Supplementary Materials

SUPPLEMENTARY METHODS

WGS and Structural Variant Detection

Sequencing libraries were prepared from genomic DNA of tissues and cell lines with 300-400 bp insert sizes, and paired-end (101 bp or 151 bp) whole genome sequencing (WGS) performed using either Illumina Hiseq 2000 or 2500 instruments. Sequencing reads were processed as described (1) LUMPY (2) was used with default parameters to detect structural variations. For public WGS data, we obtained data of 40 GC tumors from TCGA (https://gdc.cancer.gov), 100 tumors from Wang et al. (HK)(3), and 32 tumors from ICGC (https://ega-archive.org/datasets/EGAD00001003132). For GC cell lines, reads from whole-genome sequencing were processed as described (1) and structural variations were detected using LUMPY (2). To classify a structural variation as somatic, breakpoints were compared against matched normal samples and/or germline databases including GTEx (4), Database of Genomic Variants, DGV (5) using the pairToPair tool (-slop 500).

Nano-ChIP-Seq

Nano-ChIP-Seq was performed as previously described with slight modifications (6). Briefly, fresh-frozen cancer and normal tissues were dissected in liquid nitrogen to obtain ~5 mg sized piece for each ChIP. One million cells from each cell line were used for each experiment. ChIP were performed using the following antibodies: H3K27ac (ab4729, Abcam), H3K4me1 (ab8895, Abcam). After recovery of ChIP and input DNA, whole-genome-amplification was performed using the WGA4 kit (Sigma-Aldrich) and BpmI-WGA primers. Amplified DNAs were purified using PCR purification columns (QIAGEN) and digested with BpmI (New England Biolabs) to remove WGA adapters. 30ng of amplified DNA was used for each sequencing library preparation (New England Biolabs). Libraries were sequenced on a Hiseq2500 (Illumina) to an average depth of 20-30 million reads per library. Sequence reads were mapped against the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA-MEM, version 0.7.0)(7). Reads from histone ChIP-seq were trimmed at the first and last 10bp prior to alignment. Reads with low mapping quality (MAQP<10) and PCR duplications were removed using Samtools(8). Histone ChIP-enriched peaks were detected using CCAT (FDR<5%) (9). Sequencing statistics are shown in Supplementary Table 3.

Telomere Length Measurements

Genomic DNA of primary samples were obtained using the DNeasy Blood and Tissue Kit (Qiagen) and subjected to telomere length measurements. Genomic DNAs (1 ug) were digested with HphI and MnlI at 37°C for 16 hours. The digested DNA was run on a 0.8% agarose gel in 1X TAE bffer gel overnight at 65 Volts. After depurination and denaturation, the DNA was transferred to membranes and hybridized with ^{32}P-labeled $(TTAGGG)_6$ oligonucleotides as previously described(10).

Capture-C and Analysis

Capture-C was performed as previously described (11). Briefly, 1 x 10^7 cells were crosslinked by 2% formaldehyde, followed by lysis, homogenization, DpnII digestion, ligation and de-crosslinking. DNA was sonicated to 150–200 bp to produce DNA suitable for oligo capture. Three micrograms of sheared DNA was used for sequencing library preparation (New England Biolabs). *TERT* promoter associated sequences were double captured by sequential hybridization to customized biotinylated oligos (IDT) and enrichment with Dynabeads (LifeTech). Captured DNA was sequenced on an Illumina HiSeq 3000 system using the 2 x 150 bp configuration. Sequence reads were preprocessed to remove adaptor sequences using trim_galore and overlapping reads were merged using FLASH. To achieve short read mapping to the hg19 reference genome, the preprocessed reads were in-silico digested with DpnII and aligned using Bowtie (using p1, m2, best and strata settings). Aligned reads were processed using Capture-C analyser to remove PCR duplicates and to classify sub fragments as 'capture' if they were contained within the capture fragment; 'proximity exclusion' if they were within 1 Kb on either side of the capture fragment; or 'reporter' if they were outside of the 'capture' and 'proximity exclusion' regions. Normalized read counts per 100,000 interactions were visualized in bigwig format, and r3Cseq used to identify significant interactions of the viewpoint against a scaled background (q-value <0.05).

CRISPR/Cas9 Genome Editing (Deletion)

To delete the *TERT* 5' proximal region, two sgRNAs (left and right) were designed using the online tool (http://crispr.mit.edu). Two sgRNAs were cloned into the all-in-one CRISPR/Cas9 vector system as described (12). After transfection using Lipofectamine 3000 (Life Technologies), GFP-positive single cells were sorted and cultured. Single cell expended clones were subjected to DNA extraction and validated for deletion of the *TERT* 5' proximal region by PCR. *TERT* expression levels were measured using qPCR and Taqman probes (Hs00972656_m1 FAM Applied Biosystems). Clones that were transfected with gRNAs but failed to exhibit deletions were used as negative controls. Two backbones used in the all-in-one CRISPR/Cas9 vector system are: pX330A-2A-GFP-1X2 backbone (Addgene # 58766, with addition of GFP by Dr. Shang Li, Duke-NUS, Singapore); and pX330S backbone (Addgene

58778). CRISPR sgRNAs used to target the *TERT* 5' proximal region are (1) GCGGGTTCTGTCCACAGGTT; (2) CCAACACTCACATGCGTTGA.

CRISPR–Cas9 Mediated Mutation of EBF1 binding sites

To mutate EBF1-binding motifs at ChIP-seq peaks, one guide RNA for each EBF1 peak was designed using CRISPOR (13). Guides were cloned in the pSpCas9(BB)-2A-GFP (PX458) vector and sequences of the guide RNAs were verified by Sanger sequencing. YCC11 cells were transfected with either EBF1-Peak1 guide or EBF1-Peak2 guide or guides for both peaks using standard lipofectamine LTX Reagent protocol (ThermoFisher Scientific, 15338100). For each transfected condition, cells were sorted as single cells in the 96 well plates after 72h of transfection and incubated at 37^{o}C. Fresh media was added after one week, and the plates were incubated at 37^{o}C until cells confluency reached 60-70%. Single cell clones were transferred to 24 well plates and screened for the mutations within the EBF1 binding motif by RPFP analysis and Sanger sequencing. Clones harboring mutations were further expanded and harvested for Taqman qPCR.

Guide Sequence:

Guide EBF1-Peak1: TGTGACCACCTGTTATCCCA

Guide EBF1-Peak2: ATGCCGTTTATAAAGCCTGC

Silencing and Over-expression of *EBF1*

ON-TARGETplus Human siRNA SMARTpools or individual ON-TARGETplus Human individual siRNAs were used to transfect cells (8 x 10^4 cells of YCC11 per well; 4 x 10^4 cells of AGS or SNU484 per well) using Dharmafect 1 transfection reagent (Dharmacon). Cells were harvested after 72 h of RNAi treatment. Two viral systems were used for *EBF1* over-expression in GC cell lines – a lentiviral system and retroviral system. Lentiviral plasmids with a pLX304 backbone (14) were packaged in 293T cells by transfection using Lipofectamine 2000 (Thermo Fisher Scientific). GC cells were infected with lentivirus packaged with or without *EBF1* genes for 72 hours and treated with Blasticidin (7.5 µg/µl for AGS and SNU484; 5 µg/µl for IM95) for 72 hours and collected for further analysis. For the retroviral system, *EBF1* cDNAs were cloned into MIGR1 vector (#27490, Addgene) between the HpaI and Bgl 2 sites. Mutations in *EBF1* coding regions were introduced into MIGR1 overexpression plasmids (#210514, Agilent Technologies). Retrovirus packaging was performed using Platinum-A (Plat-A) retroviral packaging cell lines by transfection using Lipofectamin 2000 (Thermo Fisher Scientific). GC cells were infected with retrovirus packaged with or without *EBF1* genes for 48 hours and sorted for GFP.

DNA Methylation Analysis

Genomic DNA of gastric tumors and matched normal gastric tissues was extracted (QIAGEN) and processed for DNA methylation profiling using Illumina HumanMethylation450 BeadChips (HM450). Data of 12 CpG probes in the 5'UTR and TSS region of *EBF1* was analysed. For in-house cohort (n = 18 T/N pairs), methylation β-values were calculated and background corrected using the methylumi R BioConductor package. Normalization was performed using the BMIQ method (watermelon package in R). CpG island locations were downloaded from the UCSC genome browser. T-N Beta values for plotted for all samples. For TCGA cohort, Level 3 processed Illumina HumanMethylation450 DNA methylation data for TCGA STAD samples was downloaded from http://gdac.broadinstitute.org (20160128 standard data) and plotted for the top and bottom 25 percentile of EBF1 expressing samples (n = 191 GCs).

Real-time TRAP assay

Real-time TRAP assays were performed as previously described(15). Briefly, 1 x 10^5 cell pellets were lysed in 100 µl 1X CHAPS buffer and incubated on ice for 30 minutes. Cell lysates were spun at 14,000 rpm for 20 minutes at 4°C in a microcentrifuge. TRAP reactions were performed in 1X TRAP buffer (20 mM Tris-HCl pH8.3, 1.5 mM MgCl2, 63 mM KCl, 0.05% Tween 20 (Roche), 1 mM EGTA, 0.1 mg/ml BSA), 50 µM dNTPs, and 8 ng/µl TS Primer (5'-AATCCGTCGAGCAGAGTT-3') in a 10- µl total volume for 30 min at 30°C. Reactions were stopped by heating at 94°C for 1 minute. PCR reaction mixtures contained 1X TRAP buffer, 4 ng/µl ACX primer 5'-GCGCGGCTTA CCCTTACCCTTACCCTAACC-3', 15% glycerol, 1:2,000 SYBR Green from Invitrogen, 0.08 U/µl Taq polymerase and was initiated at 94°C for 2 min, followed by 45 cycles of 95°C for 5 s, 50°C for 6 s, and 72°C for 10 s. Threshold cycle values (Ct) were determined from semi log amplification plots. All samples were run in triplicate. Error bars show standard deviation.

Tumor xenografts in NOD_SCID mice

Female NOD_SCID mice (5 week-old) were housed in the SingHealth Experimental Medicine Centre. Mice were subcutaneously injected with 4 x 10^5 AGS cells overexpressing either *EBF1* or empty vector controls. Tumor diameters were measured twice per week with calipers, and tumor volumes were calculated based on the formula: width (mm) x width (mm) x length (mm) x 0.5. Sample size was 6 (n=6) for each group. At the end of the experiment, tumors were excised and snap frozen in liquid nitrogen for further analysis. P-values were derived using Student's t-test and P-values <0.05 were considered statistically significant.

Drug Treatments

SNU484 and AGS cells were seeded in 6-well plate (4 x 10^4 per well) and treated with GSK126 (Selleck Chemicals) and GSK 343 (Sigma-Aldrich) on the next day, at a concentration of 5 µmol/L for 72 hours. Control cells were treated with the same concentration of DMSO. Cells were then harvested for further studies. For combination drug treatments, SNU484 and AGS cells were seeded in 6-well plates (4 x 10^4 per well) and treated with 1.5 µmol/L (AGS) or 4 µmol/L (SNU484) GSK126 (Selleck Chemicals) and/or 1 µmol/L 5-aza (Sigma-Aldrich) for 72 h. For treatment of cells with TSA, 0.05 µmol/L TSA (Sigma-Aldrich) was added 24 h before cells were harvested. Cells were harvested on day 3 of drug treatment for further studies. For Vorinostat treatment, 0.5 µmol/L Vorinostat (Sigma-Aldrich) was added. Cells were harvested for further studies after 24h.

RNA-seq and data analysis

Sequencing libraries were constructed from total RNA of primary GCs, matched normal gastric tissues, and GC cell lines with 300-400 bp insert lengths, and sequencing was performed using paired-end 101bp reads. RNA-seq reads were aligned to the human genome (hg19) using TopHat2 (version 2.0.12)(16) using default parameters. Transcript abundances in units of FPKM at the gene level were estimated using Cufflinks(17). Gene expression values from primary samples were corrected for potential batch effects using ComBat(18).

CapSTARR-seq and analysis

Generation of input plasmid library

Generation of plasmid libraries was based on the method described by Vanhille et al. Briefly, SNU16 and OCUM1 genomic DNA was extracted using the PerfectPure DNA Cultured Cell kit (5 Prime). The DNA was then sheared by sonication (Covaris S2) and size-selected on 1% agarose gel (350–650 bp). Size-selected DNA fragments were ligated to adaptors (Adaptor 1: 5'-ACACTCTTTCCCTACACGACGCTCT TCCGATCT-3'; adaptor 2: 5'-GATCGGAAGAGCACACGTCT-3') after end repair and A-tailing (NEBNext Ultra End Repair/dA- Tailing and Ultra Ligation Module, New England Biolabs) and subjected to capture (NimbleGen SeqCap EZ, Roche) following manufacturer's instructions. Post-capture DNA was PCR amplified using Q5 High-Fidelity DNA polymerase (NEB) to add homology arms for subsequent cloning (forward primer: 5'-GTAATAATTCTAGAGTCGGGGCGGGAATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATCT-3'; reverse primer: 5'- TATCATGTCTGCTCGAAGCGGCATAGTGACTG GAGTTCAGACGTGTGCTCTTCCGATCT-3'; PCR program: 98 °C for 30 s; 17 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s). The vector backbone was modified from pGL4.23 [luc2/minP] (Promega). Briefly, minP promoter was replaced with SCP1 promoter and a CmR-ccdB cassette was cloned

between luc2 ORF and SV40 poly (A). The PCR product was cloned into the vector backbone (linearized by SphI and NdeI restriction enzymes) through Gibson Assembly (Gibson Assembly Master Mix, New England Biolabs) and transformed into MegaX DH10B™ T1R electrocompetent cells (ThermoFisher Scientific). A total of 8 Gibson Assembly and transformations were performed. All transformed *E. coli* cells were pooled and cultured in 4 L LB medium and harvested when OD600 reached 1.0. The input plasmid library was extracted using Plasmid Mega Kit (Qiagen) and drop dialyzed before transfection. Sequencing libraries were prepared from the input plasmid libraries by amplifying inserts from the plasmid DNA. 10 PCR reactions were set up, using Q5 High-Fidelity DNA polymerase (New England Biolabs), each with 1 ng plasmid as template (PCR program: 98 °C for 30 s; 12 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s). Reactions were pooled, cleaned up and submitted for sequencing.

Transfection of input plasmid library into cells

Input plasmid library was electroporated into cells using the BTX Electrofusion System (Harvard Apparatus) at a concentration of 1 µg DNA/1 million cells using parameters carefully optimized for each cell line. A total of 150 million were electroporated in each replicate for OCUM1 and 200 million cells for SNU16, and three biological replicates were generated per cell line. As a control, a GFP-plasmid was transfected simultaneously to monitor transfection efficiency and batch effects. After transfection, cells were plated in complete medium for 48 h until harvest.

CapSTARR-seq library preparation

Total RNA was extracted from transfected cells using PerfectPure RNA Cultured Cell Kit (5 Prime), followed by mRNA isolation using Dynabeads® Oligo (dT) kit (ThermoFisher Scientific) and treatment with TURBO DNase (ThermoFisher Scientific). Then target-specific first strand cDNA synthesis was performed with SuperScript III (ThermoFisher Scientific) to specifically reverse transcribe mRNA from the transfected plasmid library (RT primers: 5'-CAGTGACTGGAGTTCAGACG-3', 5'-GCAGCTTATAATGGT TACAAATAAAGC-3', 5'-ACATTTGTAGAGGTTTTACTTGCT-3', 5'-TCAATGTATCTTATCATGTCTGCTCGAAG-3'). After reverse transcription, reactions were treated with RNase A + H. Finally, the cDNA was PCR amplified using 5 µL cDNA reaction as template in every 50 µL PCR reaction. Q5 High-Fidelity DNA polymerase (New England Biolabs) was used, and the following PCR program: 98 °C for 30s; 20 cycles of 98 °C for 10 secs, 65 °C for 30 secs and 72 °C for 30 secs. Illumina HiSeq platform compatible primers were used (forward: 5'- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC T-3'; reverse: 5'-CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTG-3'). For each replicate, we set up separate final PCR reactions using 160 different indexing primers. PCR products for every 10 indexes were pooled and cleaned up by AmPure XP beads (Beckman Coulter; beads/reaction ratio = 0.8) and submitted for Illumina sequencing.

Data processing and analysis

CapSTARR-seq consists of two parts of data with different types of readouts. The input library represents the insert fragments in original plasmid pool, while the output library measures the abundance of self-transcribed mRNA from insert fragments of the transfected plasmid pool. Paired-end sequencing reads (2 × 100 bp) from Illumina Hiseq2000/2500 were aligned to human genome hg19 by BWA using default parameters. Only reads that mapped in proper pairs were used for downstream analysis. To exclude potential PCR amplification bias, fragments (inferred from paired-end reads) that have the same start and end positions were collapsed into distinct fragments by Picard (http://broadinstitute.github.io/picard/). An enhancer peak is called when there is significant enrichment of fragments from one region in output library than the representation of that region in input library based on Poisson distribution using MACS2. The genome coverage of the plasmid library was used as input when calculating the enrichment of CapSTARR-seq reads. CapSTARR-Seq data of cell line SNU16 is from another dataset under Gene Expression Omnibus GSE118492 (19).

Sanger Sequencing of Structural Variants

PCR primers were designed using Primer3 (v.0.4.0) (http://bioinfo.ut.ee/primer3-0.4.0/). Purified PCR products were sequenced in the forward and reverse directions using ABI BigDye Terminator Cycle Sequencing kit (Version 3.1) and an ABI 3730XL DNA Analyzer (Applied Biosystems). Sanger sequencing data was analysed using Chromas V2.6.5 (Chromas). Sequences of primer sets used are:

TERT-LMSU-Forward: 5'-GCCTAGGCTGTGGGGTAAC-3'

TERT-LMSU-Reverse: 5'-ATCGTTTTTGTGCCCAGTTC-3'

TERT-T990489-Forward: 5'-AGACTGACTGCCTCCATCGT-3'

TERT-T990489-Reverse: 5'-GCTTCTCCAGCAGCTTCCTA-3'

Standard ChIP-Seq and analysis

30 Million cells were cross-linked with 1% formaldehyde for 10 min at room temperature, and reactions were stopped by adding glycine to a final concentration of 0.2 M. Chromatin was extracted and sonicated to 500 bp. *EBF1* (sc-137065, Santa Cruz Biotechnology; H00001879-M01, Abnova) was used for chromatin immunoprecipitation (ChIP). H3K27me3 ChIP was performed by Zymo Research (ab6002, Abcam) for AGS, SNU484, IM95 and YCC11, or in-house (07-449, Millipore) for the rest of the samples. Input DNA from cells prior to immunoprecipitation were used to normalize ChIP-seq peak calling. Prior to sequencing, qPCR was used to confirm enrichments of positive and negative control ChIP regions. ChIPed DNA (10 ng) was used for ChIP with DNA sequencing (ChIP-seq) library construction following manufacturer protocols (New England Biolabs). Size distributions of the library

samples were checked using a Bio-analyzer (Agilent Technologies). Libraries were sequenced on Hiseq platform (Illumina). Sequence reads were mapped against human reference genome hg19 using Burrows- Wheeler Aligner (BWA, version 0.7.0)(7). Only high-quality mapped reads (MAPQ$\geqslant$10) were retained for downstream analyses. Potential PCR duplicated reads were removed using SAMTools (8). EBF1 binding sites were predicted using MACS2 (20). Read densities in RPKM units were computed in 50 bp windows for each library using MEDIPS (21). EBF1 motifs were predicted using detected transcription factor binding sites from ChIP-seq and HOMER (22) with default parameters. H3K27me3 ChIP-enriched peaks were detected using CCAT (FDR<5%) (9). Read densities in RPKM units were computed Using the DeepTools bamCompare package (23). Data was visualized using UCSC Genome Browser (24). Sequencing statistics are shown in Supplementary Table 3.

Detection of *EBF1* coding mutations

Exome/whole genome sequencing data of 459 GCs from public and in-house datasets were used to detect *EBF1* coding mutations. This data set includes whole genome sequences (WGS) of 212 GCs and matched normal samples (1), and additional WGS and exome data from 237 GCs from cBioPortal v1.7.3 (25, 26).

Molecular Dynamics (MD) Simulations

Initial structures of the EBF1-DNA complex, 3mlp.pdb(27), was obtained from Protein Data Bank(28). EBF1 DNA binding domain (DBD, a.a. 36-241) and 22 base-pair DNA molecules were adopted in the simulations of this study. Explicit hydrogen atoms and protonation states of titratable residues of EBF1 DBD were determined by PDB2PQR (http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/). MD simulations were prepared using AmberTools package. We adopted ff14SB force fields for the EBF1 DBD and DNA molecules. The Zinc Amber force field (ZAFF)(29) was used to model the zinc binding motif containing Zn^{2+} and zinc-binding residues (His157, Cys161, Cys164 and Cys170). The EBF1 DBD and EBF1 DBD-DNA complex were solvated in virtual cubic boxes filled with TIP3P water molecules. To maintain charge neutralization of the systems, proper numbers of counter ions (Na^{+}) were added into the simulation boxes. To release steric clashes between atoms, we energy minimized the simulation systems. The first 2000 cycles of energy minimization were performed by steepest descent method with the backbone heavy atoms of the EBF1 DBD (and DNA) restrained by a harmonic potential (the force constant was set as 10.0 kcal/mol-$Å^2$) and 2000 more cycles of energy minimization were followed without applying restraints. Periodic boundary condition was employed during dynamics simulations while long range electrostatic interactions were treated by the particle mesh Ewald method and short range nonbonded cutoff was set as 12.0 Å. SHAKE algorithm was used to constrain

bond length to allow a larger time step of 2 fs. To heat the system, we ran 100 ps simulations using time step of 1 fs. During the heating process, backbone heavy atoms of EBF1 DBD (and DNA) were restrained by a harmonic potential with force constant of 10.0 kcal/mol-$Å^2$. Dynamics simulations in this study were in the NPT ensemble. The pressure was coupled to 1.0 bar and temperature was coupled to 298.15 K by Berendsen's coupling methods with both coupling constants of 1 ps. Backbone restraints were relaxed to 5.0 kcal/mol-$Å^2$ for the following 100 ps simulation before data production. For data production, we performed 200ns simulations for all systems with larger time step of 2 fs. The temperature coupling method of production simulations was switched to Langevin thermostat with the collision frequency as 5 ps^{-1} while Berendsen's methods was maintained for pressure coupling.

EBF1 immunohistochemistry

A patient cohort of 109 GC patients treated by surgical resection with curative intent at Leeds General Infirmary, UK, between 1993 and 2004, was analyzed. Studies were performed with the approval of Local Research Ethics committees and were conducted in accordance with the Declaration of Helsinki. None of the patients received any additional adjuvant or neoadjuvant therapy. The median age of the patients was 71.3 years, ranging from 43.3 to 96 years. 64 (58.7 %) patients were male. Haematoxylin/eosin stained sections of the gastrectomy specimens were reviewed and a block with the highest tumour cell density as well as a block with normal gastric mucosa was selected for tissue microarray (TMA) construction. Three 0.6mm cores from the tumour and 3 cores from the normal mucosa were sampled. Immunohistochemistry was performed on four micron TMA sections using the Dako Autostainer Link 48. After deparaffinisation and rehydration of sections using xylene and ethanol, antigen retrieval was performed in Dako DM828 EnVision FLEX Target Retrieval Solution (pH 9) at 97 degrees Celsius over 10 minutes followed by a cool down period of 10 minutes at 80 degrees Celsius. EnVision FLEX peroxidase blocking reagent, mouse (linker), HRP and DAB were used according to the instructions provided by the manufacturer. The mouse monoclonal antibody against full length recombinant EBF1 (Abnova, Taiwan, clone 1C12) was used as primary antibody at a 1:100 dilution, incubation time 20 minutes at room temperature. Only nuclear EBF1 staining was considered as being positive staining. Positively stained lymphocytes served as internal controls. The percentage of positive tumour cells was compared to the percentage of positive normal gastric epithelium cells. Intensity was not recorded separately. Cases were classified as having a higher percentage of positive tumour cells compared to normal epithelial cells, an equal percentage of positive tumour cells and normal epithelial cells or a lower percentage of positive tumour cells compared to normal epithelial cells.

ATAC-Seq

ATAC-seq was performed according to previously published protocols (30). Briefly, 50,000 cells were pelleted at 500 g for 5 minutes. The cells were resuspended in 100 µl ATAC-resuspension buffer containing 0.5% NP40, 0.5 % Tween-20 and 0.01% Digitonin. The cells were left on ice for 5 minutes after which 1 ml of cold ATAC-RSB containing 0.1% Tween 20 was added to wash out the lysis buffer. Nuclei were pelleted at 500 g for 5 minutes, the supernatant was carefully removed and the nuclei resuspended in 50 µl transposition mixture containing 25 µl of 2x TD buffer, 16.5 µl of PBS, 5 µl nuclease free water, 0.5 µl of 1% digitonin, 0.5 µl of 10% Tween-20 and 2.5 µl of transposase (Illumina Tagment DNA enzyme 1, #20034198). The reaction was incubated at 37°C for 30 minutes in a thermoshaker with 1000 RPM. After the reaction, DNA was extracted using the NEB Monarch® PCR & DNA cleanup kit (#T1030L). PCR was performed using the Illumina/Nextera primers after which Ampure XP beads were used for library cleanup. The resulting library was sequenced on a Nextseq. Sequencing reads were mapped to the human genome assembly (hg38) with Bowtie2. Mitochondrial and viral reads were removed. ATACseq peaks were called on each sample using MACS2 (callpeak function with these parameters: -nomodel and -B). BedGraph files generated by MACS2 where normalized to read counts per million reads sequenced.

Electrophoretic Mobility Shift Assay (EMSA)

Wild type and mutated (Q196P, R209Q, R163A, S162A) EBF1 DNA binding domains (amino acids 10-250) were produced as recombinant proteins fused to N-terminal MBP tag by expression in *Escherichia coli* BL21DE3 pLysS, followed by standard amylose affinity purification. Purified proteins were incubated with double stranded FAM-labelled DNA with a EBF1 binding motif (generated by annealing of 5'-FAM-AGAGAGAGAGACTCAAGGGAATTGTGGCC and 5'-GGCCACAATTCCCTTGAGTCTCTCTCTC-3'), in binding reaction mix containing 0.8 µg poly (dI.dC), 150 mM NaCl, 10 mM Tris, 0.1 mM EDTA, 0.1 mM DTT and 5% glycerol. Tris-Glycine native gel electrophoresis was performed on the reactions and the resulting gel was imaged using the fluorescence mode of Syngene Gel doc system. R163A EBF1 has been previously reported to lose DNA binding, while S162A has been reported to retain DNA binding (31). They were used as negative and positive controls for the experiment.

MethylLight assay

1ug of genomic DNA was bisulphite converted using the EZ DNA Methylation-Gold Kit (#D5006) as per manufacturer's instructions. Human Methylated & Non-Methylated (WGA) DNA Set (#D5013) was used as controls. Methyl light assays for the romoter proximal region of *EBF1* was designed using Beacon Designer™ covering 12 CpG sites. Sequences targeting methylated *EBF1* are as

follows: F – 5' CGAGGATTGTTCGAGAGAGGT 3', R – 5' AAACACGACCTCCACGCTAAA 3', Probe - FAM 5 TAAACGCTAAACCCCGACAAATAAAACGCT 3 TAMRA. The methylated *EBF1* assay is specific for bisulphite converted and methylated DNA (data not shown). Control sequences targeting *ACTB* (32) are as follows: F- TGGTGATGGAGGAGGTTTAGTAAGT, R- AACCAATAAAACCTACTCCTCCCTTAA, Probe – FAM 5 ACCACCACCCAACACACAATAACAAACACA 3 TAMRA. Real-time PCR was performed in triplicate and each 20ul reaction contained 1X TaqMan™ Universal Master Mix II (no UNG, #4440040), 0.4uM of forward and reverse primer, 0.2 uM of probe and 1ul of template DNA. Cycling conditions were: 95C for 10 minutes, followed by 45 cycles of 95C for 15 seconds, and combined annealing and extension at 58C for 60 seconds. Delta *Ct* was calculated as *Ct* of methylated *EBF1* – *Ct* of *ACTB*. For samples showing no amplification for methylated *EBF1*, a dummy *Ct* value of 45 was used for calculations.

Western Blotting

Cells were harvested and lysed in RIPA buffer (Sigma) and incubated for 10 min on ice. Protein concentrations were measured using Pierce BCA protein assay (Thermo Scientific). *EBF1* (1:300; sc-137065, Santa Cruz Biotechnology) antibody were used to probe lysates. GAPDH (60,004-1-Ig, Proteintech Group) or Actin (A1978, Sigma) were used as loading controls.

Cell Phenotype Assays

Cell proliferation levels were measured using the CellTiter-Glo™ Luminescent Cell Viability Assay (Promega). Briefly, 1000 cells were seeded in 96-well plate and cell numbers were measured. P-values were derived using Student's t-test with P-values <0.05 considered as statistically significant. For colony formation assays, 5 x 10^3 cells were seeded on 6-well plates with medium. Cells were grown for 2 weeks and the surviving colonies were stained with gentian violet after methanol fixation. Colonies from full image areas of three replicate wells were photographed.

RT-qPCR

Total RNA was extracted from cells using RNeasy Mini Kit (QIAGEN). 200 ng of RNA was then used for RT-qPCR using TaqMan RNA to CT 1-Step kit and TaqMan probes targeting GAPDH, *TERT*, *EBF1* and *EZH2* (Hs02758991_g1 FAM, Hs00972656_m1 FAM, HS01092695_M1, Hs01016789_m1 Applied Biosystems) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). All qPCR experiments were run in triplicate, and mean values were used to determine mRNA levels. Relative quantification was performed using the comparative CT method with GAPDH as the reference gene and with the formula 2- ΔΔCT. Error bars show standard deviation.

Other public data sets

For public data sets, we used processed RNA-seq profiles (version 2, level 3) from the TCGA data portal. Gene expression microarray processed data were used as previously described (33).

SUPPLEMENTARY TABLES

Supplementary Table 1

Relationship between clinicopathological variables and EBF1 expression

	Whole series		EBF1 no loss		EBF1 loss		p-value
	n	%	n	%	n	%	
Tumour histology (Lauren)							
Intestinal	70	64.2	45	80.4	25	47.2	0.001
Diffuse	23	21.1	6	10.7	17	32.1	
Mixed/indeterminate	16	14.7	5	8.9	11	20.8	
Grade of differentiation							
G1/2	46	48.8	31	58.5	15	28.3	0.002
G3	60	51.2	22	41.5	38	71.7	
Lymph node status (pN)							
N0	41	37.6	26	46.4	15	28.3	0.052
N1/2/3	68	62.4	30	53.6	38	71.7	
Gender							
Male	64	58.7	32	57.1	32	60.4	0.733
Female	45	41.3	24	42.9	21	39.6	
Tumour location							
Proximal	30	27.5	19	33.9	11	20.8	0.599
Mid/distal	72	66.1	35	62.5	37	69.8	
Other	7	6.4	2	3.6	5	9.4	
Depth of invasion (pT)							
T1/2	31	28.4	19	33.9	12	22.6	0.194
T3/4	78	71.6	37	66.1	41	77.4	

Supplementary Table 2

Estimation of EBF1 DBD-DNA binding affinity by MM/PBSA method

	Wild type	R111W	I139T	Q196P	R209Q
ΔG (kcal/mol)	-24.8	-22.9	-20.8	-16.3	-0.7
*ΔΔG(kcal/mol)	0	1.9	4	8.5	24.1

*ΔΔG shows the energy difference from wild type.

Supplementary Table 3

Mapping statistics and quality assessment of sequencing libraries

No	Sample Name	Antibody	Total analysable reads	Total peaks (Pair-end sequencing, q<0.05; Single-end sequencing, p<0.00005)
1	AGS (EBF1 overexpression)	EBF1	50,597,877	74,937
2	HS746T	EBF1	31,871,189	638
4	YCC11	EBF1	43,508,023	1,044
7	IM95	H3K27ac	9,175,907	25736
8	KATO-III	H3K27ac	14,621,640	27,397
9	LMSU	H3K27ac	22,440,292	45,363
10	N980097	H3K27ac	20,704,698	6,505
11	SNU484	H3K27ac	21,279,506	73,634
12	SNU719	H3K27ac	12,411,536	33,239
13	T980097	H3K27ac	19,726,715	76,707
14	YCC10	H3K27ac	13,185,846	57,993
15	AGS	H3K27me3	23,344,902	141,810
16	IM95	H3K27me3	26,452,408	197,297
17	SNU484	H3K27me3	27,488,656	10,053
18	YCC11	H3K27me3	27,792,507	171,331
19	AGS (EBF1 overexpression)	Input-(control for EBF1 ChIP)	102,103,640	not applicable
20	HS746T	Input-(control for EBF1 ChIP)	31,607,929	not applicable
21	IM95	Input-(control for H3K27ac)	27,307,530	not applicable
22	KATO-III	Input-(control for H3K27ac)	15,482,488	not applicable
24	LMSU	Input-(control for H3K27ac)	28,591,488	not applicable
25	N980097	Input-(control for H3K27ac)	18,417,501	not applicable
26	SNU719	Input-(control for H3K27ac)	14,304,384	not applicable
27	T980097	Input-(control for H3K27ac)	21,537,734	not applicable
28	YCC10	Input-(control for H3K27ac)	14,817,784	not applicable
29	SNU484	Input-(control for H3K27ac)	25,115,411	not applicable
30	YCC11	Input-(control for EBF1 ChIP)	75,667,150	not applicable
31	AGS	Input-(control for H3K27me3)	31,651,175	not applicable
32	IM95	Input-(control for H3K27me3)	29,481,990	not applicable
33	SNU484	Input-(control for H3K27me3)	27,683,109	not applicable
34	YCC11	Input-(control for H3K27me3)	26,405,070	not applicable

35	AGS	EBF1	39454333	31
36	HGC27	EBF1	29704505	22799
37	AGS_EBF1-Q196P	EBF1	11417189	441
38	AGS_EBF1-R209Q	EBF1	11445469	1282
39	HGC27	INPUT-(control for EBF1 ChIP)	18343209	not applicable
40	AGS_EBF1-Q196P	INPUT	10840874	not applicable
41	AGS_EBF1-R209Q	INPUT	24073117	not applicable
42	AGS	Input (Control for EBF1 ChIP)	41910606	not applicable
43	AGS-DMSO	H3K27me3	48367545	49966
44	AGS-5AZA	H3K27me3	23698612	6714
45	AGS-GSK	H3K27me3	45466334	21
46	SNU484-DMSO	H3K27me3	31017572	7841
47	SNU484-5AZA	H3K27me3	23523434	0
48	SNU484-GSK	H3K27me3	31592530	0
49	AGS-DMSO	Input	29571221	not applicable
50	AGS-5AZA	Input	19586971	not applicable
51	AGS-GSK	Input	21401500	not applicable
52	SNU484-DMSO	Input	18076931	not applicable
53	SNU484-5AZA	Input	20427303	not applicable
54	SNU484-GSK	Input	15025058	not applicable
55	SNU16	H3K27ac	13,420,128	14,399
56	YCC11	H3K27ac	25,975,571	40,455
57	KATO-III	H3K4me1	21,003,705	34,263
58	N980097	H3K4me1	17,101,256	0
59	SNU16	H3K4me1	23,093,541	66,310
60	SNU719	H3K4me1	12,628,496	1,927
61	T980097	H3K4me1	28,898,382	30,551

62	YCC10	H3K4me1	14,183,811	84,056
63	YCC11	H3K4me1	25,504,650	0
64	KATO-III	Input-(control for H3K4me1)	15,482,488	not applicable
65	N980097	Input-(control for H3K4me1)	18,417,501	not applicable
66	SNU719	Input-(control for H3K4me1)	14,304,384	not applicable
67	T980097	Input-(control for H3K4me1)	21,537,734	not applicable
68	YCC10	Input-(control for H3K4me1)	14,817,784	not applicable
69	SNU16	Input-(control for H3K4me1,H3K27ac)	13,335,261	not applicable
70	YCC11	Input-(control for H3K4me1,H3K27ac)	26,370,632	not applicable

REFERENCES

1. Guo, Y.A., et al., *Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers.* Nature communications, 2018. **9**(1): p. 1520.
2. Layer, R.M., et al., *LUMPY: a probabilistic framework for structural variant discovery.* Genome biology, 2014. **15**(6): p. R84.
3. Wang, K., et al., *Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer.* Nature genetics, 2014. **46**(6): p. 573-582.
4. Consortium, G., *The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-660.
5. Chiang, C., et al., *The impact of structural variation on human gene expression.* Nature genetics, 2017. **49**(5): p. 692.
6. Ng, J.H., et al., *In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures.* Dev Cell, 2013. **24**(3): p. 324-33.
7. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-1760.
8. Li, H., et al., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
9. Xu, H., et al., *A signal–noise model for significance analysis of ChIP-seq with negative control.* Bioinformatics, 2010. **26**(9): p. 1199-1204.
10. Liu, C.C., et al., *Distinct responses of stem cells to telomere uncapping—a potential strategy to improve the safety of cell therapy.* Stem Cells, 2016. **34**(10): p. 2471-2484.
11. Hughes, J.R., et al., *Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment.* Nat Genet, 2014. **46**(2): p. 205-12.
12. Sakuma, T., et al., *Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system.* Scientific reports, 2014. **4**: p. 5400.
13. Haeussler, M., et al., *Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.* Genome Biol, 2016. **17**(1): p. 148.
14. Yang, X., et al., *A public genome-scale lentiviral expression library of human ORFs.* Nature methods, 2011. **8**(8): p. 659.
15. Wege, H., et al., *SYBR Green real‐time telomeric repeat amplification protocol for the rapid quantification of telomerase activity.* Nucleic Acids Research, 2002. **31**(2): p. e3-e3.
16. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome biology, 2013. **14**(4): p. R36.
17. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols, 2012. **7**(3): p. 562.
18. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-883.
19. Ooi, W.F., et al., *Integrated paired-end enhancer profiling and whole-genome sequencing reveals recurrent CCNE1 and IGF2 enhancer hijacking in primary gastric adenocarcinoma.* Gut, 2019.
20. Liu, T., *Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein–DNA interactions in embryonic stem cells.* Stem Cell Transcriptional Networks: Methods and Protocols, 2014: p. 81-95.
21. Lienhard, M., et al., *MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments.* Bioinformatics, 2013. **30**(2): p. 284-286.
22. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Molecular cell, 2010. **38**(4): p. 576-589.

23. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.
24. Kent, W.J., et al., *The human genome browser at UCSC.* Genome research, 2002. **12**(6): p. 996-1006.
25. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci. Signal., 2013. **6**(269): p. pl1-pl1.
26. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data*. 2012, AACR.
27. Treiber, N., et al., *Structure of an Ebf1: DNA complex reveals unusual DNA recognition and structural homology with Rel proteins.* Genes & development, 2010. **24**(20): p. 2270-2275.
28. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
29. Peters, M.B., et al., *Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF).* Journal of chemical theory and computation, 2010. **6**(9): p. 2935-2947.
30. Corces, M.R., et al., *An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues.* Nat Methods, 2017. **14**(10): p. 959-962.
31. Fields, S., et al., *The 'zinc knuckle' motif of Early B cell Factor is required for transcriptional activation of B cell-specific genes.* Mol Immunol, 2008. **45**(14): p. 3786-96.
32. Eads, C.A., et al., *MethyLight: a high-throughput assay to measure DNA methylation.* Nucleic Acids Res, 2000. **28**(8): p. E32.
33. Lin, S.J., et al., *Signatures of tumour immunity distinguish Asian and non-Asian gastric adenocarcinomas.* Gut, 2014: p. gutjnl-2014-308252.

FIGURE LEGENDS

Supplementary Figure 1 *EBF1* Expression in Stomach Tissues

(**A**) Distribution of *TERT* expression in GC samples and matched normal samples from in-house RNA-Seq data. Expression values from GC samples (n=18) and matched normal samples were compared and statistical significance was evaluated using Wilcoxon's rank sum test (one-sided). In each box-and-whisker plot, line within box shows median, bounds of the box show upper and lower quartiles, and bars represent minimum and maximum values. Outliers are shown as separated points. (**B**) Spectrum of *TERT* expression in GC lines. 63 cell lines were sorted in descending order of *TERT* expression levels (in units of normalized FPKM), and split by the median into two groups. Red arrows indicate twelve cell lines used for further investigation. (**C**) *EBF1* genes are expressed in many human tissues and cell types. Expression levels (RPKM) of *EBF1* in 8,555 samples (53 normal human tissues from 544 donors) analysed using RNA-Seq data from the Genotype-Tissue Expression (GTEx) Project. Expression in normal stomach is marked by the red box. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range. **(D-E)** *EBF1* expression levels in normal stomach tissues from two independent studies (RNA-seq). **(F)** Taqman-qPCR validation of *EBF1* expression levels using GC cell lines with high or low *EBF1* expression levels. **(G)** *EBF1* expression in normal gastric tissues from TCGA cohort.

Supplementary Figure 2 *EBF1* Promoter Epigenomic Alterations

(**A and B**) si-*EZH2* causes *EBF1* overexpression in AGS (A) and SNU484 (B) cells. (**C**) *EZH2* and *EBF1* expression levels in AGS cells after treatment with EZH2 inhibitors. Control cells were treated with DMSO. (**D**) Average β-values of 12 CpG probes in the 5'UTR and TSS region of *EBF1* for 18 T/N pairs of GCs. Red arrows point to GC cases with significant *EBF1* hypermethylation. Average β-values and Δβ(T-N) are shown. (**E**) Quantitative MethylLight assays were performed to validate *EBF1* promoter methylation levels in GC samples exhibiting *EBF1* hyper-methylation together with matched normal samples. Probes for the *EBF1* promoter proximal region cover 12 CpG sites. *: $p<0.05$, two sided T-test. (**F**) Drug treatments cause *EBF1* overexpression (upper panel) and *TERT* downregulation (bottom panel). GC cells (SNU484) were treated with solo, dual and triple drug-treatment combinations of GSK126, 5-Aza and TSA. Control cells were treated with DMSO. (**G**) Two HDAC inhibitors, vorinostat and TSA, cause similar levels of *EBF1* upregulation (top: SNU484; bottom: AGS). (**H**) DNA methylation analysis of two GC lines before and after 5-aza treatment by quantitative Methylation-Lightning assays. (**I**) H3K27ac enrichment at the *EBF1* locus from GC cell lines with DMSO, 5-aza, and GSK126 treatment. (**J**) H3K27ac and H3K4me1 enrichment and chromatin accessibility (ATAC-Seq) at the *EBF1* locus from GC cell lines with high or low *EBF1* expression. Blue bars highlight predicted enhancers of *EBF1*. Three of these enhancers were found to interact with *EBF1* promoter in at least one cell lines from pcHi-C database. Interactions are represented by arc lines ($p<0.01$, promoter/capture point: green). (**A-F, H**) *: $q<0.05$, two sided T-test with FDR-based multiple testing correction. RT-PCR results are presented relative to *GAPDH* and shown as mean ± SD of technical triplicates. All data are representative of three independent experiments.

Supplementary Figure 3 Negative Regulation of *TERT* by *EBF1* in GC Cells

(**A**) Pooled si-*EBF1* causes reductions in *EBF1* expression (left panel) and *TERT* overexpression (right panel) in YCC11 cells. (**B**) *EBF1* silencing causes *TERT* overexpression in HGC27 and HS746T. Two

individual *EBF1* shRNAs (left panel) resulted in *TERT* upregulation (middle panel). Telomerase activity assays show an increase of telomerase activity caused by overexpression of *TERT* via *EBF1* knockdown (right panel). Telomerase activity was measured using the Telomeric Repeat Amplification Protocol (TRAP). *: q<0.05, two sided T-test with FDR-based multiple testing correction, error bars show standard deviation. (**C**) Telomerase activity assay shows a loss of telomerase activity caused by downregulation of *TERT* via *EBF1* overexpression. (**D**) GREAT analysis demonstrating enriched pathways for *EBF1*-OE affected genes. The top 10 enriched terms ranked by p-value from the original GREAT output are listed. (**A, C**) *: p<0.05, two sided T-test. Error bars show standard deviation. RT-PCR results are presented relative to *GAPDH* and shown as mean ± SD of technical triplicates. All data are representative of three independent experiments.

Supplementary Figure 4 Characterization of EBF1 DNA-binding Domain Mutations

(**A**) Amino acid sequences at regions harbouring *EBF1* missense mutations were compared across 12 species: human, chimp, green monkey, squirrel monkey, mouse, rat, cow, horse, dog, turkey, pufferfish and zebrafish using UCSC genome browser. Yellow bars highlight amino acids with missense mutations identified in GCs. Red colour marks variant amino acids that are different from those of *Homo sapiens*. (**B**) Overexpression of wild type and mutant *EBF1* in SNU484 cells. Western blotting shows equal levels of EBF1 protein abundance after overexpressing wild type and mutant *EBF1*. Empty vector was used as a control. (**C** and **D**) Cell proliferation capacity after overexpression of wild type and mutant *EBF1* in GC cells. Proliferation assays and monolayer colony formation assays show cell proliferation capacity after overexpression of wild type and mutant *EBF1*. Error bars show standard deviation. (**E**) Heavy atom RMSD profiles over 200ns MD simulations of unbound EBF1 DBD (left panel) and EBF1 DBD-DNA complex (right panel). (**F**) Per residue RMSF profiles of the second 100ns MD simulations (from 5000 conformations) of unbound EBF1 DBD (left panel) and EBF1 DBD/DNA complex (right panel). (**G**) *TERT* expression levels in AGS cells overexpression WT EBF1, EBF1 (Q196P), or WT+Q196P. TERT levels were determined using Taqman-qPCR. *: q<0.05, two sided T-test with FDR-based multiple testing correction. RT-PCR results are presented relative to *GAPDH* and shown as mean ± SD of technical triplicates.

Supplementary Figure 5 EBF1 Occupancy Patterns at the *TERT* Locus

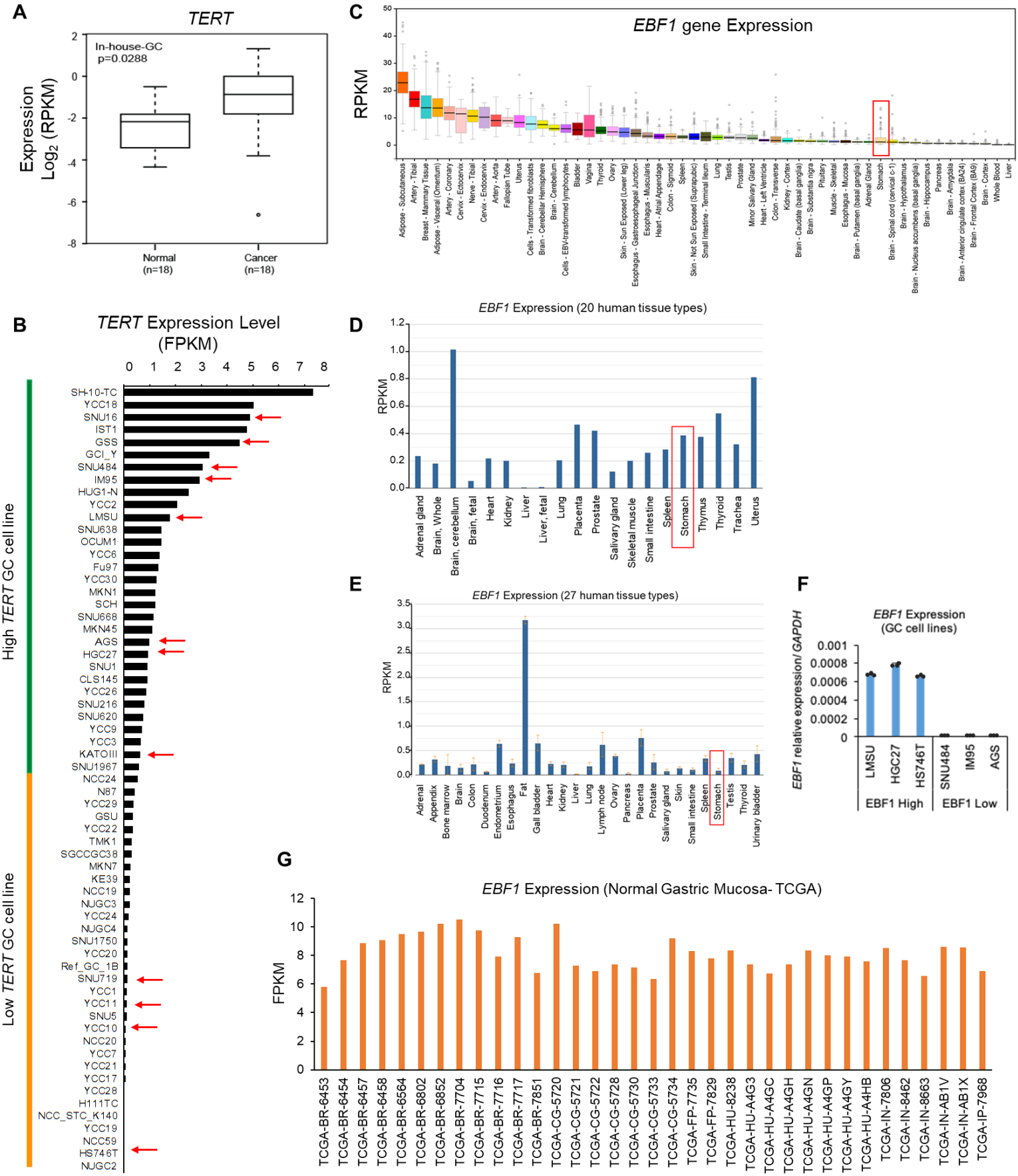
(A) EBF1-ChIPSeq data shows consistent *EBF1* binding to the *TERT* 5'proximal region (highlighted by blue bars) but not other regions in GC cell lines (HGC27, Hs746T, AGS and YCC11 with endogenous *EBF1*, as well as AGS overexpressing WT or mutant EBF1. **(B)** Chromatin accessibility assessed by ATAC-Seq profiling at the *TERT* locus from GC cell lines with high or low TERT expression. *TERT*-high cell lines show more open chromatin accessibility. **(C)** Normalized Hi-C contact maps at the *TERT* locus (Green bar) for GC cell lines with high TERT expression (SNU16), low TERT expression (SNU719), and SNU16/SNU719 interaction ratios (10 kb resolution). Color scales represent normalized contact frequencies between bins.

Supplementary Figure 6 EBF1 Binding Motifs at the *TERT* 5' Proximal Region

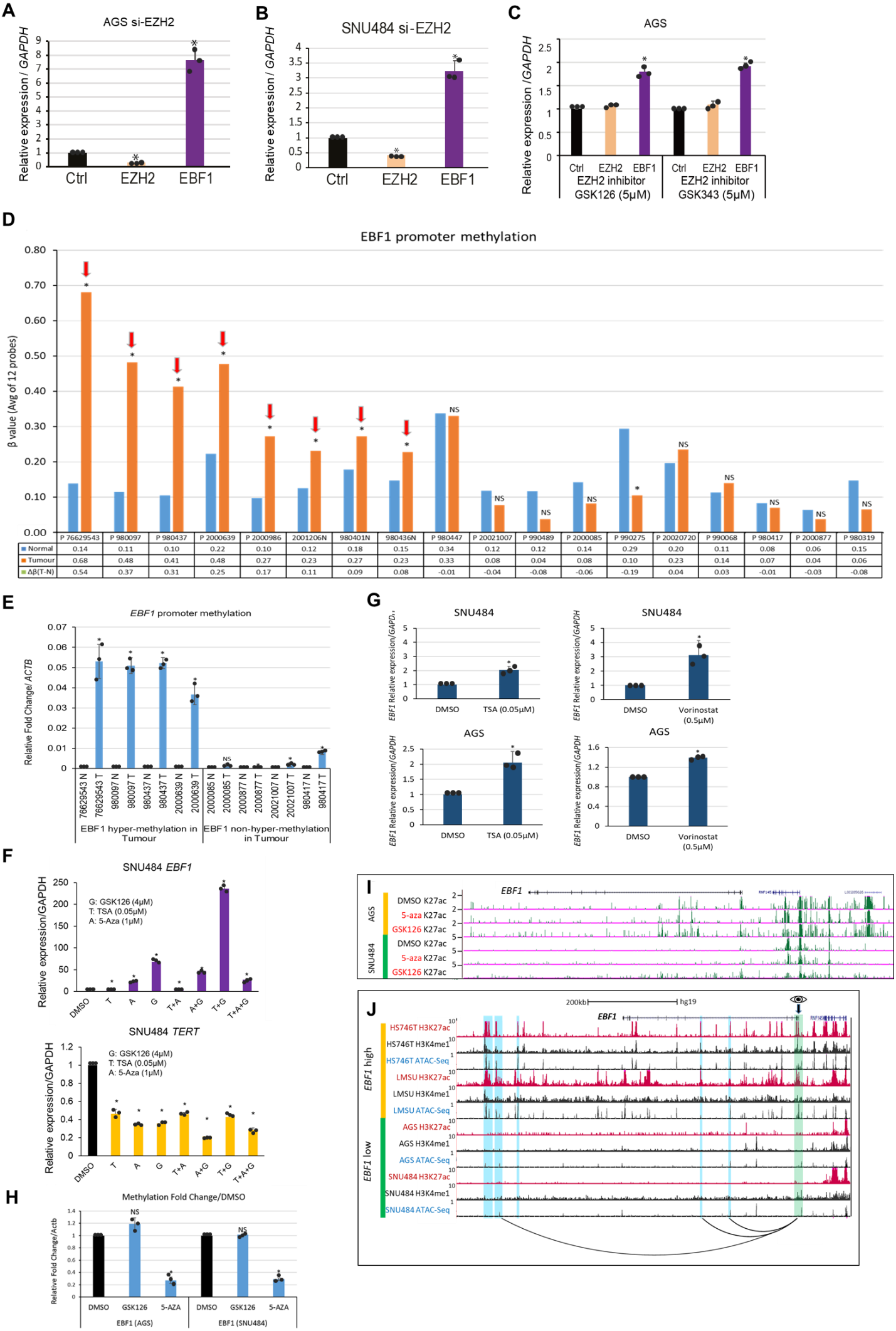
(**A**) Somatic deletion of the *TERT* 5' proximal region in GC HK-PFG144T. Aligned read coverage from WGS profiles of N-PFG144 (normal) and T-PFG144 (GC) are presented as track 1 and 3, while aligned read positions are presented as tracks 2 and 4. Red reads in track 4 have larger than expected inferred

sizes, and therefore indicate predicted deletions. (**B**) Sanger sequencing validates CRISPR-mediated mutations of EBF1 binding motifs in Peak 1 and 2. (**C**) Negative correlation between *TERT* and *EBF1* mRNA level in multiple cancer types. Besides in GC, significant negative correlation between *TERT* and *EBF1* expression level (FPKM) was also found in other cancer types in TCGA consortium (TGCT, Testicular Germ Cell Tumors; PAAD, Pancreatic adenocarcinoma; LIHC, Liver hepatocellular carcinoma; BRCA, Breast invasive carcinoma; THYM, Thymoma; PRAD, Prostate adenocarcinoma). P-values of correlation were calculated using Pearson's correlation test.

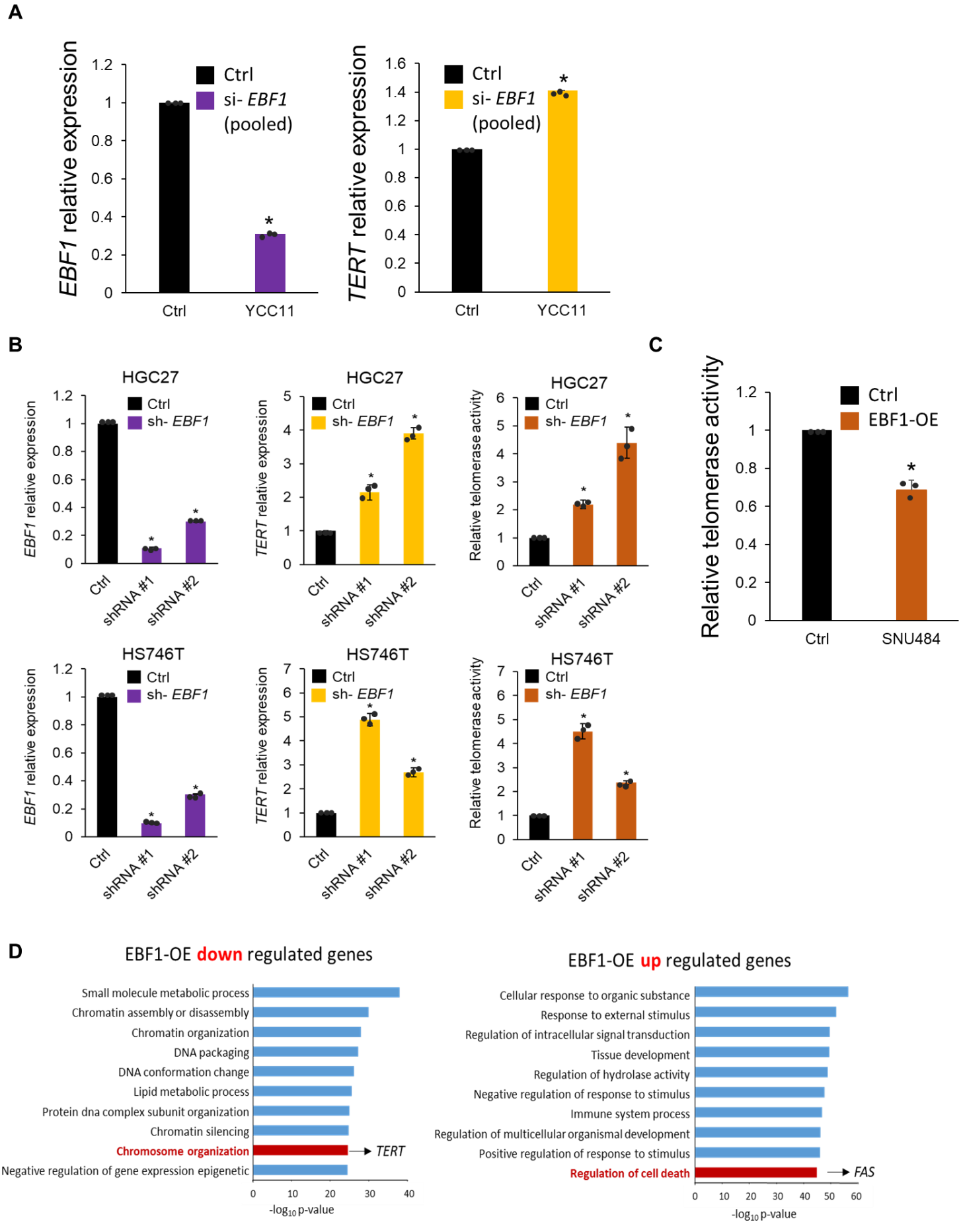
Supplementary Figure 1



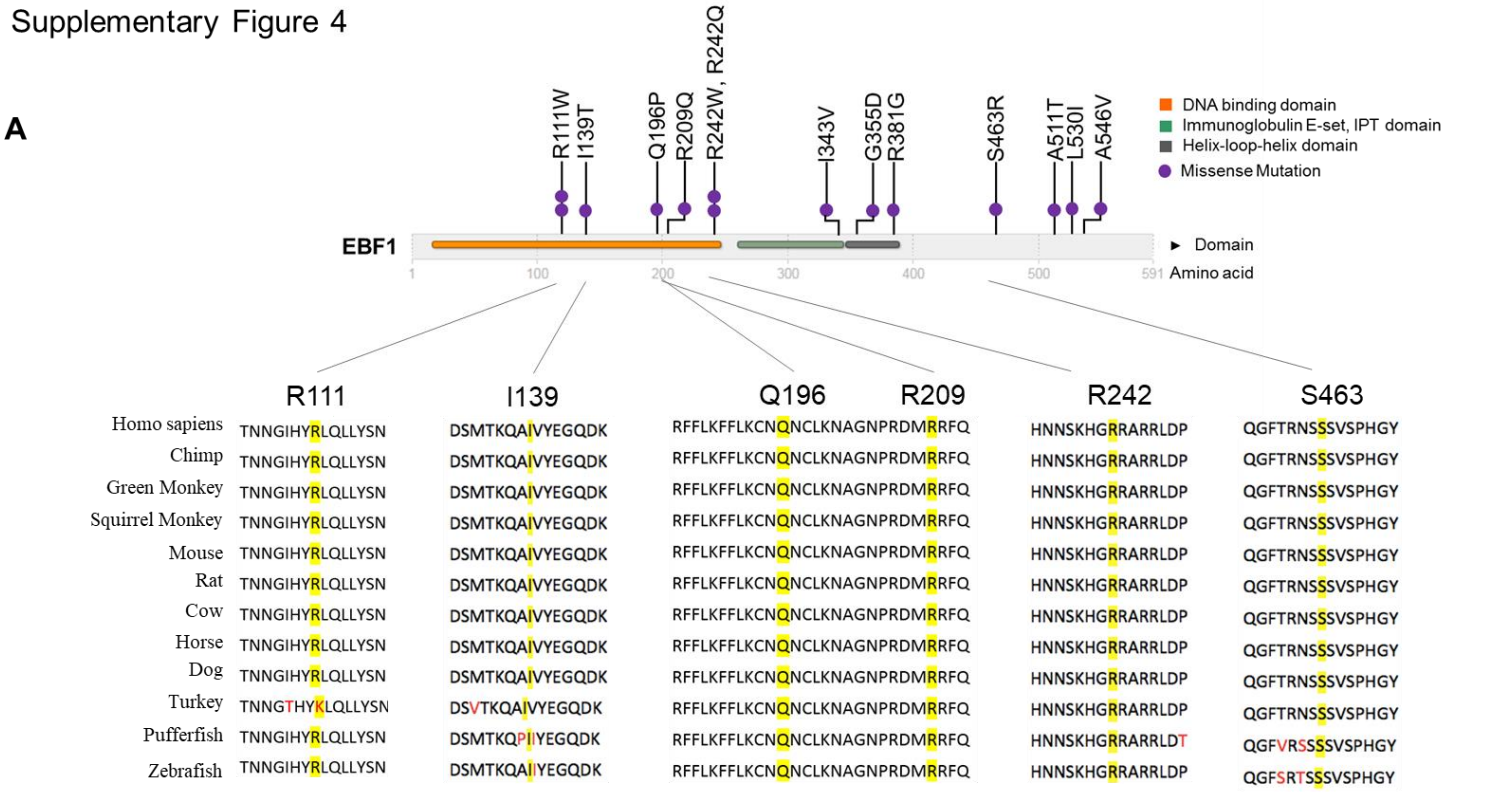
Supplementary Figure 2

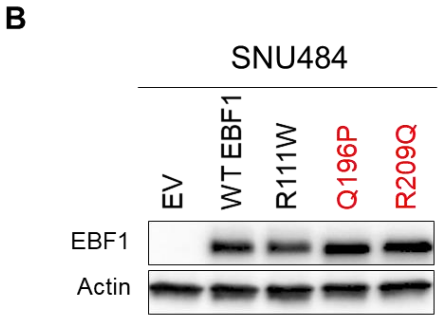


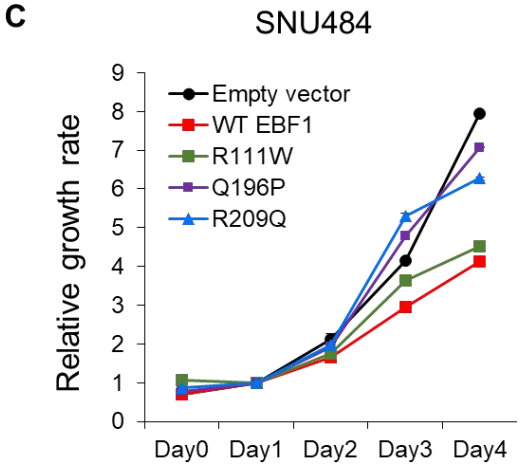
Supplementary Figure 3

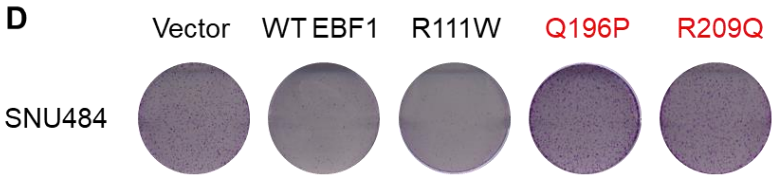


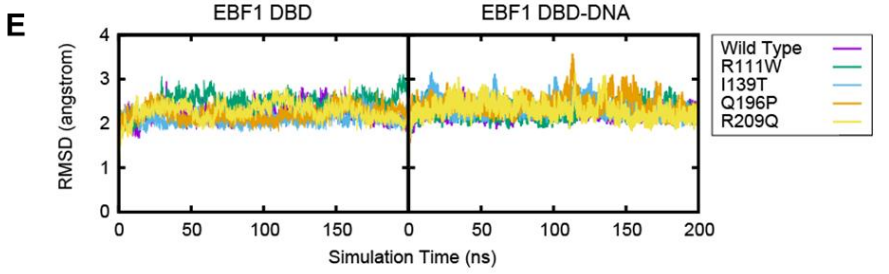
Supplementary Figure 4

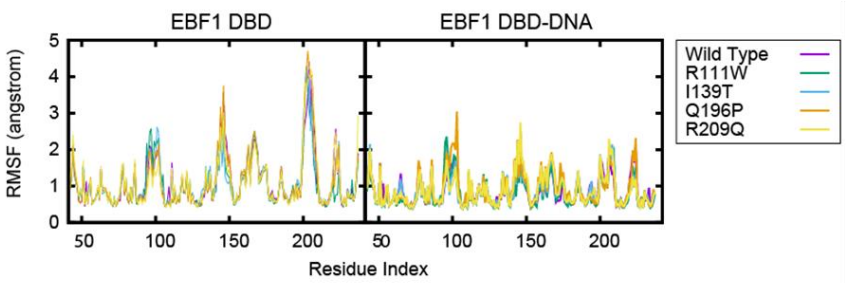


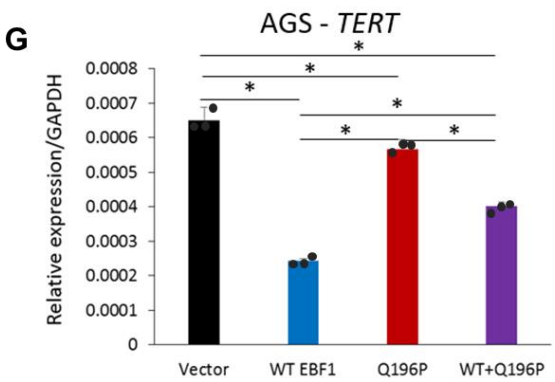




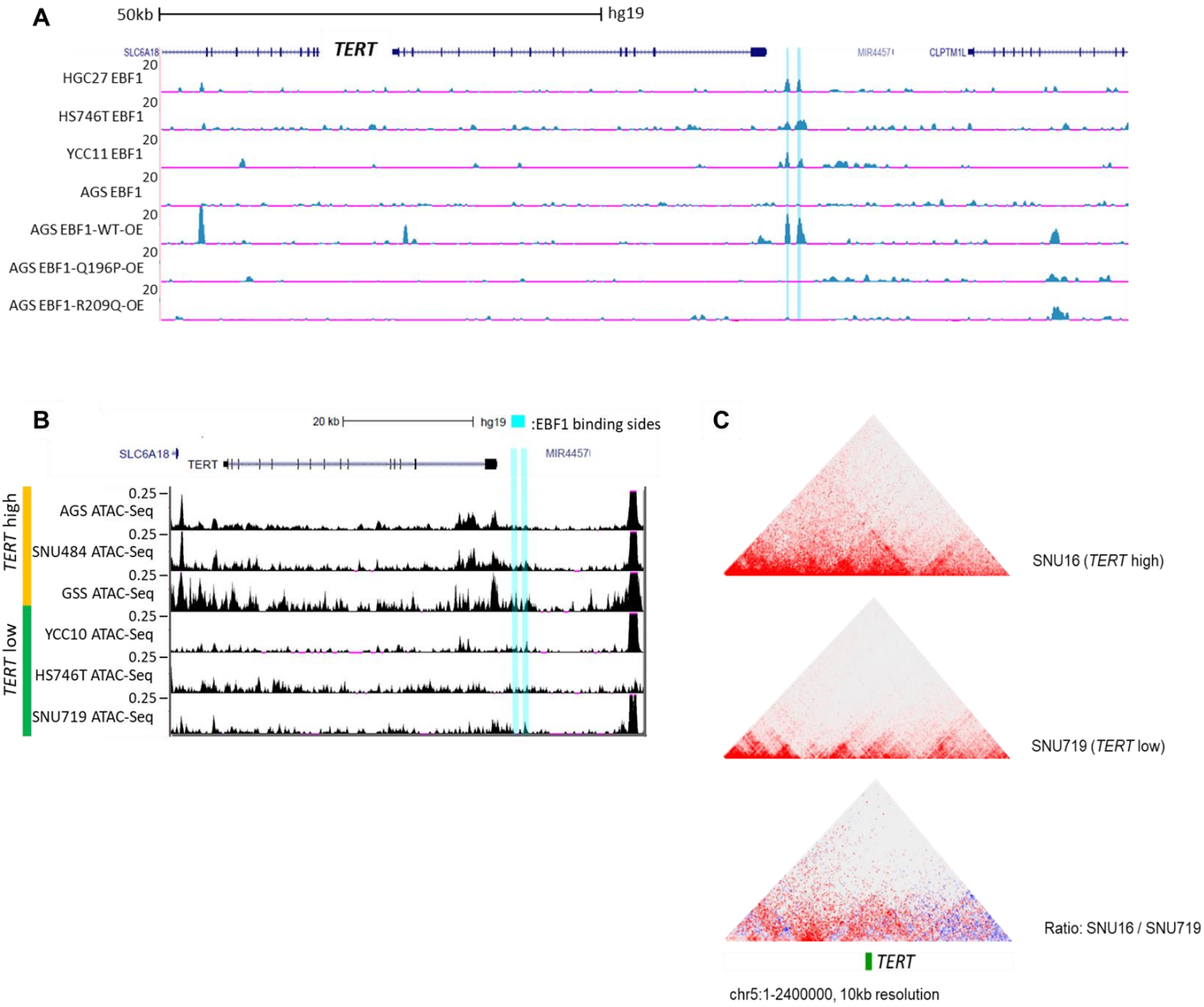








Supplementary Figure 5



Supplementary Figure 6