Supplementary Materials for

The Transcription Factor GABP Selectively Binds and Activates the Mutant TERT Promoter in Cancer

Robert J.A. Bell, H. Tomas Rube, Alex Kreig, Andrew Mancini, Shaun F. Fouse, Raman P. Nagarajan, Serah Choi, Chibo Hong, Daniel He, Melike Pekmezci, John K. Wiencke, Margaret R. Wrensch, Susan M. Chang, Kyle M. Walsh, Sua Myong, Jun S. Song, Joseph F. Costello

correspondence to: songj@illinois.edu, joseph.costello@ucsf.edu

This PDF file includes:

Supplementary Materials:

Materials and Methods:

Sample acquisition

All primary tumor samples were snap frozen in liquid nitrogen and stored at -80° C until use. Patient-matched normal samples were peripheral blood mononuclear cells or muscle tissue. GBM1 – GBM11 samples were obtained from the Neurosurgery Tissue Bank at the University of California San Francisco (UCSF). Sample use was approved by the Committee on Human Research at UCSF and research was approved by the institutional review board at UCSF. All patients provided informed written consent. Snap frozen normal human post-mortem brain tissue from two males (55 and 56 years of age respectively) was obtained from the National Disease Research Interchange (NDRI) and frontal cerebral cortex gray matter was macrodissected.

***TERT* genotyping**

Samples were genotyped for *TERT* promoter mutation status using the Roche GC-Rich Kit. Primers: *TERT_GENOTYPE* (forward: 5'-GTAAAACGACGGCCAGACGTGGCGGAGGGACTG-3'; reverse: 5'-CAGGAAACAGCTATGACAGGGCTTCCCACGTGCG-3'; Tm=60C). Samples were submitted to Sanger sequencing with the following sequencing primers *M13* (forward: 5'-GTAAAACGACGGCCAG-3'; reverse: 5'-CAGGAAACAGCTATGAC-3'; Tm=60C). A list of genotypes for all samples is given in Supplementary table S1.

RNA extraction and qRT-PCR for human primary glioblastomas and normal brain

Total RNA was isolated from frozen tumor tissue or non-diseased postmortem human brain using TRIzol (Life Technologies, Grand Island, NY). Further cleanup and on-column DNase

digestion were performed with the RNeasy kit (Qiagen, Valencia, CA). cDNA synthesis was performed with a mix of random hexamer and oligo dT using Moloney Murine Leukemia Virus Reverse Transcriptase (Life Technologies). Quantitative RT-PCR using iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) was performed with *TERT-1* and *GAPDH* primers (primer sequences given in Supplementary table S4). Melting curves were manually inspected to confirm PCR specificity. Relative expression levels were calculated using the delta delta C(T) method.

Chromatin immunoprecipitation (ChIP)

Primary GBM tissue was used to perform H3K4me3 ChIP as previously described(*24*). 80mg of frozen tissue was digested to mononucleosomes with micrococcal nuclease. Histones marked with H3K4me3 were immunoprecipitated with a monoclonal antibody (Cell Signaling, Cat # 9751) using Sepharose beads coated in protein A/G, and the DNA purified. An IgG negative control was performed in each experiment and enrichment at the *TERT* promoter was determined by qPCR with the following primer sets: *TERT-450* (forward: 5'-CTGATCCGGAGACCCAGG-3'; reverse: 5'-GTCCTCCCCTTCACGTCC-3'; Tm=60C), *TERT-639* (forward: 5'-ACTTGGGCTCCTTGACACAG-3'; reverse: 5'-TTTGGGGTGGTTTGCTCATG-3'; Tm=60C), *TERT-5329* (forward: 5'-TTAGGGCGAAAAATCCCTCT-3'; reverse: 5'-CCAAAGGCGTAAAACAGGAA-3'; Tm=60C. H3K4me3 enrichment between *TERT* mutant and wild-type samples was compared using a Wilcoxon Rank Sum test with the *TERT-450* and *TERT-639* primer sets (P-value < 0.05). ChIP for GABP, ETS1, and ETV4 was performed using the ActiveMotif High Sensitivity kit. GBM1, U251, HepG2, SK-N-SH, A375, MEL-SK-28, and MEL-SK-1239 were grown to 80% confluency in 15cm plates and fixed with 4% formaldehyde. Chromatin was sonicated to a size

range of 200-1500bp by the Diagenode Biorupter. 14-20ug of chromatin was used per GABPA (Santa Cruz Biotechnology: sc-22810), ETV4 (Aviva Systems Biology: ARP32263_P050), ETS1 (Santa Cruz Biotechnology: sc-350) and IgG control (Cell Signaling: 2729) immunoprecipitation for each cell type. Enrichment at the *TERT* promoter was determined by qPCR with the ssoAdvanded Universal SYBR Green Supermix (Biorad). The following primer sets were used: *TERT+47* (forward: 5'-GCCGGGGCCAGGGCTTCCCA-3'; reverse: 5'-CCGCGCTTCCCACGTGGCGG-3'; Tm=74C), *TERT-5329* (above), *STAT6* (forward: 5'-CACTGAGACACGCCAAGAAA-3'; reverse: 5'-AATGGGCTCTGTTTGTGAGC-3'; Tm=60C), *PPP1R12B* (forward: 5'-CAGCTTTAGGGCAAAGGTGA-3'; reverse: 5'-GTGGCCGTAAGGGAACTACA-3'; Tm=60C), *ZNF333* (forward: 5'-TGAAGACACATCTGCGAACC -3'; reverse: 5'-TCGCGCACTCATACAGTTTC-3'; Tm=60C). 1M of Resolution Solution (Roche GC-Rich kit) was added to each standard ssoAdvanced qPCR reaction. PCR was carried out on the Applied Biosystems 7900HT Fast Real-Time System. Three replicate PCR reactions were carried out for each sample.

Cell culture

GBM1, U251, and HepG2 cells were cultured in DMEM/Ham's F-12 1:1 media, 10% FBS, 1% Penicillin/Streptomycin. GBM1 cultured cells were kept under 20 passages following the primary tumor tissue dissociation. A375, MEL-SK-28, MEL-SK-1239, and SK-N-SH were cultured in RPMI-1640, 10% FBS, 1% Penicillin/Streptomycin, and 1% Sodium Pyruvate. All cells were maintained at 37 degrees Celsius, 5% CO_2.

Cell proliferation assay

GBM1 and U251 cells were transfected with siRNA for GABPA and a scramble control from Dharmacon using the standard Dharmafect 1 protocol. Briefly, cells were seeded at a density of 30,000 cells/mL in a 6-well plate. 24-hours post-seeding, cells were transfected with 50nM of siRNA and 2uL of Dharmafect 1 reagent. At t=0, 24, 48 and 72 h post-transfection, cells were trypsinized, collected and counted on a hemocytometer with trypan-blue exclusion.

Flow cytometry

Following transfection of GBM1 and U251 cells with the GABPA siRNA or the scramble siRNA control from Dharmacon, cells were trypsinized and collected at timepoints 0, 24, 48 and 72 h post-transfection. Cells were then fixed with 70% ethanol and stained with a propidium iodide/triton X-100 staining solution containing DNase-free RNaseA. Cells were acquired and analyzed using the Beckman Dickson FACSCalibur flow cytometer. At least 10,000 events were collected for U251 cells and at least 35,000 cells were collected for GBM1 cells. The cell cycle subpopulations were determined using FlowJo, utilizing the Dean-Jett-Fox algorithm.

Luciferase assays and site-directed mutagenesis

The *TERT* core promoter sequence was cloned into the pGL4.10 Promega dual luciferase vector using the Cold Fusion Cloning kit. Site-directed mutagenesis of the *TERT*-pGL4.10 construct was performed using the QuikChange Lightning Site Directed Mutagenesis kit (Agilent Technologies). Cloning and mutagenesis primers are listed in Supplementary table S2. Transfection of reporter plasmid was carried out with the XtremeGene-HP (Roche) transfection reagent. Briefly, GBM1 and U251 cells were seeded at a density of 30,000 cells/mL in a 96-well plate. 24-hours post-seeding cells were transfected with 90ng vector, 9ng pGL4.74 (renilla

control), and 0.3ul of XtremeGene-HP reagent. 48-hours post transfection, firefly luciferase activity was measured by using the Dual-Luciferase Reporter assay system (Promega) and normalized against renilla luciferase activity. All experiments were performed with six replicate wells, and each experiment was repeated at least three times. Each plate was normalized by the positive control pGL4.13 construct. Replicate experiments were scaled to the average wild-type-*TERT*-pGL4 construct.

For each deletion, 18 measurements were organized into a 3x6 table, corresponding to 3 biological replicates and 6 technical replicates. We median polished the table to obtain the overall effect of a given deletion. After removing batch effects by subtracting the row effects, standard deviation around the overall median polished effect was estimated, discarding top two and bottom two outliers. The sinusoidal fitting was performed using the model $a\sin(2\pi(x-b)/10.5)+cx+d$ in Mathematica.

siRNA knockdown and RT-qPCR

siGenome and On-Target Plus siRNA pools were obtained from Dharmacon (siRNAs listed in Supplementary table S3). siRNAs were transfected into GBM1 and U251 cells using the standard Dharmafect1 protocol. Briefly, cells were seeded at a density of 30,000 cells/mL in a 96-well plate. 24-hours post-seeding cells were transfected with 50nM of siRNA and 0.3uL of Dharmafect 1 reagent. At 24, 48, and 72 hours post-transfection, cells were lysed, cDNA was generated, and qPCR was performed to measure gene expression by the POWER SYBR Green Cells-to-Ct kit (Ambion). All experiments were performed in duplicate and each experiment was replicated at least three times. Primers used: *GUSB, TERT2, ETS1, ETS2, ELF1, ELF2, ELF4,*

ETV1, ETV3, ETV4, ETV5, ELK1, ELK3, ELK4, GABPA (Primers are listed in Supplementary table S4). Each sample was measured in triplicate on the Applied Biosystems 7900HT Fast Real-Time System. Melting curves were manually inspected to confirm PCR specificity. Relative expression levels were calculated using the deltaCT method against GUSB. Each plate was internally normalized to no-siRNA control wells. Replicate experiments were scaled to the average scramble-siRNA control values.

RNA-seq gene expression analysis

Level 3 RNA-seq data was obtained for GBM, SKCM, LIHC, and BLCA samples from the TCGA data portal(*34*). Normalized RSEM values were used to generate boxplots of gene expression for each ETS factor. Processed read counts per gene in U251 cells were obtained from Gene Expression Omnibus GSE53220(*35*) RNA-seq from GBM1 and GBM9 cultured cells was generated in-house and analyzed with TopHat and Cufflinks(*36*, *37*)**.** Exome-seq from GBM1 and GBM9 was analyzed as previously described(*38*)**.**

Allele-specific analysis of ChIP DNA and RNA-seq reads

ChIP-DNA from GABPA or H3K4me3 immunoprecipitation and matched input control DNA was used to amplify the *TERT* promoter as described above. PCR products were extracted from an agarose gel using the QiaQuik Gel Extraction Kit (Qiagen). Purified PCR products were subsequently cloned into the pCR2.1/TOPO vector (Invitrogen). At least 12 individual bacterial colonies were amplified with PCR using vector specific primers and sequenced using an ABI 3700 automated DNA sequencer. Allelic bias between ChIP DNA and matched input control

DNA was calculated by Fisher's exact test. Combined P-values were generated by the Stouffer method**.**

To measure expression in GBM1 and GBM9 we identified informative SNPs within the *TERT* gene body by analyzing existing exome-seq data. RNA-seq read pileups were calculated at these positions in the primary tumor as well as matched primary culture cells. The proportion of each allele present in RNA-seq reads was compared to that of matched exome-seq reads by Fisher's exact test. Combined P-values were generated by the Stouffer method.

Buffers and DNA constructs for single-molecule protein binding assay

DNA and GAPBA/GAPBB imaging buffer consisted of 50mM KCL in 25mM Tris (pH 7.5). The ETV4 assay utilized the previous buffer conditions supplemented with 5mM $MgCl_2$. For single molecule imaging, 0.8 mg/ml glucose oxidase, 0.625% glucose, 3 mM 6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic (Trolox), and 0.03 mg/ml catalase were added to the buffers. Oligonucleotides required to make the partial DNA duplex substrates were purchased from IDT with either Cy3 or Cy5 dyes as internal dye modification. Additionally, the CY5 strand contained a 3' biotin modification. DNA constructs were prepared by mixing the Cy3 sequence with the CY5/biotin sequence at a molar ratio of 1:1.5 in 20mMTris-HCl pH 7.5, 5mM MgCl2 and incubating at 95 C for 2 min then slowly cooling to room temperature for 2 hr.

Fluorescent labeling of GABPB and ETV4

GABPB and ETV4 (abNova) were reacted with Alexa647 (NHS-ester) in a 1:20 protein to dye ratio for 1 hour in 100 mM sodium biacarbonate buffer, pH 8.5. Excess dye was removed using a

micro Bio-spin 6 column (Biorad) twice. Alexa647 labeled proteins were used immediately after labeling.

Single molecule fluorescence data acquisition

Single molecule fluorescence experiments were carried out on quartz slides (Finkenbeiner). To minimize surface interactions with the protein, quartz slides and coverslips were coated with polyethylene glycol(*39*). Duplex DNA molecules were immobilized on the PEG-passivated surface via biotin-neutravidin interaction. Excess Cy3 molecules were washed away with reaction buffer. Imaging was initiated before protein mix was applied to capture the moment of protein binding to DNA. GABPA and Alexa647s-GABPB were premixed and incubated 5 minutes before flowing into imaging chamber. GAPBA/GAPBB and ETV4 concentrations were held at 30 nM for all experiments. All experiments and measurements were carried out at room temperature ($23 \pm 1^{o}C$).

Peak motif enrichment analysis

To determine the enrichment of motifs in ETS factor ChIP-seq peaks, we first downloaded all available ENCODE peak annotation files for 10 cell lines (A549, GM12878, Hek293, HeLa-S3, HepG2, HL-60, HTC-116, K562, MCF-7 and SK-N-SH) and five ETS factors (ELF1, ELK1, ELK4, ETS1 and GABPA, see Supplementary table S5) (*40*). To complement this, we also downloaded published ChIP-seq data for ETV1, ETV3, ETV4 and GABPA(*23*, *41*) (see Supplementary table S6) and called peaks using MACS2(*42*), generating both broadPeak and narrowPeak files.

The consensus binding motifs of these ETS factors were determined using MEME (version 4.10.0) (*43*). To identify reliable peaks, we first selected the top 500 peaks in each data set. For the ENCODE sets with replicate peak calls we trimmed this list further by taking the base pair-wise intersection across the replicates. For the non-ENCODE data sets we used the narrowPeak files. MEME was then instructed to identify up to 6 motifs with lengths 6-15bp in each peak set. Motifs corresponding to simple repeats were removed manually.

To determine the relative enrichment of the motifs CCGGAG (WT) and CCGGAA (MUT) in ChIP-seq peaks, we then counted the number $am(i)$ of occurrences of motif type m in each peak i. The local background occurrence $bm(i)$ was estimated by the motif count in peaks shifted one peak length to the left and right. To clarify the individual importance of the two motifs, we discarded all peaks containing the WT motif while calculating the MUT enrichment and vice versa. The peaks were then sorted by enrichment score (column 7 in the ENCODE files) and assigned a rank $r \in [0,1]$, highest enrichment being zero. These ordered counts were smoothed using a Gaussian smoothing kernel of width 500 peaks, giving $am(r)$ and $bm(r)$. The motif enrichment $2am(r)/bm(r)$ was finally plotted as a function of r. This analysis was performed for GABPA, ELF1, ETS1 and ETV4 using the broadPeak files. We also repeated the analysis for ETS1, comparing the consensus motif TGTAGT to the mutated motif TCTAGT, and for ETV4, comparing the consensus motif TGANTCA to the mutated motif TATAGT.

To check if the high enrichment of the MUT motif in GABPA ChIP-seq peaks was due to preferential GABPA binding to open chromatin, we calculated the frequency (per base pair) of

the WT and MUT motifs in HepG2 GABPA and ELF1 ChIP-seq peaks as well as in DNase I hypersensitive sites (see Supplementary table S5). As above, we discarded all peaks containing the WT motif when calculating the MUT frequency and vice versa. The motif frequency was calculated separately for each quintile of the peak enrichment score.

The number of CCGGAA motifs necessary for strong ETS factor binding was determined by first classifying the peaks by the number of motifs occurring within 100bp of the peak center and then plotting the distribution of the Log_{10}-transformed enrichment scores in each class of peaks using a Gaussian smoothing kernel of width 0.05. The ENCODE enrichment scores appeared to be clipped at 3100. Peaks with the palindromic sequence TTCCGGAA were discarded to avoid potential over counting. For peak annotation we used consensus peaks identified by taking the intersection between replicates and averaging the enrichment scores. The distribution of enrichment scores was then calculated for each of GABPA, ELF1 and ETS1 after pooling the consensus peaks across the cell lines A549, GM12878, K562, HepG2 and SK-N-SH (the last two were not available for ETS1).

Spacing dependence of ChIP enrichment

We determined how the distribution of motif separations differs between weak and strong ChIP peaks by first identifying peaks with exactly two motifs within 1kb of the peak center and then, using the peak enrichment score, classifying these peaks as either strong (top quartile) or weak (lower three quartiles). Peaks with overlapping motif combinations were discarded. The distribution of distance between motif centers was plotted for each class of peaks using Gaussian smoothing kernel of width 1.25bp. To calculate the null distribution of spacing, we tiled the

genome with 2kb non-overlapping windows and repeated the analysis, discarding windows overlapping ChIP-seq peaks or genome assembly gaps. We also displayed the theoretical probability density function $2(d-x)/d2$ for the spacing x between two points sampled with uniform probability in d=2kb interval. This analysis was performed both for the GABPA consensus ChIP-seq peaks in HepG2 and for each of GABPA, ELF1 and ETS1 after pooling the consensus peaks across cell lines as above.

To calculate how the enrichment score depends on the motif spacing between the motifs, we first identified all peaks with exactly two motifs within 200bp of the peak center. After sorting these peaks by motif spacing, calculated the median enrichment score for each spacing using a sliding window (in spacing) of width 3bp. A 5% confidence interval of the median was estimated using bootstrap resampling with 10,000 repetitions. This enrichment profile was calculated for each of GABPA, ELF1 and ETS1 after pooling the consensus peaks across cell lines as above.

Allele specific binding from ENCODE data

We downloaded raw ENCODE ChIP-seq and digital genomic footprinting (DGF) reads (see Supplementary table S6) and aligned them to the human reference genome (hg19) using bowtie2(*44*). For each base pair in the reference, we tabulated the occurrence of each nucleotide among the covering reads and then classified the mutation status of the nucleotides at positions 228 and 250. These categorized read-coverage counts were then displayed as bar charts.

Table S1. Summary of the *TERT* mutation status for all samples and cell lines used.

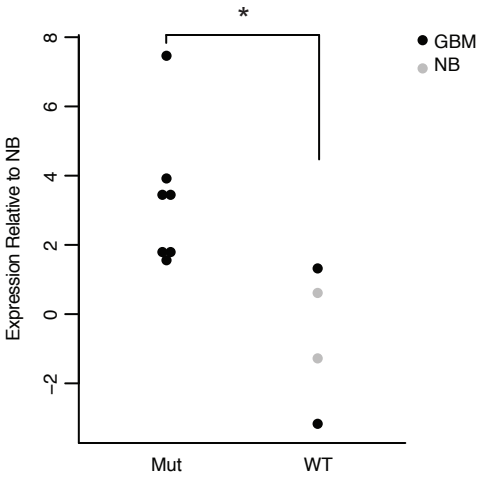
Table S2. Cloning and Mutagenesis primers.

Table S3. siRNAs used for knockdown studies.

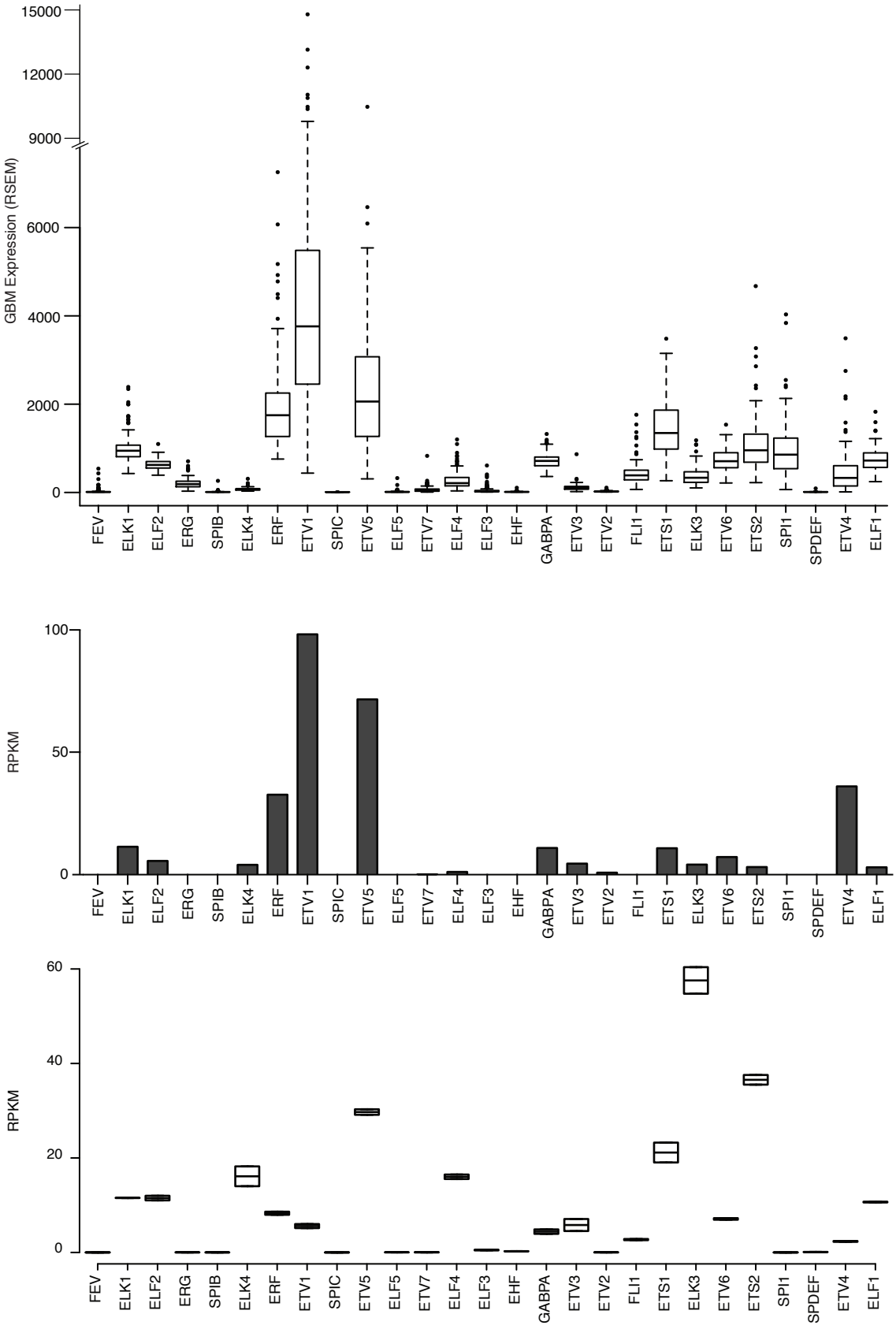
Table S4. qPCR primers used for SYBR green gene expression assays.

Table S5. ENCODE ChIP-seq peak annotations.

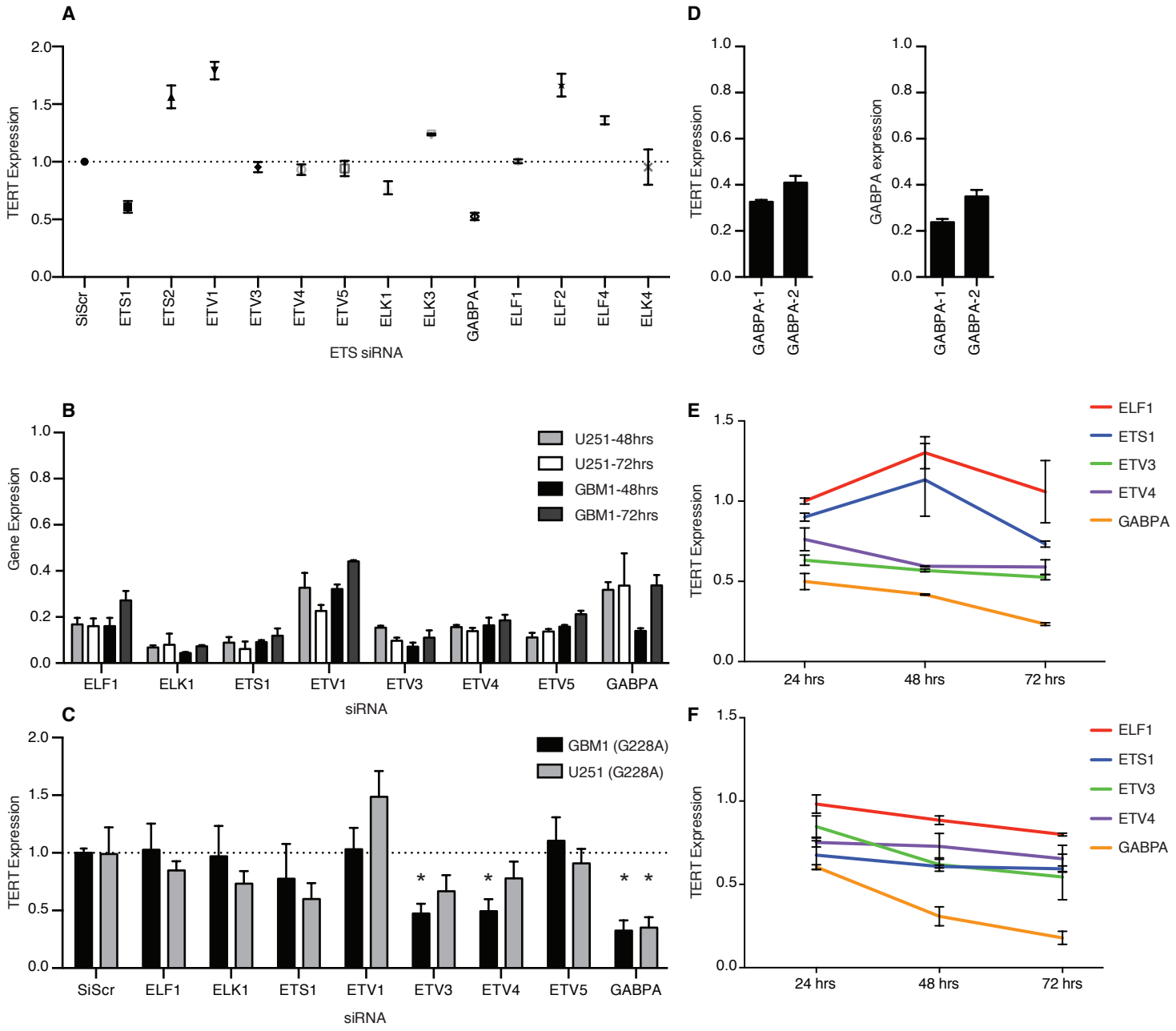
Table S6. SRA ID numbers for raw ENCODE sequencing files.



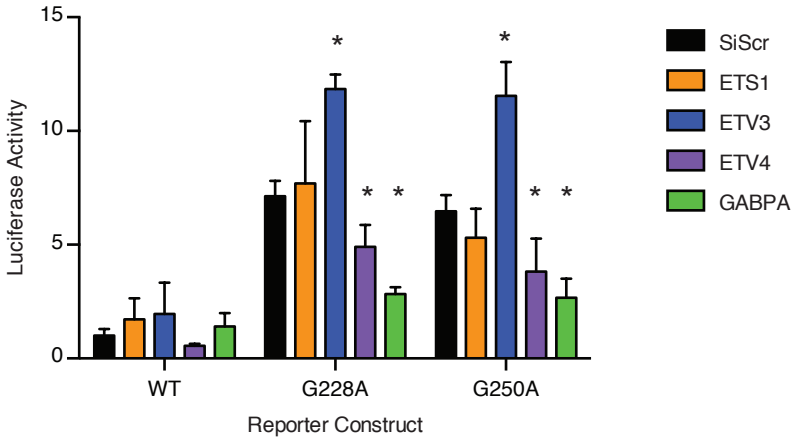
Supplementary Figure 1: *TERT* promoter mutant GBMs display elevated *TERT* expression compared to *TERT* wild-type GBM or normal brain (NB). *TERT* RT-qPCR performed from primary GBM tissue or normal brain tissue. * P <0.05, Welch two sample t-test.



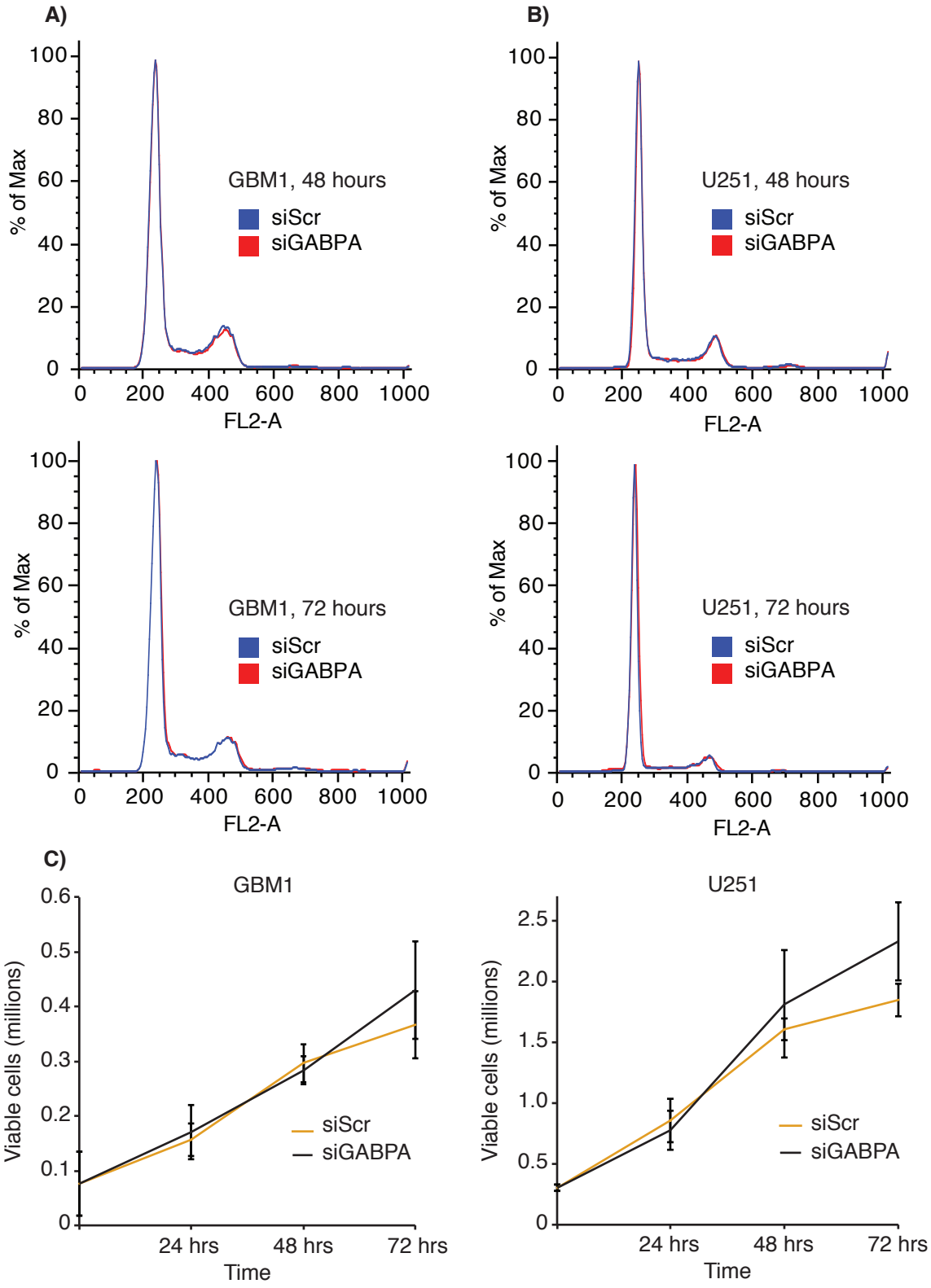
Supplementary Figure 2: ETS family gene expression in GBM. **(A)** Level 3 RNA-seq data was obtained for 169 GBM samples from the TCGA data portal. Box plots are shown for the expression of each ETS family member. **(B, C)** RPKM values of each ETS factor from GBM1 and U251 RNA-seq respectively.



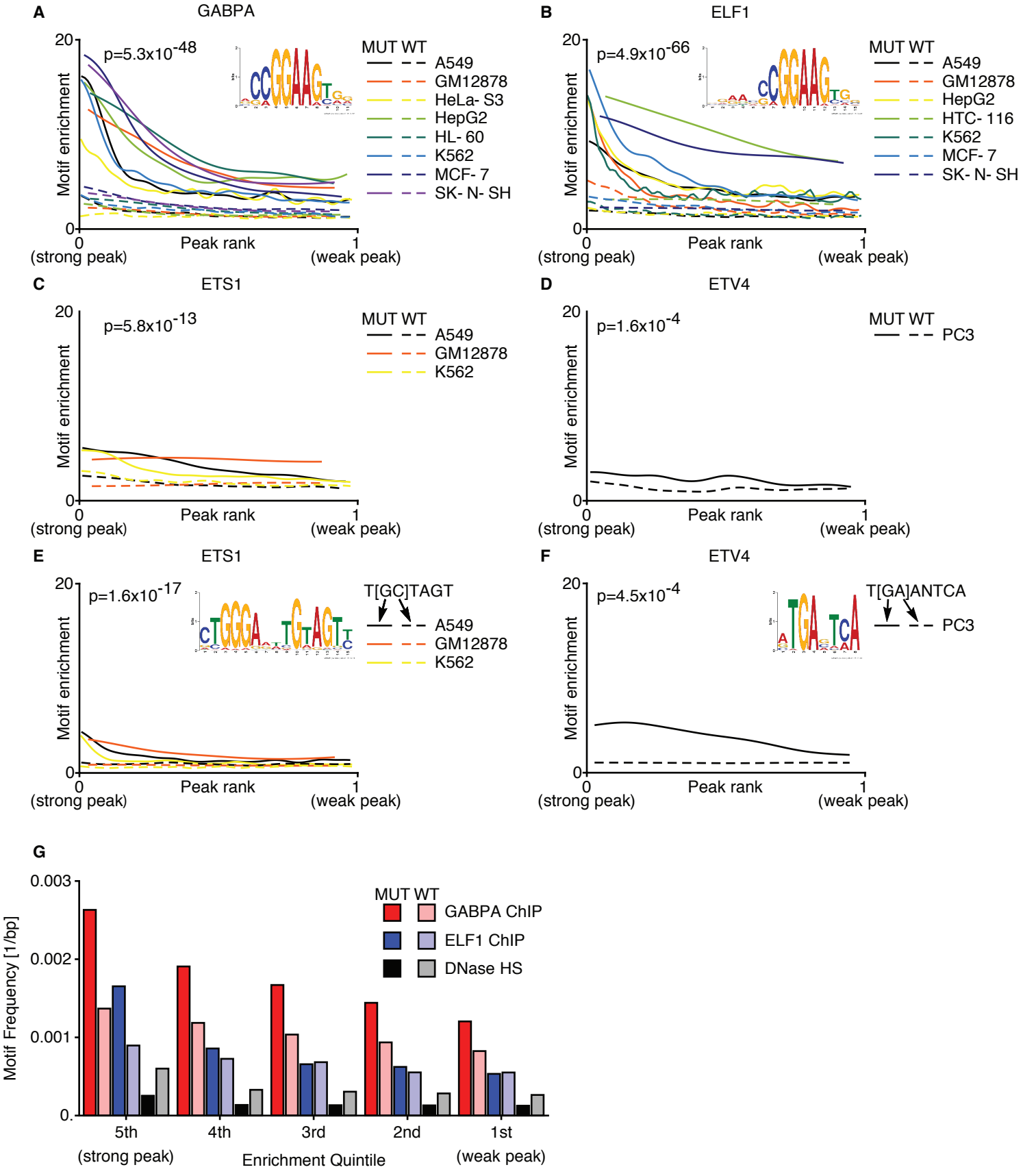
Supplementary Figure 3: *TERT* expression in response to ETS siRNA knockdown. **(A)** A preliminary siRNA screen of 13 expressed ETS factors in GBM. *TERT* expression was measured by RT-qPCR 48 hours post siRNA transfection into U251 cells. 6 ETS factors were removed from further follow-up. ETV1 was retained as an internal negative control. Values are mean ± range relative to siScramble of n=2 replicates. **(B)** ETS transcription factor expression relative to siScramble (siScr) in response to targeted siRNA knockdown. Expression was measured by RT-qPCR at 48hr and 72hr post-transfection. Values are mean ±range from 2 replicate experiments. **(C)** *TERT* expression relative to siScramble 48 hours post siRNA knockdown of 8 ETS factors. * P <0.05, Student's t-test compared to siScr. The results are an average of at least 3 independent experiments. Values are mean ± sd. **(D)** *TERT* and *GABPA* expression relative to siScramble in response to knockdown of GABPA from two independent pools of siRNAs. *TERT* and *GABPA* expression was measured by RT-qPCR at 72hr post transfection with either siGABPA-1 or siGABPA-2 in GBM1 cultured cells. Values are mean ±range from 2 replicate experiments. **(E, F)** *TERT* expression relative to siScramble at 24, 48, and 72 hours post siRNA knockdown of the top 4 ETS factor candidates in GBM1 and U251 respectively. Values are mean ±range from 2 replicate experiments.



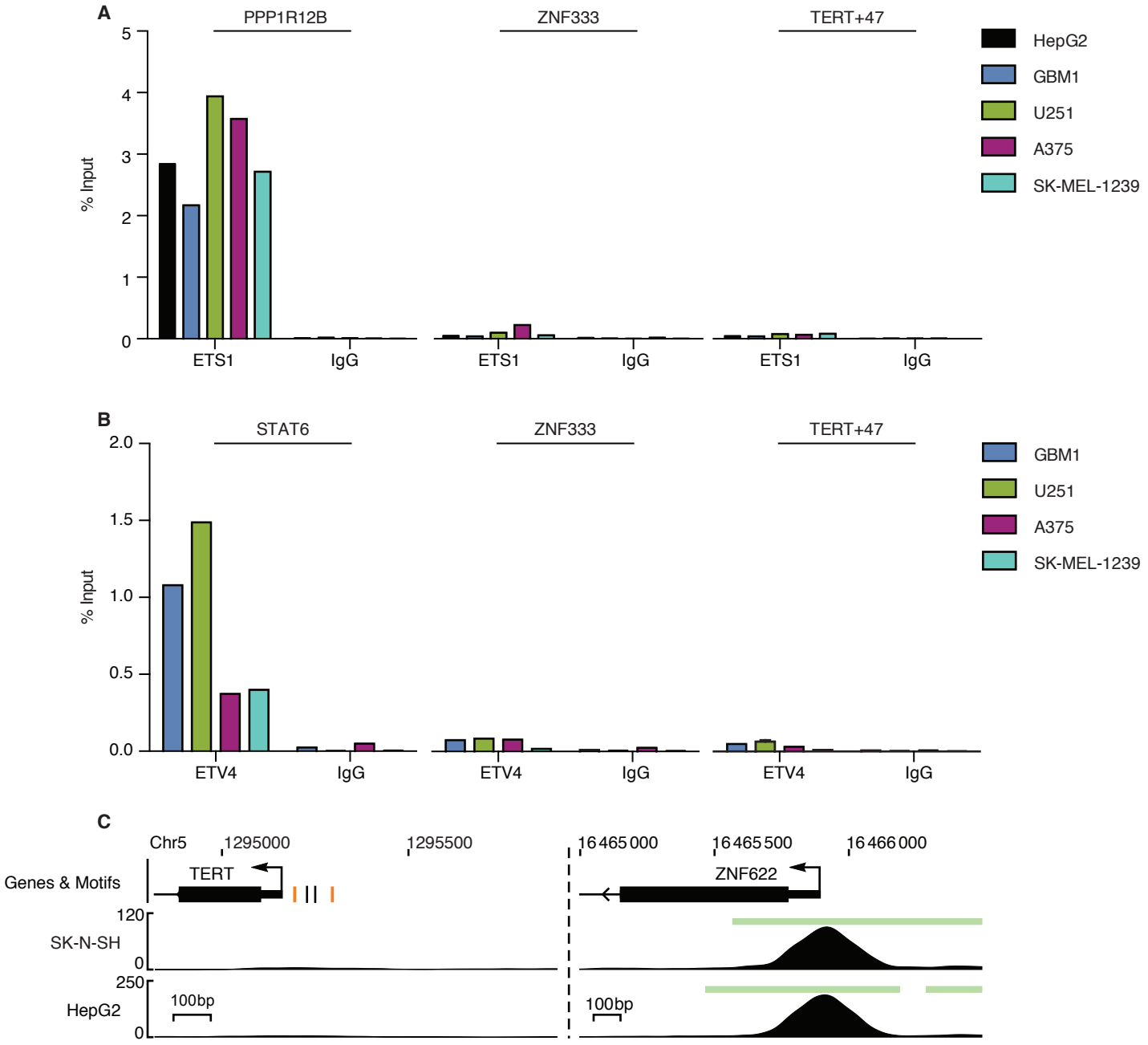
Supplementary Figure 4: GABPA selectively regulates the mutant *TERT* promoter in U251 cells. Wild-type, G228A, or G250A luciferase activity 72 hours post ETS siRNA knockdown in U251 GBM cells, scaled to wild-type siScramble (siScr). The results are an average of at least 3 independent experiments. Values are mean ± sd * P <0.05, Student's t-test compared to siScr.



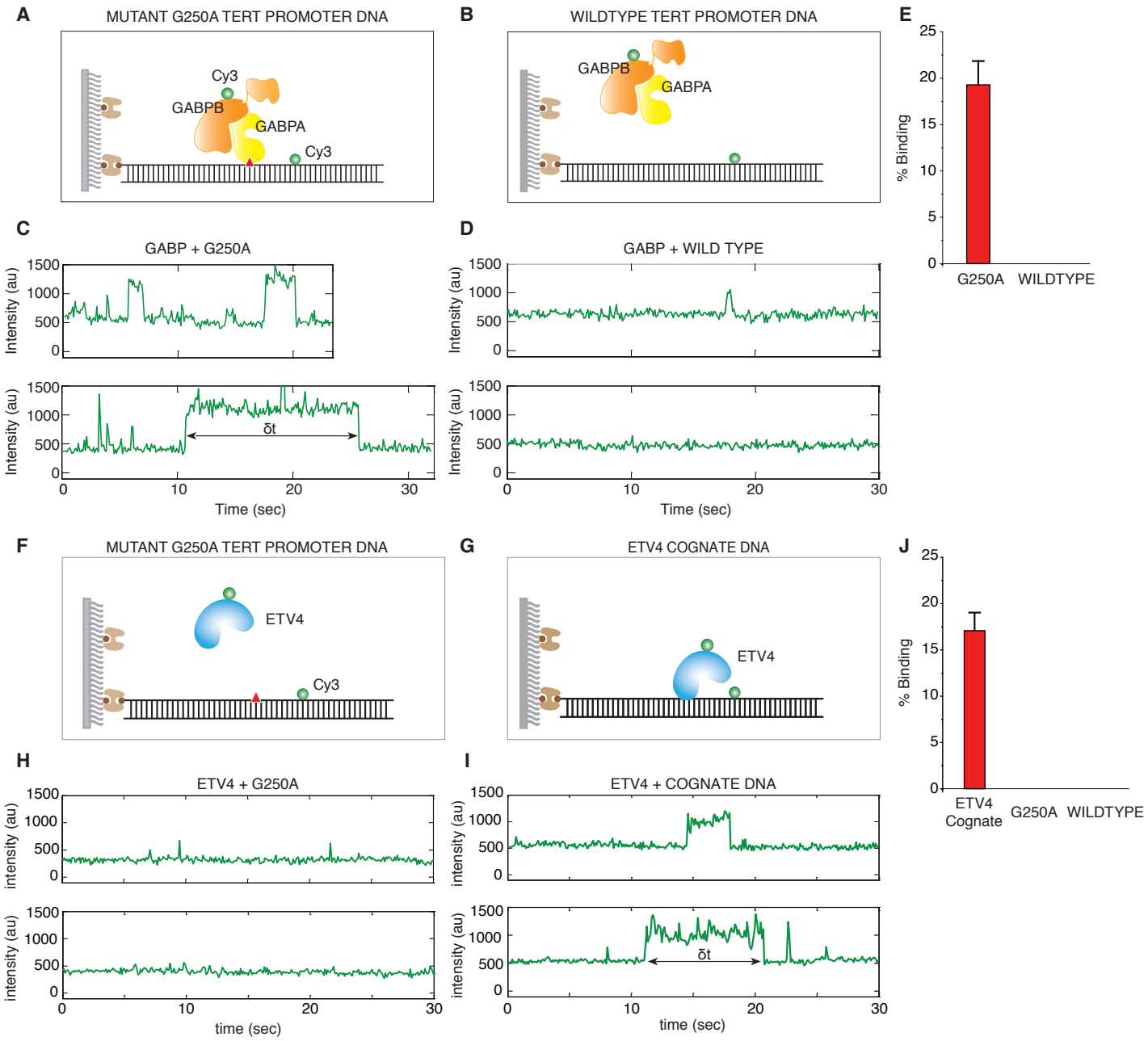
Supplementary Figure 5: Cell cycle profile and proliferation changes with GABPA knockdown in GBM cell lines. **(A, B)** flow cytometry of propidium iodine-stained genomic DNA of GBM1 **(A)** and U251 **(B)** cells fixed and analyzed at 48 h and 72 h post-transfection of GABPA siRNA and scramble control. **(C)** Cell growth curves of GBM1 and U251 generated by counting cells collected at 0, 24, 48 and 72 h post-transfection of GABPA siRNA and scramble control siRNA. Cell counts are presented as the mean value of triplicate wells that were individually transfected, and the error bars reflect s.d.



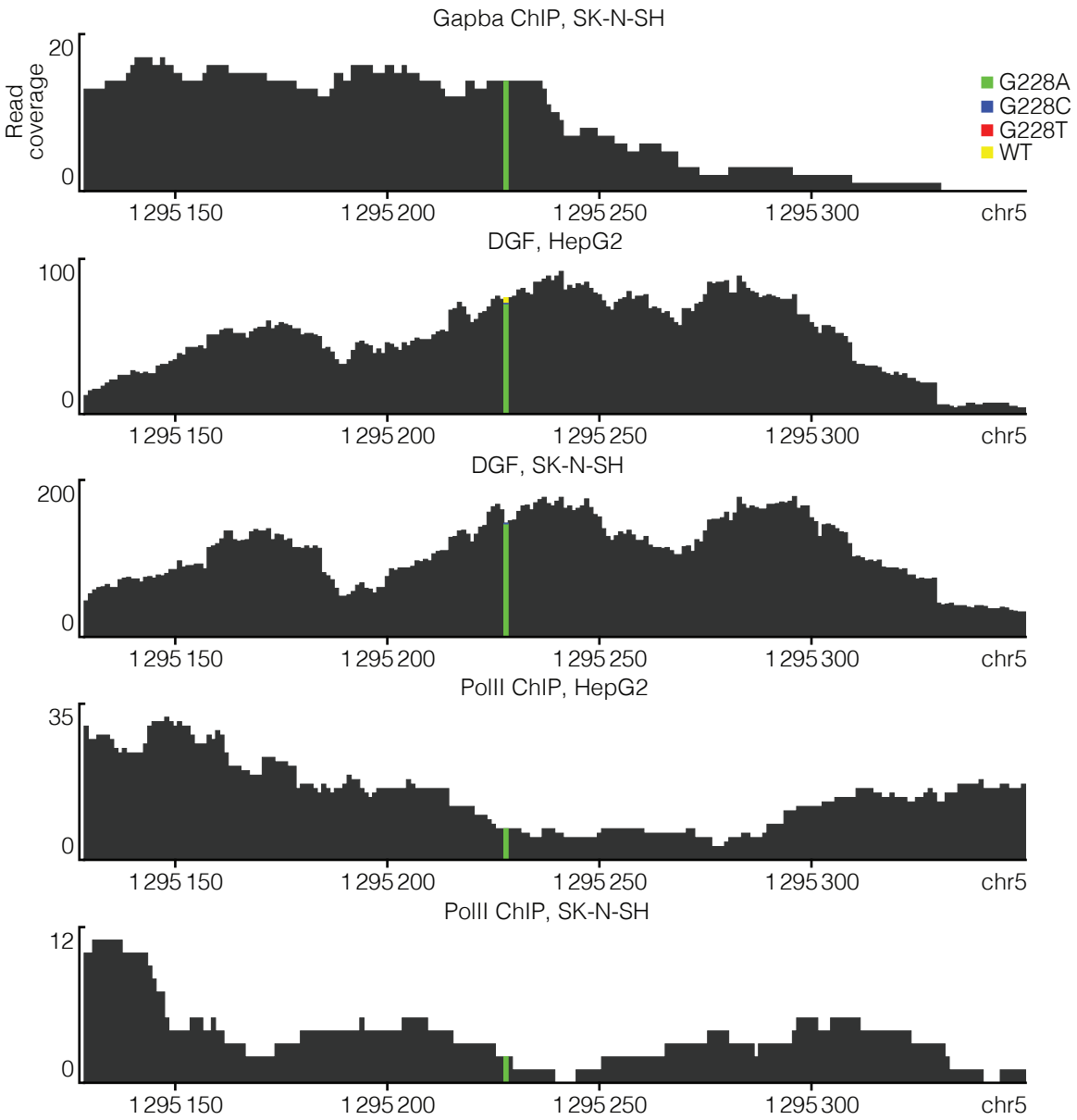
Supplementary Figure 6: Enrichment of the mutant (CCGGAA) and wild-type (CCGGAG) hexamer sequences within genome-wide **(A)** GABPA, **(B)** ELF1, **(C)** ETS1, and **(D)** ETV4 ChIP-seq peaks relative to flanking regions. P-values based on Wilcoxon rank sum test. **(E)** Enrichment of the ETS1 consensus motif and indicated mutated motif in ETS1 ChIP-seq peaks relative to flanking regions. **(F)** Enrichment of the ETV4 consensus motif and indicated mutated motif in ETV4 ChIP-seq peaks relative to flanking regions. **(G)** Frequency (per base pair) of the wild-type and mutant hexamers in HepG2 GABPA and ELF1 ChIP-seq peaks and DNase hypersensitive sites, partitioned by quintile of enrichment score.



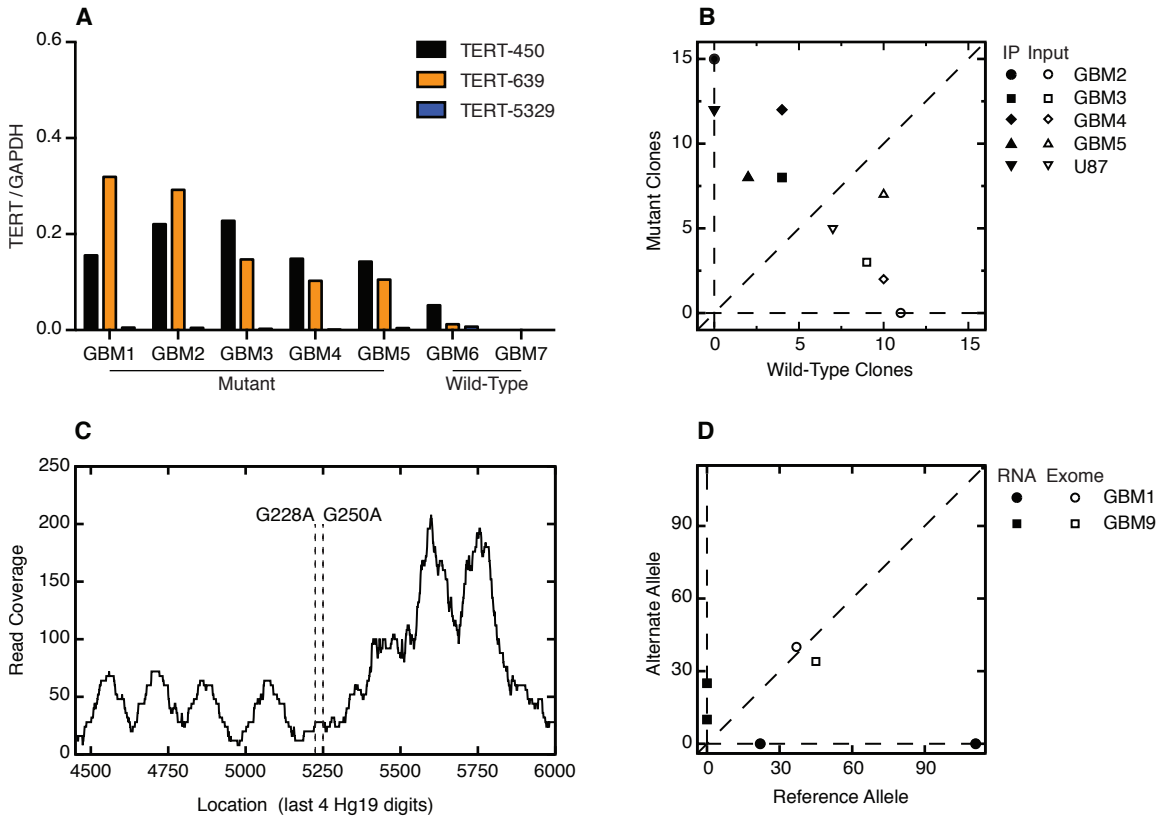
Supplementary Figure 7: Other ETS factor candidates do not bind the mutant *TERT* promoter. **(A)** ETS1 ChIP-qPCR for the *TERT* promoter and a positive and negative control locus in five cancer cell lines. **(B)** ETV4 ChIP-qPCR for the *TERT* promoter and a positive and negative control locus in four cancer cell lines. Values represent mean % input based on triplicate qPCR measurments. N=1 for each cell line. **(C)** ENCODE ELF1 ChIP-seq data at the proximal *TERT* promoter and around a distal positive control locus. Native ETS motifs and mutation positions are annotated by orange and black tick marks respectively.



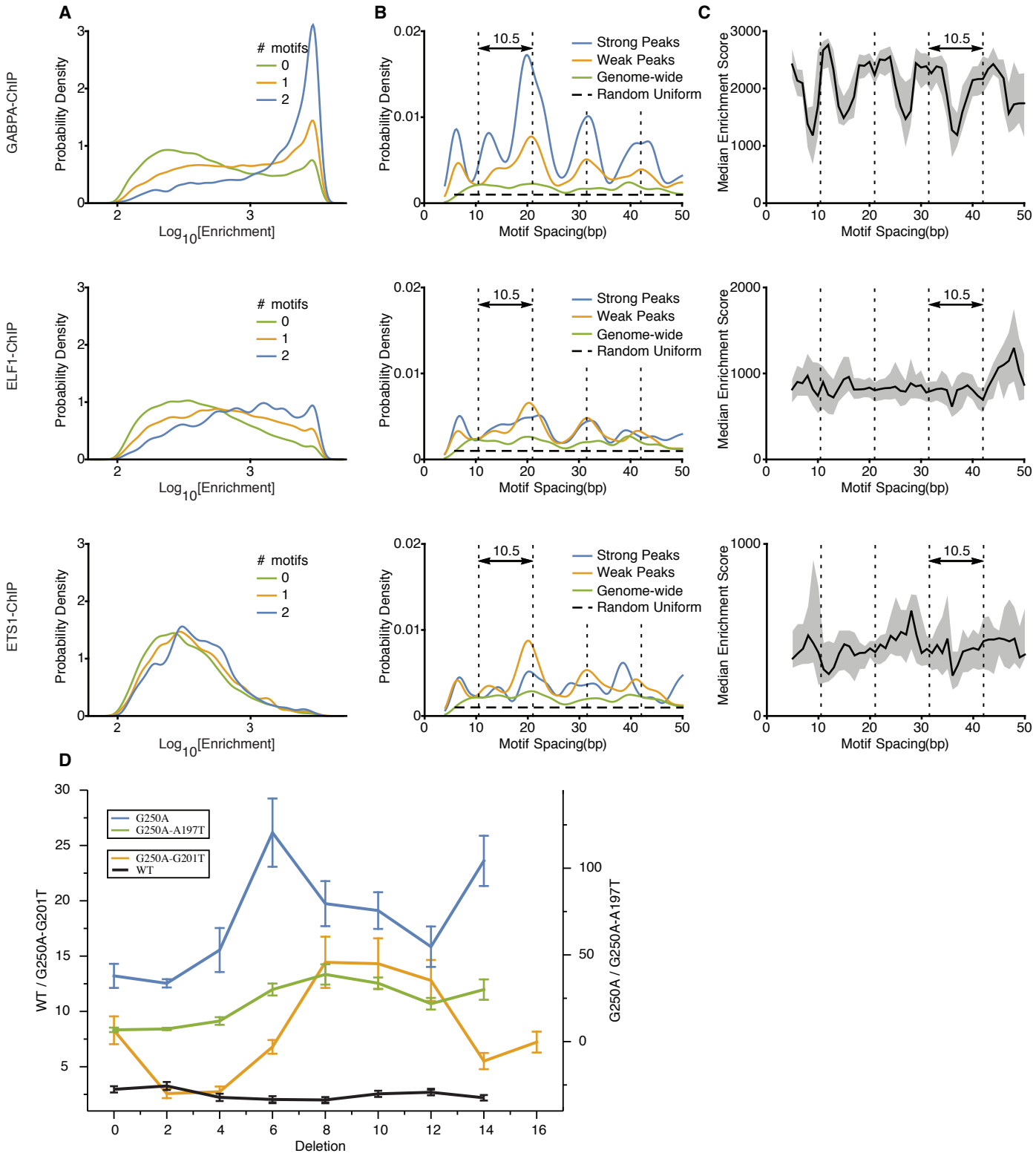
Supplementary Figure 8: GABP but not ETV4 selectively binds to mutant *TERT* promoter DNA. **(A, B)** Schematic of Cy3 labeled GABP complex binding to G250A mutant and wild-type *TERT* promoter DNA sequence, which are also labeled with Cy3. The DNA was immobilized to single molecule imaging surface (PEG-NeutrAvidin-biotin) and premixed GABPA and Cy3-GABPB was applied to the imaging chamber **(C, D)** Representative single molecule traces that show binding of GABP to G250A mutant, but not to wild-type *TERT* promoter DNA. The green signal (~500) arises from Cy3 on DNA in all experiments and the additional green intensity increase (~1200) is seen when Cy3 labeled GABP binds. Binding is not detected when Cy3 labeled GABPB is added in the absence of GABPA, suggesting that the signal increase is due to GABP complex binding. **(E)** The percentage of single molecule traces showing GABP binding in the G250A or wild-type DNA. The average dwelltime (δt) for GABP binding to G250A mutant DNA is 5.04 ± 2.78 seconds (n=255) and no such binding was detected in the wild-type DNA. The results are an average of at least 3 independent experiments. Values are mean ± sd. **(F, G)** Schematic of Cy3 labeled ETV4 binding to G250A mutant or ETV4 Cognate Sequence, which are also labeled with Cy3. **(H, I)** Representative single molecule traces show no binding of ETV4 to the G250A mutant *TERT* promoter. Binding of ETV4 is seen on ETV4 Cognate DNA. **(J)** The percentage of single molecule traces showing ETV4 binding in the ETV4 Cognate, G250A, or wild-type DNA sequence. The average dwell time (δt) for ETV4 binding to ETV4 Positive Control sequence is 5.96 ± 3.08 seconds (n=250) and no such binding was detected in the G250A or wild-type sequence. The average dwell time (δt) for ETV4 binding to ETV4 Positive Control sequence is 5.96 ± 3.08 seconds (n=250)



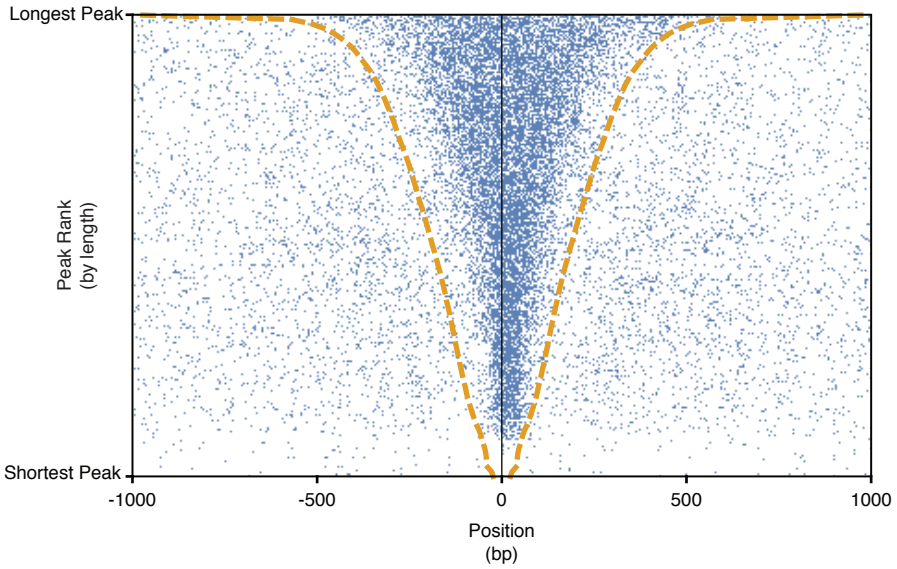
Supplementary Figure 9: Read coverage around the *TERT* promoter mutations in ENCODE GABPA ChIP-Seq, Pol II ChIP-Seq, and Digital Genomic Footprinting (DGF) data from HepG2 and SK-N-SH cells. The read coverage at the G228A mutation is color coded by the observed nucleotides to highlight allele specific binding.



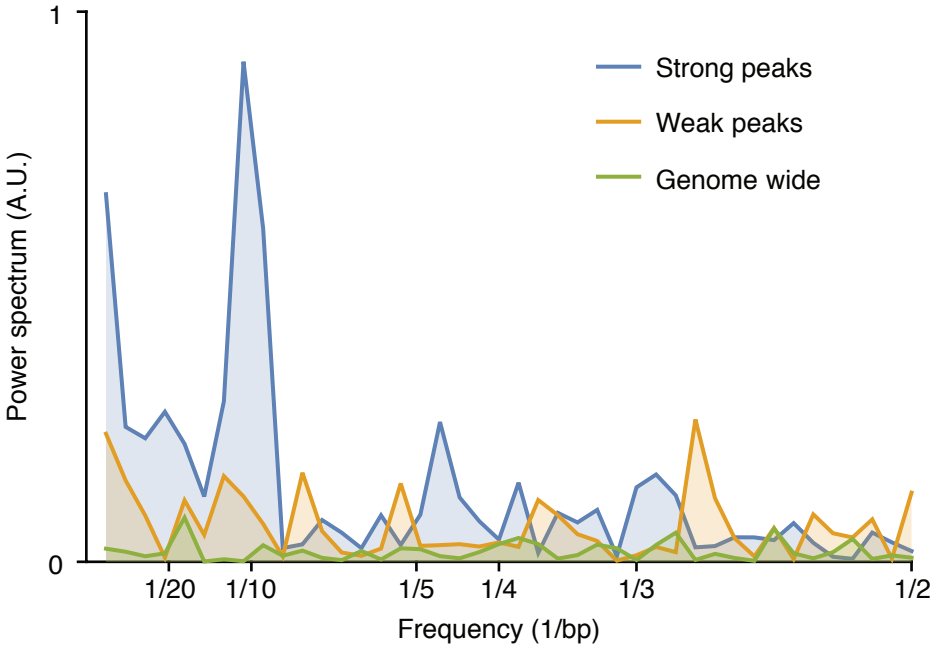
Supplementary Figure 10: GBMs harboring *TERT* promoter mutations display allele-specific H3K4me3 and gene expression. **(A)** H3K4me3 ChIP-PCR of the *TERT* promoter in five *TERT* promoter mutant and two wild-type primary GBMs. Primer set enrichments for the regions TERT-450 and TERT-639 indicated in Fig. 2 were compared between the mutant and wild-type groups (Wilcoxon Rank Sum test. P-Value = 0.0057). Values represent mean enrichment compared to GAPDH positive control primers based on triplicate qPCR measurments. N=1 for each cell line. **(B)** Allelic variant frequency in H3K4me3 (IP) or input control DNA. **(C)** Nucleosome-Positioning prediction from H3K4me3 ChIP-seq in GBM1 **(D)** Allelic *TERT* expression in GBM. RNA-seq and Exome-seq read counts are given for three informative SNPs in the *TERT* gene.



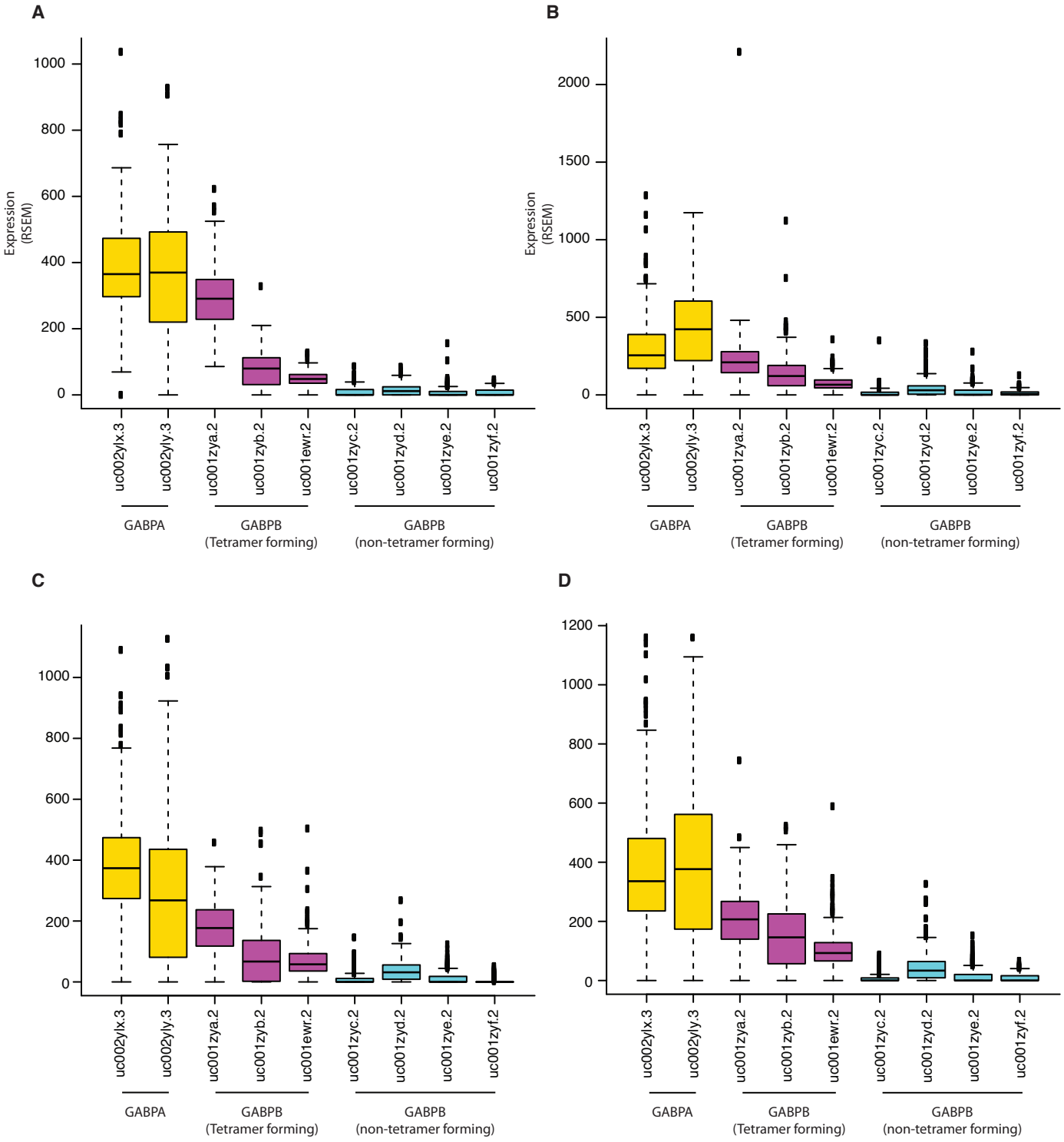
Supplementary Figure 11: Relationship between GABP motif pair distances and binding strength and *TERT* promoter activity. **(A)** Distribution of binding enrichment scores within ChIP-seq peaks containing zero, one, or two CCGGAA motifs. **(B)** Distribution of motif separations in weak and strong ChIP-seq peaks. The dashed black curve shows the theoretical null distribution and green curve shows the genome-wide distribution outside of peaks. Vertical dotted lines indicate 10.5bp periodicity. **(C)** Dependence of ChIP enrichment on motif separation using sliding window of width 3bp. The shaded area indicates 95% confidence interval of the median based on bootstrap resampling. **(D)** Site-directed mutagenesis deleting between 2 to 16 base pairs at the G228A site. Deletions were tested for promoter activity in a Wild-type, G250A, G250A+A197T, and G250A+G201T background. The results are an average of at least 3 independent experiments. Values are mean ± sd.



Supplementary Figure 12: Map of CCGGAA motifs around ENCODE GABPA ChIP-seq peaks in HepG2 cells. The peaks (outlined by dashed orange lines) were sorted by length.



Supplementary Figure 13: Fourier spectral analysis of the GABPA motif spacing distribution for separation between 15 and 100 bps from Figure 3A. Fourier analysis reveals significant periodicity around 10bp intervals.



Supplementary Figure 14: GABPB isoform expression in GBM. Level 3 isoform-specific RNA-seq data were obtained for **(A)** GBM, **(B)** melanoma, **(C)** HCC, and **(D)** bladder urothelial carcinoma samples from the TCGA data portal. Box plots are shown for the expression of each UCSC gene ID of GABPA, GABPB1, and GABPB2. Isoforms that allow for heterotetramer formation (purple) are expressed at higher levels than those that do not allow heterotetramer formation (cyan).