RESEARCH ARTICLE

A Random Screen Using a Novel Reporter Assay System Reveals a Set of Sequences That Are Preferred as the TATA or TATA-Like Elements in the *CYC1* Promoter of *Saccharomyces cerevisiae*

Kiyoshi Watanabe, Makoto Yabe, Koji Kasahara¤, Tetsuro Kokubo*

Molecular and Cellular Biology Laboratory, Graduate School of Medical Life Science, Yokohama City University, Yokohama, Kanagawa, Japan

¤ Current address: Isotope center, Faculty of Applied Biosciences, Tokyo University of Agriculture, Tokyo, Japan

* kokubo@tsurumi.yokohama-cu.ac.jp

🔓 OPEN ACCESS

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Abstract

In *Saccharomyces cerevisiae*, the core promoters of class II genes contain either TATA or TATA-like elements to direct accurate transcriptional initiation. Genome-wide analyses show that the consensus sequence of the TATA element is TATAWAWR (8 bp), whereas TATA-like elements carry one or two mismatches to this consensus. The fact that several functionally distinct TATA sequences have been identified indicates that these elements may function, at least to some extent, in a gene-specific manner. The purpose of the present study was to identify functional TATA sequences enriched in one particular core promoter and compare them with the TATA or TATA-like elements that serve as the pre-initiation complex (PIC) assembly sites on the yeast genome. For this purpose, we conducted a randomized screen of the TATA element in the *CYC1* promoter by using a novel reporter assay system and identified several hundreds of unique sequences that were tentatively classified into nine groups. The results indicated that the 7 bp TATA element (i.e., TATAWAD) and several sets of TATA-like sequences are preferred specifically by this promoter. Furthermore, we find that the most frequently isolated TATA-like sequence, i.e., TATTTAAA, is actually utilized as a functional core promoter element for the endogenous genes, e.g., *ADE5,7* and *ADE6*. Collectively, these results indicate that the sequence requirements for the functional TATA or TATA-like elements in one particular core promoter are not as stringent. However, the variation of these sequences differs significantly from that of the PIC assembly sites on the genome, presumably depending on promoter structures and reflecting the gene-specific function of these sequences.

Introduction

In eukaryotes, transcription is highly regulated by the physical interactions between transcriptional regulators that bind to the upstream activating sequence (UAS in budding yeast) or the enhancer (in higher eukaryote) and a cohort of transcription factors that assemble on the core promoter to form the pre-initiation complex (PIC) (reviewed in [1, 2]). The core promoter region contains a variety of functionally distinct small DNA segments (termed "core promoter elements") such as the TATA element, BRE[u/d], Inr/sINR, TCT motif, DPE, MTE, bridge element, DCE, and XCPE1/2 (reviewed in [3–5]). Most of these elements are recognized by general transcription factors (GTFs), including TFIIB or TFIID [6]. However, the effect of such variation in the core promoter elements on the transcriptional regulation of eukaryotic class II genes remains unclear.

The existence of functional compatibility between the UAS/enhancer and the core promoter in various organisms has been suggested [7–10]. In budding yeast, for instance, the UAS of *RPS5* can activate the *ADH1* or *CUP1* core promoter, whereas the UAS of the latter two genes cannot activate the *RPS5* core promoter [11]. In *Drosophila*, a recent genome-wide study demonstrated that the TCT motif-containing core promoter and the TATA-Inr-MTE-DPE-containing core promoter are activated by different sets of enhancers [12]. Caudal, a key regulator of the *Hox* genes in *Drosophila*, can activate the DPE-containing core promoter but cannot activate the BRE[u]-TATA-containing core promoter [13]. Similarly, Dorsal, a key regulator of the dorsal-ventral gene regulatory network in *Drosophila*, preferentially activates the DPE-containing core promoter over the TATA-containing core promoter [14]. The latter two observations might be accounted for by a recent finding that the DPE-containing core promoter is recognized not only by TBP-containing TFIID [15], but also by TRF2 (TBP-related factor 2)-containing complexes [16]. Furthermore, recent studies show that transcription of the TCT motif-containing core promoter depends on TRF2 but not on TBP/TFIID [17, 18]. Therefore, it is likely that the core transcriptional machinery recruited by a given transcriptional regulator could become fully competent only when it is recruited to the core promoter with an appropriate structure.

In budding yeast, no obvious sequence motifs in core promoter elements other than the TATA element (i.e., TATAWAWR) [19, 20] and/or the Inr element [21] have been identified [19, 20]. Although TATA function appears to be highly conserved among eukaryotes, Inr function may differ between budding yeast and metazoans; for instance, it serves as a recognition site for the TAF1-TAF2-TBP sub-complex of TFIID in the latter [22], whereas it represents the preferred initiation site for pol II in the former [20, 23, 24]. Furthermore, there are no known TBP homologues in budding yeast, such as TRFs, although the UAS still shows core promoter specificity as mentioned above [11, 25, 26]. A recent genome-wide study using the ChIP-exo method demonstrated that nearly all TATA-less promoters (i.e., corresponding to approximately 90% of all class II genes in budding yeast) contain sequences with two or less mismatches to the TATA consensus (i.e., TATAWAWR) at the PIC assembly sites [27]. This result is consistent with our previous observation that multiple AT-rich sequences could function as a core promoter element individually in the TATA-less *RPS5* core promoter [20]. Therefore, yeast class II promoters can be divided into two subclasses: one includes promoters containing TATAWAWR and the other includes those containing similar but slightly different derivatives of TATAWAWR. These two subclasses may correspond to the SAGA- and TFIID-dominated promoters, respectively [28]. These accumulating evidences suggest that budding yeast could provide a simple model system to investigate the core promoter specificity of the UAS/enhancer, since there are few types of core promoter elements (e.g., TATAWAWR and its variants) in this organism.

As reported previously (reviewed in [7]), transcriptional regulators could function selectively with specific TATA (or TATA-like) sequences. For instance, the *HIS3* core promoter contains two different types of TATA elements, namely, T_C and T_R, and only the latter can be activated by Gcn4p and Gal4p [23]. Intriguingly, three derivatives of T_R can be activated by Gcn4p but not by Gal4p [29]. Similar functional discrimination between different TATA sequences by transcriptional regulators is observed in mammals [30, 31], indicating that the regulatory system based on sequence-specific TATA function may be widely exploited among eukaryotes from yeast to mammals. In this regard, the *CYC1* promoter is particularly interesting as it contains two structurally similar but functionally distinct TATA elements: TATAβ (A*TATATATA*T) and TATAα (T*ATATAAA*A) [32]. These two TATA sequences differ only at one position (italicized base *T* vs. *A* as shown above) when comparing the underlined regions that match to the consensus TATAWAWR. However, how such a subtle difference is recognized properly by the core transcriptional machineries to execute two different functions remains obscure.

The original aim of the present study was to identify the sequence(s) that are functionally identical to TATAβ or TATAα by a randomized screen, and delineate molecular determinant(s) between these two types of TATA elements. For this purpose, we developed a novel reporter assay system that measures the amounts of polyphosphate (polyP) accumulated in yeast cells. This method is simple and cost-effective, as it uses a spectrophotometer instead of an expensive NMR machine for the measurement [33, 34]. However, the original aim to isolate two functionally distinct TATA sequences was unsuccessful, since any functional differences between TATAβ and TATAα could not be recapitulated at the reporter gene locus. However, we found that TATAα acquires Taf1p/TFIID-dependency at this locus, indicating that the recognition mode of this element by the core transcription machineries might be altered from a Taf1p-independent (e.g., SAGA) to a Taf1p-dependent (e.g., TFIID) manner. A similar screen was hitherto conducted only for the Taf1p-independent TATA element, i.e., T_R of the *HIS3* promoter [35]. Therefore, we conducted a randomized screen to identify functional derivatives of the Taf1p-dependent TATAα element of the *CYC1* promoter. More than 600 independent sequences were isolated and classified tentatively into nine distinct groups. Comparison of these sequences with functional derivatives of T_R [35] and/or TATA or TATA-like elements identified as the PIC assembly sites in many other genes [27] indicated that functional sequences may vary significantly among genes, presumably depending on promoter structures. Furthermore, the high saturation level of our screen enabled, for the first time, the identification of a possible range of functionally competent sequences in one particular promoter. Finally, we showed that the TATTTAAA sequence isolated most frequently by the screen is used as a core promoter element for the endogenous genes. These results imply that the experimental approaches developed in this work may be effective to identify a large number of TATA or TATA-like sequences that can function specifically in a subset of genes in budding yeast, and to investigate the mechanisms underlying the gene-specific function of these sequences.

Materials and Methods

Yeast strains

Standard techniques were used for yeast growth and transformation [36]. Yeast strains used in this study are listed in S1 Table. Oligonucleotide sequences used for strain construction are listed in S2 Table.

All strains used in Fig 1 were obtained from Euroscarf (BY4741 and Y00212) or described previously (YTK6352, YTK6353, YTK6356, YTK6359, YTK6362, and YTK6365) [33].
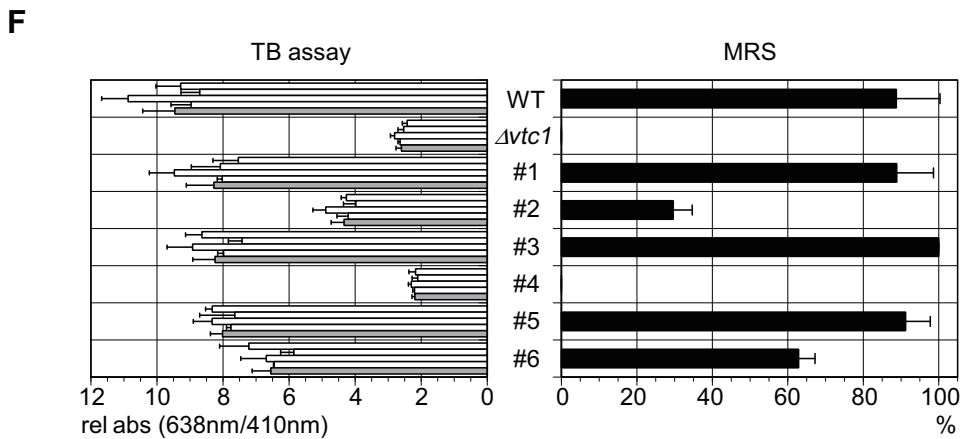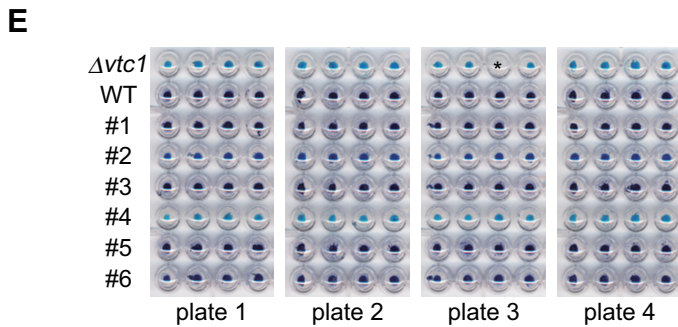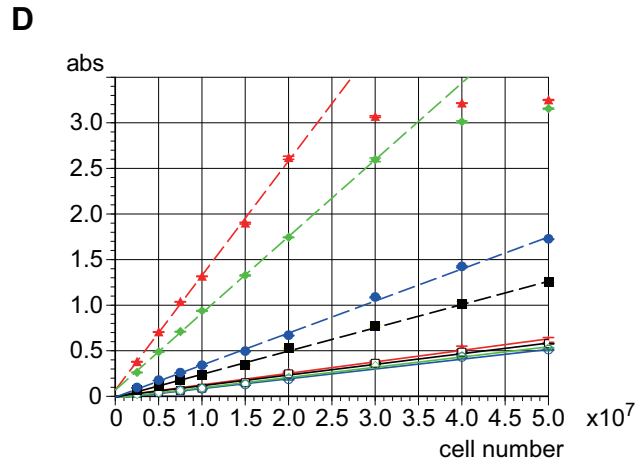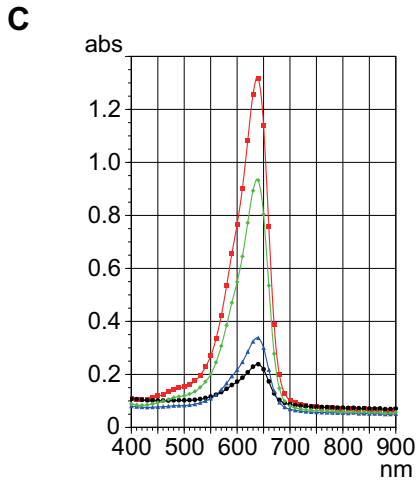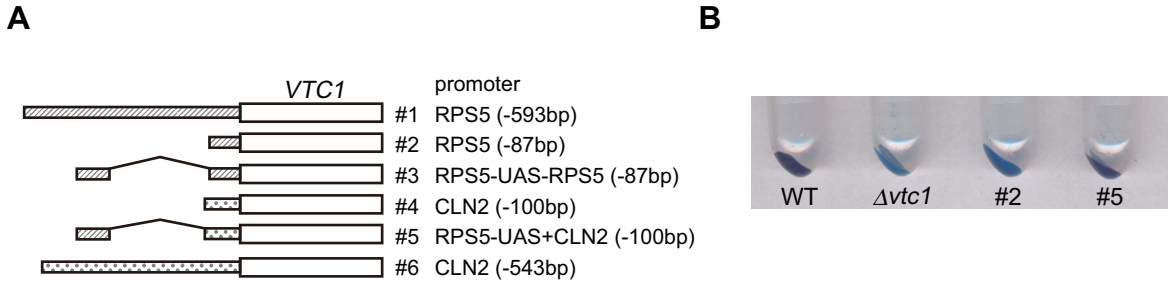
Fig 1. A novel reporter assay system measuring gene expression in yeast cells. A. Schematic diagram of the reporter constructs under the control of six different promoters, as described previously [33]. The strains tested in this figure were constructed from the wild-type strain (WT; BY4741) by replacing its endogenous *VTC1* promoter with one of these #1–6 promoters. The promoter activities (i.e., polyphosphate accumulation) of these strains were measured by MRS (magnetic resonance spectroscopy) or MRI (magnetic resonance imaging) in our previous study [33]. **B**. Photograph of the four yeast strains taken after staining with toluidine blue (TB) dye. These strains are the wild-type (WT; BY4741), *Δvtc1* (Y00212), and those carrying #2 (YTK6353) or #5 (YTK6362) promoter constructs. **C**. Absorption spectra of the set of four strains described in **B**. The spectrum of each strain is shown in a different color; WT (red), *Δvtc1* (black), #2 (blue), or #5 (green). Measurements were performed using 1×10^7 cells of each strain. **D**. Absorption at 638 nm (broken line) or 410 nm (solid line) of the set of four strains described in **B** is plotted as a function of the number of cells, as indicated in the horizontal axis. Each strain is represented by the same color described in **C**. **E**. Photograph of the eight yeast strains (WT, *Δvtc1*, and #1–6 as described in **A**), which were cultured and processed in a 96-well microplate format, taken after the TB staining. Each strain was inoculated into four wells of each plate. Here, the four plates were processed in parallel for TB staining. The third well from the left at the top row of the plate 3 (marked with an asterisk) is empty because the *Δvtc1* strain was inoculated in this well but was lost accidentally during the process. **F**. The absorption at 638 nm normalized by that at 410 nm for each strain was measured and summarized in the left panel. The average of the four wells in each plate or that of the four plates is shown as white or gray bars, respectively, along with the standard deviation. The amounts of polyphosphate in each strain were measured by MRS previously [33] and cited in the right panel as a reference.

doi:10.1371/journal.pone.0129357.g001

All strains used in other Figs were generated from BY4741 by a fusion PCR-based method [37, 38], as described below. To create YTK7548, the four sub-fragments containing *VTC1* [-320 –-11 bp] (TK7872-TK7873/BY4741), *His3MX6* (TK8297-TK10263/pFA6a-His3MX6), *CYC1* promoter [-400 –-1 bp] (TK10036-TK10037/BY4741), and *VTC1* [+1 –+320 bp] (TK9030-TK7875/BY4741) were first amplified by PCR using the primer pair/template (genomic DNA or plasmid) as described above in parenthesis, and then fused by TK7872-TK7875 to generate a 2.5 kb fragment that was used for transformation (S3 Table). This 2.5 kb-fused fragment was then inserted at the translational initiation site (A of ATG as +1) of *VTC1* of BY4741 to generate YTK7548.

To create YTK7550, the two sub-fragments containing *VTC1* [-320 –-11 bp] + *His3MX6* + *CYC1* promoter [-400 –-92 bp] (TK7872-TK10038/YTK7548) and mutated *CYC1* promoter [-116 –-1 bp] + *VTC1* [+1 –+320 bp] (TK10254-TK7875/YTK7548) were first amplified by PCR and then fused by TK2496-TK9044 to generate a 2.1 kb fragment that was inserted at the ATG site of *VTC1* of BY4741 (S3 Table). This strain (YTK7550) carried a modified *CYC1* promoter in which TATA-like sequences at sites #3 and #4 were mutated to GC-rich sequences as shown in Fig 2A.

Similarly, other strains used in Fig 2B (i.e., YTK7554, YTK7558, YTK7560, YTK7563, YTK7565, YTK7567, YTK7569, YTK7796, YTK7798, YTK7800, YTK16302, YTK16303, and YTK16304) and those used for the randomized screen (Fig 2C and 2D) were constructed by fusion PCR using the primer pair/templates described in S3 Table.

The four strains used for checking the activity of the TATTTAAA element (i.e. YTK16418, YTK16419, YTK16421, and YTK16422) were also constructed by fusion PCR, as summarized in S3 Table. Note that these strains carry the *ADE5,7* or *ADE6* promoter [-400 –-1 bp], instead of the *CYC1* promoter [-400 –-1 bp], to regulate the expression of *VTC1*.

All strains used in S1 Fig were derived from Y22.1, which carries a deletion of the chromosomal *TAF1* coding region and the wild-type *TAF1* gene in a *URA3*-based low-copy-number vector (pYN1) [39]. YTK2741 [40] and YTK3778 were generated from Y22.1 by replacing pYN1 with pM1169 (HA-tagged wild-type *TAF1*/pRS314) [41] and pM1746 (HA-tagged *taf1-N568Δ*/pRS314), respectively. The latter plasmid was constructed by site-directed mutagenesis of pM1169 using the primer TK176 [42]. To create YTK16455 and YTK16456, the three sub-fragments containing *VTC1* [-520 –-11 bp] (TK10267-TK7873/BY4741), *LEU2* (TK12262-TK6582/pUG73 [43]), or the *CYC1* promoter [-400 –-1 bp] + *VTC1* [+1 –+430 bp] (TK10036-TK4283/YTK7548) were first amplified by PCR, and then fused by TK10267-TK4283 to generate a 3.6 kb fragment that was used for transformation of YTK2741

A

```
       -180   ▼              -170 ▼        -160          -150          -140
       GCATATATATATGTGTGCGACGACACATGATCATATGGCATGCATGTGCT–
       #1(β)
          α:  TATATAAAT
          α':  TATATAAA
          m:  TACGCGTAT

       -130        ▼  -120      ▼  -110        -100          -90
       –CTGTATGTATATAAAACTCTTGTTTTCTTCTTTTCTCTAAATATTCTTTC–
          #2(α)                                    #3
              β:  TATATATAAC                    m: TAGGCCTT
              β':  TATATATATA
              m:  TAGCGCAAAC

       -80         -70          -60       -50         -40
       –CTTATACATTAGGACCTTTGCAGCATAAATTACTATACTTCTATAGACAC–
       #4
       m: TTGCGCATT

       -30
       –ACAAACACAAATACACACTAAATTAATA ATG
                                    +1
```

B
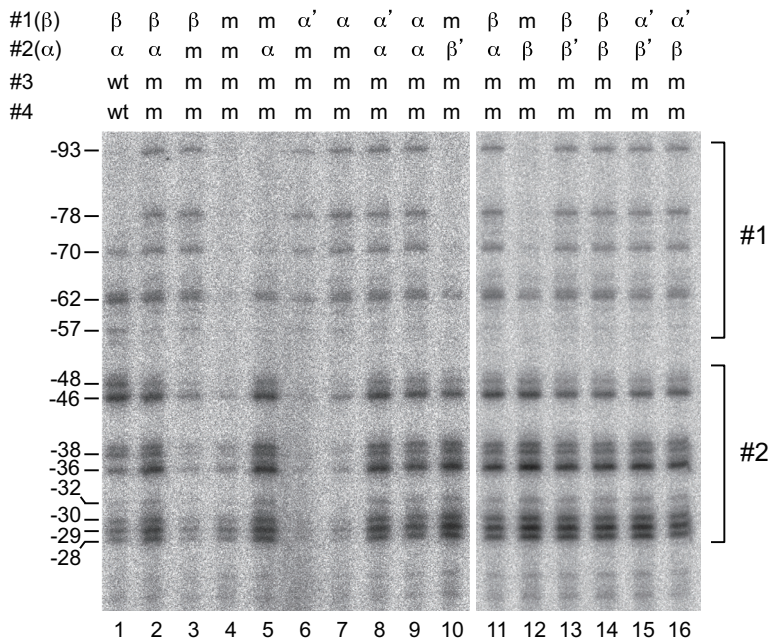
C

```
          #1m, #2random, #3m, #4m
               ▽                    ▽
          ..GTATGTATATAAAACTCT..
               #2(α)

          frame 1:  GNNNNTAAAACTC
          frame 2:  GTNNNNAAAACTC
          frame 3:  GTANNNNAAACTC
          frame 4:  GTATNNNNAACTC
          frame 5:  GTATANNNNACTC
          frame 6:  GTATATNNNNCTC
          frame 7:  GTATATANNNNTC
          frame 8:  GTATATAANNNNC
          frame 9:  GWWWWWWWWACTC
```

D

frame	1	2	3	4	5	6	7	8	9	total
screening size	8800	3520	8800	3520	8800	3520	1056	88	3520	41624
positive (>50%)	1736	290	533	108	837	860	863	88	1247	6562
% (pos/size)	19.7	8.2	6.1	3.1	9.5	24.4	81.7	100	35.4	15.8
seq determined	1390	260	412	93	702	760	805	71	1164	5657
% (seq/size)	15.8	7.4	4.7	2.6	8.0	21.6	76.2	80.7	33.1	13.6

Fig 2. Characterization and randomization of the TATA element(s) of the *CYC1* promoter at the *VTC1* locus. A. Schematic diagram of the *CYC1* promoter. The positions of the four TATA or TATA-like elements are underlined as #1(β), #2(α), #3, and #4. Each TATA or TATA-like element was replaced with a specific sequence denoted as "m" to disrupt its transcriptional activity. To replace the TATAβ element (#1(β)) with the sequence(s) derived from the TATAα element (#2(α)), the region marked with two inverted black triangles was substituted for the sequence(s) denoted as "α" or "α'". Note that "α" is identical to the TATAα element (#2(α)), whereas "α'" is identical to the sequence that had been used as the TATAα element in the original substitution study [32]. Similarly, the #2(α)-containing region marked with two inverted black triangles was substituted for the sequence(s) denoted as "β" or "β'". The substituted nucleotides that are different from those of the wild-type are shown in "bold italic". The arrows above or below the sequence indicate the transcription start site (s) (TSSs) that depend on the TATA element located at site #2 or #1, respectively, as shown in **B**. The site-specific distribution profiles of these TSSs are consistent with the results of a previous study [32]. The initiation codon of *CYC1* is marked with an open square along with the number +1 (A of ATG as +1). **B**. Identical function of the TATAβ and TATAα elements in the TSS selection of the *CYC1* promoter fused to the open reading frame (ORF) of *VTC1*. Total RNA (20μg) was isolated from strains containing wild-type (lane 1) or variously mutated (lanes 2–16) promoters (as indicated at the top) that had been cultured at 25°C in rich media (YPD) and subjected to primer extension analysis to determine the TSSs of the reporter constructs. The positions of the TSSs are indicated at the left (A of ATG as +1). Two sets of TSSs that are transcriptionally dependent on the TATA element(s) located at site #1 or #2 are marked with brackets at the right, and summarized graphically in **A**. **C**. Schematic diagram of the strategy used to construct nine randomized libraries (frames 1–9) that were screened in this study. The region containing the TATAα element (#2(α)), marked with two inverted open triangles, of the mutated *CYC1* promoter (#1m, #2(α), #3m, #4m as shown in lane 5 in **B**) was divided into nine frames (frames 1–9) and mutated randomly to "N" (A, C, G, T) or "W" (A, T), as indicated. **D**. Summary of the screen conducted in this study. The numbers of independent clones that were screened (screening size) and judged as active (positive (>50%)) are shown in the second and third lines, respectively. Of the active clones selected by the screen, the numbers of clones whose TATA sequences were successfully determined (seq determined) are shown in the fifth line. The percentages of "positive (>50%)" per "screening size" (pos/size) and those of "seq determined" per "screening size" (seq/size) were also calculated and are shown in the fourth and sixth lines, respectively.

doi:10.1371/journal.pone.0129357.g002

and YTK3778, respectively (S3 Table). Similarly, the other four pairs of *TAF1/taf1-N568Δ* strains (i.e., YTK16457/YTK16458, YTK16459/YTK16460, YTK16396/YTK16397, and YTK16398/YTK16399) were generated from YTK2741/YTK3778, as summarized in S3 Table.

Toluidine blue staining

Cells were inoculated into each well of a 96-well U-shape plate containing 100 μL of YPD media, grown overnight at 25°C without shaking, and then collected by brief centrifugation. The cell pellet was suspended in 100 μL of 80% [v/v] ethanol by shaking vigorously for 2 min. Fixed cells were collected by brief centrifugation and incubated for 10 min with 100 μL of acidic toluidine blue (TB) solution (0.1% [w/v] Toluidine Blue-O, 25% [v/v] methanol, 5% [v/v] glycerol, 5% [v/v] acetic acid, and adjusted with hydrochloric acid to pH 1.0). TB-stained cells were washed with 150 μL of DDW (deionized distilled water) three times by repeating centrifugation and resuspension. Finally, cells were suspended in 100 μL of dimethyl sulfoxide, and absorbance at 638 nm and 410 nm was measured using a Varioskan microplate reader (Thermo Scientific).

Northern blot analysis

Northern blot analysis was performed as described previously [42]. For detection of *VTC1*, *ADE5,7*, *ADE6*, and *SCR1*, DNA fragments were amplified by PCR from yeast genomic DNA, purified, and ^{32}P-labeled using random priming with the Klenow fragment (TOYOBO). The PCR primers used were as follows: *VTC1*/TK9030-TK9013, *ADE5,7*/TK8379-TK8380, *ADE6*/TK8381-TK8382, and *SCR1*/TK9507-TK10081.

Primer extension analysis

Primer extension analysis was performed as described previously [44]. The primers used were as follows: TK9613 (+27 to +3 of *VTC1*), TK8251 (+60 to +41 of *VTC1*), TK12877 (+60 to +40 of *ADE5,7*), and TK12878 (+60 to +37 of *ADE6*). The cDNA products were analyzed on a 6% polyacrylamide DNA sequencing gel. Gels were exposed to imaging plates for visualization (Typhoon FLA 7000, GE Healthcare), scanning, and quantification of electrophoretic images (ImageQuant TL software version 8.1, GE Healthcare).

Results

A novel reporter assay system to monitor gene expression in yeast cells

Vtc1p is a subunit of the vacuolar transporter chaperone (VTC) complex that mediates poly-phosphate (polyP) synthesis and transport across the membrane [45]. Our previous study showed that *VTC1* could be used as a reporter for gene expression in yeast cells through the quantification of polyP using magnetic resonance spectroscopy (MRS) and/or magnetic reso-nance imaging (MRI) [33]. Although these noninvasive methods are effective to measure gene expression in living cells [33, 34], they require expensive equipment and proficient skills for the analysis. Therefore, we devised a convenient method based on the same reporter gene system and the measurement of polyP amounts by staining with toluidine blue (TB) dye.

TB dye has been used extensively to detect polyP *in vivo* in various organisms [46, 47]. In addition, it is used *in vitro*, e.g., for visualization of polyP separated by polyacrylamide gel elec-trophoresis or for quantitative measurements of polyP synthesized enzymatically [46, 47]. As the absorption maximum of this dye is shifted to a shorter wavelength when bound to polyP (so-called metachromatic effect), the ratio of Abs_{530nm}/Abs_{630nm} is used to determine the amounts of polyP *in vitro* [46, 47]. However, such quantification has not been attempted *in vivo* because the metachromatic effect could be affected by various cellular substances such as polyanions, cations, and/or proteins [46, 47].

We previously generated a set of yeast test strains carrying *VTC1* under the control of six different strength promoters (Fig 1A) and confirmed that the *VTC1* mRNA expression levels are well correlated with the amounts of polyP measured by MRS/MRI [33]. When the two test strains (#2 and #5 in Fig 1A) and the two control strains (wild-type and *Δvtc1*) were stained with TB and washed several times with water, cells exhibited strain-specific blue colors that varied depending on the amounts of polyP (Fig 1B). Although there was no clear indication for the metachromatic shift in the absorption spectra of cell suspensions, the four strains showed significantly different Abs_{638nm} (absorption maximum) values (Fig 1C). Furthermore, the Abs_{638nm} values increased proportionally at least up to 2.5 with the cell number, and the slope of each line was strain-specific (Fig 1D). By contrast, although Abs_{410nm} values increased pro-portionally, the slopes were the same for the four strains (Fig 1D). Collectively, these results in-dicate that the ratio of Abs_{638nm}/Abs_{410nm} could represent a normalized value for the amount of polyP in a cell.

We inoculated the six test strains (#1–6 in Fig 1A) as well as the two control strains (wild-type and *Δvtc1*) in each well of 96-well plates in quadruplicate, and stained cells with TB after overnight incubation at 25°C (Fig 1E). The ratio of Abs_{638nm}/Abs_{410nm} was determined for each well and summarized in a graph (Fig 1F, left panel). The amounts of polyP assessed by this method correlated with those of polyP determined previously by MRS [33] (Fig 1F, right panel). These results indicate that the TB-staining method could be used as a novel reporter assay system to measure gene expression in yeast cells.

A random screen for the TATA or TATA-like elements that can direct transcription from the *CYC1* promoter

The *CYC1* promoter is one of the most extensively characterized promoters in *Saccharomyces cerevisiae* [32, 48–55] and has been shown to be a relatively rare (at least in budding yeast) poised promoter where TBP and RNA polymerase II are pre-loaded before activation [51–53, 55]. In addition, it contains several TATA or TATA-like elements, two of which, i.e. TATAβ (TATATATA located at site #1 from -177 to -170; abbreviated as β/#1) and TATAα (TATAT AAA located at site #2 from -123 to -116; abbreviated as α/#2), are principally responsible for

transcription from multiple transcriptional start sites (TSSs) in the upstream or downstream region, respectively [32, 48] (Fig 2A). These two TATA elements are thought to have different functions because only the upstream one is functional when the same TATA sequences are duplicated (i.e., β/#1-β/#2 or α/#1-α/#2) while both are functional when different TATA sequences are combined (i.e., β/#1-α/#2 or α/#1-β/#2) [32]. To understand this functional difference, we sought to identify any other TATA or TATA-like sequence(s) that could function as TATAβ or TATAα in the *CYC1* promoter by using a random screen based on the TB-staining method.

First, to determine whether TATAβ or TATAα have different functions also at the reporter gene (*VTC1*) locus, the wild-type or mutated *CYC1* promoter (-400 bp—+1 bp) was integrated in front of the *VTC1* coding region. The TSS(s) of the *VTC1* reporter gene were determined by primer extension analyses (Fig 2B). No differences between the function of TATAβ and TATAα at the *VTC1* locus were detected, as the TSSs of β/#1-β/#2 (lane 14) or α/#1-α/#2 (lane 9) were the same as those of β/#1-α/#2 (lane 2 and 11). As the flanking sequences were slightly modified in the original study [32], we reconstructed the strains by exploiting the same sequences as the original ones (designated here as TATAβ' and TATAα'). The results showed that the TSSs of β/#1-β'/#2 (lane 13) or α'/#1-α/#2 (lane 8) were the same as those of β/#1-α/#2 (lane 2 and 11), α'/#1-β'/#2 (lane 15), and α'/#1-β/#2 (lane 16), leading to the same conclusion that all four TATA elements (i.e., β, α, β', α') have an identical function at the *VTC1* locus, which is different from their function at the original *CYC1* locus. A previous study [32] showed that, when the *CYC1* promoter contains a single TATA at either site #1 or #2, position-specific and sequence-independent conventional TATA function is observed at the *CYC1* locus. This function was confirmed at the *VTC1* locus, as the TSSs of β/#1-mut/#2 (lane 3) or mut/#1-α/#2 (lane 5) were the same as those of [α or α']/#1-mut/#2 (lane 6, 7) and mut/#1-[β or β']/#2 (lane 10, 12), respectively. Therefore, we could recapitulate the sequence-independent function but not the sequence-dependent function of these two TATA elements at the *VTC1* locus. Furthermore, the functions of TATAβ and TATAα were found to be TFIID-dependent at the *VTC1* locus (S1 Fig), which is contrary to the previous assumption that the *CYC1* promoter is TFIID-independent [25, 42, 55, 56]. In fact, we confirmed that the endogenous *CYC1* promoter is TFIID-independent even in the strains used in this study (data not shown). Thus, the chromosomal context and/or promoter structure outside the substituted region may alter the selectivity for the core transcriptional machineries (e.g., TFIID vs. SAGA), thereby generating the aforementioned discrepancies between the previous [32] and current studies. In any case, information on preferred TATA or TATA-like sequence(s) in a TFIID-dependent promoter is limited [20, 35]. Therefore, despite its deviation from the original intention, we performed a random screen to identify functional core promoter element(s) that could substitute for TATAα in the *CYC1* promoter (mut/#1-α/#2-mut/#3-mut/#4) bearing nearly identical transcriptional activity to the wild-type (β/#1-α/#2-wt/#3-wt/#4 in Fig 2B and S1 Fig).

The TATATAAAACT (TATAα is underlined) sequence was divided into nine frames (frame 1–9), and four (N = A/C/G/T) or eight (W = A/T) nucleotides of each frame were randomized by PCR as shown in Fig 2C. The screen was conducted in a 96 multi-well plate format, each of which contained 88 (or less) distinctly randomized strains as well as the three control strains (in duplicate), i.e., wild-type (WT, positive control), a TATA mutant (TAGCGCAAA CT, negative control), and Δvtc1. The TB-staining method is quantitative enough to compare the strains processed in the same plate (i.e., under the same staining and washing conditions). However, this method could not provide an absolute value for the activity, and the results obtained at different experimental dates were difficult to compare. Therefore, the activity of each clone was evaluated by comparing it with those of the control strains processed in the same plate at the same time. In the screen, the clones showing more than the average (50%) activity

of the WT and TATA mutant were judged as potentially active (Fig 3A), and those that passed through the selection twice (Fig 3B) were stored as active clones and subjected to the sequence analysis. Active clones were significantly enriched through such multiple selection processes (compare Fig 3A and 3B). The results of the screen are summarized in Fig 2D. In total, 6,562 active clones (15.8%) were isolated from 41,624 randomized clones. Among them, 5,657 clones (13.6%) were successfully sequenced and tentatively classified into the nine groups, as described below.

Requirements for the 7th or 8th position of the TATA element in the *CYC1* promoter

A previous study revealed the consensus sequence TATAWAWR (W = A/T, R = A/G) for the TATA element [19]. However, the clones obtained in the screen for frame 8 were all active (Fig 3C), indicating that the requirements for the 8th position (i.e., "R" (R = A/G) in the consensus sequence) are greatly alleviated in the *CYC1* promoter. To confirm this, we sequenced 71 active clones and found that no sequence bias was generated during the construction process (Fig 4A, S6 Table). Thus, we conclude that the length of the TATA element in the *CYC1* promoter is 7 bp, unlike that of the consensus TATA (8 bp). A similar conclusion was reached in a previous study in which 23 TATAα mutants were assayed *in vivo* and *in vitro* [54].

In the screen for frame 7, 805 (76.2%) active and successfully sequenced clones were obtained from 1,056 randomized clones (Fig 2D). The isolation frequency of active clones was significantly higher than that expected (approx. 50%) from "W" (W = A/T) in the consensus TATA, suggesting that the requirements for the 7th position are also alleviated in the *CYC1* promoter. A heat map for these 805 clones clearly showed that the 7th position in this promoter prefers "D" (D = A/G/T) instead of "W" (W = A/T) in the consensus TATA (Fig 4B, S5 Table). Our study revealed for the first time that the *CYC1* TATA disfavors "C" at the 7th position, although previous findings showed that the *CYC1* TATA is active when it has "G" at the 7th position [54].

Classification of active TATA or TATA-like sequences obtained by the screen

In the screen for frames 1–6 and 9, 4,781 (11.8%) active and successfully sequenced clones were obtained from 40,480 randomized clones (Fig 2D). These clones comprised 601 different sequences (Fig 5B) and were summarized in a heat map with the normalized value (left panel in Fig 5A, S4 Table). Here, the normalized value for each sequence was calculated by taking into account which frame(s) could generate such a sequence, since the expected isolation frequency should vary depending on the population size of each sequence in the library (see the details in the legend for Fig 5A, S7 Table).

To further characterize 601 potentially active sequences, we sought to classify 4,781 clones into several groups. As described above, TATAWAWR (consensus TATA) and TATAWAD were both shown to be active in the *CYC1* promoter. Thus, 28 sequences (614 clones) that matched TATAWAWR were first extracted, followed by 57 sequences (1,193 clones) that matched TATAWAD successively (class I and II in Fig 5A and 5B, S2A and S2B Fig). It was previously shown that the GA element (GAAAA) is a novel type of core promoter element that does not co-occur with the TATA element [57]. Thus, we next extracted 18 sequences (29 clones) that match GAAAA (class III in Fig 5A and 5B and Nx9 in S2C Fig) from the rest (2,974 clones). Our classification analyses searched for the sequence(s) within the region from -123 to -115 (9 bp: originally TATATAAAA) that included "N" (frame 1–6) or "W" (frame 9) (Fig 2C; TATAα is underlined). As the 5′-upstream nucleotide adjacent to this region was "G"
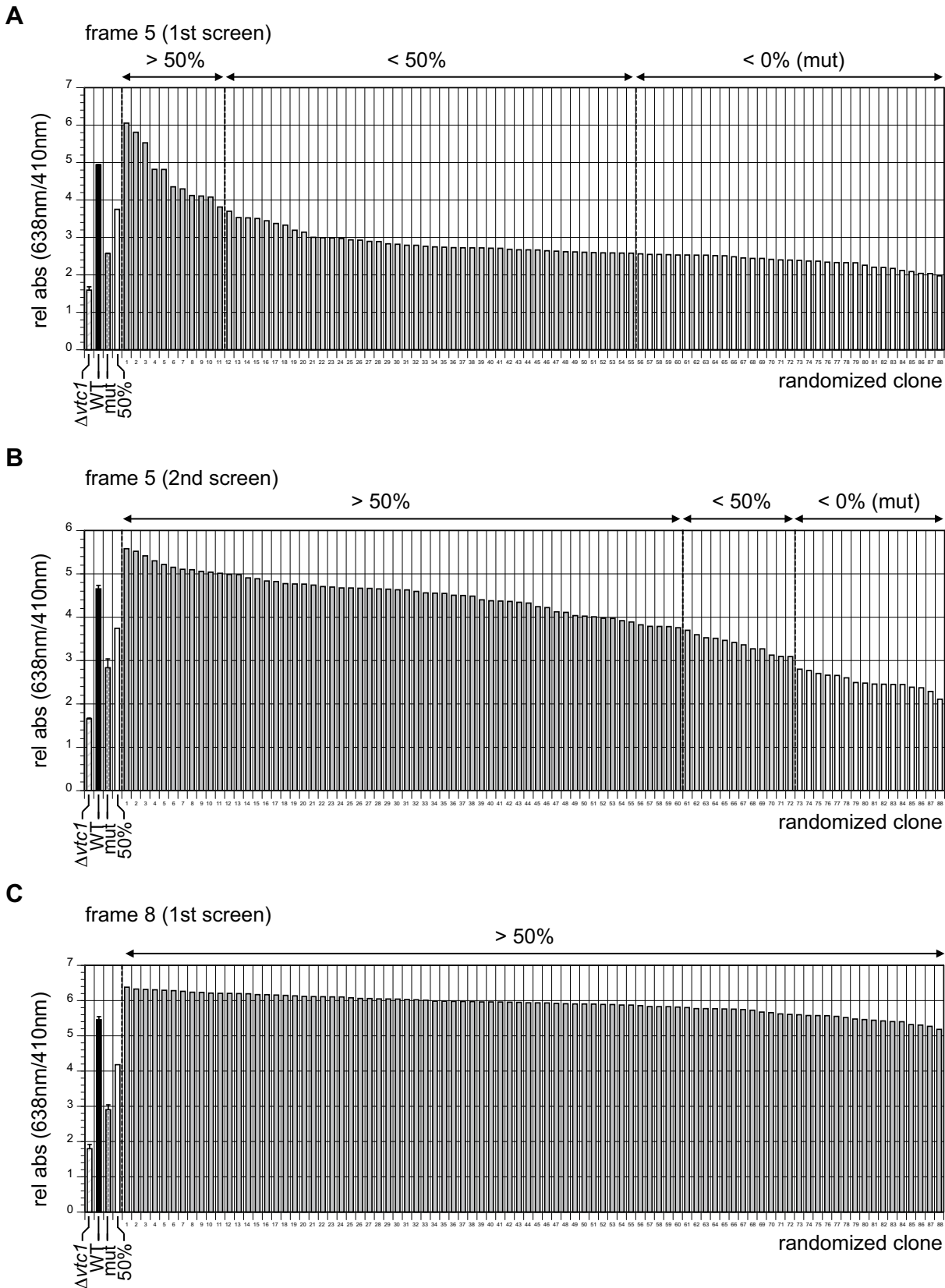
Fig 3. Active clones isolated by the screen when frame 5 or 8 was randomized. A. One part (corresponding to 88 independent clones) of the first screen for frame 5 (8,800 independent clones in total) conducted in a 96-well microplate format is shown as an example to demonstrate the profiles of the three types of clones (>50%, <50%, <0%) that can be obtained in the screen. Promoter activities (i.e., relative absorption at 638 nm/410 nm) of the wild-type strain (#1m, #2α, #3m, #4m) and the mutated strain (#1m, #2m, #3m, #4m) are set as 100% (second bar) and 0% (third bar), respectively. The clones showing >50% activity (more than the value of the fourth bar = 50%) were judged as active and pooled for the second screen as shown in **B**. The clones showing <50% activity (including those less than 0%) were judged as inactive and not used for further analyses. All clones are aligned in order of promoter activities. Note that the promoter activity of the mutated strain (0%) is not completely inactive, as it is stronger than that of the *Δvtc1* strain (first bar). Consistent with this, a significant number of clones showed promoter activities <0%. **B**. A part of the second screen for frame 5 is shown as an example to demonstrate the profiles of the three types of clones (>50%, <50%, <0%). The population size of the >50% clones becomes larger when compared with that of the first screen as shown in **A**, indicating that the active clones were enriched significantly via multiple screening. **C**. First screen for frame 8 (88 clones in total). The results indicate that all clones were active, contrary to those for frame 5 as shown in **A**.

(Fig 2C), we also searched for the sequences that matched GAAAA within the 1 bp-extended region (10 bp: GTATATAAAA) and extracted 25 sequences (54 clones) (G/Nx9 in S2C Fig). The activities of these sequences were considerably weaker than those of TATAWAWR or TATAWAD (compare the heat maps in S2 Fig). In addition, the reproducible (i.e., isolated more than twice) clone numbers for these sequences were significantly smaller than those for the latter two TATA-containing sequences, suggesting that the GA element may not be functional in the *CYC1* promoter.

Next, we searched for the TATA-containing sequence(s) other than TATAWAWR and TATAWAD by using TATAN (N = A/G/C/T) as an initial probe and found that the sequences containing TATAW (W = A/T) were considerably more active than those containing TATAS (S = G/C) by examining the color profile and clone number in the heat map (data not shown). Furthermore, TATAWD (D = A/G/T) was found to be a better sequence motif than TATAW, since the sequences containing TATAWD were significantly more active than those containing TATAWC (data not shown), although the latter matched another less active sequence motif, i.e., TATATCWD (class VII), as described later. Based on the same reason, TATAWDW was a better motif than TATAWD when the heat maps of TATAWDW (W = A/T) and TATAWDS (S = G/C) were compared (data not shown). Finally, TATAWKW (K = G/T) was chosen as a sequence motif representing class IV, as TATAWAW was already included in TATAWAD (class II), and further extension was not informative (i.e., no differences were found for TATAWKWA, -C, -G or -T) (data not shown). We extracted 71 sequences (715 clones) that matched this motif (class IV in Fig 5A and 5B and S3A Fig) and found that class IV was less active than class I and II (S2A, S2B and S3A Figs).

After exclusion of the sequences belonging to class I–IV from the total (4,781 clones), the top five most frequently isolated sequences among the remaining clones (2,230 clones) contained TTAAA (data not shown). Thus, we next searched for TTAAA-containing sequence(s) and found 945 clones containing TTAAAW (W = A/T), while no clones containing TTAAAS (S = G/C) were identified (data not shown). In addition, most of them (751 clones) had TTAAAW at the 3′-end (i.e., further extension was not informative), indicating that TTAAAW is another sequence motif. In fact, 74 sequences (945 clones) that matched this motif (class V in Fig 5A and 5B, S3B Fig) were extracted, and class V was less active than class I and II, but more active than class IV (S2A, S2B, S3A and S3B Figs).

No significant sequence features were detected in the remaining clones (1,285 clones) except that they were all W (= A/T)-rich. When 138 sequences (865 clones) that matched Wx6 (class VI) and then 135 sequences (332 clones) that matched Wx5 (class VII+VIII as described below) were extracted successively, we noticed that the latter class included a significant number of TATATC-containing sequences (data not shown). As most of these sequences could be represented as TATATCWD (D = A/G/T), the latter Wx5 class was divided into two subclasses, i.e., TATATCWD (22 sequences/115 clones) and Wx5* (= Wx5 other than

Fig 4. Graphical summary of active clones isolated by the screen for frames 7 and 8. A. Characterized properties (activity, sequence, and isolation frequency) of the active and successfully sequenced 71 clones that were isolated by the first screen for frame 8 (S6 Table) are graphically summarized in a heat map. Each clone is represented by a small square painted with an appropriate color according to its promoter activity, as indicated at the top (WT = 100%, mut = 0%), and arranged to a specific site of this heat map according to its sequence of the randomized region (TATATAAN^{1st}N^{2nd}N^{3rd}N^{4th}). Note that the squares are aligned in order of promoter activities from left to right (i.e., each bar has a red to gray color as shown more evidently in **B**, **D**). The numerals along the horizontal axis indicate the number of clones isolated by the screen. The area enlarged in **C** for readability of the sequence is marked with a bracket. **B**. Characterized properties (activity, sequence, isolation frequency) of active and successfully sequenced 805 clones that were isolated by the second screen for frame 7 (S5 Table) are summarized in a heat map. Small squares representing each clone with an appropriate color are arranged in the same way as described in **A**. The area enlarged in **D** is also marked with a bracket. **C**. The area bracketed in **A** was enlarged. **D**. The area bracketed in **B** was enlarged.
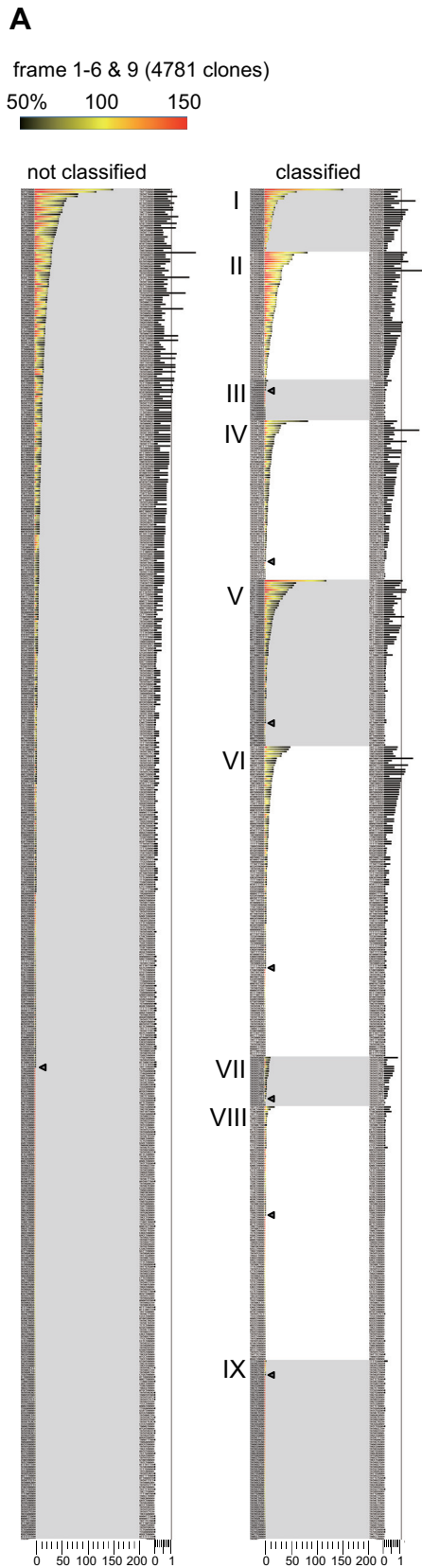
doi:10.1371/journal.pone.0129357.g004

TATATCWD; 113 sequences/217 clones) (class VII and VIII, respectively, in Fig 5A and 5B, S4B and S4C Fig). The remaining 80 sequences (88 clones) were classified as "others" (class IX in Fig 5A and 5B and S4D Fig). It is likely that class IX is not functional, since most of the clones (73 clones) were isolated only once in the screen, presumably because of experimental noise. Collectively, we conclude that active sequences/clones could be classified into seven groups, namely, class I (TATAWAWR), II (TATAWAD), IV (TATAWKW), V (TTAAAW), VI (Wx6), VII (TATATCWD), and VIII (Wx5*) in the *CYC1* promoter. However, each class may include a certain number of false-positive clones (e.g., those isolated only once) since the TB-staining method is semi-quantitative and could not yield an absolute value for the activity. Consistent with this, even the same sequences (including the wild-type "TATATAAAA" sequence that was isolated most frequently in the screen) exhibited apparently varied activities (see color profiles in Fig 5A and S2–S4 Figs).

Comparison of active sequences isolated by the screen with those identified as the PIC assembly sites throughout the genome

Recently, Rhee and Pugh determined the precise positions for a number of PICs (6,045) on the yeast genome by using a ChIP-exo method [27]. They found that nearly all of the PICs were assembled over the region including 8 bp DNA (judged from the position of TFIIB) that showed two or less mismatches to the consensus TATAWAWR even in the TATA-less promoters. Mismatched sequences are designated formally as TATA-like elements [27].

Here, we sought to compare the sequences isolated by the screen with the TATA or TATA-like elements identified previously by the ChIP-exo method. Given that the entire set of these genome-wide identified elements were not included in our library, we set up two categories for the comparison, i.e., "entire" and "isolatable" (Fig 5B and 5C). The latter set included only the sequences that could be isolated by the screen. First, we classified the "entire" set (5,946 sites) of the PIC assembly sites (http://downloads.yeastgenome.org/published_datasets/Rhee_2012_PMID_22258509/track_files/) into the same nine groups (class I–IX) and found that the population sizes of classes II (TATAWAD, 4.2%), IV (TATAWKW, 4.8%), and V (TTAAAW, 3.3%) were significantly smaller than those of classes I (TATAWAWR, 15%), VI (Wx6, 33%), and VIII (Wx5*, 13%) (Fig 5B and 5C). A similar tendency was observed for the "isolatable" set (3,448 sites) when it was classified into the same nine groups (Fig 5B and 5C). In addition, a significant number of sites were classified into class IX (others, 22%[entire] or 14%[isolatable]), contrary to the screen, in which the number of active clones classified in this class was almost zero, indicating that class IX sequences are disfavored specifically by the *CYC1* promoter. By contrast, the clones/sequences in classes II, IV, and V appeared to be significantly active in the *CYC1* promoter (Fig 5A) and thereby occupied the major fraction of the isolated clones in the screen (Fig 5B and 5C). Furthermore, the population size of the TATATCWD sequences (class

A

frame 1-6 & 9 (4781 clones)

50%   100   150

not classified   classified

I
II
III
IV
V
VI
VII
VIII
IX

0 50 100 150 200 0 1   0 50 100 150 200 0 1

B

class	seq	CYC1				genome-wide			
		clone #	normalized	%	seq #	promoter #	%	seq #	
I	TATAWAWR	614	20.9	12	28	875	15	8	entire
						875	25	8	isolatable
II	TATAWAD	1193	39.8	23	57	247	4.2	16	entire
						247	7.2	16	isolatable
III	GAAAA	29	1.8	1.1	18	281	4.7	36	entire
						200	5.8	22	isolatable
IV	TATAWKW	715	29.7	17	71	283	4.8	28	entire
						283	8.2	28	isolatable
V	TTAAAW	945	26.1	15	74	199	3.3	21	entire
						119	3.5	11	isolatable
VI	Wx6	865	40.7	23	138	1953	33	283	entire
						915	27	92	isolatable
VII	TATATCWD	115	6.6	3.8	22	57	1.0	6	entire
						57	1.7	6	isolatable
VIII	Wx5*	217	6.0	3.4	113	752	13	187	entire
						275	8.0	50	isolatable
IX	others	88	2.9	1.7	80	1299	22	518	entire
						477	14	118	isolatable
	total	4781	174.5	100	601	5946	100	1103	entire
						3448	100	351	isolatable

C

- TATAWAWR
- TATAWAD
- GAAAA
- TATAWKW
- TTAAAW
- Wx6
- TATATCWD
- Wx5*
- others
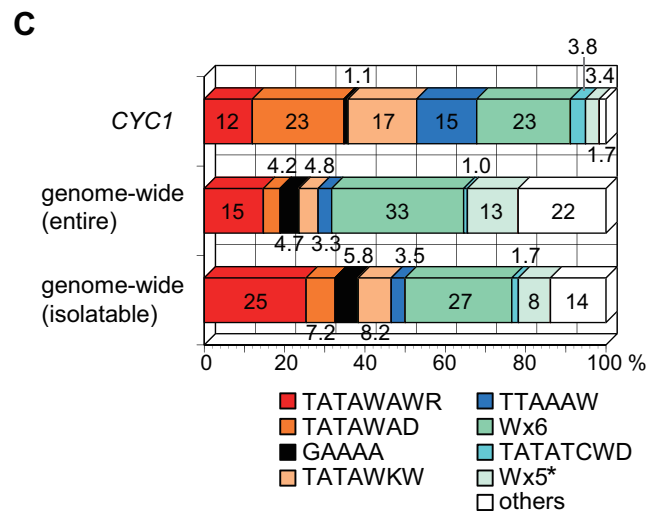
Fig 5. Classification of active clones and comparison with the PIC assembly sites on the genome. A. Graphical summary of all active clones isolated by the screen for frames 1–6 and 9 (S4 Table). In the left panel, the clones carrying the same sequences (as indicated at the left) were combined, regardless of which frame they were derived from, and aligned in order of the total clone number. The color code for each clone is the same as that described in Fig 4A. To compare the isolation frequency among the sequences, normalization is required since the chance to be isolated as an active clone is proportional to the numbers of each clone existing in each library. For this purpose, the expected number of a given clone carrying a specific sequence generated in each library was calculated and summed up across all frames (i.e., frames 1–6 and 9). The actually isolated clone number carrying each sequence was divided by each expected clone number to determine the normalized isolation frequency (S7 Table). This normalized value for each sequence is shown at the right side of the left panel with a black bar. The value of 1 (as indicated with a vertical thin line) represents 100% recovery; that is, the clones carrying a given sequence that exist in the library were all recovered as active clones by the screen. Note that some clones/sequences exceed the value of 1, indicating that there may be a certain bias to generate some specific clones/sequences more frequently than others during the randomization process. The numerals along the horizontal axis indicate the clone number (left, 0–200) or the normalized isolation frequency (right, 0–1). In the right panel, the clones presented in the left panel were classified into nine groups as summarized in **B** and aligned in order of the total clone number in each class. This panel was enlarged in S2–S4 Figs for readability of each sequence. In both panels, the boundary between the clones isolated more than twice and those isolated only once is marked with a closed triangle, highlighting the reproducibility of each clone in the screen. **B**. Summarized table for classification of the clones/sequences (9 bp) isolated by the screen and their comparison with those of the PIC assembly sites (8 bp) on the genome [27]. The number of the isolated clones (indicated as "clone #"), the sum (integral value) of the normalized isolation frequency (normalized value in S7 Table) for each sequence in each class (indicated as "normalized"), the percentage of each summed "normalized" value in total (indicated as "%"), and the number of the distinct sequences belonging to each class (indicated as "seq #") are summarized in the left-middle four columns denoted as "*CYC1*" at the top. The "entire" set of sequences (8 bp) identified previously as the PIC assembly sites [27] were classified into two groups that could be isolated by our screen ("isolatable") or not ("not isolatable"). The "entire" or "isolatable" sequences were further classified into the same nine groups as those isolated by the screen. The number of the sequences (indicated as "seq #"), the number of the genes containing such sequences in the promoter region (indicated as "promoter #"), and the percentage of the latter belonging to each class in total (indicated as "%") are summarized in the right three columns denoted as "genome-wide" at the top. Abbreviations are as follows: W = A or T; R = A or G; D = A or G or T; K = G or T; Wx6 = WWWWWW; Wx5* = WWWWW except the clones/sequences included in class VII. **C**. The percentages of the "normalized" value (*CYC1*) and those of the "promoter #" (genome-wide) were presented graphically.

VII) in the Wx5 sequences (class VII+VIII) was more than 50% (normalized value) for the screen but only 7% (= 57/809, entire) or 17% (= 57/332, isolatable) for the genome-wide analysis (Fig 5B and 5C), indicating that the TATATCWD sequences are favored specifically by the *CYC1* promoter. Altogether, these results imply that the preferred sequence(s) vary among the genes, presumably depending on the promoter structure. Notably, the population size of class III (GAAAA) was smaller in both the screen (1.1%) and the genome-wide analysis (4.7%[entire] or 5.8%[isolatable]) than the size expected based on the previous notion that the GA element is found frequently in the core promoter region (37.3%) [57]. Thus, we speculate that this element must have a different function other than serving as the PIC assembly site.

There are two types of core promoters in budding yeast that are Taf1p-enriched or Taf1p-depleted [27]. The population size of class I (consensus TATAWAWR) was significantly different for these two types of core promoters; it was much smaller in the former (9.6%[entire] or 17%[isolatable]) than in the latter (37%[entire] or 53%[isolatable]) (S5 Fig) [27]. The corresponding value (12%) obtained in the screen was also small, suggesting that the *CYC1* promoter is closer to the Taf1p-enriched type than to the Taf1p-depleted type. The significance of the difference between the screen and genome-wide "isolatable" data for each class was evaluated by using the Z test. The calculated p-values are summarized in S5A Fig. This is consistent with the fact that the *CYC1* promoter showed Taf1p-dependency when it was integrated at the *VTC1* locus (S1 Fig).

The normalized isolation frequencies (modified slightly from those in Fig 5A as described in the legend for S6 Fig) for several class I or II sequences were compared individually with the number of PIC assembly sites containing such sequences in the Taf1p-enriched and/or Taf1p-depleted core promoters (S6 Fig, note that all of these sequences are "isolatable"). As expected, the class II sequences were favored by the *CYC1* and Taf1p-enriched core promoters compared to the Taf1p-depleted core promoters. Furthermore, the normalized isolation frequency for each sequence in the screen did not correlate with its utilization frequency as the PIC assembly site on the genome. These observations suggest that the preferred sequence(s) as the core promoter element may vary among the genes, presumably depending on the promoter structure even within the Taf1p-enriched type.

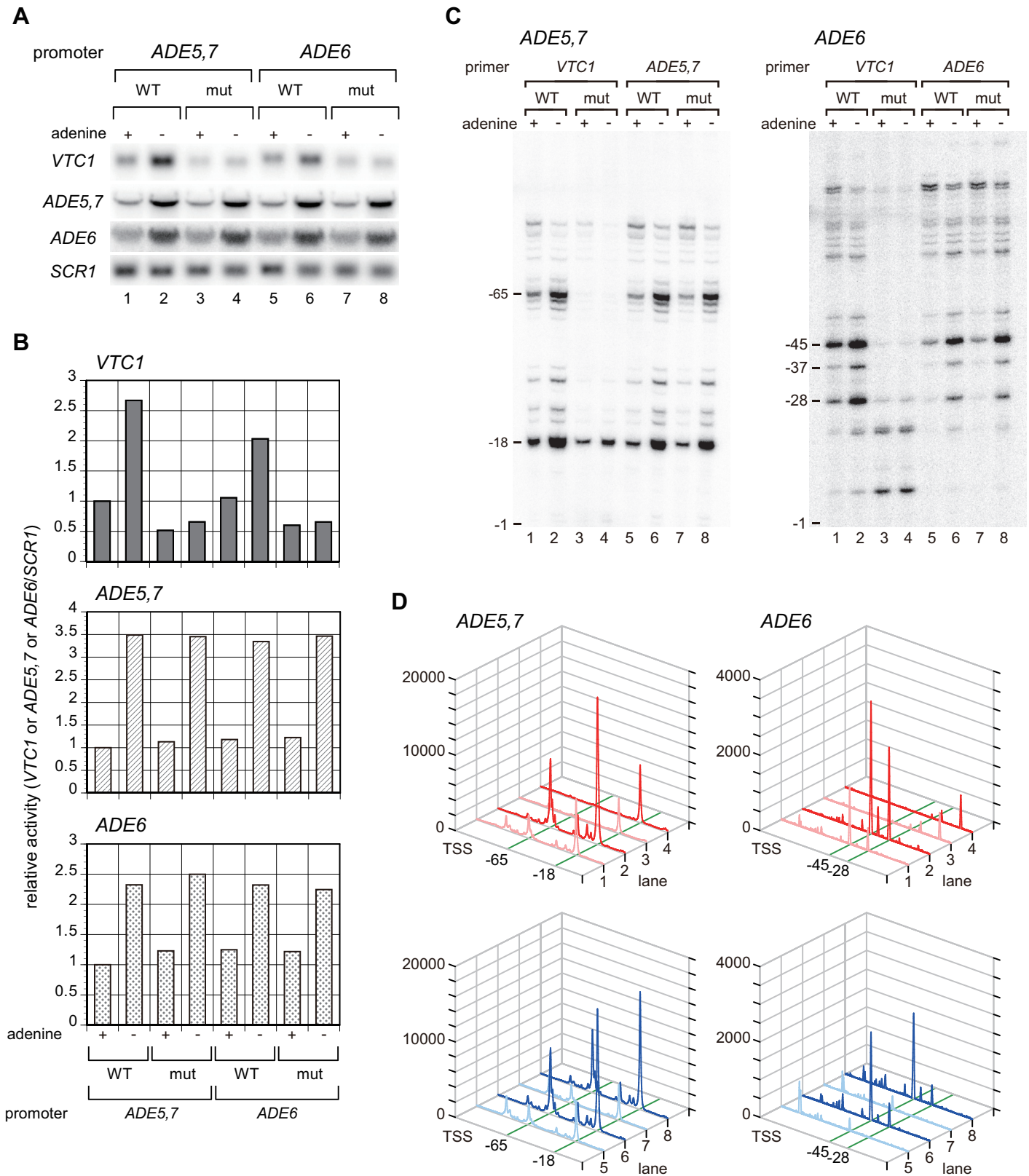## TATTTAAA is a functional core promoter element of endogenous genes

Class V represents a unique sequence motif that is significantly active, even though it does not contain any "TATA" sequences. The most frequently isolated and the most active sequence in this class was "TATTTAAAA" (9 bp) (S3B Fig). To examine whether this sequence serves as a TATA-like element even in the endogenous genes, we searched for "TATTTAAA" or "ATT-TAAAA" among the PIC assembly sites (8 bp) on the genome [27] and found that only the former functions at the 35 distinct promoters (data not shown). Among them, two promoters (i.e., *ADE5,7* and *ADE6*) were selected to determine whether the sequence "TATTTAAA" was functional (S7 Fig). Northern blot analyses showed that the reporter gene (*VTC1*) fused to these promoters could be activated at similar levels to the endogenous genes under the adenine-starved condition (Fig 6A and 6B) [58]. The "CGCCCGGG" substitution of this element (S7 Fig) significantly impaired the basal and activated expression of these two promoters (Fig 6A and 6B). Primer extension analyses also showed that this element plays a pivotal role in supporting the normal levels of basal and activated expression from the accurate TSSs in these two promoters (Fig 6C and 6D). Collectively, these results indicate that at least some of the sequences isolated by the screen are utilized as the core promoter element(s) for the endogenous genes.

Discussion

A novel reporter assay system using toluidine blue (TB) dye

In the present study, we developed a novel reporter assay system to monitor gene expression in yeast cells. This method is semi-quantitative but simple, convenient, and inexpensive, since it uses only TB dye for visualization. This enables rough estimation of promoter activities without equipment, simply by judging the color with the naked eye (Fig 1B and 1E). Several reporter assay systems using β-galactosidase (β-gal), luciferase (Luc), or GFP (or its derivatives) are well established in yeast [59–64]. These methods require expensive measuring equipment such as fluorometers, luminometers or flow cytometers. The β-gal activity can be measured by spectro-photometer when colorimetric substrates (e.g., ONPG or X-gal) are used in the assay. However, the procedures for cell disruption/permeabilization and measurement of enzymatic activities within the linear range are time-consuming and laborious. Reagents for reducing such labors are commercially available; however, they are expensive compared to those needed for the TB assay. Another distinctive feature of the present method is that the target molecule for the measurement is not the protein itself (i.e., unlike β-gal, Luc, or GFP), but rather a biochemical product (i.e., polyP) accumulated endogenously in the cell by the action of the reporter gene (i.e., *VTC1*). Thus, this method has advantages over other methods in that it does not require any exogenous substrates or intense labor efforts to adjust the assay within the linear range (unlike β-gal or Luc), or expensive equipment (unlike Luc or GFP) for the measurement. However, this method is not adequate for following the dynamic behavior of target promoters because polyP is stable after the shutoff of *VTC1* expression [33], although it is useful to measure promoter activity in a steady state as shown in this study.

Recently, massively parallel reporter assay (MPRA) systems have been developed owing to the advancement of next-generation sequencing technologies [65–70]. In these assay systems, promoter activities can be measured directly by sequencing cDNAs tagged individually with a specific barcode for identification. Such tagging system enables each promoter activity to be measured even in the mixed population, thereby greatly increasing the efficiency of the method to identify active promoter elements. Despite these advances, efforts to reveal the structure and function of the core promoter elements remain limited [65], as most studies have been focused on the functional properties of the UAS/enhancers [66–70]. Recently, a genome-wide approach

Fig 6. The TATTTAAA sequence functions as a core promoter element in the *ADE5,7* and *ADE6* promoters. A. Northern blot analyses to examine the expression of the reporter (*VTC1*) or control genes (*ADE5,7*, *ADE6*, and *SCR1*) in strains carrying the reporter driven by the wild-type (WT: lane 1, 2) or

mutated (mut: lane 3, 4) *ADE5,7* promoter, or similarly by the wt (lane 5, 6) or mut (lane 7, 8) *ADE6* promoter, as indicated at the top. Total RNA (20 μg) was isolated from these strains cultured at 30°C in synthetic complete (SC) media with (+) or without (-) adenine, blotted onto the membrane, and hybridized with the gene-specific probes indicated at the left. **B**. The raw data shown in **A** were quantified and presented graphically. Values for each transcript (*VTC1*, *ADE5,7*, *ADE6*) were normalized by *SCR1* (pol III transcript) and presented as relative to those of the strain carrying the reporter driven by the wt *ADE5,7* promoter and cultured in the presence of adenine (i.e., the values at the most left side are set as 1). **C**. Primer extension analyses to examine the transcriptional start site(s) (TSSs) of the reporter (*VTC1*) and endogenous genes (*ADE5,7*, or *ADE6*) in the same strains as described in **A**. Total RNA (20 μg) was subjected to the primer extension analysis using the specific primers for *VTC1*, *ADE5,7*, or *ADE6* as indicated at the top. The positions of major TSSs relative to the A (+1) of the start codon ATG are indicated at the left of each panel. Note that the primers for *VTC1* or endogenous genes (*ADE5,7* or *ADE6*) were designed to generate the same size of cDNA fragments when mRNAs were transcribed from the same TSSs. The TSS profiles along with the promoter sequences (wt, mut) of *ADE5,7* and *ADE6* are schematically shown in S7 Fig **D**. Each lane of the electrophenograms shown in **C** was scanned, quantified by densitometry (ImageQuant TL software version 8.1, GE Healthcare), and presented graphically. The results for the reporter (*VTC1*, lanes 1–4) are shown in the upper two panels, while those for the endogenous genes (*ADE5,7* or *ADE6*, lanes 5–8) are shown in the lower two panels.

using a STARR-seq (self-transcribing active regulatory region sequencing) method [70] demonstrated that there are three types of enhancers in *Drosophila*, namely, those that can specifically activate a "housekeeping-type" core promoter containing the TCT motif (hkCP enhancer), those that activate a "developmental-type" core promoter containing TATA, Inr, MTE, and DPE (dCP enhancer), or those that activate both (shared enhancer) [12]. This has enabled the identification of core promoter element(s) that can be activated by a specific type of UAS/enhancer in a more systematic manner. Although the MRPA and/or STARR-seq methods would be ideal for such a purpose, they are unaffordable for most researchers. In this regard, the TB method still has competitive merits, especially in budding yeast, owing to its cost-effective and high-throughput features that may be enough for the analysis of organisms carrying a compact genome equipped with fewer regulatory *cis*-elements.

TATAWAD is a functional TATA element motif for the *CYC1* promoter

A previous study demonstrated that the consensus TATA sequence is TATAWAWR (8 bp), in which the 8th position is confined to "R" (R = A/G) [19]. By contrast, we showed that any base could function almost equally at the 8th position (Figs 3C and 4A), indicating that the TATA element is 7 bp at least in the *CYC1* promoter. This assumption is consistent with a previous study showing that TATATAAN (N = A/G/C/T) are all functional at the TATAα site in the *CYC1* promoter *in vitro* and *in vivo* [54]. We also showed that the base at the 7th position is critical for function but is not confined to "W" (W = A/T), as expected from the consensus TATA sequence. This position allowed "G", but not "C", to be fully active (Fig 4B), implying that the 7th position should be assigned to "D" (D = A/G/T) at least in this promoter. Therefore, we conclude that the TATA sequence preferred by the *CYC1* promoter should be represented as "TATAWAD" (7 bp) rather than the consensus "TATAWAWR" (8 bp). To the best of our knowledge, this is the first demonstration of the "functional consensus TATA sequence" determined experimentally for one particular promoter, i.e., by using a functional screen to select for a large number of transcriptionally active clones from a randomized library. Of note, a similar randomized screen was conducted previously for the T_R element of the *HIS3* promoter [35]. However, functional consensus TATA sequence(s) could not be determined for this promoter, probably because a broader region containing the TATA element (16 bp) was replaced at once with the randomized sequence(s), contrary to our strategy in which the TATA element was divided into multiple small segments (frames 1–8) for randomization (Fig 2C).

Promoter specificity of the TATA or TATA-like sequences that were isolated by the screen as the TATAα element of the *CYC1* promoter

The populations containing the sequences that matched either TATAWAWR (class I) or TATAWAD (class II) showed similar sizes in the screen for the active *CYC1*-TATAα element

(class I+II = 35%) and the "isolatable" set of the PIC assembly sites (class I+II = 33%) ([Fig 5C]). However, the ratio of class II to I of the former (class II/I = 1.9) was significantly higher than that of the latter (class II/I = 0.28), indicating that TATAWAD is a less frequently utilized "functional TATA element motif" than TATAWAWR on the genome. Namely, the requirements for the 7th and/or 8th position(s) of the TATA element could be alleviated only for a limited number of genes. In this regard, it should be noted that the well-characterized TATA element of *HIS3* (T_R: TATAAAGT) [29] matched TATAWAD but not TATAWAWR. Furthermore, the T_R element remains active when "G" at the 7th position is mutated to "A" or "T", but not when it is changed to "C" [29, 71]. These observations suggest that both the *HIS3*-T_R and *CYC1*-TATAα elements might belong to the same TATA element subfamily for which TATAWAD could represent a "functional consensus TATA element motif". Intriguingly, a detailed structural study of the TBP-TATA complex demonstrated that "C" or "G" at the 7th position could induce a Hoogsteen-type base pair instead of a conventional Watson-Crick-type base pair upon TBP binding [72]. Thus, TATAWAD might be a novel type of core promoter element carrying a different function from that of the consensus TATAWAWR.

This study also showed that the function of the *CYC1*-TATAα element became Taf1p-dependent when it was integrated at the *VTC1* locus ([S1 Fig]). Contrary to these findings, the function of the *HIS3*-T_R element is known to be Taf1p-independent [73], arguing against the possibility that TATAWAD is a functional TATA element motif for Taf1p-dependent promoters. However, the region (TA<u>TATAAAGT</u>) containing the *HIS3*-T_R element (underlined sequence) of the *HIS3* promoter matches both TATAWAD (TA<u>TATAAAGT</u>) and TATAWAWR (<u>TATATAAA</u>GT). The Taf1p-dependency of the *HIS3*-T_R element was examined within the context of the original promoter [73], while mutational studies of the *HIS3*-T_R element were conducted for the shorter TATA segment containing TATAAAGT (TATAWAD-matched sequence) but not TATATAAA (TATAWAWR-matched sequence) [29, 71]. Therefore, it remains possible that the function of the *HIS3*-T_R element itself (TATAAAGT) is Taf1p-dependent, and, if so, TATAWAD might be a functional TATA element motif for the Taf1p-dependent promoters. This assumption is consistent with the observation that the utilization ratio of TATAWAD to TATAWAWR at the PIC assembly site is significantly higher in the Taf1p-enriched promoters than in the Taf1p-depleted promoters ([S5 Fig]). Further analyses are needed to test this possibility more directly.

Another remarkable observation of the present study is that TTAAAW-containing sequences (class V) could function as active core promoter elements in the *CYC1* promoter ([Fig 5A], [S2]–[S4] Figs). This motif also appeared to be gene-specific by comparison with the genome-wide data ([Fig 5C]). We paid special attention to the sequence "TAT<u>T</u>TAAAA" (1 bp mismatched from the wild-type "TATATAAAA") because it was the most frequently isolated, not only in class V but also across the entire classes (class I–IX) except the wild-type sequence itself ([Fig 5A], [S2]–[S4] Figs). Furthermore, this sequence was one of the most active sequences among the 23 specifically mutated *CYC1*-TATAα elements [54]. Notably, a similar but non-class V variant, "TAT<u>T</u>TAGT" (2 bp mismatch) of the *HIS3*-T_R element, was shown to be as active as the wild-type [29], and another variant, "TAT<u>TA</u>AT" (termed R2; 3 bp mismatch), was isolated as one of the most active *HIS3*-T_R elements by a randomized screen [35]. These observations suggest that "TATTTARW" (R = A/G, W = A/T) provides a novel core promoter element motif that is shared by the *HIS3* (TATTTAGT, TATTTAAT) and *CYC1* (TATTTAAA*A*) promoters. However, as the sequences containing this motif were isolated less frequently by our screen (i.e., <u>TATTTAAT</u>A/class VI [10 clones/normalized value 0.73], <u>TTATTTAAA</u>/class VI [8 clones/normalized value 0.58], A<u>TATTTAAA</u>/class VI [4 clones/normalized value 0.29] and <u>TATTTAGA</u>A/class VI [2 clones/normalized value 0.15] other than <u>TATTTAAA</u>A/class V [118 clones/normalized value 1.07]), and the isolated clones appeared to be rather inactive

except the "TATTTAAAA" sequence itself (S4 and S7 Tables), we assume that "TTAAAW" (class V) is a more appropriate core promoter element motif than others, such as "TATTTA RW", at least for the *CYC1* promoter. Notably, three class V sequences, namely, "CACCGCTA TTTAAATCCC" (termed R15), "CTACTACTATTAAAACCCA" (termed R33), and "TTAAAAGCGTCCCATTTC" (termed R51), were isolated as *HIS3*-T_R derivatives by a randomized screen, although their activities were very weak [35], supporting the aforementioned view that class V may represent a gene-specific core promoter element motif.

We found that the "TATTTAAA" (8 bp) sequence was utilized as the PIC assembly site in 35 endogenous genes [27] and demonstrated that this sequence could function as a *bona fide* core promoter element at least in the *ADE6* and *ADE5,7* promoters (Fig 6). Of note, when the frame containing this 8 bp sequence was extended downstream by 1 bp, the "TATTTAAAC" and "TATTTAAAT" sequences resided in the *ADE5,7* and *ADE6* promoters, respectively (S7 Fig). The PIC was also assembled on the same "TATTTAAA" (8 bp) sequence in the *ADE4* promoter [27], and a similar 1 bp extension revealed the "TATTTAAAG" sequence in this promoter. These observations suggest that these three *ADE* genes (i.e., *ADE4*, *ADE5,7*, and *ADE6*), which are involved in the early steps of the purine *de novo* synthesis pathway [74], might be regulated coordinately in a "TATTTAAA" motif-dependent manner rather than in a "TTAAAW" motif-dependent manner. Our screen could isolate "TATTTAAAA" but not "TATTTAAA[C/G/T]", as the latter was not included in the library, implying that the 9th position of the "TATTTAAAA" sequence may not be important for *CYC1* promoter activity, similar to the *ADE* promoters. However, "TATTTAAA"-containing sequence(s) (i.e., TTATTTAAA/class VI [8 clones/ normalized value 0.58] and ATATTTAAA/class VI [4 clones/normalized value 0.29]) were isolated less frequently and showed weaker activities than "TATTTAAAA" as described above (S4 and S7 Tables), indicating that "TATTTAAAA" (9 bp) is a more appropriate functional core promoter element motif for the *CYC1* promoter than "TATTTAAA" (8 bp). Furthermore, "TATTTAAA" (8 bp) may be another distinct gene-specific core promoter element motif, since it was active in the *ADE* promoters but not in the *CYC1* and *HIS3* (e.g., aforementioned "R15/CACCGCTATTTAAATCCC") promoters. Therefore, it is likely that not only the sequence but also the length of the core promoter element(s) play a critical role in their gene-specific function, as evidenced by the relationship between TATTTAAA (8 bp) and TATTTAAAA (9 bp), or that between TATAWAD (7 bp) and TATAWAWR (8 bp).

Conclusion

In the present study, we identified a large number of sequences that could function as the *CYC1*-TATAα element by using a novel reporter assay system. This assay system is convenient, sensitive, and reliable, as shown by its ability to isolate all active 18 variants among the 23 specifically mutated *CYC1*-TATAα elements [54] (S4–S7 Tables). Our results showed that TATA-WAD is a functional consensus TATA element motif for the *CYC1* promoter and suggested that several other sequence motifs function in a gene-specific manner. The isolated sequences were classified tentatively into nine groups; however, they showed considerable variation in transcriptional activity even within a single class (Fig 5A, S2–S4 Figs). To further delineate the consensus sequence(s) for the functional core promoter element motif(s), more detailed statistical and/or mutational analyses need to be performed in a large number of sequences in each class. In addition, a similar screen may need to be conducted more extensively and systematically in a wide variety of promoters. Nevertheless, we believe that these experimental approaches could provide valuable information for understanding not only the structures of the core promoter element(s), but also the mechanisms underlying the functional compatibilities between the UAS and the core promoter.

Supporting Information

S1 Fig. Transcription from the *CYC1* promoter at the *VTC1* locus is Taf1p-dependent. A.
Northern blot analyses to examine the expression of the reporter (*VTC1*) or control gene (*SCR1*)
in *TAF1* or *taf1-N568Δ* strains carrying the reporter driven by wild-type (β, α, wt, wt at site #1, 2,
3, 4, respectively) or variously mutated *CYC1* promoters as indicated at the top. The details of
the promoter sequences tested here were shown in Fig 2A. Total RNA (20 μg) was isolated from
these strains 2 hours after temperature shift to 37°C or continuously cultured at 25°C over the
same time period in rich media (YPD), and blotted onto the membrane and hybridized with the
gene-specific probes indicated at the right. **B**. The raw data shown in **A** were quantified and pre-
sented graphically. Values for each *VTC1* transcript were normalized by *SCR1* (pol III transcript)
and presented as relative to those derived from the *TAF1* strain carrying the reporter driven by
the wild-type *CYC1* promoter (β, α, wt, wt) and cultured at 25°C or 37°C (i.e., the values at 1st or
11th lane are set as 1 for 25°C or 37°C, respectively). To emphasize the contribution of the TATA
element(s) at site #1 or #2 to transcriptional activity, the background levels represented by the
(m, m, m, m) promoter (i.e., 5th, 10th, 15th and 20th lanes in each set) are colored in gray.
(EPS)

**S2 Fig. Enlarged Figures of the right panel in Fig 5A for readability of each sequence of
the isolated clones (class I, II and III).** The scales at the horizontal axis are 10 (clone number)
or 0.1 (normalized value) per division, respectively. **A**. Class I (TATAWAWR). **B**. Class II
(TATAWAD). **C**. Class III (GAAAA). The GAAAA-containing sequences were searched for
the isolated clones by using two distinct frames; one is Nx9 (9 bp; TATATAAAA, upper panel)
and the other is G/Nx9 (10 bp; GTATATAAAA, lower panel). Note that only the former is
shown in Fig 5A. The boundary between the clones isolated more than twice and those isolated
only once is marked with a closed triangle as described in Fig 5A.
(EPS)

**S3 Fig. Enlarged Figures of the right panel in Fig 5A for readability of each sequence of the
isolated clones (class IV and V).** The scales and closed triangles are as described in S2 Fig. **A**.
Class IV (TATAWKW). **B**. Class V (TTAAAW).
(EPS)

**S4 Fig. Enlarged Figures of the right panel in Fig 5A for readability of each sequence of the
isolated clones (class VI, VII, VIII and IX).** The scales are as described in S2 Fig. Only repro-
ducible clones isolated more than twice are shown in this figure. **A**. Class VI (Wx6). **B**. Class
VII (TATATCWD). **C**. Class VIII (Wx5*). **D**. Class IX (others).
(EPS)

**S5 Fig. Classification and comparison of the clones/sequences isolated by the screen with
the sequences utilized as the PIC assembly sites in the Taf1p-enriched or Taf1p-depleted
promoters. A**. Summarized table for classification of the clones/sequences isolated by the
screen and comparison of their sequences with those of the PIC assembly sites on the genome
as described in Fig 5B, except that the latter (indicated as "all") was further classified into two
groups (indicated as "Taf1p-enriched" and "Taf1p-depleted") [27]. The significance of the dif-
ferences between the screen and genome-wide "isolatable" data for each class was evaluated by
using the Z test. The calculated p-values are shown in the 10th, 14th and 18th columns, respec-
tively. Abbreviations are the same as those described in Fig 5B except that pro # = promoter #
and N/A = not applicable. **B**. The percentages of the "normalized" value (*CYC1*) and those of
the "promoter #" (genome-wide) were presented graphically as described in Fig 5C.
(EPS)

S6 Fig. Comparison of the normalized* isolation frequencies of the individual sequences belonging to class I (TATAWAWR) or II (TATAWAD) that were isolated by the screen with the utilization frequencies as the PIC assembly sites on the genome. The results were summarized in a table (**A**) or four graphs (**B**). In both **A** and **B**, the sequences are aligned according to the promoter number in the "all" category of the PIC assembly sites. Note that the normalized* value in this figure is not the same as the normalized value in [Fig 5](#) and [S2](#)–[S5](#) Figs. Each sequence (8 bp: TATAWAWT #1–8 & TATAWAD #1–16) in the list as summarized in **A** includes 1–6 distinct isolated sequences (9 bp). For instance, TATATATA (TATAWAWR #1/ 107 clones) includes six distinct sequences (TATATATA<u>A</u>/ 61 clones, TATATATA<u>T</u>/ 3 clones, TATATATA<u>G</u>/ 8 clones, TATATATA<u>C</u>/ 3 clones, <u>A</u>TATATATA/ 15 clones, <u>T</u>TATATATA/ 17 clones), whereas TATAAATG (TATAWAWR #8/ 20 clones) contains only one sequence (TATAAATG<u>A</u>/ 20 clones) ([S7 Table](#)). The normalized* value (e.g., TATATATA: 0.74, TATAAATG: 0.58) was calculated by dividing the total number of isolated clones (9 bp) containing a given sequence (8 bp) (e.g., TATATATA: 61+3+8+3+15+17 = 107, TATAAATG: 20) by the expected number of the clones containing the same 8 bp sequence that exist in the libraries (e.g., TATATATA: 75.625+13.75+13.75+13.75+13.75+13.75 = 144.375, TATAAATG: 34.375, as listed in [S7 Table](#)), not by summing up the normalized value determined individually for each sequence (9 bp) as described in the legend of [Fig 5](#) (e.g., TATATATA: 61/75.625+3/ 13.75+8/13.75+3/13.75+15/13.75+17/13.75 = 4.15, TATAAATG: 20/34.375 = 0.58). The normalized value (e.g., TATATATA: 4.15, TATAAATG: 0.58) deviates significantly from the normalized* value that does not exceed 1.0 theoretically (e.g., TATATATA: 0.74, TATAAATG: 0.58), especially when target sequence(s) are generated from multiple libraries. We suppose that the normalized value is more appropriate for the purpose of [Fig 5](#) and [S2](#)–[S5](#) Figs than the normalized* value, since the integral of the values that have been normalized individually for each sequence is more informative for the comparison between several classes, each of which includes a huge number of sequences. On the contrary, the normalized* value is more appropriate for the purpose of this figure than the normalized value, since the latter is affected significantly by the number of isolated sequences (9 bp) included in each target sequence (8 bp). Such artificial effects derived from the library construction strategy need to be minimized for the comparison attempted here. Thus, we exploit the normalized* value in this figure, in which all isolated sequences (9 bp) containing the same 8 bp sequence were treated evenly and subjected as a set to the calculation for normalization.
(EPS)

S7 Fig. Sequences of the *ADE5,7* (A) and *ADE6* (B) promoters. The TATTTAAA sequence and its mutated derivative (mut: CGCCCGGG) are marked with a thick black line. The major transcriptional start sites (TSSs) determined by primer extension analyses as described in [Fig 6C](#) are indicated below the sequence with a black arrow, while the TSSs determined in a previous study [[75](#)] are indicated above the sequence with a gray arrow as a reference. The initiation codon of each gene is marked with an open square along with the number of +1 (A of ATG as +1).
(EPS)

S1 Table. *Saccharomyces cerevisiae* strains used in this study.
(DOC)

S2 Table. Oligonucleotides used in this study.
(DOC)

S3 Table. PCR primers used for the construction of yeast strains in this study.
(XLSX)

S4 Table. Promoter activities of the clones isolated from the libraries of frame 1–6 & 9.
(XLSX)

S5 Table. Promoter activities of the clones isolated from the library of frame 7.
(XLSX)

S6 Table. Promoter activities of the clones isolated from the library of frame 8.
(XLSX)

S7 Table. Normalized values of the clones isolated from the libraries of frame 1–6 & 9.
(XLS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: KW TK. Performed the experiments: KW KK. Analyzed the data: KW MY TK. Contributed reagents/materials/analysis tools: MY. Wrote the paper: KW TK.

References

1. Levine M, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. Cell. 2014; 157(1):13–25. doi: 10.1016/j.cell.2014.02.009 PMID: 24679523; PubMed Central PMCID: PMC4059561.

2. Hahn S, Young ET. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics. 2011; 189(3):705–36. doi: 10.1534/genetics.111.127019 PMID: 22084422; PubMed Central PMCID: PMC3213380.

3. Kadonaga JT. Perspectives on the RNA polymerase II core promoter. Wiley interdisciplinary reviews Developmental biology. 2012; 1(1):40–51. doi: 10.1002/wdev.21 PMID: 23801666; PubMed Central PMCID: PMC3695423.

4. Muller F, Tora L. Chromatin and DNA sequences in defining promoters for transcription initiation. Biochim Biophys Acta. 2014; 1839(3):118–28. doi: 10.1016/j.bbagrm.2013.11.003 PMID: 24275614.

5. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet. 2012; 13(4):233–45. doi: 10.1038/nrg3163 PMID: 22392219.

6. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 2006; 41(3):105–78. PMID: 16858867.

7. Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev. 2002; 16(20):2583–92. PMID: 12381658.

8. Muller F, Demeny MA, Tora L. New problems in RNA polymerase II transcription initiation: Matching the diversity of core promoters with a variety of promoter recognition factors. J Biol Chem. 2007. PMID: 17395580.

9. Kobayashi A, Watanabe Y, Akasaka K, Kokubo T. Real-time monitoring of functional interactions between upstream and core promoter sequences in living cells of sea urchin embryos. Nucleic Acids Res. 2007; 35(14):4882–94. PMID: 17626044.

10. Ohler U, Wassarman DA. Promoting developmental transcription. Development. 2010; 137(1):15–26. Epub 2009/12/22. 137/1/15 [pii] doi: 10.1242/dev.035493 PMID: 20023156; PubMed Central PMCID: PMC2796937.

11. Li XY, Bhaumik SR, Zhu X, Li L, Shen WC, Dixit BL, et al. Selective recruitment of TAFs by yeast upstream activating sequences. Implications for eukaryotic promoter structure. Curr Biol. 2002; 12 (14):1240–4. PMID: 12176335.

12. Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature. 2015; 518 (7540):556–9. doi: 10.1038/nature13994 PMID: 25517091.

13. Juven-Gershon T, Hsu JY, Kadonaga JT. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes Dev. 2008; 22(20):2823–30. Epub 2008/10/17. 22/20/2823 [pii] doi: 10.1101/gad.1698108 PMID: 18923080; PubMed Central PMCID: PMC2569877.

14. Zehavi Y, Kuznetsov O, Ovadia-Shochat A, Juven-Gershon T. Core promoter functions in the regulation of gene expression of Drosophila dorsal target genes. J Biol Chem. 2014; 289(17):11993–2004. doi: 10.1074/jbc.M114.550251 PMID: 24634215; PubMed Central PMCID: PMC4002106.

15. Kadonaga JT. The DPE, a core promoter element for transcription by RNA polymerase II. Experimental & molecular medicine. 2002; 34(4):259–64. doi: 10.1038/emm.2002.36 PMID: 12515390.

16. Kedmi A, Zehavi Y, Glick Y, Orenstein Y, Ideses D, Wachtel C, et al. Drosophila TRF2 is a preferential core promoter regulator. Genes Dev. 2014; 28(19):2163–74. doi: 10.1101/gad.245670.114 PMID: 25223897; PubMed Central PMCID: PMC4180977.

17. Wang YL, Duttke SH, Chen K, Johnston J, Kassavetis GA, Zeitlinger J, et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev. 2014; 28(14):1550–5. doi: 10.1101/gad.245662.114 PMID: 24958592; PubMed Central PMCID: PMC4102762.

18. Parry TJ, Theisen JW, Hsu JY, Wang YL, Corcoran DL, Eustice M, et al. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev. 2010; 24(18):2013–8. doi: 10.1101/gad.1951110 PMID: 20801935; PubMed Central PMCID: PMC2939363.

19. Basehoar AD, Zanton SJ, Pugh BF. Identification and distinct regulation of yeast TATA box-containing genes. Cell. 2004; 116(5):699–709. PMID: 15006352.

20. Sugihara F, Kasahara K, Kokubo T. Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of Saccharomyces cerevisiae. Nucleic Acids Res. 2011; 39(1):59–75. doi: 10.1093/nar/gkq741 PMID: 20805245; PubMed Central PMCID: PMC3017598.

21. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene. 2007; 389(1):52–65. PMID: 17123746.

22. Chalkley GE, Verrijzer CP. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250- TAF (II)150 complex recognizes the initiator. EMBO J. 1999; 18(17):4835–45. PMID: 10469661

23. Struhl K. Molecular mechanisms of transcriptional regulation in yeast. Annu Rev Biochem. 1989; 58:1051–77. Epub 1989/01/01. doi: 10.1146/annurev.bi.58.070189.005155 PMID: 2673007.

24. Smale ST, Kadonaga JT. The RNA Polymerase II Core Promoter. Annu Rev Biochem. 2003. PMID: 12651739.

25. Cheng JX, Floer M, Ononaji P, Bryant G, Ptashne M. Responses of four yeast genes to changes in the transcriptional machinery are determined by their promoters. Curr Biol. 2002; 12(21):1828–32. PMID: 12419182.

26. Mencia M, Moqtaderi Z, Geisberg JV, Kuras L, Struhl K. Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. Mol Cell. 2002; 9(4):823–33. PMID: 11983173.

27. Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature. 2012; 483(7389):295–301. doi: 10.1038/nature10799 PMID: 22258509; PubMed Central PMCID: PMC3306527.

28. Huisinga KL, Pugh BF. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol Cell. 2004; 13(4):573–85. PMID: 14992726.

29. Harbury PA, Struhl K. Functional distinctions between yeast TATA elements. Mol Cell Biol. 1989; 9 (12):5298–304. Epub 1989/12/01. PMID: 2685558; PubMed Central PMCID: PMC363694.

30. Simon MC, Fisch TM, Benecke BJ, Nevins JR, Heintz N. Definition of multiple, functionally distinct TATA elements, one of which is a target in the hsp70 promoter for E1A regulation. Cell. 1988; 52 (5):723–9. PMID: 2830991.

31. Wefald FC, Devlin BH, Williams RS. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. Nature. 1990; 344(6263):260–2. PMID: 2156167.

32. Li WZ, Sherman F. Two types of TATA elements for the CYC1 gene of the yeast Saccharomyces cerevisiae. Mol Cell Biol. 1991; 11(2):666–76. PMID: 1846668; PubMed Central PMCID: PMC359717.

33. Ki S, Sugihara F, Kasahara K, Tochio H, Okada-Marubayashi A, Tomita S, et al. A novel magnetic resonance-based method to measure gene expression in living cells. Nucleic Acids Res. 2006; 34(6):e51. PMID: 16598072.

34. Ki S, Sugihara F, Kasahara K, Tochio H, Shirakawa M, Kokubo T. Magnetic resonance-based visualization of gene expression in mammalian cells using a bacterial polyphosphate kinase reporter gene. Biotechniques. 2007; 42(2):209–15. PMID: 17373486.

35. Singer VL, Wobbe CR, Struhl K. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev. 1990; 4(4):636–45. Epub 1990/04/01. PMID: 2163345.

36. Lundblack V. 13. Saccharomyces cerevisiae. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., editors. Current protocols in molecular Biology. 2. New York: John Wiley & Sons; 1998. p. 13.0.1.–..9.

37. Longtine MS, McKenzie A 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast. 1998; 14(10):953–61. PMID: 9717241.

38. Kitazono AA, Tobe BT, Kalton H, Diamant N, Kron SJ. Marker-fusion PCR for one-step mutagenesis of essential genes in yeast. Yeast. 2002; 19(2):141–9. PMID: 11788969.

39. Kokubo T, Swanson MJ, Nishikawa JI, Hinnebusch AG, Nakatani Y. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA- binding protein. Mol Cell Biol. 1998; 18(2):1003–12. PMID: 9447997

40. Takahata S, Kasahara K, Kawaichi M, Kokubo T. Autonomous function of the amino-terminal inhibitory domain of TAF1 in transcriptional regulation. Mol Cell Biol. 2004; 24(8):3089–99. PMID: 15060133.

41. Takahata S, Ryu H, Ohtsuki K, Kasahara K, Kawaichi M, Kokubo T. Identification of a novel TATA element-binding protein binding region at the N terminus of the Saccharomyces cerevisiae TAF1 protein. J Biol Chem. 2003; 278(46):45888–902. PMID: 12939271.

42. Tsukihashi Y, Miyake T, Kawaichi M, Kokubo T. Impaired core promoter recognition caused by novel yeast TAF145 mutations can be restored by creating a canonical TATA element within the promoter region of the TUB2 gene. Mol Cell Biol. 2000; 20(7):2385–99. PMID: 10713163

43. Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res. 2002; 30(6):e23. PMID: 11884642; PubMed Central PMCID: PMC101367.

44. Kasahara K, Ki S, Aoyama K, Takahashi H, Kokubo T. Saccharomyces cerevisiae HMO1 interacts with TFIID and participates in start site selection by RNA polymerase II. Nucleic Acids Res. 2008; 36 (4):1343–57. PMID: 18187511. doi: 10.1093/nar/gkm1068

45. Hothorn M, Neumann H, Lenherr ED, Wehner M, Rybin V, Hassa PO, et al. Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. Science. 2009; 324(5926):513–6. Epub 2009/04/25. 324/5926/513 [pii] doi: 10.1126/science.1168120 PMID: 19390046.

46. Kuraev I.S. VVM, Kulakovskaya T.V. Methods of polyphosphate assay in biological materials. In: Kuraev I.S. VVM, Kulakovskaya T.V., editor. The Biochemistry of Inorganic Polyphosphates. Second Edition ed. England: John Wiley & Sons, Ltd; 2004. p. 15–35.

47. Lorenz B. SHC. Methods for investigation of inorganic polyphosphates and polyphosphate-metabolizing enzymes. In: Schröder H.C. MWEG, editor. Inorganic Polyphosphates. Progress in Molecular and Subcellular Biology. Germany: Springer-Verlag Berlin Heiderberg; 1999. p. 217–39.

48. Hahn S, Hoar ET, Guarente L. Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1985; 82 (24):8562–6. Epub 1985/12/01. PMID: 3001709; PubMed Central PMCID: PMC390957.

49. Hahn S, Buratowski S, Sharp PA, Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc Natl Acad Sci U S A. 1989; 86 (15):5718–22. PMID: 2569738.

50. Lue NF, Kornberg RD. Accurate initiation at RNA polymerase II promoters in extracts from Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1987; 84(24):8839–43. PMID: 3321057; PubMed Central PMCID: PMC299646.

51. Chen J, Ding M, Pederson DS. Binding of TFIID to the CYC1 TATA boxes in yeast occurs independently of upstream activating sequences. Proc Natl Acad Sci U S A. 1994; 91(25):11909–13. PMID: 7991556; PubMed Central PMCID: PMC45345.

52. Kuras L, Struhl K. Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. Nature. 1999; 399(6736):609–13. PMID: 10376605

53. Martens C, Krett B, Laybourn PJ. RNA polymerase II and TBP occupy the repressed CYC1 promoter. Mol Microbiol. 2001; 40(4):1009–19. PMID: 11401707.

54. Bjornsdottir G, Myers LC. Minimal components of the RNA polymerase II transcription apparatus determine the consensus TATA box. Nucleic Acids Res. 2008; 36(9):2906–16. PMID: 18385157. doi: 10.1093/nar/gkn130

55. Lee SK, Fletcher AG, Zhang L, Chen X, Fischbeck JA, Stargell LA. Activation of a poised RNAPII-dependent promoter requires both SAGA and mediator. Genetics. 2010; 184(3):659–72. doi: 10.1534/genetics.109.113464 PMID: 20048049; PubMed Central PMCID: PMC2845336.

56. Tsukihashi Y, Kawaichi M, Kokubo T. Requirement for yeast TAF145 function in transcriptional activation of the RPS5 promoter that depends on both core promoter structure and upstream activating sequences. J Biol Chem. 2001; 276(28):25715–26. Epub 2001/05/05. doi: 10.1074/jbc.M102416200M102416200 [pii]. PMID: 11337503.

57. Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, Soding J, et al. A conserved GA element in TATA-less RNA polymerase II promoters. PLoS One. 2011; 6(11):e27595. doi: 10.1371/journal.pone.0027595 PMID: 22110682; PubMed Central PMCID: PMC3217976.

58. Koehler RN, Rachfall N, Rolfes RJ. Activation of the ADE genes requires the chromatin remodeling complexes SAGA and SWI/SNF. Eukaryot Cell. 2007; 6(8):1474–85. PMID: 17573544.

59. de Almeida RA, Burgess D, Shema R, Motlekar N, Napper AD, Diamond SL, et al. A Saccharomyces cerevisiae cell-based quantitative beta-galactosidase assay compatible with robotic handling and high-throughput screening. Yeast. 2008; 25(1):71–6. doi: 10.1002/yea.1570 PMID: 17957822.

60. Bovee TF, Helsdingen RJ, Koks PD, Kuiper HA, Hoogenboom RL, Keijer J. Development of a rapid yeast estrogen bioassay, based on the expression of green fluorescent protein. Gene. 2004; 325:187–200. PMID: 14697524.

61. Kainth P, Sassi HE, Pena-Castillo L, Chua G, Hughes TR, Andrews B. Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. Methods. 2009; 48(3):258–64. doi: 10.1016/j.ymeth.2009.02.015 PMID: 19269327.

62. Ainsworth WB, Rome CM, Hjortso MA, Benton MG. Construction of a cytosolic firefly luciferase reporter cassette for use in PCR-mediated gene deletion and fusion in Saccharomyces cerevisiae. Yeast. 2012; 29(12):505–17. doi: 10.1002/yea.2931 PMID: 23172625.

63. Zeevi D, Sharon E, Lotan-Pompan M, Lubling Y, Shipony Z, Raveh-Sadka T, et al. Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. Genome Res. 2011; 21(12):2114–28. doi: 10.1101/gr.119669.110 PMID: 22009988; PubMed Central PMCID: PMC3227101.

64. Gertz J, Siggia ED, Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature. 2009; 457(7226):215–8. doi: 10.1038/nature07521 PMID: 19029883; PubMed Central PMCID: PMC2677908.

65. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol. 2009; 27(12):1173–5. doi: 10.1038/nbt.1589 PMID: 19915551; PubMed Central PMCID: PMC2849652.

66. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol. 2012; 30(3):265–70. doi: 10.1038/nbt.2136 PMID: 22371081; PubMed Central PMCID: PMC3402344.

67. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol. 2012; 30(3):271–7. doi: 10.1038/nbt.2137 PMID: 22371084; PubMed Central PMCID: PMC3297981.

68. Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol. 2012; 30(6):521–30. doi: 10.1038/nbt.2205 PMID: 22609971; PubMed Central PMCID: PMC3374032.

69. Mogno I, Kwasnieski JC, Cohen BA. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. Genome Res. 2013; 23(11):1908–15. doi: 10.1101/gr.157891.113 PMID: 23921661; PubMed Central PMCID: PMC3814890.

70. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013; 339(6123):1074–7. doi: 10.1126/science.1232542 PMID: 23328393.

71. Wobbe CR, Struhl K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol Cell Biol. 1990; 10(8):3859–67. PMID: 2196437; PubMed Central PMCID: PMC360896.

72. Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. 1999; 13(24):3217–30. PMID: 10617571

73. Moqtaderi Z, Bai Y, Poon D, Weil PA, Struhl K. TBP-associated factors are not generally required for transcriptional activation in yeast [see comments]. Nature. 1996; 383(6596):188–91. PMID: 8774887

74. Ljungdahl PO, Daignan-Fornier B. Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae. Genetics. 2012; 190(3):885–929. doi: 10.1534/genetics.111.133306 PMID: 22419079; PubMed Central PMCID: PMC3296254.

75. Park D, Morris AR, Battenhouse A, Iyer VR. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res. 2014; 42(6):3736–49. doi: 10.1093/nar/gkt1366 PMID: 24413663; PubMed Central PMCID: PMC3973313.