RESEARCH ARTICLE

Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs

Michael Allevato[1], Eugene Bolotin[2], Mark Grossman[1¤], Daniel Mane-Padros[2], Frances M. Sladek[2*], Ernest Martinez[1*]

1 Department of Biochemistry, University of California Riverside, Riverside, California, United States of America, **2** Department of Cell Biology and Neuroscience, University of California Riverside, Riverside, California, United States of America

¤ Current address: Division of Infectious Diseases, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, United States of America
* ernest.martinez@ucr.edu (E.M.); frances.sladek@ucr.edu (F.M.S.)

Abstract

The MYC oncoprotein regulates transcription of a large fraction of the genome as an obligatory heterodimer with the transcription factor MAX. The MYC:MAX heterodimer and MAX:MAX homodimer (hereafter MYC/MAX) bind Enhancer box (E-box) DNA elements (CANNTG) and have the greatest affinity for the canonical MYC E-box (CME) CACGTG. However, MYC:MAX also recognizes E-box variants and was reported to bind DNA in a "non-specific" fashion *in vitro* and *in vivo*. Here, in order to identify potential additional non-canonical binding sites for MYC/MAX, we employed high throughput *in vitro* protein-binding microarrays, along with electrophoretic mobility-shift assays and bioinformatic analyses of MYC-bound genomic loci *in vivo*. We identified all hexameric motifs preferentially bound by MYC/MAX *in vitro*, which include the low-affinity non-E-box sequence AACGTT, and found that the vast majority (87%) of MYC-bound genomic sites in a human B cell line contain at least one of the top 21 motifs bound by MYC:MAX *in vitro*. We further show that high MYC/MAX concentrations are needed for specific binding to the low-affinity sequence AACGTT *in vitro* and that elevated MYC levels *in vivo* more markedly increase the occupancy of AACGTT sites relative to CME sites, especially at distal intergenic and intragenic loci. Hence, MYC binds diverse DNA motifs with a broad range of affinities in a sequence-specific and dose-dependent manner, suggesting that MYC overexpression has more selective effects on the tumor transcriptome than previously thought.

Citation: Allevato M, Bolotin E, Grossman M, Mane-Padros D, Sladek FM, Martinez E (2017) Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs. PLoS ONE 12(7): e0180147. https://doi.org/10.1371/journal.pone.0180147

Editor: Finbarr Hayes, University of Manchester, UNITED KINGDOM

Received: February 21, 2017

Accepted: June 9, 2017

Published: July 18, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. In addition, PBM data are publicly available on the Nuclear Receptor DNA Binding website (http://nrdbs.ucr.edu), Accession number PBM7.1MAX09072010 and PBM7.1MYCMAX09072010.

Funding: This work was supported by grants from the NIH (R01CA158540 to E.M. and R01DK094707 to F.M.S). M.A. was supported by a MARC U-STAR training grant from NIH (T34GM062756).

Introduction

The oncoprotein MYC is an essential DNA-binding transcription factor of the basic helix-loop-helix leucine zipper (bHLHZ) family that regulates transcription of a large number of genes in metazoans and controls cell cycle, cell growth, metabolism, cell proliferation, differentiation, apoptosis, and cell transformation and is overexpressed in many types of cancer [1, 2]. MYC binds DNA as a heterodimer with an obligatory bHLHZ partner called MAX. In contrast to MYC, the MAX protein can also homodimerize. The MYC:MAX heterodimer and MAX:

Competing interests: The authors have declared that no competing interests exist.

MAX homodimer recognize Enhancer box (E-box) DNA elements (CANNTG) and have the highest affinity for the palindromic E-box sequence CACGTG, referred to hereafter as the canonical MYC E-box (CME) [3–9]. Heterodimerization with MAX is essential for most MYC biological functions, including transcription regulation and cell transformation [3, 6, 10, 11]. In contrast MAX:MAX homodimers do not have a transcription regulatory domain and interfere with MYC:MAX functions via recognition of the same E-box motifs [12, 13].

Crystal structures of the bHLHZ domains of MYC:MAX and MAX:MAX complexes (referred to as MYC/MAX complexes hereafter) bound to a CME-containing oligonucleotide showed that three conserved amino acids within the basic region of MYC and MAX make identical contacts with four specific bases located symmetrically on either arm of the palindromic CME sequence CACGTG (base contacts underlined for only one strand) [14, 15]. *In vitro* binding site selection studies showed that MYC/MAX complexes can also bind non-canonical motifs, albeit with reduced affinities, including the E-box sequence CATGTG and several E-box variants such as CACGCG, CACGAG, CATGCG, and CACGTT (non canonical base underlined) [5, 9].

Genome-wide chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) in mammalian cells confirmed that the CME (CACGTG) is the most significantly enriched sequence at MYC-bound loci *in vivo* followed by the variant sequences CACGCG, CATGCG, CACGAG, and CATGTG previously identified *in vitro* [16–19]. Both *in vitro* and *in vivo* studies also indicated that MYC/MAX affinity for core E-box elements is influenced by nucleotides immediately flanking the core hexamer [5, 17]. ChIP-seq analyses further revealed that MYC often co-localizes with CpG islands [19, 20]; in the genome of mouse fibroblasts, for example, about half of MYC-bound sites occur within 1 kb of a transcription start site (TSS), a majority (~75%) of which are CpG islands [17]. The genomic occupancy of MYC:MAX has also been shown to correlate with an active/open chromatin conformation [21, 22] and with the occupancy of the RNA polymerase II (RNAPII) machinery [23]. This is consistent with not only the known interactions between MYC and the RNAPII machinery, other DNA-binding regulators and chromatin-modifying coregulators [1, 24–26], but also with the observation that MYC stimulates the release of pre-engaged and paused RNAPII at many promoters in embryonic stem cells [27]. More recently, it was noted that under certain *in vitro* conditions MYC:MAX complexes display non-specific DNA-binding activity and that MYC binding to chromatin *in vivo* does not generally correlate with the presence of high affinity motifs, including the top 12 motifs identified *in vitro* [23]. This led to the proposal that MYC genomic occupancy is largely independent of sequence specificity and that MYC:MAX complexes are primarily recruited by other chromatin-associated factors and/or epigenetic marks [23, 25, 28].

In this study, in order to address the issue of nonspecific DNA binding by MYC/MAX, we used recombinant MYC/MAX complexes in a protein-binding microarray (PBM) approach with which we could examine binding to 5000 unique DNA sequences, without interference of other host factors. We show that MYC/MAX complexes bind in a sequence-specific manner to not only high-affinity E-boxes and close variants but also to more degenerate and lower affinity non-E-box DNA sequences, such as the palindromic hexamer AACGTT, which we confirmed by electrophoretic mobility-shift assays (EMSAs). Bioinformatic analyses of published ChIP-seq datasets indicated that the vast majority (87%) of MYC-occupied genomic loci in human cells contained at least one of the top 21 MYC-bound motifs identified in the PBM, including the low-affinity AACGTT hexamer. Furthermore, we found differential effects of MYC expression levels on binding to high-affinity CME versus low-affinity non-E-box AACGTT motifs as well as promoter versus distal regions. In addition, the top biological processes preferentially regulated via MYC-bound CME genes (e.g., ribosome biogenesis, metabolic processes) are not enriched for in genes with MYC-associated AACGTT motifs. Hence, we propose that low-affinity binding sites could play an important role in conveying specific MYC functions under conditions of overexpression, which occur during transformation to a cancerous state.

Materials and methods

Recombinant MYC/MAX complexes

Recombinant six histidine (6his)-tagged full-length human MYC (p64; UniProtKB P01106-1) and MAX (p21; UniProtKB P61244-2) proteins were expressed in *E. coli*, purified and reconstituted to form functional MYC:MAX heterodimeric complexes, essentially as previously described [29], with the exception that RosettaBlue(DE3)pLacI cells were used to achieve higher expression of recombinant 6his-MYC. MYC:MAX complexes were verified to be devoid of MAX:MAX dimers by electrophoretic mobility shift assay (EMSA).

MYC/MAX protein binding microarray (PBM)

Protein binding microarrays (PBMs) were carried out essentially as previously described [30] except that purified recombinant MYC/MAX proteins complexes were used. A custom-designed array was ordered from Agilent and contained 15,000 spots of single-stranded DNA probes with ~5000 unique DNA sequences each spotted in triplicate as 12-nucleotide test sequences and attached to a glass slide (designed as in Fig 1A). The variable 12-nucleotide sequence was located between the linker (`5′- CGTCGATATAGTAATCTTAGCTATTAA-3′`) and the cap (`5′-GCCGG-3′`). All the unique DNA sequences of the array are provided in supplemental S1 Table. The DNA probes were made into double-stranded duplexes using a primer to a common linker sequence, dNTPs (GE Healthcare) and Thermo Sequenase (Affymetrix, Cat# 78500). The slides were subsequently blocked for 90 min with 2% non-fat milk in PBS plus 0.01% Tween 20, washed three times (all washes were done with PBS plus 0.01% Tween 20 unless otherwise specified) and then incubated with 550 ng MAX:MAX or MYC: MAX complexes in binding buffer (15 mM Tris-HCl pH 7.5, 100 mM KCl, 0.075% NP40, 7.5 mM beta-mercatoethanol, 375 ng/μl BSA and 8 ng/μl poly(dA–dT):(dA–dT) for 60 min. After three washes the slides were incubated overnight with a primary antibody to MAX diluted 1:200 (rabbit anti-MAX C-17X, Santa Cruz), washed three times and then incubated with donkey anti-rabbit Dylight 549 (Jackson ImmunResearch) at 1:50 dilution for 60 min. Primary and secondary antibodies were diluted in 2% milk plus 0.01% Tween20. All reactions were carried out at room temperature. After washing three times, plus a final wash in PBS, the slide was dried and scanned at 543 nm using a GenePix Axon 4000B scanner (Molecular Devices, Sunnyvale, CA) at the UCR Genomics Core Facility. Extraction and normalization of the data were as described previously [30]. MEME (meme-suite.org) was used for unbiased motif discovery and generation of position weight matrices as sequence logos [31]. PBM data are publicly available on the Nuclear Receptor DNA Binding website (http://nrdbs.ucr.edu), Accession numbers PBM7.1MAX09072010 and PBM7.1MYCMAX09072010.

Electrophoretic mobility-shift assay (EMSA)

Complementary pairs of oligonucleotides consisting of the Consensus MYC E-box (CME), Non-E-box (NE), derivative sequence variants, and Control (Ctrl.) sequences were purchased from Eurofins Genomics, annealed and labeled by Klenow filling with [α-32P]dGTP to a specific activity of ~4 × 10^5 cpm/pmol. DNA-binding with recombinant MYC/MAX complexes were performed in 10-μl reactions containing 0.5 pmol labeled probe, 80 ng poly(dA–dT): (dA–dT), 375 ng/μl BSA, 15 mM Tris–HCl (pH 7.9), 100 mM KCl, 15% glycerol, 0.15 mM

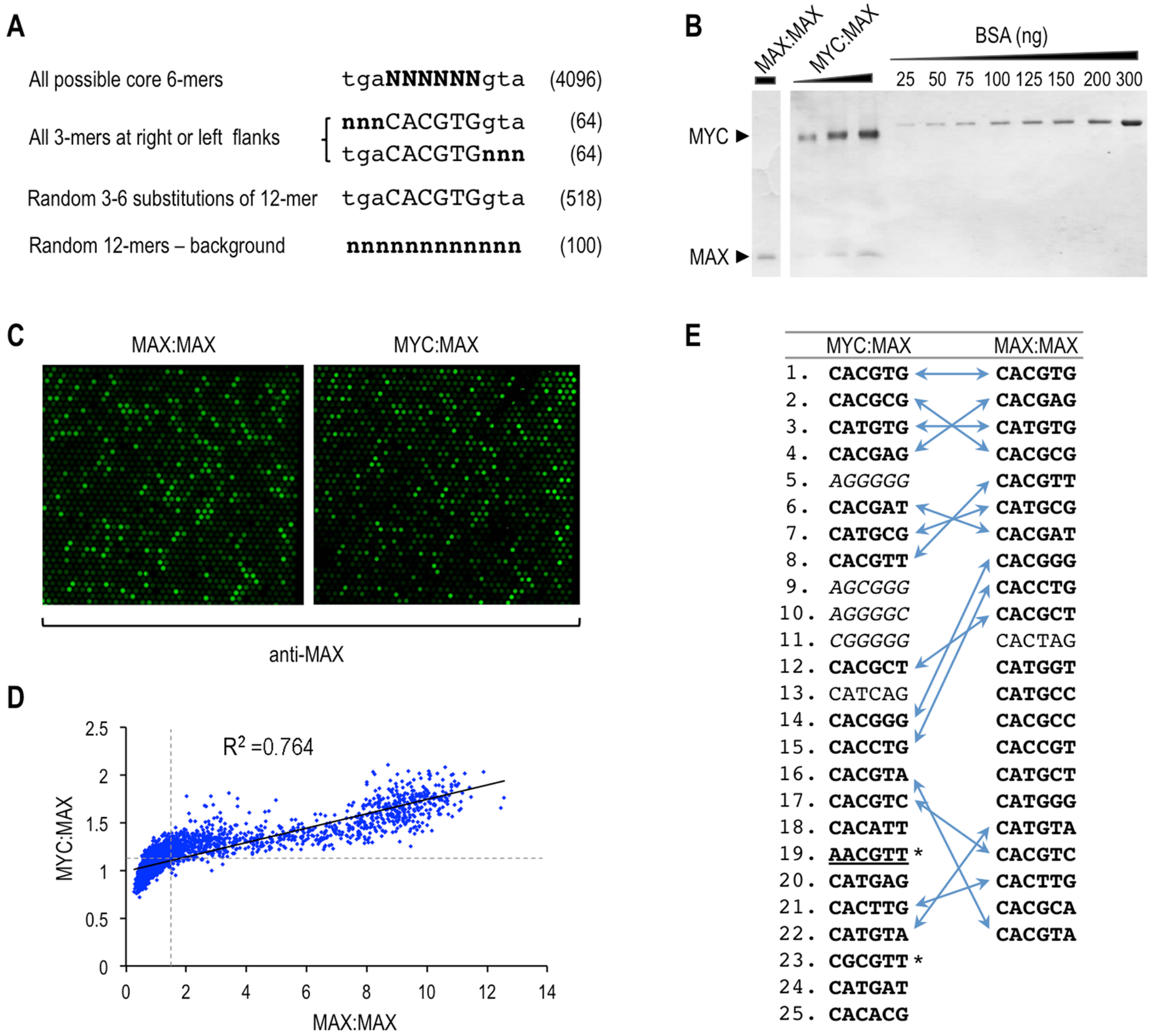


Fig 1. PBM analysis of the DNA-binding specificity of MYC/MAX complexes. (A) The custom PBM design contained immobilized DNA probes in triplicates with the indicated sequences (N and n indicate the varying positions having any of the 4 nucleotides). The total number of different sequences is shown in brackets. **(B)** Coomassie-stained SDS-PAGE showing the purified recombinant MYC/MAX complexes. **(C)** Images of fluorescent spot signals on PBM slides bound by MYC:MAX and MAX:MAX complexes and probed with anti-MAX antibody. **(D)** Scatter plot illustrating the correlation between MYC:MAX and MAX:MAX binding scores (normalized fluorescent intensities) for each individual probe spotted on the PBM ($R^2 = 0.764$). Dashed lines indicate the threshold values for binding scores above 80% of random probes. **(E)** Ranked order list of top-scoring motifs in PBM probes bound by MYC:MAX and MAX:MAX (with binding scores above 95% of random probes). Double arrows indicate identical motifs. Itallics are the low complexity G/C-rich motifs. Asterisks indicate non-E-box motifs, including the AACGTT motif (underlined). Motifs highlighted in bold were bound by both MYC:MAX and MAX:MAX above the 80% random threshold.

https://doi.org/10.1371/journal.pone.0180147.g001

EDTA, 0.075% NP-40, and 7.5 mM 2-mercaptoethanol, essentially as previously described [29]. DNA-binding competition assays were performed similarly, but unlabeled competitor oligos were mixed with the radiolabeled CME probe before the proteins were added. DNA-binding reactions were analyzed on a 6% native polyacrylamide gel, as described [29]. After drying the gel was autoradiographed and signals were quantified with a PhosphoImager (Typhoon 9410; GE Healthcare). The fraction (%) of bound probe remaining (Bound) is relative to no competitor and is the mean ± S.D. of more than three independent experiments. Statistical significance was determined by two-tailed Student's t-test and $P<0.05$.

Bioinformatics analyses

Datasets, filtering, and quality assessment. MYC-bound genomic sequences in the MYC-inducible human B cell line P493-6 were obtained from Gene Expression Omnibus (GEO) ChIP-seq reads derived from cells under either "high" (GSM1234501) or "low" (GSM1234500) levels of MYC [22]. GEO sample GSM894093 was used as the input control. ChIP-seq data sets of MYC-bound sequences in U2OS cells were obtained from GEO: GSE77356 [28]. SRA files were converted to FASTQ via NCBI SRA toolkit and uploaded to the public Galaxy server (http://usegalaxy.org) [32–34]. ChIP-seq reads were filtered using FASTQ Groomer and evaluated using the FASTQC tool.

Analysis of ChIP-seq data. For peak finding, ChIP-seq reads were aligned to the UCSC human genome (hg19) using Bowtie for Illumina. Unmapped reads were removed and the SAM files were converted to BAM files. The BAM files were transferred to a Galaxy server dedicated to ChIP-seq analysis, Cistrome (http://cistrome.org) [35]. MYC peaks were called using the MACS software version 2. Only significant peaks with P values below 1e-08 were recognized. MACs peaks were ranked by fold enrichment and sorted into three subsets: "Low", "Medium" and "High" enriched regions. The apex of each MYC peak was expanded by +/-100 base pairs to obtain MYC summits.

For motif enrichment and annotation, the genomic locations of all E-boxes, the Non-E-box (AACGTT) and control sequences were determined using the HOMER (http://homer.salk.edu) motif discovery software [36]. The normalized frequencies of motifs in MYC peaks and summits were obtained by intersecting the genomic positions of ChIP-seq peaks/summits and motifs, and then normalizing to the occurrence of each motif in the genome. Homer's peak annotation program was used to associate MYC peaks and summits with nearby genes and to identify peaks in the promoter (+/- 2 kb of TSS), intergenic, or intragenic regions of a genomic feature. Homer's peak annotation tool was also used to determine the spatial distribution of the CME and NE motifs in MYC summits (+/-100bp) and summits extended by 250bp on either side (total region +/-350bp).

For the gene ontology analyses, the genes associated with MYC peaks or summits containing only CME or NE motifs were analyzed by using Metascape (http://metascape.org) [37]. The annotation field "Biological Process" was used in the analyses.

Results

Identification of DNA sequence motifs preferentially bound by MYC/MAX complexes *in vitro*

In order to identify in an unbiased manner the DNA sequences preferentially bound by MYC/MAX complexes, we designed a protein binding microarray (PBM) that contained ~5000 different DNA probes (all in triplicate) and included all possible 6-mer motifs (tgaNNNNNNgta, 4096 total; Fig 1A). In addition, the PBM contained 128 probes with all 3-mer substitutions in either the left or the right flank of the CME core (nnnCACGTGgta and tgaCACGTGnnn) and 518 random substitutions of the 12-mer CME motif tgaCACGTGgta (minimum of three and maximum of six substitutions within a given 12-mer sequence). The PBM also contained 100 random 12-mer probes, which were used to evaluate non-specific binding (see Fig 1A). The PBM was incubated with purified recombinant full-length MYC:MAX or MAX:MAX protein complexes (Fig 1B) and analyzed by immunofluorescence using an anti-MAX antibody

(Fig 1C). The MYC:MAX complexes were devoid of any MAX:MAX dimer as determined by EMSA (see below).

Global analysis of the ~5000 DNA sequences on the PBM indicated, as expected, that the binding sequence preferences of MYC:MAX strongly correlate to those of MAX:MAX (R^2 = 0.764, Fig 1D). When considering only the most significantly bound probes in the PBM (binding scores greater than 95% of the random 12-mer sequences, corresponding to >2 standard deviations from mean random score), we identified 378 probes with 25 different motifs bound by MYC:MAX and 466 probes with 22 different motifs bound by MAX:MAX (Fig 1E, S1 Table). Among these, a total of 14 different 6-mer motifs were common to both MYC:MAX and MAX:MAX (Fig 1E, double arrows). The top six common MYC/MAX motifs matched the top six motifs identified by Guo et al. [23] in an independent PBM analysis reported while our study was in progress (see S1 Table). Two other MYC-bound motifs identified by Guo et al. in their top-12 list (Guo's motifs #7 and #10) were also in our list of top binders (see S1 Table). As expected, the CME motif CACGTG had the highest binding score for both MYC:MAX and MAX:MAX (Fig 1E). Nearly all the remaining top-ranked MYC/MAX motifs were either E-boxes (CANNTG) or variants having at least one half E-box sequence of the type CAC or CAT (Fig 1E). Notable exceptions were the motifs AACGTT and CGCGTT (asterisks in Fig 1E). These motifs diverge from typical CANNTG E-boxes (and from the CME) by two substitutions in both the CA and TG arms that define the palindromic E-box (underlined above), and hence were referred to as "non-E-box" sequences hereafter. In addition, the low complexity G/C-rich motifs AGGGGG, AGCGGG, AGGGGC, and CGGGGG (italics in Fig 1E) were also found in top-ranked probes bound by MYC:MAX. However, these motifs were not bound by MAX:MAX and, therefore, were not investigated further (see Discussion). Using a less stringent binding threshold (binding scores greater than 80% of random sequences) the non-E-box palindromic sequence AACGTT and all other motifs indicated in bold in Fig 1E (a total of 27 top-ranked motifs) were bound by both MYC:MAX and MAX:MAX. These represent high-confidence MYC/MAX motifs.

We further selectively analyzed the 4096 probes having every 6-mer motif flanked by identical nucleotides (i.e., tgaNNNNNNgta) and found ~360 probes specifically bound by MAX: MAX (Fig 2A) and MYC:MAX (Fig 2B) (with scores higher than 80% of random sequences). These probes contained most of the top 6-mer motifs for MYC:MAX and MAX:MAX identified in Fig 1E (see also S1 Table), and included the 27 motifs bound by both MYC:MAX and MAX:MAX (shown in bold in Fig 1E). As expected, the motifs of the top 50 bound probes were generally close variants of the canonical CME sequence CACGTG for both MAX:MAX and MYC:MAX (Fig 2A and 2B; logos 1–50). MYC:MAX also bound the low complexity G/C-rich sequence CCCCCT (or AGGGGG) and variants thereof (Fig 2B; logos 1–50). The next 50 probes with intermediate scores (Fig 2A and 2B; logos 51–100) had more degenerate motifs but often retained one half E-box and a central CG or TG (e.g., CACGNN or CATGNN). The remaining 150 probes with the lowest binding scores (Fig 2A and 2B, logos 101–350) had highly degenerate motifs that often had the first C replaced by A, but maintained the central CG or TG dinucleotides (e.g., CACGNN, CATGNN, or AACGNN). Notably, this analysis also identified the palindromic sequence AACGTT as a new non-E-box motif bound by both MYC:MAX and MAX:MAX complexes (Fig 2A and 2B, left panels).

The PBM data further indicated that in the context of the CME motif CACGTG there was a moderate influence of the flanking nucleotides, with possible differential effects on MYC:MAX and MAX:MAX (S1–S4 Figs). For instance, the top 25% of CME motifs most avidly bound by MAX:MAX were enriched in flanking sequences with a T or G at position n^1 of the 12-mer $n^1n^2n^3$CACGTGgta, while MYC:MAX-bound motifs often had an A at position n^3 (S1 and S3 Figs). However, both MYC:MAX and MAX:MAX complexes bound poorly to sequences with

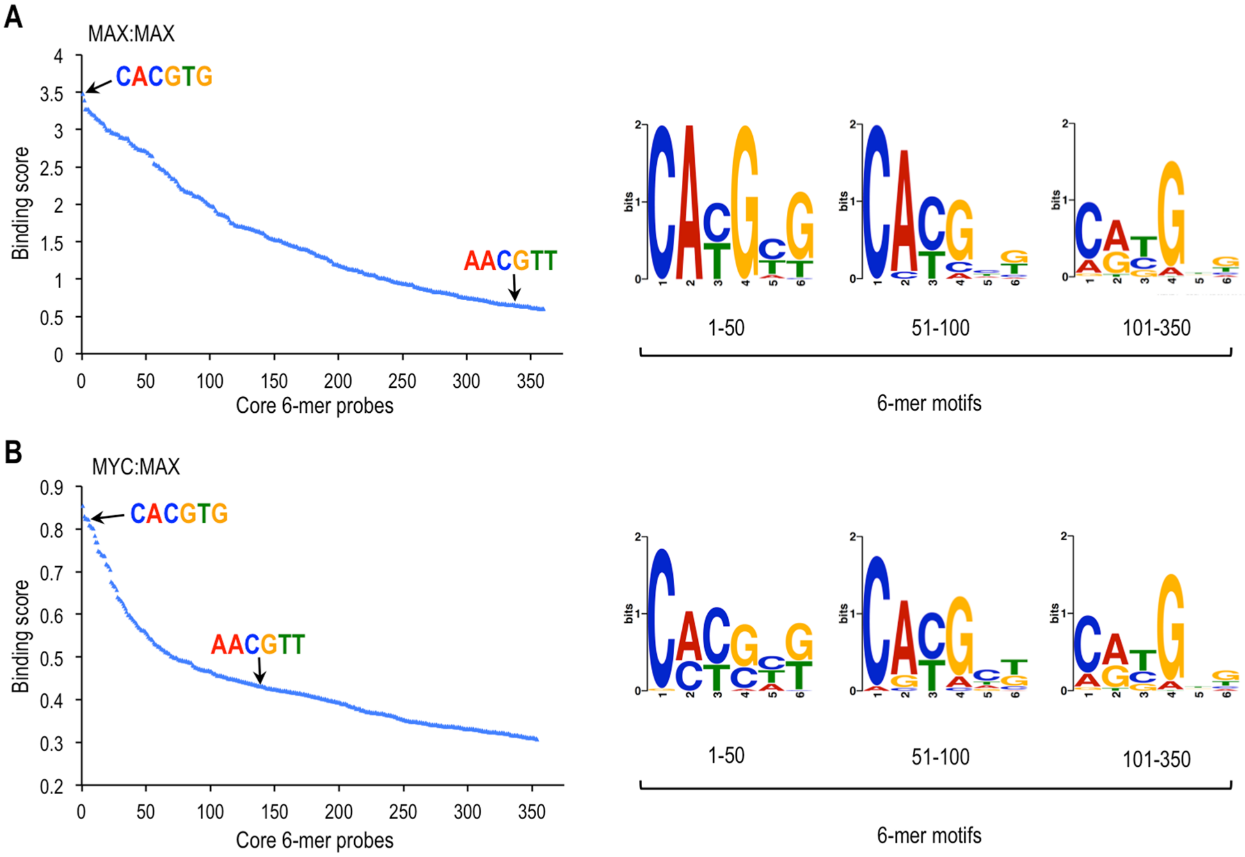


Fig 2. Position weight matrices of ranked core 6-mers bound by MAX:MAX and MYC:MAX. Core 6-mer probes of the PBM were ranked according to their binding scores (log2 of normalized fluorescent intensities) for MAX:MAX **(A)** and MYC:MAX **(B)**. Only probes scoring above 80% of random sequences are shown. The highest scoring probes containing the CME (CACGTG) and the NE (AACGTT) are indicated with arrows. Position weight matrices of 6-mer motifs present in the top 1–50, 51–100, and 101–350 bound probes were obtained with MEME and are represented as logos. The 8-mer sequences of bound probes (i.e., aNNNNNNg) were used as input for motif discovery.

https://doi.org/10.1371/journal.pone.0180147.g002

an A at position n^{10} of the 12-mer tgaCACGTGn^{10}n^{11}n^{12}, as those sequences were most often found in the group of bound probes with the lowest binding scores (bottom 25%, S2 Fig). Further analysis of bound probes having three to six random substitutions of the 12-mer tgaCACGTGgta also suggested distinct sequence preferences between MYC:MAX and MAX:MAX (S4 Fig). The significance of this differential binding is as of yet unclear, but merits further investigation. The complete ranked lists of the PBM probe sequences and binding scores for MYC:MAX and MAX:MAX are provided as supplemental information (S1 Table).

Sequence-specific binding of MYC/MAX to the non-E-box motif AACGTT

Recognition of the canonical MYC E-box sequence (CME) by MYC/MAX complexes involves important symmetrical contacts with specific bases (underlined for only the top strand CACGTG), including at the first and last positions of the core hexamer [14, 15]. Consequently, it was intriguing to find that the non-E-box (NE) palindromic sequence AACGTT, which has substitutions at both the first and last positions, was specifically bound by MYC/MAX in the PBM. To further verify the specificity of this interaction we used EMSA to analyze the binding of MYC:MAX and MAX:MAX complexes to radiolabeled CME, NE and non-specific DNA (Ctrl) probes (Fig 3A). Both MYC:MAX and MAX:MAX complexes bound specifically to the CME and NE probes (no binding to the Ctrl probe was observed) but the binding to the NE probe required higher protein concentrations, indicative of a lower affinity (Fig 3B). To measure the relative affinities of MYC:MAX and MAX:MAX for the CME and NE motifs we performed EMSA competition experiments in which different amounts of cold CME, NE or Ctrl oligonucleotides were co-incubated with the labeled NE probe and MYC/MAX complexes (Fig 3C). Compared to the CME, the NE motif had a significantly lower affinity for MYC: MAX and MAX:MAX (about 20- and 33-fold lower, respectively), while the Ctrl oligo did not compete, as anticipated (Fig 3C).

To determine which specific nucleotide positions within the NE motif are important, we performed additional EMSA competition experiments with NE variant (NE-V) sequences and MAX:MAX complexes. Substitution of the first and/or last base pair of the NE sequence AACGTT as in NE variants NE-V1 to V3, or inversion of the central CG nucleotides to GC (i.e., AAGCTT in NE-V4) (Fig 4A) eliminated the competition by the unlabeled oligos indicating that they do not bind MAX:MAX (Fig 4B and 4C). In contrast, a substitution at the first position that restored a perfect half E-box (CACGTT, EV-1) increased competition, and hence binding, and served as a positive control (Fig 4B). Interestingly, a substitution of a flanking nucleotide that generated the palindromic sequence CAACGTTG, which resembles an "extended" CME, did not increase the binding affinity (Fig 4C, NE(C) motif). Finally, the NE had similar binding affinity as several E-box variants that contained one perfect E-box half site CAC/GTG or CAT/ATG (E-V2 to V4, Fig 5) and are known to be specifically bound by MYC: MAX [9]. The relative affinities of the motifs analyzed by EMSA (summarized in Fig 6) were consistent with the binding scores of these sequences in the PBM and identified motifs with "high", "medium", and "low" affinity, and several motifs that do not bind MYC/MAX specifically (Fig 6).

MYC binding to AACGTT-containing genomic loci is highly dependent on increased MYC expression levels

To address whether MYC associates with genomic sites containing the NE sequence AACGTT *in vivo*, we mined previously published MYC ChIP-seq peaks from the human B cell line P493-6 [22]. P493-6 cells carry a tetracycline-repressible ectopic MYC transgene allowing for a range of MYC expression; in the absence of tetracycline the cells express high levels of MYC and display a Burkitt lymphoma-like phenotype [38]. We first compared, in MYC-overexpressing cells, the normalized frequencies (i.e., enrichment) of E-box sequences (CANNTG) and the NE motif AACGTT for all MYC-bound sites (all MYC peaks) or for only the top 33% most significantly bound loci (top 33% MYC peaks, ranked by p-value; Fig 7A). While the CME was the most enriched sequence in both categories, the NE sequence was also enriched under MYC peaks, and was more frequently found than all other E-boxes, with the exception of CAGCTG (Fig 7A). Since MYC ChIP-seq peaks are broad (covering on average ~500 bp), we also examined a more restricted DNA-bound region within 100 bp of the apex (MYC summits), which should contain the primary MYC binding site(s). Within MYC summits, the NE was also more enriched than many E-boxes, but less so than the CME and three other E-boxes, including the sequence CAGCTG (Fig 7B). The latter is intriguing because the CAGCTG sequence was not significantly bound by MYC/MAX in the PBM, consistent with the fact that another class of bHLH proteins (e.g., AP4, E12, E47 and MYOD) recognizes this E-box [39], suggesting an indirect binding of MYC to this element *in vivo*. Notably, the relative *in vitro* binding affinities of MYC/MAX for the specific motifs analyzed by EMSA were generally

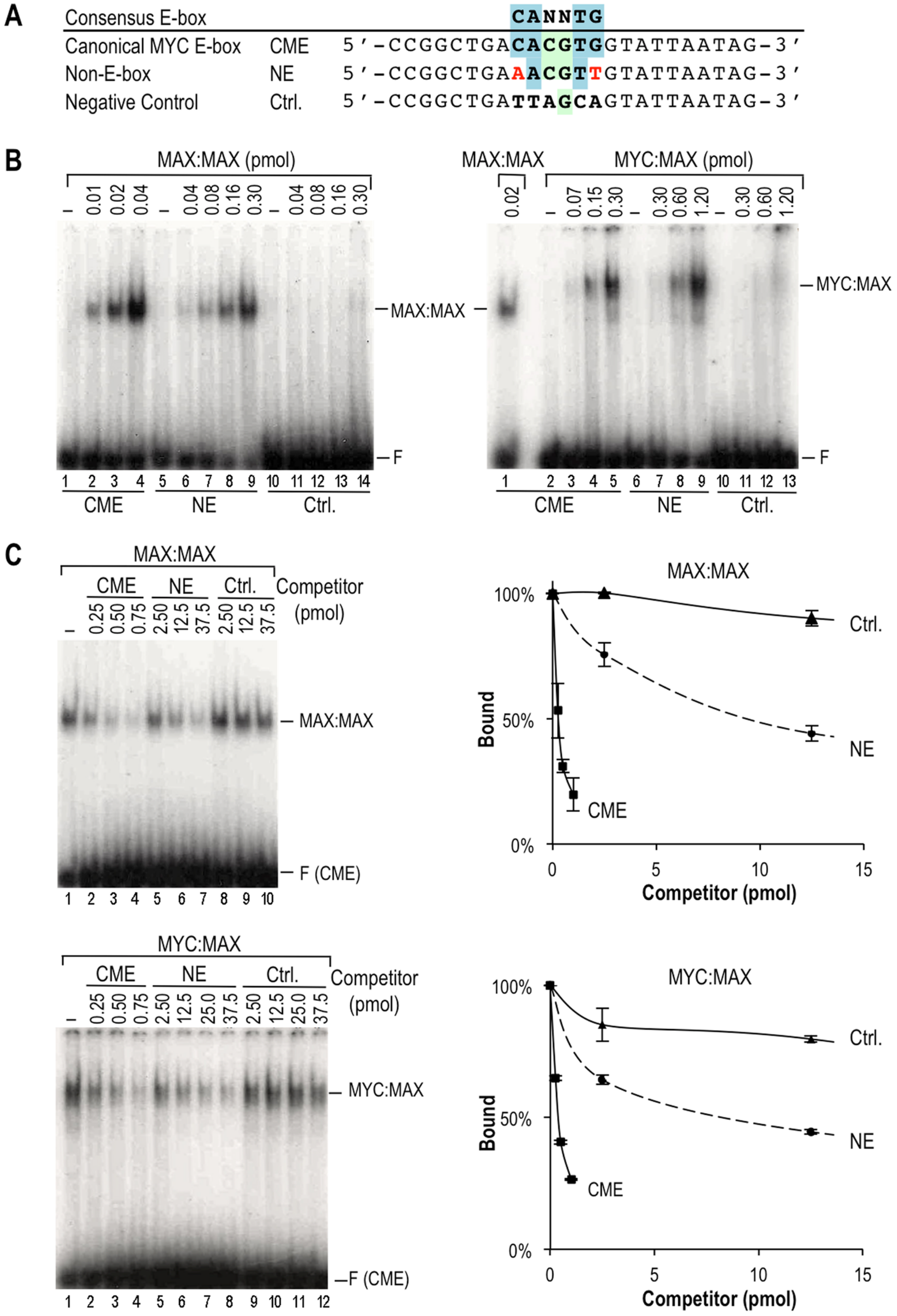
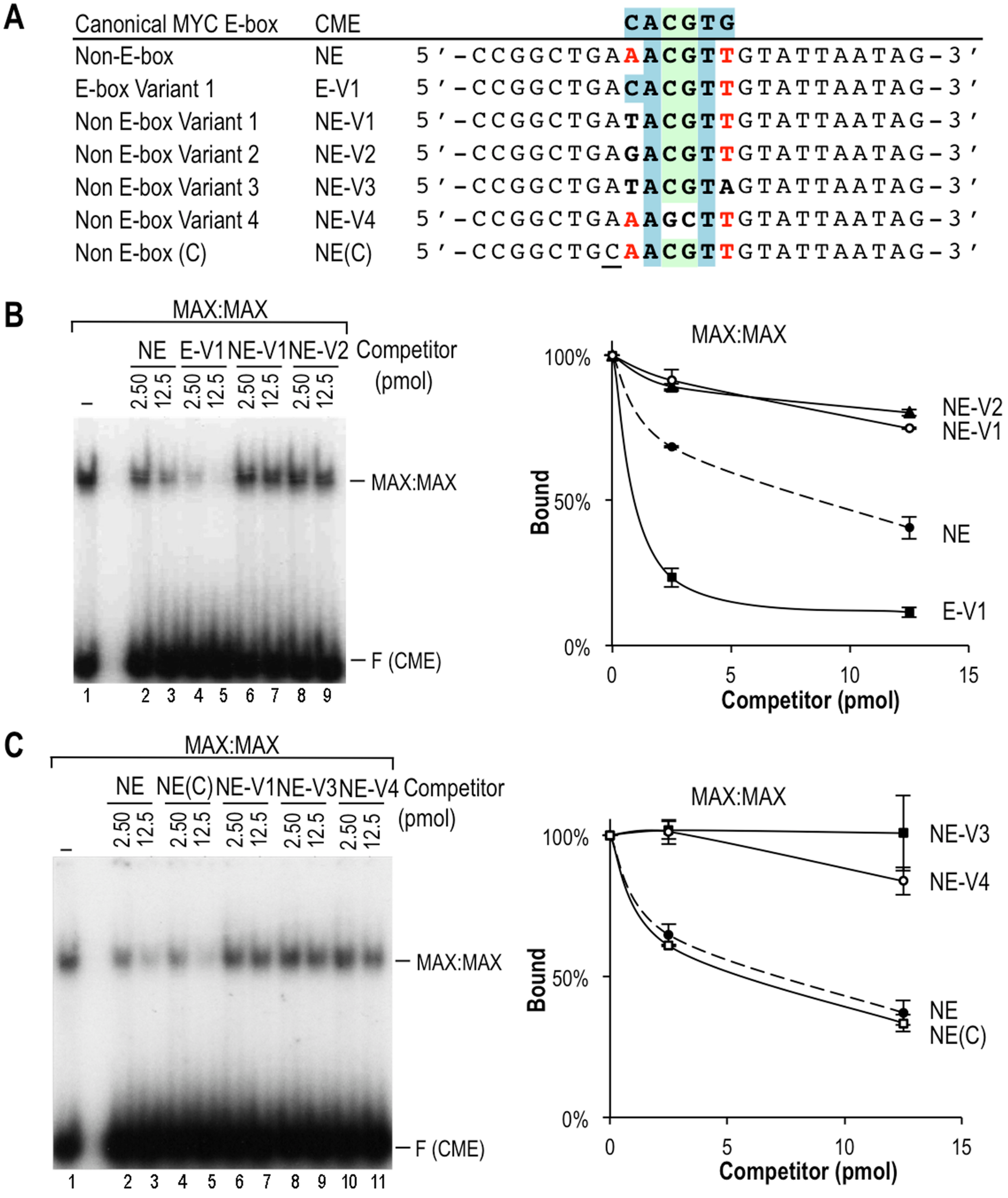


Fig 3. Specific binding of MYC/MAX complexes to the non-E-box (NE) motif AACGTT. (A) Sequences of the probes used in EMSA. Nucleotides in red indicate positions in the NE probe that deviate from an E-box. **(B)** EMSA with the CME, NE and Ctrl probes and with the indicated amounts (pmol) of MAX:MAX and MYC:MAX. The position of the unbound/free probe is shown (F). **(C)** EMSA competition assays were performed with the labeled CME probe and the indicated MYC/MAX complex and the indicated amounts (pmol) of either CME, NE or Ctrl competitor oligos or no competitor (-). Two representative gels are shown (left). Results from three independent replicates were quantitated and the means (± S.D.) were plotted (right) and significantly differed ($p<0.05$) between the competitor oligos.

https://doi.org/10.1371/journal.pone.0180147.g003



A

Canonical MYC E-box	CME	CACGTG
Non-E-box	NE	5′-CCGGCTGAAACGTTGTATTAATAG-3′
E-box Variant 1	E-V1	5′-CCGGCTGACACGTTGTATTAATAG-3′
Non E-box Variant 1	NE-V1	5′-CCGGCTGATACGTTGTATTAATAG-3′
Non E-box Variant 2	NE-V2	5′-CCGGCTGAGACGTTGTATTAATAG-3′
Non E-box Variant 3	NE-V3	5′-CCGGCTGATACGTAGTATTAATAG-3′
Non E-box Variant 4	NE-V4	5′-CCGGCTGAAAGCTTGTATTAATAG-3′
Non E-box (C)	NE(C)	5′-CCGGCTGCAACGTTGTATTAATAG-3′

Fig 4. Sequence requirements for MYC/MAX binding to the NE motif AACGTT. (A) Sequences of the oligo competitors. **(B, C)** EMSA competition assays with MAX:MAX and the labeled CME probe and the indicated amounts (pmol) of competitor oligos. Three independent experiments were quantitated as in Fig 3C. Significant differences were observed between NE and NE-V1/-V2/V3/V4 ($p<0.005$ by two-tailed t-test).

https://doi.org/10.1371/journal.pone.0180147.g004

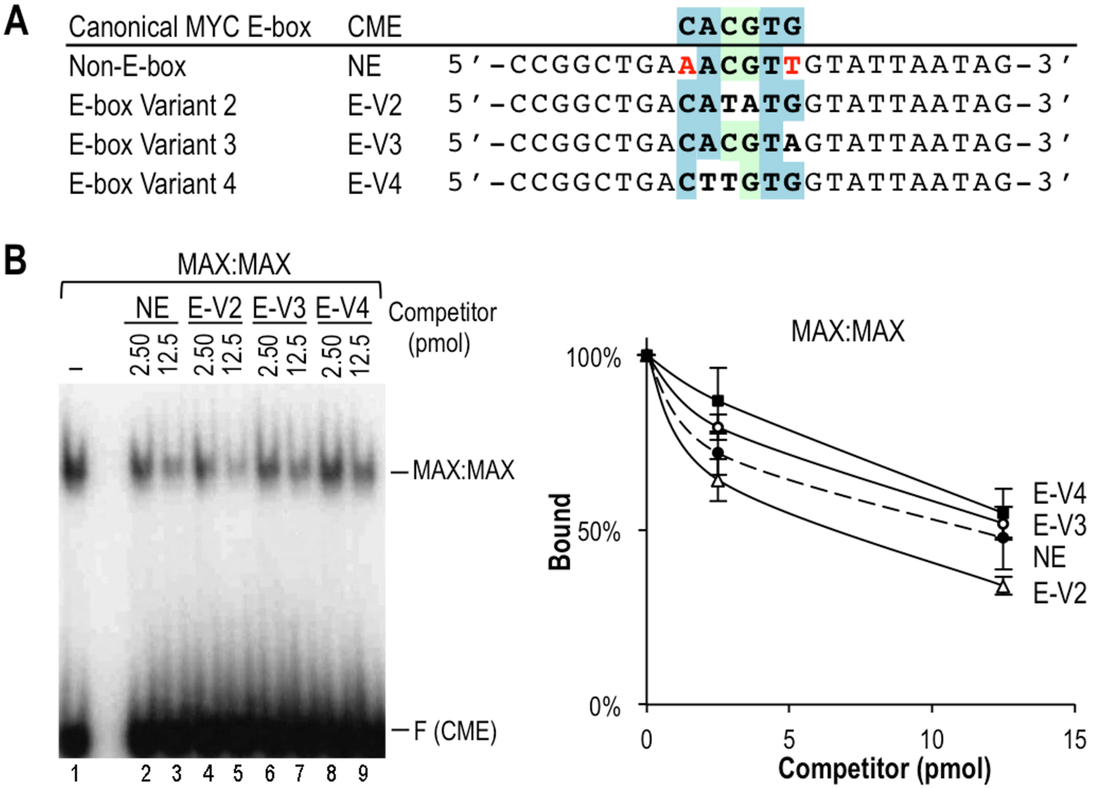
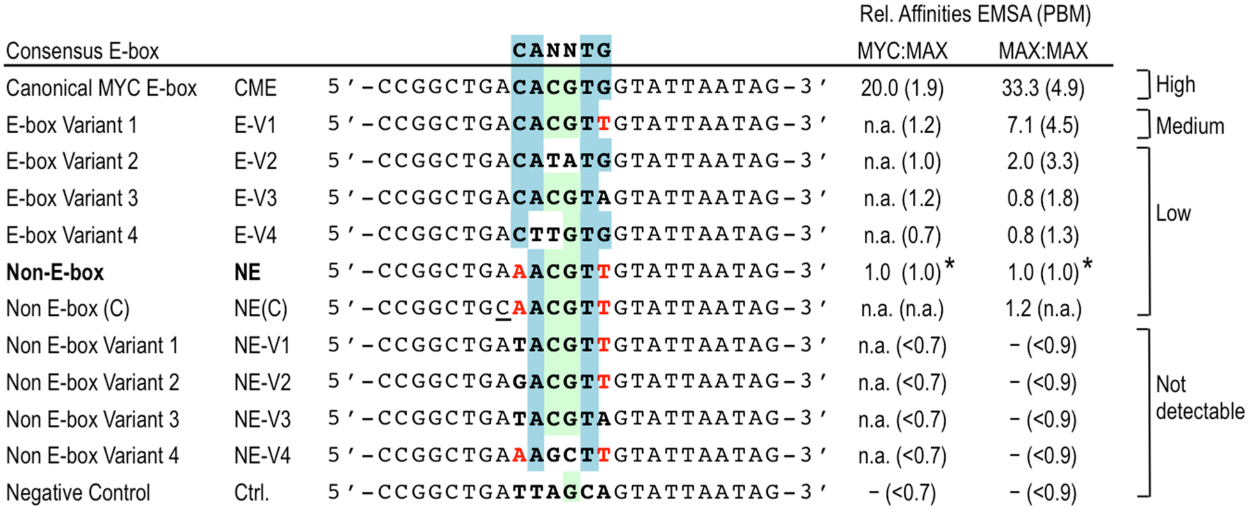


Fig 5. The NE motif AACGTT and several E-box variants have a similar affinity for MAX:MAX. (A) Sequences of the oligo competitors. **(B)** EMSA competition assays with MAX:MAX and the labeled CME probe and the indicated amounts (pmol) of competitor oligos. Three independent experiments were quantitated as in Fig 3C. No statistically significant difference was observed between NE and the different competitors (p>0.05 by two-tailed t-test).

https://doi.org/10.1371/journal.pone.0180147.g005



			Rel. Affinities EMSA (PBM)		
Consensus E-box		CANNTG	MYC:MAX	MAX:MAX	
Canonical MYC E-box	CME	5'-CCGGCTGACACGTGGTATTAATAG-3'	20.0 (1.9)	33.3 (4.9)	High
E-box Variant 1	E-V1	5'-CCGGCTGACACGTTGTATTAATAG-3'	n.a. (1.2)	7.1 (4.5)	Medium
E-box Variant 2	E-V2	5'-CCGGCTGACATATGGTATTAATAG-3'	n.a. (1.0)	2.0 (3.3)	Low
E-box Variant 3	E-V3	5'-CCGGCTGACACGTAGTATTAATAG-3'	n.a. (1.2)	0.8 (1.8)	
E-box Variant 4	E-V4	5'-CCGGCTGACTTGTGGTATTAATAG-3'	n.a. (0.7)	0.8 (1.3)	
Non-E-box	**NE**	5'-CCGGCTGAAACGTTGTATTAATAG-3'	1.0 (1.0)*	1.0 (1.0)*	
Non E-box (C)	NE(C)	5'-CCGGCTGCAACGTTGTATTAATAG-3'	n.a. (n.a.)	1.2 (n.a.)	
Non E-box Variant 1	NE-V1	5'-CCGGCTGATACGTTGTATTAATAG-3'	n.a. (<0.7)	- (<0.9)	Not detectable
Non E-box Variant 2	NE-V2	5'-CCGGCTGAGACGTTGTATTAATAG-3'	n.a. (<0.7)	- (<0.9)	
Non E-box Variant 3	NE-V3	5'-CCGGCTGATACGTAGTATTAATAG-3'	n.a. (<0.7)	- (<0.9)	
Non E-box Variant 4	NE-V4	5'-CCGGCTGAAAGCTTGTATTAATAG-3'	n.a. (<0.7)	- (<0.9)	
Negative Control	Ctrl.	5'-CCGGCTGATTAGCAGTATTAATAG-3'	- (<0.7)	- (<0.9)	

Fig 6. Summary of relative affinities of MYC:MAX and MAX:MAX complexes for different E-box and non-E-box sequences. Affinities were normalized relative to the affinity of the NE sequence, which was arbitrarily set to 1 (*). Relative affinities from the PBM scores are indicated between brackets. Sequences not analyzed by EMSA are indicated by "n.a." and non-detectable binding is indicated by "-".

https://doi.org/10.1371/journal.pone.0180147.g006

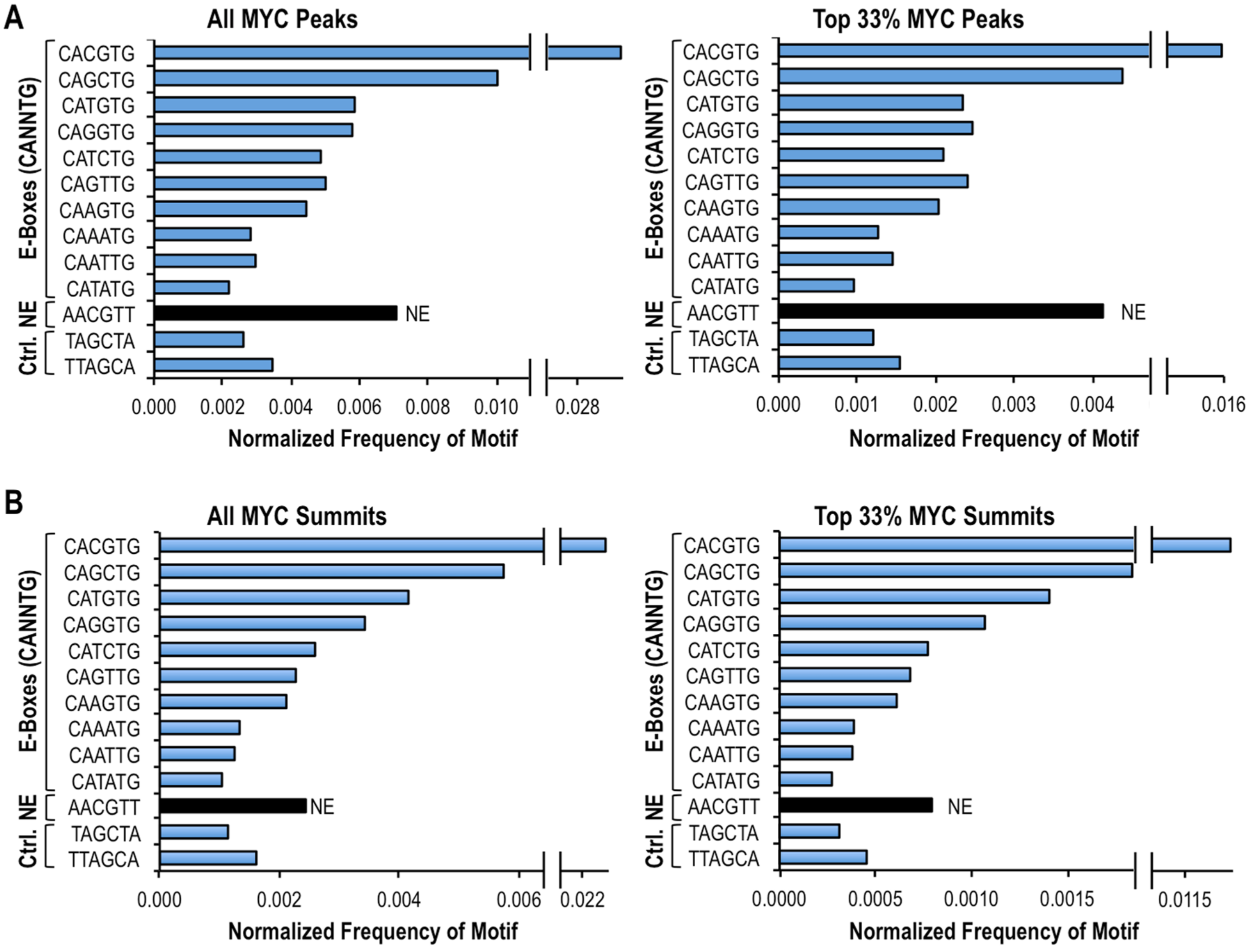


Fig 7. Normalized Frequencies of E-boxes and of the NE motif AACGTT at MYC-occupied genomic loci in human P493-6 B-cells. MYC-bound sequences were obtained from ChIP-seq data of P493-6 cells overexpressing MYC [22]. The frequencies of all ten possible E-boxes (CANNTG), the NE motif AACGTT, and two control (Ctrl) sequences under MYC ChIP-seq peaks **(A)** and summits **(B)** were normalized to the occurrence of each motif in the human genome. The summits are the ±100 bp region from the apex of ChIP-seq peaks. The normalized frequencies are shown for all MYC peaks/summits (left) and for the most significant MYC peaks/summits (top 33% based on p-value; right).

https://doi.org/10.1371/journal.pone.0180147.g007

consistent with the normalized frequencies of these sequences within MYC summits *in vivo* (S5 Fig). Two exceptions were the E-box CATATG, which had a measurable affinity for MYC/MAX *in vitro* but was rarely found in MYC summits *in vivo*; and the NE variant GACGTT (NE-V2), which was not detectably bound *in vitro* but was found at MYC summits *in vivo* (S5 Fig).

The above analyses indicate that the NE motif AACGTT is found within MYC-bound sequences *in vivo* at a higher normalized frequency (more enriched) than many E-box motifs. Furthermore, for both the MYC peaks and the more restricted MYC summits the vast majority of MYC-bound sequences with NE motifs (801 out of 1132 and 329 out of 390, respectively) did not contain the CME (Fig 8A). Altogether these results suggest that the NE might be recognized by overexpressed MYC *in vivo*. However, it should be noted that with the exception of the highest affinity and most enriched CME (CACGTG) the other MYC motifs examined, including medium-affinity/medium-enriched motifs (CATGTG and CACGTT) and the low-affinity E-box (CATATG) and NE (AACGTT) did not generally cluster at the center of ChIP-

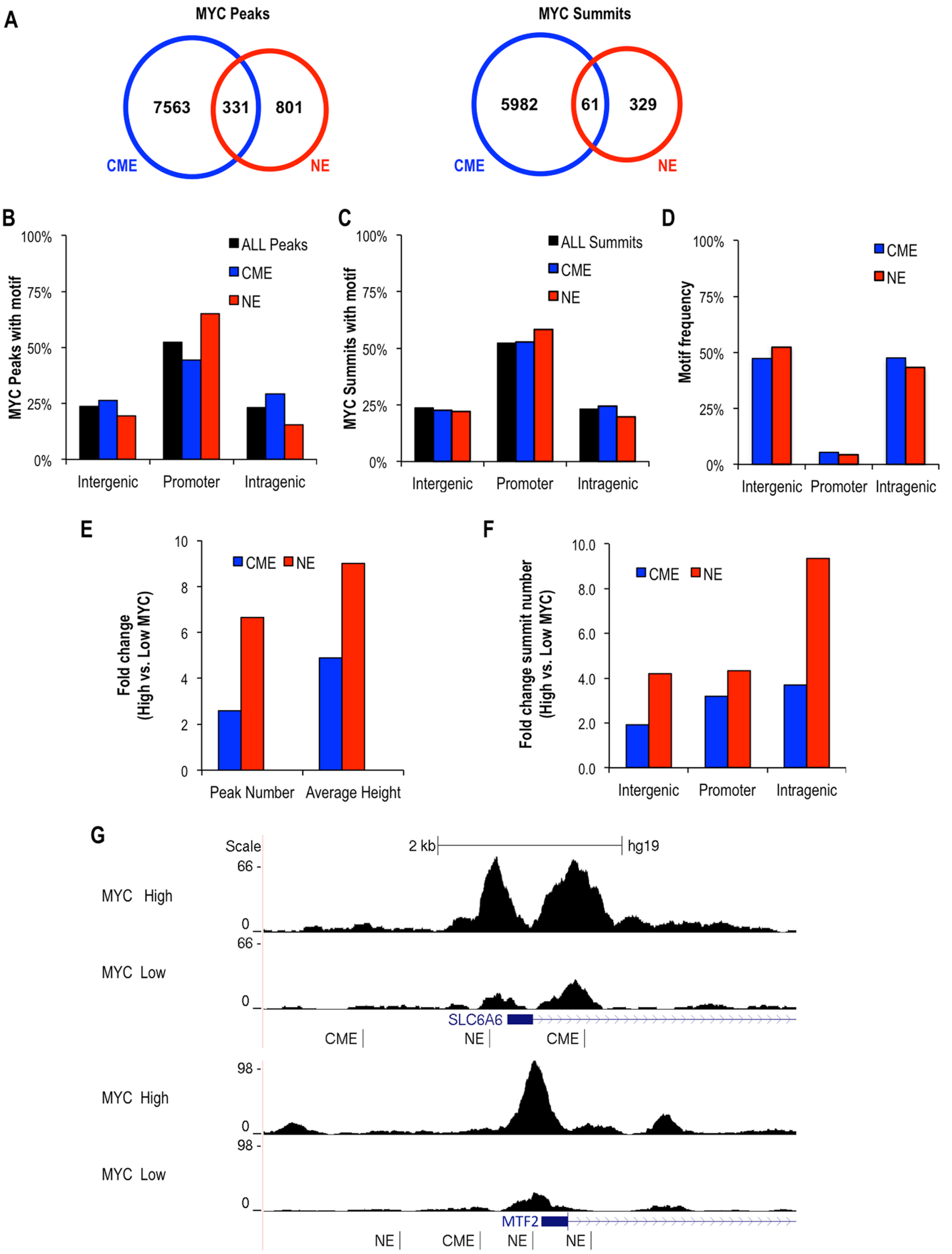


Fig 8. Characterization of MYC-occupied CME (CACGTG) and NE (AACGTT) genomic loci and influence of MYC expression levels in P493-6 cells. **(A)** Venn diagrams show the number of MYC ChIP-seq peaks and summits that contain the CME, the NE or both motifs in high MYC-expressing P493-6 cells. **(B, C)** Frequency distribution plots show the fraction (%) of all MYC ChIP-seq peaks and summits and those specifically containing the CME or NE motifs that are located within promoters (± 2 kb from a TSS), intergenic or intragenic regions. **(D)** Frequency distribution of the CME and NE motifs in the human genome. **(E)** Effect of MYC overexpression on the number and average (mean) height of MYC ChIP-seq summits is shown as fold change (high MYC vs. low MYC) for the summits containing the CME or NE motifs. **(F)**

Effect of MYC overexpression on the number of MYC ChIP-seq summits at promoters, intergenic and intragenic regions is shown as fold change (high MYC vs. low MYC) for CME- or NE-containing summits. **(G)** Examples of NE and CME ChIP-seq peaks from human genome browser in P493-6 cells expressing low and high MYC levels.

https://doi.org/10.1371/journal.pone.0180147.g008

seq summits (S5 Fig). This could suggest that recruitment of MYC to non-CME loci may often involve cooperative interactions with other chromatin-associated factors [21, 23, 25, 28, 40] in addition to sequence-specific binding to low-affinity DNA motifs.

We next characterized the location of different categories of MYC ChIP-seq peaks (all peaks/summits or peaks/summits containing either a CME or NE) relative to annotated genes in MYC-overexpressing cells (Fig 8B and 8C). The NE-containing MYC peaks and summits were mostly (~55–60%) localized to promoter proximal regions (+/-2 kb from a TSS) with the remaining NE-containing MYC peaks/summits equally distributed between intergenic and intragenic locations. A similar distribution was found for all peaks/summits and CME-containing peaks/summits (Fig 8B and 8C). However, this distribution differed remarkably from the global genomic location of CME and NE sites, which were both found primarily in inter- and intragenic regions (Fig 8D), indicating a preferential binding of MYC to both the CME and the NE within promoter regions.

In order to determine whether the cellular concentration of MYC impacts its binding to CME or NE sites *in vivo*, we compared the ChIP-seq peak number and height, as well as position relative to the gene body, in cells expressing high versus low levels of MYC. Consistent with the low affinity of MYC:MAX for the NE *in vitro* (Fig 6), the switch from low to high MYC expression resulted in a notable increase in both the total number and average height of NE-containing summits (~7-fold and ~9-fold, respectively) (Fig 8E). In contrast, the number of peaks and average height of CME-containing summits were considerably less sensitive to MYC protein concentration (~2-fold and ~5-fold, respectively) (Fig 8E). Furthermore, the differential effects of MYC expression levels on the number of NE versus CME sites occupied were more pronounced at distal intergenic and intragenic sites compared to promoter regions (Fig 8F). Nonetheless, induction of high MYC levels enhanced the intensity of MYC binding to NE motifs within both promoters and distal regions. Examples of an increase in size of NE-containing peaks in high MYC-expressing cells are provided in Fig 8G (for the promoters of *SLC6A6* and *MTF2* genes) and in S6 Fig (for other promoter, intergenic, and intragenic regions). These results suggest that MYC overexpression is important for binding NE sites *in vivo*, especially at distal regions.

To assess whether high-affinity CME and low-affinity NE sequences mediate the same or distinct MYC functions, we performed gene ontology analyses of promoters with MYC-bound CME motifs (CME summits lacking the NE) or MYC-bound NE motifs (NE summits lacking the CME). As expected, MYC-bound CME promoter sites are found in genes involved in a variety of biological processes, including most prominently ribosome biogenesis and cellular metabolism. In contrast, genes with MYC-occupied NE promoter sequences are not generally involved in these top biological processes, but instead appear to have a preference for the DNA damage response (S7 Fig and S2 Table).

Importantly, we repeated all the above analyses of MYC-bound genomic sequences in a different human cell line (U2OS) by re-analyzing published ChIP-seq datasets obtained under low endogenous MYC levels and after doxycycline–induced MYC overexpression [28] and obtained similar results (S5 and S8 Figs, and S3 Table).

Discussion

Using an unbiased PBM approach we have identified a set of 27 high-confidence hexamer motifs that are specifically bound by both MYC:MAX and MAX:MAX complexes *in vitro* (sequences in bold in Fig 1E). These include not only the CME sequence CACGTG and other E-boxes and E-box variants, but also non-E-box motifs such as the palindromic sequence AACGTT (referred to as NE), which was not previously known to be a specific MYC/MAX binding element. Notably, we verified by EMSA that both MYC:MAX and MAX:MAX complexes bound *in vitro* to the new AACGTT motif in a sequence-specific manner despite the ~20–30 fold lower affinity of MYC/MAX complexes for this sequence compared to the CME. The PBM results further indicated a contribution of the nucleotides immediately flanking the core 6-mer motifs, which supports previous observations made *in vitro* [5] and *in vivo* [17, 23]. In particular, the PBM indicated that CACGTGa (or its reverse complement tCACGTG) is disfavored by both MYC:MAX and MAX:MAX complexes, which is in accord with early site selection and EMSA experiments [5].

In addition, the PBM analyses identified low complexity G/C-rich sequences (AGGGGG, AGCGGG, AGGGGC, and CGGGGG) that were bound relatively strongly by MYC:MAX. However, these probes were not bound by MAX:MAX. This suggests that MYC may bind to these PBM probes in a sequence-independent manner that may not involve the basic region DNA-binding residues conserved between MYC and MAX. Although this observation is intriguing given the fact that *in vivo* MYC often binds genomic loci associated with CpG islands [17, 19, 20], the mode of recognition of these G/C-rich probes (sequence versus structural context recognition) needs to be further investigated by additional methods. Thus, our *in vitro* binding studies demonstrate that MYC/MAX complexes bind to a variety of non-canonical sequences and recognize low-affinity motifs, such as the non-E-box motif AACGTT, in a sequence-specific manner. The recognition of low-affinity non-E-box sequences is distinct from the non-specific DNA binding activity of MYC:MAX that was observed *in vitro* at high protein and DNA concentrations in the absence of non-specific competitor DNA [23].

The crystal structures of MYC:MAX and MAX:MAX complexes bound to DNA indicate that both MYC and MAX make identical contacts with the bases on either arm of the CME palindromic sequence [14, 15]. These include hydrogen bonding with the first and last base pairs of the CME (CACGTG), which are substituted in the NE motif (AACGTT). Molecular modeling of these substitutions in the MAX:MAX-DNA co-crystal structure [14] suggests a possible explanation as to why the NE motif is recognized by MYC/MAX complexes, albeit with a lower affinity. We propose that the first and last base pairs (AACGTT) of the NE may form a strong hydrogen bond (2.99 Å) between the thymine (T) and the His28 residue of MAX, while the complementary adenine (A) makes only a weak van der Waals contact with Glu32 of MAX (4.26Å). Other base pairs at these positions (distinct from CME or NE) would either clash with the structure or be too distant to make contact with MAX (see S9 Fig).

Analysis of chromatin loci bound by MYC in the genome of human P493-6 B-cells indicated that ~87% of all MYC-occupied sequences (ChIP-seq summits) contain at least one of the top 21 MYC:MAX-bound motifs identified by the PBM (excluding the low complexity G/C-rich motifs, Fig 1E) in both low- and high-MYC expressing cells (S10 Fig). Notably, we found that the NE motif AACGTT is present within MYC-bound genomic sequences (in both P493-6 and U2OS cells) at frequencies comparable to those of many E-box motifs. These observations support the previous finding that MYC often binds genomic loci that lack high affinity E-boxes [23]. However, that does not mean that such binding is necessarily "non-specific", as previously suggested [23]. Rather, we propose that MYC binding to low-affinity sites such as AACGTT may contribute to cooperative DNA binding with other transcription factors or chromatin-associated factors, in accord with recent findings [21, 23, 25, 28, 40].

In human cells (P493-6 and U2OS) MYC bound both the CME (CACGTG) and the AACGTT motif preferentially within promoter regions and MYC overexpression increased binding to both types of promoter motifs. Notably, however, the most prevalent functions of genes with MYC-bound CME promoter motifs (ribosome biogenesis and metabolic processes) were not enriched in the genes with MYC-occupied AACGTT promoter motifs, which tended to be involved in the DNA damage response. Moreover, MYC overexpression preferentially increased the number of occupied AACGTT sites relative to CME sites, especially at distal inter/intragenic loci (Fig 8F and S8 Fig). Hence, MYC expression levels and recognition of distinct DNA sequences with different affinities might influence distinct genetic programs, as was recently proposed [28].

In conclusion, our results suggest that overexpression of MYC allows for sequence-specific binding to the NE motif AACGTT, and raise the intriguing possibility that the marked induction of *de novo* MYC binding to low-affinity AACGTT-containing genomic loci (especially those distal to promoters) could be important for the oncogenic effects of MYC overexpression that is characteristic of many cancers.

Supporting information

S1 Fig. Effects of left-flanking 3-mer sequences on binding of MAX:MAX and MYC:MAX to the core CME. CME-containing PBM probes with all possible left-flanking 3-mers were ranked according to their binding to MYC:MAX and MAX:MAX. Logos of position weight matrices were obtained with MEME for the top 25%, 25–50%, 50–75%, and 75–100% of bound probes.
(PDF)

S2 Fig. Effects of right-flanking 3-mer sequences on binding of MAX:MAX and MYC: MAX to the core CME. CME-containing PBM probes with all possible right-flanking 3-mers were ranked according to their binding to MYC:MAX and MAX:MAX. Logos of position weight matrices were obtained with MEME for the top 25%, 25–50%, 50–75%, and 75–100% of bound probes.
(PDF)

S3 Fig. Effects of flanking 3-mer sequences on binding of MAX:MAX versus MYC:MAX. Scatter plots of MYC:MAX and MAX:MAX binding scores for CME-containing PBM probes with all possible left-flanking (top) and right-flanking (bottom) 3-mers are shown. Probes in red indicate those with trimer sequences that may favor binding by one of the complexes. Underlined are favored left-flanking nucleotides supported by the position weigh matrices of S1 Fig.
(PDF)

S4 Fig. Effects of mutations in tgaCACGTGgta on binding of MAX:MAX versus MYC: MAX. Scatter plots show MYC:MAX and MAX:MAX binding scores for PBM probes containing random 3–6 substitutions in 12-mer tgaCACGTGgta (probe shown in black). Top plot shows all probes with binding score above 80% random threshold and the double dashed line indicates the 95% binding threshold. Bottom shows only probes above 95% random threshold. Red indicates probes with mutations that preferentially decrease binding to either MYC:MAX or MAX:MAX. Mutations are underlined.
(PDF)

S5 Fig. Frequency and distribution of motifs within MYC ChIP-seq peaks in two human cell lines overexpressing MYC. (A) Comparison of the MYC/MAX affinity for different motifs (verified by EMSA) with the normalized frequency of motifs under MYC ChIP-seq summits (-100/+100 bp) in the genome of P493-6 and U2OS cells overexpressing MYC. The in vitro affinities High, Medium, Low and non-detectable (n.d.) relate to Fig 6. **(B)**

Comparison of motif frequency distribution within CME and NE summits (-100/+100 bp) versus their flanking 500bp regions (summits extended +/- 250 bp on each flank). **(C)** Frequency distribution of the indicated motifs within their MYC summits (-100/+100 bp). Only the CME shows a reproducible clustering at the center of the summit region in both cell lines. Motif frequency distributions were obtained with Homer peak annotation tool.
(PDF)

S6 Fig. Examples of MYC ChIP-seq peaks in the genome of human P493-6 cells under low and high MYC expression conditions. Tracks are from Genome Browser. Relates to Fig 8G.
(PDF)

S7 Fig. Gene ontology analyses of promoters with MYC-associated CME or NE motifs. Enrichment of biological processes associated with gene promoters having MYC-bound CME **(A)** or NE **(B)** motifs (ChIP-seq summits) in MYC-overexpressing P493-6 cells were obtained with Metascape and are ranked by significance (-log10 P value). See also S2 Table for the lists of genes.
(PDF)

S8 Fig. Characterization of MYC-occupied CME (CACGTG) and NE (AACGTT) genomic loci and influence of MYC expression levels in U2OS cells. (A) MYC-bound sequences were obtained from published ChIP-seq datasets of U2OS cells expressing low endogenous MYC (low-MYC) or overexpressing MYC (High-MYC) [28]. The frequencies of all ten possible E-boxes (CANNTG), the NE motif AACGTT, and two control (Ctrl) sequences under MYC ChIP-seq summits were normalized to the occurrence of each motif in the human genome. The summits are the ±100 bp region centered at the apex of ChIP-seq peaks. **(B)** Venn diagram shows the number of MYC ChIP-seq summits that contain the CME, the NE or both motifs in high MYC-expressing U2OS cells. **(C)** Frequency distribution plots show the fraction (%) of all MYC ChIP-seq summits and those specifically containing the CME or NE motifs that are located within promoters (± 2 kb from a TSS), intergenic or intragenic regions. **(D)** Effect of MYC overexpression on the number and average (mean) height of MYC ChIP-seq summits is shown as fold change (high MYC vs. low MYC) for the summits containing the CME or NE motifs. **(E)** Effect of MYC overexpression on the number of MYC ChIP-seq summits at promoters, intergenic and intragenic regions is shown as fold change (high MYC vs. low MYC) for the specific CME- or NE-containing summits, and for all summits. **(F)** Gene ontology analyses of promoters with MYC-associated CME (top) or NE (bottom) summits in MYC-overexpressing U2OS cells. Enrichment of biological processes was obtained with Metascape and processes are ranked by significance (-log10 P value). See also S3 Table for the lists of genes associated with each biological process.
(PDF)

S9 Fig. Diagram illustrating a possible mechanism for sequence-specific recognition of the NE motif by MYC/MAX complexes. (Left) The contacts between MAX:MAX and the CME motif are depicted based on published X-ray crystallographic results [14] (PDBe 1an2). Arrows indicate H-bonds and the corresponding distance between base and amino acid residues is indicated. Only the contacts by one MAX monomer with on one half of the symmetric palindrome are shown. **(Right)** Proposed interactions of MAX with the NE motif were modeled by using the Coot (Crystallographic Object-Oriented Toolkit) software. The postulated new hydrogen bond between His28 (H28) of MAX and O4 of Thymine (T) is indicated with a red arrow. A potential van der Waals contact is indicated with a dashed red arrow.
(PDF)

S10 Fig. Most MYC-bound genomic sites in P493-6 cells contain at least one of the top 21 MYC:MAX-bound motifs identified by the PBM. (A) Total number of MYC ChIP-seq summits (all summits) and summits having one of the top 21 motifs for MYC:MAX identified in vitro by the PBM (i.e., MYC:MAX motifs of Fig 1E excluding GC-rich motifs in italics) under low and high MYC expression. **(B)** Fraction (%) of MYC summits with one of the top 21 MYC:MAX-bound motifs identified *in vitro*.
(PDF)

S1 Table. PBM probes sequences and MYC/MAX binding scores.
(XLSX)

S2 Table. Gene ontology of MYC-bound CME and NE sites in P493-6 cells.
(XLSX)

S3 Table. Gene ontology of MYC-bound C'ME and NE sites in U2OS cells.
(XLSX)

Acknowledgments

We thank all the members of our laboratories for encouragement and critical discussions. We also thank Jikui Song for assistance with molecular modeling.

Author Contributions

Conceptualization: EM FMS MA.

Data curation: MA EB FMS EM.

Formal analysis: MA EB EM.

Funding acquisition: EM FMS.

Investigation: MA EB MG DM-P EM.

Methodology: FMS EM MA EB.

Project administration: EM FMS.

Resources: EM FMS.

Supervision: EM FMS.

Validation: MA EB MG DM-P.

Visualization: MA EB FMS EM.

Writing – original draft: MA EM.

Writing – review & editing: MA EB MG DM-P FMS EM.

References

1. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu Rev Cell Dev Biol. 2000; 16:653–99. https://doi.org/10.1146/annurev.cellbio.16.1.653 PMID: 11031250.
2. Dang CV. MYC on the path to cancer. Cell. 2012; 149(1):22–35. https://doi.org/10.1016/j.cell.2012.03.003 PMID: 22464321.

3. Amati B, Dalton S, Brooks MW, Littlewood TD, Evan GI, Land H. Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature. 1992; 359(6394):423–6. https://doi.org/10.1038/359423a0 PMID: 1406955.

4. Blackwood EM, Eisenman RN. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science. 1991; 251(4998):1211–7. PMID: 2006410.

5. Solomon DL, Amati B, Land H. Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. Nucleic Acids Res. 1993; 21(23):5372–6. PMID: 8265351.

6. Prendergast GC, Lawe D, Ziff EB. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell. 1991; 65(3):395–407. PMID: 1840505.

7. Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN, Weintraub H. Sequence-specific DNA binding by the c-Myc protein. Science. 1990; 250(4984):1149–51. PMID: 2251503.

8. Berberich S H-D N, Espenshade P, Cole M. Max encodes a sequence-specific DNA-binding protein and is not regulated by serum growth factors. Oncogene. 1992; 7:775–9. PMID: 1565473

9. Blackwell TK, Huang J, Ma A, Kretzner L, Alt FW, Eisenman RN, et al. Binding of myc proteins to canonical and noncanonical DNA sequences. Mol Cell Biol. 1993; 13(9):5216–24. PMID: 8395000.

10. Amati B, Littlewood TD, Evan GI, Land H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. EMBO J. 1993; 12(13):5083–7. PMID: 8262051.

11. Kretzner L, Blackwood EM, Eisenman RN. Myc and Max proteins possess distinct transcriptional activities. Nature. 1992; 359(6394):426–9. https://doi.org/10.1038/359426a0 PMID: 1406956.

12. Lindeman GJ, Harris AW, Bath ML, Eisenman RN, Adams JM. Overexpressed max is not oncogenic and attenuates myc-induced lymphoproliferation and lymphomagenesis in transgenic mice. Oncogene. 1995; 10(5):1013–7. PMID: 7898919.

13. Canelles M, Delgado MD, Hyland KM, Lerga A, Richard C, Dang CV, et al. Max and inhibitory c-Myc mutants induce erythroid differentiation and resistance to apoptosis in human myeloid leukemia cells. Oncogene. 1997; 14(11):1315–27. https://doi.org/10.1038/sj.onc.1200948 PMID: 9178892.

14. Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature. 1993; 363(6424):38–45. https://doi.org/10.1038/363038a0 PMID: 8479534.

15. Nair SK, Burley SK. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. Cell. 2003; 112(2):193–205. PMID: 12553908.

16. Grandori C, Mac J, Siebelt F, Ayer DE, Eisenman RN. Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites in vivo. EMBO J. 1996; 15(16):4344–57. PMID: 8861962.

17. Perna D, Faga G, Verrecchia A, Gorski MM, Barozzi I, Narang V, et al. Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. Oncogene. 2012; 31(13):1695–709. https://doi.org/10.1038/onc.2011.359 PMID: 21860422.

18. Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, et al. Transcriptional amplification in tumor cells with elevated c-Myc. Cell. 2012; 151(1):56–67. https://doi.org/10.1016/j.cell.2012.08.026 PMID: 23021215.

19. Zeller KI, Zhao X, Lee CW, Chiu KP, Yao F, Yustein JT, et al. Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc Natl Acad Sci U S A. 2006; 103(47):17834–9. https://doi.org/10.1073/pnas.0604129103 PMID: 17093053.

20. Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S, Greene J, et al. Genomic targets of the human c-Myc protein. Genes Dev. 2003; 17(9):1115–29. https://doi.org/10.1101/gad.1067003 PMID: 12695333.

21. Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, et al. Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol. 2006; 8(7):764–70. https://doi.org/10.1038/ncb1434 PMID: 16767079.

22. Sabo A, Kress TR, Pelizzola M, de Pretis S, Gorski MM, Tesi A, et al. Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. Nature. 2014; 511(7510):488–92. https://doi.org/10.1038/nature13537 PMID: 25043028.

23. Guo J, Li T, Schipper J, Nilson KA, Fordjour FK, Cooper JJ, et al. Sequence specificity incompletely defines the genome-wide occupancy of Myc. Genome Biol. 2014; 15(10):482. https://doi.org/10.1186/s13059-014-0482-3 PMID: 25287278.

24. Tu WB, Helander S, Pilstal R, Hickman KA, Lourenco C, Jurisica I, et al. Myc and its interactors take shape. Biochim Biophys Acta. 2015; 1849(5):469–83. https://doi.org/10.1016/j.bbagrm.2014.06.002 PMID: 24933113.

25. Kress TR, Sabo A, Amati B. MYC: connecting selective transcriptional control to global RNA production. Nat Rev Cancer. 2015; 15(10):593–607. https://doi.org/10.1038/nrc3984 PMID: 26383138.

26. Cowling VH, Cole MD. Myc Regulation of mRNA Cap Methylation. Genes Cancer. 2010; 1(6):576–9. https://doi.org/10.1177/1947601910378025 PMID: 21170289.

27. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, et al. c-Myc regulates transcriptional pause release. Cell. 2010; 141(3):432–45. https://doi.org/10.1016/j.cell.2010.03.030 PMID: 20434984.

28. Lorenzin F, Benary U, Baluapuri A, Walz S, Jung LA, von Eyss B, et al. Different promoter affinities account for specificity in MYC-dependent gene regulation. Elife. 2016;5. https://doi.org/10.7554/eLife.15161 PMID: 27460974.

29. Farina A, Faiola F, Martinez E. Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in Escherichia coli. Protein Expr Purif. 2004; 34(2):215–22. https://doi.org/10.1016/j.pep.2003.11.021 PMID: 15003254.

30. Bolotin E, Liao HL, Ta TC, Yang CH, Hwang-Verslues W, Evans JR, et al. Integrated Approach for the Identification of Human Hepatocyte Nuclear Factor 4 alpha Target Genes Using Protein Binding Microarrays. Hepatology. 2010; 51(2):642–53. https://doi.org/10.1002/hep.23357 PMID: 20054869

31. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36. PMID: 7584402.

32. Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11 (8):R86. https://doi.org/10.1186/gb-2010-11-8-r86 PMID: 20738864.

33. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010; Chapter 19:Unit 19 0 1–21. https://doi.org/10.1002/0471142727.mb1910s89 PMID: 20069535.

34. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005; 15(10):1451–5. https://doi.org/10.1101/gr.4086505 PMID: 16169926.

35. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol. 2011; 12(8):R83. https://doi.org/10.1186/gb-2011-12-8-r83 PMID: 21859476.

36. 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(4):576–89. https://doi.org/10.1016/j.molcel.2010.05.004 PMID: 20513432.

37. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. Cell Host Microbe. 2015; 18(6):723–35. https://doi.org/10.1016/j.chom.2015.11.002 PMID: 26651948.

38. Pajic A, Spitkovsky D, Christoph B, Kempkes B, Schuhmacher M, Staege MS, et al. Cell cycle activation by c-myc in a burkitt lymphoma model cell line. Int J Cancer. 2000; 87(6):787–93. PMID: 10956386.

39. Dang CV, Dolde C, Gillison ML, Kato GJ. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. Proc Natl Acad Sci U S A. 1992; 89(2):599–602. PMID: 1731330.

40. Thomas LR, Wang Q, Grieb BC, Phan J, Foshage AM, Sun Q, et al. Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. Mol Cell. 2015; 58(3):440–52. https://doi.org/10.1016/j.molcel.2015.02.028 PMID: 25818646.

Supporting Information

S1 Fig. Effects of left-flanking 3-mer sequences on binding of MAX:MAX and MYC:MAX to the core CME. CME-containing PBM probes with all possible left-flanking 3-mers were ranked according to their binding to MYC:MAX and MAX:MAX. Logos of position weight matrices were obtained with MEME for the top 25%, 25-50%, 50-75%, and 75-100% of bound probes.

S2 Fig. Effects of right-flanking 3-mer sequences on binding of MAX:MAX and MYC:MAX to the core CME. CME-containing PBM probes with all possible right-flanking 3-mers were ranked according to their binding to MYC:MAX and MAX:MAX. Logos of position weight matrices were obtained with MEME for the top 25%, 25-50%, 50-75%, and 75-100% of bound probes.

S3 Fig. Effects of flanking 3-mer sequences on binding of MAX:MAX versus MYC:MAX. Scatter plots of MYC:MAX and MAX:MAX binding scores for CME-containing PBM probes with all possible left-flanking (top) and right-flanking (bottom) 3-mers are shown. Probes in red indicate those with trimer sequences that may favor binding by one of the complexes. Underlined are favored left-flanking nucleotides supported by the position weigh matrices of S1 Fig.

S4 Fig. Effects of mutations in tgaCACGTGgta on binding of MAX:MAX versus MYC:MAX. Scatter plots show MYC:MAX and MAX:MAX binding scores for PBM probes containing random 3-6 substitutions in 12-mer tgaCACGTGgta (probe shown in black). Top plot shows all probes with binding score above 80% random threshold and the double dashed line indicates the 95% binding threshold. Bottom shows only probes above 95% random threshold. Red indicates probes with mutations that preferentially decrease binding to either MYC:MAX or MAX:MAX. Mutations are underlined.

S5 Fig. Frequency and distribution of motifs within MYC ChIP-seq peaks in two human cell lines overexpressing MYC (A) Comparison of the MYC/MAX affinity for different motifs (verified by EMSA) with the normalized frequency of motifs under MYC ChIP-seq summits (-100/+100 bp) in the genome of P493-6 and U2OS cells overexpressing MYC. The in vitro affinities High, Medium, Low and non-detectable (n.d.) relate to Fig 6. **(B)** Comparison of motif frequency distribution within CME and NE summits (-100/+100 bp) versus their flanking 500bp regions (summits extended +/- 250 bp on each flank). **(C)** Frequency distribution of the indicated motifs within their MYC summits (-100/+100 bp). Only the CME shows a reproducible clustering at the center of the summit region in both cell lines. Motif frequency distributions were obtained with Homer peak annotation tool.

S6 Fig. Examples of MYC ChIP-seq peaks in the genome of human P493-6 cells under low and high MYC expression conditions. Tracks are from Genome Browser. Relates to Fig 8G.

S7 Fig. Gene ontology analyses of promoters with MYC-associated CME or NE motifs. Enrichment of biological processes associated with gene promoters having MYC-bound CME **(A)** or NE **(B)** motifs (ChIP-seq summits) in MYC-overexpressing P493-6 cells were obtained with Metascape and are ranked by significance (-log10 P value). See also S2 Table for the lists of genes.

S8 Fig. Characterization of MYC-occupied CME (CACGTG) and NE (AACGTT) genomic loci and influence of MYC expression levels in U2OS cells. (A) MYC-bound sequences were obtained from published ChIP-seq datasets of U2OS cells expressing low endogenous MYC (low-MYC) or overexpressing MYC (High-MYC) [28]. The frequencies of all ten possible E-boxes (CANNTG), the NE motif AACGTT, and two control (Ctrl) sequences under MYC ChIP-seq summits were normalized to the occurrence of each motif in the human genome. The summits are the ±100 bp region centered at the apex of ChIP-seq peaks. **(B)** Venn diagram shows the number of MYC ChIP-seq summits that contain the CME, the NE or both motifs in high MYC-expressing U2OS cells. **(C)** Frequency distribution plots show the fraction (%) of all MYC ChIP-seq summits and those specifically containing the CME or NE motifs that are located within promoters (± 2 kb from a TSS), intergenic or intragenic regions. **(D)** Effect of MYC overexpression on the number and average (mean) height of MYC ChIP-seq summits is shown as fold change (high MYC vs. low MYC) for the summits containing the CME or NE motifs. **(E)** Effect of MYC overexpression on the number of MYC ChIP-seq summits at promoters, intergenic and intragenic regions is shown as fold change (high MYC vs. low MYC) for the specific CME- or NE-containing summits, and for all summits. **(F)** Gene ontology analyses of promoters with MYC-associated CME (top) or NE (bottom) summits in MYC-overexpressing U2OS cells. Enrichment of biological processes was obtained with Metascape and processes are ranked by significance (-log10 P value). See also S3 Table for the lists of genes associated with each biological process.

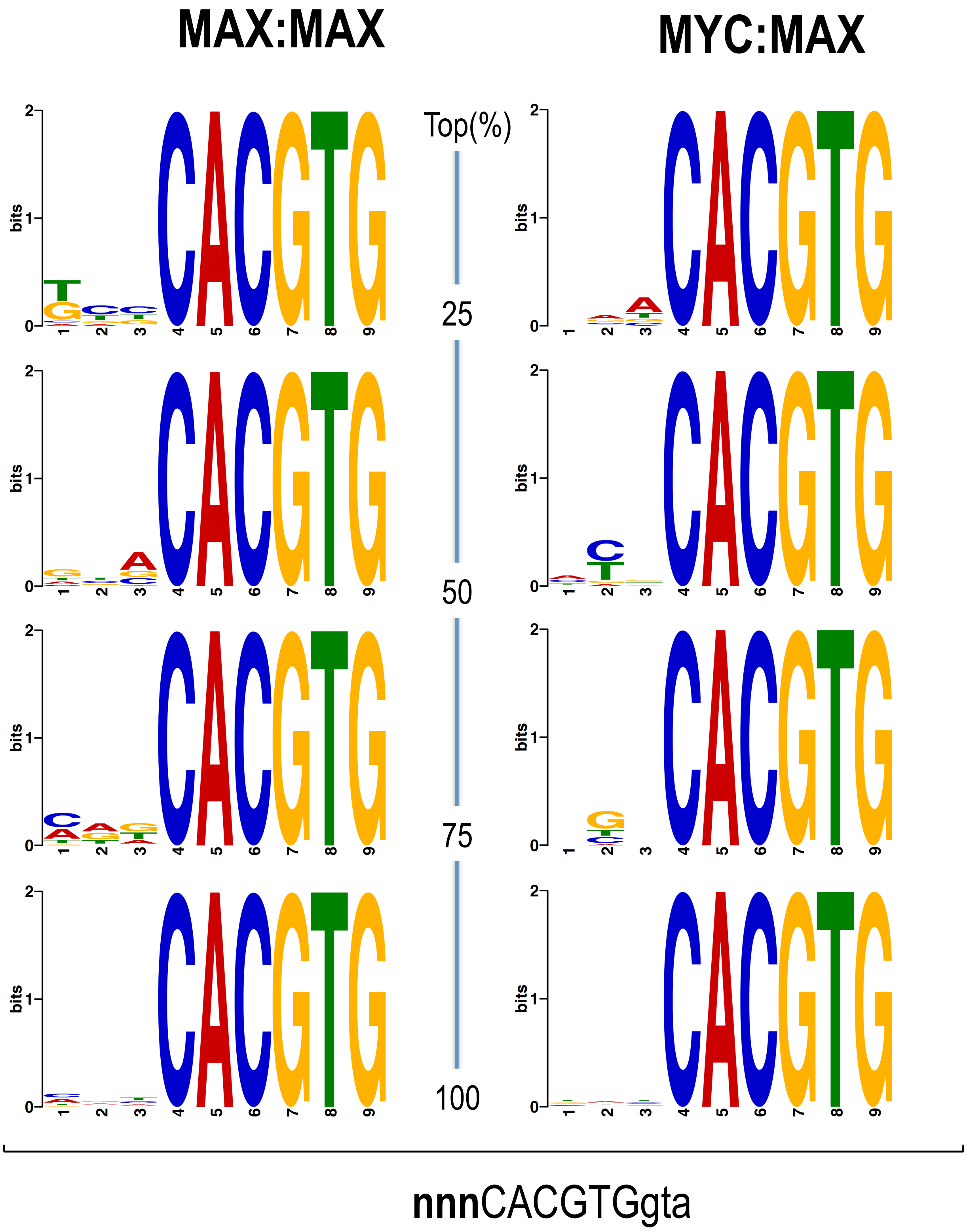
S9 Fig. Diagram illustrating a possible mechanism for sequence-specific recognition of the NE motif by MYC/MAX complexes. (Left) The contacts between MAX:MAX and the CME motif are depicted based on published X-ray crystallographic results [14] (PDBe 1an2). Arrows indicate H-bonds and the corresponding distance between base and amino acid residues is indicated. Only the contacts by one MAX monomer with on one half of the symmetric palindrome are shown. **(Right)** Proposed interactions of MAX with the NE motif were modeled by using the Coot (Crystallographic Object-Oriented Toolkit) software. The postulated new hydrogen bond between His28 (H28) of MAX and O4 of Thymine (T) is indicated with a red arrow. A potential van der Waals contact is indicated with a dashed red arrow.

S10 Fig. Most MYC-bound genomic sites in P493-6 cells contain at least one of the top 21 MYC:MAX-bound motifs identified by the PBM. (A) Total number of MYC ChIP-seq summits (all summits) and summits having one of the top 21 motifs for MYC:MAX identified in vitro by the PBM (i.e., MYC:MAX motifs of Fig 1E excluding GC-rich motifs in italics) under low and high MYC expression. **(B)** Fraction (%) of MYC summits with one of the top 21 MYC:MAX-bound motifs identified *in vitro*.

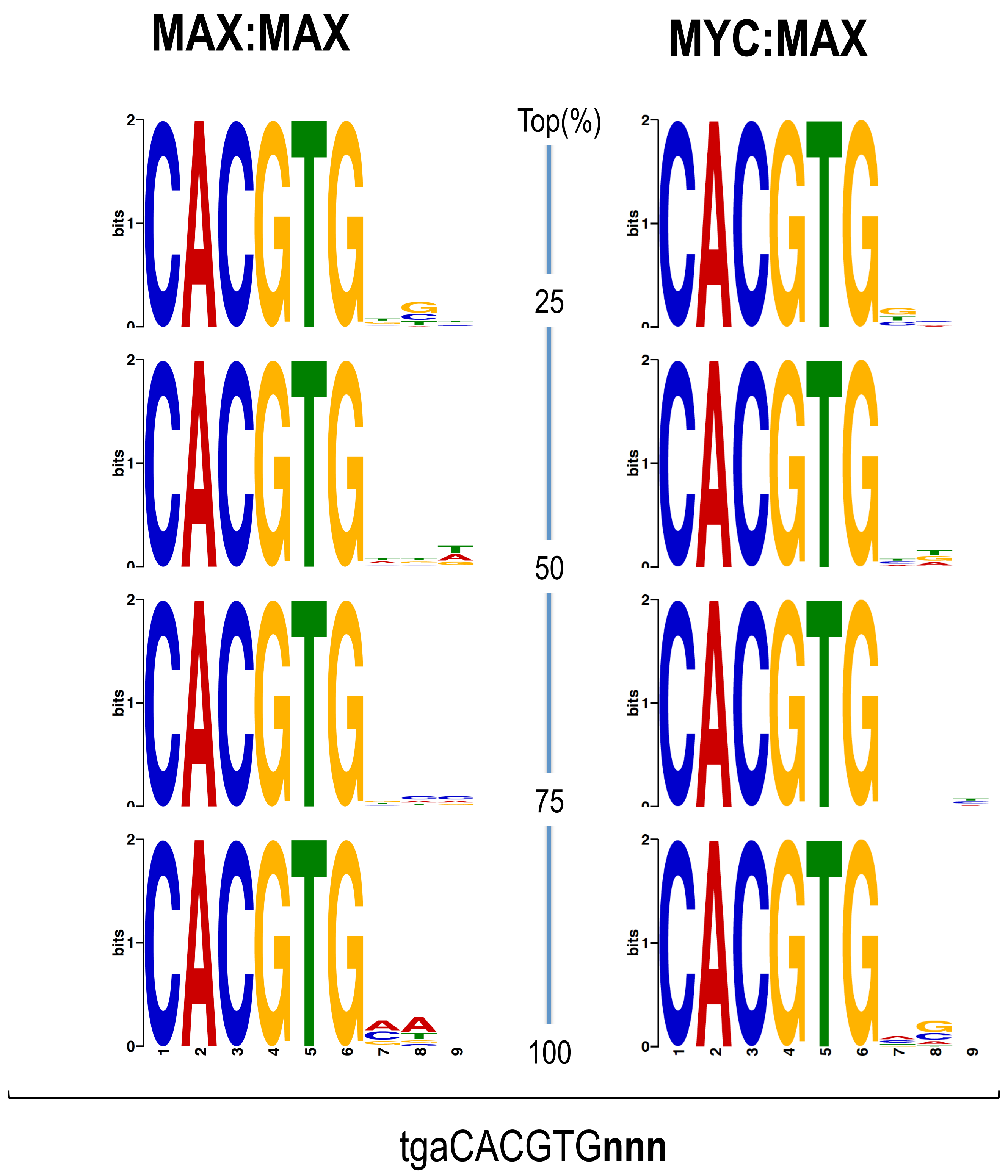
S1 Table. PBM probes sequences and MYC/MAX binding scores

S2 Table. Gene ontology of MYC-bound CME and NE sites in P493-6 cells

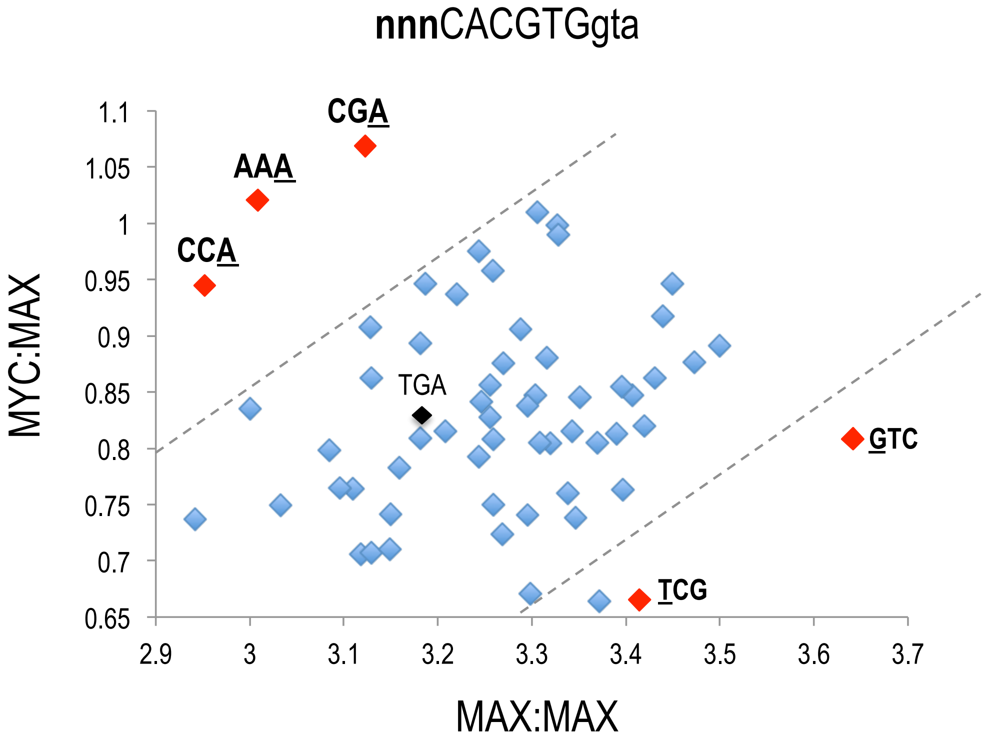
S3 Table. Gene ontology of MYC-bound CME and NE sites in U2OS cells

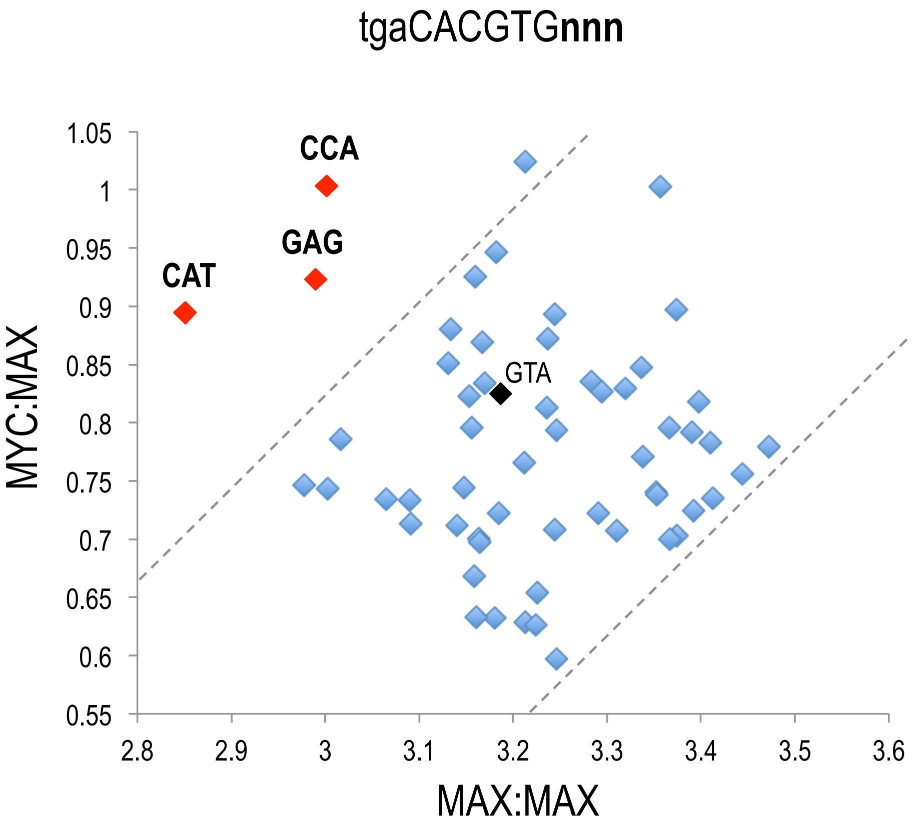

MAX:MAX
MYC:MAX
Top(%)
25
50
75
100
bits
0
1
2
1
2
3
4
5
6
7
8
9
nnnCACGTGgta

S1 Fig


MAX:MAX
MYC:MAX
Top(%)
25
50
75
100
bits
0
1
2
1
2
3
4
5
6
7
8
9
tgaCACGTGnnn

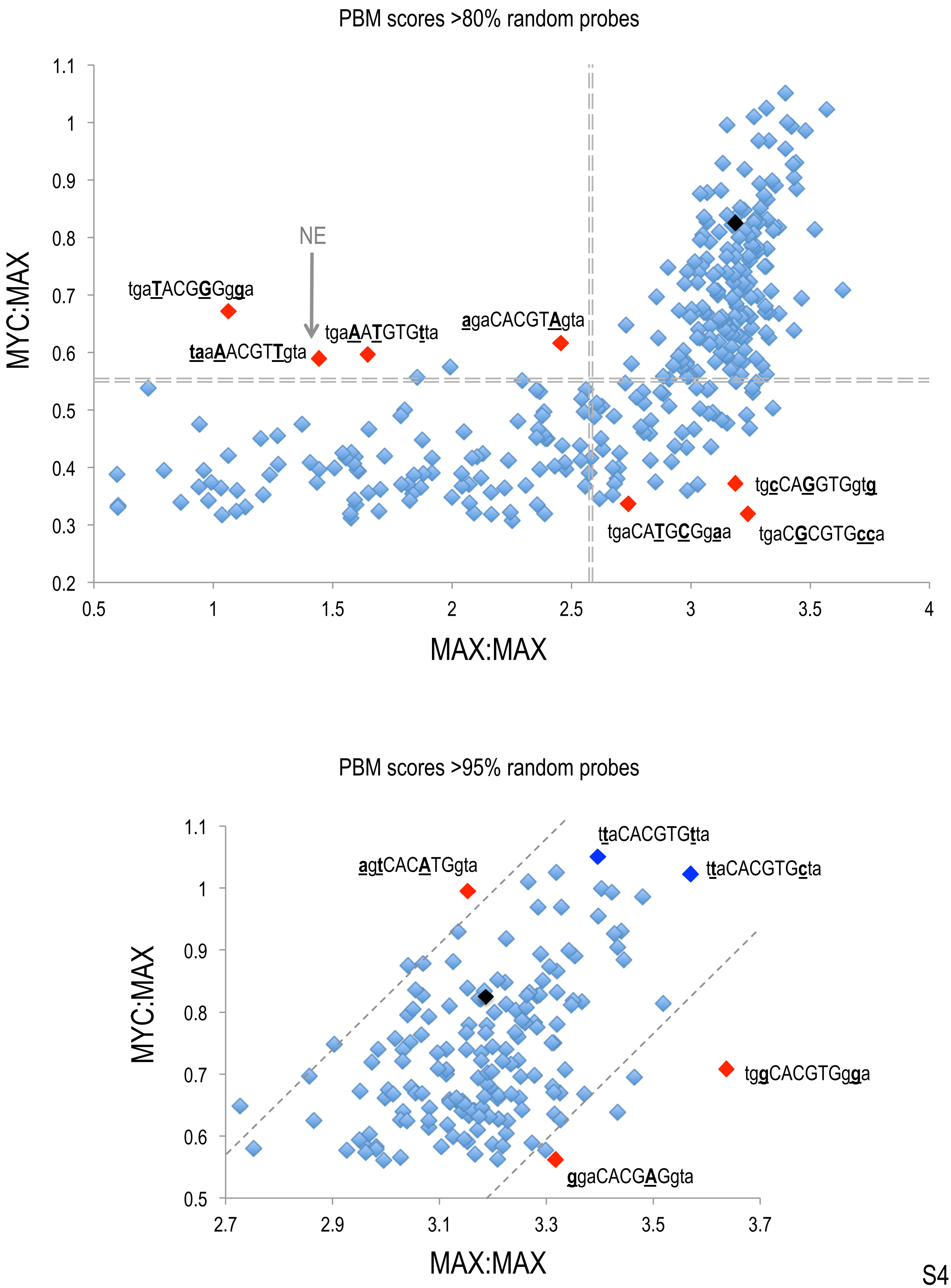
S2 Fig


nnnCACGTGgta
MYC:MAX
MAX:MAX
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
2.9
3
3.1
3.2
3.3
3.4
3.5
3.6
3.7
CGA
AAA
CCA
TGA
GTC
TCG

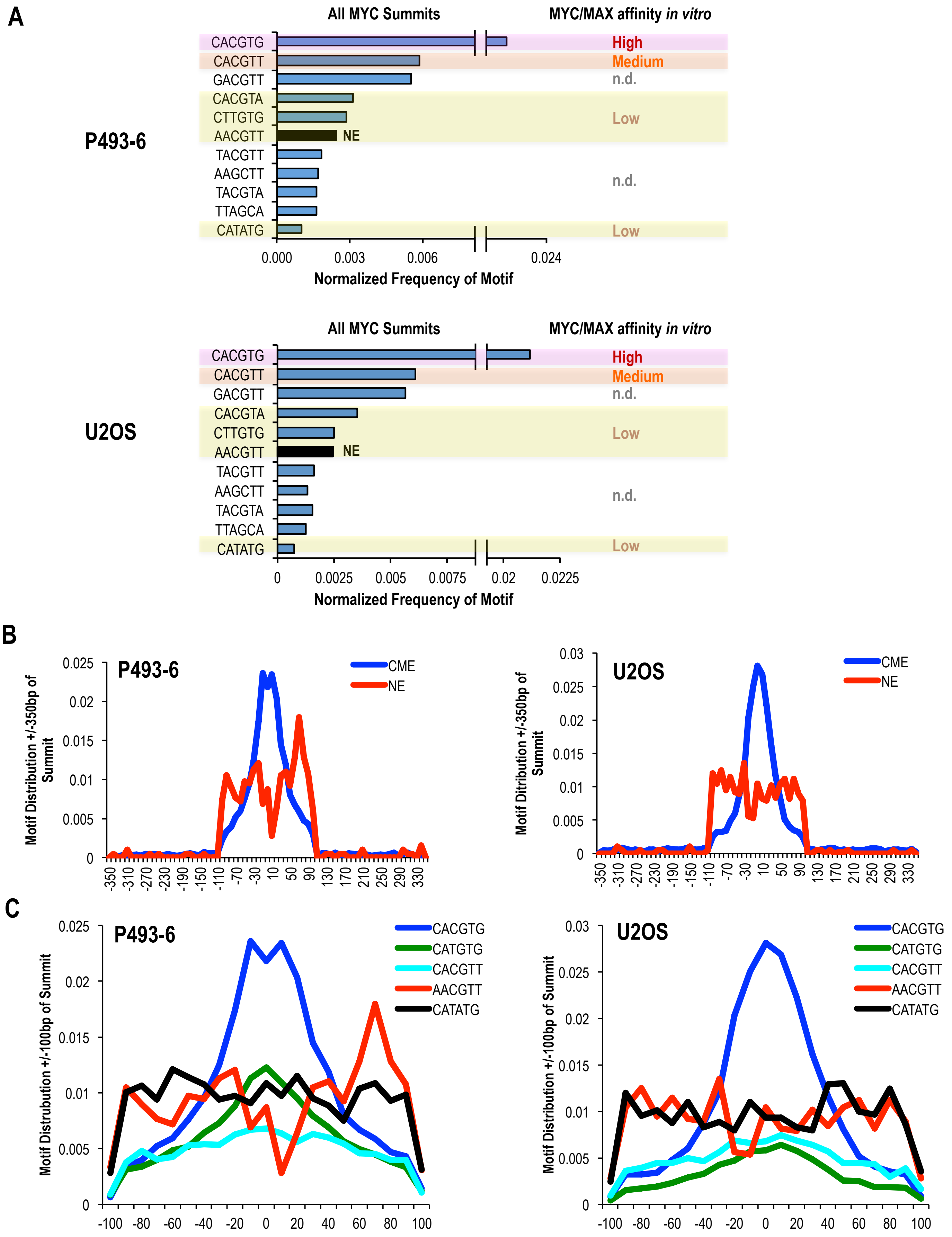

tgaCACGTG**nnn**
MYC:MAX
MAX:MAX
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
2.8
2.9
3
3.1
3.2
3.3
3.4
3.5
3.6
CCA
GAG
CAT
GTA

S3 Fig

Effects of mutations in 12-mer tgaCACGTGgta

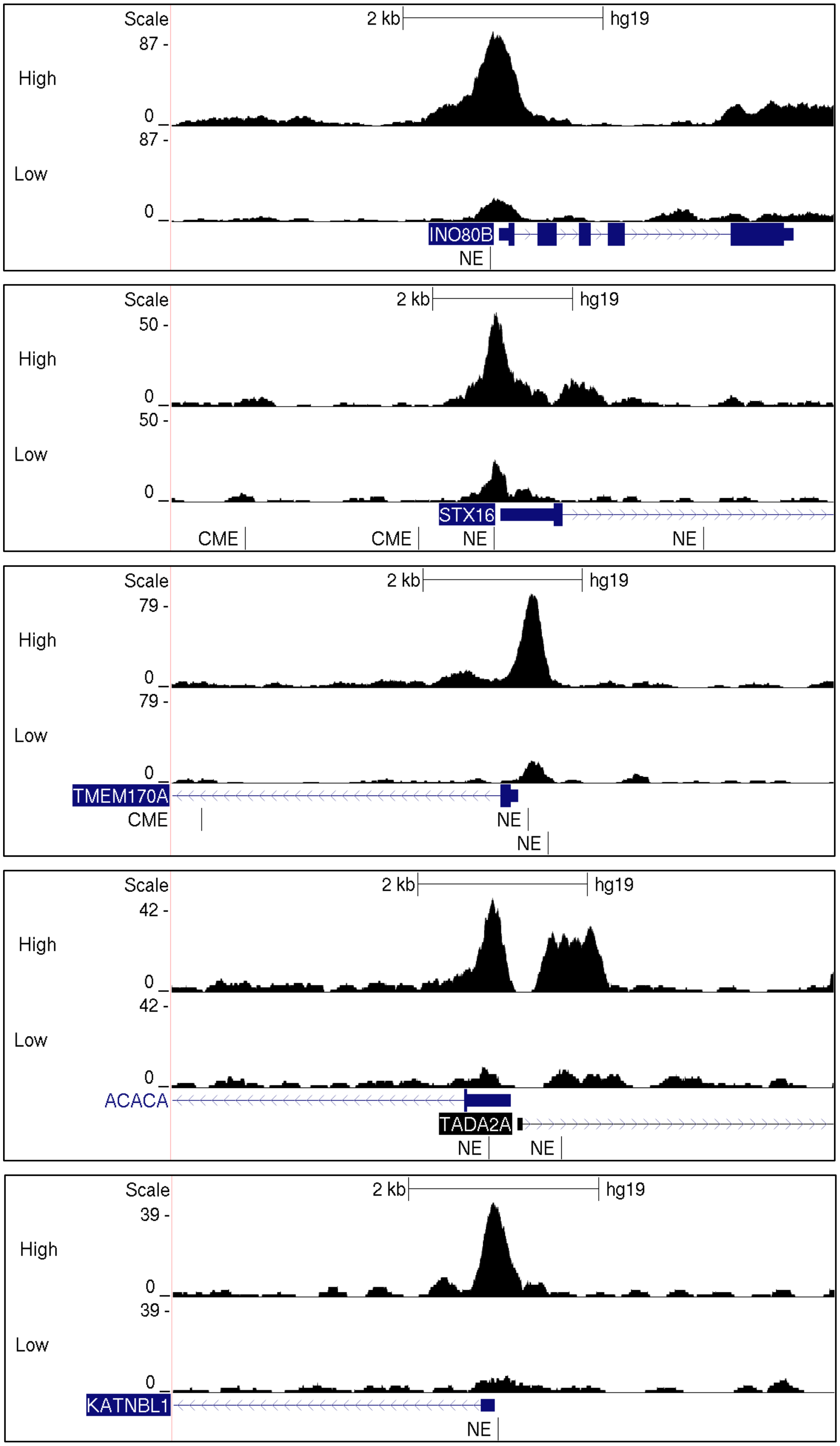


S4 Fig


A
All MYC Summits
MYC/MAX affinity in vitro
P493-6
CACGTG
CACGTT
GACGTT
CACGTA
CTTGTG
AACGTT
NE
TACGTT
AAGCTT
TACGTA
TTAGCA
CATATG
High
Medium
n.d.
Low
n.d.
Low
0.000
0.003
0.006
0.024
Normalized Frequency of Motif
All MYC Summits
MYC/MAX affinity in vitro
U2OS
CACGTG
CACGTT
GACGTT
CACGTA
CTTGTG
AACGTT
NE
TACGTT
AAGCTT
TACGTA
TTAGCA
CATATG
High
Medium
n.d.
Low
n.d.
Low
0
0.0025
0.005
0.0075
0.02
0.0225
Normalized Frequency of Motif
B
P493-6
Motif Distribution +/-350bp of Summit
CME
NE
0.025
0.02
0.015
0.01
0.005
0
-350
-310
-270
-230
-190
-150
-110
-70
-30
10
50
90
130
170
210
250
290
330
U2OS
Motif Distribution +/-350bp of Summit
CME
NE
0.03
0.025
0.02
0.015
0.01
0.005
0
-350
-310
-270
-230
-190
-150
-110
-70
-30
10
50
90
130
170
210
250
290
330
C
P493-6
Motif Distrubution +/-100bp of Summit
CACGTG
CATGTG
CACGTT
AACGTT
CATATG
0.025
0.02
0.015
0.01
0.005
0
-100
-80
-60
-40
-20
0
20
40
60
80
100
U2OS
Motif Distribution +/-100bp of Summit
CACGTG
CATGTG
CACGTT
AACGTT
CATATG
0.03
0.025
0.02
0.015
0.01
0.005
0
-100
-80
-60
-40
-20
0
20
40
60
80
100

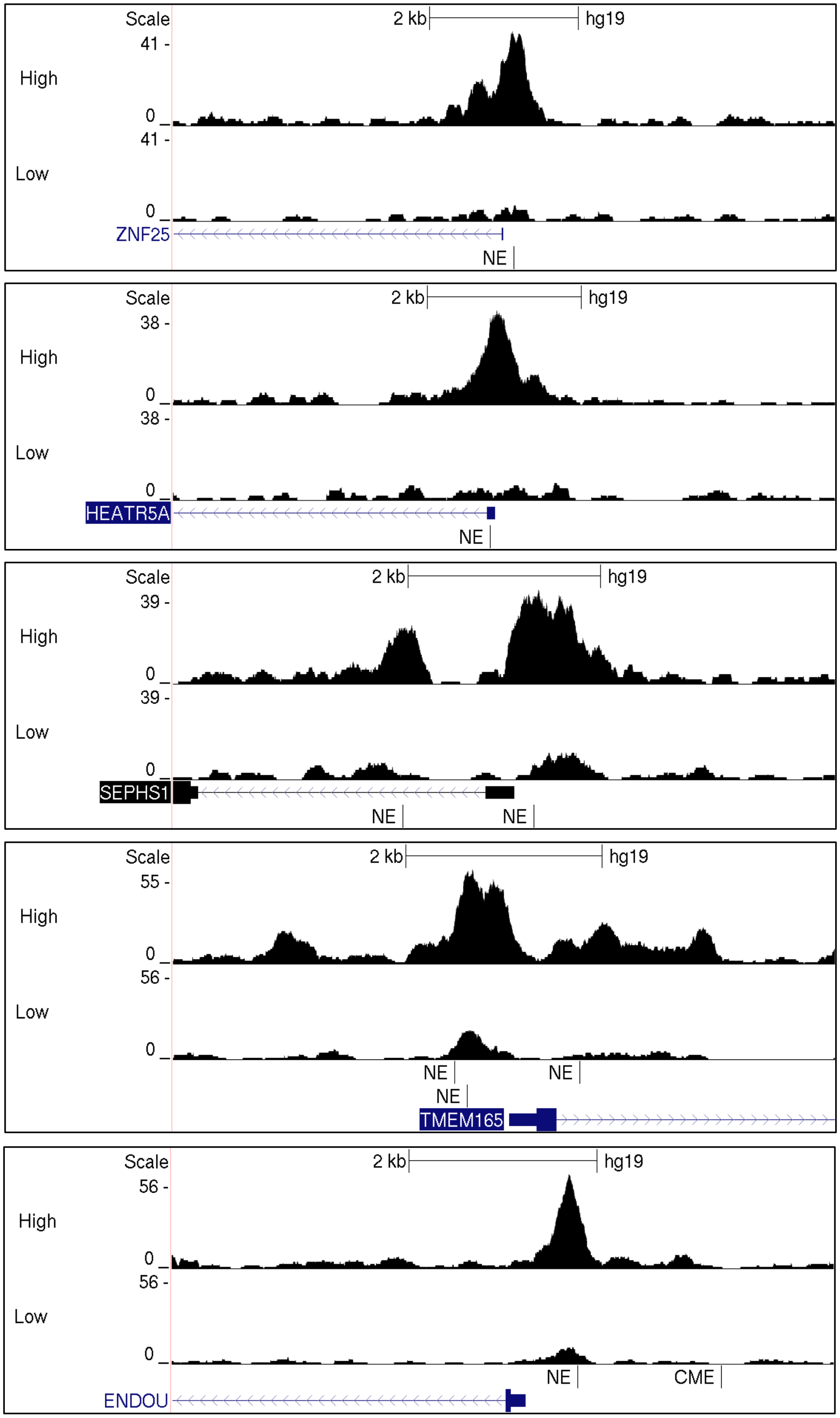
S5 Fig

Promoter +/-2kB



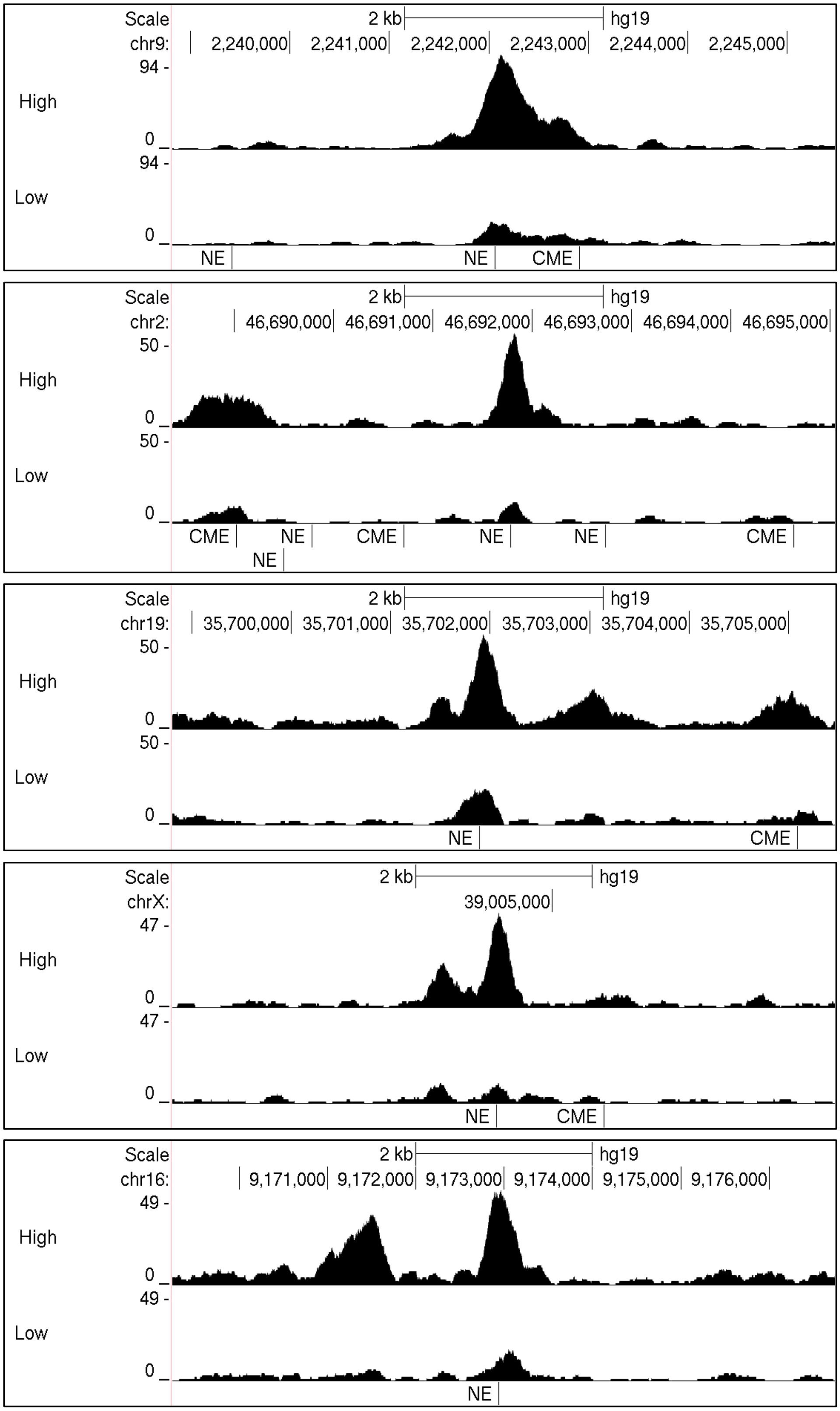
S6 Fig

Promoter +/-2kB



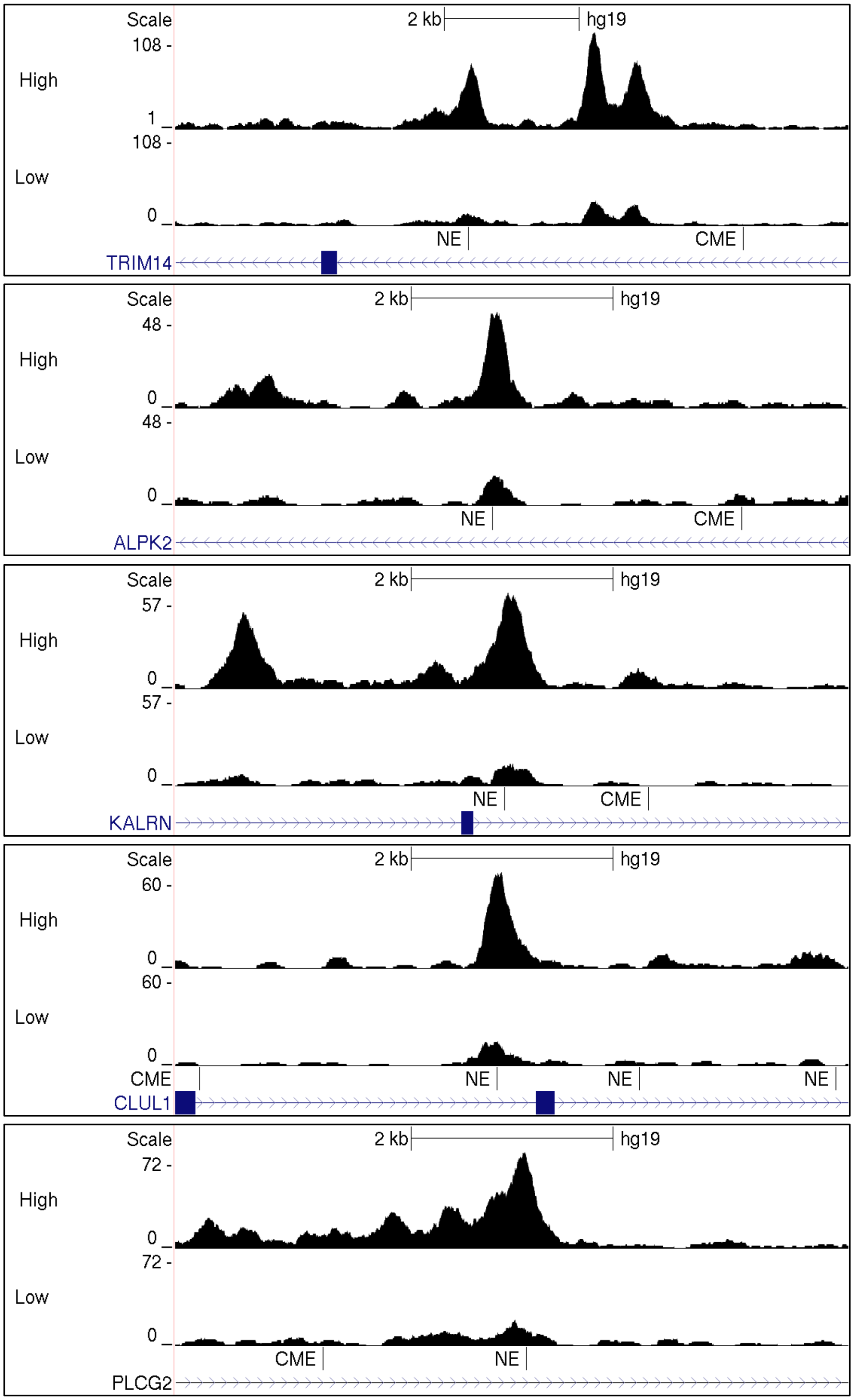
S6 Fig

Intergenic

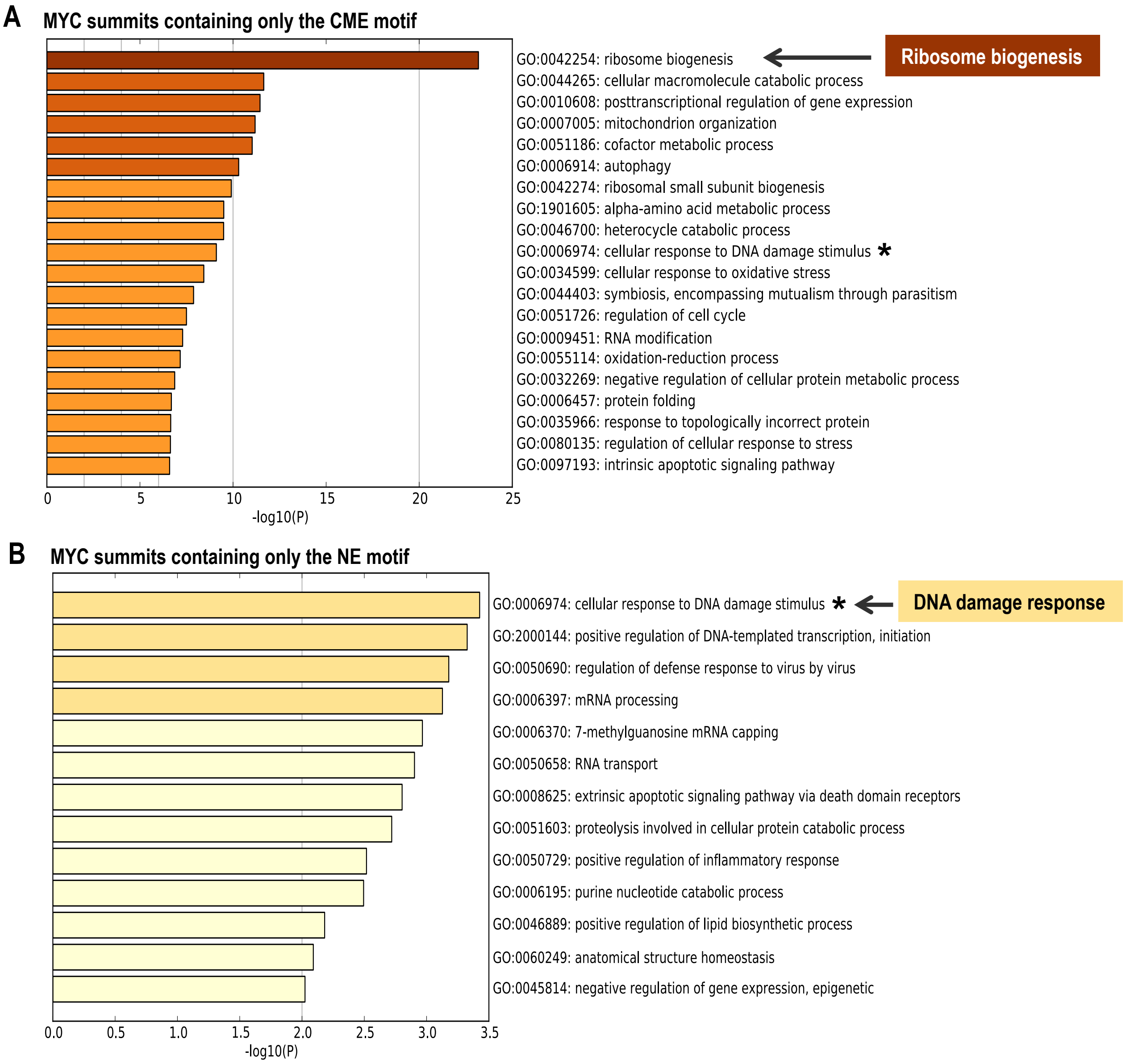


S6 Fig

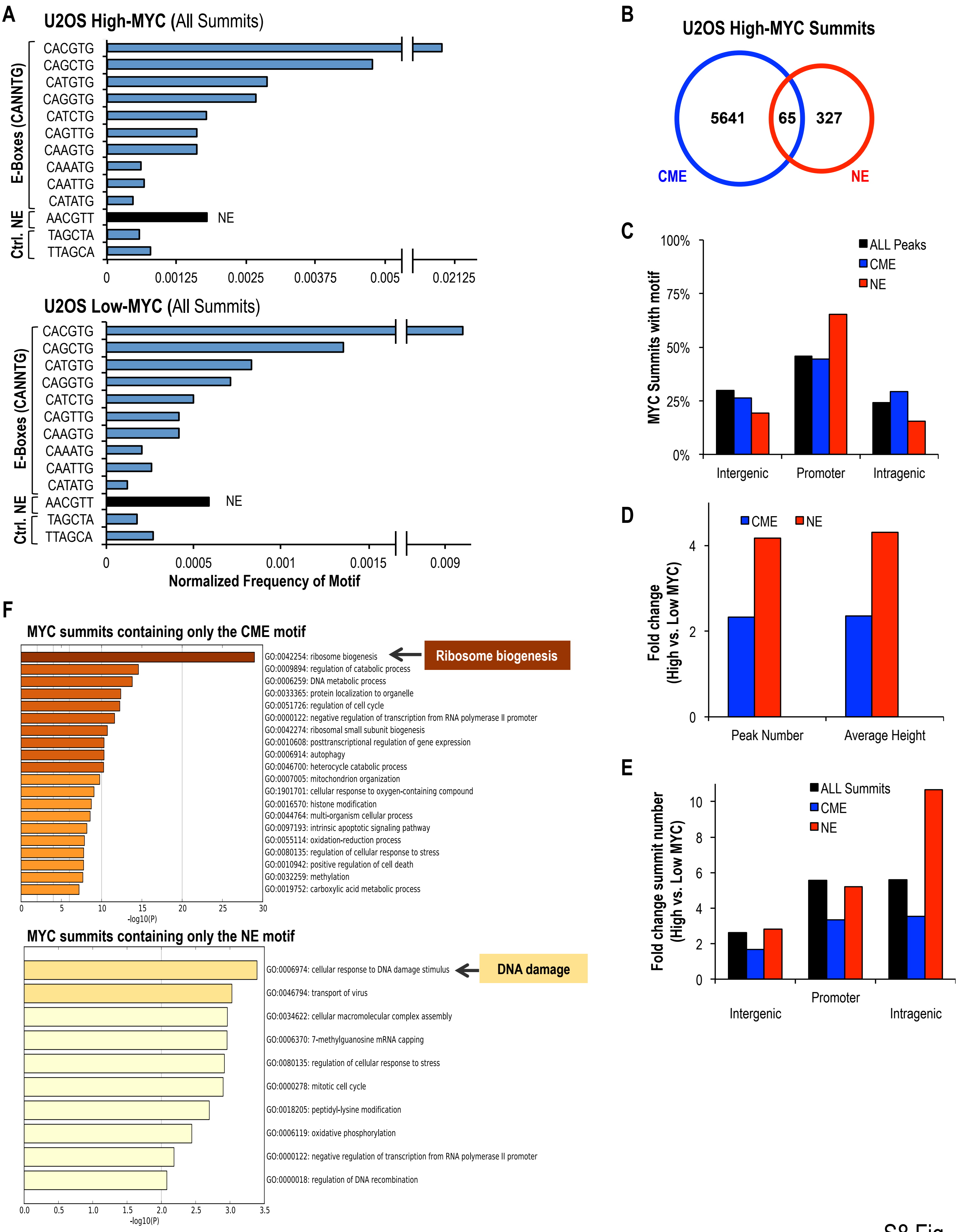
Intragenic



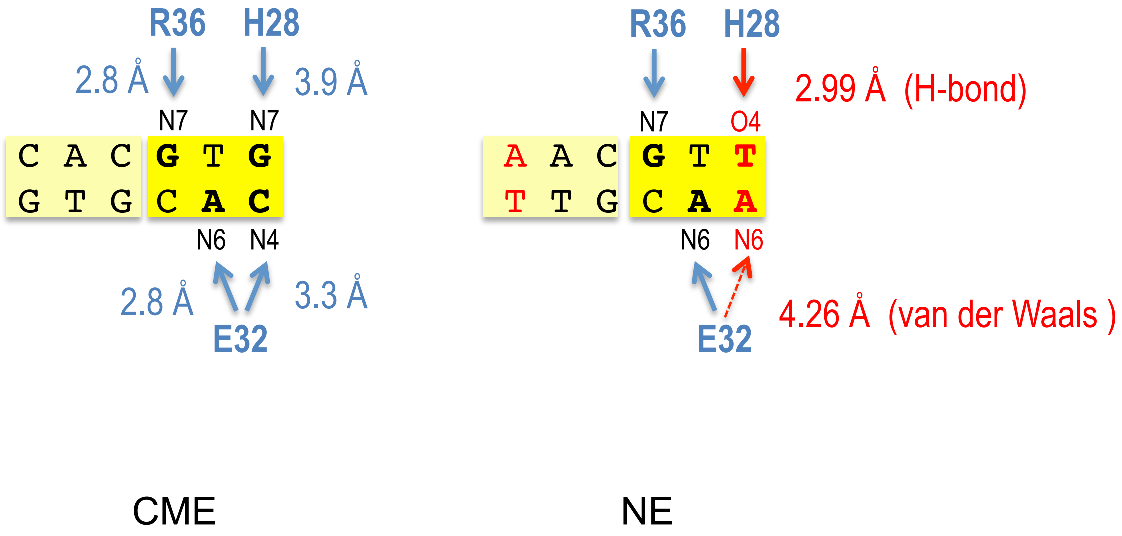
S6 Fig


A
MYC summits containing only the CME motif
GO:0042254: ribosome biogenesis
Ribosome biogenesis
GO:0044265: cellular macromolecule catabolic process
GO:0010608: posttranscriptional regulation of gene expression
GO:0007005: mitochondrion organization
GO:0051186: cofactor metabolic process
GO:0006914: autophagy
GO:0042274: ribosomal small subunit biogenesis
GO:1901605: alpha-amino acid metabolic process
GO:0046700: heterocycle catabolic process
GO:0006974: cellular response to DNA damage stimulus *
GO:0034599: cellular response to oxidative stress
GO:0044403: symbiosis, encompassing mutualism through parasitism
GO:0051726: regulation of cell cycle
GO:0009451: RNA modification
GO:0055114: oxidation-reduction process
GO:0032269: negative regulation of cellular protein metabolic process
GO:0006457: protein folding
GO:0035966: response to topologically incorrect protein
GO:0080135: regulation of cellular response to stress
GO:0097193: intrinsic apoptotic signaling pathway
0
5
10
15
20
25
-log10(P)
B
MYC summits containing only the NE motif
GO:0006974: cellular response to DNA damage stimulus *
DNA damage response
GO:2000144: positive regulation of DNA-templated transcription, initiation
GO:0050690: regulation of defense response to virus by virus
GO:0006397: mRNA processing
GO:0006370: 7-methylguanosine mRNA capping
GO:0050658: RNA transport
GO:0008625: extrinsic apoptotic signaling pathway via death domain receptors
GO:0051603: proteolysis involved in cellular protein catabolic process
GO:0050729: positive regulation of inflammatory response
GO:0006195: purine nucleotide catabolic process
GO:0046889: positive regulation of lipid biosynthetic process
GO:0060249: anatomical structure homeostasis
GO:0045814: negative regulation of gene expression, epigenetic
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-log10(P)

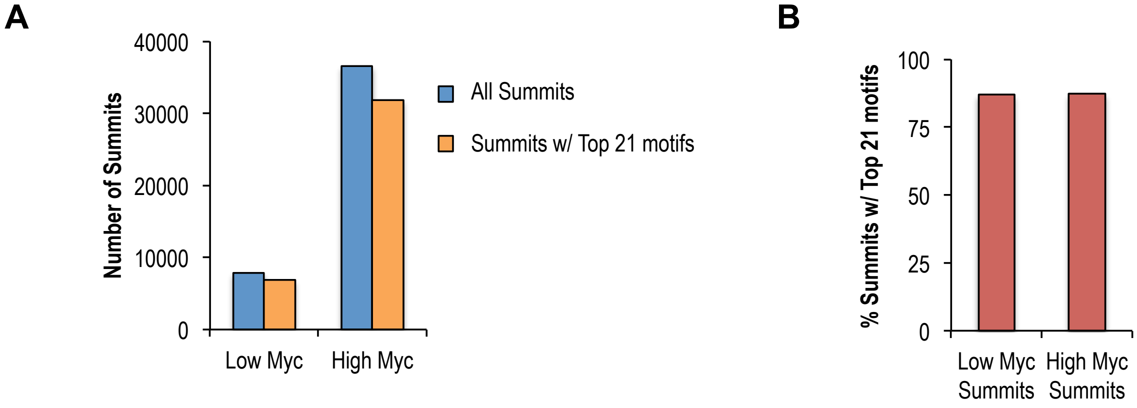
S7 Fig


A
U2OS High-MYC (All Summits)
E-Boxes (CANNTG)
CACGTG
CAGCTG
CATGTG
CAGGTG
CATCTG
CAGTTG
CAAGTG
CAAATG
CAATTG
CATATG
NE
AACGTT
NE
Ctrl.
TAGCTA
TTAGCA
0
0.00125
0.0025
0.00375
0.005
0.02125
U2OS Low-MYC (All Summits)
E-Boxes (CANNTG)
CACGTG
CAGCTG
CATGTG
CAGGTG
CATCTG
CAGTTG
CAAGTG
CAAATG
CAATTG
CATATG
NE
AACGTT
NE
Ctrl.
TAGCTA
TTAGCA
0
0.0005
0.001
0.0015
0.009
Normalized Frequency of Motif
B
U2OS High-MYC Summits
5641
65
327
CME
NE
C
MYC Summits with motif
100%
75%
50%
25%
0%
ALL Peaks
CME
NE
Intergenic
Promoter
Intragenic
D
Fold change (High vs. Low MYC)
CME
NE
4
2
0
Peak Number
Average Height
E
Fold change summit number (High vs. Low MYC)
ALL Summits
CME
NE
10
8
6
4
2
0
Intergenic
Promoter
Intragenic
F
MYC summits containing only the CME motif
GO:0042254: ribosome biogenesis
GO:0009894: regulation of catabolic process
GO:0006259: DNA metabolic process
GO:0033365: protein localization to organelle
GO:0051726: regulation of cell cycle
GO:0000122: negative regulation of transcription from RNA polymerase II promoter
GO:0042274: ribosomal small subunit biogenesis
GO:0010608: posttranscriptional regulation of gene expression
GO:0006914: autophagy
GO:0046700: heterocycle catabolic process
GO:0007005: mitochondrion organization
GO:1901701: cellular response to oxygen-containing compound
GO:0016570: histone modification
GO:0044764: multi-organism cellular process
GO:0097193: intrinsic apoptotic signaling pathway
GO:0055114: oxidation-reduction process
GO:0080135: regulation of cellular response to stress
GO:0010942: positive regulation of cell death
GO:0032259: methylation
GO:0019752: carboxylic acid metabolic process
Ribosome biogenesis
0
5
10
15
20
25
30
-log10(P)
MYC summits containing only the NE motif
GO:0006974: cellular response to DNA damage stimulus
GO:0046794: transport of virus
GO:0034622: cellular macromolecular complex assembly
GO:0006370: 7-methylguanosine mRNA capping
GO:0080135: regulation of cellular response to stress
GO:0000278: mitotic cell cycle
GO:0018205: peptidyl-lysine modification
GO:0006119: oxidative phosphorylation
GO:0000122: negative regulation of transcription from RNA polymerase II promoter
GO:0000018: regulation of DNA recombination
DNA damage
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-log10(P)

S8 Fig


R36
H28
2.8 Å
3.9 Å
N7
N7
C A C G T G
G T G C A C
N6
N4
2.8 Å
3.3 Å
E32
CME
R36
H28
2.99 Å (H-bond)
N7
O4
A A C G T T
T T G C A A
N6
N6
E32
4.26 Å (van der Waals)
NE

S9 Fig


A
Number of Summits
40000
30000
20000
10000
0
Low Myc
High Myc
All Summits
Summits w/ Top 21 motifs
B
% Summits w/ Top 21 motifs
100
75
50
25
0
Low Myc Summits
High Myc Summits

S10 Fig