Supplemental Material

Promoter-specific dynamics of TATA-binding protein association with the human genome

Yuko Hasegawa and Kevin Struhl

Table of Contents

SUPPLEMENTAL METHODS

Plasmid construction

The DNA fragment coding 3xHA was prepared by annealing oligos (Supplementary Table 2,) and inserted between the *Bam*HI and *Eco*RI sites of pBudCE4.1 (Invitrogen). The ERT2 DNA fragment was PCR amplified from pCAG-Cre-ERT2 (Addgene #14797) using 1Fw and 1Rv primers and inserted between *Sal*I and *Bam*HI sites of pBudCE4.1-3xHA. ERT2-3HA fragment was then PCR amplified using 2Fw and 2Rv primers and inserted between *Bmt*I and *Bam*HI sites of pmCherry-C1 (Clontech). The TBP coding sequence was PCR amplified from human cDNA and was cloned into *Sal*I site of this plasmid containing ERT2-3HA, yielding pCMV-TBP-ERT2-3HA.

Cell culture

pCMV-TBP-ERT2-3HA construct was transfected to HEK293 (ATCC) cells, and the stable line was established by G418 selection. Cells were routinely cultured in phenol-red free DMEM containing 10% FBS.

Nuclear and nucleolar fractionation

Cells cultured in 6 cm dish were washed twice with ice-cold PBS twice, then scraped and collected in 1.5 ml tubes. After brief centrifugation, cells were suspended in 250 μl of CE buffer (10 mM Hepes-KOH (pH7.5), 1.5 mM MgCl2, 10 mM KCl, 0.05% NP-40) and incubated for 3 min on ice. Cells were again centrifuged for 3 min at 3,000 rpm, and the pellet were suspended in 250 μl of RIPA buffer (50 mM Tris-HCl (pH8.0), 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 0.5% Na-Deoxycholate, 0.1% SDS) and sonicated (Misonix 3000,

level 2, on 30 sec off 30 sec, total 2 min). After centrifugation for 12,000 rpm for 10 min, the supernatant was collected and used as the nuclear fraction. For quantification of the protein, we performed western blotting and detected TBP-ERT2 and endogenous TBP using anti-TBP antibody (Abcam #ab51841). The signals were detected by LAS3000 imaging system and analyzed by ImageJ software.

To isolate nucleoli, cells in five 150-mm dishes were trypsinized and collected in the 15 ml tubes, then suspended in ice-cold CE buffer without detergent (10 mM Hepes-KOH (pH7.9), 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT) and incubated for 5 min on ice. The cell suspension was transferred to pre-chilled 7 ml Dounce tissue homogenizer, and homogenized to obtain nuclei. After centrifuge and removed supernatant, the nuclear pellet was suspended in 3 ml of S1 buffer (0.25 M Sucrose, 10 mM $MgCl_2$) and then put on the top of 3 ml S2 buffer (0.35 M Sucrose, 0.5 mM MgCl2). The tube was centrifuged at 2,500 rpm for 5 min at 4°C, and after that supernatant was removed. The pellet was suspended in 3 ml of S2 buffer and sonicated for 3 x 10 sec using a probe sonicator. The sonicated solution was layered onto 3 ml of S3 buffer (0.88 M Sucrose, 0.5 mM $MgCl_2$), and centrifuged for 10 min at 3,500 rpm, 4°C. The pellet containing nucleolus was washed with 500 µl of S2 buffer, spun down at 3,500 rpm for 5 min and then used for the assay.

Sample preparation during the time course

TBP-ERT2 nuclear translocation was induced by 4OHT (Sigma-Aldrich, H7904) addition at the final concentration of 100 nM to culture medium. Cells were retained in the CO2 incubator for indicated times. For western blotting, cells were immediately put on ice, and proceed to the subcellular fractionation. For ChIP, culture medium was removed and then exchanged by fixing solution (see below) at room temperature to avoid over crosslinking. This step takes 2.5 minutes, thus we summed up this time gap and incubation time and used as real time points (7.5, 12.5, 17.5, 32.5, 62.5, 92.5, 362.5, 1442.5 min) for simulating turnover rate.

Chromatin immunoprecipitation

Chromatin was prepared from two 150 mm dishes. After removing the culture medium, cells were fixed at room temperature with fixing solution (1% formaldehyde, methanol-free in DMEM without serum) for 10 min. Fixation was stopped by adding 125 mM glycine. Cells were washed with ice-cold PBS twice, and then scraped from dishes and collected. Cells were suspended in 2 ml Lysis Buffer 1 (50 mM Hepes-KOH (pH7.5), 140 mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100, with protease inhibitor and phosphatase inhibitor), rotated for 10 min at 4°C. After collecting cells by centrifugation, cells were re-suspended in 2 ml Lysis Buffer 2 (10 mM Tris-HCl (pH8.0), 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, with protease inhibitor and phosphatase inhibitor) and rotated for 10 min at 4°C. Cells were collected by centrifugation and re-suspended in 1.3 ml Lysis Buffer 3 (10 mM Tris-HCl (pH8.0), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Sodium deoxycholate, 0.5% N-Lauroylsarcosine, with protease inhibitor and phosphatase inhibitor), and then sonicated (Misonix 3000 sonicator, level 2, ON 30 sec, OFF 30 sec, 7 cycles (total sonication time is 3.5 min)). Majority of the fragment sizes were approximately 200-500 bp. Cell lysates were centrifuged at 12,000 rpm for 10 min at 4°C after adding Triton X-100 (1% final concentration), and the supernatants were used for the analysis. The protein concentration of the lysates was measured by BCA Protein Assay Kit (Thermo Scientific) and the amounts of the lysates yielding 1.3 μg total protein were used for immunoprecipitation. The volume of the lysates were adjusted to 1 ml by adding Lysis Buffer 3. For sequencing samples, yeast chromatin prepared from HA-TBP expressing strain were added at the constant ratio (0.039 μg total protein in the lysate, that corresponds to 3% of total human protein amount in the lysates) to the human cell lysates as a spike-in control. 1/10 volumes of the mixed lysates were kept as input samples. 0.3 μg of anti-HA antibody (Santa Cruz #sc-7392) or 3 μg of anti-TBP antibody (Abcam #ab51841) was added to the lysates, and the mixtures were rotated overnight at 4°C. Immunoprecipitation was performed at 4°C for 3 hr with 50 μl protein G dynabeads (Thermo Fisher). Beads were washed five times with ice-cold Wash buffer 1 (50 mM Hepes-KOH

(pH7.6), 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Sodium deoxycholate) and rinsed with TE (10 mM Tris-HCl (pH8.0), 1 mM EDTA) with 50 mM NaCl. Immunoprecipitates were eluted by incubating the beads in 75 μl Elution buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS) at 65°C for 15 min and collecting the supernatants. This step was repeated twice. The eluted samples were incubated overnight at 65°C for de-crosslinking. After treating with RNase A and Proteinase K, DNA was purified with PCR purification kit (QIAGEN) by adding 100 μl water. 55.5 μl of eluted DNA from IP samples and 2.5 μl of eluted DNA (diluted to 1/10) from input samples were used for the library preparation. For qPCR analysis, 3.3 μl of eluted DNA from IP samples and four dilution series of input samples for standards were used. Primers are listed in Supplementary Table 2.

Yeast chromatin preparation

HA-TBP yeast strain was grown in YPD at 30 °C to an 0.4 OD600 and fixed for 20 min at room temperature with final 1% formaldehyde. Then the fixation was quenched with final 250 mM glycine for 5 min at room temperature. Cells were collected and washed with PBS once, and were disrupted in Mini Bead Beater (BioSpec Products, Inc.) with 3 cycles of 5 min run at maximum speed and 1 min rest on ice following suspending in 400 μl of FA lysis buffer (50 mM Hepes-KOH (pH7.5), 150 mM NaCl, 1% Triton X-100, 0.1% Sodium deoxycholate, 0.1% SDS). After centrifugation and removal of the supernatant, the pellet was suspended in 480 μl of FA lysis buffer and sonicated (Misonix 3000 sonicator, level 6, ON 10 sec, OFF 10 sec, 21 cycles (total sonication time is 3.5 min) x 6 times with 1 min interval). The sample was centrifuged at maximum speed at 4°C for 20 min, and then the supernatant was collected and used as yeast chromatin.

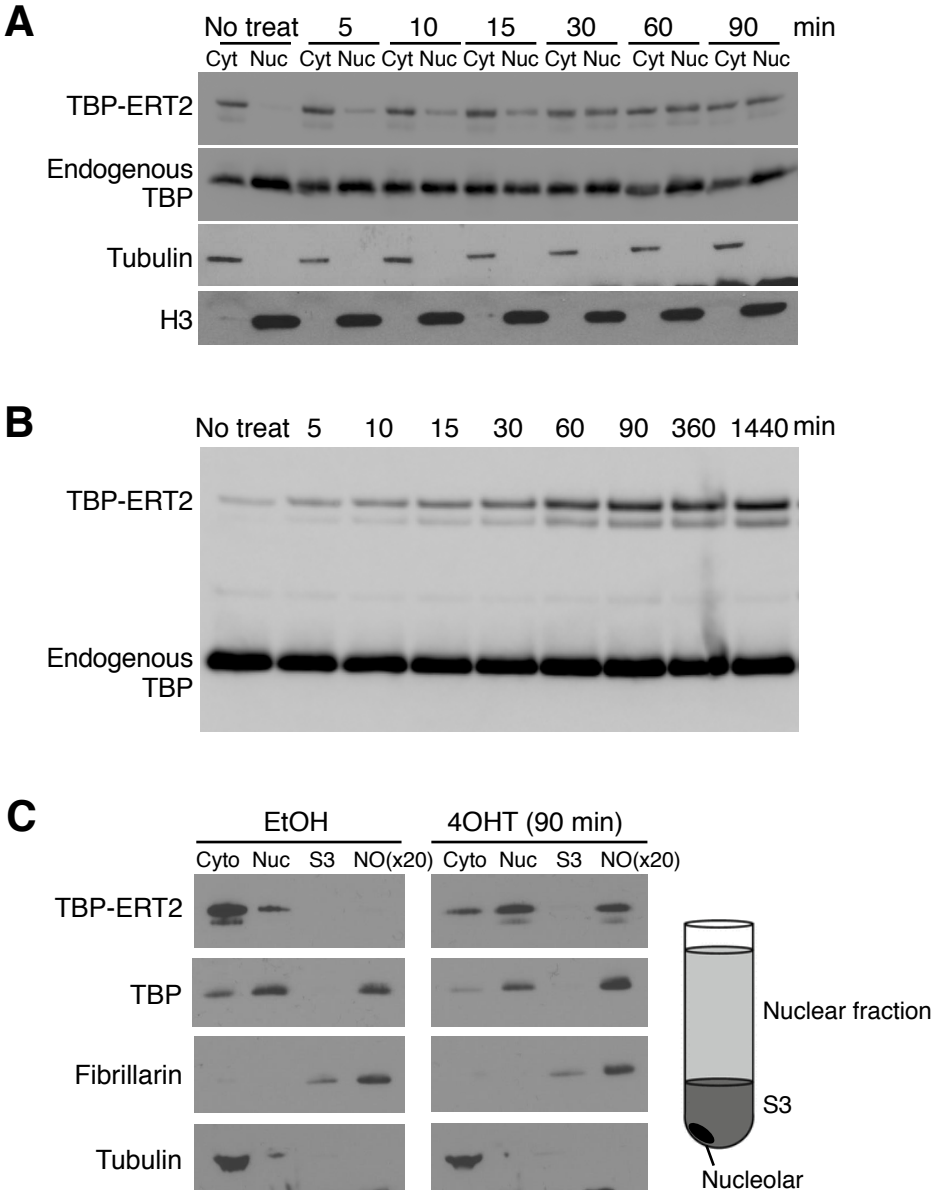
Library preparation and sequencing

Sequencing library preparation was performed with NEBNext Ultra DNA library prep kit (NEB #E7370S) and NEBNext Multiplex Oligos (NEB #E7335S) following the

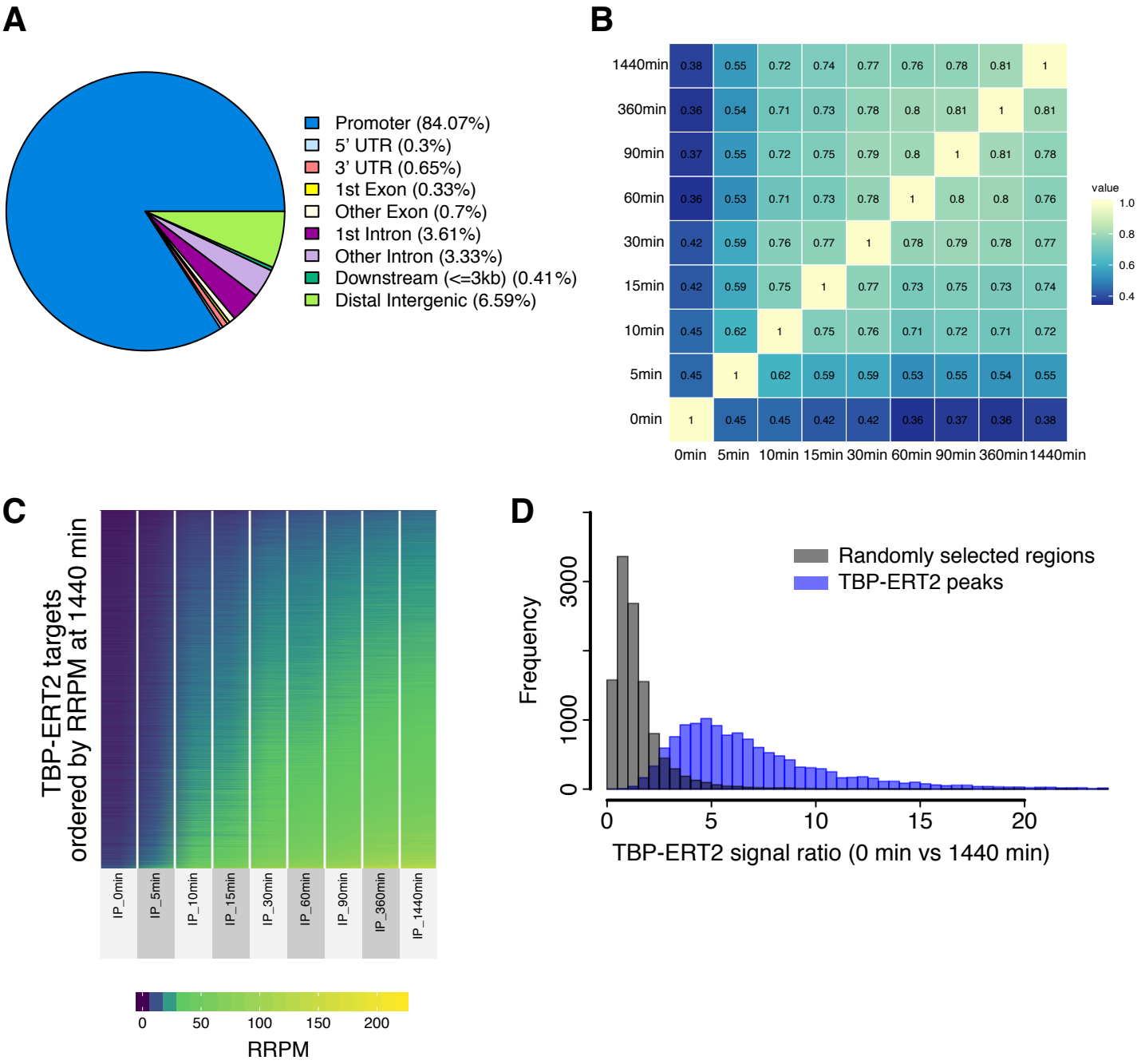
manufacturer's instructions. PCR cycles for amplification of adaptor-ligated DNA were 9 cycles for input samples and 14 cycles for IP samples. Large fragments were removed using 0.4 x AMPure beads (Beckman Coulter). Libraries were sequenced on NextSeq 500.

ChIP-seq computational analysis

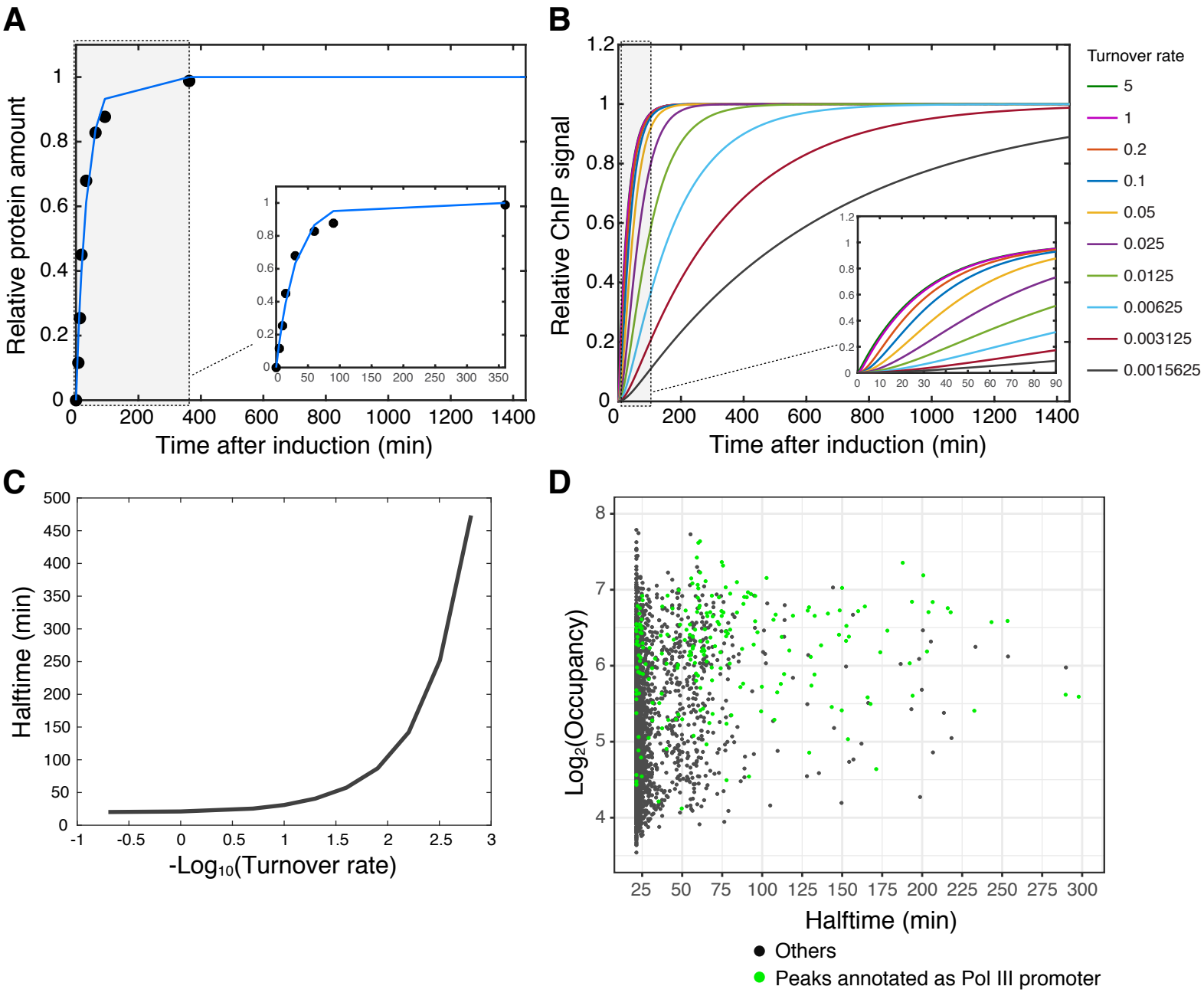
We concatenated the genome sequences of human (hg38) and cerevisiae (sacCer3) to generate a combined genome sequence. We added "_s" to all yeast chromosome names to avoid chromosome name duplication, and built a custom Bowtie2 index from this combined genome sequence. Sequenced reads were aligned to this custom library with default parameters (--sensitive). The resulting SAM files were then split into the two files containing reads mapped to human chromosomes and mapped to yeast chromosome. Duplicates were removed by Picard, and only reads exceed mapping quality 30 were retained except for specific analysis (repetitive genes including ribosomal DNA). The number of reads mapped to human or yeast genomes and that passed the quality filter are listed in Supplementary Table 1. We calculated RRPM (reference adjusted RPM) by dividing the number of the reads mapped to human genome loci by the total reads number mapped to yeast genome (the numbers in the MQ>30 column in Supplementary Table 1, and that in the MQ>0 column for repetitive gene analysis). This RRPM value was used as the intensity of the ChIP signal in this study. Peak calling was performed using MACS2 with a *P*-value threshold of 0.01. We extracted 13,148 peaks that were called at least two samples in the later time points (60, 90, 360, 1440 min). Peak annotation was performed using UROPA with custom GTF file combined GRCh38.94 and tRNA GTF files downloaded from Ensembl and UCSC genome browser respectively.



Supplemental Fig. S1. TBP-ERT2 localizes to the nucleus upon tamoxifen treatment. (*A*) Western blotting analysis of TBP-ERT2, endogenous TBP, Tubulin alpha (cytoplasmic marker), and Histone H3 (nuclear marker) in the cytoplasmic (Cyt) and the nuclear (Nuc) fraction following the given time after tamoxifen treatment. (*B*) Western blotting analysis of TBP-ERT2 and endogenous TBP in nuclear fraction. (*C*) Western blotting analysis of TBP-ERT2, endogenous TBP, Fibrillarin (nucleolar marker), and Tubulin alpha (cytoplasmic marker) in the cytoplasmic (Cyt), nuclear (Nuc), S3, and nucleolar (NO) fraction before (EtOH) and 90 min after tamoxifen treatment (4OHT). Nucleolar fraction was concentrated to 20-fold.

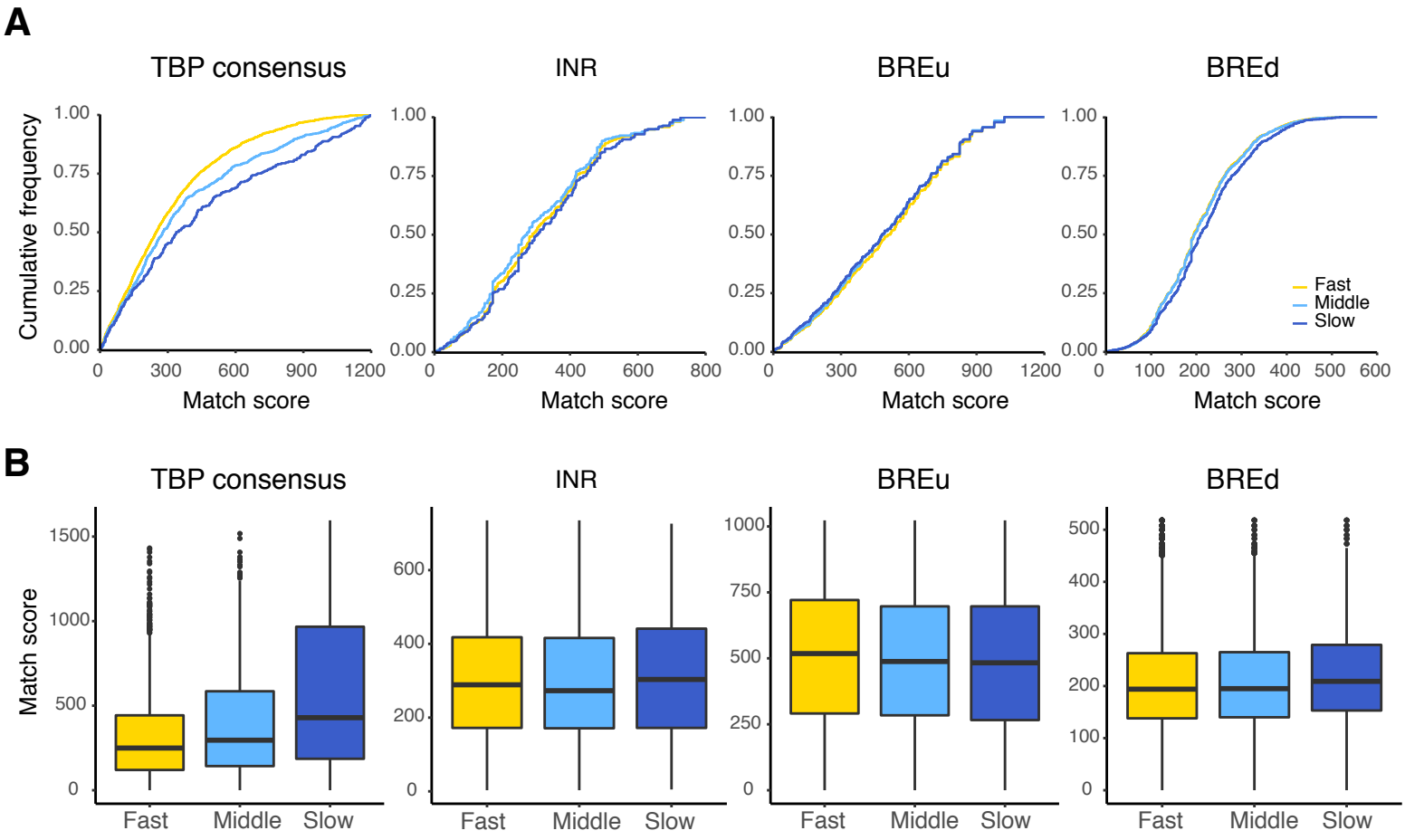


Supplemental Fig. S2. TBP-ERT2 ChIP signal gradually increases after tamoxifen treatment. (*A*) Location of TBP-ERT2 peaks. (*B*) Spearman correlation of ChIP-seq signal at peaks between samples at a given time point after tamoxifen induction. (*C*) ChIP-seq signal (RRPM: Reference-adjusted RPM) at each target site through time-course. (*D*) Histogram of the frequency (number) of regions or TBP-ERT2 peaks having indicated ChIP-seq signal ratio between 0 min and 1440 min samples.



Supplemental Fig. S3. ChIP signal increasing is reflected in the binding turnover rate. (*A*) Change of the nuclear TBP-ERT2 amount over the time-course. Markers are real data measured by western blotting (means of the three independent analyses) and the line is simulated change of the protein amount based on the equation describing nuclear translocation of the protein (Methods). Small window indicates the enlargement of the data at the early time points. (*B*) Simulated increase of the relative ChIP signal with various Lambda (binding turnover rate) values. Small window indicates the enlargement of the data at the early time points. (*C*) Relation between Lambda (binding turnover rate) and halftime (time point where the relative ChIP signal reaches to 0.5). (*D*) Scatter plot of TBP-ERT2 peaks in log2 occupancy vs halftime. Green color indicates peaks annotated as pol III genes. Black are the others.

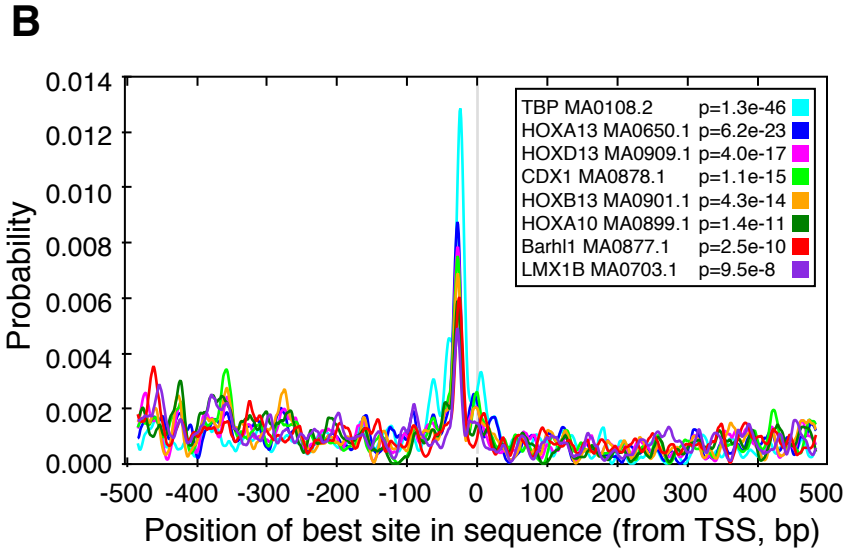
Supplemental Fig. S4. Cutoff for peaks that we set based on Goodness-of-fitting.

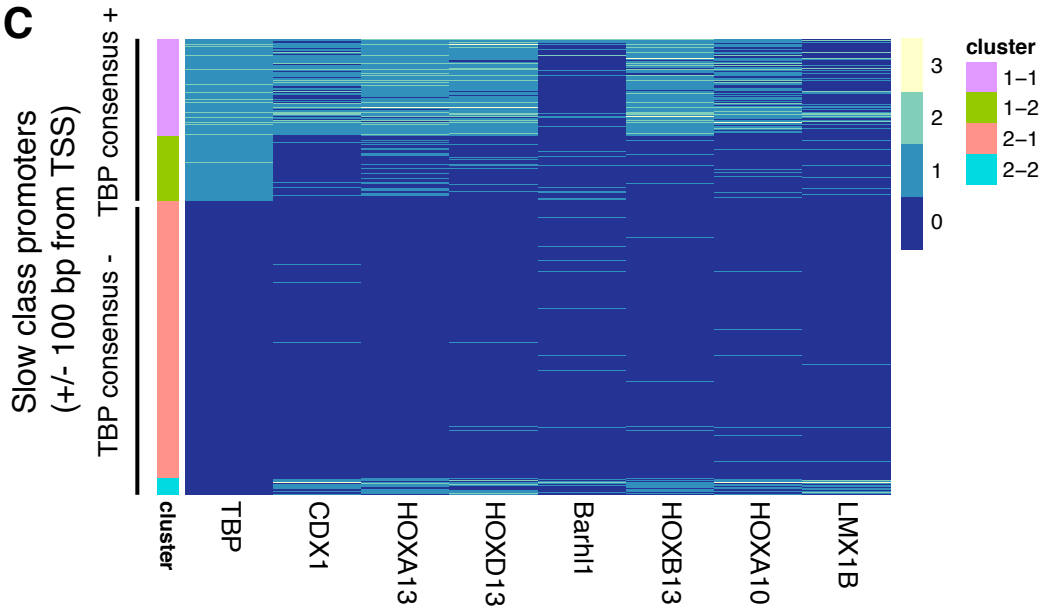


Supplemental Fig. S5. Analysis of motifs in Pol II core promoter. (*A*) Cumulative frequencies of match score of hit sequences (forward direction) to each motif in ±100 bp region (±10 bp region for INR) from gene start site. (*B*) Boxplots of the highest match score of hit sequences (forward direction).

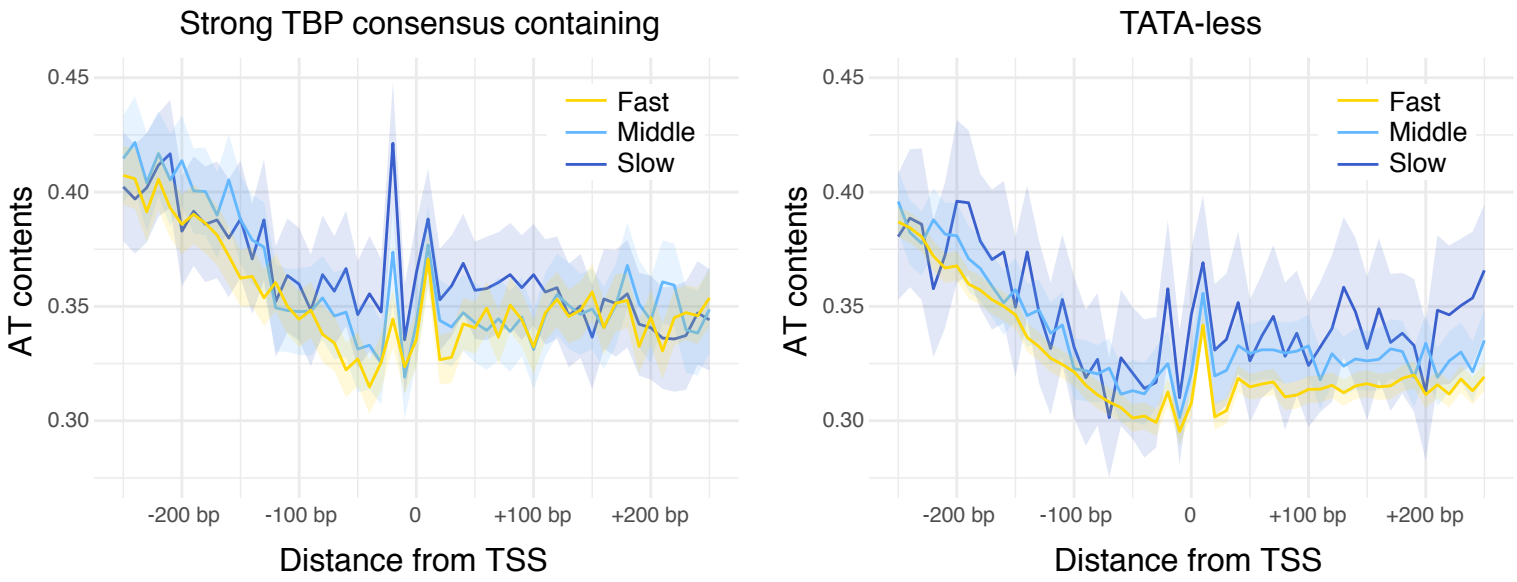
A

Name	Consensus sequence	Fisher E-value
TBP	STATAAAWRSVVVBN	2.8e-38
CDX1	GYMATAAAA	1.5e-8
HOXA13	KRCCACGCCCMCYT	1.5e-8
HOXD13	CYMATAAAAH	3.8e-7
Barhl1	NNTAAWTGNB	1.0e-4
HOXB13	CCAATAAAAH	1.1e-3
HOXA10	DGYMATAAAAH	4.3e-3
LMX1B	NYAATTAA	1.9e-2

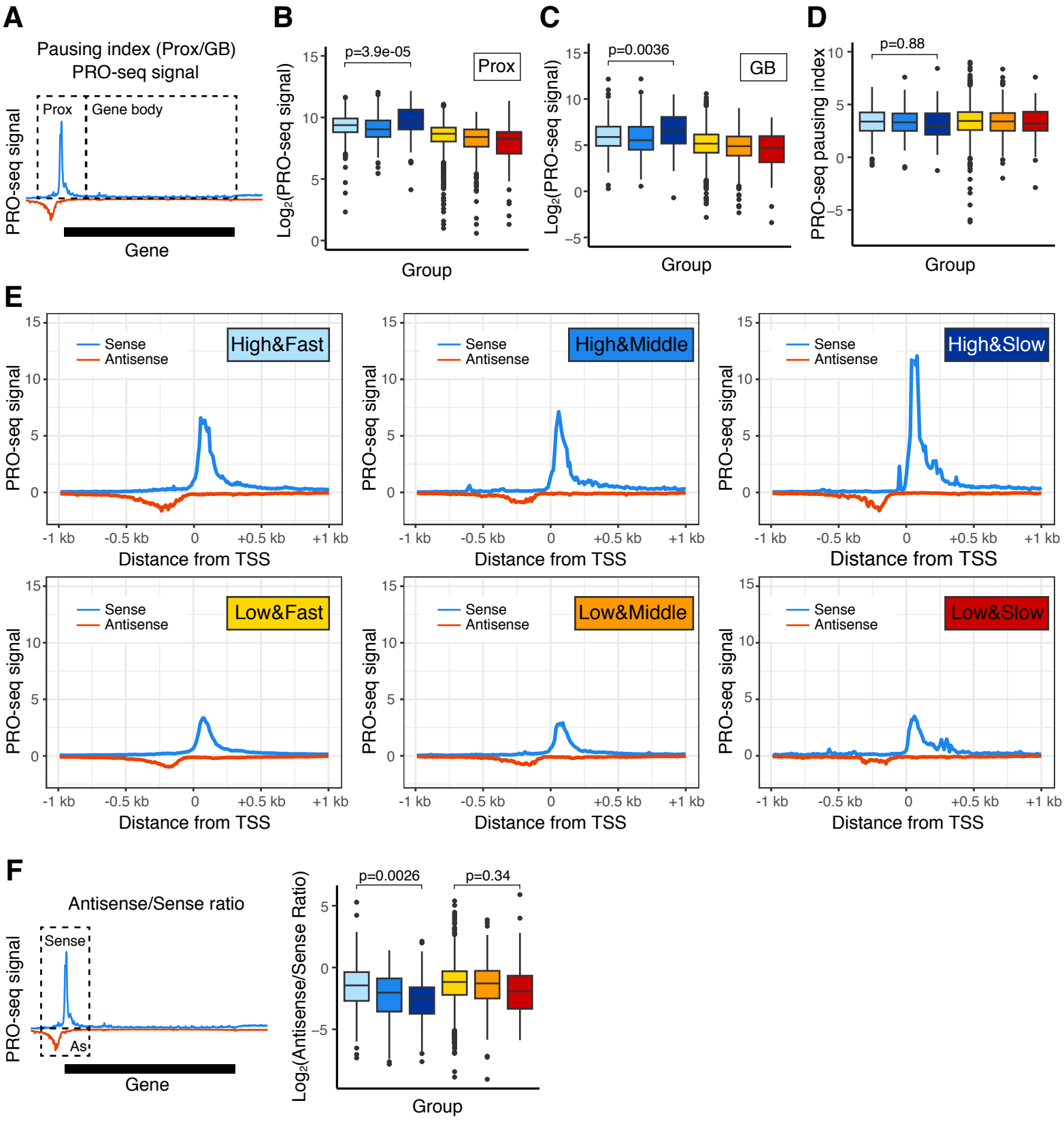




Supplemental Fig. S6. Known TF binding motifs enriched in slow class promoters searched by CentriMo (http://meme-suite.org/doc/centrimo.html). (*A*) The list of TFs of which binding motifs are significantly enriched in Slow class promoters (Fisher E-value < 0.05, ± 500 bp from TSS). (*B*) Location of the best hit sites in the sequences. (*C*) Occurrences of TF binding motifs in Slow class promoters (±100 bp from TSS).

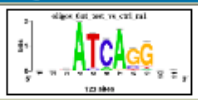
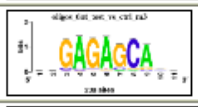
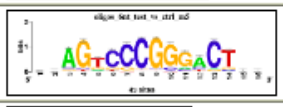
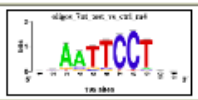
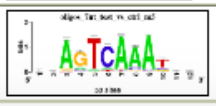


Supplemental Fig. S7. TATA-less pol II promoters in Slow class tend to have higher AT content. Average AT-contents in the promoters containing strong TBP consensus (left) or TATA-less promoters (right). Promoters that have strong TBP consensus (*P*-value < 0.0001, PWMtools) or TATAWAW (RSAT, pattern matching) are categorized as Strong TBP consensus containing (n=1459, Fast n=895, Middle n=415, Slow n=149). Rest were categorized in TATA-less (n=4164, Fast n=3201, Middle n=700, Slow n=263). Shadow areas represent 95% confidence intervals for the means.

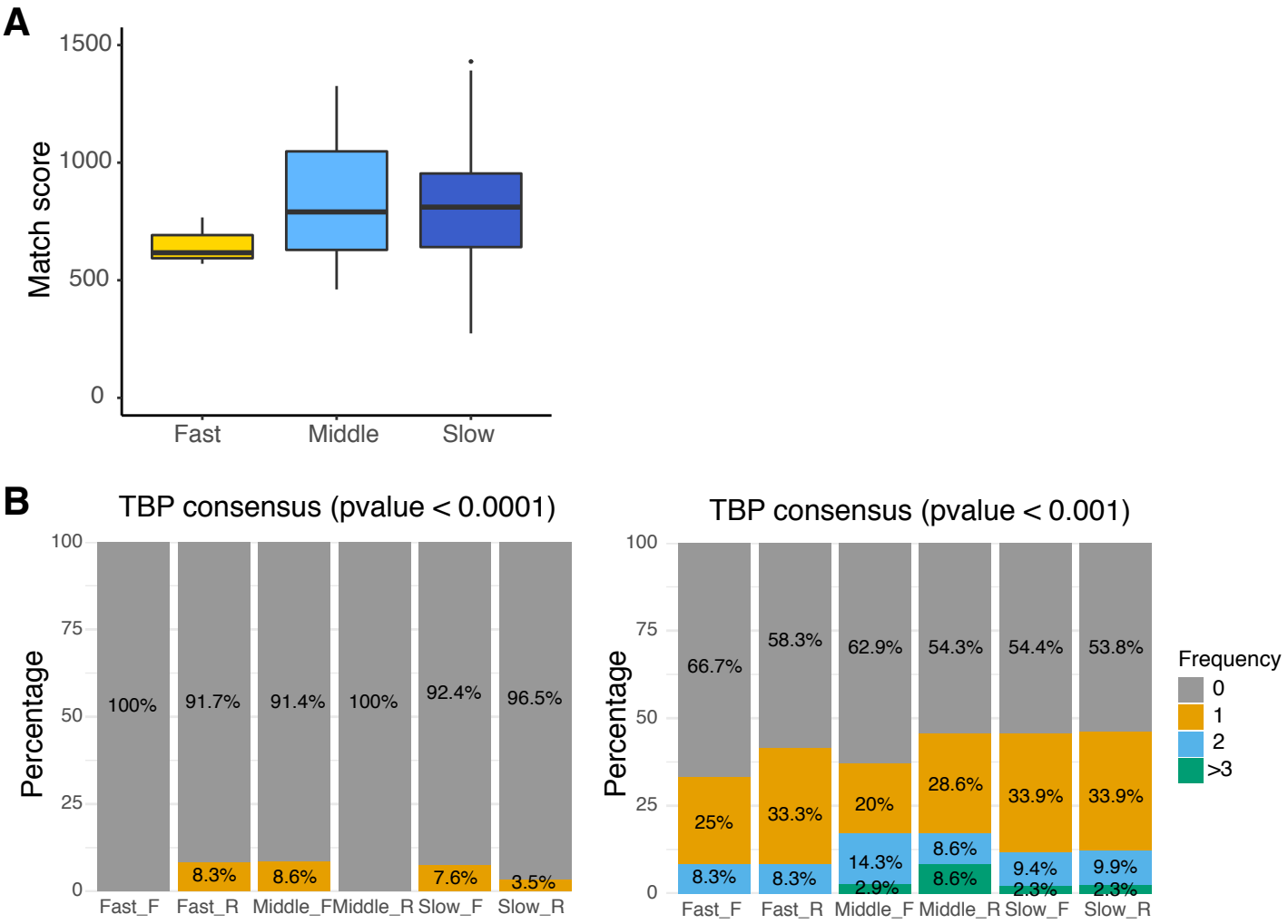


Supplemental Fig. S8. Polymerase pausing does not associate with the binding turnover rate of TBP, and High and Slow class Pol II promoters show strong transcription in sense direction. (*A*) Schematic illustration of the proximal region (Prox: from -500-bp to +500-bp of TSS) and the gene body region (GB: from +500-bp of TSS to TES). (*B*) PRO-seq signal (Woo et al. 2018*) in the proximal region. (*C*) PRO-seq signal in the gene body (GB) region. (*D*) Pausing index. *P*-values are *t*-test results. (*E*) PRO-seq signal in sense (blue) and anti-sense (red) direction. (*F*) Schematic illustration of the region (from -500-bp to +500-bp of TSS) for counting PRO-seq signal (left), and $\log_2$ ratio between anti-sense and sense PRO-seq signal (right). *P*-values are *t*-test results.

*Woo YM, Kwak Y, Namkoong S, Kristjansdottir K, Lee SH, Lee JH, Kwak H. 2018. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Rep* 24: 3630-3641 e3637.

Motif	Logo	3 Top hits in databases
oligos_6nt_test_vs_ctrl_m1 E-value=9.1e-12		versus jaspar_core_nonredundant_vertebrates: Dux, SIX1, GATA5,
oligos_6nt_test_vs_ctrl_m2 E-value=9.3e-07		versus jaspar_core_nonredundant_vertebrates: EWSR1_FLI1,
oligos_6nt_test_vs_ctrl_m3 E-value=2.6e-06		versus jaspar_core_nonredundant_vertebrates: no match
oligos_6nt_test_vs_ctrl_m4 E-value=3.7e-05		versus jaspar_core_nonredundant_vertebrates: no match
oligos_6nt_test_vs_ctrl_m5 E-value=6.3e-05		versus jaspar_core_nonredundant_vertebrates: no match
oligos_7nt_test_vs_ctrl_m1 E-value=3.5e-22		versus jaspar_core_nonredundant_vertebrates: YY2, NFIX,
oligos_7nt_test_vs_ctrl_m2 E-value=4.1e-20		versus jaspar_core_nonredundant_vertebrates: Hand1_Tcf3,
oligos_7nt_test_vs_ctrl_m3 E-value=4.4e-18		versus jaspar_core_nonredundant_vertebrates: Dux, NFYA,
oligos_7nt_test_vs_ctrl_m4 E-value=3.9e-14		versus jaspar_core_nonredundant_vertebrates: TEAD4, TEAD1, ELF5,
oligos_7nt_test_vs_ctrl_m5 E-value=3.9e-14		versus jaspar_core_nonredundant_vertebrates: LIN54, Nr2e1, Dux,

Supplemental Fig. S9. Discovered motifs enriched in high TBP occupancy and fast TBP turnover class of promoters that are strongly transcribed (Pol II-ser5 ChIP signal of Fig. 4D > 1; n = 219) compared to high occupancy and slow turnover with similar transcriptional activity (n = 125).



Supplemental Fig. S10. TBP consensus motif in Pol III genes. (*A*) Boxplots of the highest match score of hit sequences in forward direction. (*B*) Percentages of the pol III promoters (pseudogenes were removed, n=217, Fast n=11, Middle n=35, Slow=171) that have hit sequences to TBP consensus motif (left: *P*-value < 0.0001, right: *P*-value < 0.001, PWMtools).

Sample	Total reads	Mapped 0 times	% in the total reads	Mapped	% in the total reads	Mapped to hg38	% in the mapped reads	Mapped to sacCer3	% in the mapped reads
input_0min	32,834,384	544,285	1.7%	32,290,099	98.3%	32,251,821	99.9%	38,278	0.1%
input_5min	40,468,937	565,707	1.4%	39,903,230	98.6%	39,828,313	99.8%	74,917	0.2%
input_10min	34,332,896	531,837	1.5%	33,801,059	98.5%	33,735,832	99.8%	65,227	0.2%
input_15min	45,183,241	640,534	1.4%	44,542,707	98.6%	44,471,017	99.8%	71,690	0.2%
input_30min	54,189,936	727,886	1.3%	53,462,050	98.7%	53,330,184	99.8%	131,866	0.2%
input_60min	43,575,407	678,361	1.6%	42,897,046	98.4%	42,787,995	99.7%	109,051	0.3%
input_90min	39,602,962	646,229	1.6%	38,956,733	98.4%	38,866,732	99.8%	90,001	0.2%
input_360min	41,574,884	680,870	1.6%	40,894,014	98.4%	40,797,127	99.8%	96,887	0.2%
input_1440min	44,255,025	680,279	1.5%	43,574,746	98.5%	43,482,779	99.8%	91,967	0.2%
IP_0min	61,463,578	1,412,705	2.3%	60,050,873	97.7%	56,694,056	94.4%	3,356,817	5.6%
IP_5min	48,153,540	970,311	2.0%	47,183,229	98.0%	45,245,093	95.9%	1,938,136	4.1%
IP_10min	62,715,387	1,127,539	1.8%	61,587,848	98.2%	59,532,301	96.7%	2,055,547	3.3%
IP_15min	46,491,870	1,030,356	2.2%	45,461,514	97.8%	43,406,502	95.5%	2,055,012	4.5%
IP_30min	59,597,348	1,242,744	2.1%	58,354,604	97.9%	56,383,307	96.6%	1,971,297	3.4%
IP_60min	56,181,699	1,250,257	2.2%	54,931,442	97.8%	52,472,684	95.5%	2,458,758	4.5%
IP_90min	55,323,612	1,373,512	2.5%	53,950,100	97.5%	51,364,702	95.2%	2,585,398	4.8%
IP_360min	62,691,237	1,596,832	2.5%	61,484,366	98.1%	58,894,000	95.8%	2,590,366	4.2%
IP_1440min	64,460,688	1,206,871	1.9%	63,129,110	97.9%	60,192,825	95.3%	2,936,285	4.7%

Supplemental Table S1. Numbers of reads mapped to human genome (hg38) and yeast genome (sacCer3)

Experiment	Primer Name	Sequence
Plasmid construction	3xHA Fw	GATCCTACCCTTATGACGTCCCCGACTACGCTGGTTACCCTTATGACGTCCCCGACTATGCCGGCTACCCCTACGATGTCCCGGATTACGCATGAG
Plasmid construction	3xHA Rv	AATTCTCATGCGTAATCCGGGACATCGTAGGGGTAGCCGGCATAGTCGGGGACGTCATAAGGGTAACCAGCGTAGTCGGGGACGTCATAAGGGTAG
Plasmid construction	1Fw	ATTCCTGCAGGTCGACTCTGCTGGAGACATGAGAGCTGC
Plasmid construction	1Rv	CATAAGGGTAGGATCCGGCCGCAGCTGTGGCAGGGAAACC
Plasmid construction	2Fw	GAACCGTCAGATCCGCTAGCGTCGACGGACCATCTGCTGGAGACATG
Plasmid construction	2Rv	TAGATCCGGTGGATCCGAATTCTCATGCGTAATCCGGGAC
qPCR	qPCR-45S Fw	CCGGGGGAGGTATATCTTT
qPCR	qPCR-45S Rv	CCAACCTCTCCGACGACA
qPCR	qPCR-5S Fw	CCTGAACGCGCCCGATCTC
qPCR	qPCR-5S Rv	AAGCCTACAGCACCCGGTATTC
qPCR	qPCR-RAB5B Fw	GGTCAGTAGCGGCCTTCCTA
qPCR	qPCR-RAB5B Rv	GAGGCGGGTCAAAGTGAAAC
qPCR	qPCR-RPS9 Fw	TCTCGTGACGTTTTCACGCACC
qPCR	qPCR-RPS9 Rv	CTCCCTACTTCTCAGGTTTCCACT
qPCR	qPCR-chr8NC Fw	CCATCCCACTAGGGAGAAAA
qPCR	qPCR-chr8NC Rv	ATGCAACAGCTGTGAGGTGA

Supplemental Table S2. Primer list

GEO accession	Experiment	Cell line	Reference
GSE133729	TBP-ERT2 ChIP-seq	HEK293	This study
DRX000062 (SRA)	MNase-seq	HEK293	Sugano et al. 2010.
GSE31477 (sample GSM935534)	pol II ChIP-seq (Ab: 8WG16)	HEK293	ENCODE project
GSE97827 (sample GSM2579073)	pol II Ser5-P ChIP-seq	HEK293	Fong et al. 2017.
GSE97827 (sample GSM2579070)	pol II Ser7-P ChIP-seq	HEK293	Fong et al. 2017.
GSE97827 (sample GSM2579063)	pol II Ser2-P ChIP-seq	HEK293	Fong et al. 2017.
GSE31477 (sample GSM935302)	KAT2A(GCN5) ChIP-seq	HeLa-S3	ENCODE project
GSE127427 (sample GSM3634283)	EP300 ChIP-seq	HeLa-S3	ENCODE project
GSE31477 (sample GSM935511)	SMARCA4(BRG1) ChIP-seq	HeLa-S3	ENCODE project
GSE32465 (sample GSM803455)	TAF1 ChIP-seq	HeLa-S3	ENCODE project
GSE37403 (sample GSM917814)	SNAPC1 ChIP-seq	MCF10A	Baillat et al. 2012.
GSE32970 (sample GSM1008573)	DNase-seq	HEK293T	ENCODE project
GSE118739 (sample GSM3346248)	POLR3A ChIP-seq	HEK293	Choquet et al. 2019.
GSE103719 (sample GSM2779677)	PRO-seq	HEK293	Woo et al. 2018.

Supplemental Table S3. GEO accession numbers of publicly available datasets used in this study