Supplemental Figure Legends

Supplemental Figure 1. Identification of Super enhancers by ROSE. **A:** Super enhancer tracks in IGV overlaid with H3K27Ac tracks from measured cell lines at the *EGFR* gene. Super enhancers were identified by stitching together peaks of H3K27Ac histone marks with the ROSE software package. (Black bars) Super enhancers as called by ROSE. **B:** The *EGFR* intron 1 super enhancers were ranked based on H3K27Ac signal at the shown intron 1 super enhancers in the indicated GBM and HNSCC cell lines using the ROSE software package.

Supplemental Figure 2. Schematic of positioning of enhancer segments 1-20 analyzed in the pGL4.24 vector. CE/CE2 (pink outline) are highlighted.

Supplemental Figure 3. Targeting dCas9-KRAB to the CE's decreases *EGFR* gene transcription. **A:** (Left) H3K27Ac IGV track of SF767 cells showing the position of gRNA's targeting the *EGFR* intron 1 enhancers and off-target (O-T) control. (Right) Western blot for dCas9-KRAB expression in SF767 cells after transduction with lenti-dCas9-KRAB-blast. β-Actin was used as a loading control. **B:** H3K27Ac enrichment at the targeted enhancer regions before and after dCas9-KRAB targeting was analyzed by ChIP-qPCR. Primers were designed around the targeted regions as well as a PCR negative control (Ctrl) from a H3K27Ac-negative region within *EGFR* intron 1. **C:** (Top) *EGFR* expression in dCas9-KRAB expressing cell lines was analyzed by RT-qPCR. *EGFR* transcript level was first normalized to *GAPDH* and subsequently calculated as fold change relative to off-target control. (Bottom) Western blot for EGFR protein levels. β-Actin was used as a loading control. **D:** Cell proliferation curves were generated by measuring ATP levels every two days over 9 days. Significance is measured relative to O-T for knockdown cells. **B-D:** (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, n = 3 independent experiments, Student's *t* test).

Supplemental Figure 4. AP-1 family transcription factors are identified as possible modulators of *EGFR* intron 1 enhancers. **A:** Top 5 HOMER *de novo* motif search results in ATAC-seq peaks located within TSS-distal enhancers. % of Targets = Percentage of input peaks containing the indicated motif. % of

Background = Percentage of randomly generated background sequences which contain the indicated motif. The number of background regions is 2x the total number of input peaks. STD(Bg STD) = average distance from the peak center where the motif was identified. **B:** Tracks of ENCODE ChIP-seq peaks from c-Jun and c-Fos ChIP-seq experiments in HeLa and HUVEC cells overlaid with tracks for H3K27Ac and ATAC-seq in glioma and HNSCC cell lines. Darker bars indicate stronger binding signals. **C:** Analysis of EGFR, JunDN-HA, c-Jun, and c-Fos protein expression in U87 cells by western blot after transduction with pMIEG3-JunDN-HA. β-Actin was used as a loading control. **D:** Analysis of JunDN efficacy on a luciferase reporter containing a trimerized AP-1 binding site (***$p < 0.0005$, n = 3 independent experiments, Student's *t* test).

Supplemental Figure 5. JQ1 treatment reduces *EGFR* transcription through inhibition of transcription factor activity. **A:** Analysis of EGFR, BRD4, BRD3, BRD2, c-Fos, c-Myc, and c-Jun protein expression in SF767 cells by western blot after treatment with 0.5μM JQ1 for 24 hours. β-Actin was used as a loading control. **B:** *EGFR* expression was analyzed by RT-qPCR in SF767 cells treated with 0.5μM JQ1 for 24 hours. *EGFR* transcript level was first normalized to *GAPDH* and subsequently calculated as fold change relative to DMSO control. **C:** Analysis of EGFR, BRD4, BRD3, BRD2, c-Fos, and c-Jun protein expression in the indicated cell lines by western blot. β-Actin was used as a loading control. **D:** BET bromodomain protein and histone acetylation ChIP-seq tracks from the liposarcoma cell line LPS141. **E:** Fold changes in enrichment of the indicated factors after 24 hours of 0.5μM JQ1 was measured at the indicated regions by ChIP-qPCR. ChIP enrichment is normalized to a negative control primer in chr12. **F:** Relative interaction frequency by 3C of each restriction fragment (F1-10) was calculated as described in the experimental procedures and was plotted against genomic location of the EcoRI restriction site. Significant differences in interaction are indicated for SF767 +JQ1 relative to control (SF767 +DMSO). **G:** Analysis of EGFR, BRD4, BRD3, BRD2 and β-Actin protein expression in SF767 cells by western blotting after treatment with indicated siRNA. A scrambled siRNA was used as treatment control and β-Actin was used as a loading control. **B,E-F:** (*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, n = 3 independent experiments, Student's *t* test)

Supplemental Figure 6. Clinical sample data shows activity of *EGFR* intron 1 enhancers and associated transcription factors. **A:** (Left) RNA-seq FPKM values for *EGFR* in the indicated GSC or GBM primary tumor ordered from highest (top) to lowest (bottom). Samples highlighted in red have known *EGFR* gene amplification. (Right) H3K27Ac tracks for the indicated GSC or GBM primary tumor. Data range is equal for each track. **B:** *EGFR* RNA-seq FPKM plotted against the RNA-seq FPKM of each indicated transcription factor. The relationship was analyzed using Spearman's rank order correlation. **C:** TCGA data was accessed using the Gene Expression Profiling Interactive Analysis (GEPIA) database. Tumor versus normal samples were compared for GBM and HNSCC for the indicated transcription factors.

Supplemental Methods

ChIP-seq analysis. Raw reads from Illumina Hi-seq 4000 were aligned to the human genome (hg19) using Bowtie2 software (1) (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) with default parameters. Non-uniquely mapped reads were removed with "awk" command and unique reads were used for secondary analysis. Genome-wide read coverage was calculated by igvtools count and visualized using the Integrative Genomics Viewer (IGV) (http://software.broadinstitute.org/software/igv/home). Peaks were called by first creating TagDirectories of H3K27Ac and Input samples using the "makeTagDirectory" command in the HOMER (Hypergeometric Optimization of Motif EnRichment) suite of tools (http://homer.ucsd.edu/homer/index.html) (2). Peaks were identified from the H3K27Ac tag directories using the input directories for normalization with the "findPeaks" command from HOMER with the following parameters: -style histone -size 250 -minDist 250. ChIP-seq tracks are generated from single biological replicates with a minimum of 10 million reads.

ATAC-seq analysis. Nextera adapters were trimmed from the raw fastq files by using cutadapt (3) (https://cutadapt.readthedocs.io/en/stable/index.html) with parameters "-m 5 -e 0.10 -a CTGTCTCTTATA -A CTGTCTCTTATA" and then aligned to human reference genome hg38 using bowtie2 with parameters "-X2000 --mm --local". Next, the improperly mapped, poorly mapped and unmatched reads were filtered from the resultant raw bam files using samtools view with parameters "-F 1804 -f 2 -q30". Duplicates were marked with Picard (https://broadinstitute.github.io/picard/index.html) with the command "picard MarkDuplicate"

and removed by “samtools view”. Final bam files were generated after removing mitochondrial reads by “awk” command. Replicate Bam files were merged using “samtools merge” and converted to tagAlign format using the bedtools (https://bedtools.readthedocs.io/en/latest/) command “bamtobed” with parameters “-bedpe -mate1 -l” and “awk”. To account for the cutting offset of Tn5 transposase, mapping position was shifted using “awk”. Peaks were called from tagAlign files using the Model-based Analysis of ChIP-Seq (MACS2) (4) (https://github.com/taoliu/MACS) callpeak command with parameters “-g hs -p 0.01 --nomodel --shift -75 --extsize 150 -B --SPMR --keep-dup all --call-summits”. Bedgraph files output from MACS2 were converted to BigWig files using bedGraphToBigWig and visualized using IGV.

Super enhancer identification. The ROSE (Rank Ordering of Super-Enhancers) algorithm was downloaded (https://bitbucket.org/young_computation/rose/src) and run from the command line (5,6). Bam files of H3K27Ac and Input ChIP-seq reads for each cell line were generated from raw files with bowtie2 and sorted with the samtools (http://samtools.sourceforge.net/) command “samtools sort”. GFF files were created from peak files for each cell line using the “awk” command. ROSE was called using the command “python ROSE_main.py” with the following parameters: -g HG19 -i [GFF_File] -r [H3K27Ac_ChIP.sorted.bam] -o [Output_Folder] -c [Input.sorted.bam] -t 3000.

TF Motif identification. Tag directories were created from H3K27Ac alignment files using HOMER makeTagDirectory. Peaks from the H3K27Ac ChIP-seq were identified from the tag directories using HOMER findPeaks with parameters “-style histone -size 250 -minDist 250”. NarrowPeak files from individual ATAC-seq replicates and pooled peak files from MACS2 were sorted for high P-value peaks (>= 13 -log10 [p < 0.05]) with “awk”. In order to get a set of high P-value, replicated peaks, high P-value pooled peaks were intersected with high P-value peaks from each replicate using bedtools intersect with the -u parameter. In order to exclude promoter proximal peaks, replicated high P-value peaks were sorted for. The Gencode hg38 annotation (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_29/gencode.v29.annotation.gtf.gz) was downloaded and TSS +/- 2kb were filtered out with “awk”. High p-value peaks were intersected with the Gencode TSS annotation with bedtools intersect with the -v parameter. H3K27Ac peaks were

modified from hg19 to hg38 using the hgLiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver) for intersecting with ATAC-seq peaks. Promoter distal high p-value ATAC peaks were intersected with the H3K27Ac ChIP-seq peaks to identify ATAC peaks located within regions of H3K27Ac using bedtools intersect with the -wa parameter. Motifs were identified from these peaks using HOMER findMotifsGenome with the parameter "-size given". *De novo* motifs were identified from html files output by HOMER.

GEPIA analysis: Gene Expression Profiling Interactive Analysis (GEPIA) (7) was accessed from the internet (http://gepia.cancer-pku.cn/index.html). Plots were generated with the following parameters: Differential Methods, ANOVA; |Log2FC| Cutoff 1; q-value Cutoff 0.01; Log scale no; Match TCGA normal and GTEx data.

Supplemental References

1. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
2. 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.
3. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011;17:3.
4. 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:R137.
5. Lovén 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;153:320–34.
6. Whyte W a, Orlando D a, 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;153:307–19.
7. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45:W98–102.