Supplementary material for

RoboCOP: Jointly computing chromatin occupancy profiles for numerous factors from chromatin accessibility data

May 27, 2021

Abf1	Abf2	Ace2	Adr1	Aft1	Aft2
Aro80	Asg1	Azf1	Bas1	Cad1	Cat8
Cbf1	Cep3	Cha4	Cin5	Crz1	Cst6
Cup9	Dal80	Dal82	Ecm22	Ecm23	Fhl1
Fkh1	Fkh2	Fzf1	Gal4	Gat1	Gat3
Gat4	Gcn4	Gcr1	Gis1	Gln3	Gsm1
Gzf3	Hac1	Hal9	Hap1	Hcm1	Hmlalpha2
Hmra2	Hsf1	Leu3	Lys14	Matalpha2	Mbp1
Mcm1	Met31	Met32	Mga1	Mig1	Mig2
Mig3	Mot3	Msn1	Msn2	Msn4	Ndt80
Nhp10	Nhp6a	Nhp6b	Nrg1	Nrg2	Oaf1
Pbf1	Pbf2	Pdr1	Pdr3	Phd1	Pdr8
Pho2	Pho4	Put3	Rap1	Rdr1	Rds1
Rds2	Reb1	Rei1	Rfx1	Rgm1	Rgt1
Rim101	Rox1	Rph1	Rpn4	Rsc3	Rsc30
Rtg3	Sfl1	Sfp1	Sig1	Sip4	Skn7
Sko1	Smp1	Sok2	Spt15	Srd1	Stb3
Stb4	Stb5	Ste12	Stp1	Stp2	Stp3
Stp4	Sum1	Sut1	Sut2	Swi4	Swi5
Tbf1	Tbs1	Tea1	Tec1	Tos8	Tye7
Uga3	Ume6	Upc2	Usv1	Vhr1	Xbp1
Yap1	Yap3	Yap6	YBR033W	YBR239C	YDR520C
YER064C	YER130C	YER184C	YGR067C	YKL222C	YLL054C
YLR278C	YML081W	YNR063W	Yox1	YPR013C	YPR015C
YPR022C	YPR196W	Yrm1	Yrr1	Zap1	Zms1

Supplementary Table S1. List of the 150 TFs used by RoboCOP. The 69 TFs colored in gray have no binding sites reported by MacIsaac *et al.* (1), leaving us with 81 TFs when we validate binding sites. TFs colored in red have additional, newer validation sets from ChIP-exo (2) and ORGANIC (3) protocols. TFs colored in blue are known to bind the genome in some experimental condition and the binding is known to increase under some specific experimental conditions. These TFs are classified as conditionally-enabled by (4), and may not be expressed in our own data.

Model for predicting nucleosome positions	AUPR
RoboCOP$_{\text{MNase}}$	0.645
RoboCOP$_{\text{MNase+ATAC}}$	0.623
DANPOS2 (MNase)	0.518
RoboCOP$_{\text{ATAC}}$	0.437
COMPETE$_{\text{RoboCOP}}$	0.294
random_1	0.264
random_2	0.264
random_3	0.263
random_4	0.265
random_5	0.265
COMPETE$_{\text{16Kd}}$	0.197
COMPETE$_{\text{8Kd}}$	0.195
COMPETE$_{\text{Kd}}$	0.196
NucleoATAC	0.137

Supplementary Table S2. AUPR values that result from predicting nucleosome positions under different models, sorted by decreasing AUPR. Various RoboCOP and COMPETE models are compared with DANPOS2 applied to MNase data, and NucleoATAC applied to ATAC data. We have inserted into the list results from five trials of applying the greedy algorithm to random genomic locations. Note that COMPETE$_{\text{RoboCOP}}$-predicted nucleosome positions are only marginally better than randomly predicted nucleosome positions, and when COMPETE doesn't have access RoboCOP-learned transition probabilities, its predictions are worse. NucleoATAC's low AUPR is because it has low recall, reporting only 13,344 nucleosomes of the 67,543 that are present in the reference set; among the nucleosomes it reports, its precision is quite good (see Fig. S12).

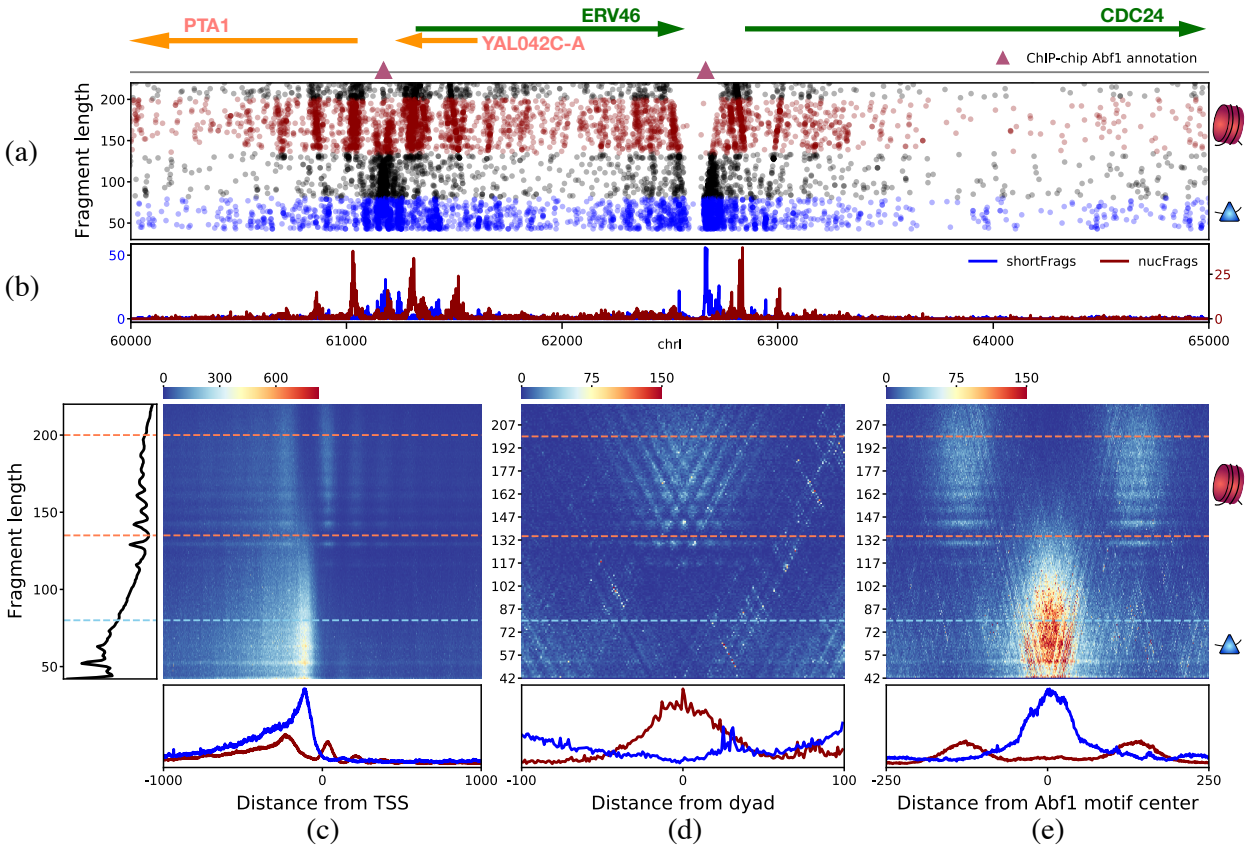
Model with higher TF AUPRs	Model with lower TF AUPRs	Significance
RoboCOP$_{\text{MNase}}$	RoboCOP$_{\text{MNase+ATAC}}$	0.22
RoboCOP$_{\text{MNase}}$	RoboCOP$_{\text{ATAC}}$	0.22
RoboCOP$_{\text{MNase}}$	COMPETE$_{\text{RoboCOP}}$	0.14
RoboCOP$_{\text{MNase}}$	COMPETE$_{\text{Kd}}$	0.15
RoboCOP$_{\text{MNase}}$	COMPETE$_{\text{8Kd}}$	0.11
RoboCOP$_{\text{MNase}}$	COMPETE$_{\text{16Kd}}$	0.11
RoboCOP$_{\text{MNase}}$	**FIMO-MNase**	**0.03***
RoboCOP$_{\text{MNase}}$	**FIMO-ATAC**	**0.001***
RoboCOP$_{\text{MNase+ATAC}}$	RoboCOP$_{\text{ATAC}}$	0.48
RoboCOP$_{\text{MNase+ATAC}}$	COMPETE$_{\text{RoboCOP}}$	0.38
RoboCOP$_{\text{MNase+ATAC}}$	COMPETE$_{\text{Kd}}$	0.36
RoboCOP$_{\text{MNase+ATAC}}$	COMPETE$_{\text{8Kd}}$	0.31
RoboCOP$_{\text{MNase+ATAC}}$	COMPETE$_{\text{16Kd}}$	0.31
RoboCOP$_{\text{MNase+ATAC}}$	FIMO-MNase	0.07
RoboCOP$_{\text{MNase+ATAC}}$	**FIMO-ATAC**	**0.01***
RoboCOP$_{\text{ATAC}}$	COMPETE$_{\text{RoboCOP}}$	0.38
RoboCOP$_{\text{ATAC}}$	COMPETE$_{\text{Kd}}$	0.34
RoboCOP$_{\text{ATAC}}$	COMPETE$_{\text{8Kd}}$	0.30
RoboCOP$_{\text{ATAC}}$	COMPETE$_{\text{16Kd}}$	0.30
RoboCOP$_{\text{ATAC}}$	FIMO-MNase	0.07
RoboCOP$_{\text{ATAC}}$	**FIMO-ATAC**	**0.006***
COMPETE$_{\text{RoboCOP}}$	COMPETE$_{\text{Kd}}$	0.44
COMPETE$_{\text{RoboCOP}}$	COMPETE$_{\text{8Kd}}$	0.40
COMPETE$_{\text{RoboCOP}}$	COMPETE$_{\text{16Kd}}$	0.40
COMPETE$_{\text{RoboCOP}}$	FIMO-MNase	0.11
COMPETE$_{\text{RoboCOP}}$	**FIMO-ATAC**	**0.01***
COMPETE$_{\text{Kd}}$	COMPETE$_{\text{8Kd}}$	0.45
COMPETE$_{\text{Kd}}$	COMPETE$_{\text{16Kd}}$	0.45
COMPETE$_{\text{Kd}}$	FIMO-MNase	0.14
COMPETE$_{\text{Kd}}$	**FIMO-ATAC**	**0.01***
COMPETE$_{\text{8Kd}}$	COMPETE$_{\text{16Kd}}$	0.48
COMPETE$_{\text{8Kd}}$	FIMO-MNase	0.15
COMPETE$_{\text{8Kd}}$	**FIMO-ATAC**	**0.01***
COMPETE$_{\text{16Kd}}$	FIMO-MNase	0.15
COMPETE$_{\text{16Kd}}$	**FIMO-ATAC**	**0.01***
FIMO-MNase	FIMO-ATAC	0.22

Supplementary Table S3. Pairwise comparison of the sets of 81 AUPR values that result from validating TF binding site predictions under different models. Various RoboCOP and COMPETE models are compared with FIMO-MNase and FIMO-ATAC. In each row, AUPR values for the model in the first column are higher than those of the model in the second column, and the significance of this comparison is shown the third column. Significance is computed as the p-value of a Mann-Whitney U test over the 81 AUPR values for the two models.

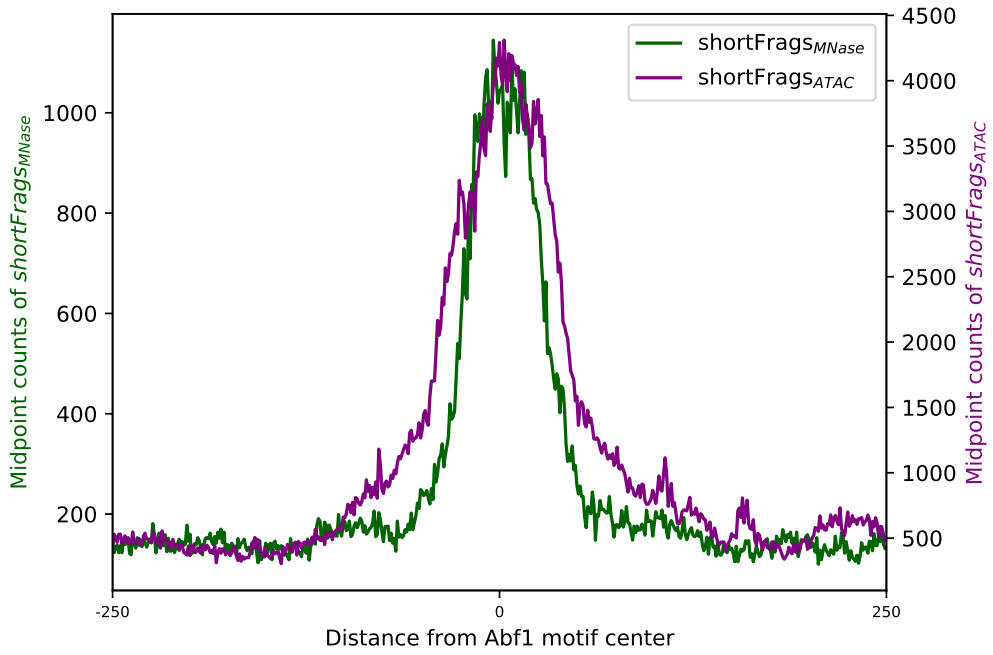
upmost 100 genes	downmost 100 genes	constant 100 genes
HSP12 (6.30)	RPL9A (−4.59)	AIF1 (0.00)
MET3 (5.87)	RPS22A (−4.51)	YMR272W-B (0.00)
PDC6 (5.69)	SSB1 (−4.43)	YBR013C (0.00)
YDL124W (5.59)	RPS31 (−4.42)	YNR070W (0.00)
HSP26 (5.41)	RPS3 (−4.36)	COS12 (0.00)
MET14 (5.41)	ASC1 (−4.34)	UBS1 (0.00)
DDR48 (5.15)	RPL3 (−4.34)	IPL1 (0.00)
AHP1 (4.99)	YEF3 (−4.31)	SEN54 (0.00)
MET10 (4.85)	RPL8B (−4.30)	OSW1 (0.00)
STF2 (4.70)	RPS9B (−4.27)	YJL043W (0.00)
GRE1 (4.69)	RPL5 (−4.25)	FIG1 (0.00)
SPG4 (4.60)	RPS5 (−4.20)	DYN3 (0.00)
ECM17 (4.46)	RPP2B (−4.16)	BCK1 (0.00)
PRB1 (4.26)	RPS12 (−4.11)	CTF3 (0.00)
CYS3 (4.17)	RPL12A (−4.11)	IST2 (0.00)
PNC1 (4.09)	RPS2 (−4.01)	VEL1 (0.00)
YCT1 (3.99)	RPP2A (−3.97)	YLR406C-A (0.00)
BDS1 (3.97)	RPP0 (−3.96)	SMA2 (0.00)
PBI2 (3.94)	RPP1A (−3.86)	MEI5 (0.00)
ZWF1 (3.92)	RPS15 (−3.82)	MF(ALPHA)1 (0.00)
TMA10 (3.92)	RPS0B (−3.81)	MMS1 (0.00)
RTC3 (3.90)	RPL28 (−3.80)	ECM12 (0.00)
TIS11 (3.84)	RPL38 (−3.79)	HPR1 (0.00)
KAR2 (3.79)	RPS14A (−3.75)	PEX25 (0.00)
BNA3 (3.78)	ILV5 (−3.75)	NPL6 (0.00)
HSP42 (3.71)	RPL16B (−3.72)	YPL009C (0.00)
UBI4 (3.68)	RPL22A (−3.71)	ESC8 (0.00)
TSA1 (3.67)	RPL2A (−3.70)	RRN5 (0.00)
PEP4 (3.67)	RPL1B (−3.68)	YHR214C-E (0.00)
LAP4 (3.67)	RPL2B (−3.67)	GET1 (0.00)
MET16 (3.60)	RPL32 (−3.64)	SWI3 (0.00)
ICY2 (3.59)	SSB2 (−3.64)	GNT1 (0.00)
PHO89 (3.58)	RPL1A (−3.64)	YOL019W-A (0.00)
YNL208W (3.54)	RPS16B (−3.64)	HIS3 (0.00)
MET6 (3.53)	RPL18A (−3.63)	GPI15 (0.00)
SEO1 (3.52)	RPL25 (−3.61)	KAR5 (0.00)
FIT3 (3.48)	RPS13 (−3.61)	SPT23 (0.00)
AGP3 (3.42)	RPL11A (−3.60)	SPT10 (0.00)
HSP104 (3.39)	RPL21A (−3.60)	SPR28 (0.00)
TDH1 (3.37)	RPL26B (−3.58)	SOG2 (0.00)
HBN1 (3.34)	RPL17A (−3.58)	DCG1 (0.00)
FIT2 (3.31)	RPL15A (−3.57)	STH1 (0.00)
GLK1 (3.30)	RPL30 (−3.57)	THP1 (0.00)
MET1 (3.30)	RPS11A (−3.57)	HST4 (0.00)
SSA1 (3.29)	RPS8A (−3.54)	EST1 (0.00)
AIM17 (3.19)	RPS1B (−3.52)	YKR041W (0.00)
YNL134C (3.17)	RPS4A (−3.52)	MRPS5 (0.00)
GRE3 (3.11)	RPS26B (−3.51)	OST1 (0.00)
PGK1 (3.11)	RPL19B (−3.51)	YDL133W (0.00)
HSP82 (3.11)	RPL31A (−3.49)	POP1 (0.00)
TSL1 (3.11)	RPS7A (−3.49)	SED5 (0.00)
ERO1 (3.09)	SNU13 (−3.47)	MSH4 (0.00) *cont.*

upmost 100 genes	downmost 100 genes	constant 100 genes
PRX1 (3.07)	RPS4B (−3.46)	CCZ1 (0.00)
YMR315W (3.04)	RPS20 (−3.44)	SPT8 (0.00)
TPS1 (3.03)	RPL42B (−3.44)	MDM1 (0.00)
TFS1 (3.00)	RPL27A (−3.42)	HMI1 (0.00)
MXR1 (2.99)	RPS18B (−3.42)	VID24 (0.00)
PDI1 (2.99)	RPS0A (−3.39)	STB6 (0.00)
SAM2 (2.99)	RPL12B (−3.37)	CDC9 (0.00)
TRX2 (2.97)	RPS10A (−3.37)	YOR387C (0.00)
YBR085C-A (2.97)	RPL13B (−3.36)	SPC110 (0.00)
MRP8 (2.97)	RPL6B (−3.36)	STE5 (0.00)
FMO1 (2.95)	RPL8A (−3.35)	RTC4 (0.00)
MET2 (2.95)	RPL24A (−3.34)	TOP1 (0.00)
YOR285W (2.94)	NSR1 (−3.33)	YCR090C (0.00)
MUP3 (2.93)	RPS24B (−3.30)	YPR027C (0.00)
GSP2 (2.92)	RPS18A (−3.30)	PSR2 (0.00)
ALD4 (2.92)	RPS6B (−3.30)	BUB1 (0.01)
GRX1 (2.90)	RPL11B (−3.29)	YOR072W-B (0.01)
CMK2 (2.88)	RPS6A (−3.28)	LDB16 (0.01)
ENO1 (2.88)	RPL6A (−3.26)	YDR042C (0.01)
YLR194C (2.86)	RPL20A (−3.23)	ATS1 (0.01)
SER33 (2.86)	RPS24A (−3.22)	LOH1 (0.01)
ARA1 (2.85)	RPL43A (−3.22)	TAF13 (0.01)
GRE2 (2.84)	RPS8B (−3.21)	YER188C-A (0.01)
GPD1 (2.83)	RPL33A (−3.21)	MSI1 (0.01)
MUP1 (2.82)	NOP1 (−3.20)	YDR194W-A (0.01)
HOR7 (2.82)	RPL4A (−3.19)	RUD3 (0.01)
DDR2 (2.82)	RPL14A (−3.17)	YHR100C (0.01)
GPM1 (2.80)	RPS19B (−3.17)	YOR293C-A (0.01)
MET22 (2.79)	TMA19 (−3.17)	SIN3 (0.01)
WTM1 (2.69)	RPL16A (−3.17)	FYV12 (0.01)
ARN2 (2.68)	RPL37A (−3.16)	TRP3 (0.01)
RTS3 (2.66)	RPL19A (−3.16)	SMK1 (0.01)
CPR1 (2.63)	RPL7A (−3.16)	PFD1 (0.01)
YER067W (2.62)	RPS17A (−3.15)	YGR146C-A (0.01)
GSH1 (2.61)	RPL17B (−3.14)	KAR9 (0.01)
PGM2 (2.61)	RPS23B (−3.12)	FIG2 (0.01)
GSC2 (2.61)	RPS27B (−3.11)	PGA1 (0.01)
PRC1 (2.60)	RPL20B (−3.11)	YKR051W (0.01)
MET8 (2.56)	RPS11B (−3.10)	SMC2 (0.01)
HOM6 (2.56)	RPL23A (−3.10)	SNF6 (0.01)
STF1 (2.55)	RPS28A (−3.08)	YLR412C-A (0.01)
GOR1 (2.55)	STM1 (−3.07)	FUS1 (0.01)
HNT1 (2.54)	RPL40B (−3.05)	MTW1 (0.01)
YET1 (2.52)	RPL34B (−3.05)	CNA1 (0.01)
YHR138C (2.52)	RPP1B (−3.04)	CDC12 (0.01)
MMP1 (2.51)	RPL39 (−3.03)	MSC7 (0.01)
HSP78 (2.51)	ILV3 (−3.02)	SMC6 (0.01)
HXK1 (2.51)	RPL23B (−2.99)	YMC1 (0.01)

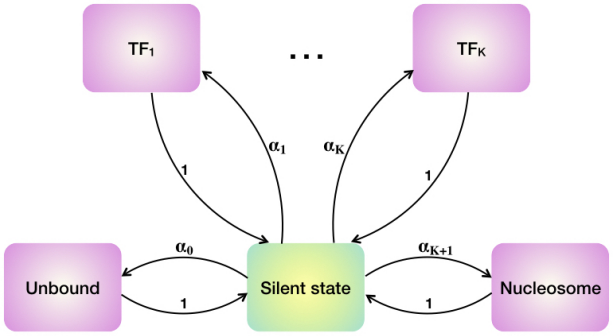
Supplementary Table S4. Lists of 100 most up-regulated, 100 most down-regulated, and 100 most constant genes after 60 minutes of cadmium treatment. Upmost and downmost are sorted with greatest change first (log-fold expression change in parentheses). Most constant are sorted with least absolute change first (absolute log-fold expression change in parentheses).



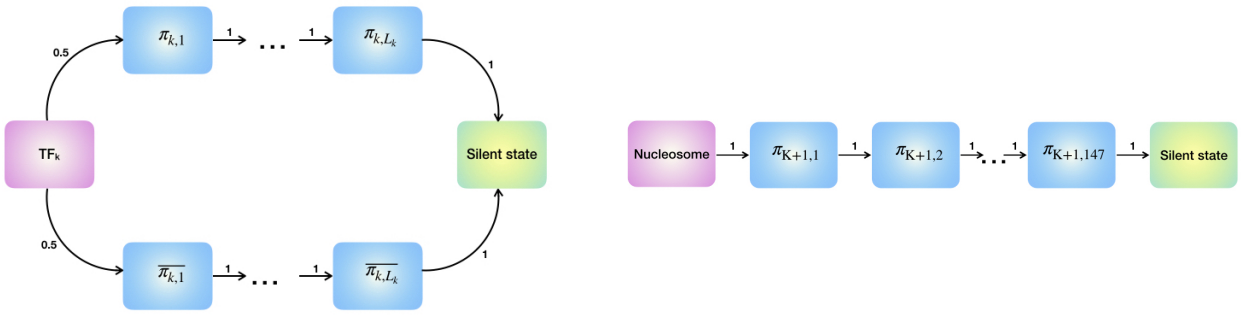
Supplementary Figure S1. Paired-end ATAC-seq data is informative about the binding of both nucleosomes and smaller DBFs, such as transcription factors, but primarily in open chromatin regions of the genome, which in yeast is mostly in and around promoter sequences (see also Fig. S11c). (a) Two-dimensional plot of ATAC-seq fragments from positions 60,000 to 65,000 of yeast chromosome I. Each fragment is plotted based on its length (y-axis) and the genomic location of its midpoint (x-axis). Nucleosome-sized fragments (`nucFrags`, length 135–200) are colored red, while shorter fragments corresponding to smaller proteins (`shortFrags`, length ≤ 80) are colored blue. Above the plot are genomic annotations for this region, with Watson strand genes depicted as green arrows and Crick strand genes as orange. Below the gene annotations, known TF binding sites (1) are indicated using triangles. This region contains two annotated binding sites for Abf1 (pink). (b) Aggregate numbers of red and blue dots at each genomic position in (a), resulting in the one-dimensional `nucFrags` and `shortFrags` signals, respectively. (c) Composite heatmap of ATAC-seq fragments around all yeast genes, centered on each gene's TSS. Panels along the side and bottom show marginal densities. The heatmap and the side panel shows bands of fragments ranging in length between 125–200 in nucleosome occupied regions. The aggregate of `nucFrags` (red) in the bottom panel has peaks on each side of the promoter region. The promoter region is enriched with `shortFrags` (blue curve in bottom panel). The heatmap reveals many fragments in the promoter region but decreases within gene bodies. (d) Composite heatmap of ATAC-seq fragments around the 2,000 most well-positioned nucleosomes in the yeast genome (5), centered on each nucleosome's dyad. The `nucFrags` signal peaks at the dyad and decreases roughly symmetrically in either direction. (e) Composite heatmap of ATAC-seq fragments around all annotated Abf1 binding sites (1) in the yeast genome, centered on each site's motif. Note the clear enrichment of `shortFrags` near Abf1 sites.



Supplementary Figure S2. Midpoint counts of MNase-seq and ATAC-seq `shortFrags` around annotated Abf1 binding sites. The MNase-seq signal is slightly sharper and more spatially precise than the ATAC-seq signal around these annotated binding sites.

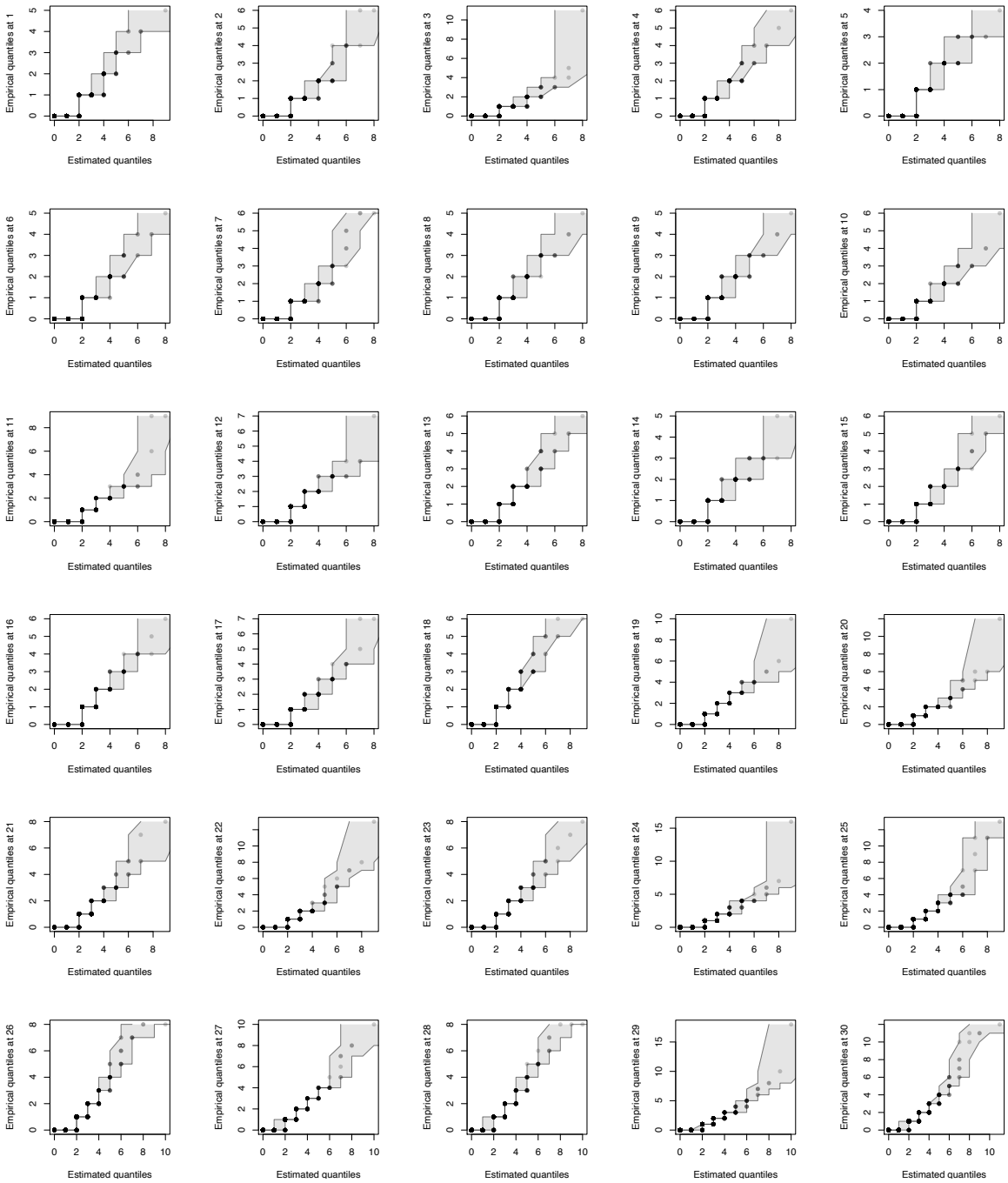


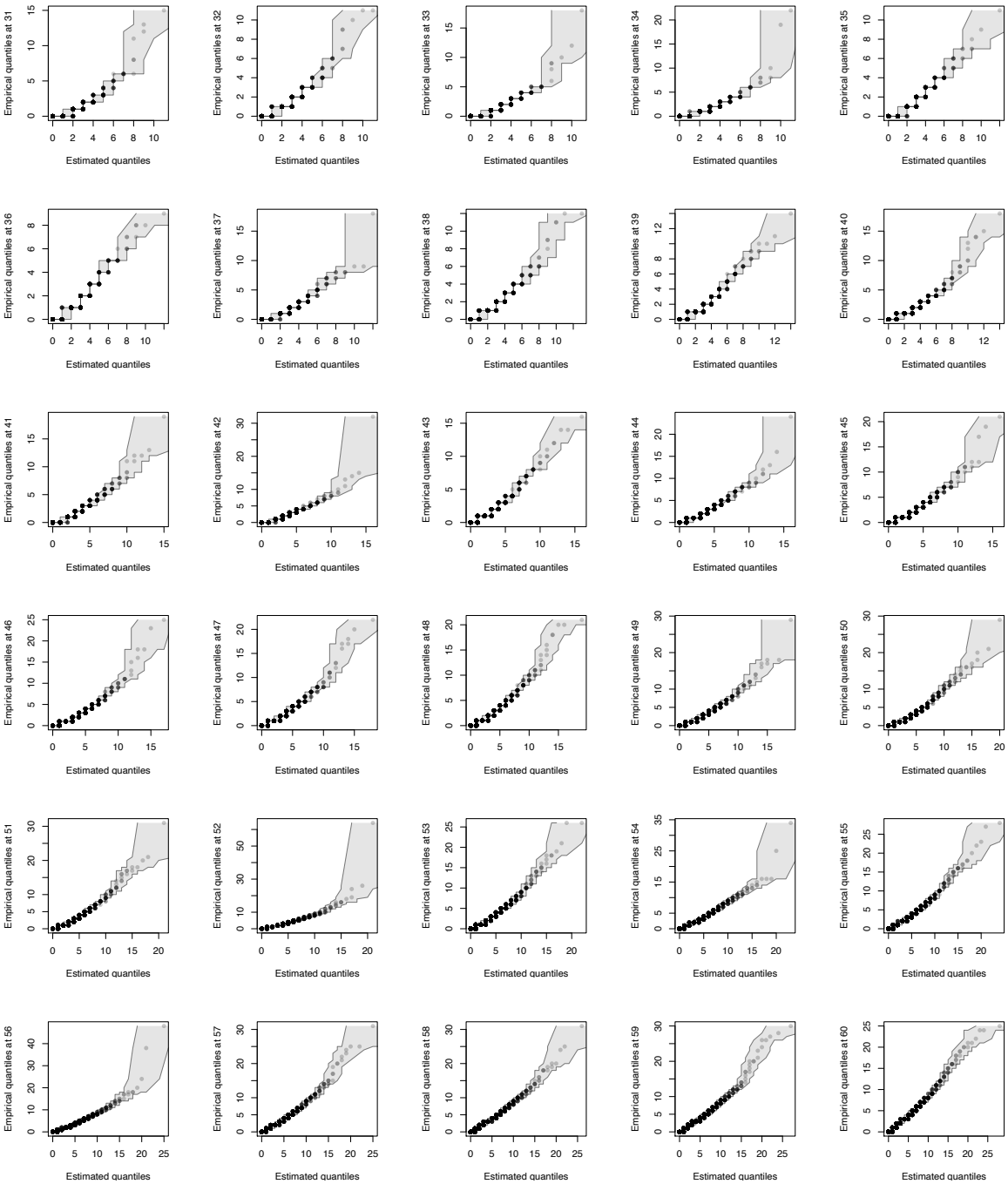
Supplementary Figure S3. Transition diagram for the different DBFs in RoboCOP: a nucleosome, a collection of K distinct TFs, and an unbound nucleotide. From the silent state, the model can transition to any DBF (pink), denoted π_k, with probability α_k. After traversing through that DBF, the model then comes back to the silent state (green) with probability 1. From the silent state, it can then transition to another DBF. Transitions among the DBFs are therefore independent of one another.

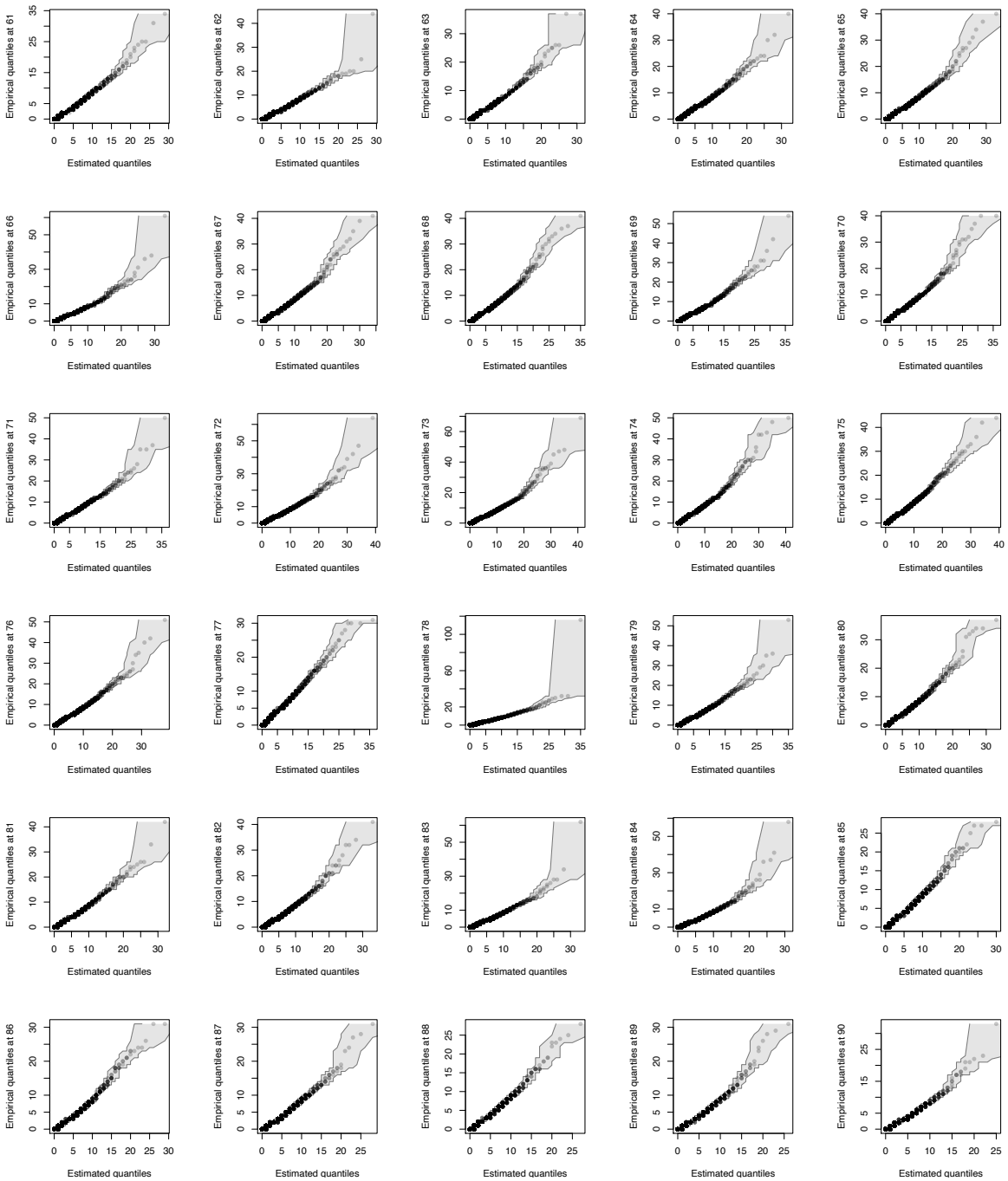


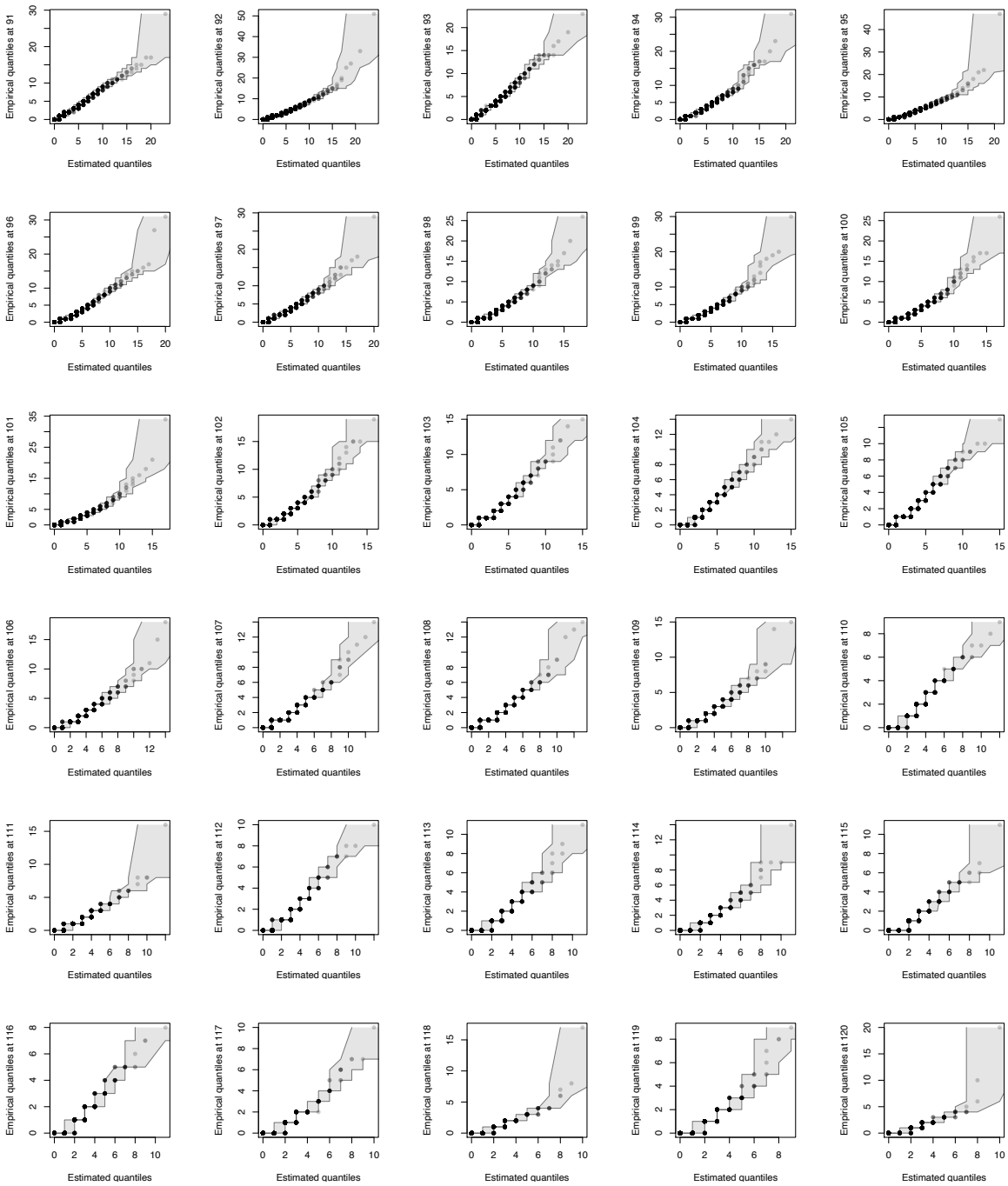
(a) (b)

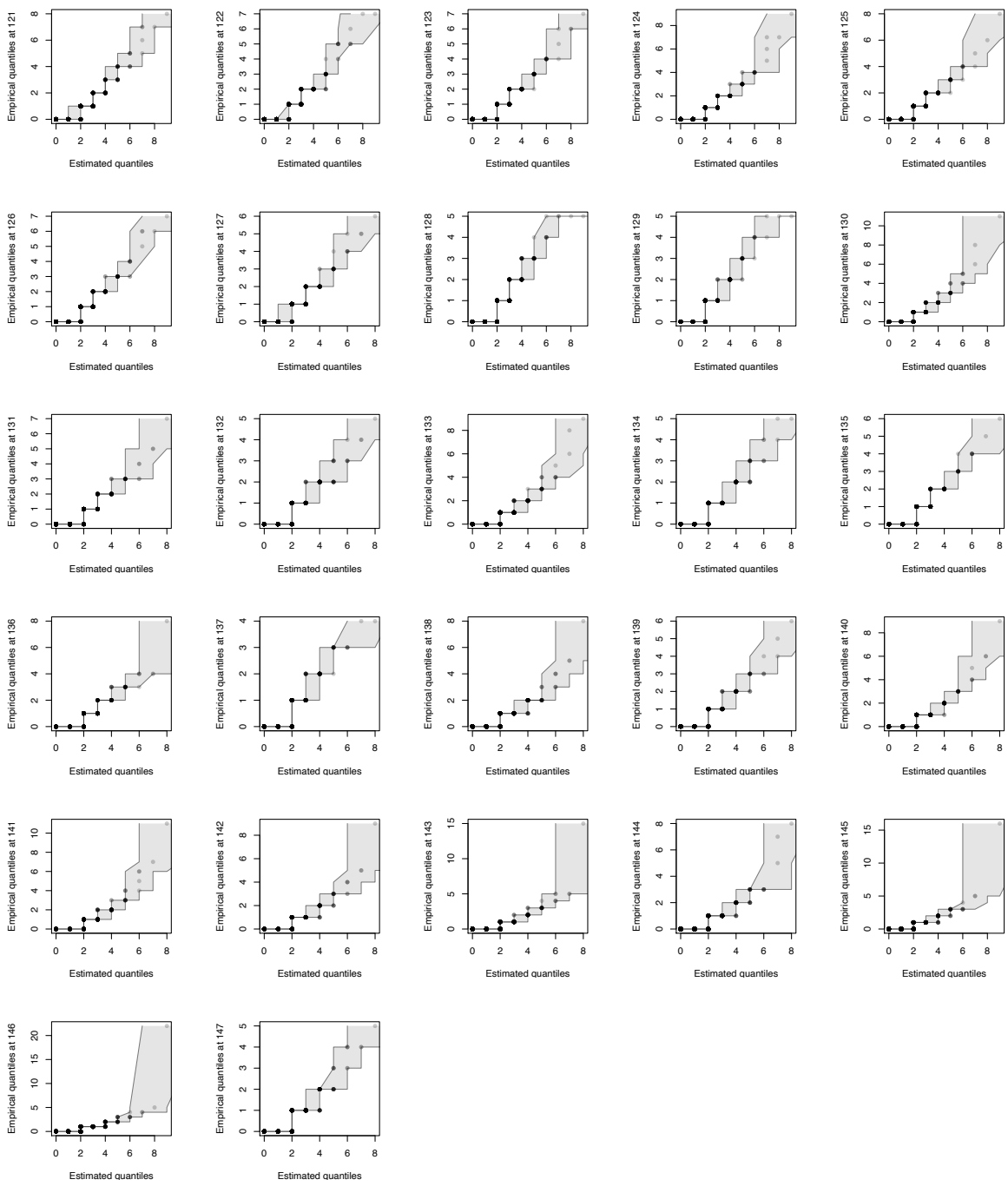
Supplementary Figure S4. (a) State transition diagram for a TF, π_k for $k \in \{1, \ldots, K\}$. From the first hidden silent state (pink) of the TF, the model can transition to either the forward motif (top) or the reverse motif (bottom) with equal probability of 0.5. It then transitions through the entire length, L_k, of the motif. After transitioning through the entire motif, the model transitions back to the central silent state (green) with probability 1. (b) State transition diagram for a nucleosome, π_{K+1}. Nucleosomes are modeled to be 147 bases long. After transitioning through the 147 nucleosome positions, the model transitions back to the central silent state with probability 1.


Empirical quantiles at 1
Empirical quantiles at 2
Empirical quantiles at 3
Empirical quantiles at 4
Empirical quantiles at 5
Empirical quantiles at 6
Empirical quantiles at 7
Empirical quantiles at 8
Empirical quantiles at 9
Empirical quantiles at 10
Empirical quantiles at 11
Empirical quantiles at 12
Empirical quantiles at 13
Empirical quantiles at 14
Empirical quantiles at 15
Empirical quantiles at 16
Empirical quantiles at 17
Empirical quantiles at 18
Empirical quantiles at 19
Empirical quantiles at 20
Empirical quantiles at 21
Empirical quantiles at 22
Empirical quantiles at 23
Empirical quantiles at 24
Empirical quantiles at 25
Empirical quantiles at 26
Empirical quantiles at 27
Empirical quantiles at 28
Empirical quantiles at 29
Empirical quantiles at 30
Estimated quantiles

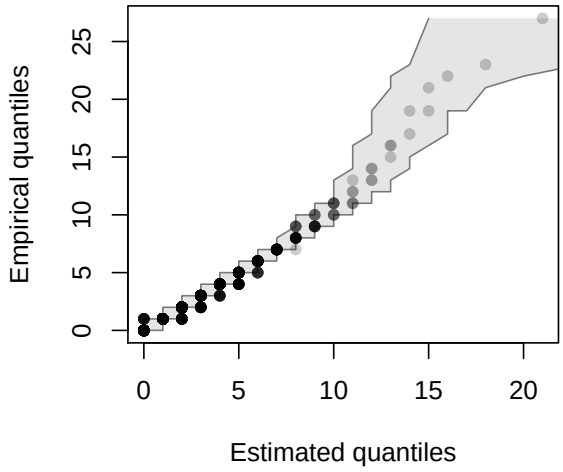

Empirical quantiles at 31
Empirical quantiles at 32
Empirical quantiles at 33
Empirical quantiles at 34
Empirical quantiles at 35
Empirical quantiles at 36
Empirical quantiles at 37
Empirical quantiles at 38
Empirical quantiles at 39
Empirical quantiles at 40
Empirical quantiles at 41
Empirical quantiles at 42
Empirical quantiles at 43
Empirical quantiles at 44
Empirical quantiles at 45
Empirical quantiles at 46
Empirical quantiles at 47
Empirical quantiles at 48
Empirical quantiles at 49
Empirical quantiles at 50
Empirical quantiles at 51
Empirical quantiles at 52
Empirical quantiles at 53
Empirical quantiles at 54
Empirical quantiles at 55
Empirical quantiles at 56
Empirical quantiles at 57
Empirical quantiles at 58
Empirical quantiles at 59
Empirical quantiles at 60
Estimated quantiles


Empirical quantiles at 61
Empirical quantiles at 62
Empirical quantiles at 63
Empirical quantiles at 64
Empirical quantiles at 65
Empirical quantiles at 66
Empirical quantiles at 67
Empirical quantiles at 68
Empirical quantiles at 69
Empirical quantiles at 70
Empirical quantiles at 71
Empirical quantiles at 72
Empirical quantiles at 73
Empirical quantiles at 74
Empirical quantiles at 75
Empirical quantiles at 76
Empirical quantiles at 77
Empirical quantiles at 78
Empirical quantiles at 79
Empirical quantiles at 80
Empirical quantiles at 81
Empirical quantiles at 82
Empirical quantiles at 83
Empirical quantiles at 84
Empirical quantiles at 85
Empirical quantiles at 86
Empirical quantiles at 87
Empirical quantiles at 88
Empirical quantiles at 89
Empirical quantiles at 90
Estimated quantiles

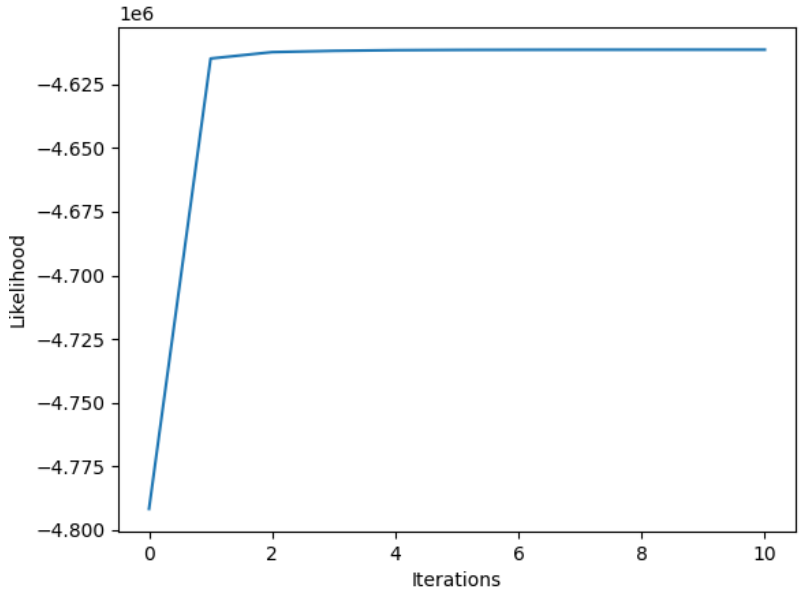

Empirical quantiles at 91
Empirical quantiles at 92
Empirical quantiles at 93
Empirical quantiles at 94
Empirical quantiles at 95
Empirical quantiles at 96
Empirical quantiles at 97
Empirical quantiles at 98
Empirical quantiles at 99
Empirical quantiles at 100
Empirical quantiles at 101
Empirical quantiles at 102
Empirical quantiles at 103
Empirical quantiles at 104
Empirical quantiles at 105
Empirical quantiles at 106
Empirical quantiles at 107
Empirical quantiles at 108
Empirical quantiles at 109
Empirical quantiles at 110
Empirical quantiles at 111
Empirical quantiles at 112
Empirical quantiles at 113
Empirical quantiles at 114
Empirical quantiles at 115
Empirical quantiles at 116
Empirical quantiles at 117
Empirical quantiles at 118
Empirical quantiles at 119
Empirical quantiles at 120
Estimated quantiles



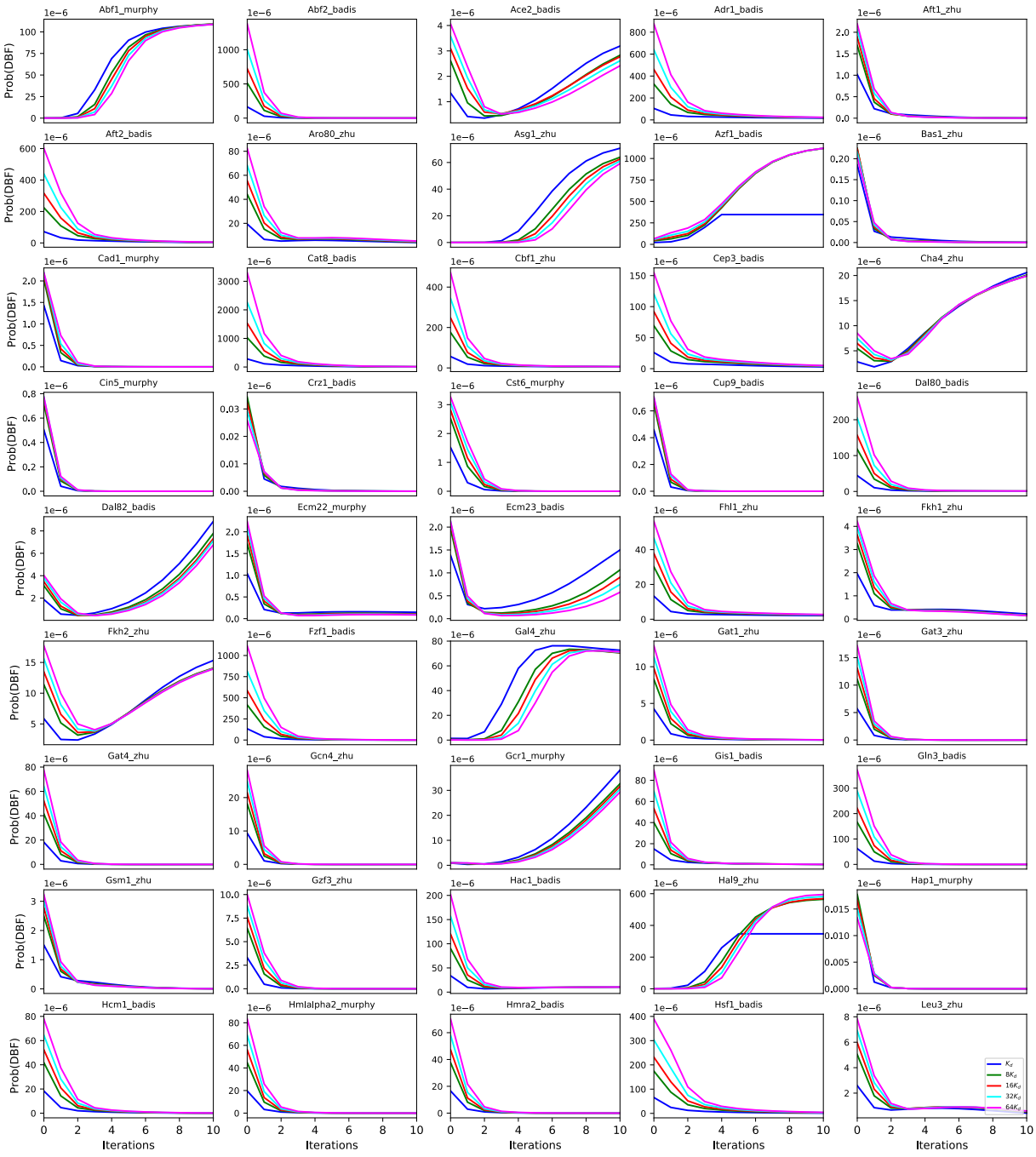
Supplementary Figure S5. Quantile-quantile plots assessing the fits of the negative binomial distributions for MNase-seq midpoint counts of `nucFrags` at each of the 147 positions of a nucleosome. The shaded gray region corresponds to the 95% confidence interval of the plots. The plots show the estimated negative binomial distributions are good fits to the midpoint counts of `nucFrags`.

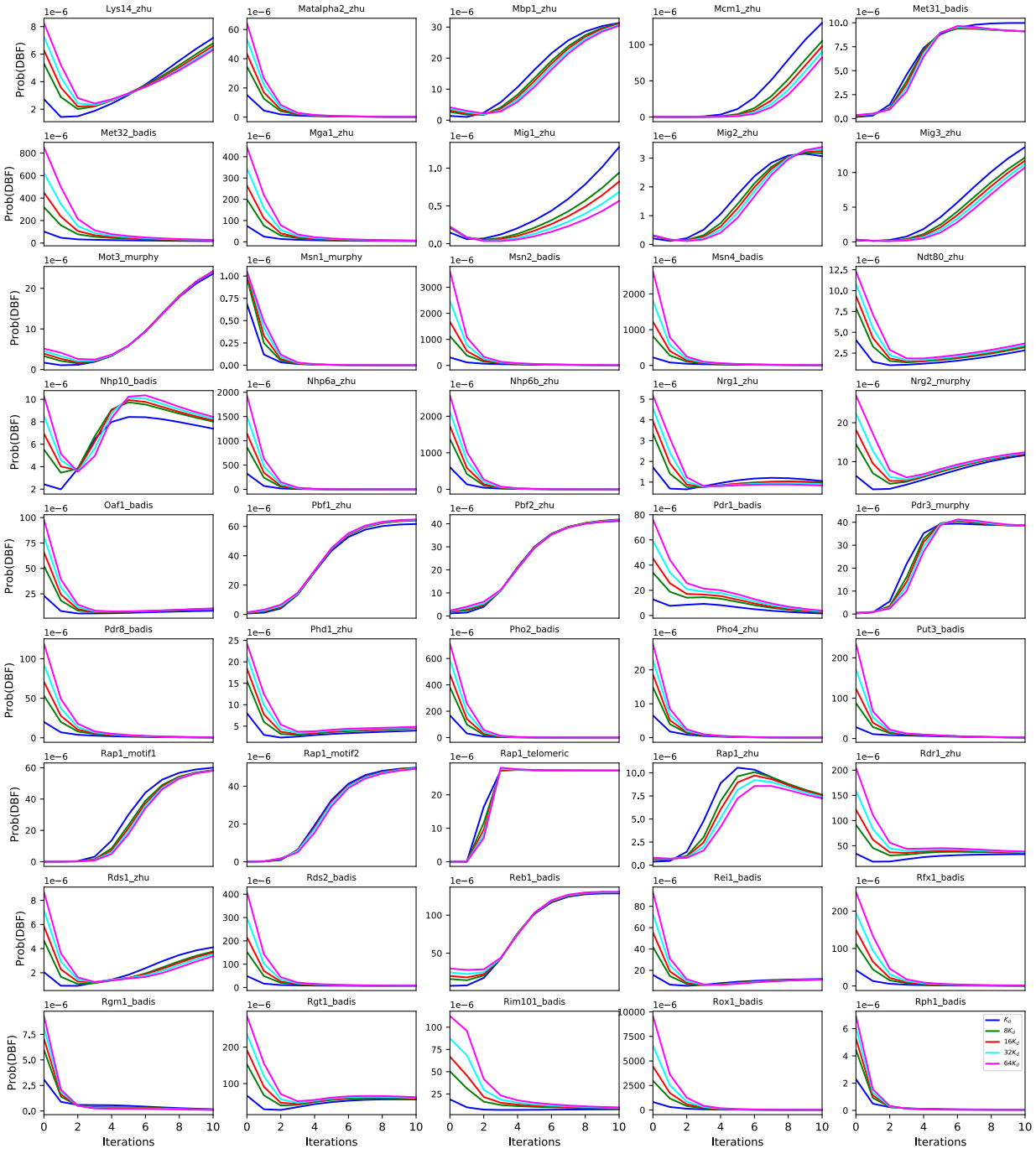


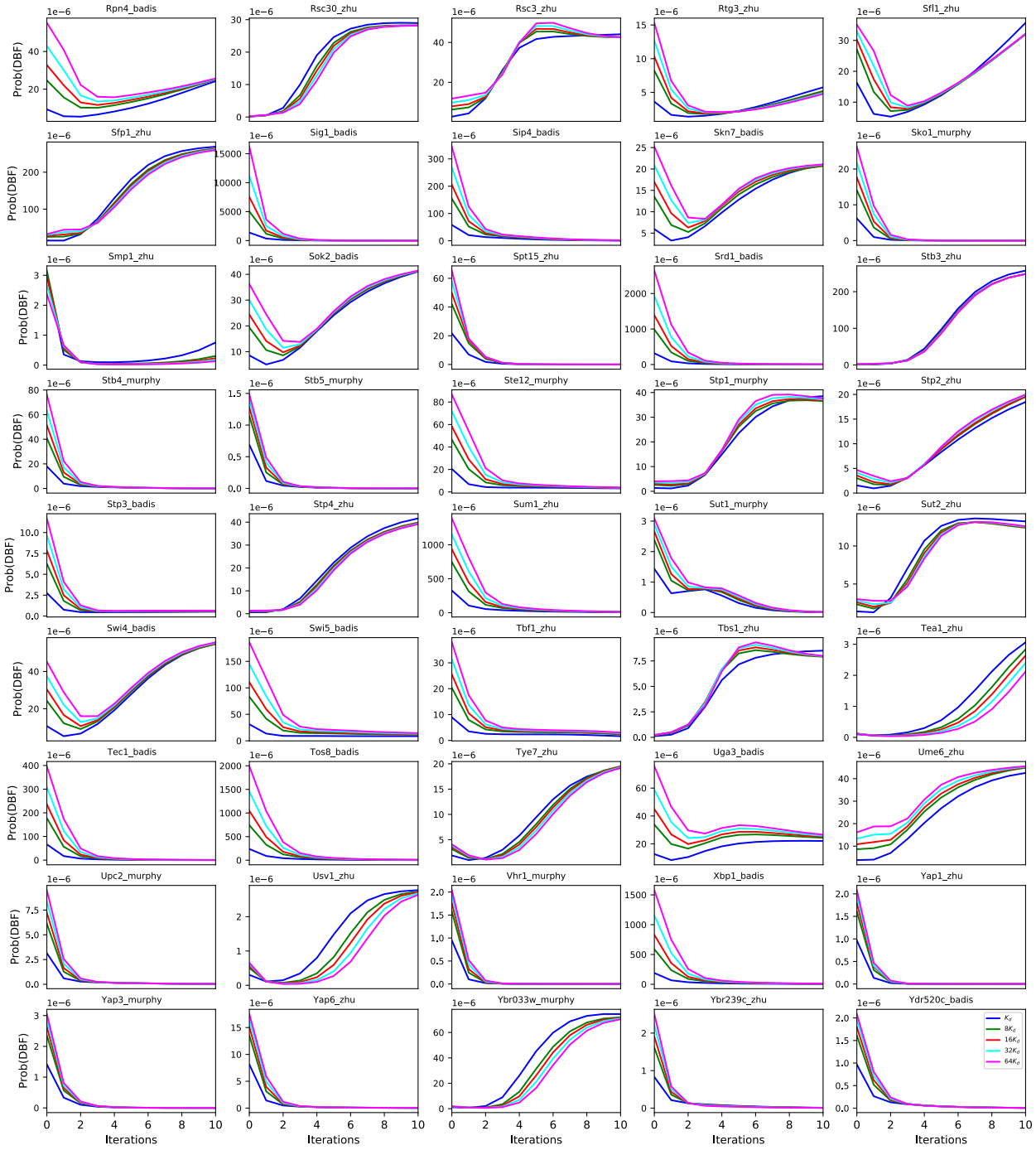
Supplementary Figure S6. Quantile-quantile plot assessing the fit of the negative binomial distribution for MNase-seq midpoint counts of `shortFrags` at annotated Abf1 and Reb1 binding sites (1). The plot shows the estimated negative binomial distribution is a good fit to the midpoint counts of `shortFrags`.

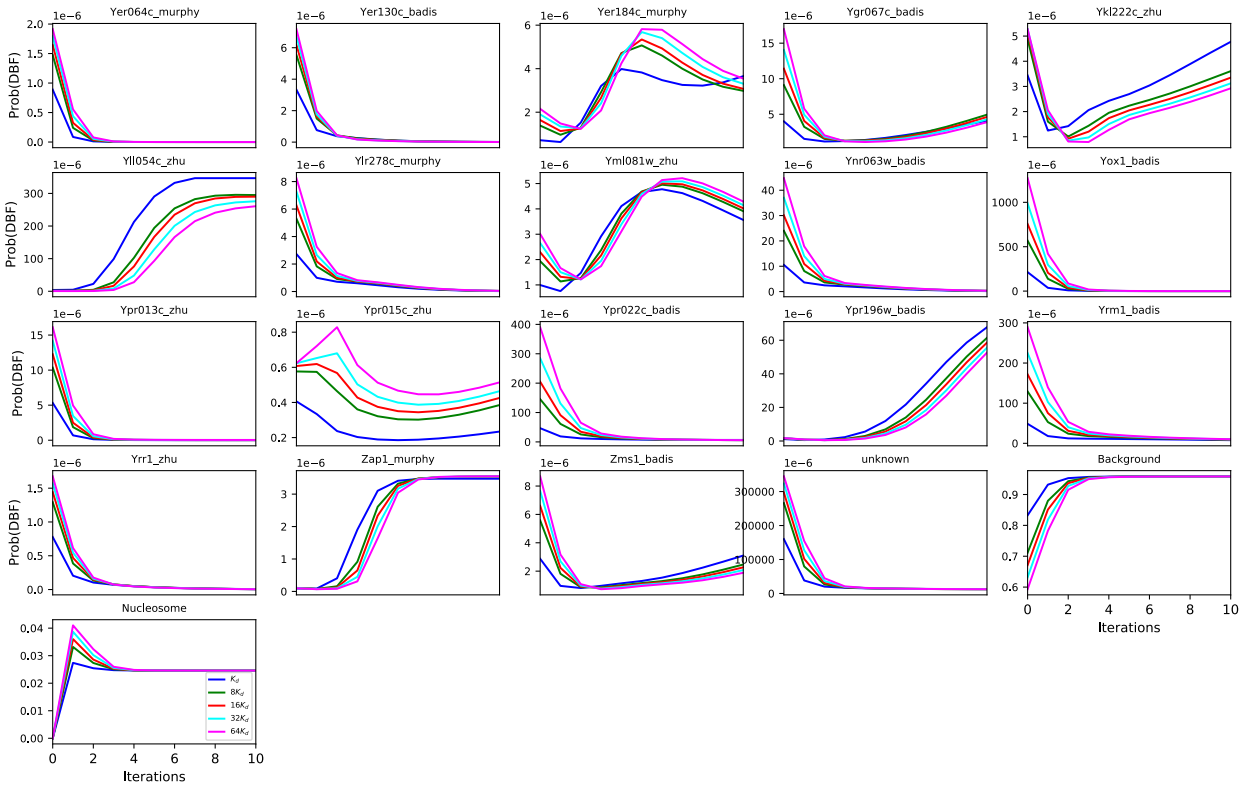


Supplementary Figure S7. Plot of the likelihood before and during the 10 iterations of EM of RoboCOP$_{\text{MNase}}$. Iteration 0 represents the likelihood at initialization, before the first round of EM. The likelihood appears to converge after just a few iterations, but we continue to 10.

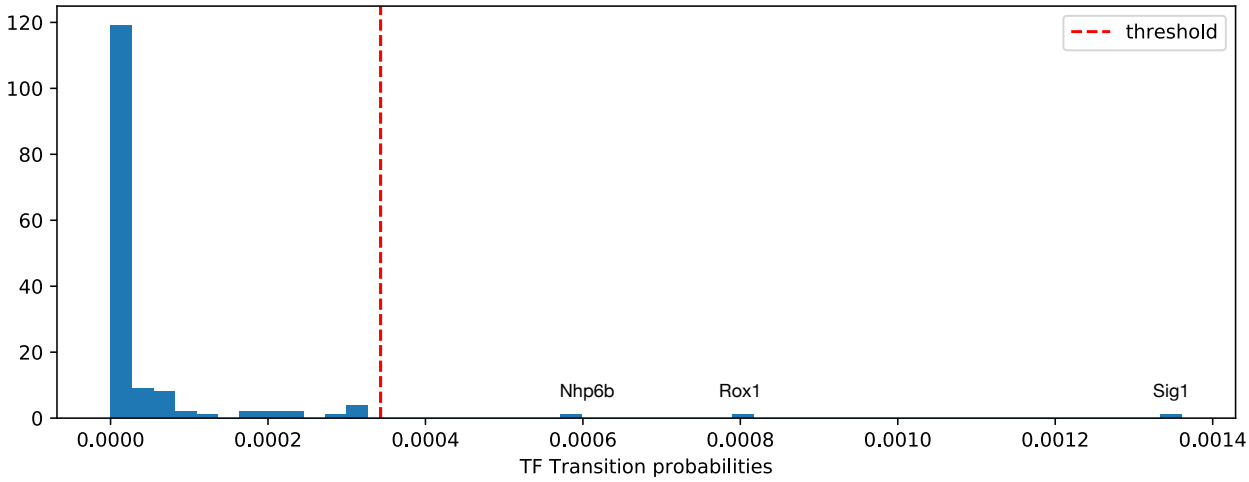

Abf1_murphy
Abf2_badis
Ace2_badis
Adr1_badis
Aft1_zhu
Aft2_badis
Aro80_zhu
Asg1_zhu
Azf1_badis
Bas1_zhu
Cad1_murphy
Cat8_badis
Cbf1_zhu
Cep3_badis
Cha4_zhu
Cin5_murphy
Crz1_badis
Cst6_murphy
Cup9_badis
Dal80_badis
Dal82_badis
Ecm22_murphy
Ecm23_badis
Fhl1_zhu
Fkh1_zhu
Fkh2_zhu
Fzf1_badis
Gal4_zhu
Gat1_zhu
Gat3_zhu
Gat4_zhu
Gcn4_zhu
Gcr1_murphy
Gis1_badis
Gln3_badis
Gsm1_zhu
Gzf3_zhu
Hac1_badis
Hal9_zhu
Hap1_murphy
Hcm1_badis
Hmlalpha2_murphy
Hmra2_badis
Hsf1_badis
Leu3_zhu
Prob(DBF)
Iterations
1e−6
K_d
$8K_d$
$16K_d$
$32K_d$
$64K_d$


Lys14_zhu
Matalpha2_zhu
Mbp1_zhu
Mcm1_zhu
Met31_badis
Met32_badis
Mga1_zhu
Mig1_zhu
Mig2_zhu
Mig3_zhu
Mot3_murphy
Msn1_murphy
Msn2_badis
Msn4_badis
Ndt80_zhu
Nhp10_badis
Nhp6a_zhu
Nhp6b_zhu
Nrg1_zhu
Nrg2_murphy
Oaf1_badis
Pbf1_zhu
Pbf2_zhu
Pdr1_badis
Pdr3_murphy
Pdr8_badis
Phd1_zhu
Pho2_badis
Pho4_zhu
Put3_badis
Rap1_motif1
Rap1_motif2
Rap1_telomeric
Rap1_zhu
Rdr1_zhu
Rds1_zhu
Rds2_badis
Reb1_badis
Rei1_badis
Rfx1_badis
Rgm1_badis
Rgt1_badis
Rim101_badis
Rox1_badis
Rph1_badis
1e−6
Prob(DBF)
Iterations
K_d
$8K_d$
$16K_d$
$32K_d$
$64K_d$

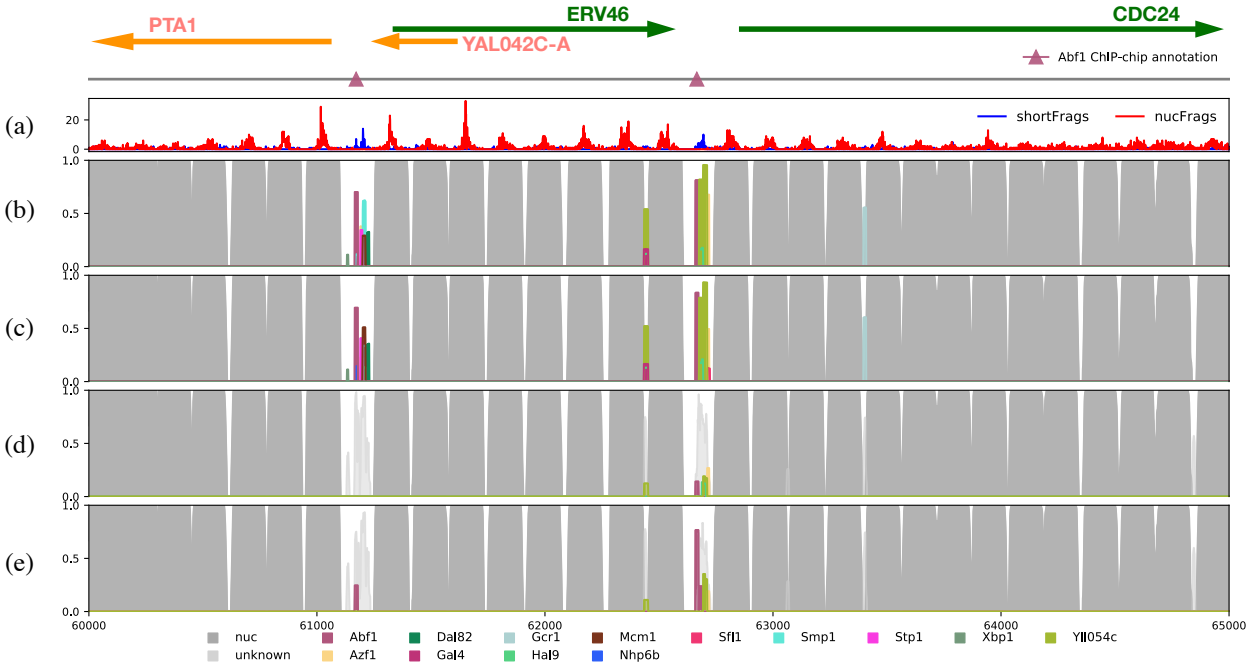

Rpn4_badis
Rsc30_zhu
Rsc3_zhu
Rtg3_zhu
Sfl1_zhu
Sfp1_zhu
Sig1_badis
Sip4_badis
Skn7_badis
Sko1_murphy
Smp1_zhu
Sok2_badis
Spt15_zhu
Srd1_badis
Stb3_zhu
Stb4_murphy
Stb5_murphy
Ste12_murphy
Stp1_murphy
Stp2_zhu
Stp3_badis
Stp4_zhu
Sum1_zhu
Sut1_murphy
Sut2_zhu
Swi4_badis
Swi5_badis
Tbf1_zhu
Tbs1_zhu
Tea1_zhu
Tec1_badis
Tos8_badis
Tye7_zhu
Uga3_badis
Ume6_zhu
Upc2_murphy
Usv1_zhu
Vhr1_murphy
Xbp1_badis
Yap1_zhu
Yap3_murphy
Yap6_zhu
Ybr033w_murphy
Ybr239c_zhu
Ydr520c_badis
1e−6
Prob(DBF)
Iterations
K_d
$8K_d$
$16K_d$
$32K_d$
$64K_d$



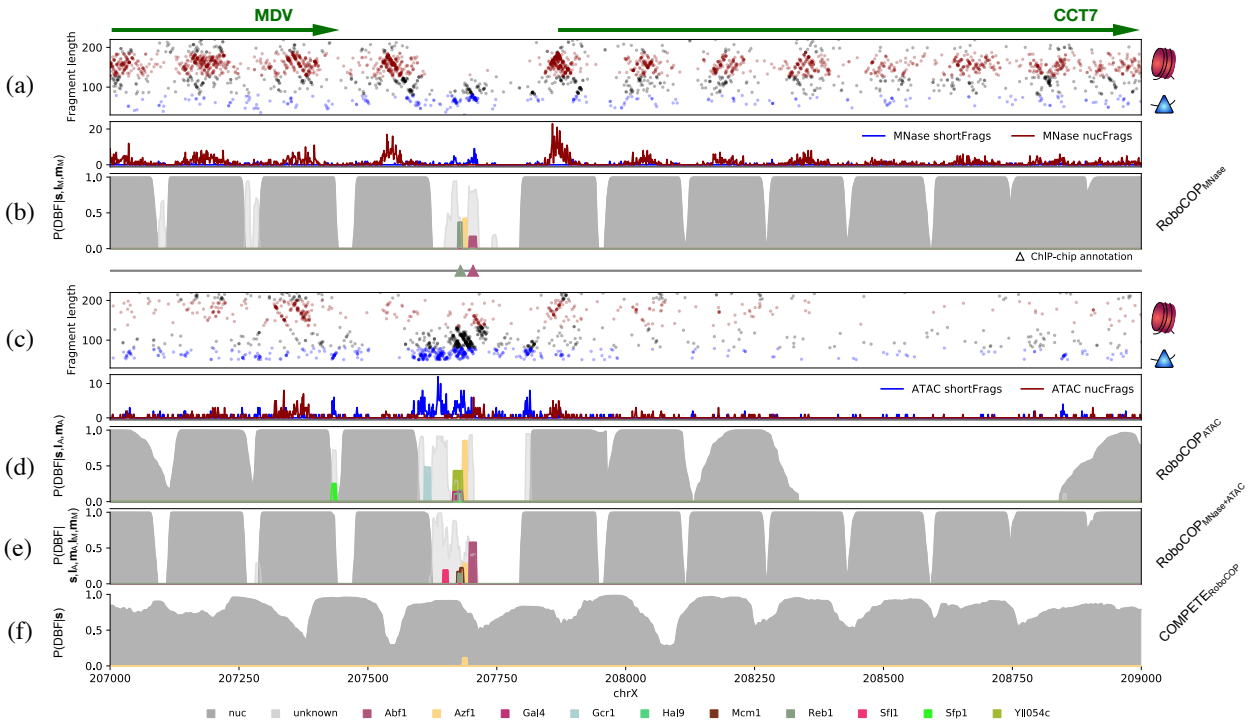
Supplementary Figure S8. For each DBF, a graph of the evolution of its transition probabilities before and during the 10 iterations of EM of RoboCOP$_{\text{MNase}}$ when initializing TF weights using various multiples of K_d.



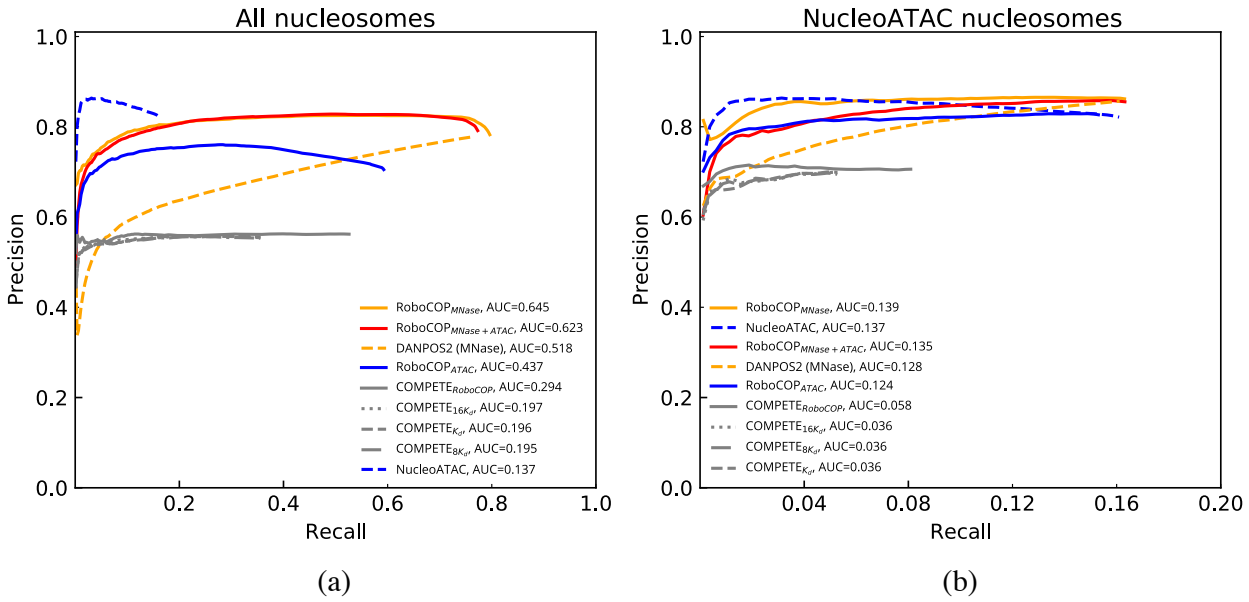
Supplementary Figure S9. Choosing a threshold (dotted red line) for the maximum transition probability of any TF. The threshold value is calculated to be two standard deviations above the mean of all initial TF transition probabilities. Only three TFs (Nhp6b, Rox1, and Sig1) have an initial transition probability above the threshold.



Supplementary Figure S10. (a) Aggregate of MNase-seq `nucFrags` and `shortFrags` from positions 60,000 to 65,000 of yeast chromosome I. Above the plot are genomic annotations for this region, with Watson strand genes depicted as green arrows and Crick strand genes as orange. Below the gene annotations, known TF binding sites (1) are indicated using triangles. This region contains two annotated binding sites for Abf1 (pink). (b) RoboCOP$_{\text{MNase}}$ without a threshold and without an 'unknown' factor, (c) with a threshold but without an 'unknown' factor, (d) without a threshold but with an 'unknown' factor, and (e) with a threshold and with an 'unknown' factor (i.e., the final model). In (b,c), we observe large TF clusters that result in increased false positive predictions. In (d), without the threshold, we see that most TFs are predicted as being an 'unknown' factor. In (e), by adding an 'unknown' factor and a threshold on how large a TF transition probability can grow, we achieve a happy medium.

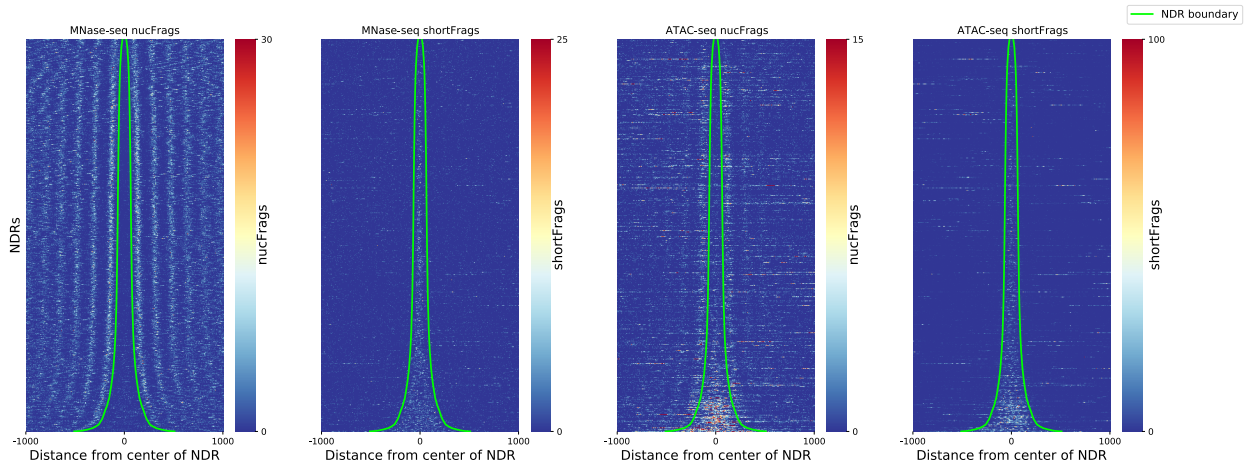


Supplementary Figure S11. Chromatin accessibility landscape from positions 207,000 to 209,000 of yeast chromosome X as assayed by (a) MNase-seq and (c) ATAC-seq. RoboCOP utilizes the chromatin accessibility information to generate a probabilistic occupancy landscape using (b) MNase-seq alone ($\text{RoboCOP}_{\text{MNase}}$), (d) ATAC-seq alone ($\text{RoboCOP}_{\text{ATAC}}$), or (e) both MNase-seq and ATAC-seq ($\text{RoboCOP}_{\text{MNase+ATAC}}$). (f) $\text{COMPETE}_{\text{RoboCOP}}$ when run with the DBF weights learned from $\text{RoboCOP}_{\text{MNase}}$ is unable to generate an accurate landscape due to the absence of chromatin accessibility data in the model.

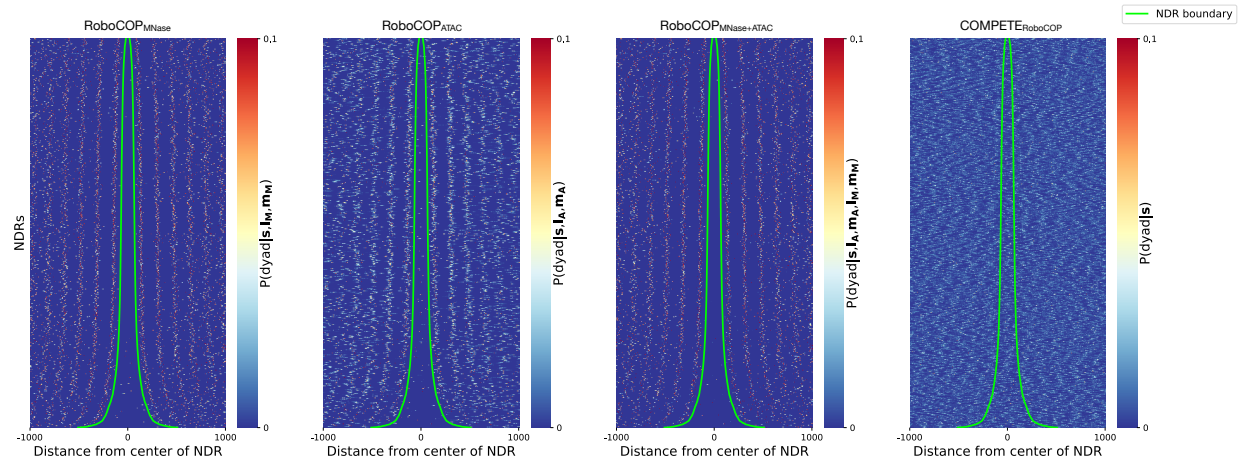


(a) (b)

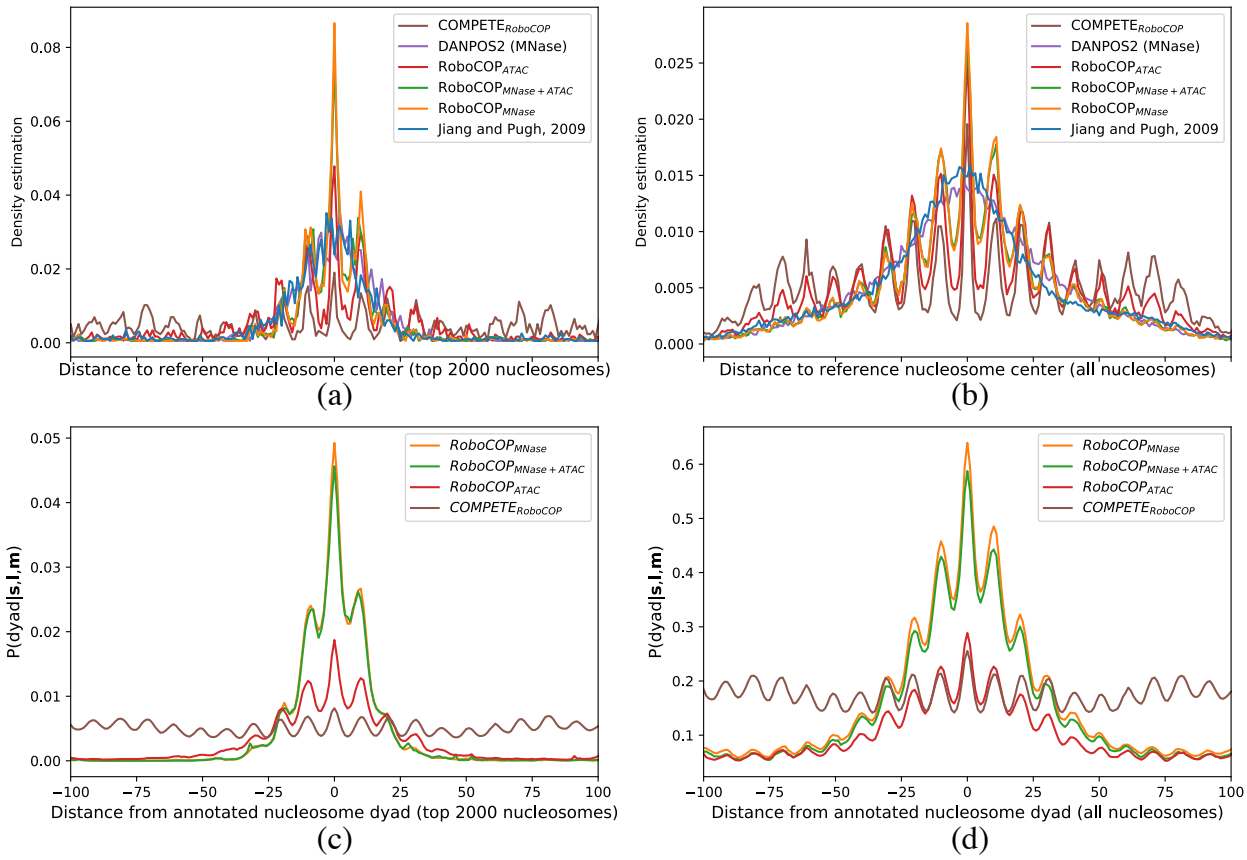
Supplementary Figure S12. (a) Precision-recall curves to assess the ability of different models to recover reference nucleosome positions. Various RoboCOP and COMPETE models are compared with DANPOS2 applied to MNase data, and NucleoATAC applied to ATAC data. Nucleosomes reported by NucleoATAC are primarily limited to those in or adjacent to promoter regions, among the easiest to identify. As a result, NucleoATAC exhibits low recall but high initial precision. (b) To compare the methods based on their ability to predict the nucleosomes reported by NucleoATAC, we filtered the predictions of each method to keep only those that are close to (within 55 bp of) a nucleosome reported by NucleoATAC, and then compared those to the reference. The precision-recall curves for these filtered lists are much closer to one another, confirming that these nucleosomes are easier for all methods to identify. Even COMPETE models, which do not incorporate chromatin accessibility information at all, fare a bit better, suggesting that the genomic sequences underlying these nucleosomes may have better matches to the dinucleotide specificity model used by COMPETE, or perhaps that the 'barrier elements' of predicted TF binding sites within promoters may help COMPETE to better position nucleosomes adjacent to the promoter.



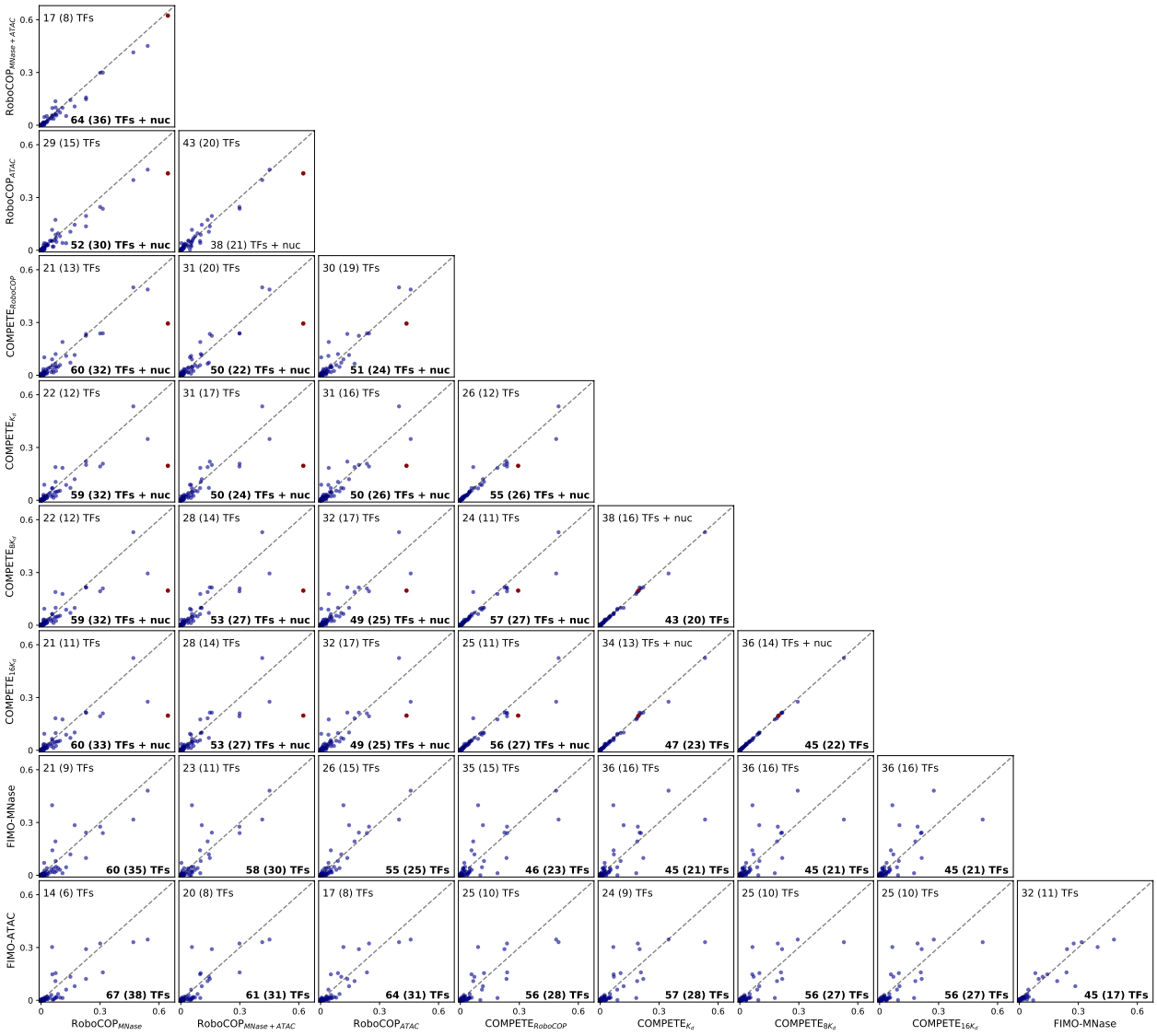
Supplementary Figure S13. Chromatin accessibility signal around NDRs for different kinds of chromatin accessibility data (MNase-seq and ATAC-seq). (a) Midpoint counts of MNase-seq `nucFrags` are depleted within the NDRs and occur in periodic intervals outside of them. (b) Midpoint counts of MNase-seq `shortFrags` are enriched within the NDRs and largely absent outside of them. (c) Midpoint counts of ATAC-seq `nucFrags` are not only enriched within the NDRs but have only a weak periodicity outside of them. (d) Midpoint counts of ATAC-seq `shortFrags` are highly enriched within the NDRs and largely absent outside of them.



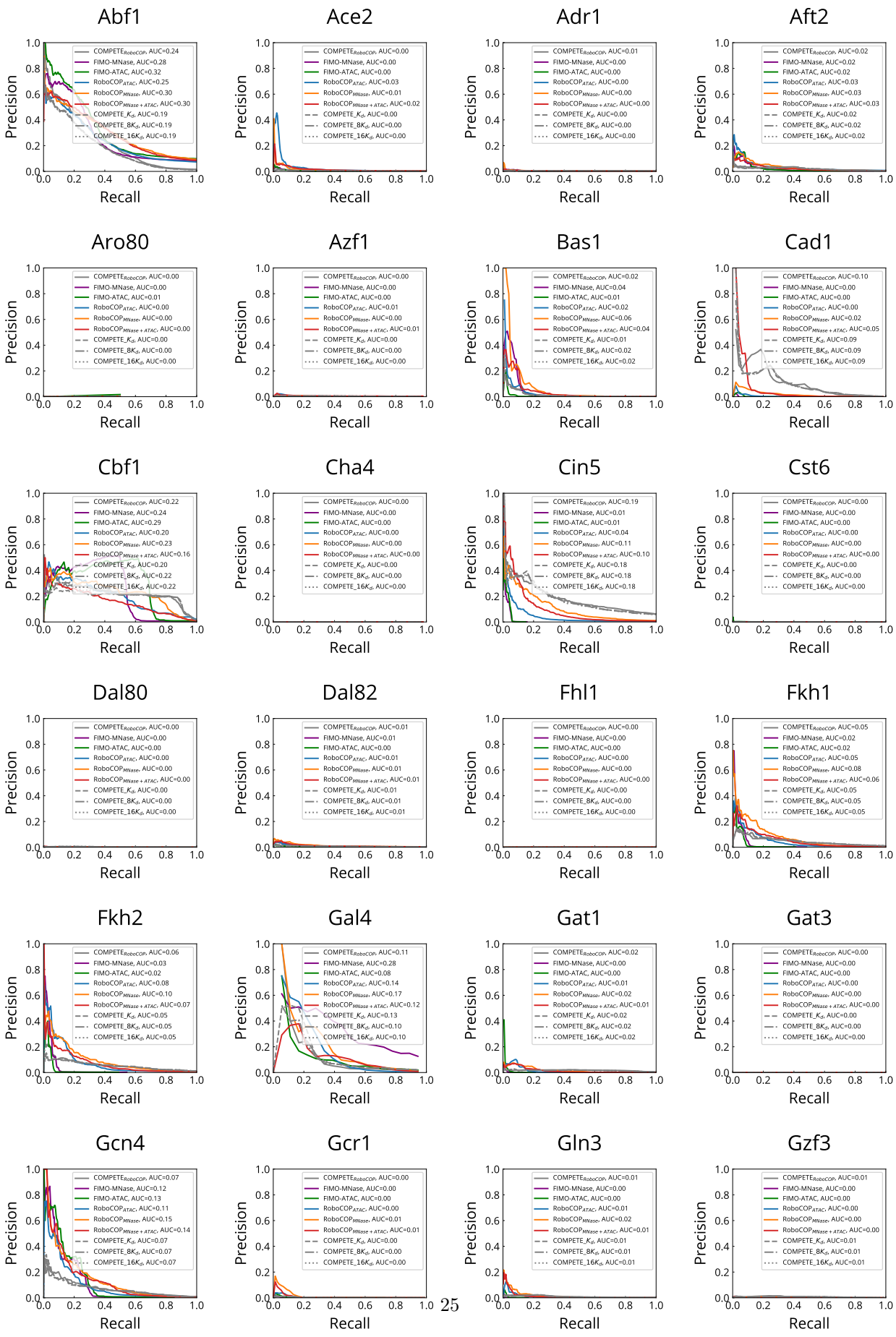
Supplementary Figure S14. Nucleosome dyad predictions around NDRs as predicted by RoboCOP when run with (a) MNase-seq, (b) ATAC-seq, or (c) both MNase-seq and ATAC-seq, as well as (d) by COMPETE which uses neither MNase-seq nor ATAC-seq data. The nucleosome predictions are periodic and well-phased around the NDRs for the three RoboCOP outputs in (a–c) even though there are differences in the signals from MNase-seq and ATAC-seq (Fig. S13). COMPETE fails to distinguish between nucleosome occupied regions and NDRs due to the absence of chromatin accessibility data.

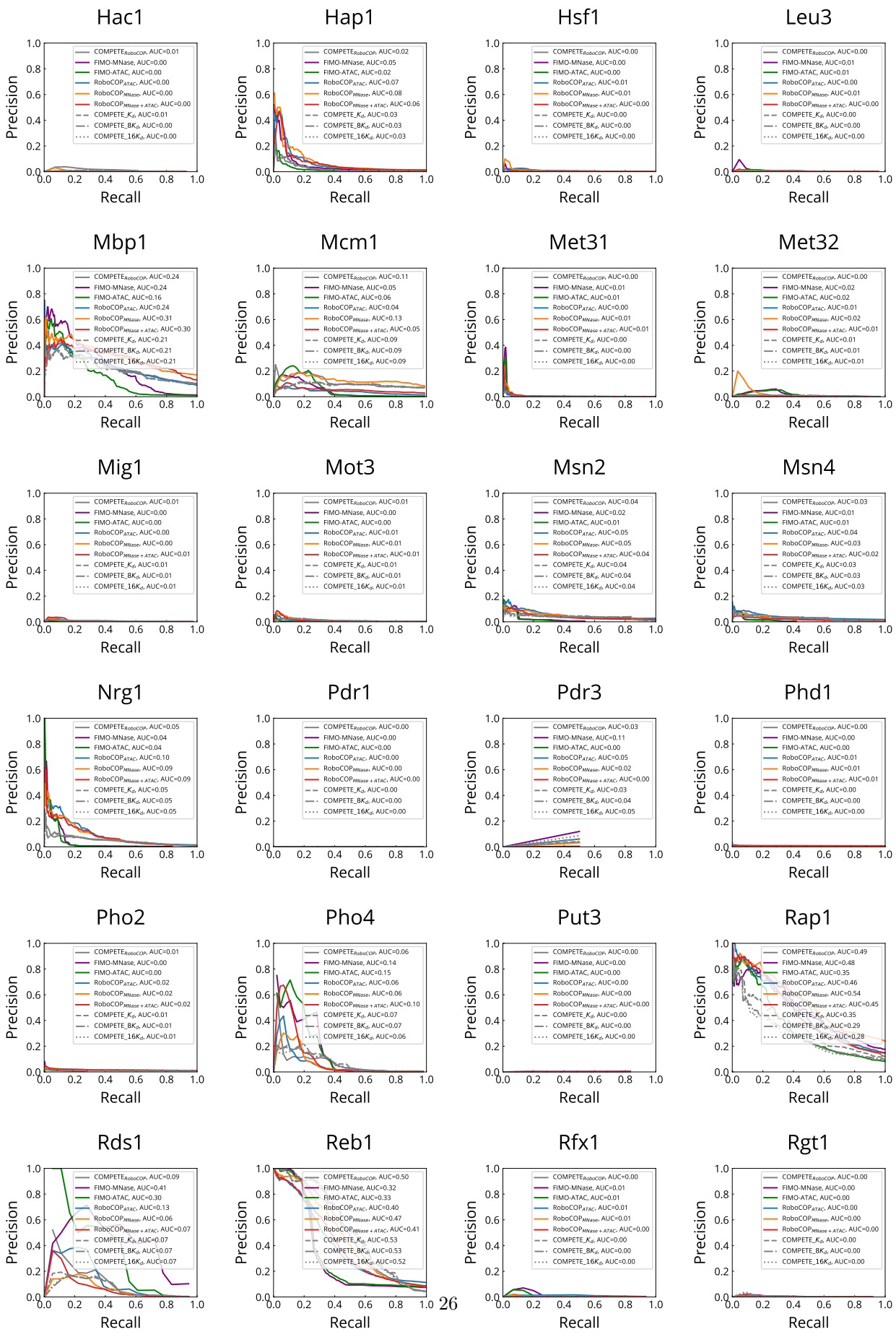


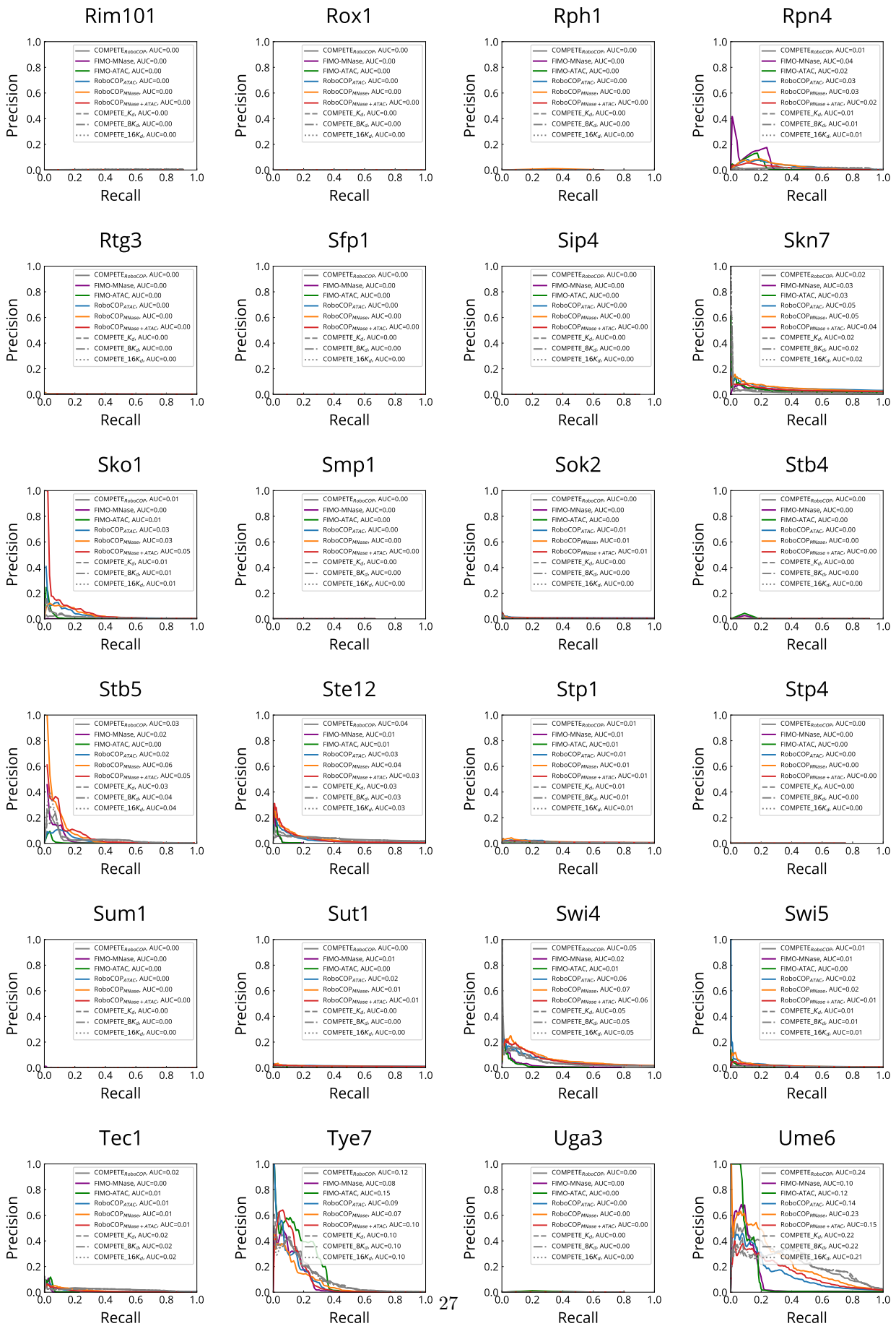
Supplementary Figure S15. Comparison of the distances between nucleosome dyads predicted by different models and those of the reference map of the (a) top 2000 well-positioned nucleosomes from (5) and (b) all nucleosomes in (5). The distribution of distances is plotted for each of the three RoboCOP models, DANPOS2, COMPETE$_{\text{RoboCOP}}$, and the previously published consensus nucleosome reference map of Jiang and Pugh (6). Nucleosome predictions of the RoboCOP models line up better with the reference nucleosome positions compared to the other methods. The distances for DANPOS2 and for Jiang and Pugh do not reflect the fact that alternative nucleosome positions tend to exhibit a 10 bp oscillation in their spacing. Curves of the posterior probability of nucleosome dyad as predicted by the RoboCOP models and COMPETE across the (c) 2,000 most well-positioned nucleosomes and (d) all nucleosomes. The signal is sharper for the top 2000 nucleosomes in (c) and becomes wider in (d) upon including all the nucleosome positions that have greater uncertainty about their positions.

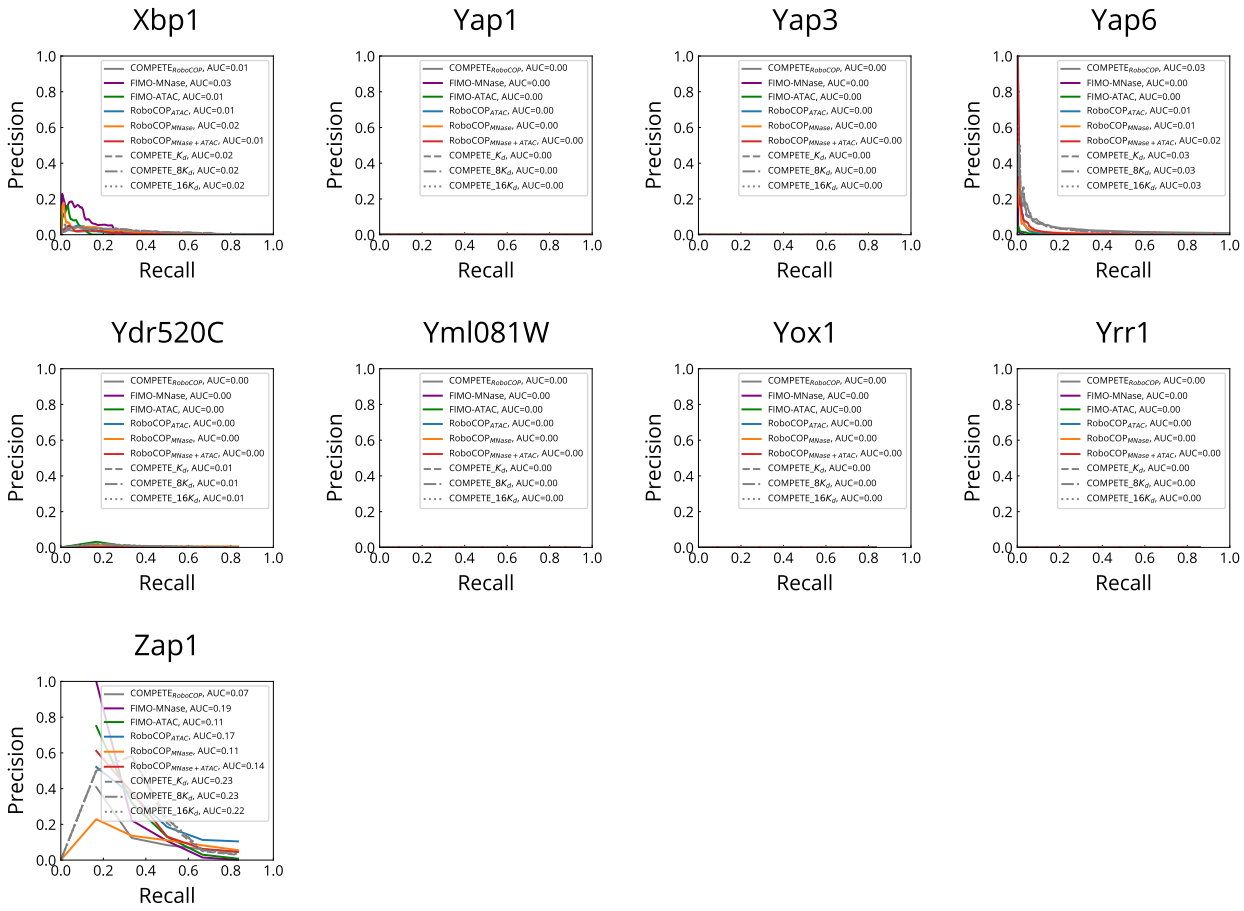


Supplementary Figure S16. Pairwise scatter plots of the AUPR values of TFs (blue) and nucleosome (red). FIMO-MNase and FIMO-ATAC only predict TFs, so plots involving FIMO-MNase and FIMO-ATAC do not include AUPR for nucleosome. Points below the diagonal indicate DBFs for which the model on the x-axis performs better than the model on the y-axis, and *vice versa*. Numbers in the corners reflect the number of DBFs for which the model on the given axis has higher AUPR values (bold if greater than the numbers on the other axis).

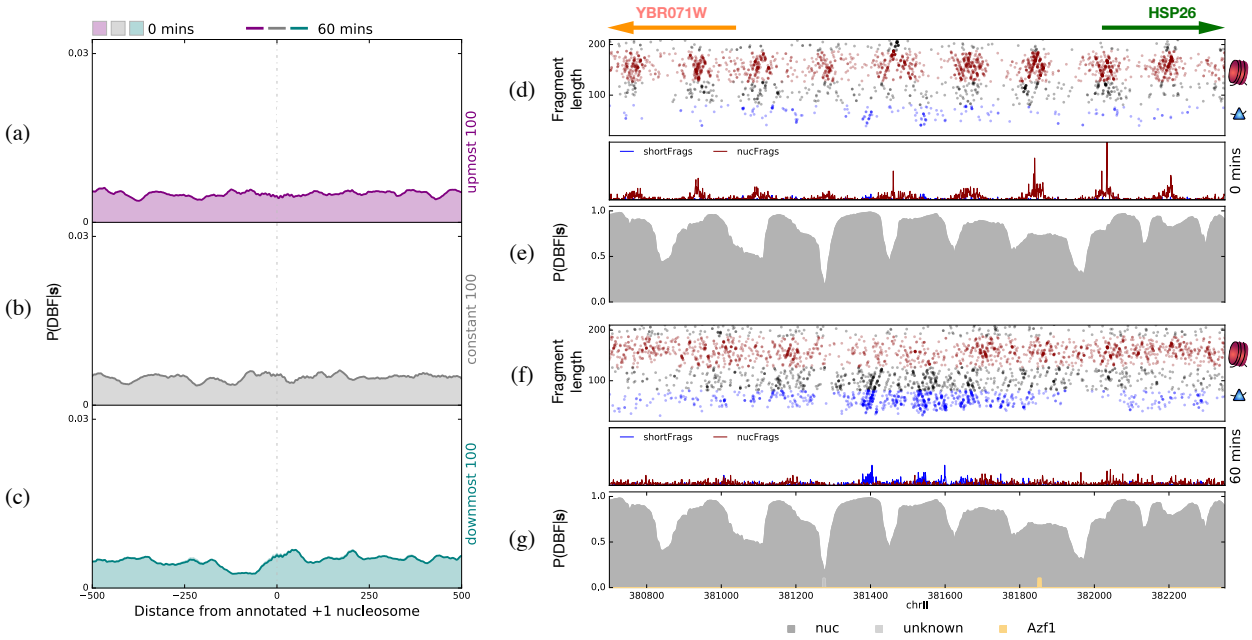

Abf1
COMPETE$_{RoboCOP}$, AUC=0.24
FIMO-MNase, AUC=0.28
FIMO-ATAC, AUC=0.32
RoboCOP$_{ATAC}$, AUC=0.25
RoboCOP$_{MNase}$, AUC=0.30
RoboCOP$_{MNase+ATAC}$, AUC=0.30
COMPETE_K_d, AUC=0.19
COMPETE_8K_d, AUC=0.19
COMPETE_16K_d, AUC=0.19
Ace2
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.03
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase+ATAC}$, AUC=0.02
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Adr1
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Aft2
COMPETE$_{RoboCOP}$, AUC=0.02
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.02
RoboCOP$_{ATAC}$, AUC=0.03
RoboCOP$_{MNase}$, AUC=0.03
RoboCOP$_{MNase+ATAC}$, AUC=0.03
COMPETE_K_d, AUC=0.02
COMPETE_8K_d, AUC=0.02
COMPETE_16K_d, AUC=0.02
Aro80
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Azf1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Bas1
COMPETE$_{RoboCOP}$, AUC=0.02
FIMO-MNase, AUC=0.04
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.02
RoboCOP$_{MNase}$, AUC=0.06
RoboCOP$_{MNase+ATAC}$, AUC=0.04
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.02
COMPETE_16K_d, AUC=0.02
Cad1
COMPETE$_{RoboCOP}$, AUC=0.10
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase+ATAC}$, AUC=0.05
COMPETE_K_d, AUC=0.09
COMPETE_8K_d, AUC=0.09
COMPETE_16K_d, AUC=0.09
Cbf1
COMPETE$_{RoboCOP}$, AUC=0.22
FIMO-MNase, AUC=0.24
FIMO-ATAC, AUC=0.29
RoboCOP$_{ATAC}$, AUC=0.20
RoboCOP$_{MNase}$, AUC=0.23
RoboCOP$_{MNase+ATAC}$, AUC=0.16
COMPETE_K_d, AUC=0.20
COMPETE_8K_d, AUC=0.22
COMPETE_16K_d, AUC=0.22
Cha4
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Cin5
COMPETE$_{RoboCOP}$, AUC=0.19
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.04
RoboCOP$_{MNase}$, AUC=0.11
RoboCOP$_{MNase+ATAC}$, AUC=0.10
COMPETE_K_d, AUC=0.18
COMPETE_8K_d, AUC=0.18
COMPETE_16K_d, AUC=0.18
Cst6
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Dal80
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Dal82
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase+ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Fhl1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Fkh1
COMPETE$_{RoboCOP}$, AUC=0.05
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.02
RoboCOP$_{ATAC}$, AUC=0.05
RoboCOP$_{MNase}$, AUC=0.08
RoboCOP$_{MNase+ATAC}$, AUC=0.06
COMPETE_K_d, AUC=0.05
COMPETE_8K_d, AUC=0.05
COMPETE_16K_d, AUC=0.05
Fkh2
COMPETE$_{RoboCOP}$, AUC=0.06
FIMO-MNase, AUC=0.03
FIMO-ATAC, AUC=0.02
RoboCOP$_{ATAC}$, AUC=0.08
RoboCOP$_{MNase}$, AUC=0.10
RoboCOP$_{MNase+ATAC}$, AUC=0.07
COMPETE_K_d, AUC=0.05
COMPETE_8K_d, AUC=0.05
COMPETE_16K_d, AUC=0.05
Gal4
COMPETE$_{RoboCOP}$, AUC=0.11
FIMO-MNase, AUC=0.28
FIMO-ATAC, AUC=0.08
RoboCOP$_{ATAC}$, AUC=0.14
RoboCOP$_{MNase}$, AUC=0.17
RoboCOP$_{MNase+ATAC}$, AUC=0.12
COMPETE_K_d, AUC=0.13
COMPETE_8K_d, AUC=0.10
COMPETE_16K_d, AUC=0.10
Gat1
COMPETE$_{RoboCOP}$, AUC=0.02
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase+ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.02
COMPETE_8K_d, AUC=0.02
COMPETE_16K_d, AUC=0.02
Gat3
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Gcn4
COMPETE$_{RoboCOP}$, AUC=0.07
FIMO-MNase, AUC=0.12
FIMO-ATAC, AUC=0.13
RoboCOP$_{ATAC}$, AUC=0.11
RoboCOP$_{MNase}$, AUC=0.15
RoboCOP$_{MNase+ATAC}$, AUC=0.14
COMPETE_K_d, AUC=0.07
COMPETE_8K_d, AUC=0.07
COMPETE_16K_d, AUC=0.07
Gcr1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase+ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Gln3
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase+ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Gzf3
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase+ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Precision
Recall


Hac1
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Hap1
COMPETE$_{RoboCOP}$, AUC=0.02
FIMO-MNase, AUC=0.05
FIMO-ATAC, AUC=0.02
RoboCOP$_{ATAC}$, AUC=0.07
RoboCOP$_{MNase}$, AUC=0.08
RoboCOP$_{MNase + ATAC}$, AUC=0.06
COMPETE_K_d, AUC=0.03
COMPETE_8K_d, AUC=0.03
COMPETE_16K_d, AUC=0.03
Hsf1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Leu3
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Mbp1
COMPETE$_{RoboCOP}$, AUC=0.24
FIMO-MNase, AUC=0.24
FIMO-ATAC, AUC=0.16
RoboCOP$_{ATAC}$, AUC=0.24
RoboCOP$_{MNase}$, AUC=0.31
RoboCOP$_{MNase + ATAC}$, AUC=0.30
COMPETE_K_d, AUC=0.21
COMPETE_8K_d, AUC=0.21
COMPETE_16K_d, AUC=0.21
Mcm1
COMPETE$_{RoboCOP}$, AUC=0.11
FIMO-MNase, AUC=0.05
FIMO-ATAC, AUC=0.06
RoboCOP$_{ATAC}$, AUC=0.04
RoboCOP$_{MNase}$, AUC=0.13
RoboCOP$_{MNase + ATAC}$, AUC=0.05
COMPETE_K_d, AUC=0.09
COMPETE_8K_d, AUC=0.09
COMPETE_16K_d, AUC=0.09
Met31
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Met32
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.02
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase + ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Mig1
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase + ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Mot3
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Msn2
COMPETE$_{RoboCOP}$, AUC=0.04
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.05
RoboCOP$_{MNase}$, AUC=0.05
RoboCOP$_{MNase + ATAC}$, AUC=0.04
COMPETE_K_d, AUC=0.04
COMPETE_8K_d, AUC=0.04
COMPETE_16K_d, AUC=0.04
Msn4
COMPETE$_{RoboCOP}$, AUC=0.03
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.04
RoboCOP$_{MNase}$, AUC=0.03
RoboCOP$_{MNase + ATAC}$, AUC=0.02
COMPETE_K_d, AUC=0.03
COMPETE_8K_d, AUC=0.03
COMPETE_16K_d, AUC=0.03
Nrg1
COMPETE$_{RoboCOP}$, AUC=0.05
FIMO-MNase, AUC=0.04
FIMO-ATAC, AUC=0.04
RoboCOP$_{ATAC}$, AUC=0.10
RoboCOP$_{MNase}$, AUC=0.09
RoboCOP$_{MNase + ATAC}$, AUC=0.09
COMPETE_K_d, AUC=0.05
COMPETE_8K_d, AUC=0.05
COMPETE_16K_d, AUC=0.05
Pdr1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Pdr3
COMPETE$_{RoboCOP}$, AUC=0.03
FIMO-MNase, AUC=0.11
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.05
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.03
COMPETE_8K_d, AUC=0.04
COMPETE_16K_d, AUC=0.05
Phd1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.01
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Pho2
COMPETE$_{RoboCOP}$, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.02
RoboCOP$_{MNase}$, AUC=0.02
RoboCOP$_{MNase + ATAC}$, AUC=0.02
COMPETE_K_d, AUC=0.01
COMPETE_8K_d, AUC=0.01
COMPETE_16K_d, AUC=0.01
Pho4
COMPETE$_{RoboCOP}$, AUC=0.06
FIMO-MNase, AUC=0.14
FIMO-ATAC, AUC=0.15
RoboCOP$_{ATAC}$, AUC=0.06
RoboCOP$_{MNase}$, AUC=0.06
RoboCOP$_{MNase + ATAC}$, AUC=0.10
COMPETE_K_d, AUC=0.07
COMPETE_8K_d, AUC=0.07
COMPETE_16K_d, AUC=0.06
Put3
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Rap1
COMPETE$_{RoboCOP}$, AUC=0.49
FIMO-MNase, AUC=0.48
FIMO-ATAC, AUC=0.35
RoboCOP$_{ATAC}$, AUC=0.46
RoboCOP$_{MNase}$, AUC=0.54
RoboCOP$_{MNase + ATAC}$, AUC=0.45
COMPETE_K_d, AUC=0.35
COMPETE_8K_d, AUC=0.29
COMPETE_16K_d, AUC=0.28
Rds1
COMPETE$_{RoboCOP}$, AUC=0.09
FIMO-MNase, AUC=0.41
FIMO-ATAC, AUC=0.30
RoboCOP$_{ATAC}$, AUC=0.13
RoboCOP$_{MNase}$, AUC=0.06
RoboCOP$_{MNase + ATAC}$, AUC=0.07
COMPETE_K_d, AUC=0.07
COMPETE_8K_d, AUC=0.07
COMPETE_16K_d, AUC=0.07
Reb1
COMPETE$_{RoboCOP}$, AUC=0.50
FIMO-MNase, AUC=0.32
FIMO-ATAC, AUC=0.33
RoboCOP$_{ATAC}$, AUC=0.40
RoboCOP$_{MNase}$, AUC=0.47
RoboCOP$_{MNase + ATAC}$, AUC=0.41
COMPETE_K_d, AUC=0.53
COMPETE_8K_d, AUC=0.53
COMPETE_16K_d, AUC=0.52
Rfx1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP$_{ATAC}$, AUC=0.01
RoboCOP$_{MNase}$, AUC=0.01
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Rgt1
COMPETE$_{RoboCOP}$, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP$_{ATAC}$, AUC=0.00
RoboCOP$_{MNase}$, AUC=0.00
RoboCOP$_{MNase + ATAC}$, AUC=0.00
COMPETE_K_d, AUC=0.00
COMPETE_8K_d, AUC=0.00
COMPETE_16K_d, AUC=0.00
Precision
Recall

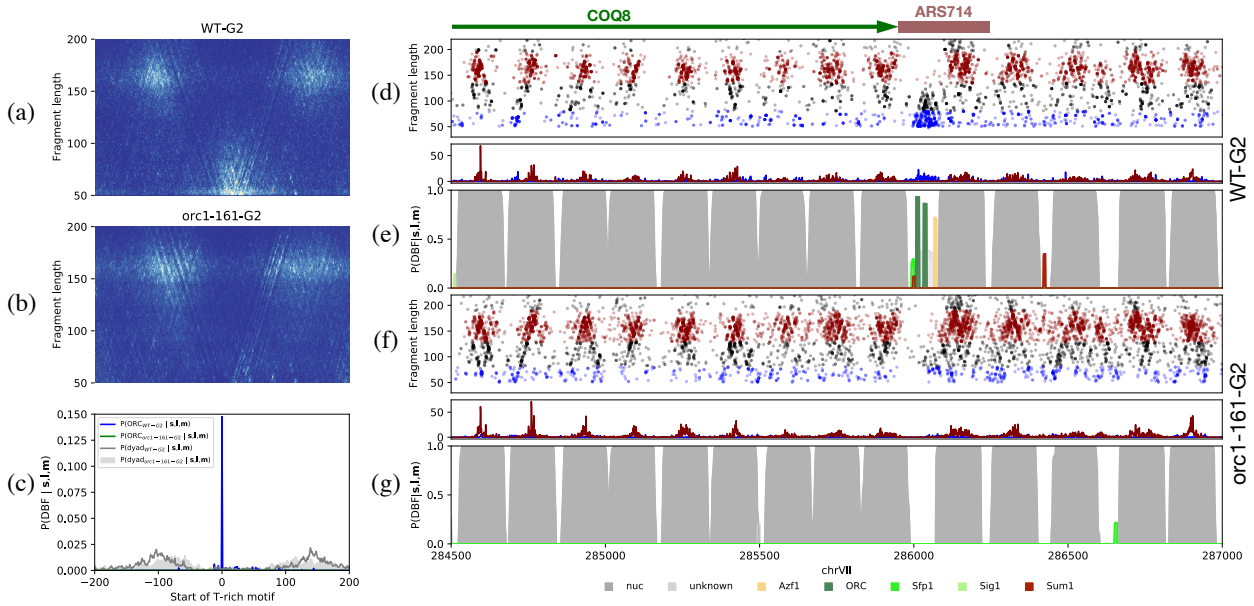

Rim101
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Rox1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Rph1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Rpn4
COMPETE_RoboCOP, AUC=0.01
FIMO-MNase, AUC=0.04
FIMO-ATAC, AUC=0.02
RoboCOP_ATAC, AUC=0.03
RoboCOP_MNase, AUC=0.03
RoboCOP_MNase + ATAC, AUC=0.02
COMPETE_Kd, AUC=0.01
COMPETE_8Kd, AUC=0.01
COMPETE_16Kd, AUC=0.01
Rtg3
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Sfp1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Sip4
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Skn7
COMPETE_RoboCOP, AUC=0.02
FIMO-MNase, AUC=0.03
FIMO-ATAC, AUC=0.03
RoboCOP_ATAC, AUC=0.05
RoboCOP_MNase, AUC=0.05
RoboCOP_MNase + ATAC, AUC=0.04
COMPETE_Kd, AUC=0.02
COMPETE_8Kd, AUC=0.02
COMPETE_16Kd, AUC=0.02
Sko1
COMPETE_RoboCOP, AUC=0.01
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.01
RoboCOP_ATAC, AUC=0.03
RoboCOP_MNase, AUC=0.03
RoboCOP_MNase + ATAC, AUC=0.05
COMPETE_Kd, AUC=0.01
COMPETE_8Kd, AUC=0.01
COMPETE_16Kd, AUC=0.01
Smp1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Sok2
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.01
RoboCOP_MNase, AUC=0.01
RoboCOP_MNase + ATAC, AUC=0.01
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Stb4
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Stb5
COMPETE_RoboCOP, AUC=0.03
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.02
RoboCOP_MNase, AUC=0.06
RoboCOP_MNase + ATAC, AUC=0.05
COMPETE_Kd, AUC=0.03
COMPETE_8Kd, AUC=0.04
COMPETE_16Kd, AUC=0.04
Ste12
COMPETE_RoboCOP, AUC=0.04
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP_ATAC, AUC=0.03
RoboCOP_MNase, AUC=0.04
RoboCOP_MNase + ATAC, AUC=0.03
COMPETE_Kd, AUC=0.03
COMPETE_8Kd, AUC=0.03
COMPETE_16Kd, AUC=0.03
Stp1
COMPETE_RoboCOP, AUC=0.01
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.01
RoboCOP_ATAC, AUC=0.01
RoboCOP_MNase, AUC=0.01
RoboCOP_MNase + ATAC, AUC=0.01
COMPETE_Kd, AUC=0.01
COMPETE_8Kd, AUC=0.01
COMPETE_16Kd, AUC=0.01
Stp4
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Sum1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Sut1
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.02
RoboCOP_MNase, AUC=0.01
RoboCOP_MNase + ATAC, AUC=0.01
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Swi4
COMPETE_RoboCOP, AUC=0.05
FIMO-MNase, AUC=0.02
FIMO-ATAC, AUC=0.01
RoboCOP_ATAC, AUC=0.06
RoboCOP_MNase, AUC=0.07
RoboCOP_MNase + ATAC, AUC=0.06
COMPETE_Kd, AUC=0.05
COMPETE_8Kd, AUC=0.05
COMPETE_16Kd, AUC=0.05
Swi5
COMPETE_RoboCOP, AUC=0.01
FIMO-MNase, AUC=0.01
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.02
RoboCOP_MNase, AUC=0.02
RoboCOP_MNase + ATAC, AUC=0.01
COMPETE_Kd, AUC=0.01
COMPETE_8Kd, AUC=0.01
COMPETE_16Kd, AUC=0.01
Tec1
COMPETE_RoboCOP, AUC=0.02
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.01
RoboCOP_ATAC, AUC=0.01
RoboCOP_MNase, AUC=0.01
RoboCOP_MNase + ATAC, AUC=0.01
COMPETE_Kd, AUC=0.02
COMPETE_8Kd, AUC=0.02
COMPETE_16Kd, AUC=0.02
Tye7
COMPETE_RoboCOP, AUC=0.12
FIMO-MNase, AUC=0.08
FIMO-ATAC, AUC=0.15
RoboCOP_ATAC, AUC=0.09
RoboCOP_MNase, AUC=0.07
RoboCOP_MNase + ATAC, AUC=0.10
COMPETE_Kd, AUC=0.10
COMPETE_8Kd, AUC=0.10
COMPETE_16Kd, AUC=0.10
Uga3
COMPETE_RoboCOP, AUC=0.00
FIMO-MNase, AUC=0.00
FIMO-ATAC, AUC=0.00
RoboCOP_ATAC, AUC=0.00
RoboCOP_MNase, AUC=0.00
RoboCOP_MNase + ATAC, AUC=0.00
COMPETE_Kd, AUC=0.00
COMPETE_8Kd, AUC=0.00
COMPETE_16Kd, AUC=0.00
Ume6
COMPETE_RoboCOP, AUC=0.24
FIMO-MNase, AUC=0.10
FIMO-ATAC, AUC=0.12
RoboCOP_ATAC, AUC=0.14
RoboCOP_MNase, AUC=0.23
RoboCOP_MNase + ATAC, AUC=0.15
COMPETE_Kd, AUC=0.22
COMPETE_8Kd, AUC=0.22
COMPETE_16Kd, AUC=0.21
Precision
Recall



Supplementary Figure S17. Precision-recall curves to assess the ability of different models to recover the binding sites of the 81 TFs whose binding sites have been reported (1). Various RoboCOP and COMPETE models are compared with FIMO-MNase and FIMO-ATAC.



Supplementary Figure S18. (a–c) Aggregate nucleosome dyad probability, as computed by COMPETE$_{\text{RoboCOP}}$, around annotated +1 nucleosomes (7) of (a) the 100 most up-regulated genes (purple), (b) the 100 genes least changed in transcription (gray), and (c) the 100 most down-regulated genes (teal), before and 60 minutes after treating cells with cadmium. We find that COMPETE$_{\text{RoboCOP}}$ does not identify any change in these nucleosome occupancy profiles after cadmium treatment. (d) Two-dimensional plot of MNase-seq fragments near the HSP26 promoter (positions 380,700 to 382,350 of yeast chromosome II are shown) before treatment with cadmium (`nucFrags` in red; `shortFrags` in blue), along with the `nucFrags` and `shortFrags` signals that result from aggregating those midpoint counts. Gene annotations depicted with arrows at the top (Watson strand in green; Crick strand in orange). (e) COMPETE$_{\text{RoboCOP}}$-predicted occupancy profile of this region before treatment with cadmium. (f,g) The same as (d,e), respectively, but 60 minutes after cadmium treatment. Even though the MNase-seq data have changed significantly after treatment, COMPETE$_{\text{RoboCOP}}$ does not make use of this data (except via updates to the RoboCOP-learned transition probabilities), so fails to capture these changes in its chromatin occupancy profile.



Supplementary Figure S19. Composite heatmaps of MNase-seq fragments around all active origins of replication, centered on the start of T-rich ORC motifs from (8), for (a) wild-type yeast cells under G_2 arrest, and (b) *orc1-161* mutant cells under G_2 arrest. The mutant cells lack an enrichment of small fragments around the annotated origins. (c) The composite RoboCOP-computed probability of ORC binding is sharply peaked at the start of the ORC motif when applied to data from the wild-type cells, but is virtually absent when applied to data from the *orc1-161* mutant cells. Moreover, the composite probability of the nucleosome dyad around these regions indicates that nucleosomes move closer to the origin in mutant cells than in wild-type cells. (d) MNase-seq fragment plot around a specific origin, ARS714 (positions 284,500 to 287,000 of chromosome VII), shows that wild-type cells have an enrichment of `shortFrags` within the ACS portion of the ARS, while these `shortFrags` are missing in the *orc1-161* mutant, shown in (f). RoboCOP output in (e) indicates that the `shortFrags` in the ACS correspond to two bound ORC motifs, but this binding is missing in the RoboCOP output from the mutant, shown in (g). The two flanking nucleosomes around the ACS have also moved closer to each other in the mutant cells.

Supplementary References

1. MacIsaac, K. D., Wang, T., Gordon, D. B., Gifford, D. K., Stormo, G. D., and Fraenkel, E. (December, 2006) An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics,* **7**(1), 113.

2. Rhee, H. S. and Pugh, B. F. (December, 2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell,* **147**(6), 1408–1419.

3. Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K., and Henikoff, S. (February, 2014) High-resolution mapping of transcription factor binding sites on native chromatin. *Nature Methods,* **11**(2), 203–209.

4. Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., MacIsaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J.-B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., and Young, R. A. (September, 2004) Transcriptional regulatory code of a eukaryotic genome. *Nature,* **431**(7004), 99–104.

5. Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (June, 2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature,* **486**(7404), 496–501.

6. Jiang, C. and Pugh, B. F. (2009) A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biology,* **10**, R109.

7. Chereji, R. V., Ramachandran, S., Bryson, T. D., and Henikoff, S. (February, 2018) Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biology,* **19**(1), 19.

8. Belsky, J. A., MacAlpine, H. K., Lubelsky, Y., Hartemink, A. J., and MacAlpine, D. M. (January, 2015) Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes & Development,* **29**(2), 212–224.