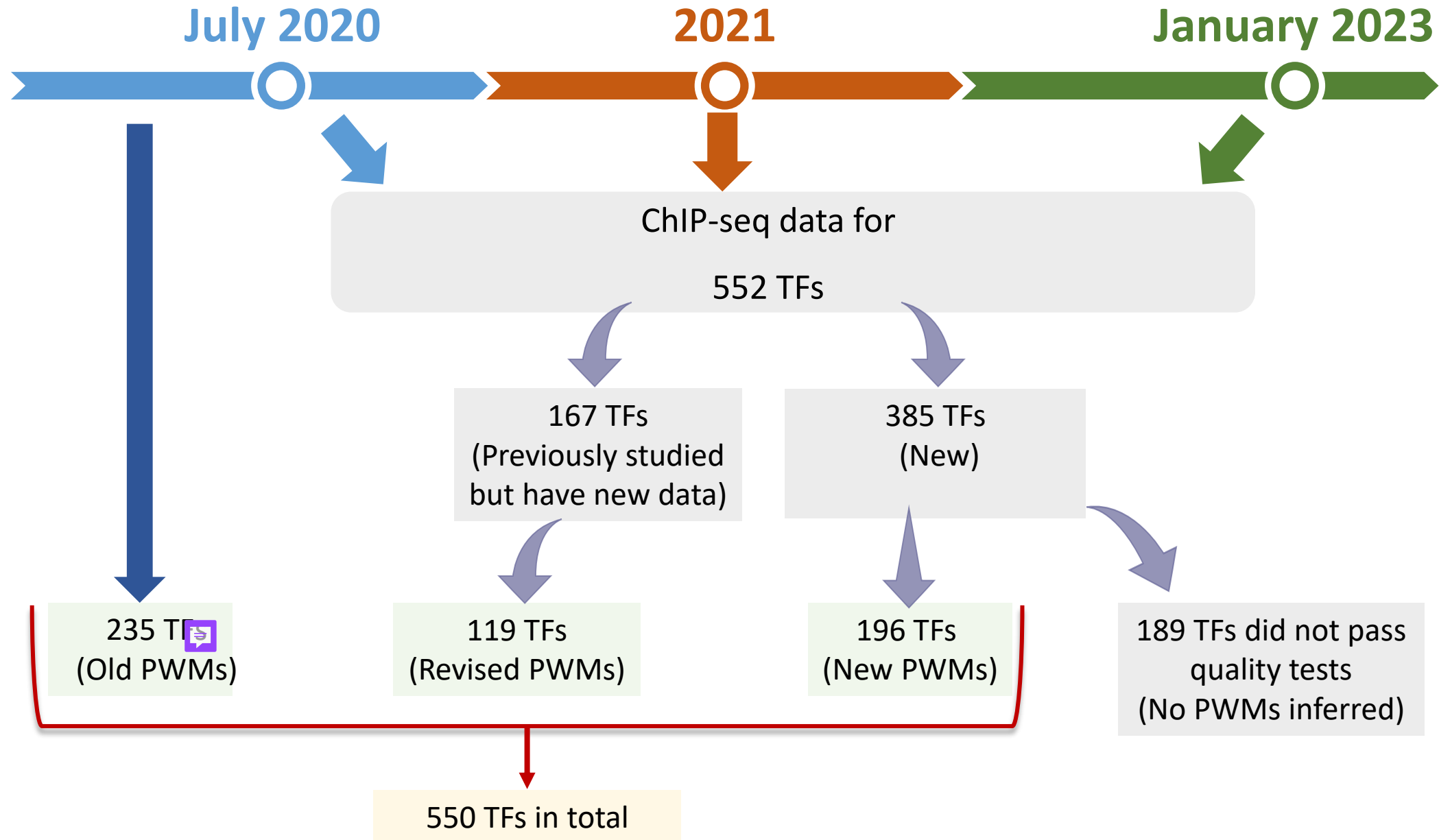


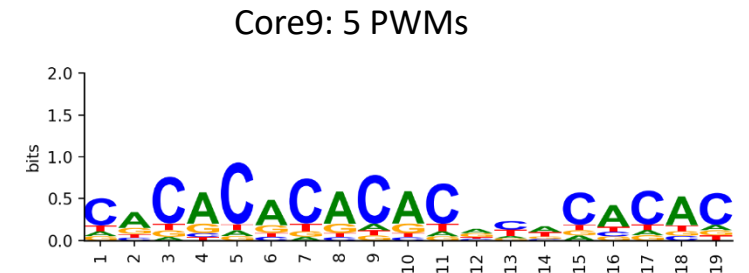
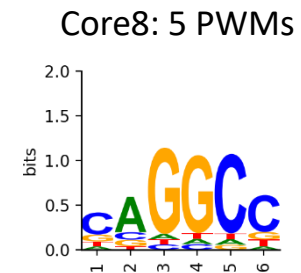
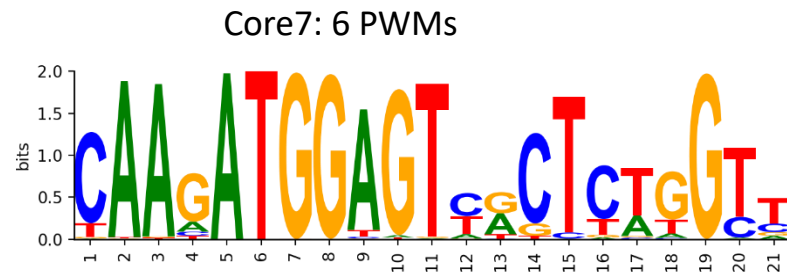
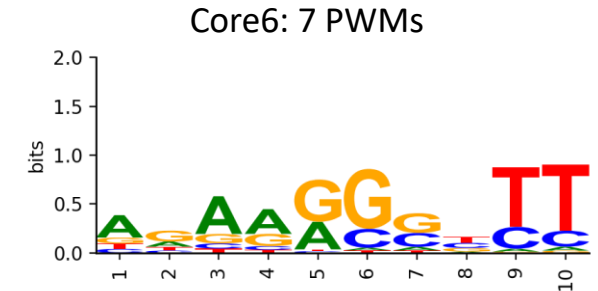
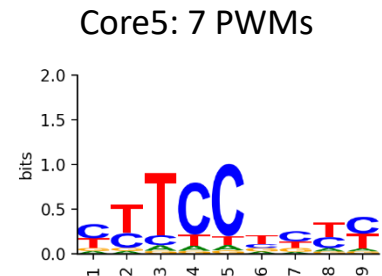
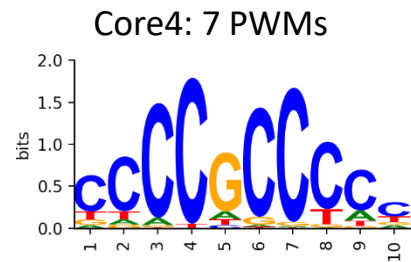
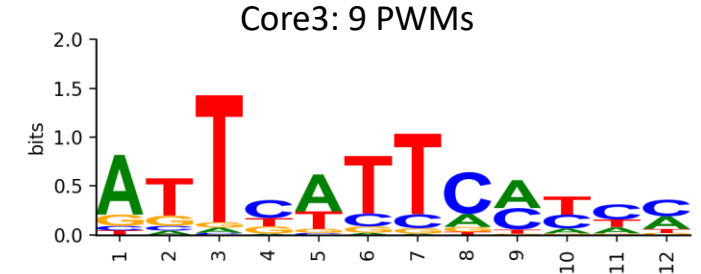
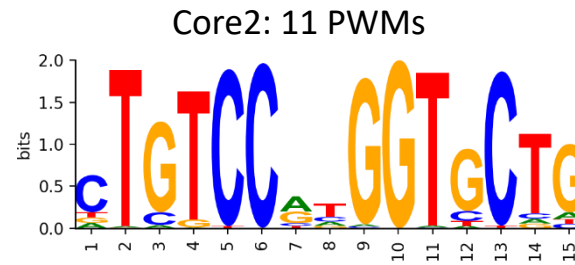
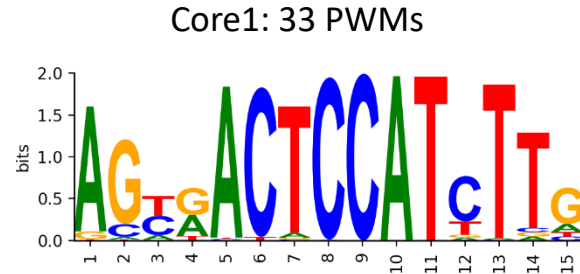
S1 Fig: analysis flowchart



S2 Fig: Core motifs

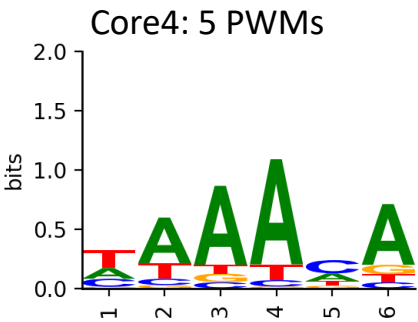
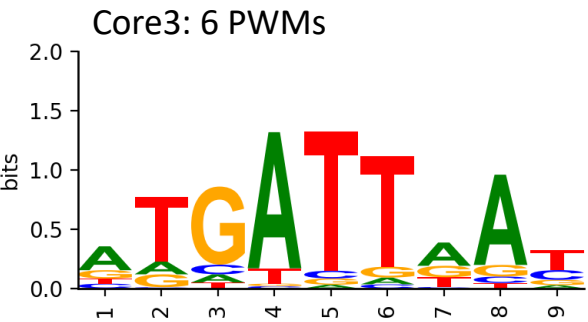
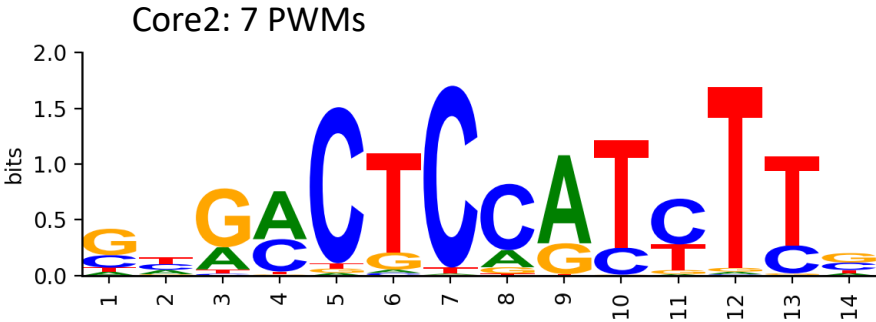
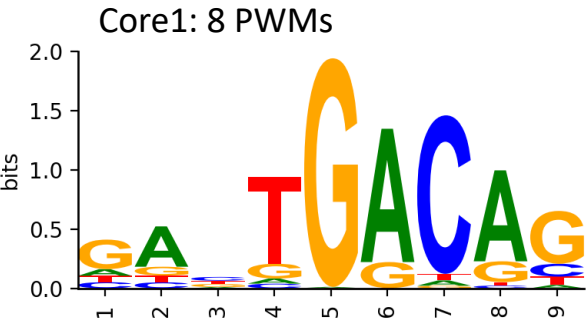
C2H2-ZF

- Total of canonical PWMs: 244



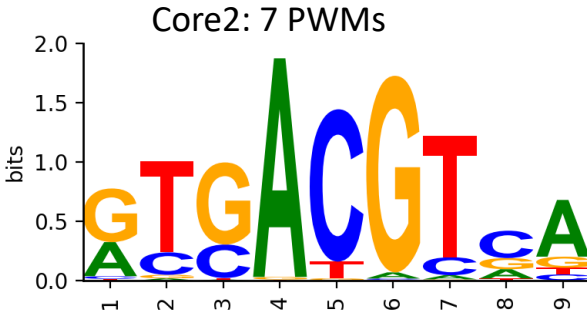
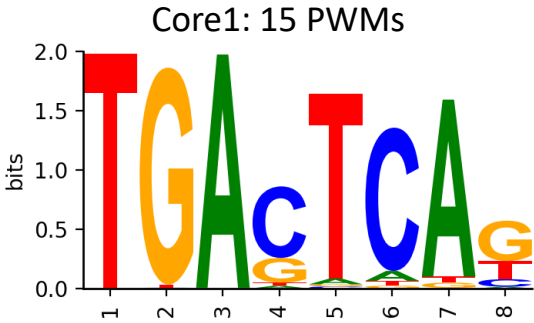
Homeodomain

- Total of canonical PWMs: 27

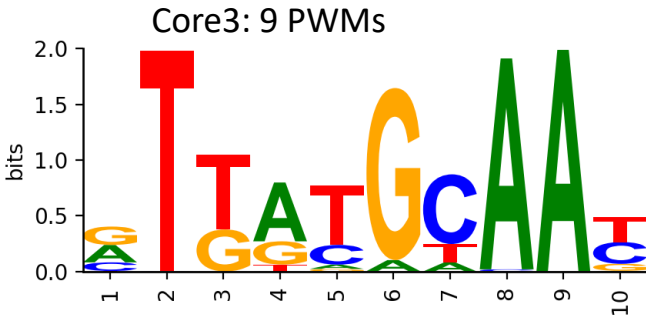
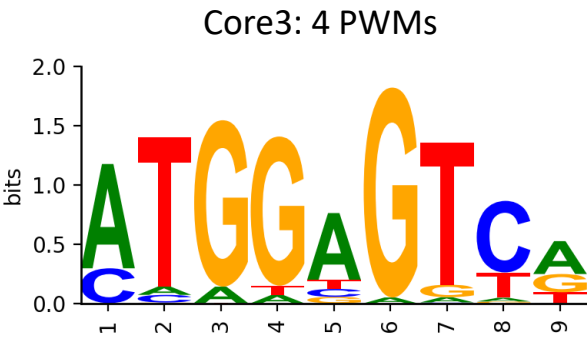
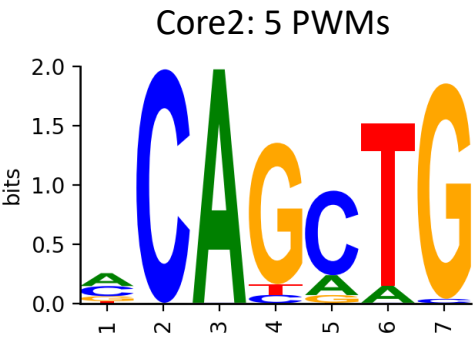
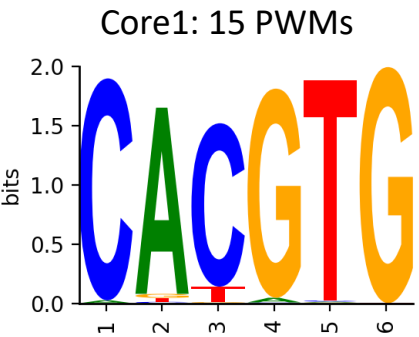


bZIP

- Total of canonical PWMs: 33



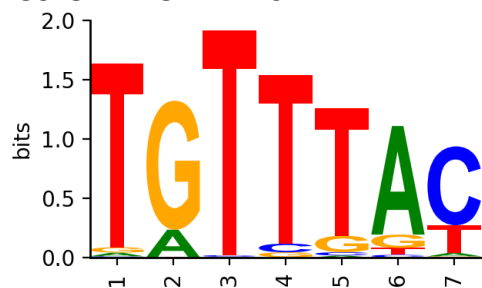
bHLH Total of canonical PWMs: 26



Forkhead

- Total of canonical PWMs: 14

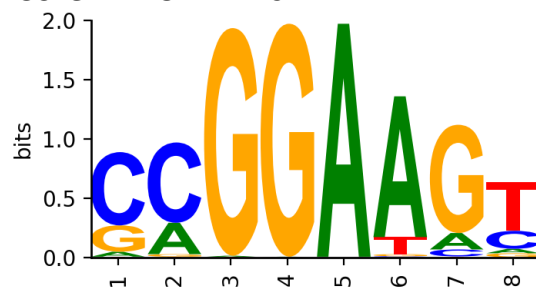
Core 1: 13 PWMs



Ets

- Total of canonical PWMs: 14

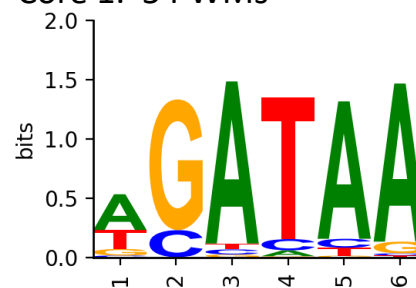
Core 1: 13 PWMs



GATA

- Total of canonical PWMs: 5

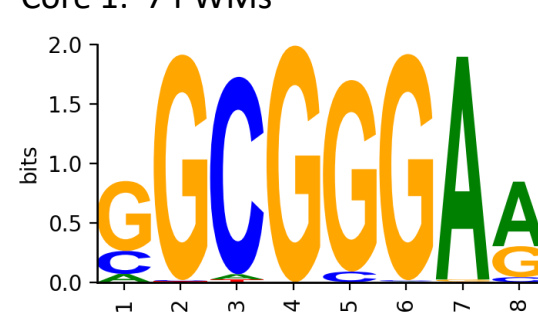
Core 1: 5 PWMs



E2F

- Total of canonical PWMs: 8

Core 1: 7 PWMs



CUT; Homeodomain

- Total of canonical PWMs: 3

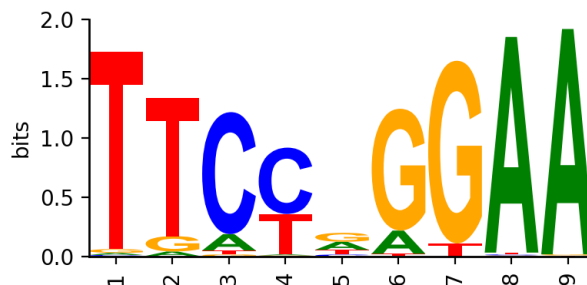
Core 1: 3 PWMs



STAT

- Total of canonical PWMs: 3

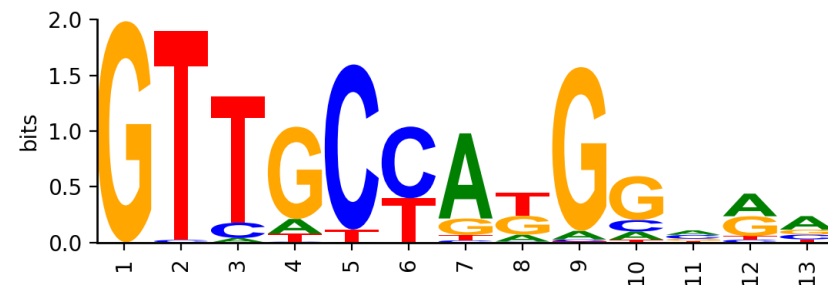
Core 1: 3 PWMs



RFX

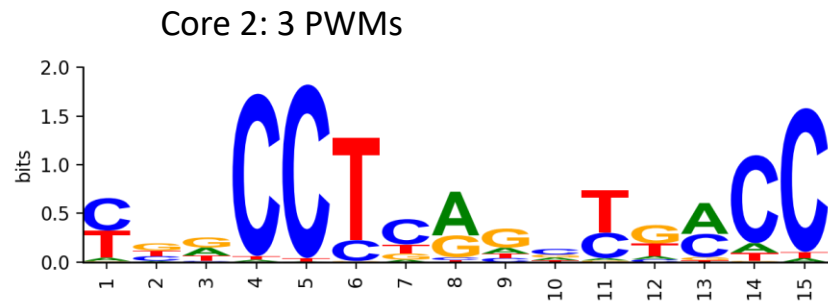
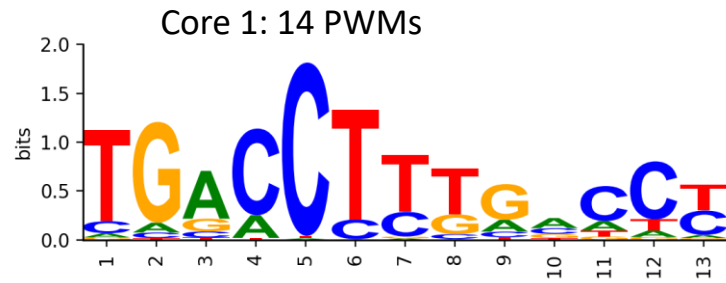
- Total of canonical PWMs: 3

Core 1: 3 PWMs



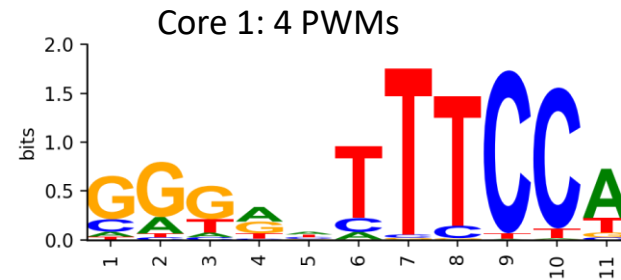
Nuclear receptor

- Total of canonical PWMs: 21



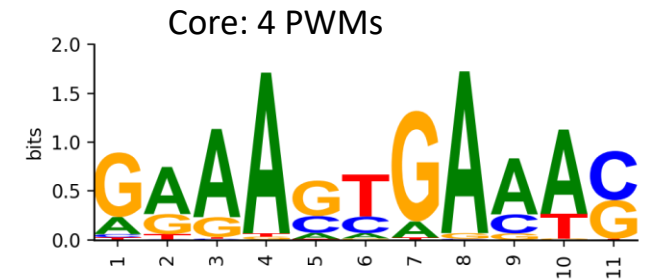
Rel

- Total of canonical PWMs: 4



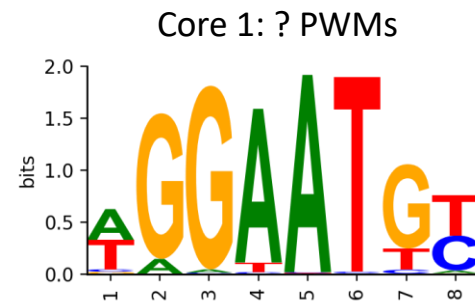
IRF

- Total of canonical PWMs: 4



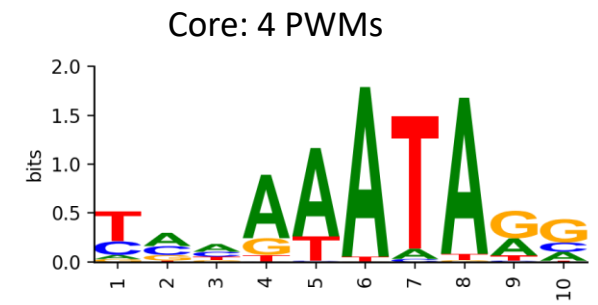
TEA

- Total of canonical PWMs: ?



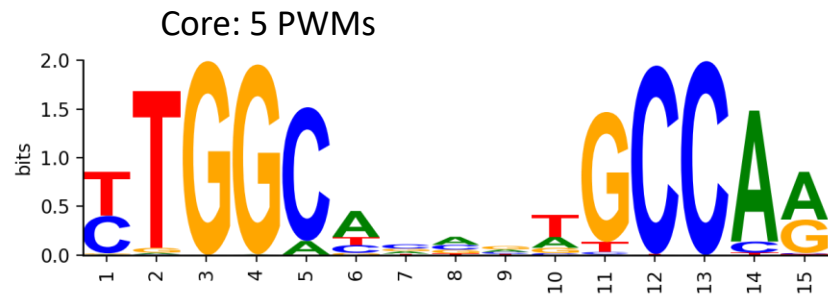
MADS box

- Total of canonical PWMs: 4



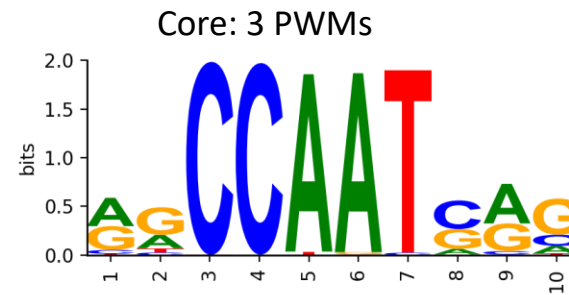
SMAD

- Total of canonical PWMs: 7



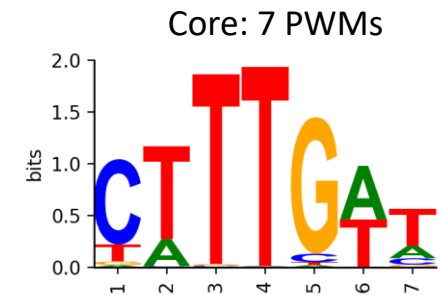
Unknown

- Total of canonical PWMs: 6



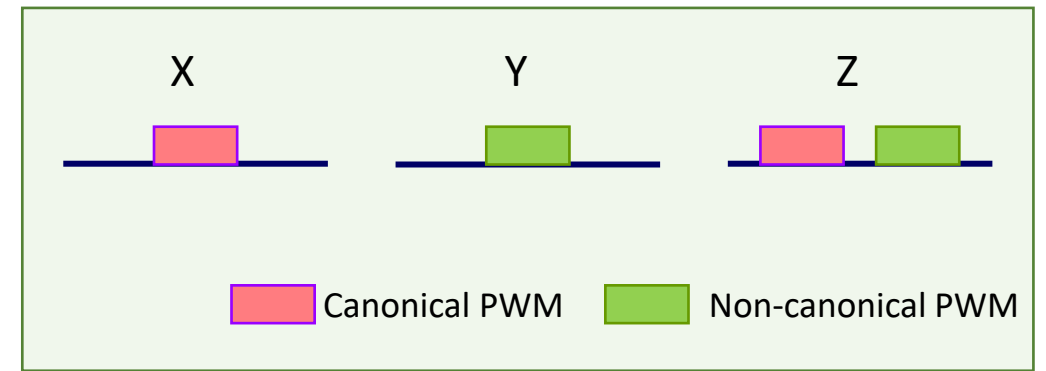
HMG/Sox

- Total of canonical PWMs: 7



S3 Fig. **Co-binding analysis.** Computing the counts of canonical binding, tethered binding and co-binding in a set of ChIP-seq data

Method

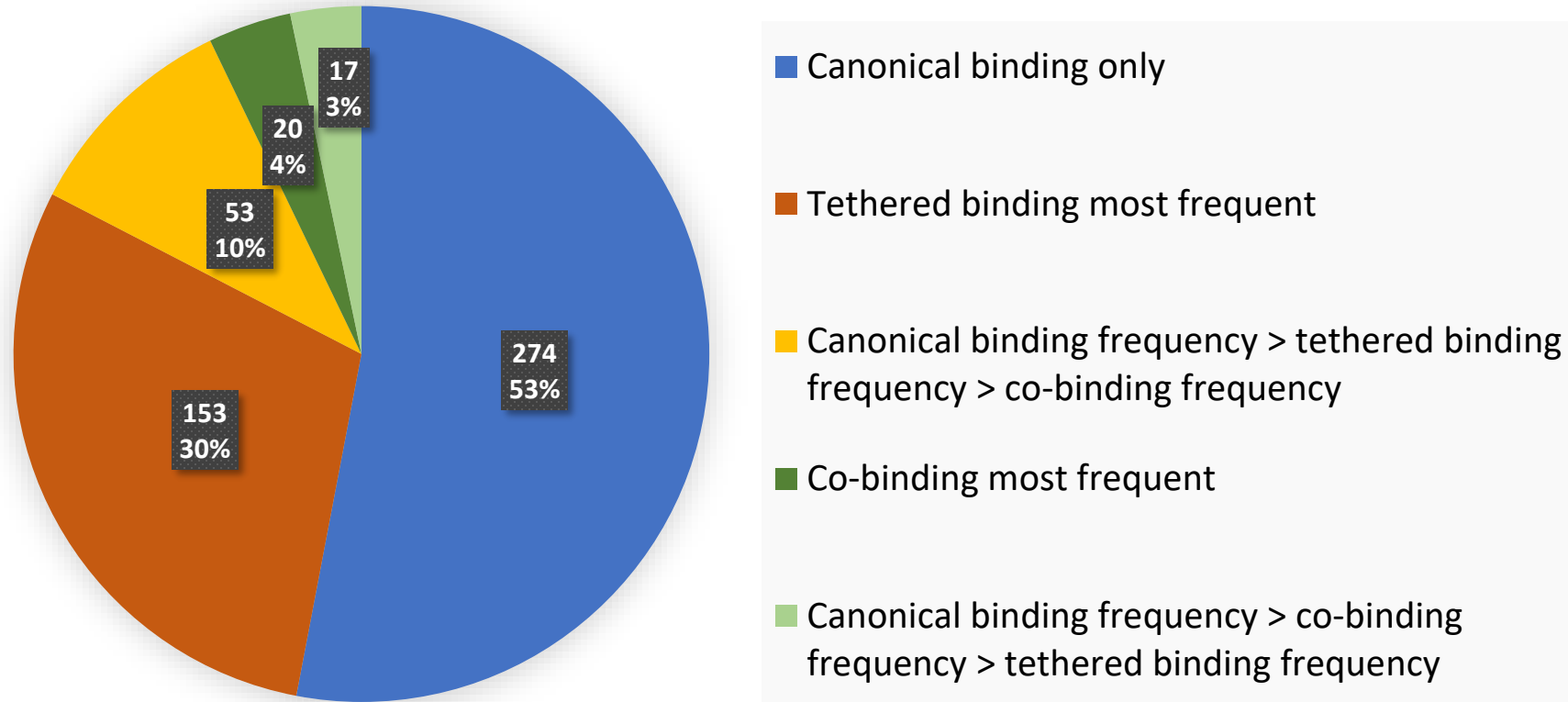


- In the above figure, there are three types of the ChIP-seq peaks: (1) The peak contains only a canonical motif (a red rectangle), (2) the peak contains only a non-canonical motif (a green rectangle), and (3) the peak contains a canonical motif and a non-canonical motif.
- From a set of ChIP-seq data for a TF, one can compute the following counts:
 - X: Number of peaks containing only the canonical motif.
 - Y: Number of peaks each containing at least one non-canonical (co-occurring) motif but no canonical motif.
 - Z: Number of peaks each containing both the canonical PWM and a non-canonical PWM.
- Note:
 - A non-canonical PWM is a PWM that is similar to the canonical PWM of a TF that belong to another TF family.
 - The presence of a motif is determined using the PWM and FIMO with a p-value < 0.0001 to screen the genome; a FIMO hit is retained only if it overlaps with a ChIP-seq peak.

Classification of TFs into five groups

- 1) $X > 0, Y = Z = 0$: canonical binding only
- 2) $X > Y > Z$: canonical binding frequency > tethered binding frequency > co-binding frequency
- 3) $X > Z > Y$: canonical binding frequency > co-binding frequency > tethered binding frequency
- 4) $Y > X$ and $Y > Z$: tethered binding most frequent
- 5) $Z > X$ and $Z > Y$: co-binding most frequent

Pie chart of the five groups of TFs



S4 Fig. Identification of canonical PWN for a TF by the ranking method

Inferred PWMs from ChIP-seq data for **GATA3**

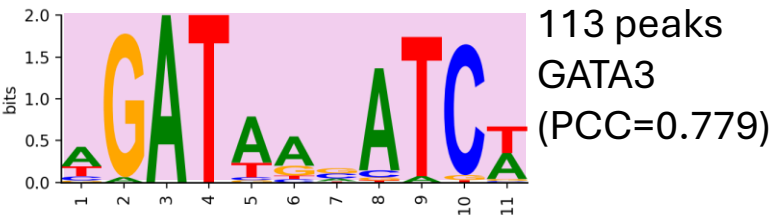
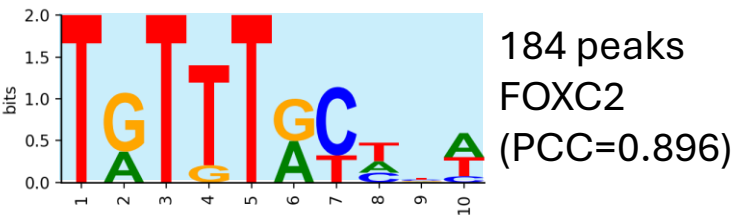
Experiment

Top

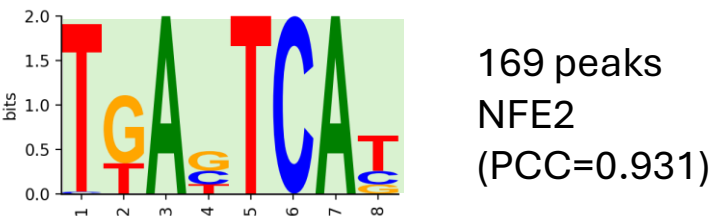
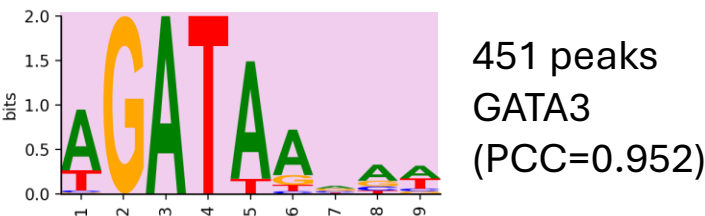
Second

Reference

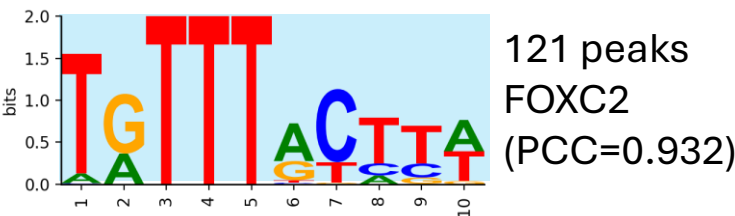
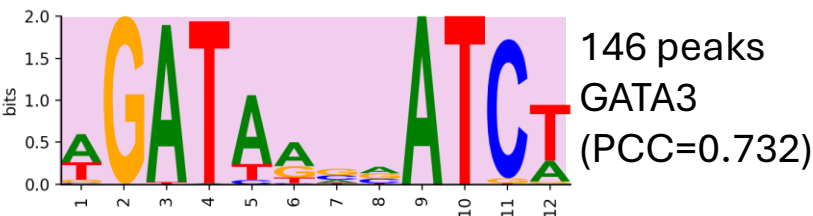
T47D
(ENCSR000BMX)



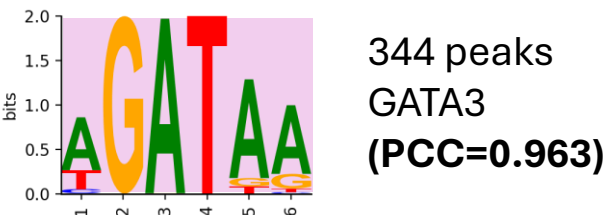
SK-N-SH
(ENCSR000BTH)



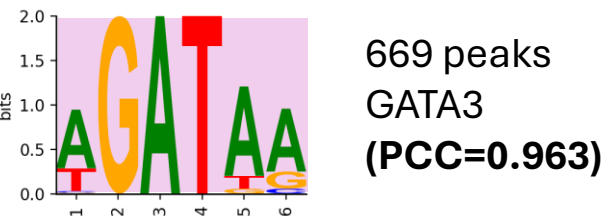
MCF-7
(ENCSR423RTK)



MCF-7
(ENCSR000EWS)



Multiple
experiments
(ranking)

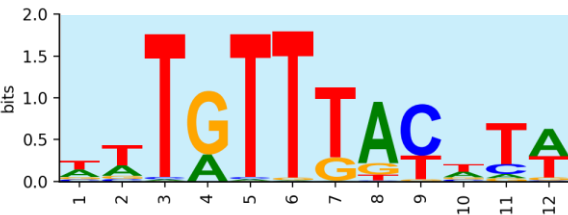


No secondary

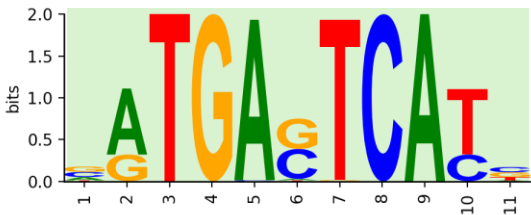
(A) GATA3 (SELEX: M03063_2.00)



(B) FOXC2 (SELEX: M03038_2.00)



(C) NFE2 (SELEX : M04266_2.00)



Inferred PWMs from ChIP-seq data for PBX3

Experiment

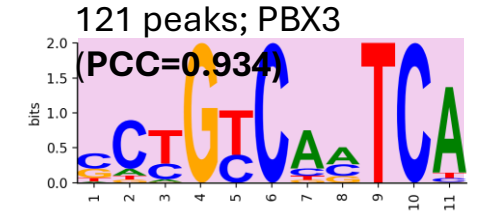
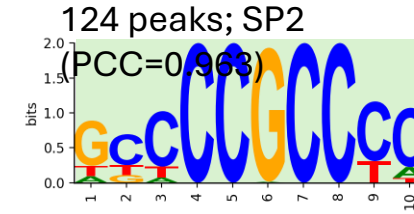
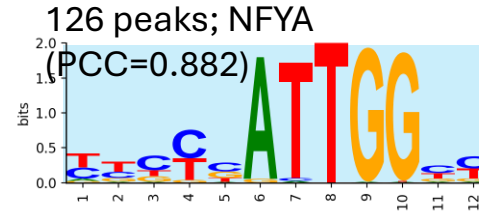
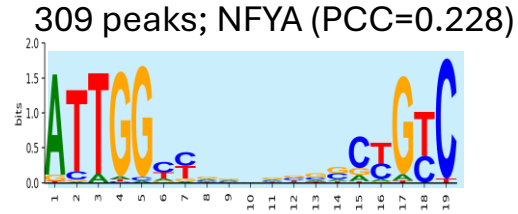
Top

Second

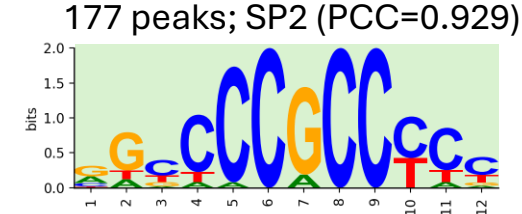
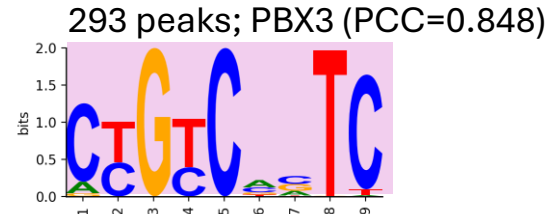
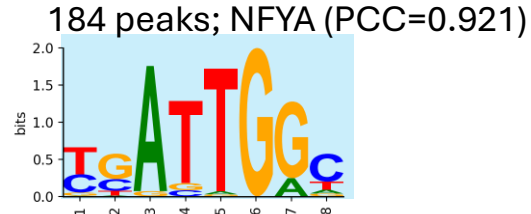
Third

Fourth

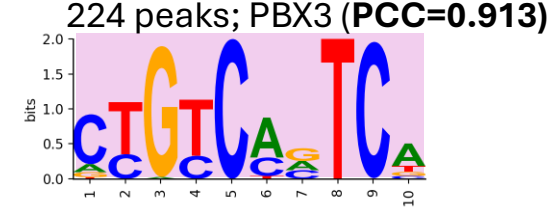
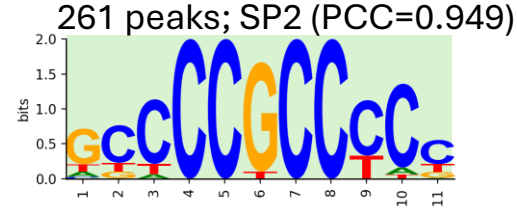
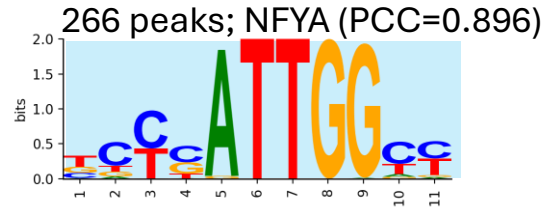
A549
(ENCSR000BTN)



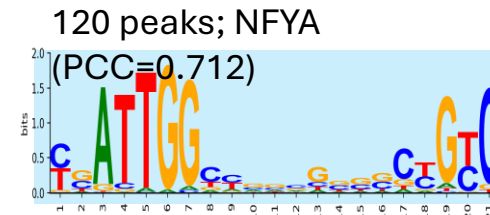
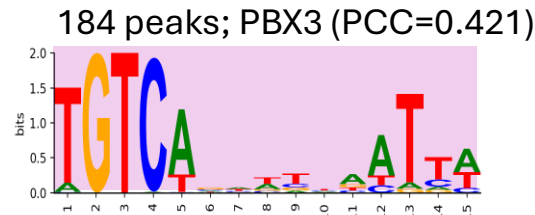
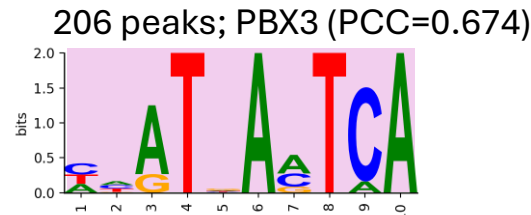
GM12878
(ENCSR000BGR)



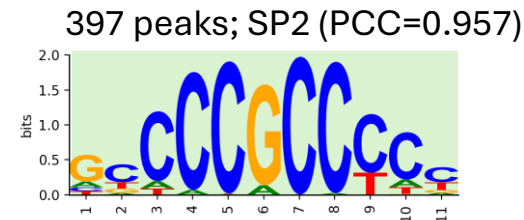
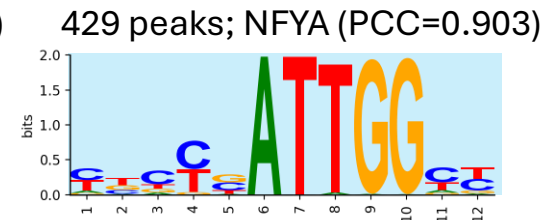
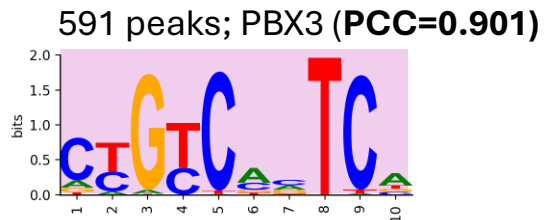
HEK293
(ENCSR865HXK)



SK-N-SH
(ENCSR000BVE)

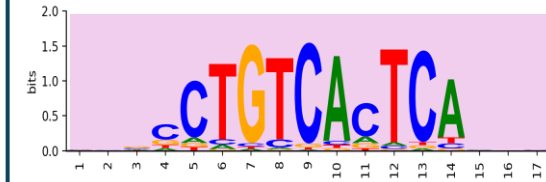


Multiple
experiments
(ranking)

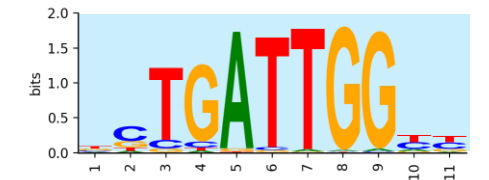


Reference

(A) PBX3 (JASPAR: MA1114.1)



(B) NFYA (JASPAR: MA0060.3)



(C) SP2 (JASPAR: MA0516.3)

