# Different SP1 binding dynamics at individual genomic loci in human cells

Yuko Hasegawa and Kevin Struhl*

Dept. Biological Chemistry and Molecular Pharmacology

Harvard Medical School

Boston, MA 02115

SI Appendix

## MATERIALS AND METHODS

### Plasmid construction

gRNA sequences (Dataset S3, SP1gRNA) were inserted between two BbsI sites of pX330 (Addgene #42230) to obtain pX330-SP1gRNA#1 and pX330-SP1gRNA#2 vectors. A PCR-amplified fragment from pX330-SP1gRNA#2 with primers pX330XbaI and pX330EcoRV was inserted between the XbaI and EcoRV sites of pX330-SP1gRNA#1 to express two gRNAs from one vector. To construct a knock-in template, a PCR fragment amplified from K562 genomic DNA with SP1 cloning primers was inserted between EcoRV and EcoRI sites in pBluescript (pBlue-SP1genome), and genomic sequences corresponding to PAM were mutated by site-directed mutagenesis. ERT2-3xHA was amplified from pCMV-TBP-ERT2-3HA with ERT2_Fw and ERT2_Rv primers, and P2A-PuroR was amplified from lentiCRISPR v2 (Addgene #52961) with PuroR_Fw and PuroR_Rv primers. ERT2-3xHA and P2A-PuroR fragments were joined to PCR product amplified from pBlue-SP1genome with SP1vector_Fw and SP1vector_Rv primers by In-Fusion. We used this construct as a knock-in template for inserting ERT2-3xHA-PuroR after 3' end of SP1 coding region. Primers are listed in Dataset S3.

### Cell line and culture

The K562 cell line was purchased from ATCC (ATCC CCL-243) and cultured with IMDM a modification of Dulbecco's Modified Eagle Medium, (GIBCO) with 10% FBS. For establishing the SP1-ERT2-3HA expressing cell line, pX330 with gRNA sequences and the knock-in template described above were transfected to K562 cells by Nucleofector (Amaxa). Cells were selected by puromycin in TCS medium (STEMCELL technologies) and individual colonies were picked and expanded. We detected the ERT2-3HA insertion by PCR using Genotyping Fw and Rv primers (Dataset S3). To induce expression of SP1-ERT2-3HA, 4-hydroxy tamoxifen (4OHT; Sigma, H7904) was added to culture medium to a

final concentration of 100 nM for the indicated times. For western blotting, cells were immediately placed on ice, and subjected to the subcellular fractionation. For ChIP, culture medium was exchanged with fixing solution (described below) at room temperature.

**Nuclear fractionation**

Nuclear fractionation was performed as described previously (1). Cells were cultured in T25 flasks and washed with ice-cold PBS twice. Cells were suspended in 250 ul of CE buffer (10 mM Hepes-KOH (pH7.5), 1.5 mM MgCl2, 10 mM KCl, 0.1% NP-40) and incubated for 3 min on ice. Cells were again centrifuged for 3 min at 700 g, and the pellet was suspended in 250 ul 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) followed by a brief sonication (Misonix 3000, level 2, on 30 sec off 30 sec, total 2 min). The lysate was centrifuged at 12,000 rpm for 10 min, and the supernatant was used as the nuclear fraction. Western blotting was performed to detect SP1-ERT2-3HA and endogenous SP1 using anti-HA antibody (Abcam #ab51841) or anti-SP1 antibody (Millipore #07-645). Protein signals were detected by the LAS3000 imaging system and analyzed by ImageJ software.

**Chromatin immunoprecipitation (ChIP)**

K562 cells in two T75 flasks were treated with 4OHT at 100 nM for indicated times, and ChIP was performed essentially as described previously (1). Cells were fixed with fixing solution (1% formaldehyde in IMDM without serum) for 10 min at room temperature. After stopping fixation by adding 125 mM glycine, cells were washed with ice-cold PBS twice. 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), and 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, total sonication time 8 min). 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 1.7 $\mu$g total protein was used for the immunoprecipitation step. Yeast chromatin were prepared from HA-TBP expressing strain as following and added to the lysates at a constant ratio (3% of total protein amount in the K562 cell lysates) as a spike-in control. For immunoprecipitation, 0.3 $\mu$g of anti-HA antibody (Santa Cruz #sc-7392) or 3 $\mu$g of anti-SP1 antibody (Millipore #07-645) were used, mixed with the chromatin lysate and rotated overnight. Antibody-protein complexes were isolated with 50 $\mu$l protein G dynabeads (Thermo Fisher). DNA was purified with PCR purification kit (QIAGEN) by adding 100 $\mu$l water. Primers for quantitative PCR are listed in Dataset S3.

**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 $\mu$l 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 $\mu$l 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. Libraries were sequenced on an Illumina HiSeq 2500 and NextSeq 500 sequencing systems.

**ChIP-seq data analysis**

ChIP-seq data analysis was done as previously described (1). In brief, we mapped sequence reads to the custom reference genome built with human (hg38) and yeast (sacCer3) genome by the default parameters of Bowtie2 (--sensitive). Peak calling was performed using MACS2 with a p-value threshold of 0.01, and we extracted 14,436 peaks of which overlapping peak lengths between 60, 90 and 360 min samples are at least 50 bp. 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. This RRPM value was used as the SP1-ERT2 ChIP-seq signal intensity. For *k*-medoids clustering, we determined the optimal number of clusters by using factoextra package and performed *k*-medoids clustering with cluster package in R. Other publicly available data used in this study (2-8) are listed in Dataset S4.

**Defining categories of SP1 binding sites**

We consolidated the chromatin states defined by GenoSTAN (9) into five classes: Promoters, Prom.11, PromWF.5, and PromF.3; Enhancers, Enh.15, EnhWF.2, EnhF.10, ReprEnh.4; Gene bodies (GBs), Gen5'.13 and Elon'.14; Polycomb-repressed, ReprD.17, Repr.7 and ReprW.9; Others, Low.1, Low.8, Low.12 and Low.16. We excluded Elon.6 and Low.18

classes because no SP1 binding sites overlap these chromatin states. CpG islands locations are downloaded from UCSC table browser. CpG methylation levels in K562 cell line (RRBS_K562_11.05) was obtained from https://computational-epigenetics.de/compepiweb/reprogramming_leukemia/rnbeads_run_human/tracks_and_tables.html (reference doi: 10.1038/ncomms8091) and converted into hg38 coordinate by CrossMap program.

**Motif and Gene Ontology (GO) analysis**

Motif enrichment analysis between Faster peaks and Slow peaks was performed using AME (http://meme-suite.org/tools/ame) by setting each other as user-provided control sequences. We combined Fast peaks and Middle fast peaks and used them as Faster peaks because the number of Fast peaks was smaller than the other classes. The TF motifs from JASPAR2018_CORE_vertebrates_non-redundant database and of which ChIP-seq data are publicly available were chosen for the figures. For Gene Ontology analysis, we used CistromeGO (http://go.cistrome.org/) by changing the setting of peak number to use for using all the peaks.

**Data access**

All raw and processed sequencing data generated in this paper have been submitted to the NCBI Gene Expression Omnibus under accession number GSE162811.

**Supplementary References**


1. Hasegawa Y & Struhl K (2019) Promoter-specific dynamics of TATA-binding protein association with the human genome. *Genome Res*. **29**:1939-1950.

2. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

3. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, Cooper GM, Reddy TE, Crawford GE & Myers RM (2013) Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol*. *Cell* **52**: 25-36.

4. Neyret-Kahn H, Benhamed M, Ye T, Le Gras S, Cossec JC, Lapaquette P, Bischof O, Ouspenskaia M, Dasso M, Seeler J, Davidson I & Dejean A (2013) Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res*. **23**: 1563-1579.

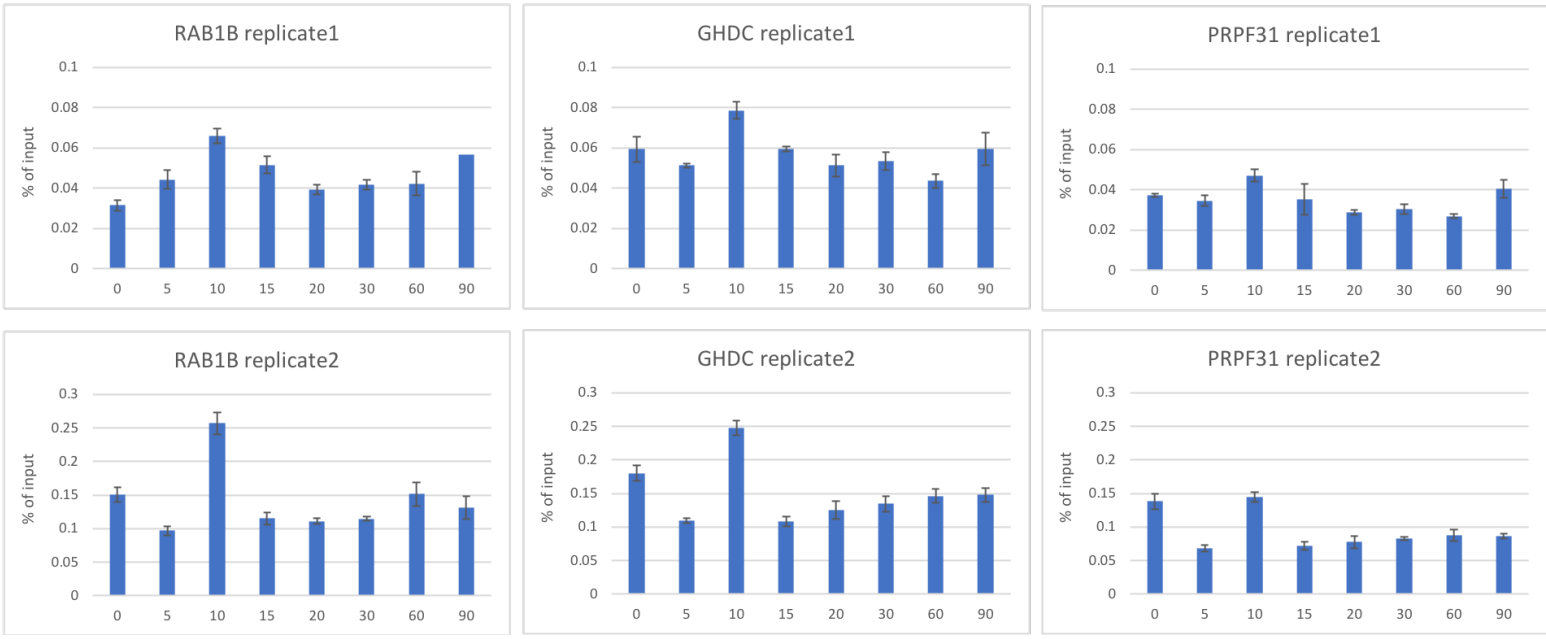
5. Thurman RE *et al* (2012) The accessible chromatin landscape of the human genome. *Nature* **489**: 75-82.

6. Natarajan A, Yardimci GG, Sheffield NC, Crawford GE & Ohler U (2012) Predicting cell-type–specific gene expression from regions of open chromatin. *Genome Res*. **22**: 1711-1722.

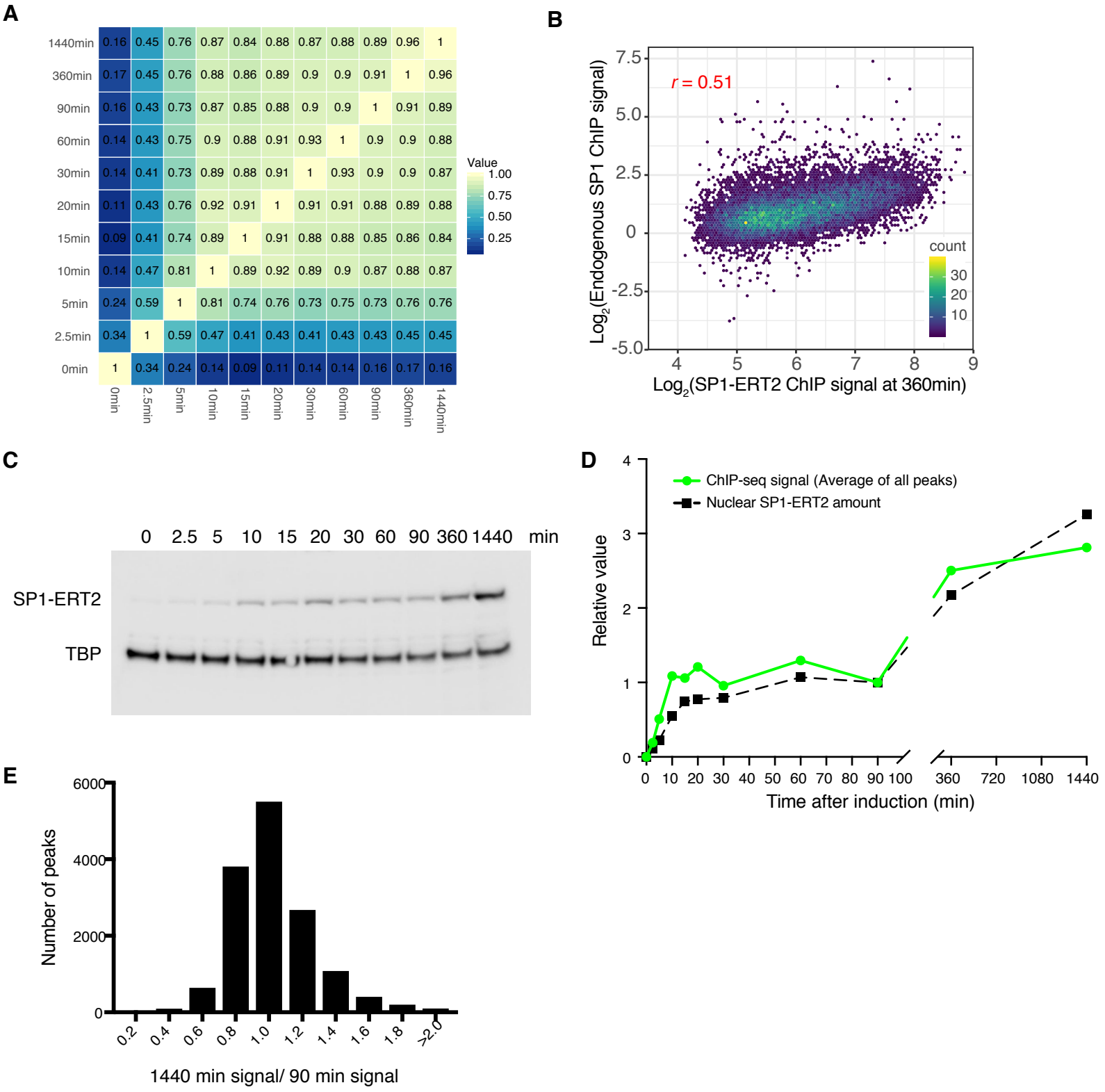
7. Karabacak Calviello A, Hirsekorn A, Wurmus R, Yusuf D & Ohler U (2019) Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol*. **20**: 1-13.

8. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A & Lis JT (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat*. *Genetics* **46**: 1311-1320.

9. Zacher B, *et al*. (2017) Accurate promoter and enhancer identification in 127 ENCODE and roadmap epigenomics cell types and tissues by GenoSTAN. *PLoS One* 12:e0169249.

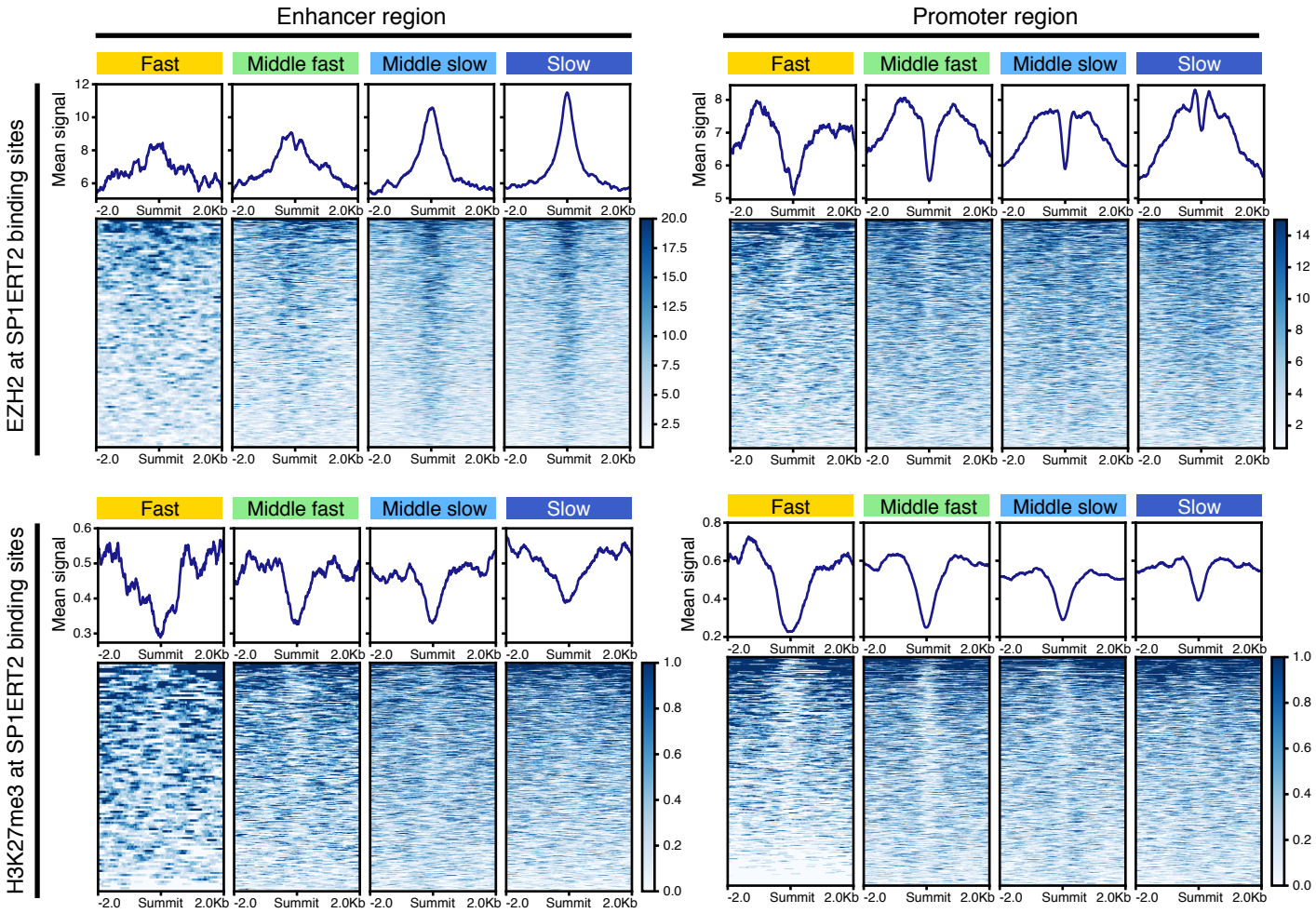



Supplemental Figure S1. Total SP1 ChIP signal over the time-course. Numbers on the x-axis represents time after tamoxifen treatment (min). Replicate means independent biological replicate. Error bar shows SD (n=3).

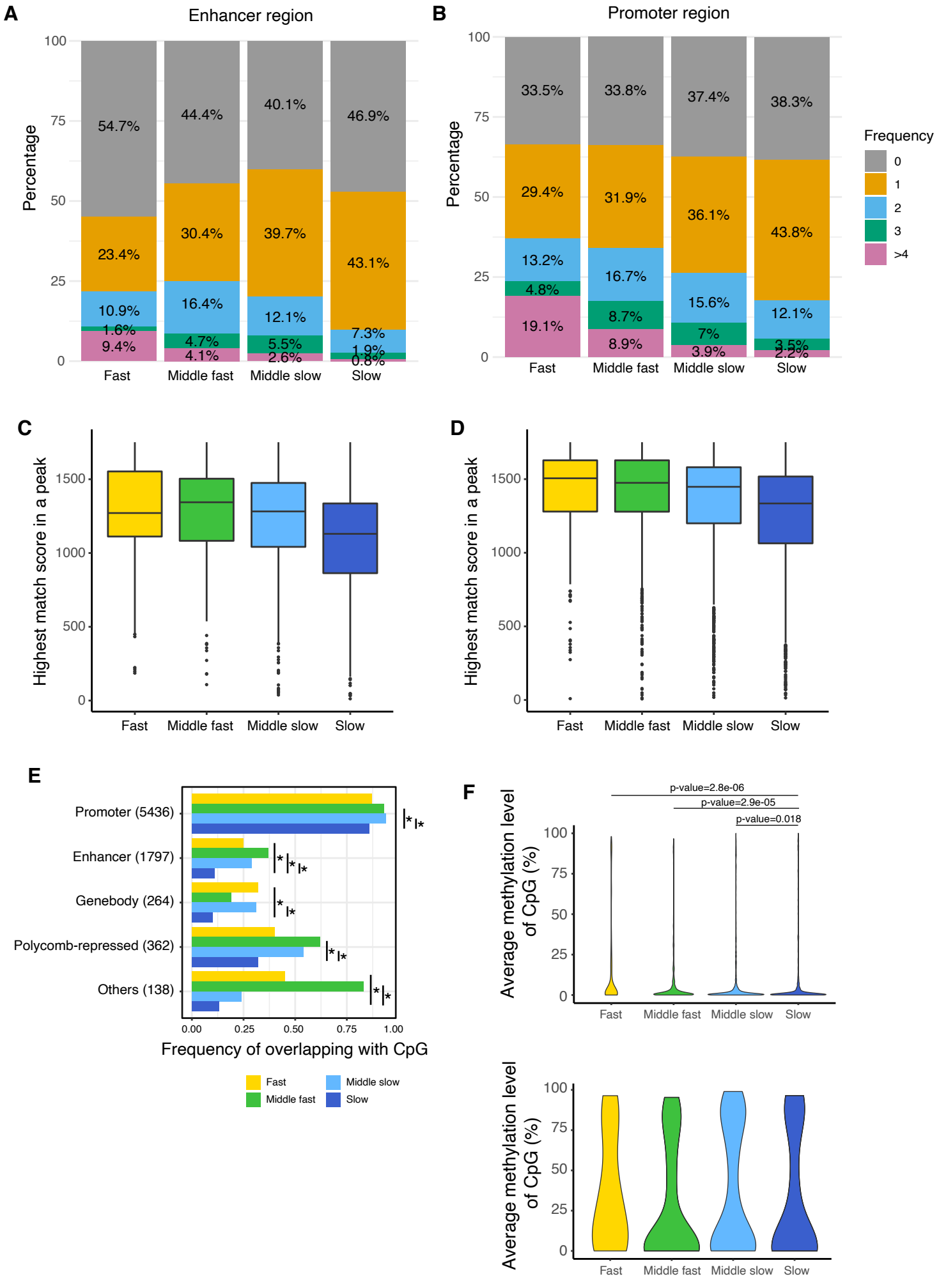



Supplemental Figure S2. Nuclear translocation of SP1-ERT2-3HA and time-course ChIP-seq.

(A) Spearman correlation coefficients between ChIP signal of each sample. (B) Correlation between SP1-ERT2 ChIP signal and endogenous SP1 ChIP signal. r is Spearman correlation coefficient. (C) Time-course western blotting analysis of SP1-ERT2-3HA detected with anti-HA antibody, and TBP (internal control) detected with anti-TBP antibody. (D) Relative amount of nuclear SP1-ERT2-3HA (dashed black) and average ChIP-seq signal of SP1-ERT2-3HA (green). (E) SP1 peak numbers in each bin. 90 min and 1440 min SP1 ChIP signals are normalized by SP1-ERT2 protein level.

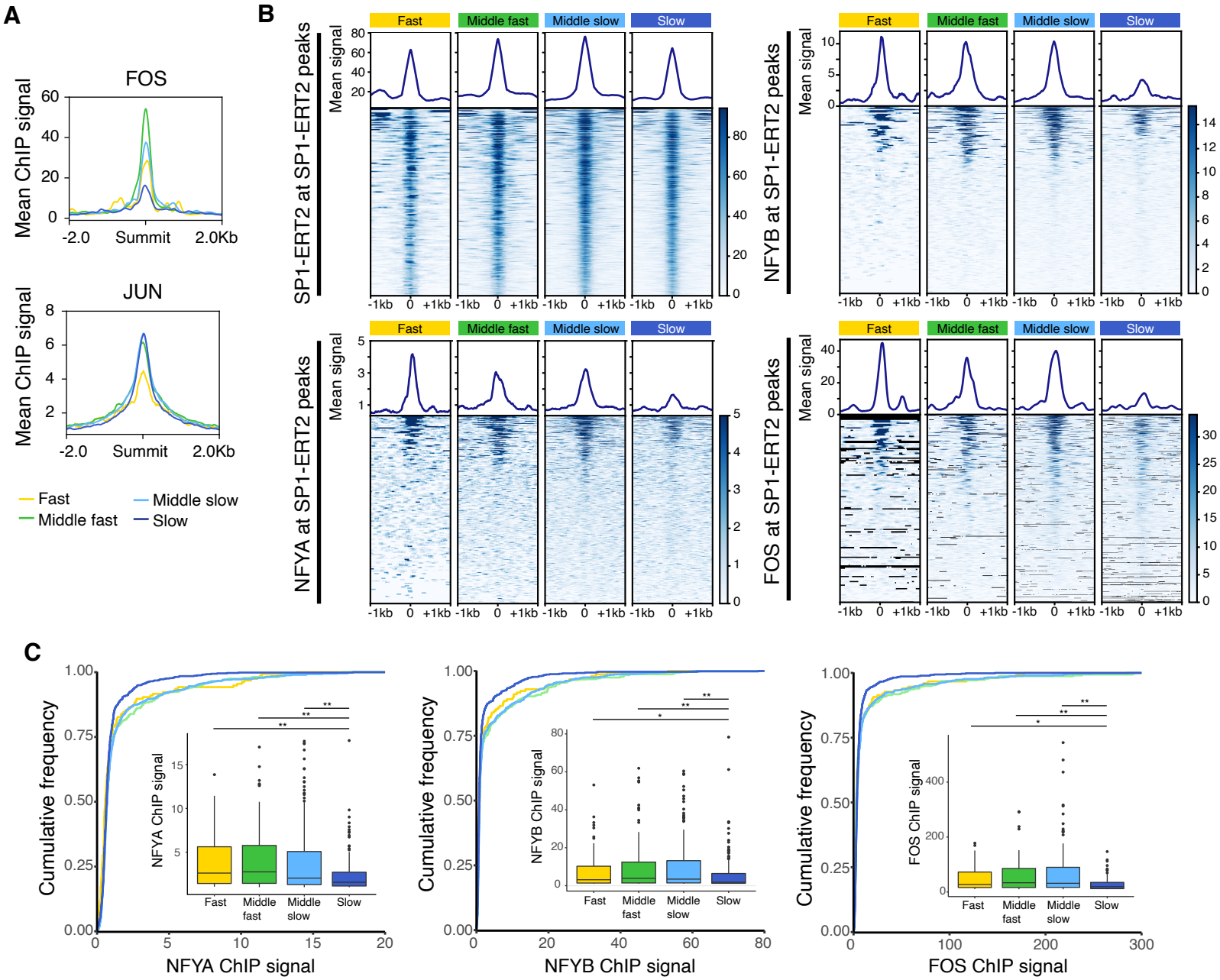



Supplemental Figure S3. EZH2 and H3K27me3 signal around SP1 binding sites.


A
Enhancer region
Percentage
100
75
50
25
0
54.7%
23.4%
10.9%
1.6%
9.4%
44.4%
30.4%
16.4%
4.7%
4.1%
40.1%
39.7%
12.1%
5.5%
2.6%
46.9%
43.1%
7.3%
1.9%
0.8%
Fast
Middle fast
Middle slow
Slow
B
Promoter region
Percentage
33.5%
29.4%
13.2%
4.8%
19.1%
33.8%
31.9%
16.7%
8.7%
8.9%
37.4%
36.1%
15.6%
7%
3.9%
38.3%
43.8%
12.1%
3.5%
2.2%
Frequency
0
1
2
3
>4
C
Highest match score in a peak
1500
1000
500
0
Fast
Middle fast
Middle slow
Slow
D
Highest match score in a peak
1500
1000
500
0
Fast
Middle fast
Middle slow
Slow
E
Promoter (5436)
Enhancer (1797)
Genebody (264)
Polycomb-repressed (362)
Others (138)
0.00
0.25
0.50
0.75
1.00
Frequency of overlapping with CpG
Fast
Middle fast
Middle slow
Slow
F
p-value=2.8e-06
p-value=2.9e-05
p-value=0.018
Average methylation level of CpG (%)
100
75
50
25
0
Fast
Middle fast
Middle slow
Slow
Average methylation level of CpG (%)
100
75
50
25
0
Fast
Middle fast
Middle slow
Slow

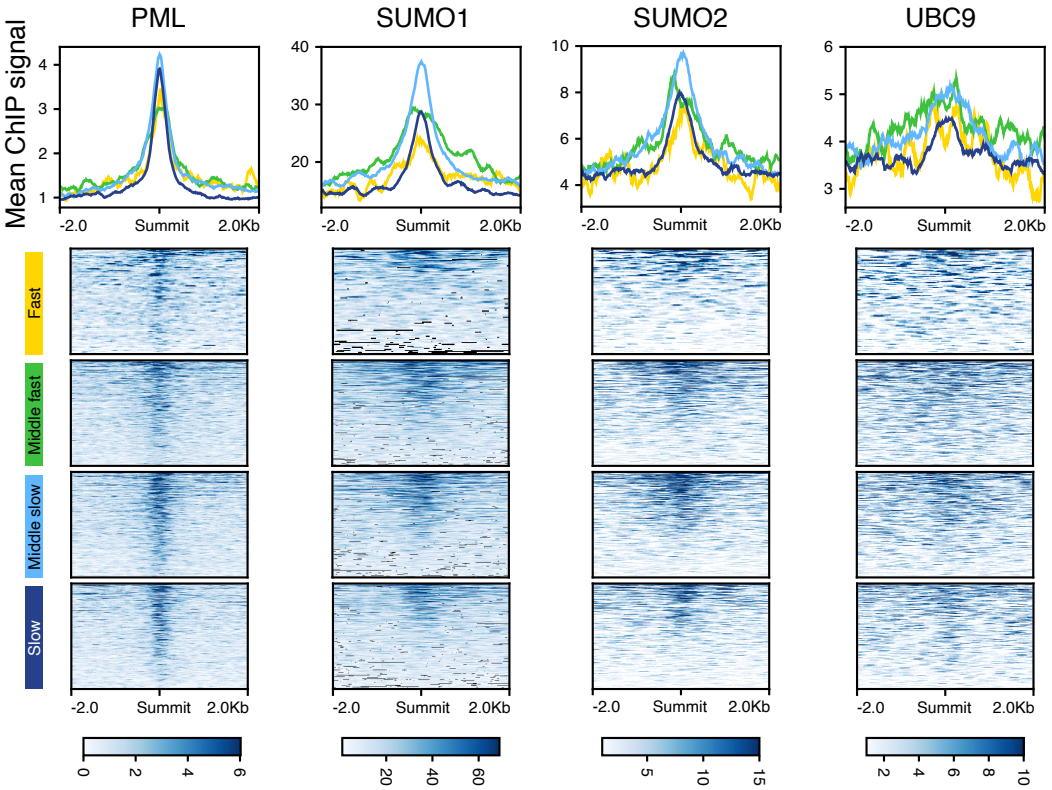
Supplemental Figure S4. SP1 consensus motif in SP1 binding sites in promoter and enhancer regions.

(A) Percentage of SP1 peaks in enhancer region that have strong hit sequences ($p < 0.0001$) to SP1 consensus motif, with colors indicating the number of SP1 motifs in each peak. (B) Percentage of SP1 peaks in promoter region that have strong hit sequences ($p < 0.0001$) to SP1 consensus motif, with colors indicating the number of SP1 motifs in each peak. (C) Box plot represents the highest match score to the position weight matrix of SP1 consensus motif in each peak located in enhancer region. (D) Box plot represents the highest match score to the position weight matrix of SP1 consensus motif in each peak located in promoter region. (E) Frequency of SP1 binding sites overlapping with UCSC CpG islands. asterisk (*) represents $p \leq 0.05$ in Fisher's exact test. (F) Average DNA methylation level of CpG islands overlapping with SP1 binding sites in promoter region (top) and enhancer region (bottom) in each class. P-values are results of Wilcoxon test.

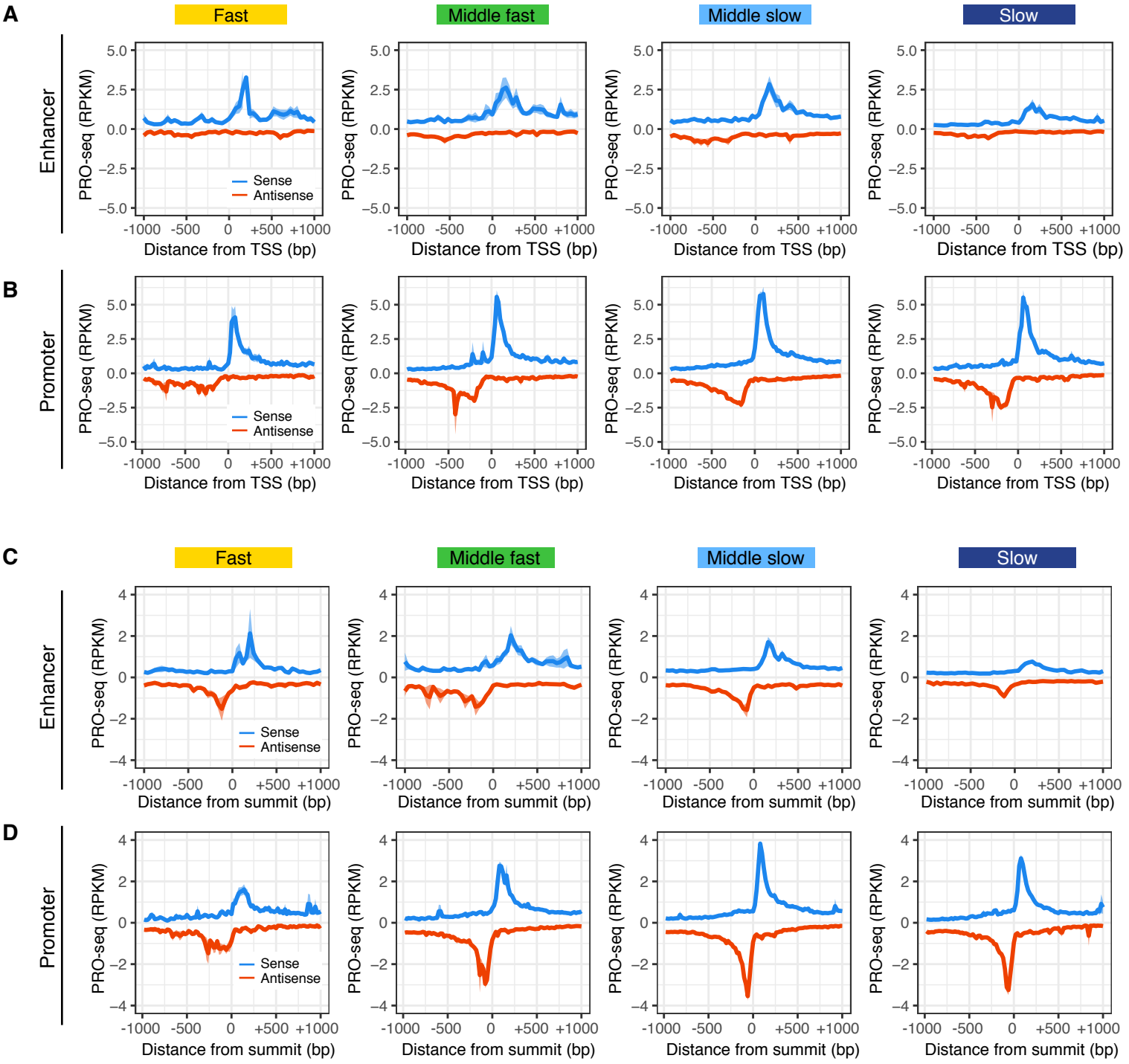



Supplemental Figure S5. NFY and FOS ChIP signals are higher in Fast class.

(A) Means and heatmap of FOS and JUN binding level at SP1 binding sites. (B) Means and heatmap of indicated transcription factors. (C) Box plots show ChIP signal intensity of the indicated factors between -0.5 kb to +0.5 kb from SP1-ERT2 peak summit in promoter region. Asterisk, $p < 0.05$; two asterisks, $p < 0.001$.




Supplemental Figure S6. PML, SUMO1, SUMO2, and UBC9 occupancy at SP1 binding sites in enhancer regions.




Supplemental Figure S7. Transcriptional activity of the nearest genes from SP1 binding sites and SP1 binding sites in each class.

(A)(B) PRO-seq signal (RPKM) of the nearest genes from SP1 binding sites in (A) enhancer regions and (B) promoter regions. (C)(D) PRO-seq signal (RPKM) at the SP1 binding sites in (C) enhancer regions and (D) promoter regions. Shaded areas represent mean ± SE.