013)to define an initial set and filtered for predictions with at least 1.75-fold enrichment of H3K4me1 or H3K27ac. To identify putative enhancers in tissues and tumor samples with H3K27ac data alone, we used MACS to find H3K27ac peaks (Zhang et al., 2008). Predictions within 5kb of a transcription start site were removed. To identify enhancers specifically active in a set of EGFRvIII positive samples $P = \{p_1, p_2, \ldots\}$ as compared

to a set of EGFRvIII negative samples $N = \{n_1, n_2, \ldots\}$, we devised the following general strategy. For a given sample *s* and histone mark $h \in H$, let $E_{s,h,l}$ denote the enrichment of the histone mark for the sample at locus *l* compared to input, as defined above. As the histone marks examined here are enriched as punctuated peaks, we assume that the genome-wide distribution of binned (1kb) data is, to a first approximation, Gaussian. Therefore, the random variables represent the difference between histone modification enrichment between an EGFRvIII+ sample p_i compared to an EGFRvIII- sample n_j is $D_{p_i,n_j,h,l} = E_{p_i,h,l} - E_{n_j,h,l}$, which is also Gaussian with mean and standard deviation denoted respectively as $\mu_{p_i-n_j,h}$ and $\sigma_{p_i-n_j,h}$. Therefore, the random variable $G_{p_i,n_j,h,l} = \frac{D_{p_i,n_j,h,l} - \mu_{p_i-n_j,h}}{\sigma_{p_i-n_j,h}}$ follows a Gaussian distribution with mean 0 and variance 1. Finally, let the test statistic $S_{h,l} = \sum_{p_i \in P} \sum_{n_j \in N} \left(G_{p_i,n_j,h,l}\right)^2$, which by definition follows a Chi-squared distribution with degrees of freedom equal to $|P| \times |N|$. Higher values of $S_{h,l}$ indicate increased deviation of samples in *P* from samples in *N* for locus *l*. Since $S_{h,l}$ is unsigned, we create the indicator random variable $C_{i,j,h,l} = \begin{cases} 1 \; if \; G_{p_i,n_j,h,l} > 0 \\ 0 \; otherwise \end{cases}$. Finally, to identify EGFRvIII-specific enhancers having a histone mark *h*, we keep loci *l* where $\sum_i \sum_j C_{i,j,h,l} > C_{cutoff}$ and where the one-sided Chi-Squared p-value for $S_{h,l}$ is less than p_{cutoff}, which is determined empirically with randomized data such that p_{cutoff} corresponds to a false discovery rate of *F*. For the cell line comparison of two U87vIII and two U87 samples, we let $C_{cutoff} = 3$ and *F* = 10%. For the more extensive comparison of six tumor biopsies and seven control specimens, we let $C_{cutoff} = 0.5 \times |P| \times |N| = 21$ and *F* = 1%.

Gene Expression Analysis

Total RNA was extracted using the Qiagen RNeasy mini kit (Qiagen). For the quantitative polymerase chain reaction (qPCR) experiments, total RNA was reverse-transcribed to single-stranded cDNA using the SuperScript VILO cDNA synthesis kit (Life Technologies) and qPCR

was performed using the KAPA Sybr green qPCR kit (KAPA Biosystems). For the experiment shown in Figure S6A, total RNA was reverse-transcribed using the High Capacity cDNA Reverse Transcription Kit and qPCR was performed using Taqman Gene Expression Assays (Life Technologies). PCR primer information is shown in Table S3.

Western Blot

Cell lysates were extracted in RIPA buffer. Protein concentrations were determined by the BCA protein assay kit (Thermo Scientific). For Western blots, 20 µg of total protein per sample was electrophoresized in 4-12% gradient SDS-PAGE gel and transferred to PDVL membrane according to Invitrogen protocol (Invitrogen). The membrane was incubated with primary antibodies at 4 °C overnight, followed by incubation with HRP-conjugated secondary antibodies at room temperature for 2 hours. Chemical fluorescence signals were detected by ChemiDoc (Bio-Rad). Antibodies used include rabbit anti-SOX9 (Millipore, Cat#AB5535), rabbit anti-FOXG1 (Active Motif, Cat# 31211), mouse anti- ACTIN (Sigma, Cat# A4700), rabbit anti-EGFR (Millipore, Cat#: 06847MI), rabbit anti-phospho-EGFR Y1068 (Cell Signaling, Cat# 3777S), rabbit anti-ERK1/2 (p44/42) antibody (Cell Signaling, Cat #9102), rabbit anti-phospho-ERK (Thr202/Tyr204) (Cell Signaling, Cat# 9101S), mouse anti-S6 (Cell Signaling, Cat# 2317S), rabbit anti-phospho-S6 (Ser235/236) (Cell Signaling, Cat# 4858S), HRP-linked anti-rabbit IgG antibody (Cell Signaling, Cat#7074S), HRP-linked anti-mouse IgG antibody (Cell Signaling, Cat# 7076S). All antibodies were diluted 1-2,000, except that the mouse anti-Actin antibody was diluted 1:10,000.

Immunohistochemistry

Formalin-fixed, paraffin-embedded (FFPE) tissue sections were prepared by the Histology Core Facility at UCSD Moores Cancer Center. Immunohistochemistry was performed according to standard procedures. Antigen was retrieved by boiling slides in 0.01 M of sodium citrate (pH 6.0) in a microwave for 15 min. Sections were incubated with primary antibodies at 4 °C

overnight, followed by incubation with biotinylated secondary antibodies at room temperature for 30 min. Antibodies used include: rabbit anti-SOX9 (Millipore, Cat# AB5535), rabbit anti-FOXG1 (Thermo Scientific, Cat# PA5-26794), anti-Ki67 (Thermo Scientific, Cat# PA5-16785). phospho-PDGFRbeta (Y751) (Cell Signaling, Cat# 3161L), and phospho-Met (Y1234/1235) (Cell Signaling, Cat# 3077P). All antibodies were used in 1:200 ~1:500 dilutions.

Plasmid Constructs

The infra-Red Fluorescent Protein 720 (iRFP720) Cdna (Shcherbakova and Verkhusha, 2013) was PCR-cloned into the *Nhe* I and *Xho* I restriction enzyme sites of the lentiviral plasmid vector pLV. Short hairpin constructs were purchased from Sigma: shSOX9#1 (TRC#: TRCN0000020384; target sequence: 5'-GCATCCTTCAATTTCTGTA-3'), shSOX9#2 (TRC#: TRCN0000020386 ; target sequence: 5'-CTCCACCTTCACCTACATG-3'), shFOXG1#1 (TRC#: TRCN0000013952; target sequence: 5'-CATGAAGAACTTCCCTTAC-3'), shFOXG1#2 (TRC#: TRCN0000013949; target sequence: 5'-CCACAATCTGTCCCTCAAC-3'). shBRD4 (TRC#: TRCN0000196576; target sequence: 5'-GCCAAATGTCTACACAGTA-3').

Lentivirus Production and Infection

Lentiviral plasmid constructs were co-transfected with lentivirus packaging plasmids to Lenti-X 293T cells (Clontech), using X-tremeGENE DNA transfection reagents (Roche). Supernatants containing high titer lentiviruses were collected between 24-72 hours after transfection and were filtered through a 0.45 μm syringe filter before use. For infection experiments, $5x10^5$ U87 and U87EGFRvIII cells were seeded in 10 cm culture dish. On the next day, 1/10 volume of high titer lentivirus and 12.5 ng/ml (final concentration) of polybrene were mixed and added to cultured cells. After 8 hours of incubation in 37 °C, the supernatant containing virus was replaced by fresh culture media (DMEM/10%FBS). Infected cells were selected by puromycin (1 μg /ml) or blasticidin (2 μg /ml).

siRNA Transfection

Transient knockdown experiments using siRNA was performed as described (Akhavan et al., 2013). siRNAs were purchased from Life Technologies. siEGFR#1 (Ambion Silencer Select ID: s564). siEGFR#2 (Ambion Silencer Select ID: s565).

Caspase Assay

Cells (2000/well for U87 and U87 EGFRvIII kinase dead, 1500/well for U87 EGFRvIII) were seeded in optical 96-well plates (Greiner, 655090) and incubated for 16 hours at 37°C, 5% CO_2. JQ1 (DMSO stock; dose response from 0.03 μM to 10 μM) and 4 μM CellEvent Caspase-3/7 Green Reagent (Life Technologies, C10423) were then added to the culture medium. After 96 hours at 37°C, 5% CO_2, NucBlue Live ReadyProbes Reagent (Hoechst 33342; Life Technologies, R37605) was added 30 minutes prior to imaging on a Cell Voyager 7000 spinning disk confocal imaging system (Yokogawa Electric Corporation) with a 20X 0.75 NA U-PlanApo objective and 2560x2160 sCMOS camera (Andor) at 2x2 binning. The imaging chamber was maintained at 37°C and 5% CO_2. Image acquisition and data analysis were performed using the CV7000 software. 6-8 fields/well were imaged, with duplicate wells for each condition. 3 x 2 μm z-sections in the blue (40 % power, 300 ms, 2.2X gain) and green (40 % power, 300 ms, 2.2X gain) channels were captured in each field. JQ1-dependent caspase activity per cell was calculated by dividing the total fluorescence at 530 nm (corresponding to the cleaved caspase reporter reagent) by the number of nuclei (Hoechst staining) and then subtracting background fluorescence per cell (as determined from the DMSO-treated wells). EC_{50}s were computed using this normalized data and a 4-parameter, variable slope fit with Prism 5 (GraphPad). In control studies (Figure S6B), cells were plated as above. After 16 hours at 37°C, 5% CO_2, JQ1 (1 μM) was added to the culture medium. After an additional 96 hours at 37°C, 5% CO_2, the cells were treated with 30 μM Z-DEVD-FMK (Tocris) for 1 hour prior to staining with the CellEvent Caspase 3/7 Green and NucBlue Live ReadyProbes Reagents for 1 hour. Imaging

and data analysis were performed as above.

Cell Proliferation and Viability Assay

Cells were seeded in 12-well culture plates at 15,000 cells per well. After three days, total and live cells in each well were counted by Trypan blue assay using a TC10 automatic cell counter (Bio-Rad).

Flow Cytometry

Cells were seeded at a density of 37,500 in each well of 6-well plate. The following day, DMSO or 1 μM JQ1 was added in DMEM media containing 2.5% FBS. Annexin V/Propidium Iodide (PI) staining was carried out 5 days following JQ1 treatment using the FITC Annexin V Apoptosis Detection Kit II (BD Biosciences). Flow cytometric analyses were performed using the BD LSR II Flow Cytometer System (BD Biosciences). 10,000 cells within the gated population were analyzed.

Supplemental References

Akhavan, D., Pourzia, A.L., Nourian, A.A., Williams, K.J., Nathanson, D., Babic, I., Villa, G.R., Tanaka, K., Nael, A., Yang, H.*, et al.* (2013). De-repression of PDGFRbeta transcription promotes acquired resistance to EGFR tyrosine kinase inhibitors in glioblastoma patients. Cancer discovery *3*, 534-547.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoyannopoulos, J., Ernst, J., Kellis, M., and Ren, B. (2013). RFECS: a random-forest based algorithm for enhancer identification from chromatin state. PLoS computational biology *9*, e1002968.

Shcherbakova, D.M., and Verkhusha, V.V. (2013). Near-infrared fluorescent proteins for multicolor in vivo imaging. Nature methods *10*, 751-754.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A.*, et al.* (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nature methods *9*, 969-972.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.