Lessons learned from transcription factor co-association analysis

The enhancer-promoter interaction model encoded by the transcription factors AP-1 and NF-Y

Workshop on Bioinformatics of Gene Regulation
Göttingen, March 7-9 2018

Institute of Bioinformatics
Director: Prof. Dr. Edgar Wingender
University Medical Center Göttingen
Georg-August-University

Martin Haubrock
mhaubro@uni-goettingen.de

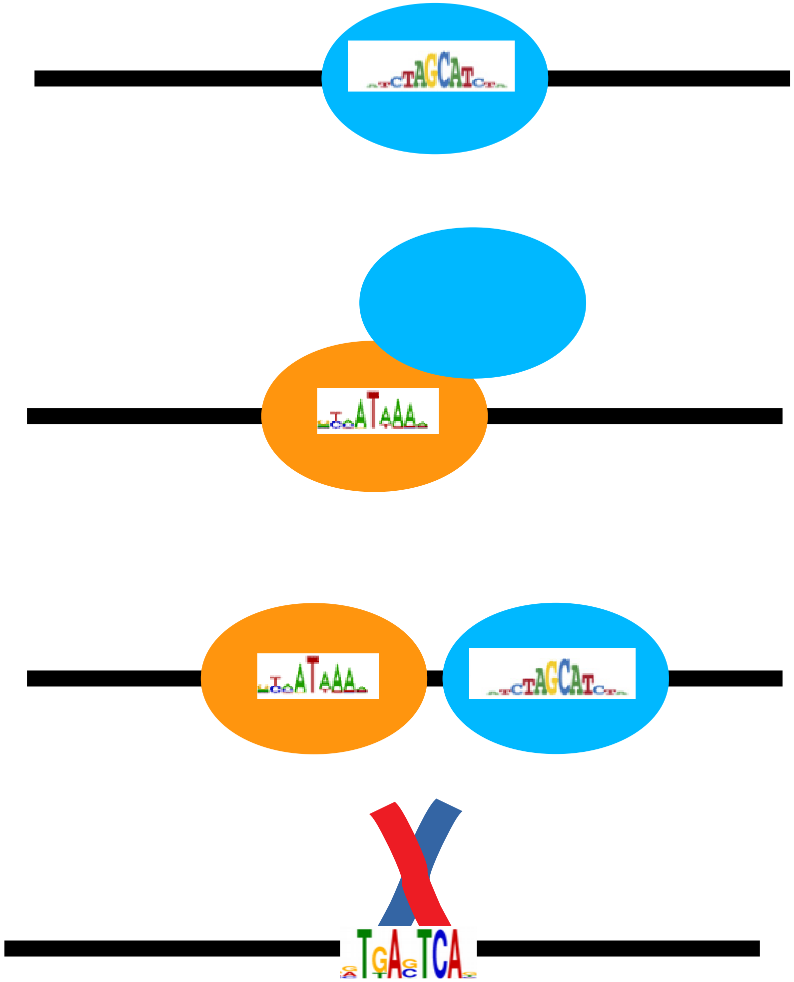
UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Binding mode: TF-DNA interactions

- **Direct and indirect factor binding**

Transcription factors (TF) are proteins that control the efficiency of gene expression

Most of them act by recognizing cis-regulatory elements (TFBS) in gene proximal (promoter) or distal (enhancer) regions



Direct

(direct *motif*)

Tethering

(*indirect motif*)

Co-binding

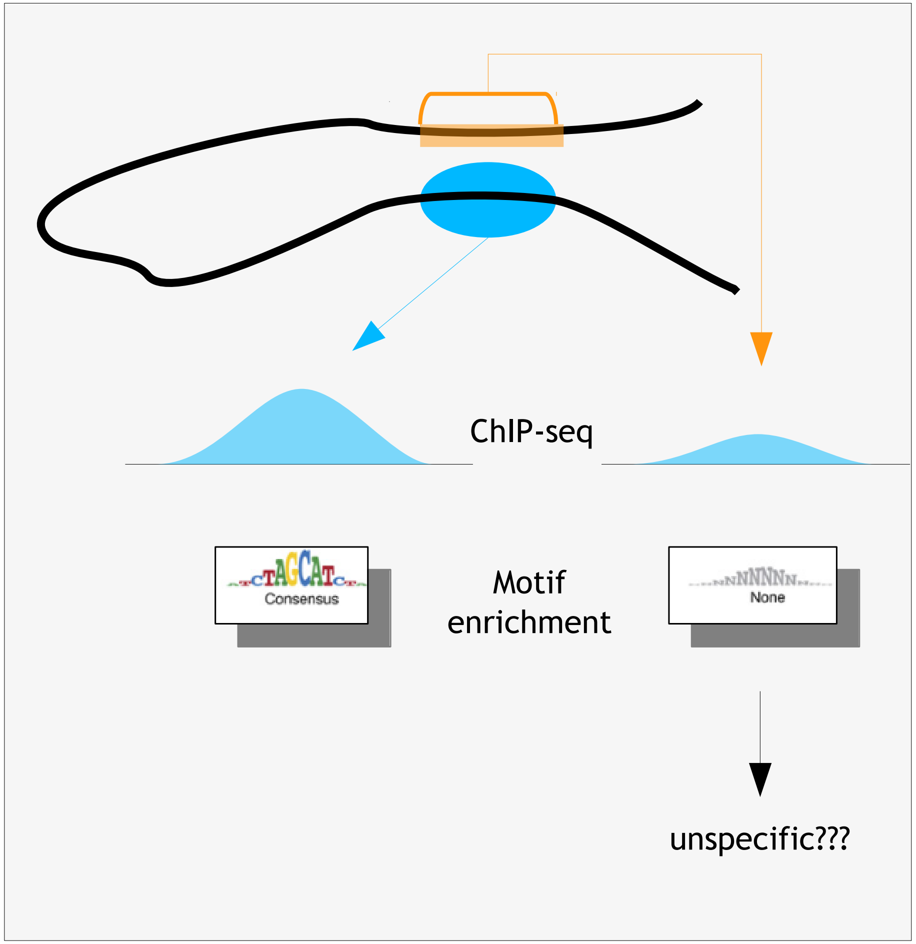
(*two motifs*)

Dimerization

(*one motif*)

Interpreting ChIP-seq: Motif finding

- **Regulatory and transcriptional complexity resulting from ChIP-seq experiments:**
  - ChIP-seq of a transcription factor results in immunoprecipitation of bound DNA sequences and DNA sequences in close proximity (Mecer and Mattick 2013)

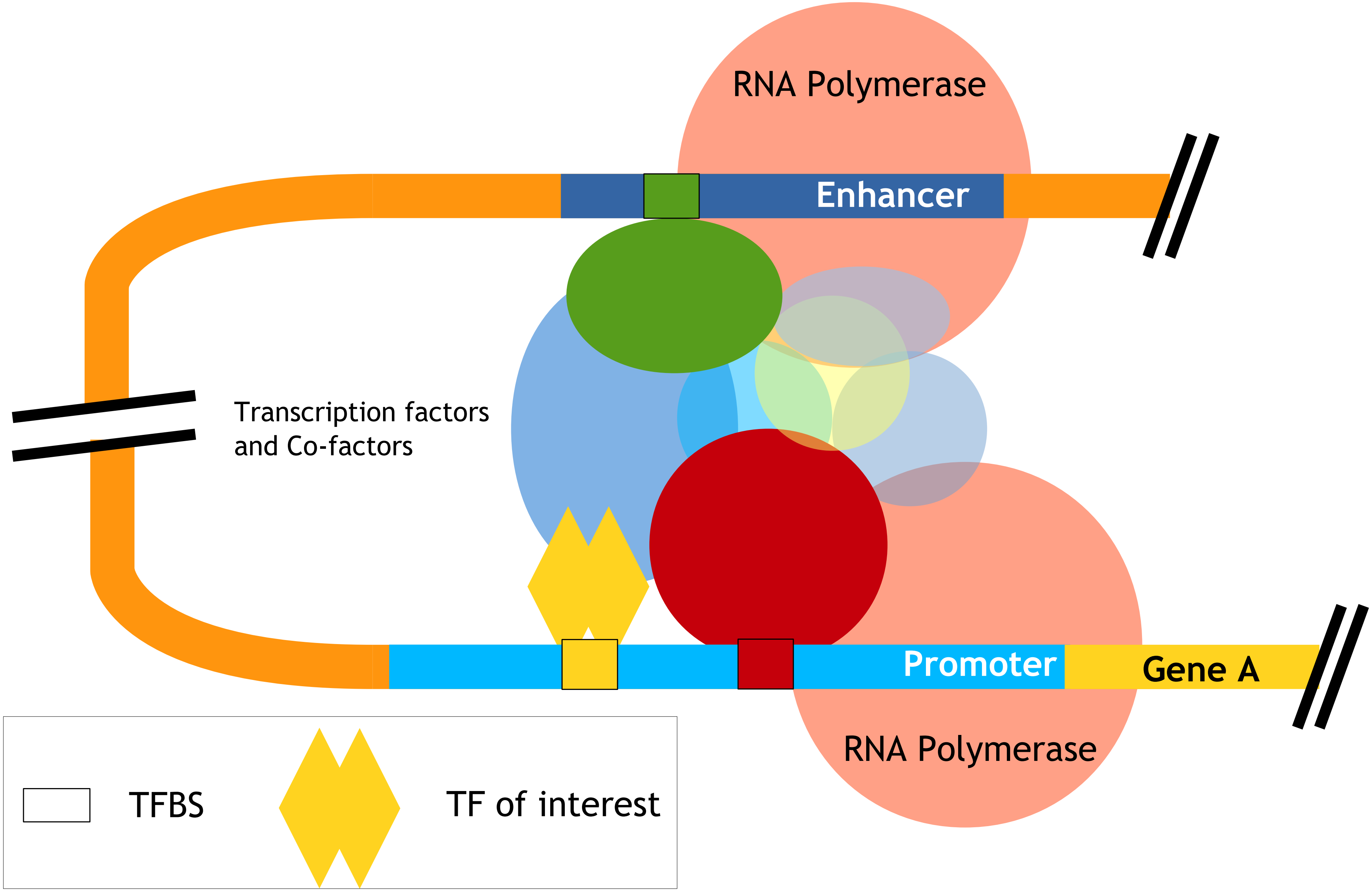


Hypothesis:

Close spatial proximity of distal and proximal ChIP-seq regions exhibit functional enhancer/promoter relations.

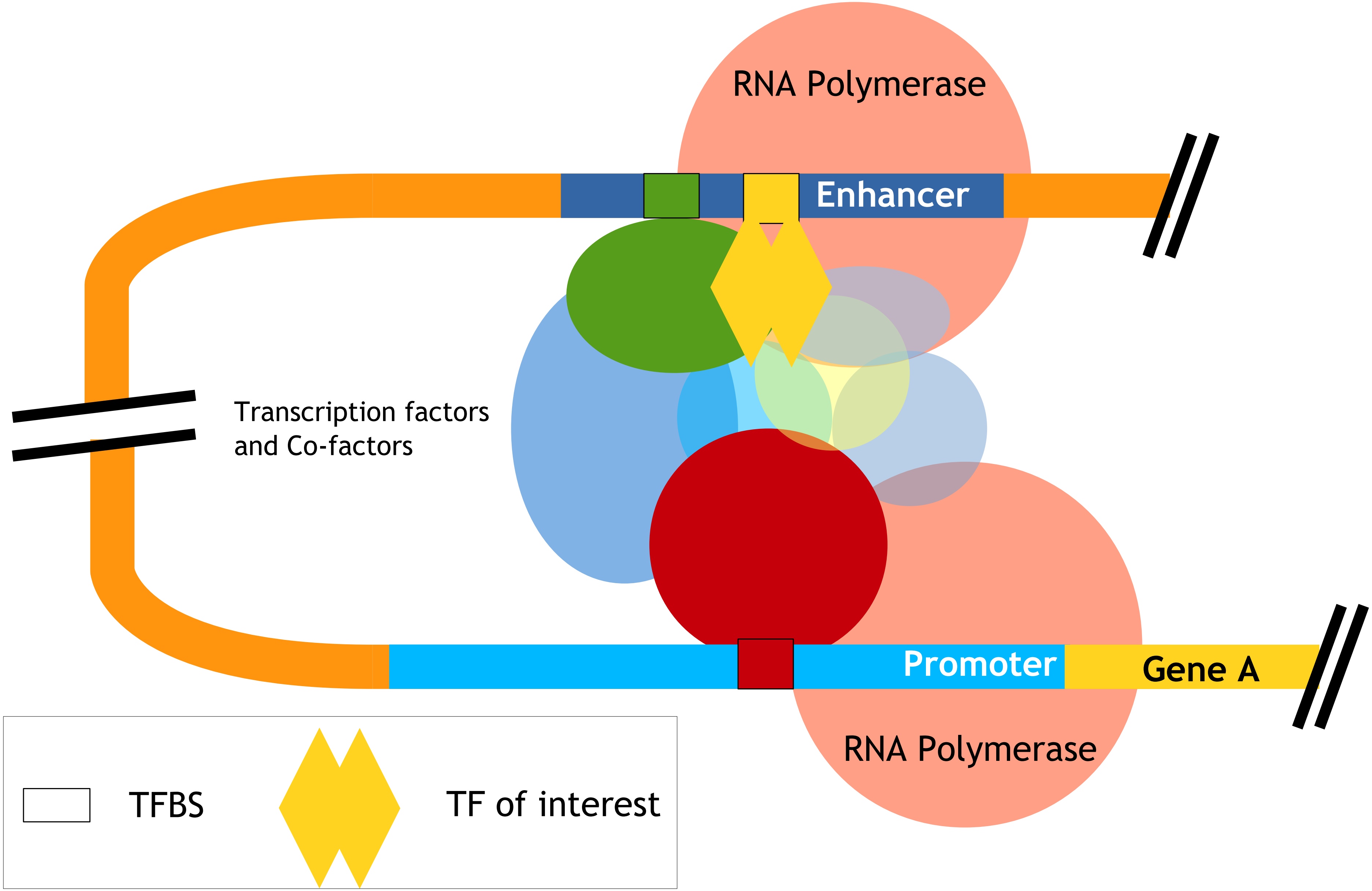
Enhanceosome

- Direct TF-DNA interaction



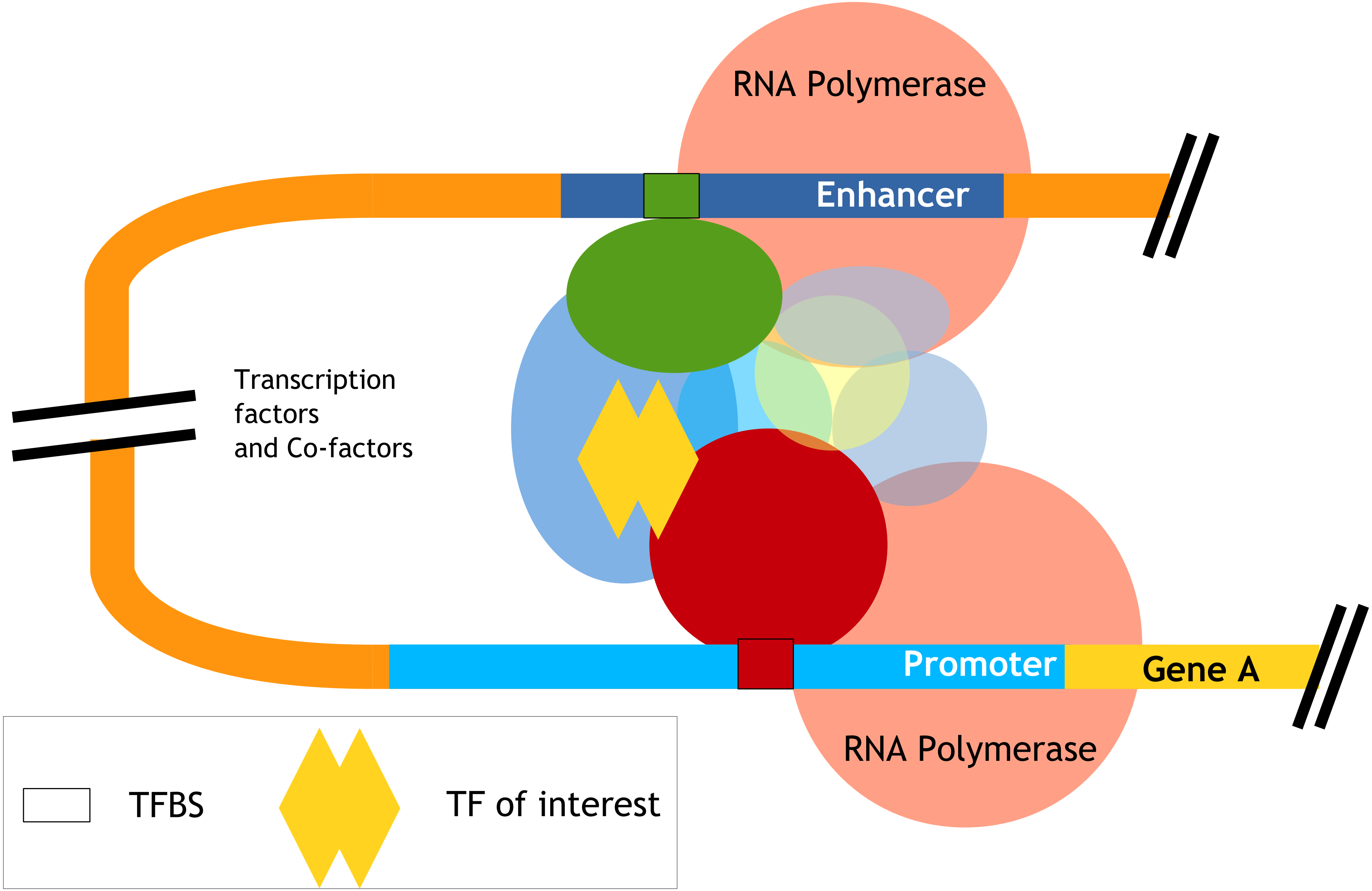
Enhanceosome

- **Direct TF-DNA interaction**



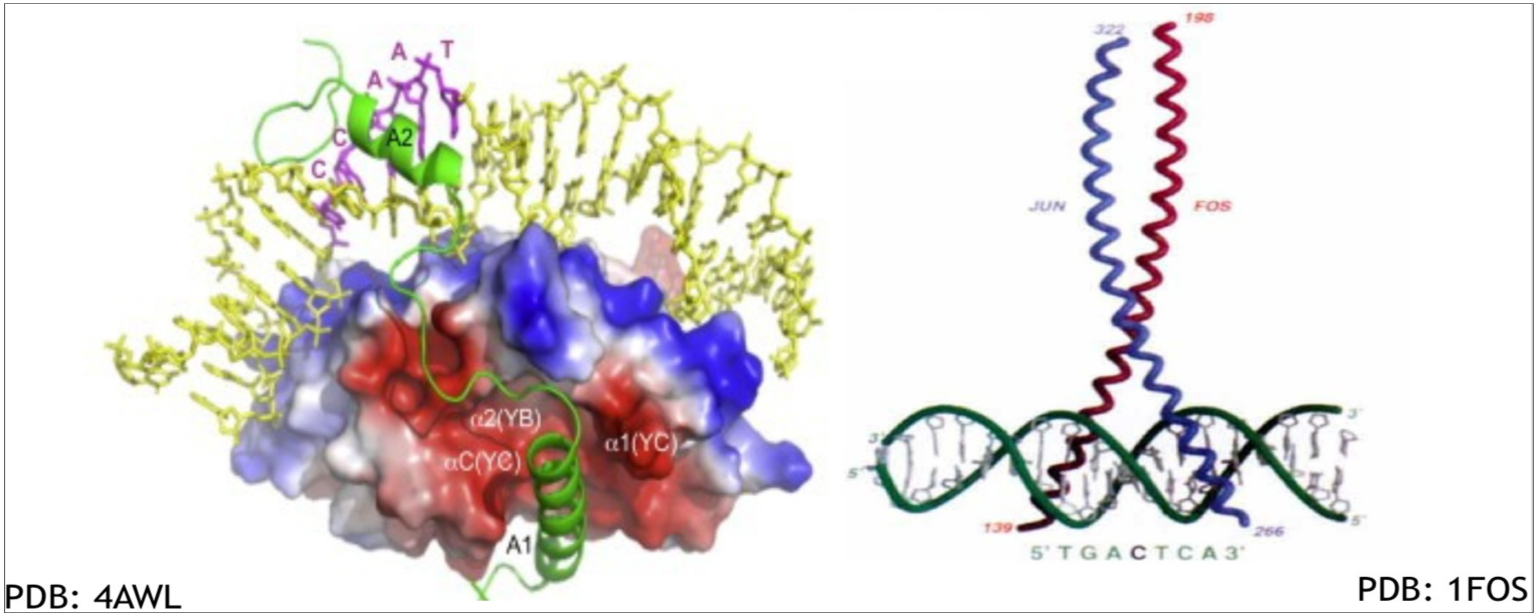
Enhanceosome

- **Indirect TF-DNA interaction**



Study

- **Enhancer Promoter Analysis**



PDB: 4AWL

PDB: 1FOS

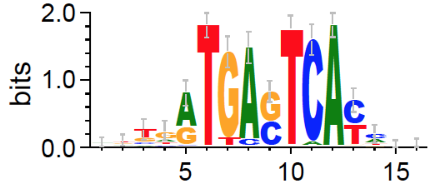
- **Motif enrichment in ChIP-seq data interacting regions using Receiver Operating Characterisitic (AUROC) using PWM library from TRANSFAC**
- **ChIP-seq data from ENCODE**
  - HUVEC, K562, HeLa S3, and GM12878

Binding site composition for c-Fos ChIP-seq

- HeLa, K562, and GM12878 promoter exhibit NF-Y binding motifs

a

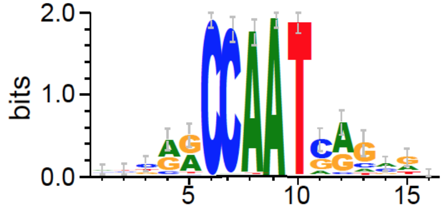
AP-1



TRANSFAC: V$JUNDM2_04

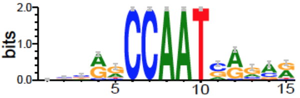
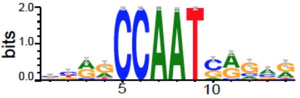
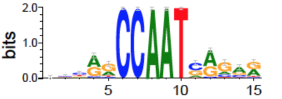
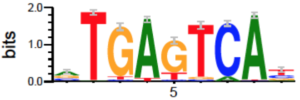
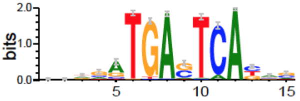
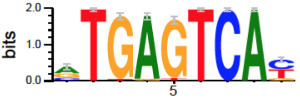
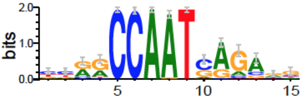
b

NF-Y



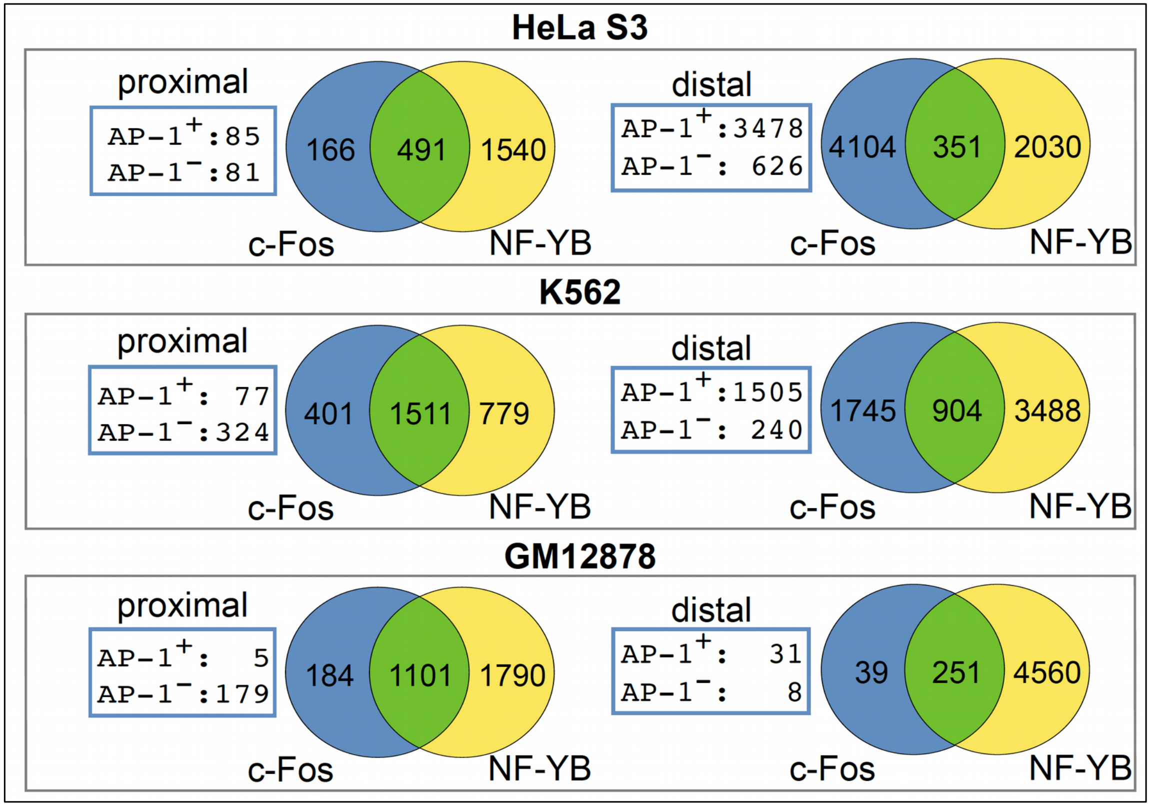
TRANSFAC: V$NFY_01

c

class	cell-line	MEME motif	AP-1$_{AUC}$	NF-Y$_{AUC}$
proximal	HUVEC	 3.9e-192 411/657	0.884^{1}	0.639
	HeLa S3	 1.3e-654 531/657	0.685	0.878^{2}
	K562	 7.8e-741 578/657	0.623	0.915^{1}
	GM12878	 7.1e-786 595/657	0.634	0.916^{1}
distal	HUVEC	 4.4e-366 657/657	0.942^{1}	0.646
	HeLa S3	 1.8e-683 657/657	0.891^{1}	0.614
	K562	 4.0e-431 410/657	0.810^{1}	0.731
	GM12878	 1.1e-319 247/290	0.575	0.904^{1}

Overlap of c-Fos and NF-YB ChIP-seq intervals

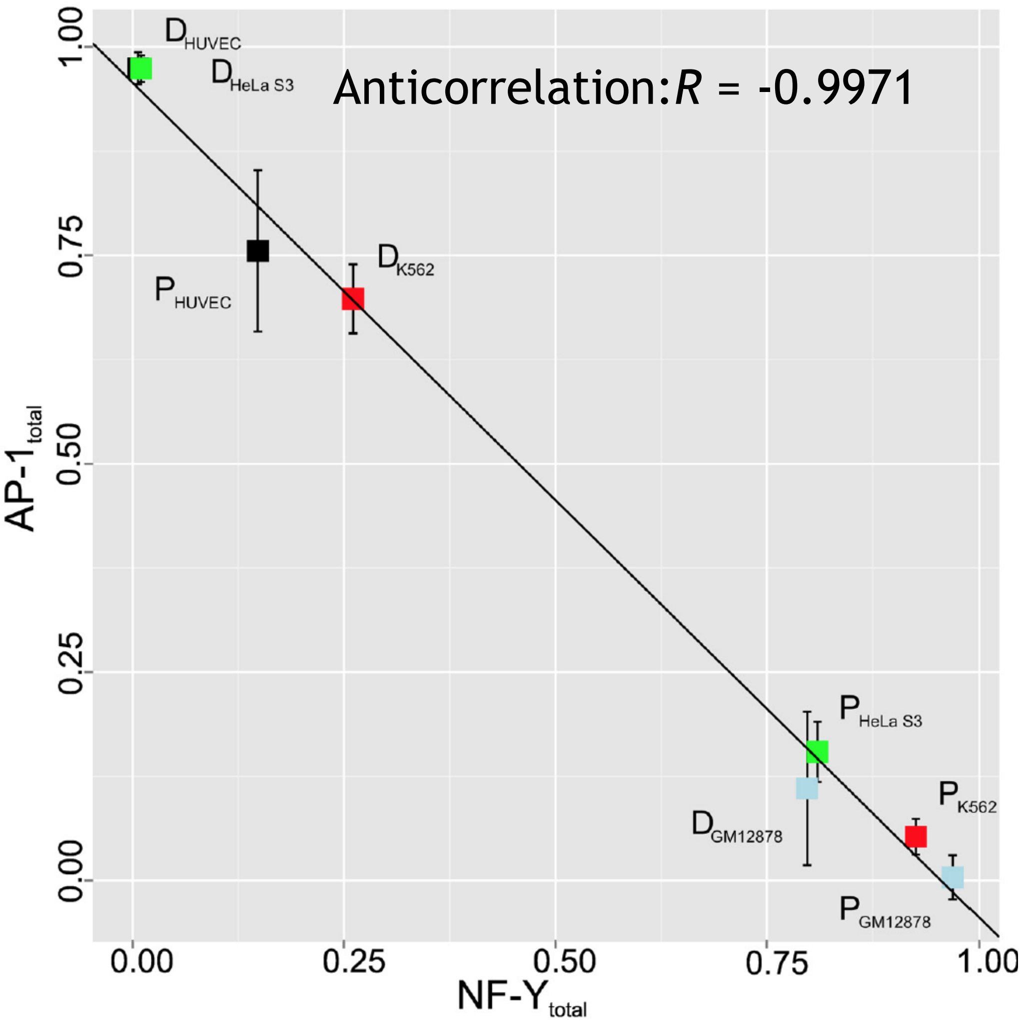
- c-Fos precipitated regions largely co-localize with NF-YB genomic intervals

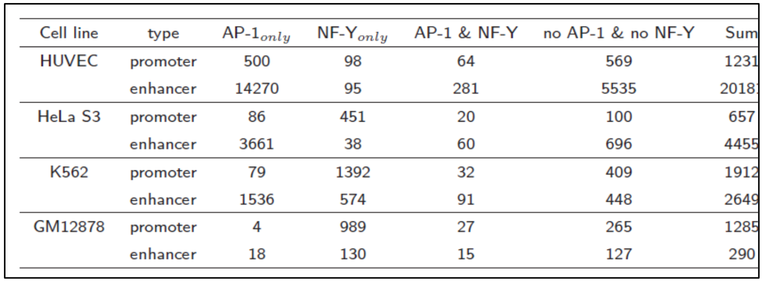


Cell line	type	c-Fos	c-Fos+NF-YB	NF-YB
HeLa S3	p	166 (7.6%)	491 (22.3%)	1540 (70.1%)
	d	4104 (63,3%)	351 (5.4%)	2030 (31.3%)
K562	p	401 (14.9%)	1511 (56.2%)	779 (28.9%)
	d	1745 (28.4%)	904 (14.7%)	3488 (56.8%)
GM12878	p	184 (6.0%))	1101 (35.8%)	1790 (58.2%)
	d	39 (1.0%)	251 (5.2%)	4560 (94.0%)

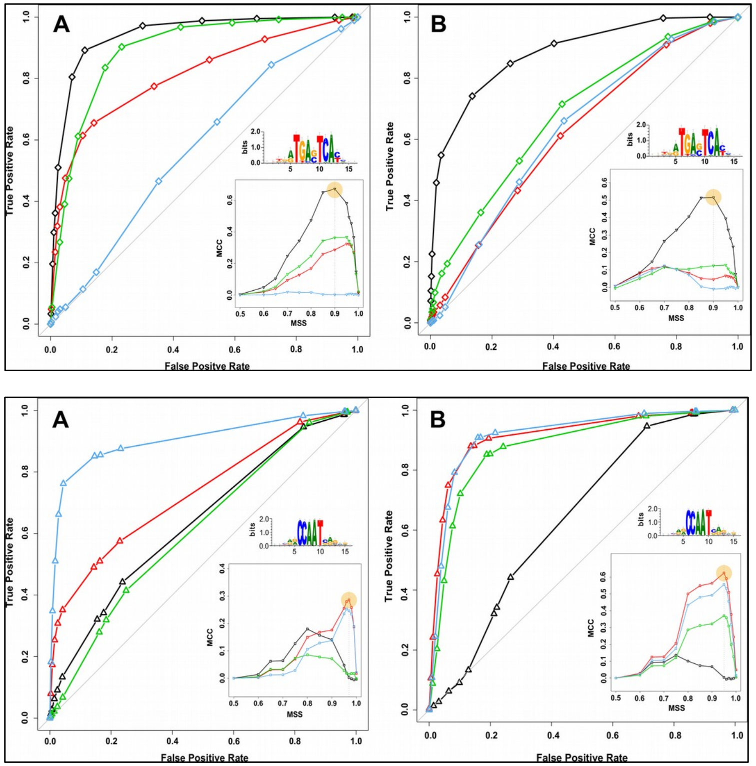
Mutual exclusivity of AP-1 and NF-Y motifs

- NF-Y and AP-1 motifs are mutually exclusive in c-Fos precipitated regions (both in distal and in proximal intervals)





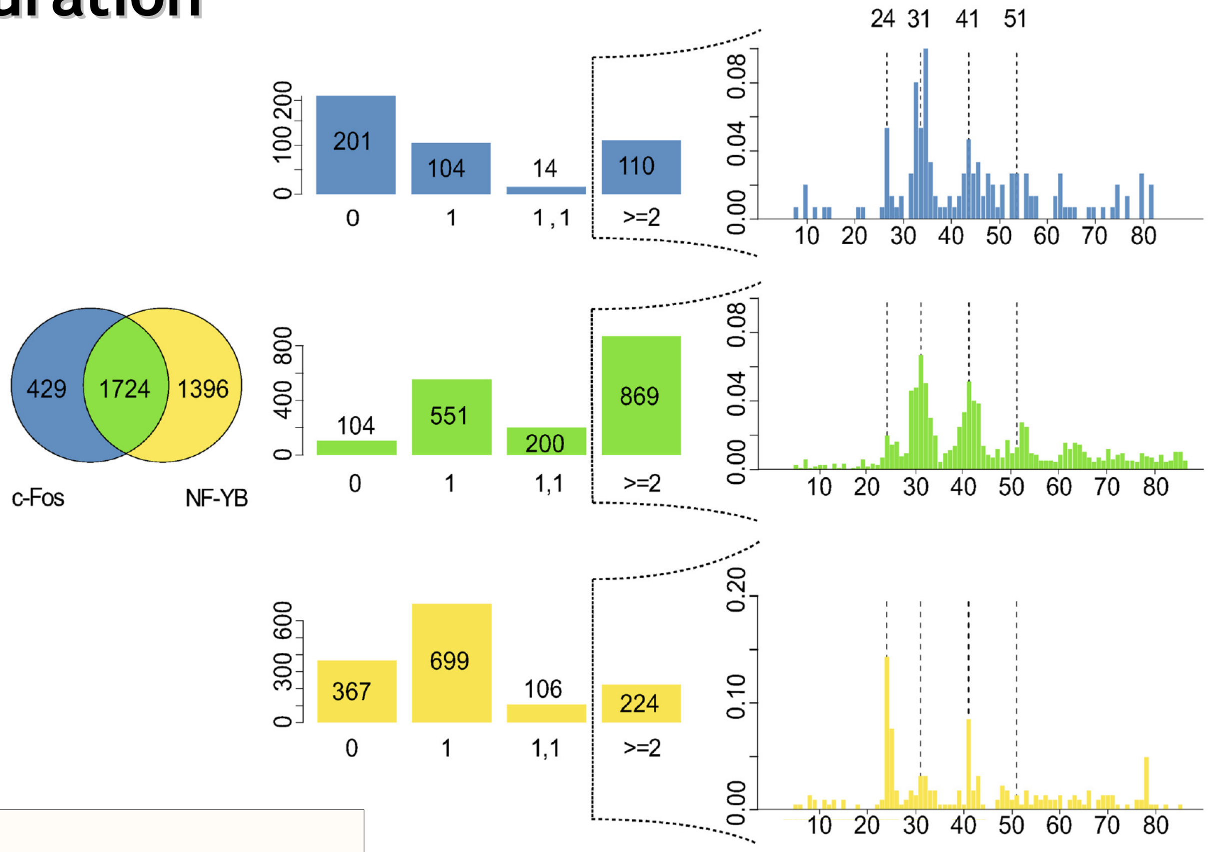
Cell line	type	AP-1$_{only}$	NF-Y$_{only}$	AP-1 & NF-Y	no AP-1 & no NF-Y	Sum
HUVEC	promoter	500	98	64	569	1231
	enhancer	14270	95	281	5535	2018
HeLa S3	promoter	86	451	20	100	657
	enhancer	3661	38	60	696	4455
K562	promoter	79	1392	32	409	1912
	enhancer	1536	574	91	448	2649
GM12878	promoter	4	989	27	265	1285
	enhancer	18	130	15	127	290

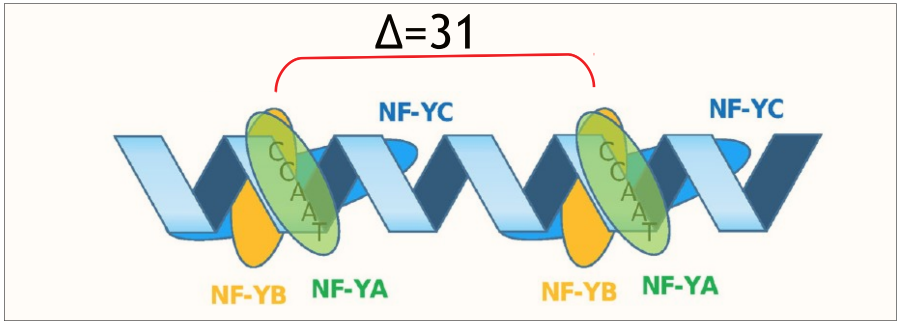


(A) distal, (B) proximal

CCAAT motif dimers in proximal regions

- **Many c-Fos-bound genomic intervals exhibit CCAAT dimers in a specific configuration**



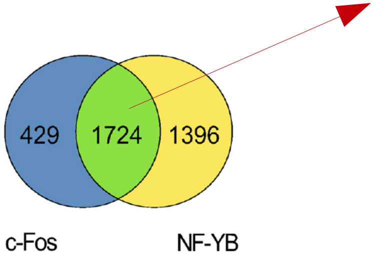


Enhancer/Promoter distance distribution

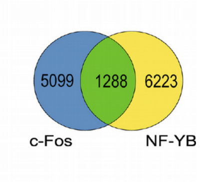
- **AP-1 positive enhancers tend to be close to c-Fos targeted promoters**

Promoter set

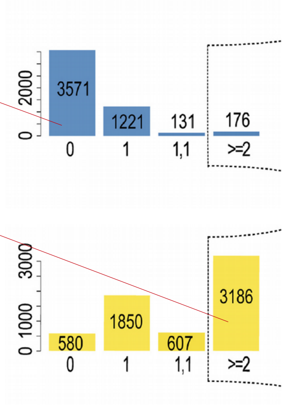
- P(c-Fos+NF-YB):1724

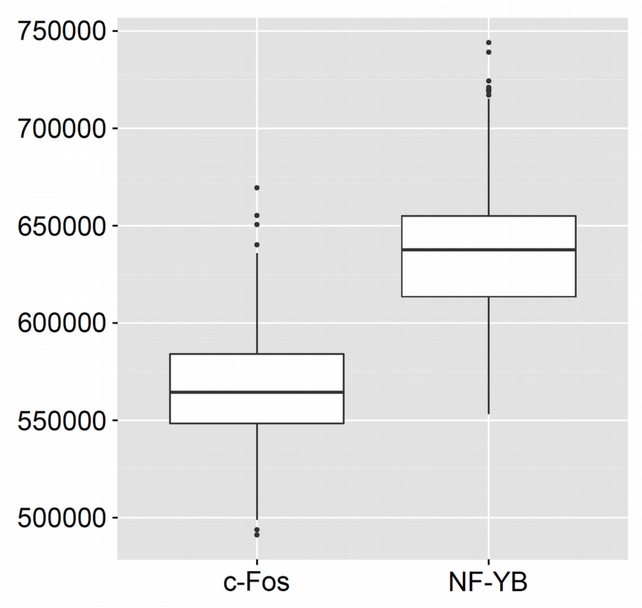


Enhancer set



- c-Fos: 3571
- NF-YB: 3186



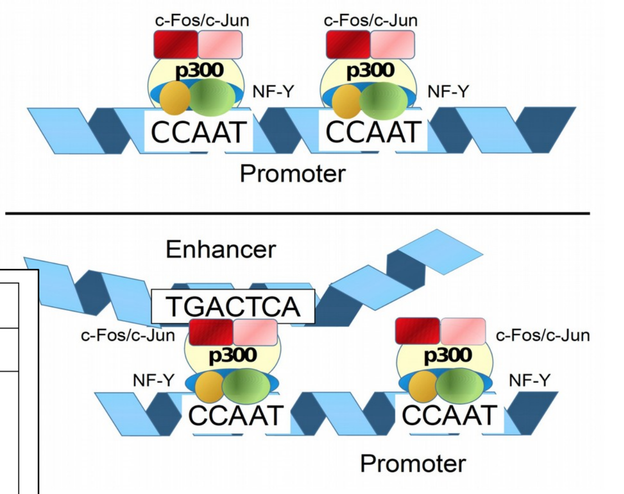


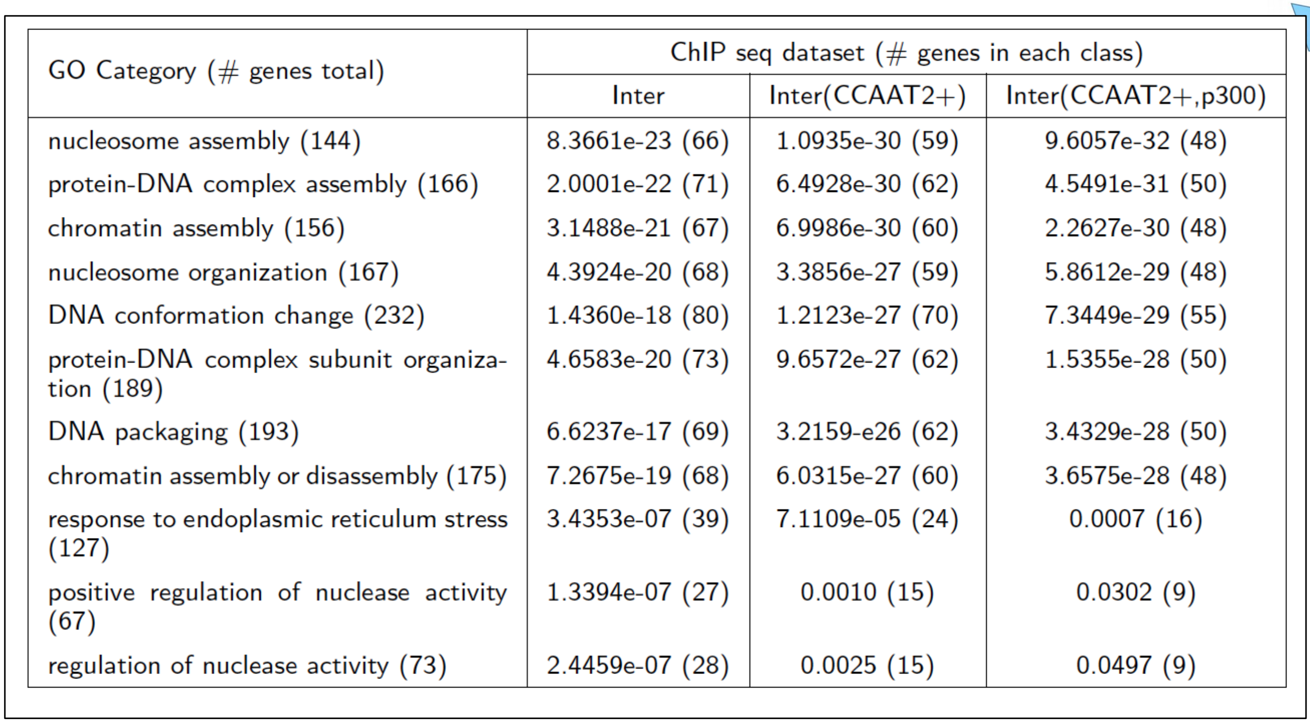
c-Fos-bound and AP-1 motif shaped enhancers show a clear tendency to be closer to c-Fos-bound and CCAAT box characterized promoters than the control (Mann-Whitney Test: P = 3.31703e − 281)

CCAAT direct repeats regulate defined gene set

- **NF-Y binding site architecture defines a c-Fos targeted promoter class**

Model: c-Fos/AP-1 and NF-YB/NF-Y interaction





GO Category (# genes total)	ChIP seq dataset (# genes in each class)		
	Inter	Inter(CCAAT2+)	Inter(CCAAT2+,p300)
nucleosome assembly (144)	8.3661e-23 (66)	1.0935e-30 (59)	9.6057e-32 (48)
protein-DNA complex assembly (166)	2.0001e-22 (71)	6.4928e-30 (62)	4.5491e-31 (50)
chromatin assembly (156)	3.1488e-21 (67)	6.9986e-30 (60)	2.2627e-30 (48)
nucleosome organization (167)	4.3924e-20 (68)	3.3856e-27 (59)	5.8612e-29 (48)
DNA conformation change (232)	1.4360e-18 (80)	1.2123e-27 (70)	7.3449e-29 (55)
protein-DNA complex subunit organization (189)	4.6583e-20 (73)	9.6572e-27 (62)	1.5355e-28 (50)
DNA packaging (193)	6.6237e-17 (69)	3.2159-e26 (62)	3.4329e-28 (50)
chromatin assembly or disassembly (175)	7.2675e-19 (68)	6.0315e-27 (60)	3.6575e-28 (48)
response to endoplasmic reticulum stress (127)	3.4353e-07 (39)	7.1109e-05 (24)	0.0007 (16)
positive regulation of nuclease activity (67)	1.3394e-07 (27)	0.0010 (15)	0.0302 (9)
regulation of nuclease activity (73)	2.4459e-07 (28)	0.0025 (15)	0.0497 (9)

For more details:

RESEARCH ARTICLE

NF-Y Binding Site Architecture Defines a C-Fos Targeted Promoter Class

Martin Haubrock, Fabian Hartmann, Edgar Wingender

Published: August 12, 2016 • https://doi.org/10.1371/journal.pone.0160803