

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

Table S1: List of 150 TFs used by RoboCOP. All TFs excepting the ones colored in gray have binding sites reported in MacIsaac *et al.* (2006). The TFs colored in red have alternate validation sets from ChIP-exo (Rhee and Pugh, 2011) and ORGANIC (Kasinathan *et al.*, 2014) protocols. The ones colored in blue are known to express more under specific experimental conditions (Harbison *et al.*, 2004).

<b>top100</b>	<b>bottom100</b>	<b>mid100</b>
HXK1	RPL9A	TCM62
HSP78	RPS22A	MCK1
MMP1	SSB1	AGC1
YHR138C	RPS31	GPB2
YET1	RPS3	HEH2
HNT1	ASC1	YPL168W
GOR1	RPL3	MRK1
STF1	YEF3	SEN2
HOM6	RPL8B	MTC2
MET8	RPS9B	FCY21
PRC1	RPL5	CDD1
GSC2	RPS5	CDC24
PGM2	RPP2B	COX11
GSH1	RPS12	TFC8
YER067W	RPL12A	IME1
CPR1	RPS2	YBP1
RTS3	RPP2A	YEL1
ARN2	RPP0	IST3
WTM1	RPP1A	PKC1
MET22	RPS15	DTD1
GPM1	RPS0B	ADY4
DDR2	RPL28	VPS52
HOR7	RPL38	YGR153W
MUP1	RPS14A	YPL199C
GPD1	ILV5	FRT1
GRE2	RPL16B	LEM3
ARA1	RPL22A	PFA3
SER33	RPL2A	CUE2
YLR194C	RPL1B	PSY3
ENO1	RPL2B	THI72
CMK2	RPL32	SHU2
GRX1	SSB2	NSP1
ALD4	RPL1A	RAD53
GSP2	RPS16B	YKL065W-A
MUP3	RPL18A	RAD24
YOR285W	RPL25	ATG16
MET2	RPS13	VRP1
FMO1	RPL11A	MLH2
MRP8	RPL21A	MSY1
YBR085C-A	RPL26B	ADE2
TRX2	RPL17A	BTS1
SAM2	RPL15A	EXO84
PDI1	RPL30	YNL165W
MXR1	RPS11A	ASF2
TFS1	RPS8A	IOC2
TPS1	RPS1B	MNT3
YMR315W	RPS4A	CBF1 (contd.)

<b>top100</b>	<b>bottom100</b>	<b>mid100</b>
PRX1	RPS26B	CSN12
ERO1	RPL19B	AYT1
TSL1	RPL31A	MSS11
HSP82	RPS7A	KIP1
PGK1	SNU13	CCT3
GRE3	RPS4B	EAF6
YNL134C	RPS20	PUS5
AIM17	RPL42B	DSS1
SSA1	RPL27A	RPB11
MET1	RPS18B	YNL176C
GLK1	RPS0A	NTE1
FIT2	RPL12B	SGF73
HBN1	RPS10A	PRK1
TDH1	RPL13B	TRR1
HSP104	RPL6B	APC5
AGP3	RPL8A	HOS2
FIT3	RPL24A	MSE1
SEO1	NSR1	QDR1
MET6	RPS24B	MTC5
YNL208W	RPS18A	YIR035C
PHO89	RPS6B	GLE1
ICY2	RPL11B	RTT103
MET16	RPS6A	MNS1
LAP4	RPL6A	NHP10
PEP4	RPL20A	FTR1
TSA1	RPS24A	FLC1
UBI4	RPL43A	RTC5
HSP42	RPS8B	PLB3
BNA3	RPL33A	PAU15
KAR2	NOP1	FIT1
TIS11	RPL4A	BIO3
RTC3	RPL14A	CTK3
TMA10	RPS19B	AIM33
ZWF1	TMA19	MMS21
PBI2	RPL16A	YLR253W
BDS1	RPL37A	LRC6
YCT1	RPL19A	ISC1
PNC1	RPL7A	SGM1
CYS3	RPS17A	RSM28
PRB1	RPL17B	RRP43
ECM17	RPS23B	CFT2
SPG4	RPS27B	YMR166C
GRE1	RPL20B	GYL1
STF2	RPS11B	NUP49
MET10	RPL23A	GGA2
AHP1	RPS28A	YPL257W
DDR48	STM1	MEC1
MET14	RPL40B	SMP3
HSP26	RPL34B	EAP1
YDL124W	RPP1B	CDC11
PDC6	RPL39	CDC26
MET3	ILV3	SAG1
HSP12	RPL23B	YHC1

Table S2: List of 100 most upregulated, most downregulated genes. And 100 genes with almost no changes in transcription.

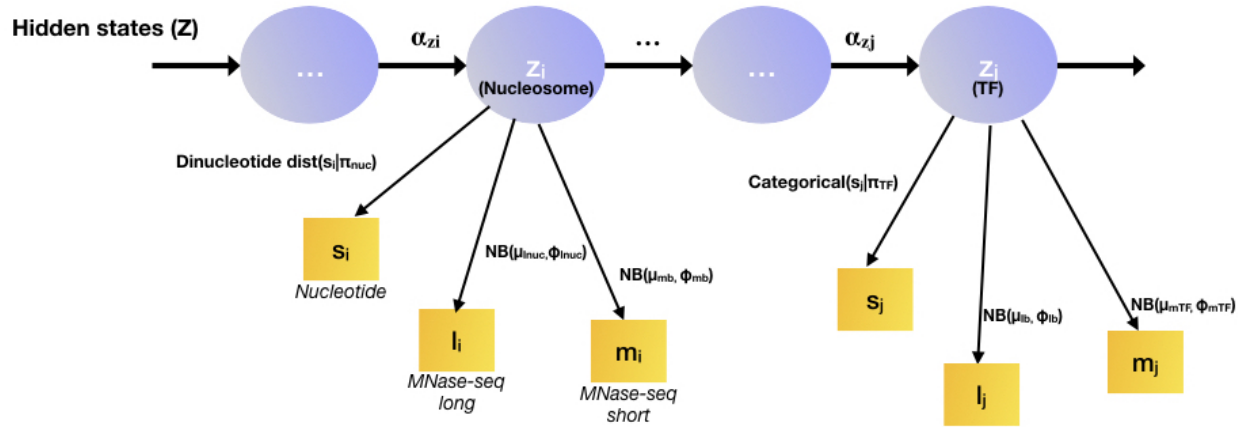


Figure S1: Hidden state sequence of RoboCOP. Each state emits a nucleotide using a categorical distribution, number of midpoints of nucFragms using a negative binomial distribution, and number of midpoints of shortFragms using a negative binomial distribution.

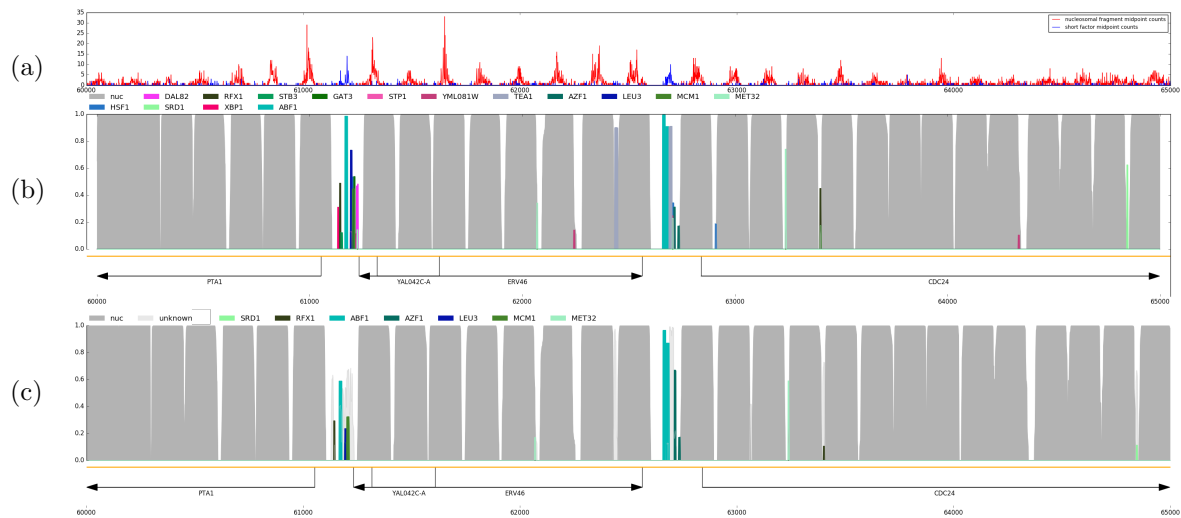
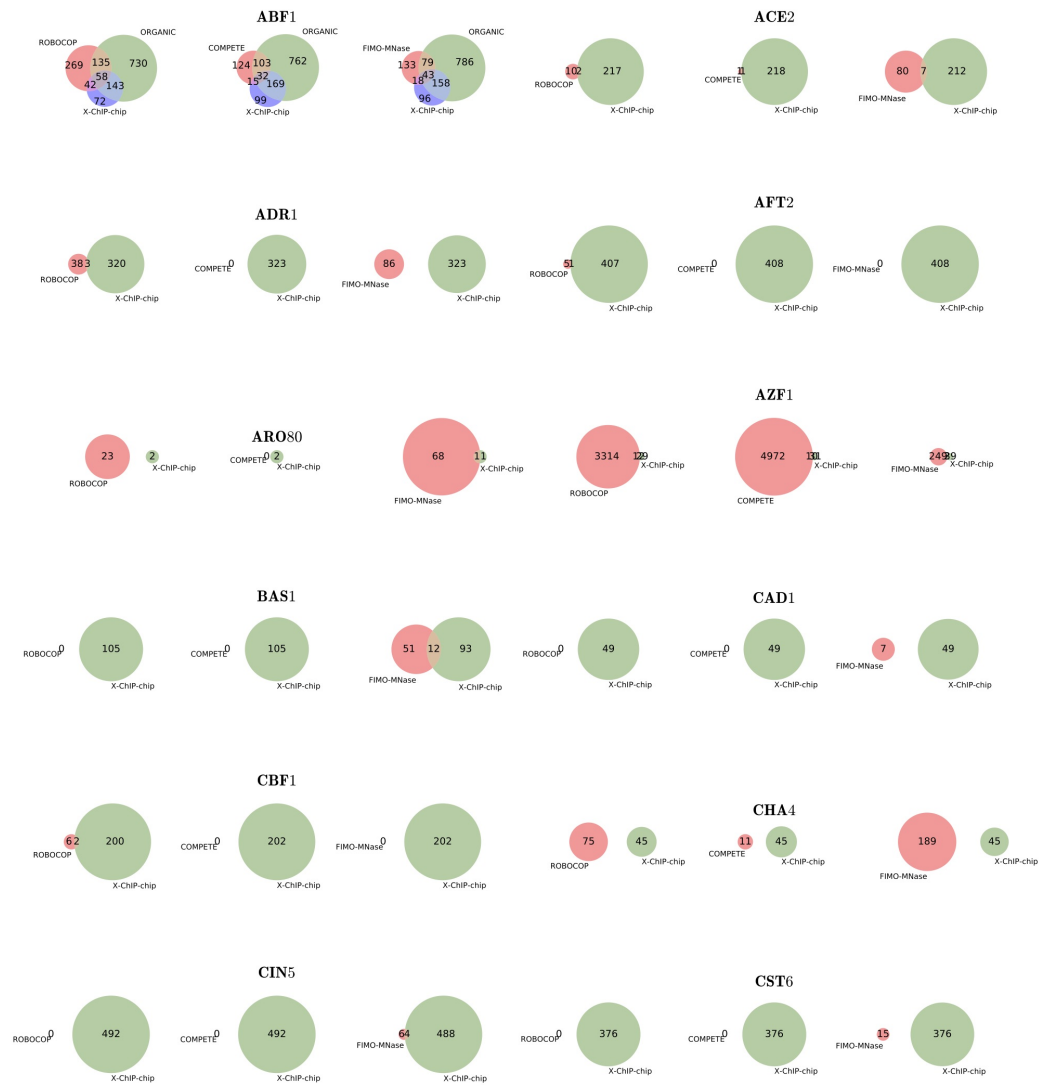
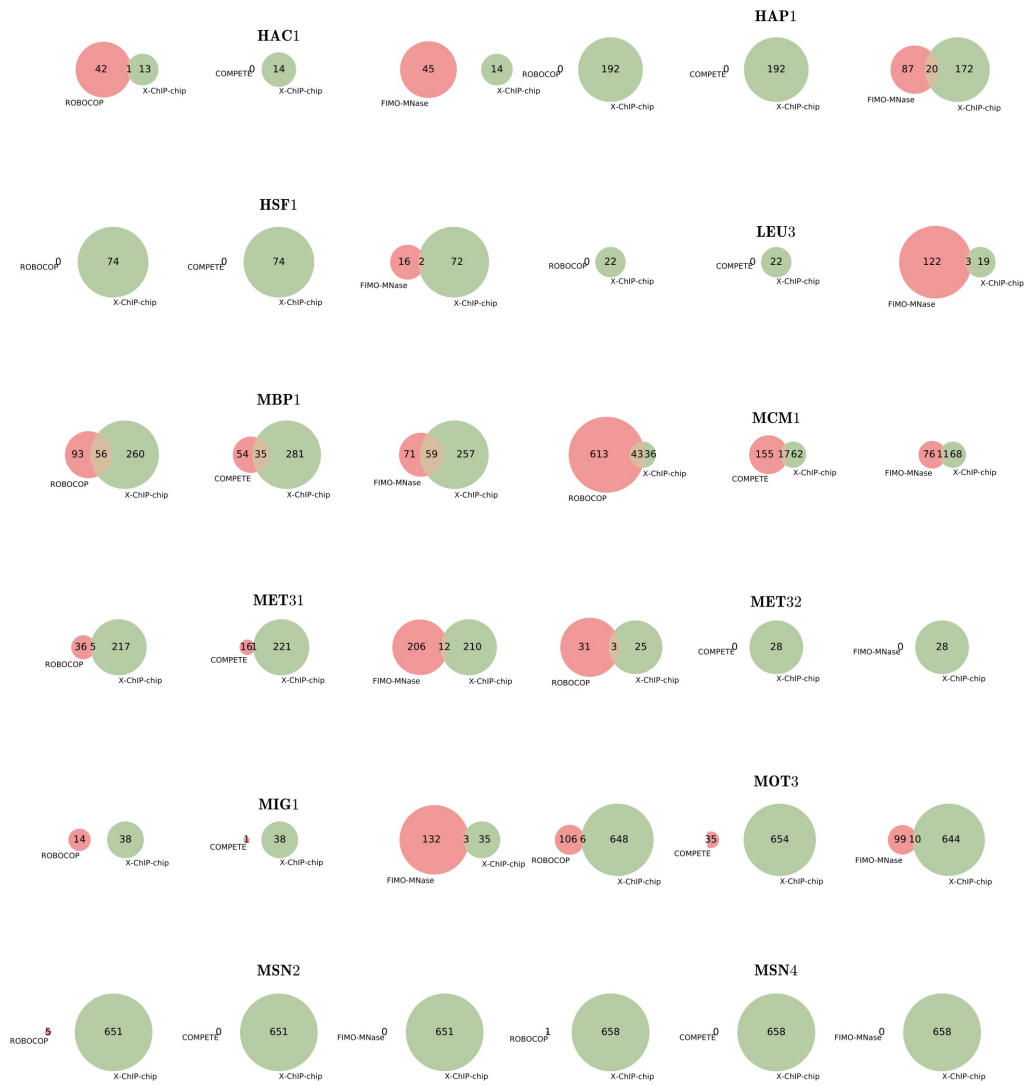


Figure S2: (a) MNase-seq fragment midpoint counts. Running RoboCOP (b) with and (c) without “unknown” factor. In (c) we observe more clustering of TFs means there are more false positives. Adding “unknown” factor as DBF reduces TF clustering thereby reducing number of false positive predictions.















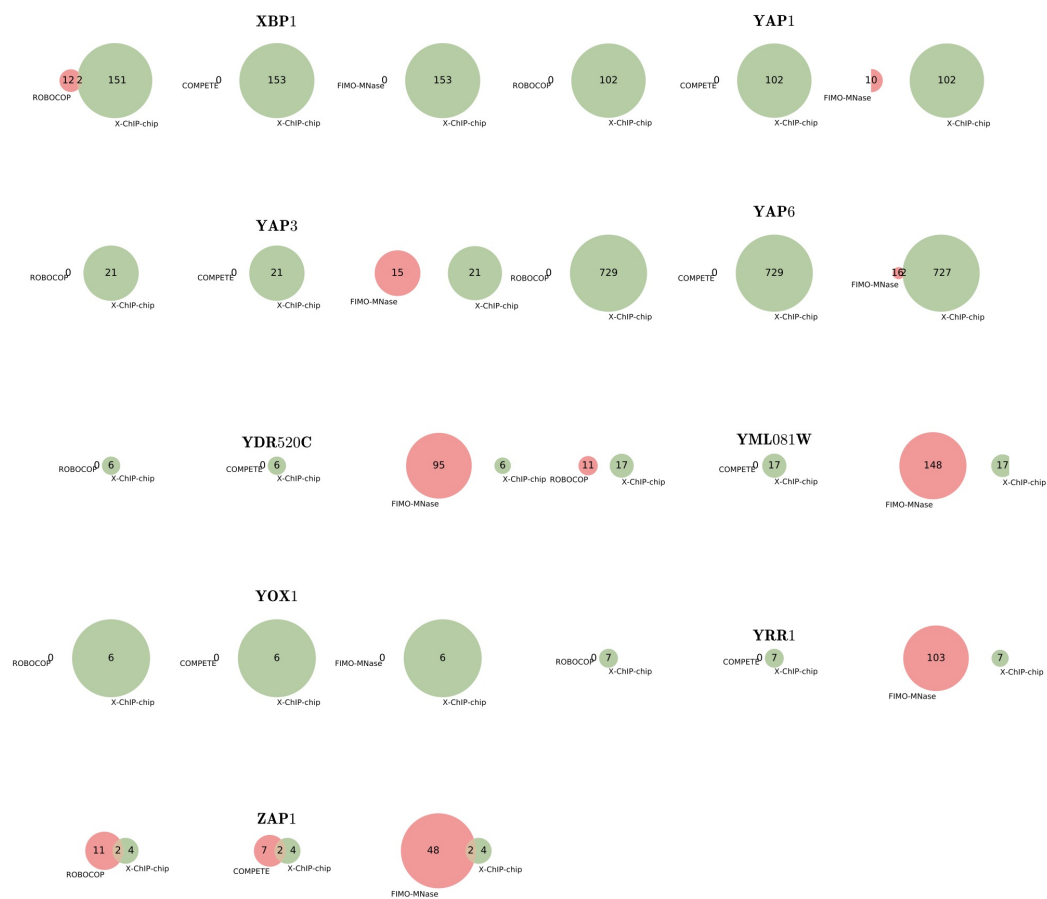


Figure S3: Venn digrams comparing the TF binding site predictions of RoboCOP, COMPETE, and FIMO-MNase with reported binding sites from X-ChIP-chip (MacIsaac *et al.*, 2006), ChIP-exo (Rhee and Pugh, 2011) and ORGANIC (Kasinathan *et al.*, 2014).

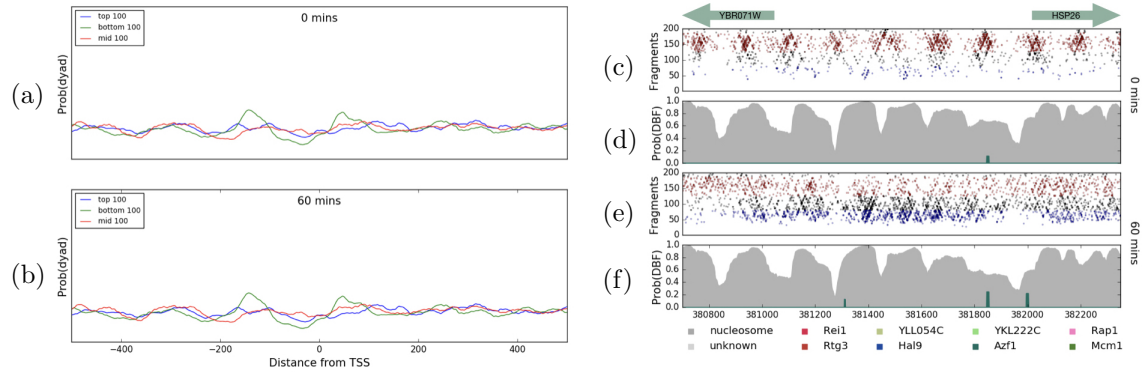


Figure S4: Aggregate of  $P(\text{dyad})$  computed by COMPETE around the TSS of genes that are most up-regulated (blue), most down-regulated (green) and no change in transcription (red) (a) before, and (b) after treating cells with cadmium. We find that COMPETE is unable to capture the chromatin dynamics around the TSS. MNase-seq midpoint plot and COMPETE predicted binding landscape of HSP26 promoter at chrII:380700-382350 (c) before and (e) after treatment with cadmium. As MNase-seq is not an input to COMPETE, it fails to capture the changes in chromatin under cadmium treatment.

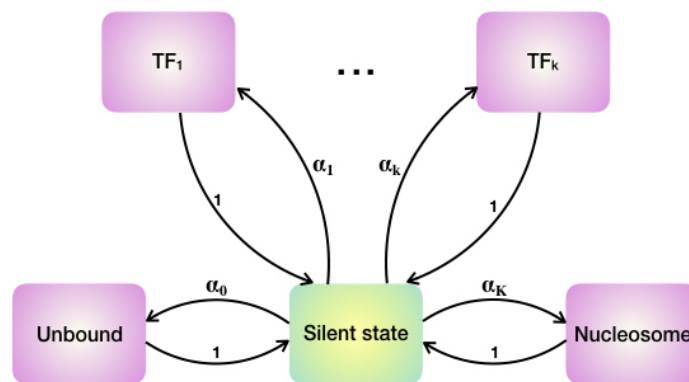


Figure S5: Transition diagram among the different DBFs. The DBFs are nucleosome, TFs and unbound state. The model transitions to a DBF  $k$  with probability  $\alpha_k$  and then comes back to the silent state with a probability of 1.

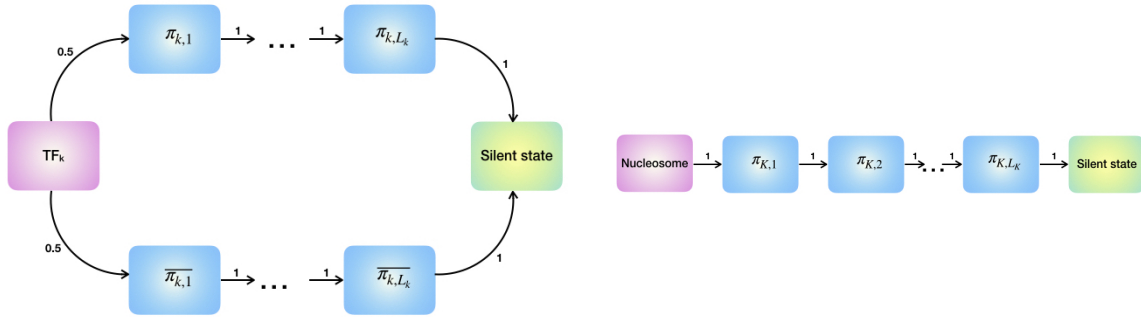
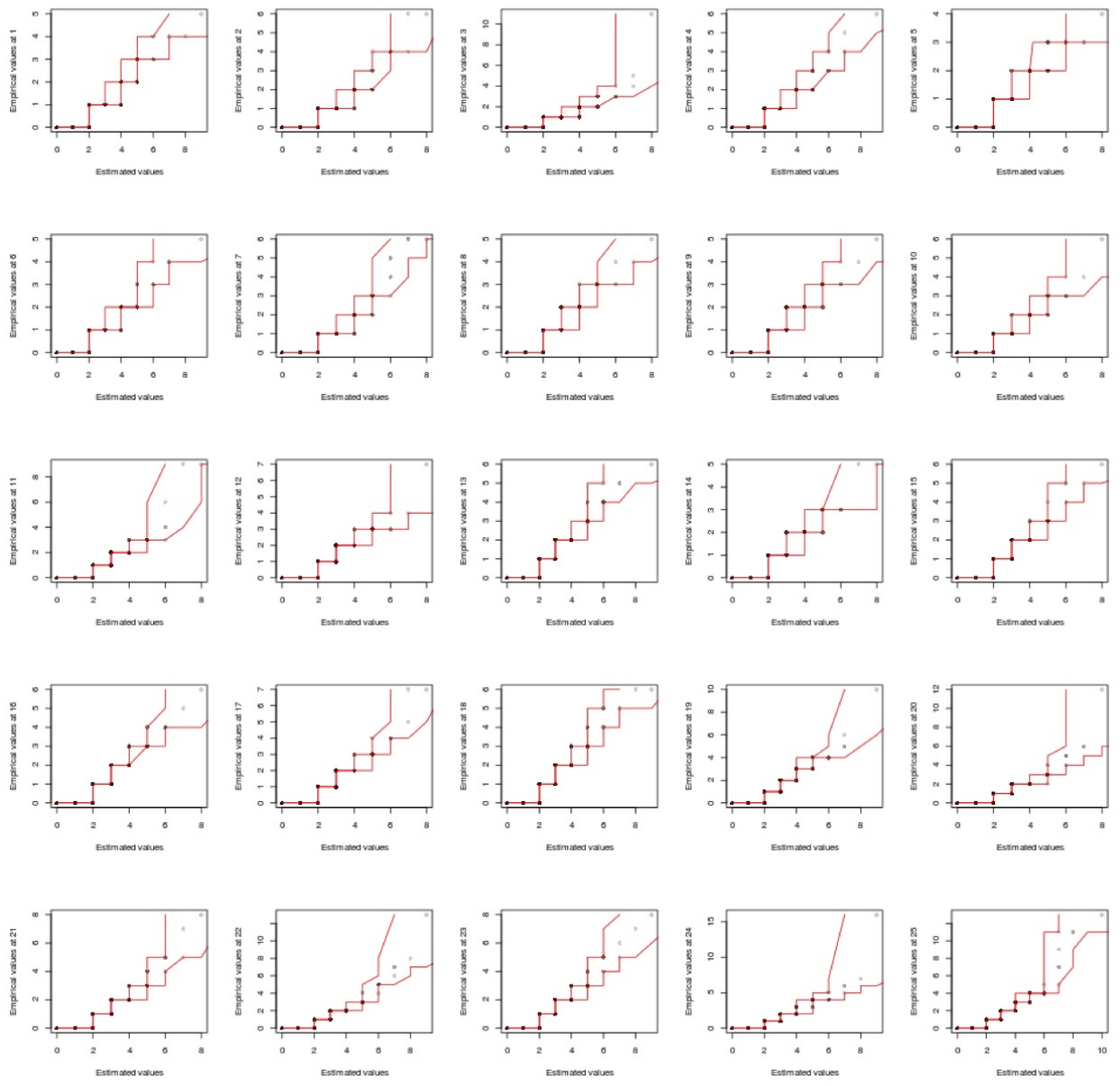
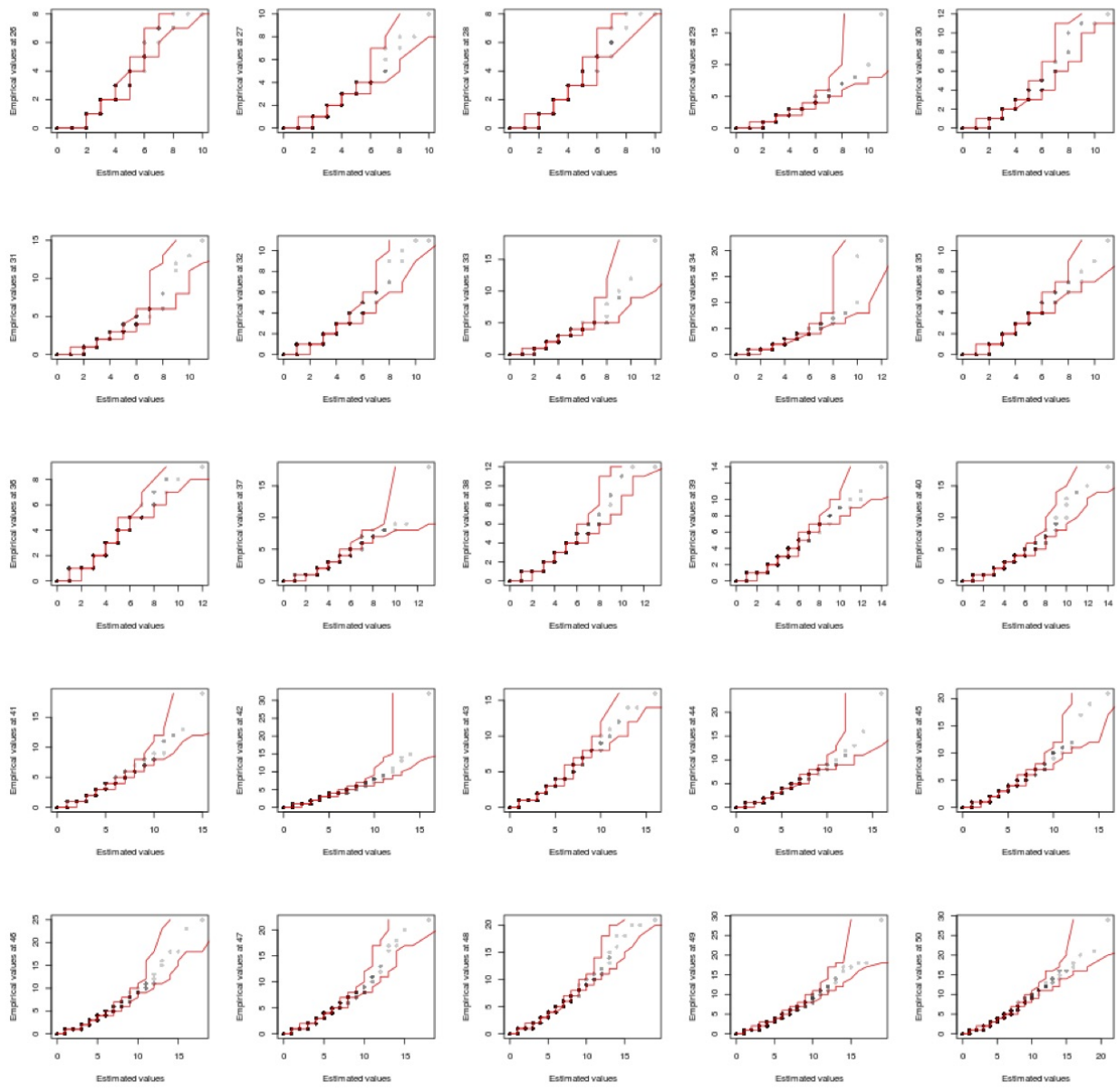


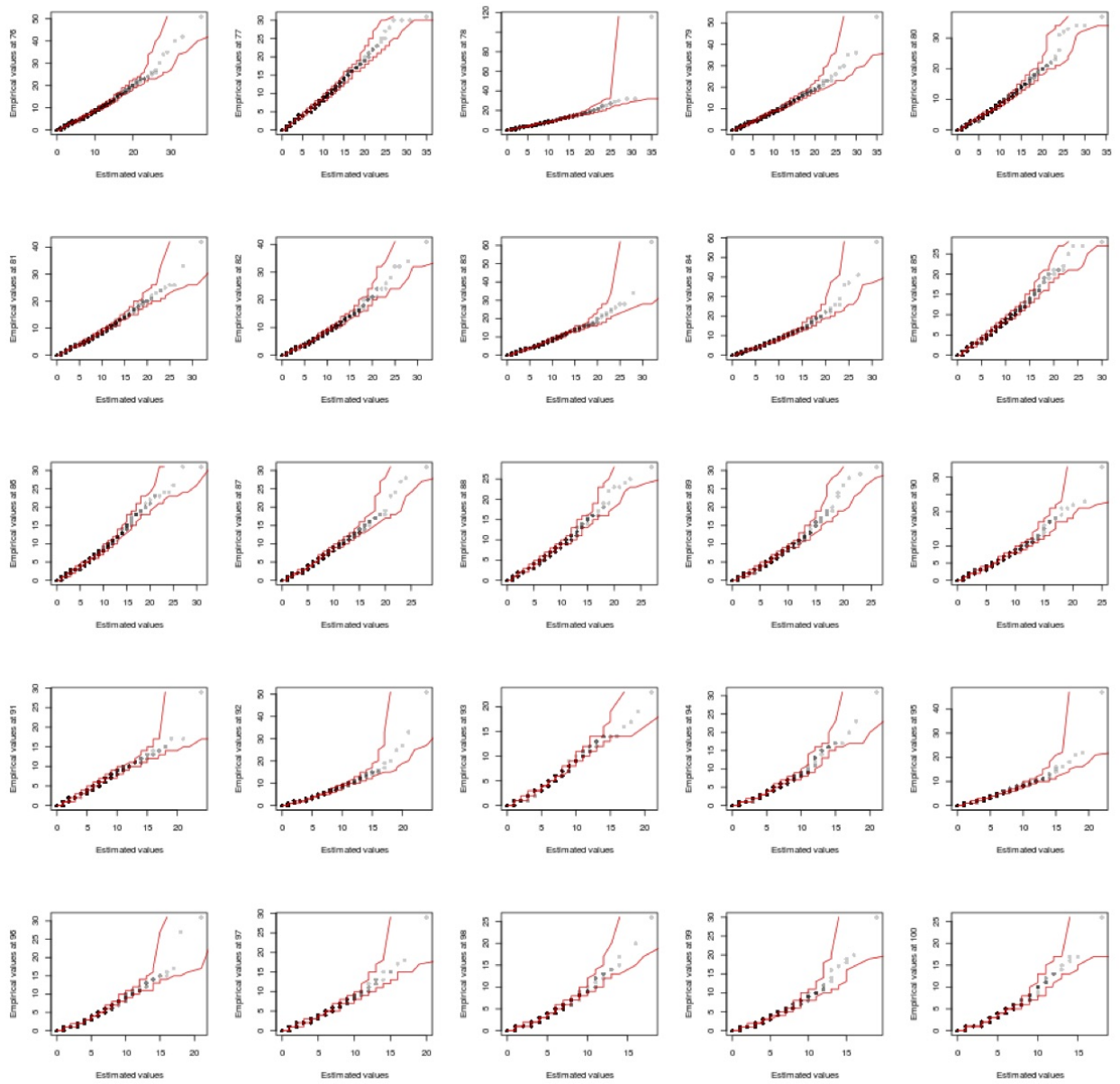
Figure S6: (a) State transition diagram of a TF. The model puts equal weight to transition into the motif and the reverse complement of the motif. After transitioning through the entire motif, it transitions back to the central silent state with probability of 1. (b) State transition diagram of the nucleosome model. Nucleosomes are modeled to be 147 bases long. After transitioning through the 147 nucleosome states the model transitions back to the central silent state.

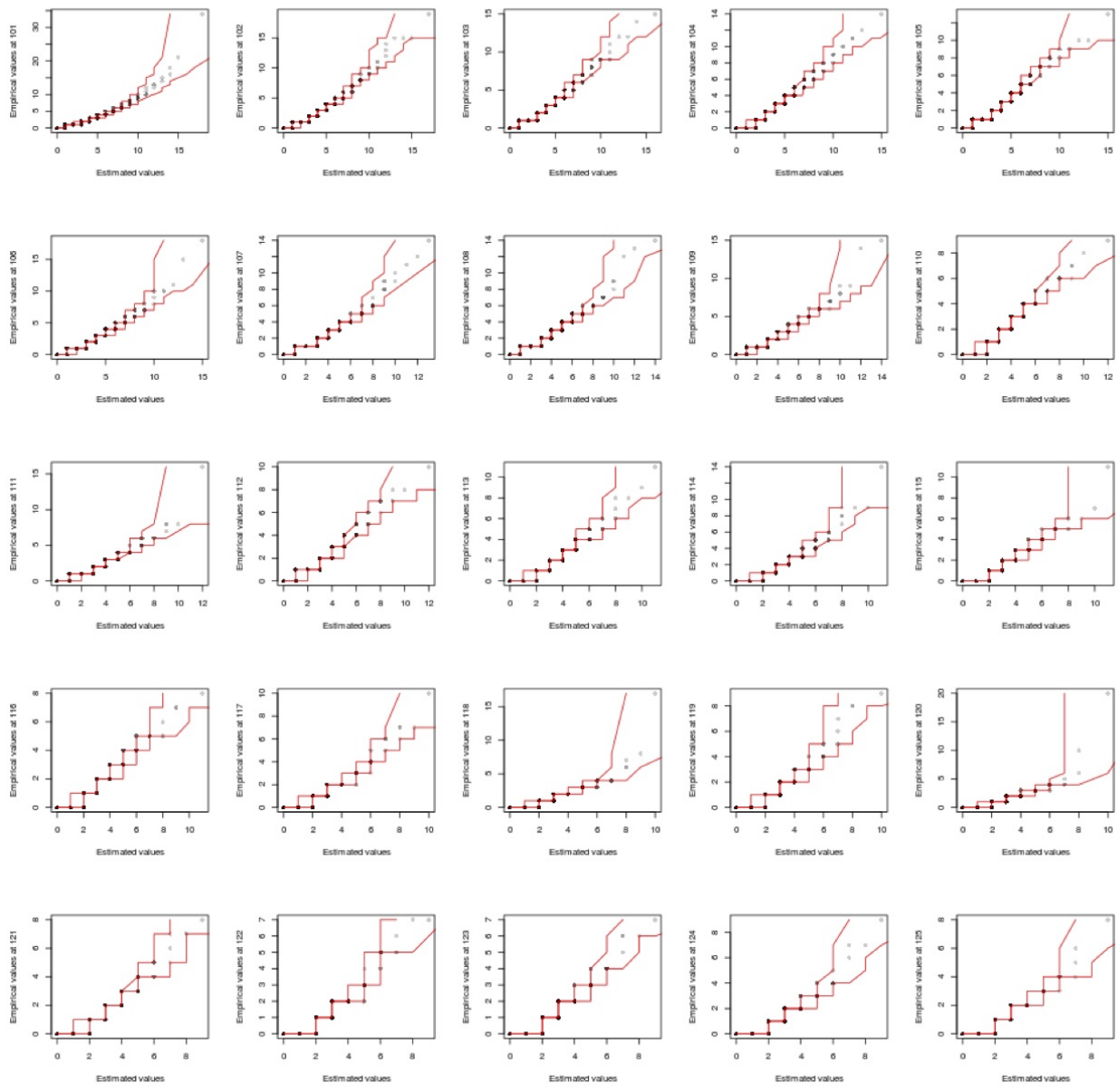












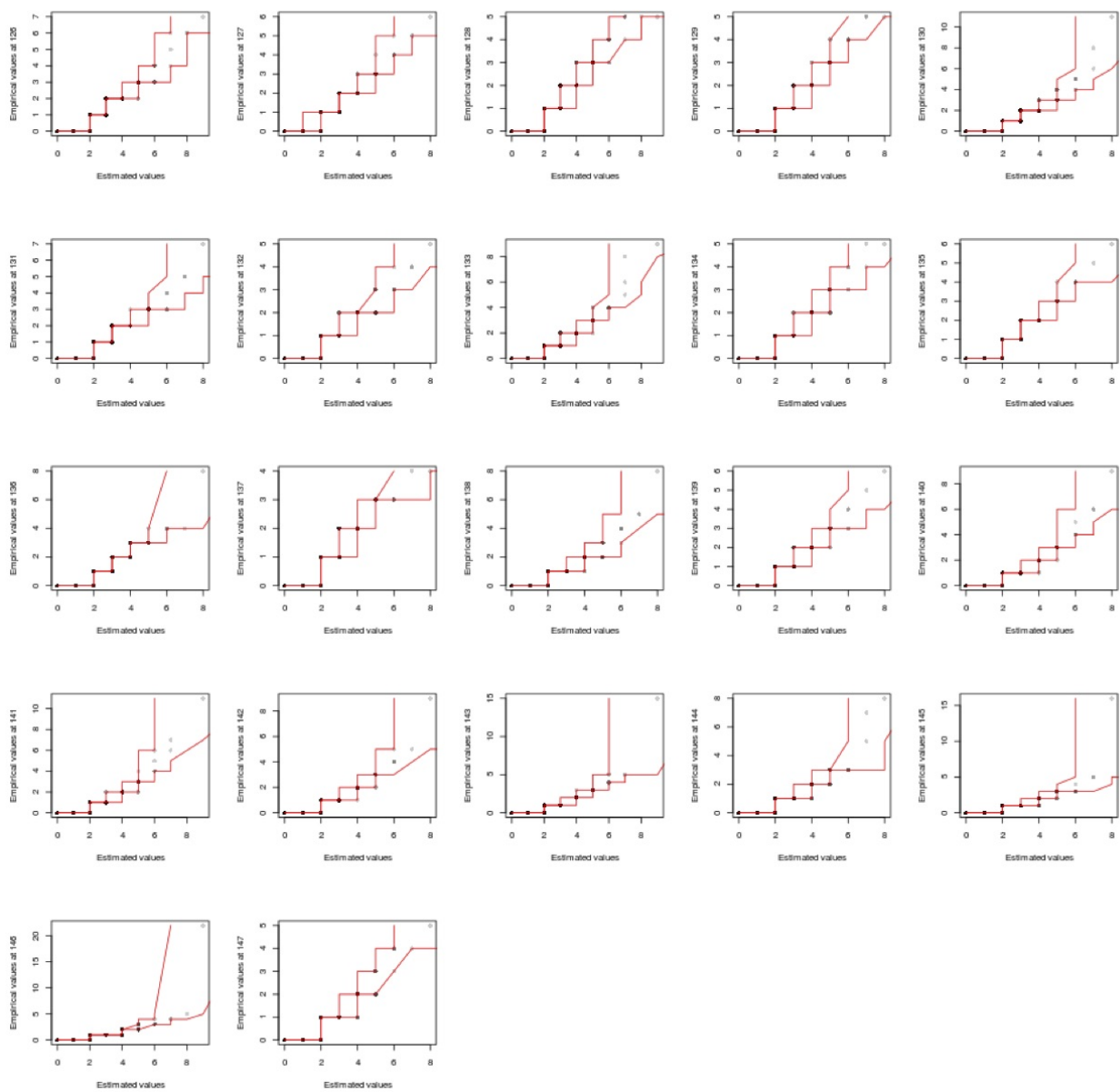


Figure S7: QQ-plots of negative binomial distributions for *nucleosomal fragment* midpoint counts of all 147 positions of nucleosome.

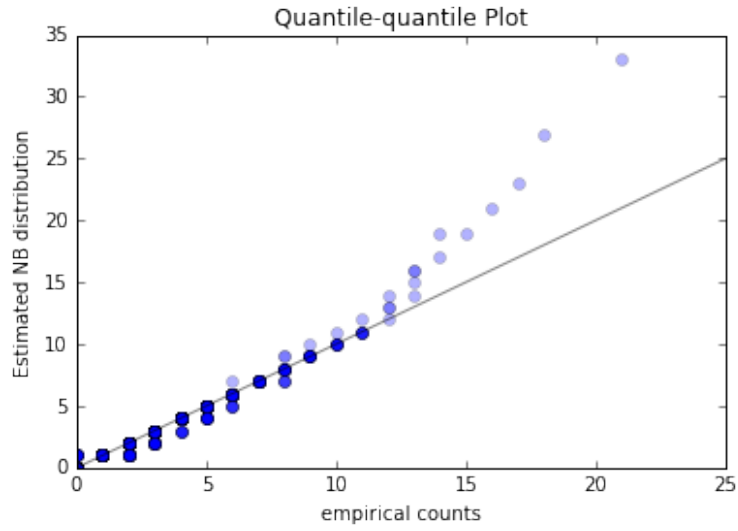


Figure S8: Quantile quantile plot of NB distribution of midpoint counts of *short factor fragments* to predict TF binding sites.

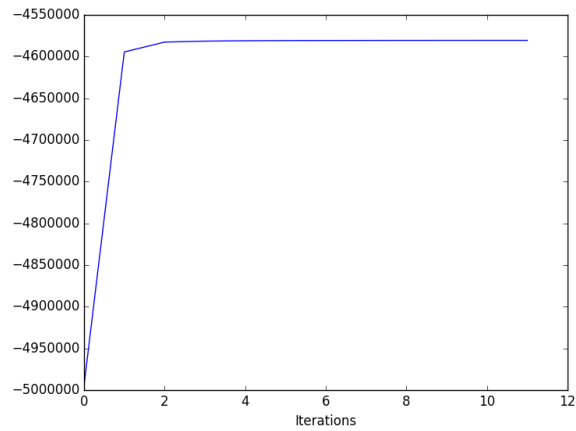
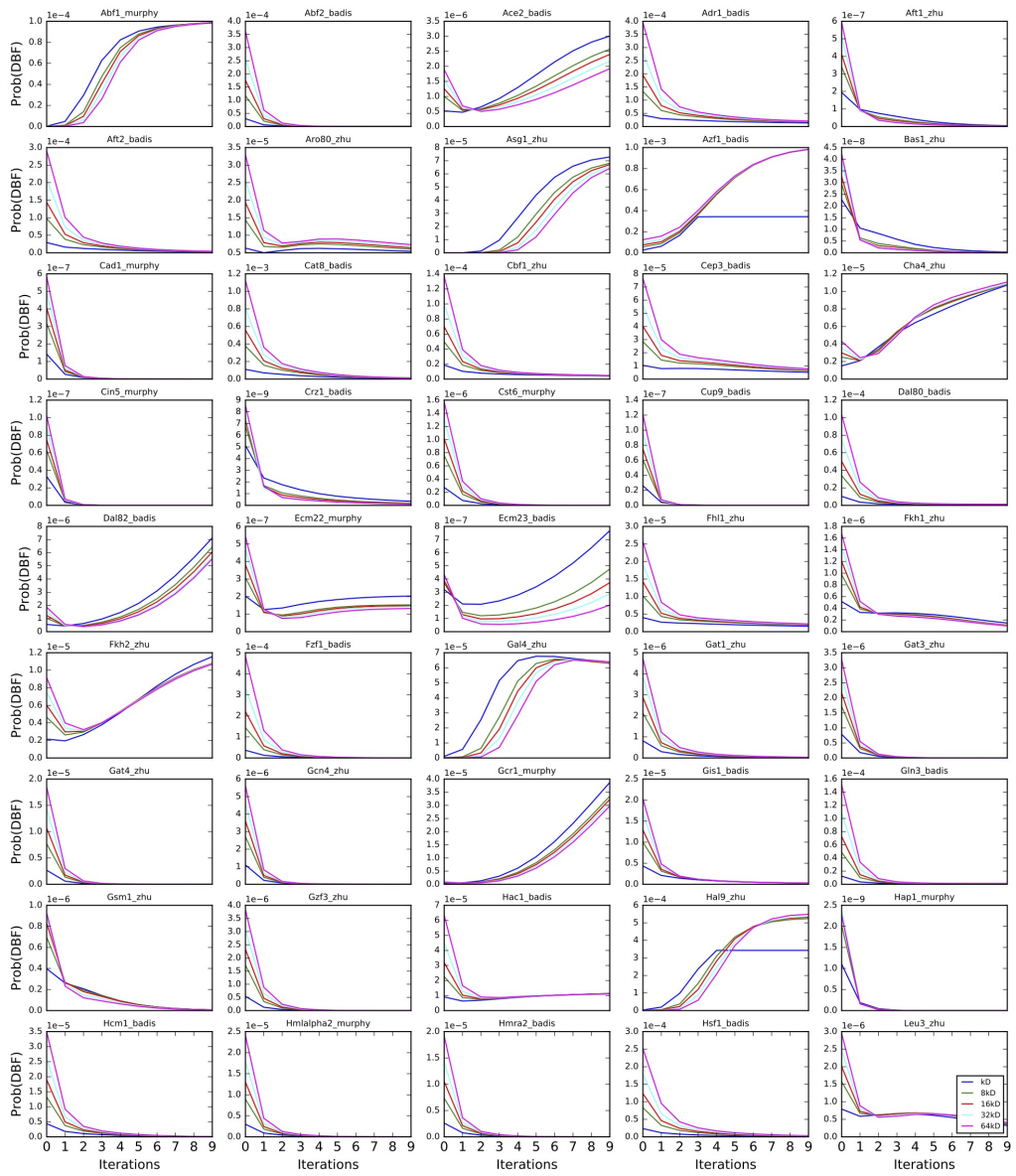
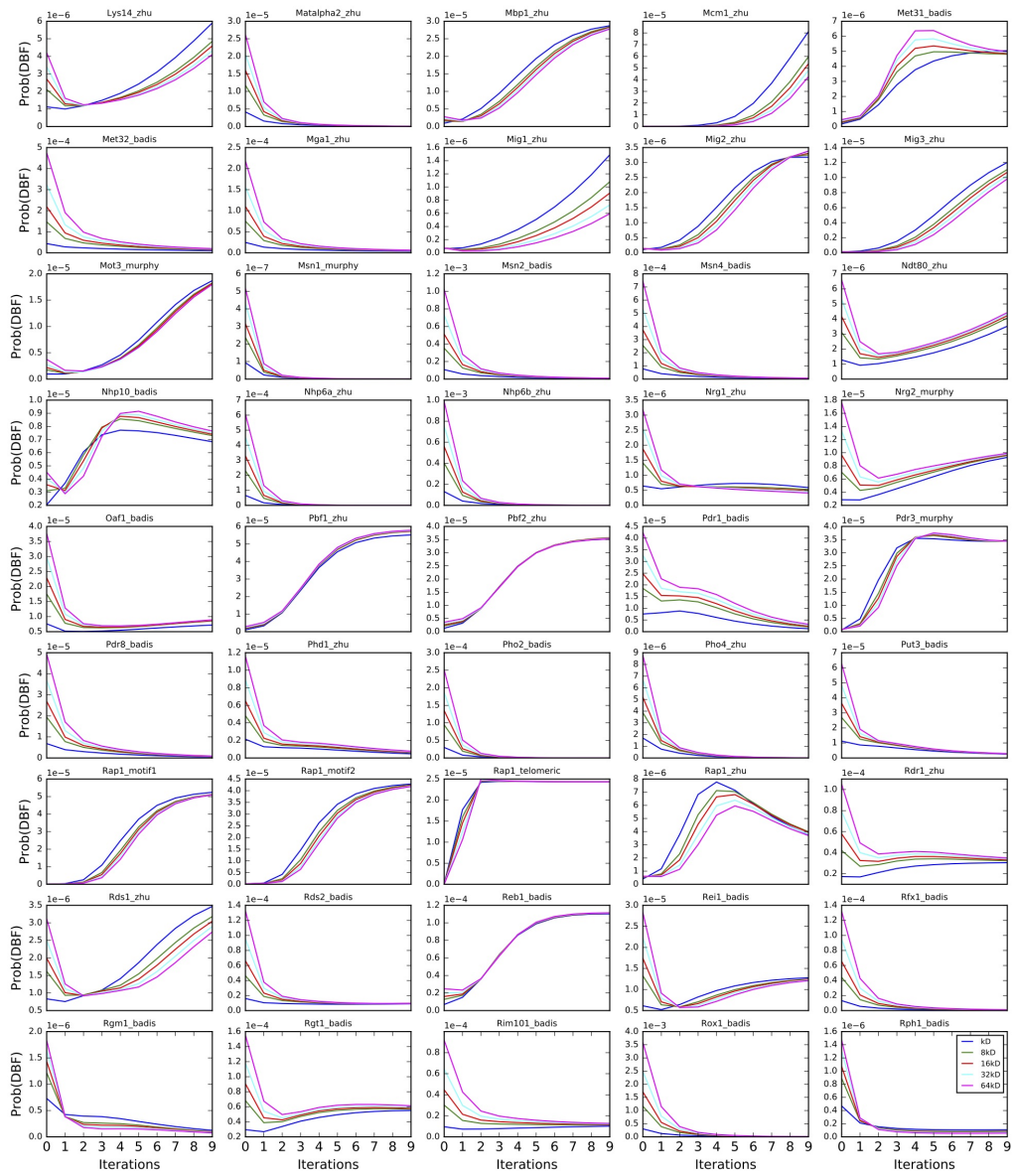
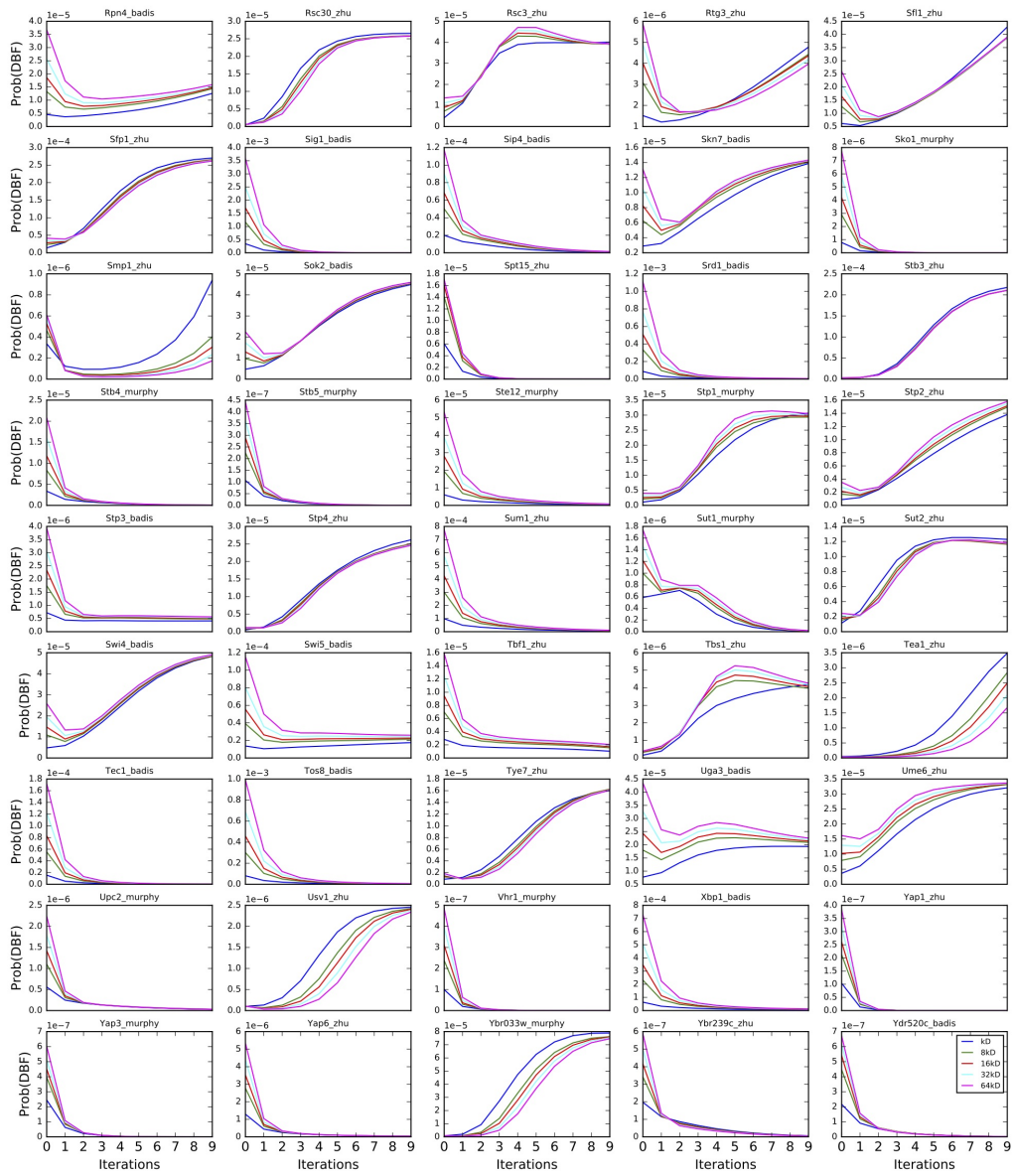


Figure S9: Likelihood in the 10 iterations of EM of RoboCOP.







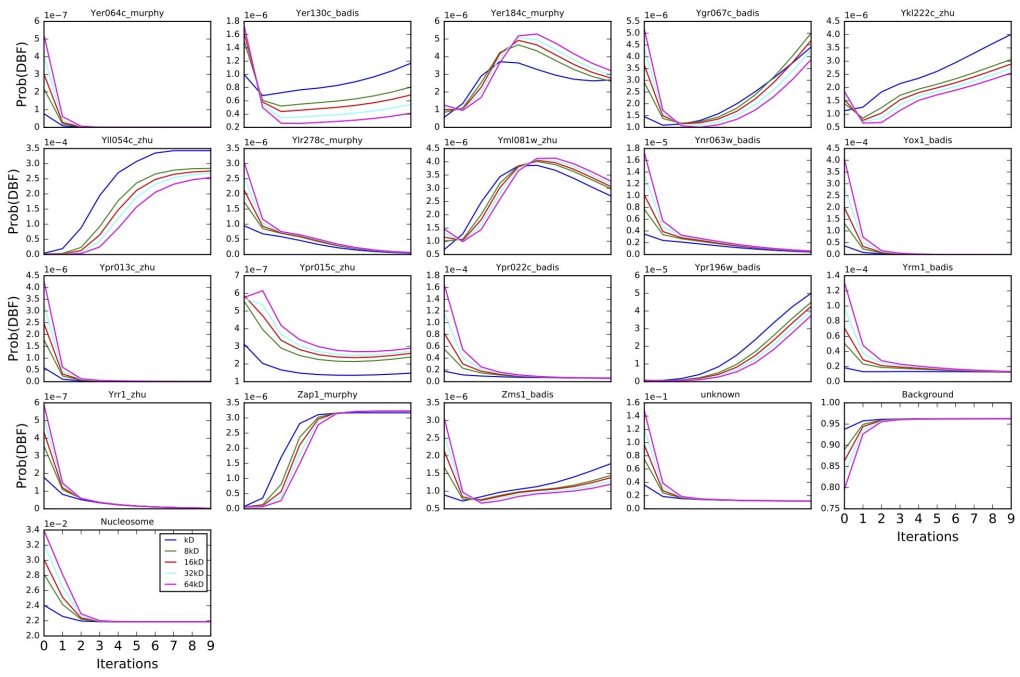


Figure S10: Transition weights of DBFs in every iteration of EM when for initializing using multiples of  $K_D$ .



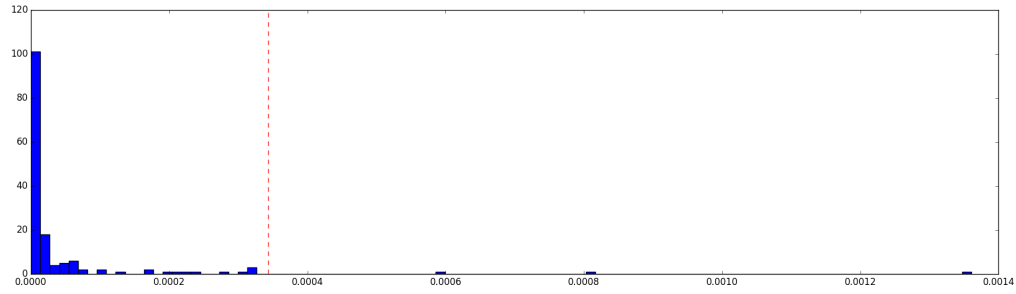


Figure S11: Choosing threshold for transition weight of TFs.

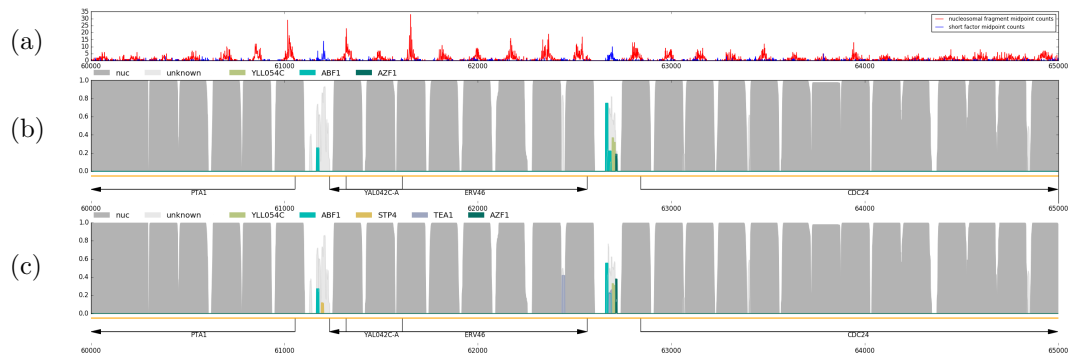


Figure S12: (a) MNase-seq fragment midpoint counts. Running RoboCOP (b) with and (c) without threshold cutoff of TF weights. In (c) we see more clustering of TFs means there are more false positives. RoboCOP gives fewer false positives when TF transition weights are updated with threshold.

## References

- 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. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**(7004), 99–104.
- Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K., and Henikoff, S. (2014). High-resolution mapping of transcription factor binding sites on native chromatin. *Nature Publishing Group*, **11**(2), 203–209.
- MacIsaac, K. D., Wang, T., Gordon, D. B., Gifford, D. K., Stormo, G. D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics*, **7**(1), 113.
- Rhee, H. S. and Pugh, B. F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, **147**(6), 1408–1419.