

**Figures S1, S2, S3. Output of MuMoD compared with MEME, Weeder, PeakMotifs, and ChipMunk.**

Figure S1 shows results for proteins profiled in HL1, (A) p300, (B) Tbx5, (C) Srf, (D) Nkx2-5, (E) Gata4, and (F) Mef2. Figure S2 shows results for Gata3 profiled in (A) CD4, (B) CD8, (C) DN, (D) DP, (E) NKT, (F) Th1, (G) Th17, (H) Th2, (I) nTreg. Figure S3 shows results for (A) cohesin subunit Smc1a in ES, (B) mediator subunit Med12 in ES, (C) mediator subunit Med12 in MEF, (D) Smc1a and Med12 in ES, and (E) Sox2 and CTCF in ES.

**Figure S4. ChIP enrichment of sequences without a motif in the p300 set**

(A) Density of three modes identified in the p300 set along the sequences sorted by ChIP enrichment is illustrated in red, blue and, black, respectively. The density of sequences with no mode is shown in magenta. (B) Boxplots of the ChIP enrichment score or peak heights are shown for each mode. Both figures indicate that sequences without a motif are more likely to have a lower ChIP enrichment.

**Figure S5. Proximity of modes in mediator set to TSS.**

Five modes are identified in the Med12-bound sequences in ES cells. The boxplot of distances to the closest TSS of the sequences in each mode are shown in red, which all other modes is in blue. Modes 1 and 2 are more enriched close to TSS, while the other three modes are more likely to be farther away.

**Table S1. The variation in the number of identified modes with  $\lambda$**

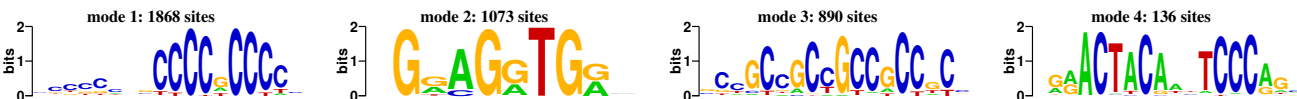
The number of optimal modes changes with the value of  $\lambda$  that is used to penalize complex models. A large value  $\geq 20$  gives at the most two modes. A low value of 1 typically returns more modes. Since the maximum number of modes allowed was 6, we do not see numbers greater than that.





## S1(B) Tbx5 in HL1; top 5000 sequences

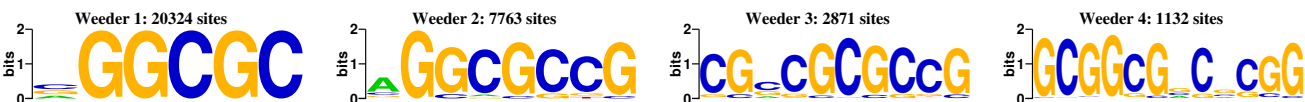
## MuMoD



## MEME



## WEEDER



## PEAKMOTIFS



## CHIPMUNK

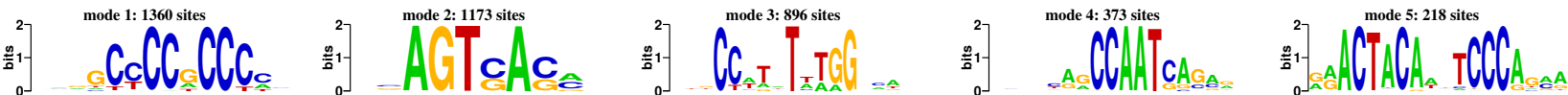


Literature consensus

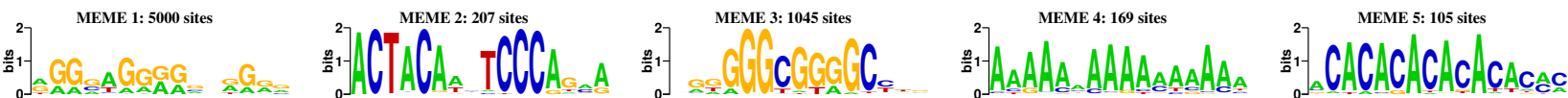


S1(C) Srf in HL1; top 5000 sequences

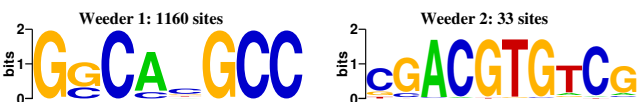
MuMod



MEME



WEEDER



PEAKMOTIFS



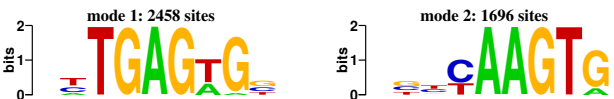
CHIPMUNK



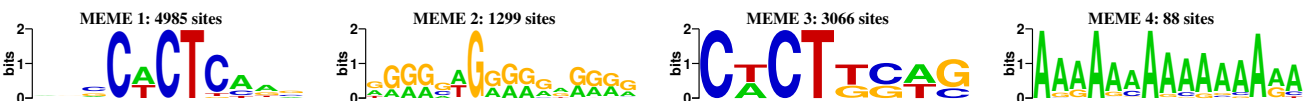


## S1(D) Nkx2-5 in HL1; top 5000 sequences

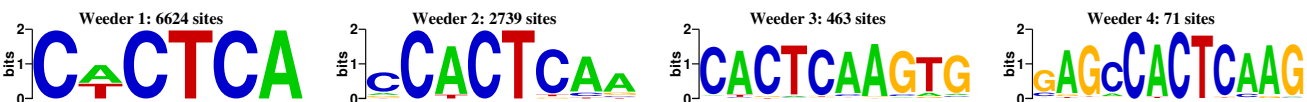
## MuMoD



## MEME



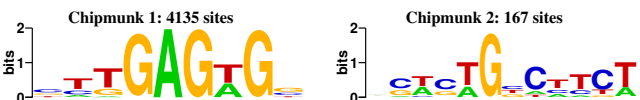
## WEEDER



## PEAKMOTIFS

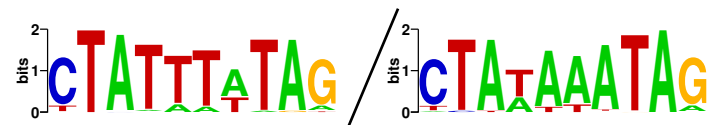


## CHIPMUNK





Literature consensus

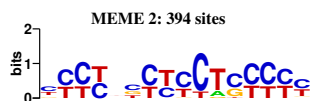
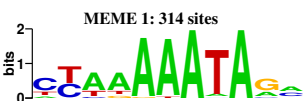


S1(F) Mef2 in HL1; 1248 sequences

MuMoD



MEME



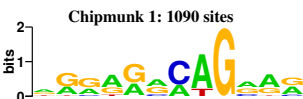
WEEDER



PEAKMOTIFS



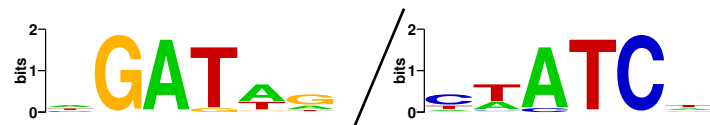
CHIPMUNK







Literature consensus

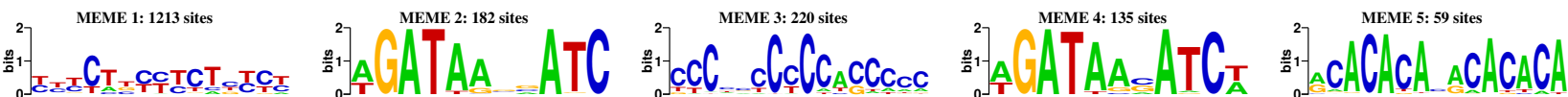


## S2(B) Gata3 in CD8; 1213 sequences

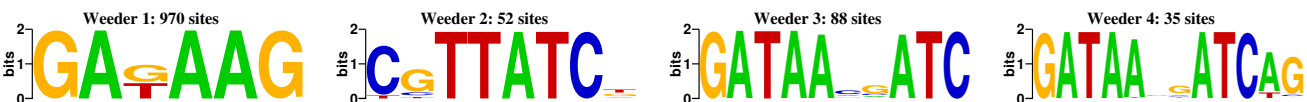
### MuMod



### MEME



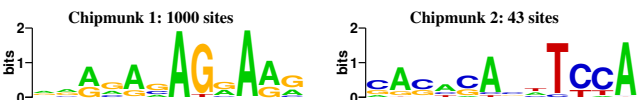
### WEEDER



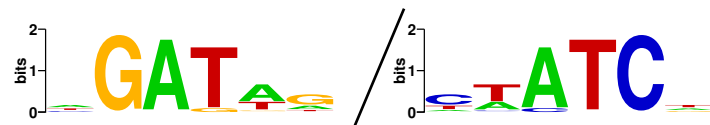
### PEAKMOTIFS



### CHIPMUNK

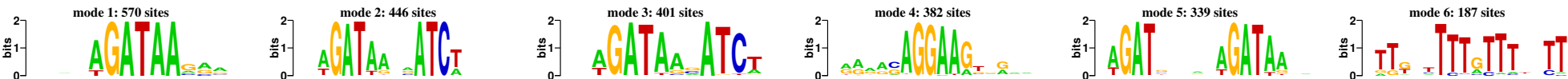


Literature consensus

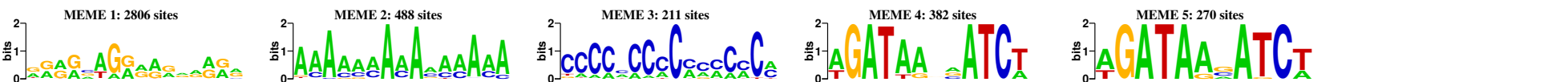


S2(C) Gata3 in DN; 2807 sequences

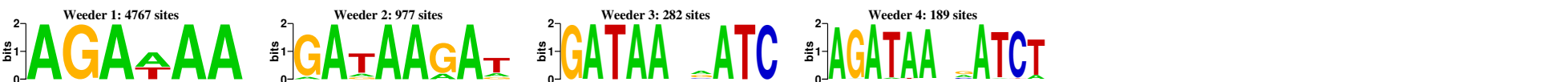
MuMod



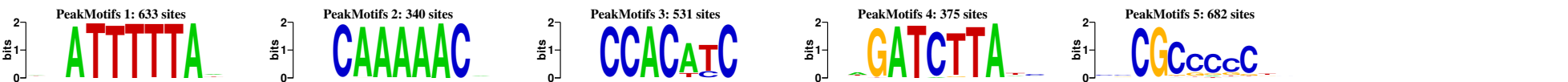
MEME



WEEDER

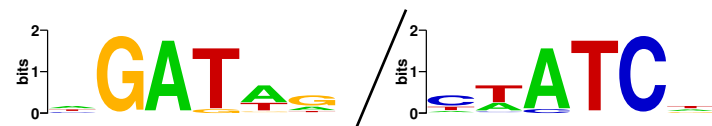


PEAKMOTIFS



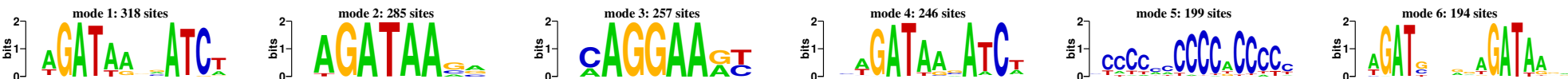
CHIPMUNK



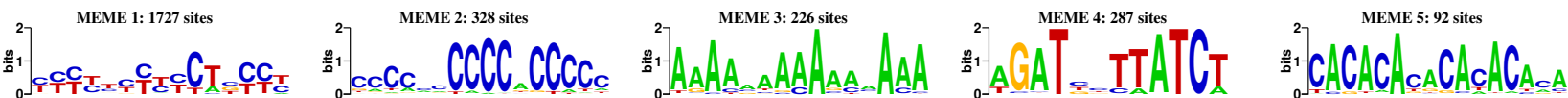


## S2(D) Gata3 in DP; 1736 sequences

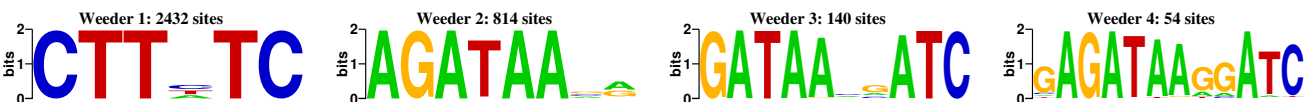
## MuMoD



## MEME



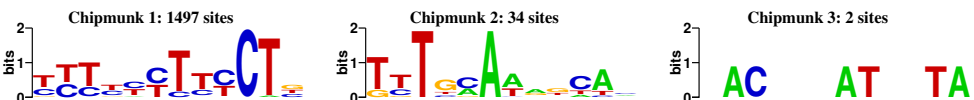
## WEEDER

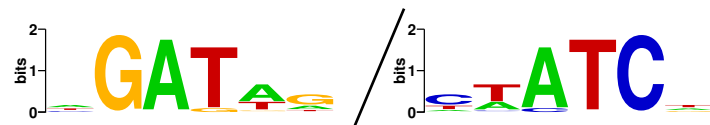


## PEAKMOTIFS



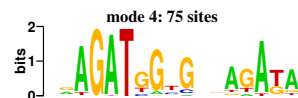
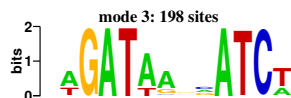
## CHIPMUNK



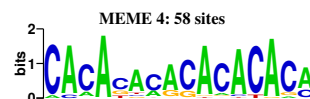
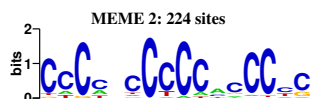
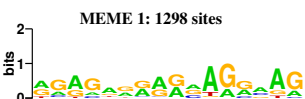


## S2(E) Gata3 in NKT; 1298 sequences

## MuMoD



## MEME



## WEEDER



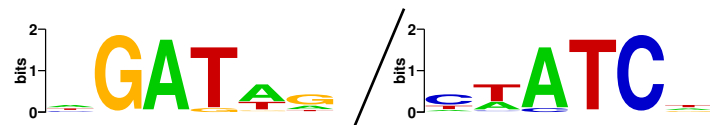
## PEAKMOTIFS



## CHIPMUNK

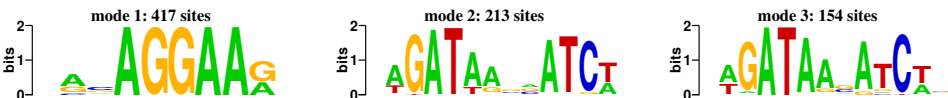


Literature consensus



## S2(F) Gata3 in Th1; 921 sequences

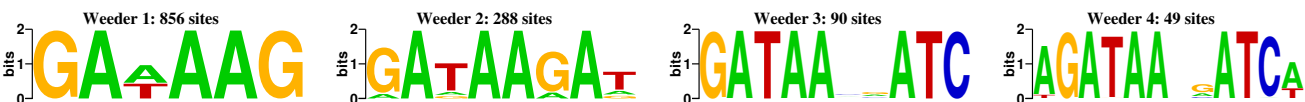
### MuMod



### MEME



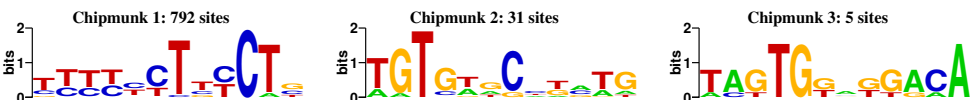
### WEEDER



### PEAKMOTIFS



### CHIPMUNK



Literature consensus

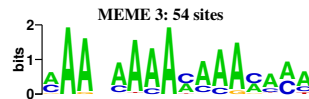
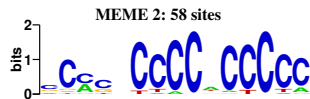
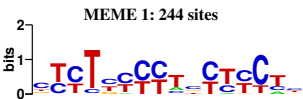


## S2(G) Gata3 in Th17; 250 sequences

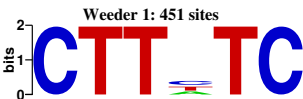
### MuMod



### MEME



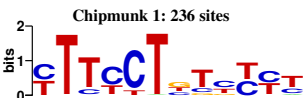
### WEEDER

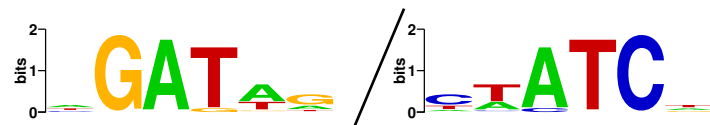


### PEAKMOTIFS



### CHIPMUNK



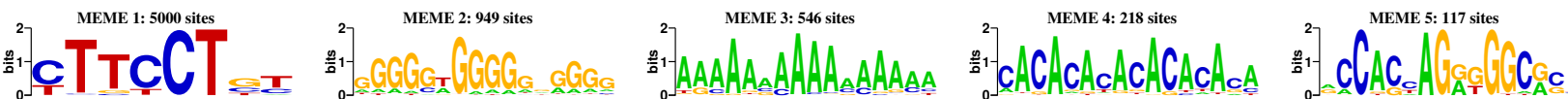


## S2(H) Gata3 in Th2; top 5000 sequences

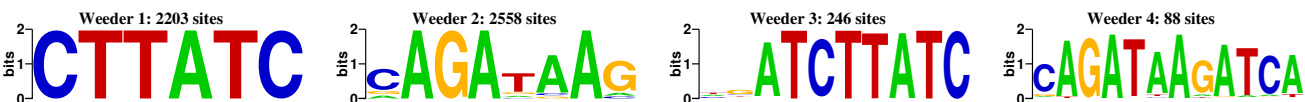
## MuMoD



## MEME



## WEEDER

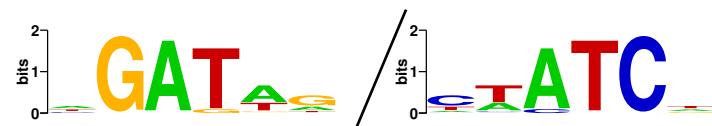


## PEAKMOTIFS



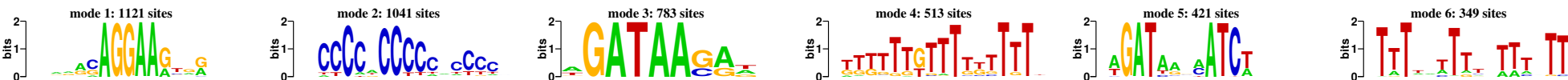
## CHIPMUNK



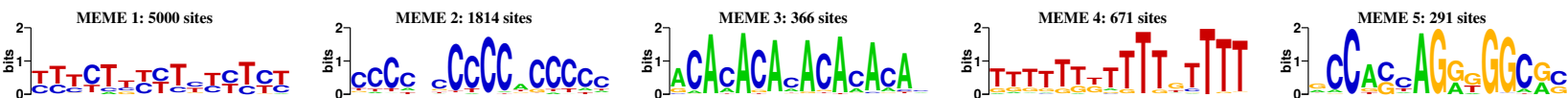


## S2(I) Gata3 in iTreg; top 5000 sequences

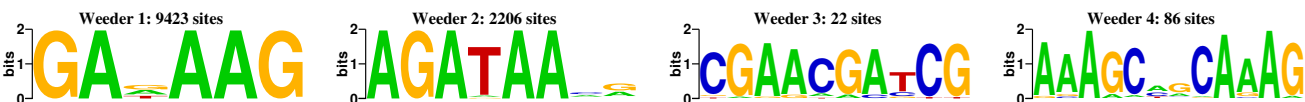
## MuMod



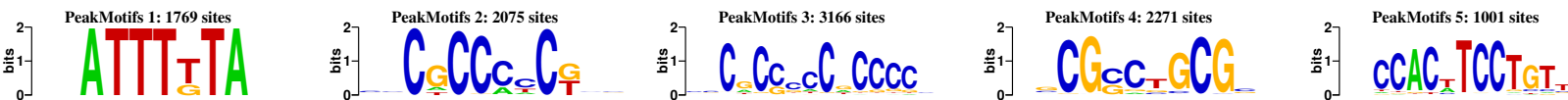
## MEME



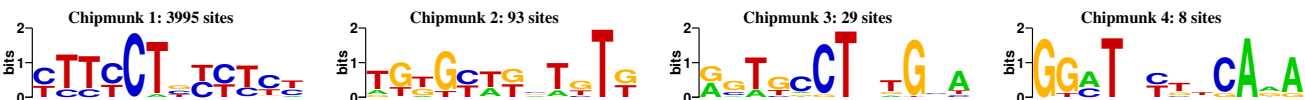
## WEEDER



## PEAKMOTIFS

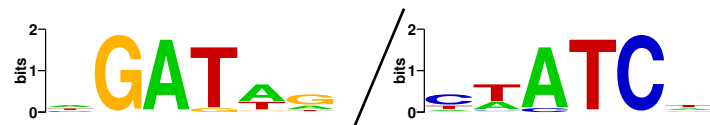


## CHIPMUNK



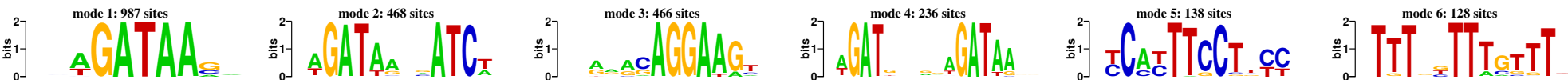


Literature consensus

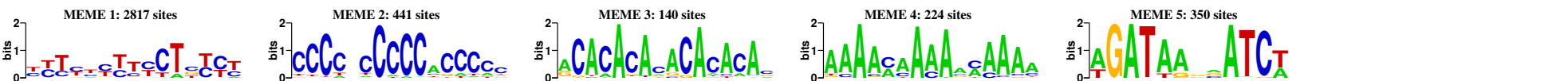


## S2(J) Gata3 in nTreg; 2817 sequences

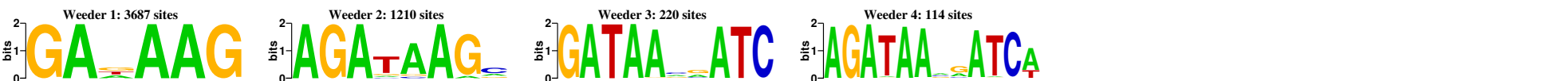
### MuMod



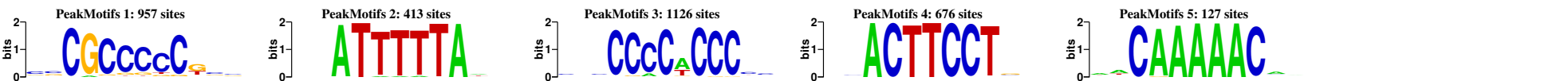
### MEME



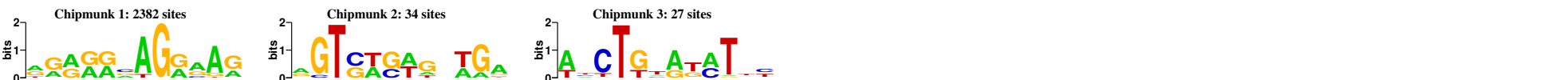
### WEEDER



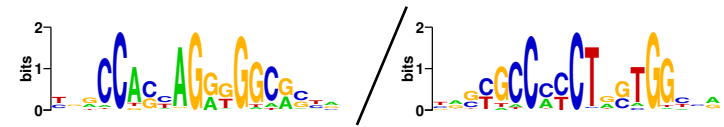
### PEAKMOTIFS



### CHIPMUNK

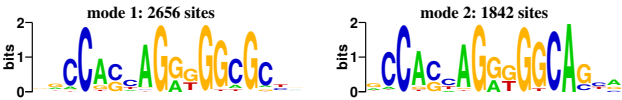


Literature consensus for CTCF

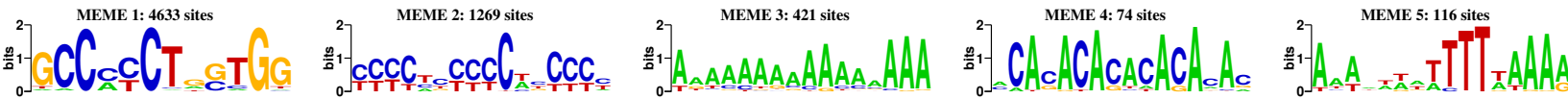


S3(A) Cohesin subunit Smc1a in ES; top 5000 sequences

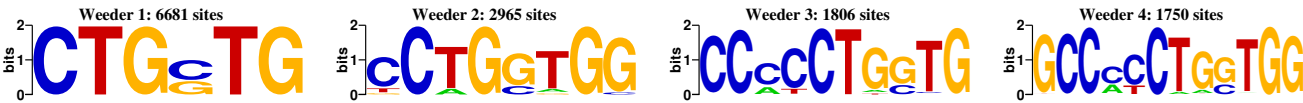
MuMod



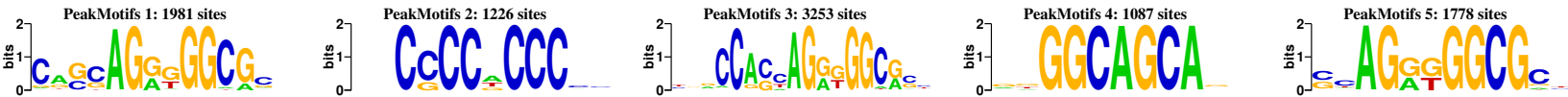
MEME



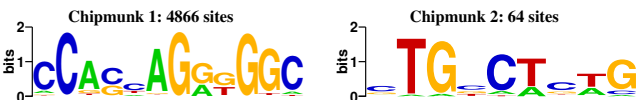
WEEDER



PEAKMOTIFS



CHIPMUNK

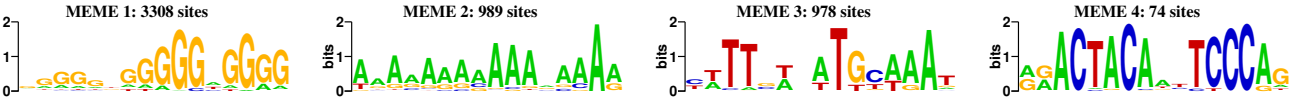


# S3(B) Mediator subunit Med12 in ES; top 5000 sequences

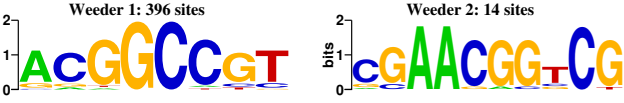
## MuMoD



## MEME



## WEEDER



## PEAKMOTIFS



## CHIPMUNK

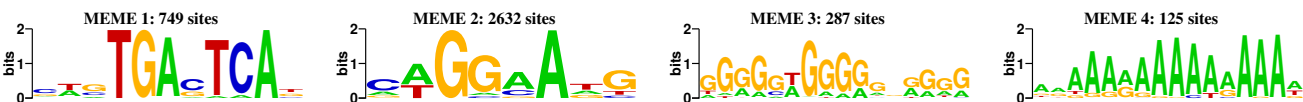


## S3(C) Mediator subunit Med12 in MEF; 2632 sequences

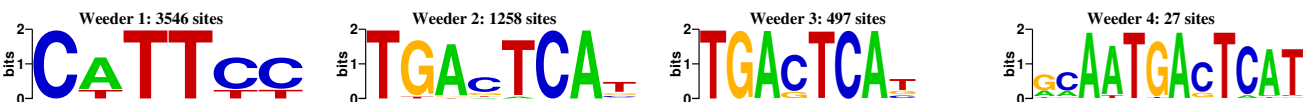
### MuMoD



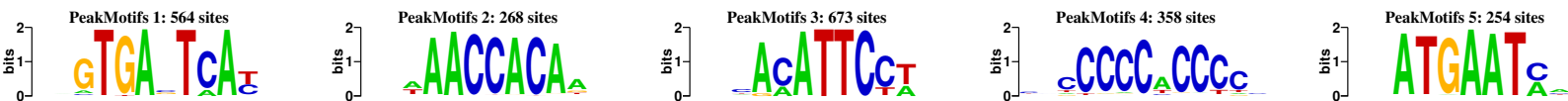
### MEME



### WEEDER



### PEAKMOTIFS

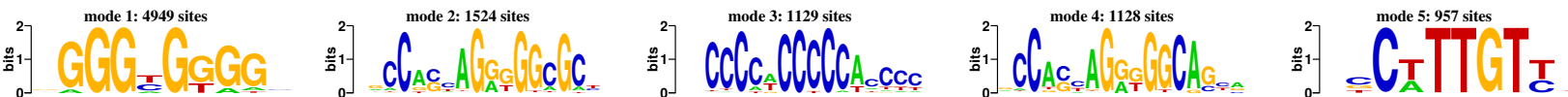


### CHIPMUNK

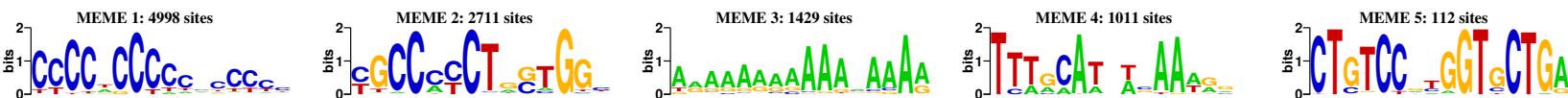


## S3(D) Smc1a and Med12 *both* in ES; 11865 sequences

### MuMoD



### MEME



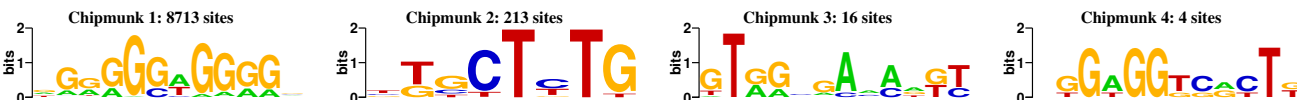
### WEEDER



### PEAKMOTIFS

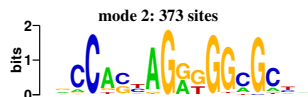
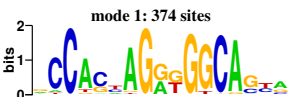


### CHIPMUNK

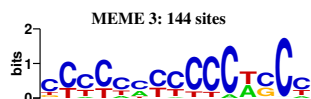
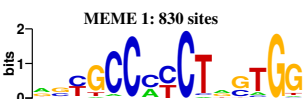


## S3(E) Sox2 and CTCF *both* in ES; 1035 sequences

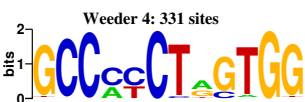
### MuMoD



### MEME



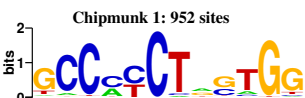
### WEEDER



### PEAKMOTIFS

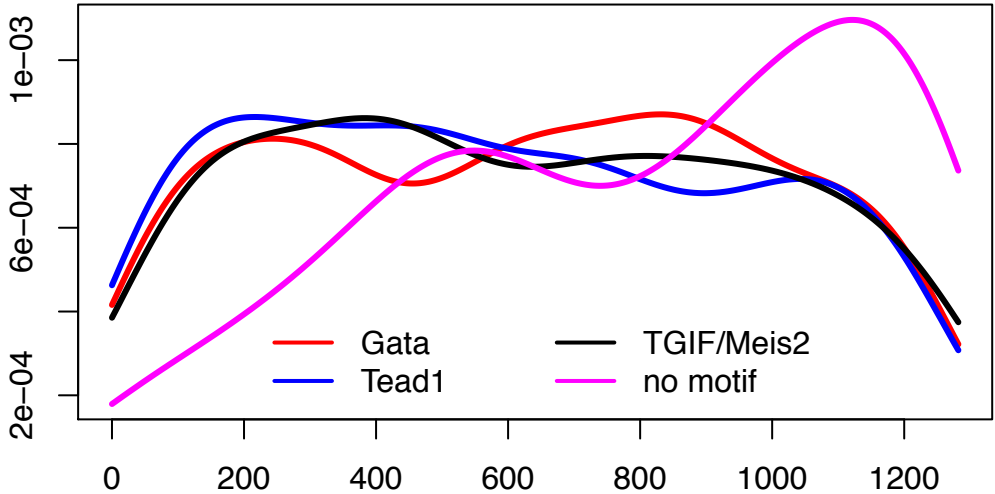


### CHIPMUNK

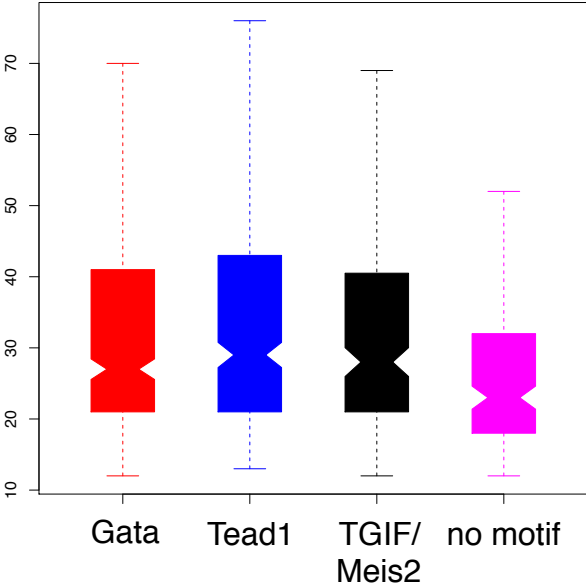


### S4. ChIP enrichment of sequences without a motif in the p300 set

(A) Density of modes



(B) ChIP enrichment of modes



# S5. Proximity of modes in mediator set to TSS

