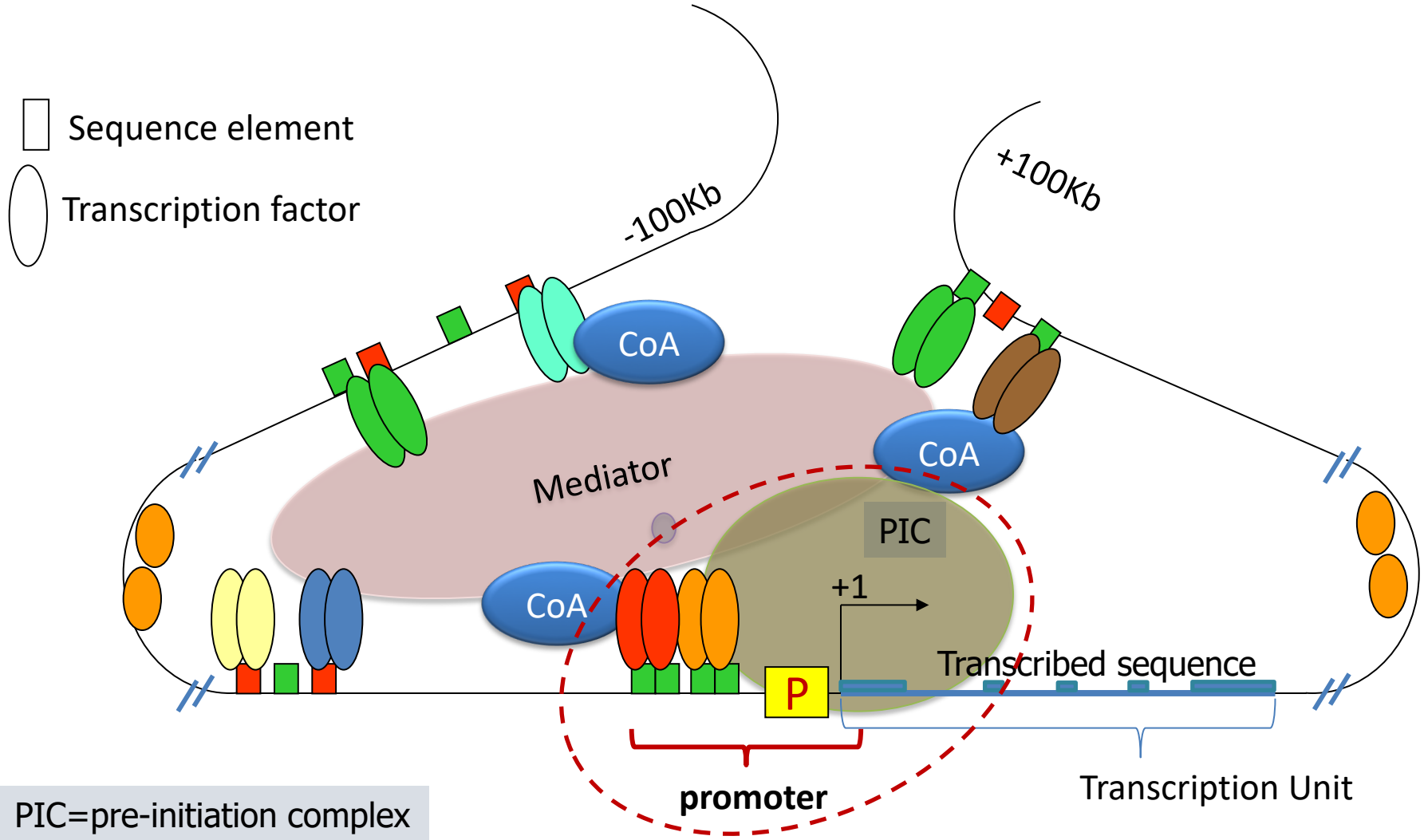
L4.2 – Transcriptional regulation (Promoters and RNAP)

Schematics of eukaryotic gene regulatory sequences and proteins



PIC=pre-initiation complex

AGENDA

1. How to define and identify promoters
2. Different types of gene promoters
3. Transcriptional initiation: PIC and RNAPolII

Textbook

Seminars in Cell & Developmental Biology 57 (2016) 11–23

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Promoter architectures and developmental gene regulation

Vanja Haberle [a,b,1], Boris Lenhard [a,*]

[a] Institute of Clinical Sciences and MRC Clinical Sciences Center, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK
[b] Department of Biology, University of Bergen, Thormøhlensgate 53A, N-5008 Bergen, Norway

ARTICLE INFO

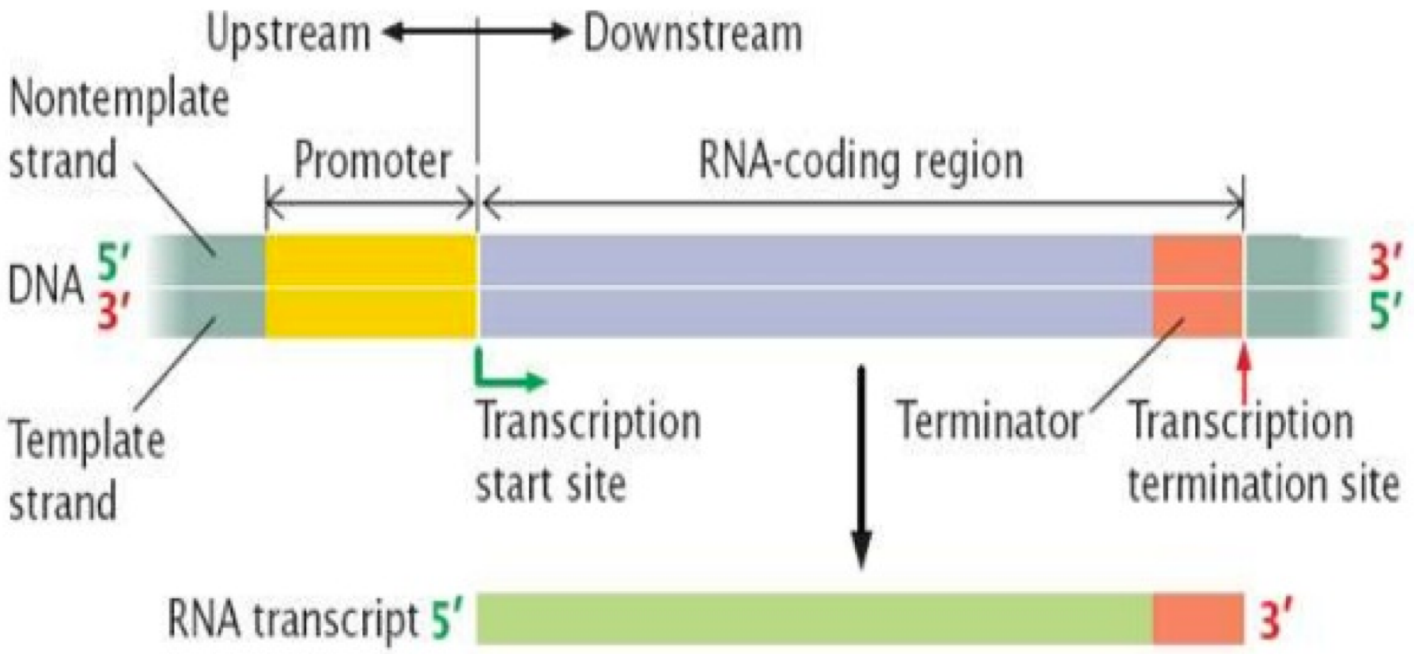
Article history:
Received 5 November 2015
Received in revised form 8 January 2016
Accepted 11 January 2016
Available online 16 January 2016

Keywords:
Transcriptional regulation
Core promoter
Transcription start sites
CAGE
Promoter types
Overlapping codes

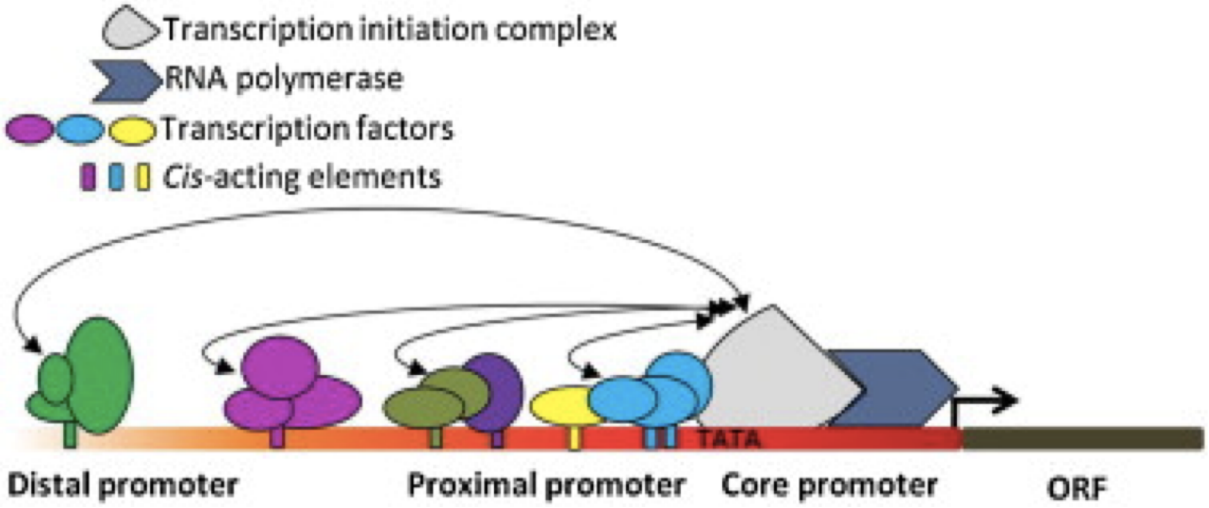
ABSTRACT

Core promoters are minimal regions sufficient to direct accurate initiation of transcription and are crucial for regulation of gene expression. They are highly diverse in terms of associated core promoter motifs, underlying sequence composition and patterns of transcription initiation. Distinctive features of promoters are also seen at the chromatin level, including nucleosome positioning patterns and presence of specific histone modifications. Recent advances in identifying and characterizing promoters using next-generation sequencing-based technologies have provided the basis for their classification into functional groups and have shed light on their modes of regulation, with important implications for transcriptional regulation in development. This review discusses the methodology and the results of genome-wide studies that provided insight into the diversity of RNA polymerase II promoter architectures in vertebrates and other Metazoa, and the association of these architectures with distinct modes of regulation in embryonic development and differentiation.

Transcription unit



Promoter: Where the RNA polymerase and associated factors bind to initiate transcription (important to distinguish between core promoter vs proximal promoter vs bioinformatics definition)



Promoter = the minimal sequence sustaining transcription and correct initiation, usually 50-150 bp 5'-upstream TSS

Upstream regulatory sequence: sequences 5'-adjacent to promoter that regulate promoter utilization (500-2,000 bp, usually a downstream part to +100 is also included). Also sometimes indicated as «UAS», «proximal regulatory element» or «proximal enhancer».

Enhancers: regulatory sequences or «modules» laying virtually at any distance and position from the regulated («cognate») TSS or promoter. Note: even though «enhancer» means «something that increases», enhancers may display repressing activity.

PRMOTERS IDENTIFICATION and DEFINITION

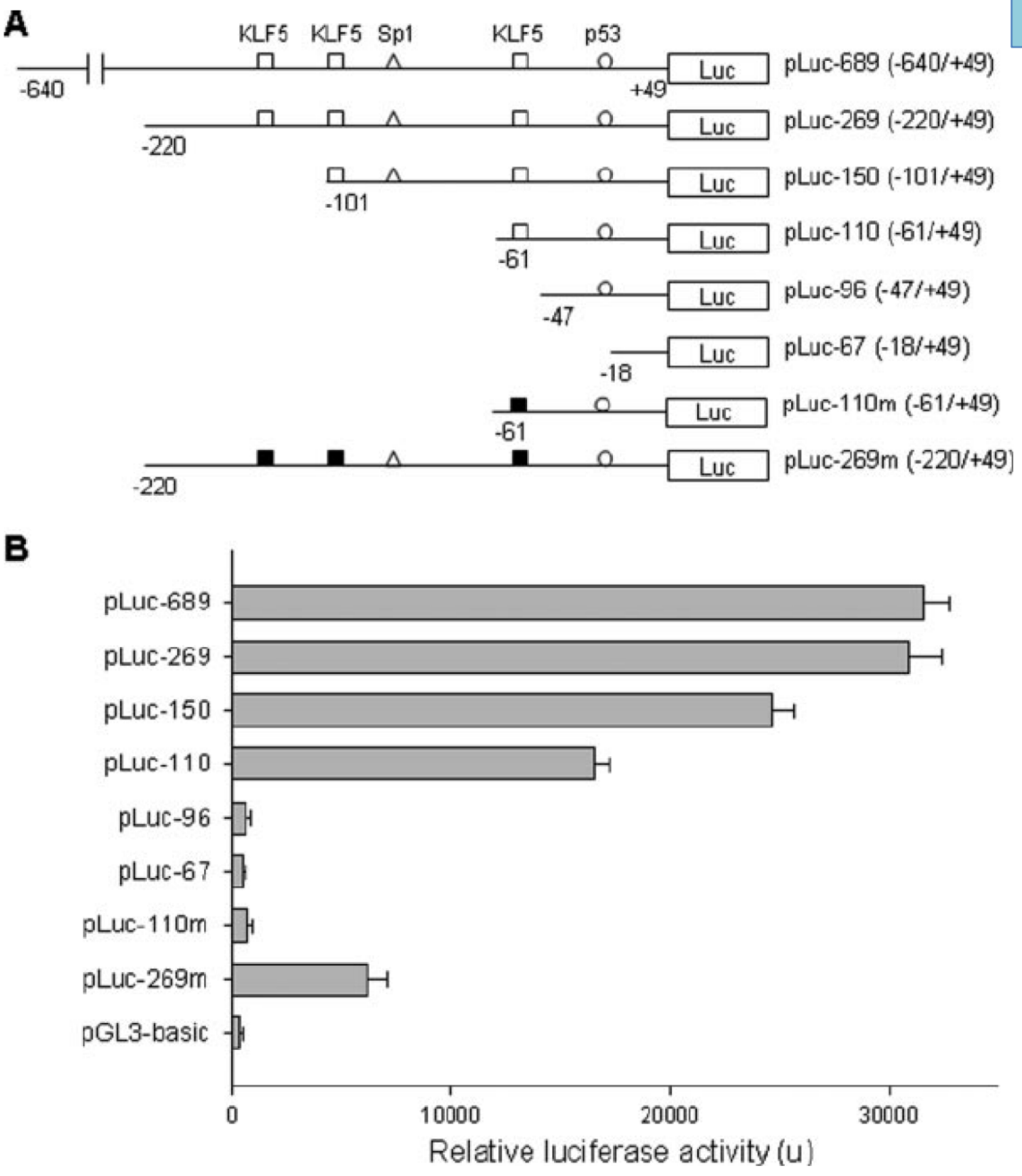
To identify promoters (one-by-one):

- Reporter assay
- EMSA
- DNase footprinting
- In vitro transcription

To identify promoters (genome-wide):

- CAGE
- 5’-SAGE
- PIC-component ChIP-Seq
- Histone PTMs/variants ChIP-Seq
- Bioinformatics

Examples of original data



Zhu et al., JBC 2006

KLF5 and p53 cooperate to regulate Survivin expression

Examples of original data

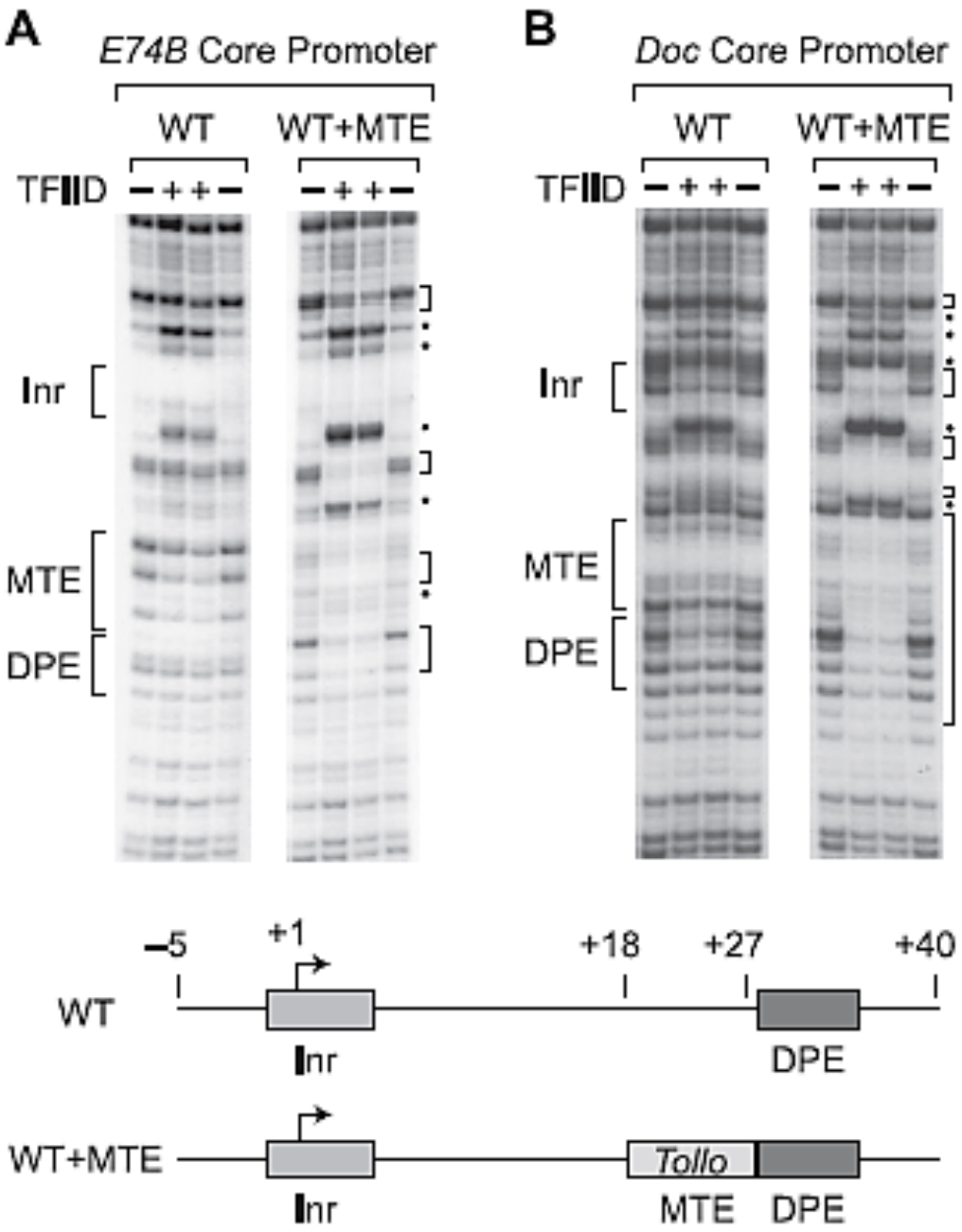


FIG. 3. The addition of an MTE increases the affinity of TFIID for the core promoter.

DNase I footprinting

Brackets=protected
*=hypersensitive

Theisen et al. 2010

Core promoter elements

- **TATA box** … ~ 30 bp upstream, consensus TATA(A/T)A(A/T)
- Instead of a TATA box, some eukaryotic (TATA-less) genes contain initiator (**Inr**) … surrounds TSS, extremely degenerate consensus sequence YYAN(T/A)YYY (A – TSS, N – any nucleotide)
- Promoters with both TATA and Inr also exist.
- **DPE** (downstream promoter element) in TATA-less
  - Present in some $TATA^-$, Inr^+ promoters, 30 bp downstream. consensus: RGWCGTG (W = A or T)

Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **2002**; *16* (20):2583-92.

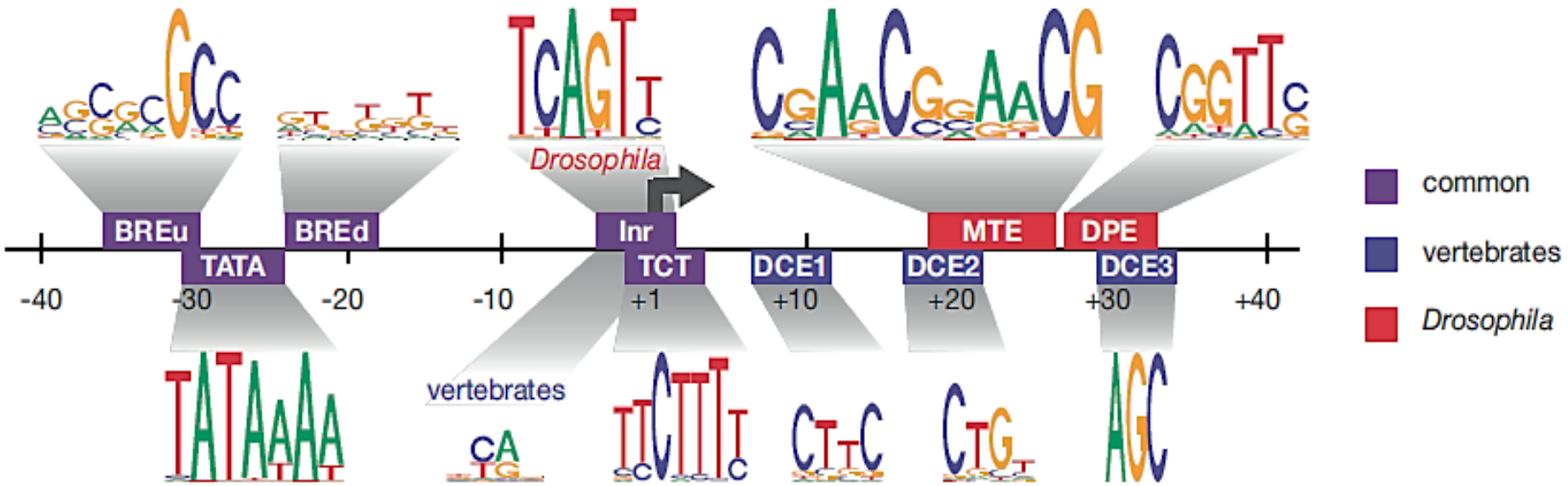
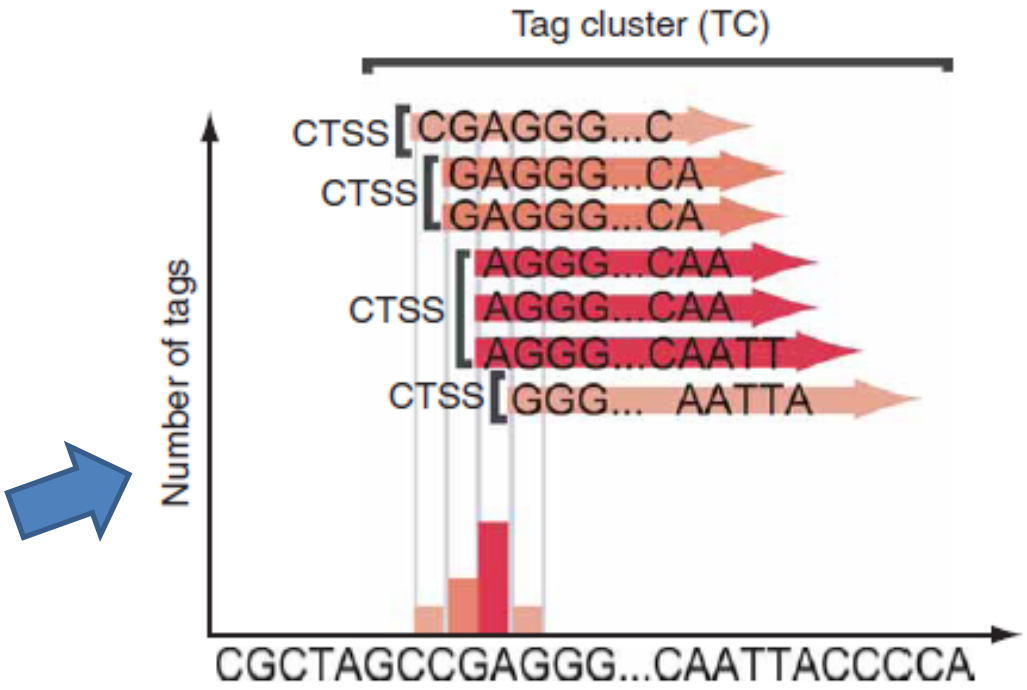


Fig. 1. Summary of most prevalent core promoter elements positionally constrained with respect to transcription start site (TSS; marked as +1 position). The location of elements relative to the TSS is shown as coloured boxes, where the colour indicates whether the element is Drosophila-specific (red), vertebrate-specific (blue) or common(purple). Associated sequence logos are based on motifs from [125] and [6] for Drosophila and motifs from the JASPAR database for vertebrates. The initiator motif (Inr) differs between Drosophila and vertebrates and both sequence logos are shown. Most promoters only have one or a few of these elements, and some elements are mostly found in certain species. BRE, TFIIB recognition element; DCE, downstream core element; DPE, downstream promoter element; Inr, initiator; MTE, motif ten element; TATA, TATA-box element; TCT, TCT initiator. IMPORTANT: hardly any real promoter contains all or even most of the above elements – on the contrary, different elements are associated with different promoter architectures and their co-occurrence in individual promoters are strongly underrepresented compared to chance.

Studies by CAGE and RNA-Seq have redefined functional groups of promoters genome-wide.

Example of results from CAGE analysis

Carninci et al., 2006



(gene sequence)

Figure 1 Definition and characteristics of CAGE tag clusters. (a) Tag clusters are produced by grouping overlapping tags on the same strand. Hence, tag clusters are defined by a start and end position, a count of tags and a distribution of these counts. Unique tag starts within the tag cluster form CAGE tag starting sites (CTSSs).

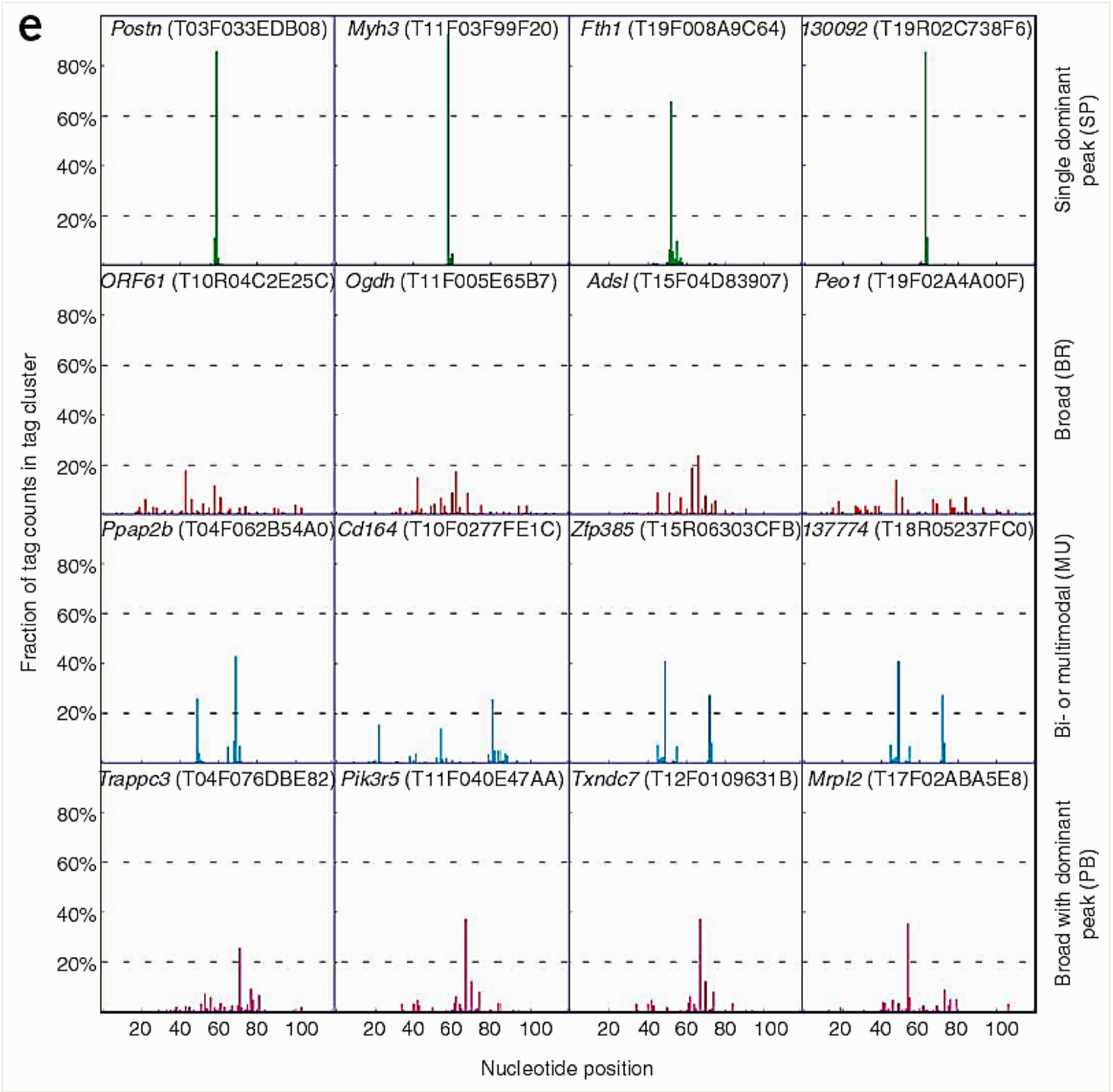
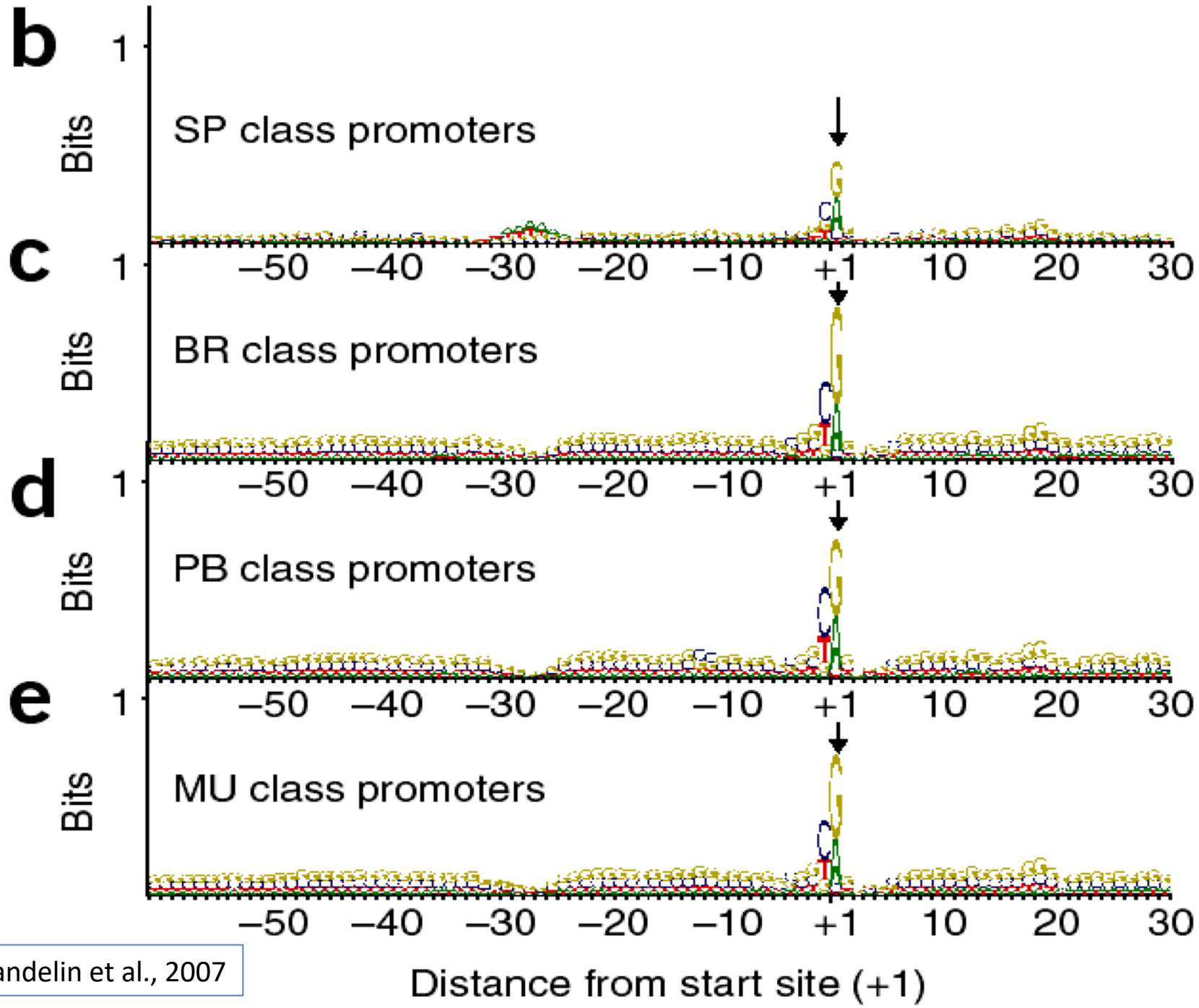


Figure 1. (e) Arrays of representative tag clusters for different shape classes. Histograms indicate the fraction of tags in the tag cluster mapping into each position in a 120-bp window centered on the tag cluster. The single peak (SP) class is characterized by a sharp peak, indicative of a single, well-defined TSS. The broad (BR) shape indicate multiple, weakly defined TSSs. The bimodal/multimodal (MU) shape class implies multiple welldefined TSSs within one cluster. Combination of a well-defined TSS surrounded by weaker TSSs results in a broad with dominant peak shape (PB). HUGO gene names or transcriptional unit identifiers for cognate genes and tag cluster identifiers are shown above each tag cluster.

Carninci et al., 2006



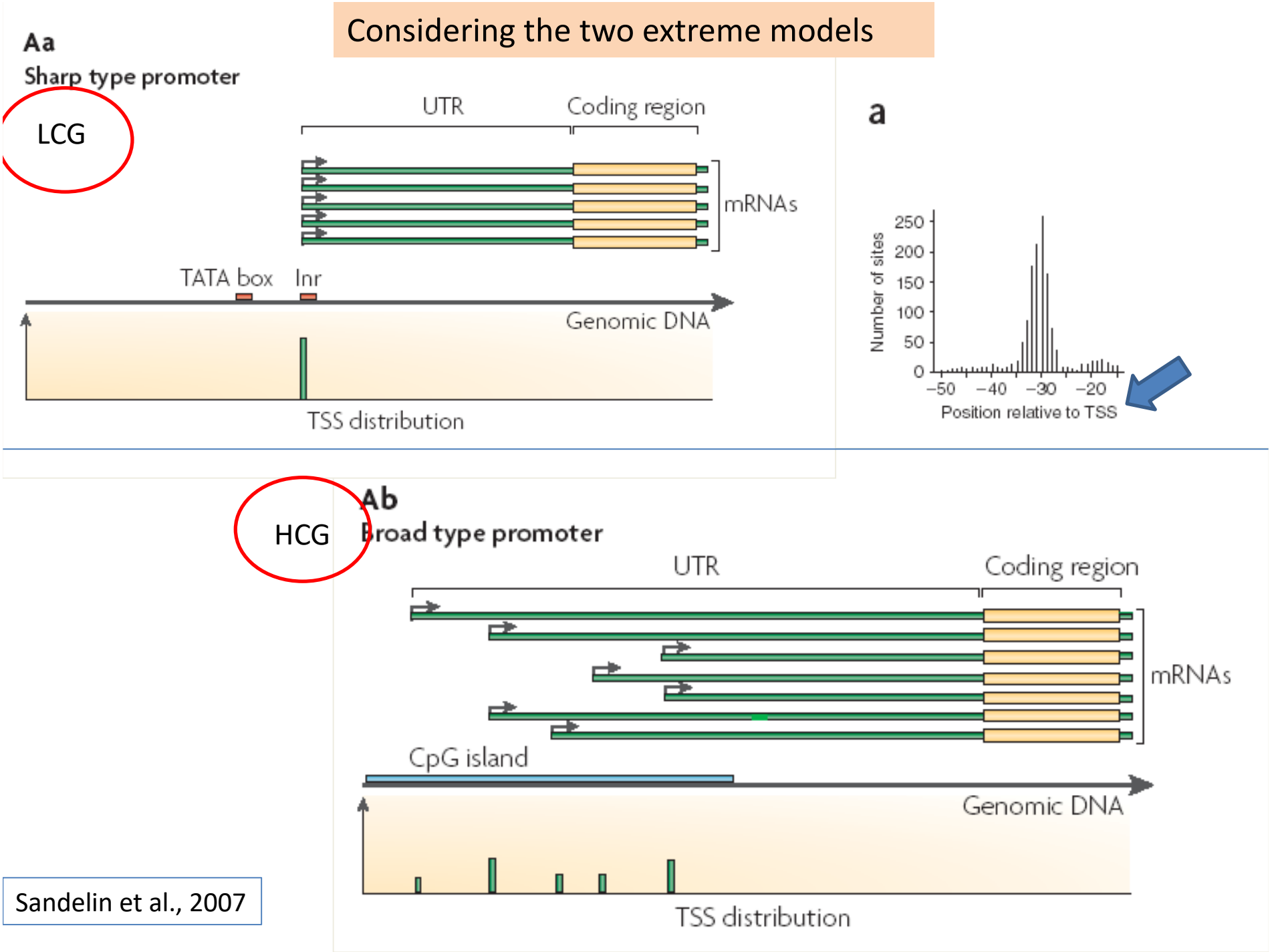
Sandelin et al., 2007

← Figure legend

Legend to the previous slide

Figure 2. TATA-box and TSS spacing definition and consensus.
(b–e) Sequence logos for promoter sequences aligned at the TSSs constructed by counting each tag and its flanking region as one sequence, divided by promoter shape class. The y axis shows the information content (measured in bits). In all cases, there is a clear preference for a pyrimidine-purine initiation site at –1,+1. A TATA-like motif is visible around the –30 position in the SP class promoters (b). In the BR class promoters, as most of those promoters are overlapped by CpG islands, the entire region is GC-rich; there is anisotropy of nucleotide content: there are more guanine than cytosine nucleotides in the plus strand upstream of the TSS (c). The logos of PB (d) and MU (e) class promoters look similar to this, indicating that these two ambiguous two categories are more likely to share the common initiation mechanism with BR promoters than with the SP ones. The PB class has a certain proportion of mixed cases, with both a CpG island and a TATA-box.

Considering the two extreme models



Sandelin et al., 2007

Genomic studies have partially changed our knowledge of promoters.

Studies oriented to define the **TSS** genome-wide, such as CAGE and 5' - SAGE, were especially instructive.

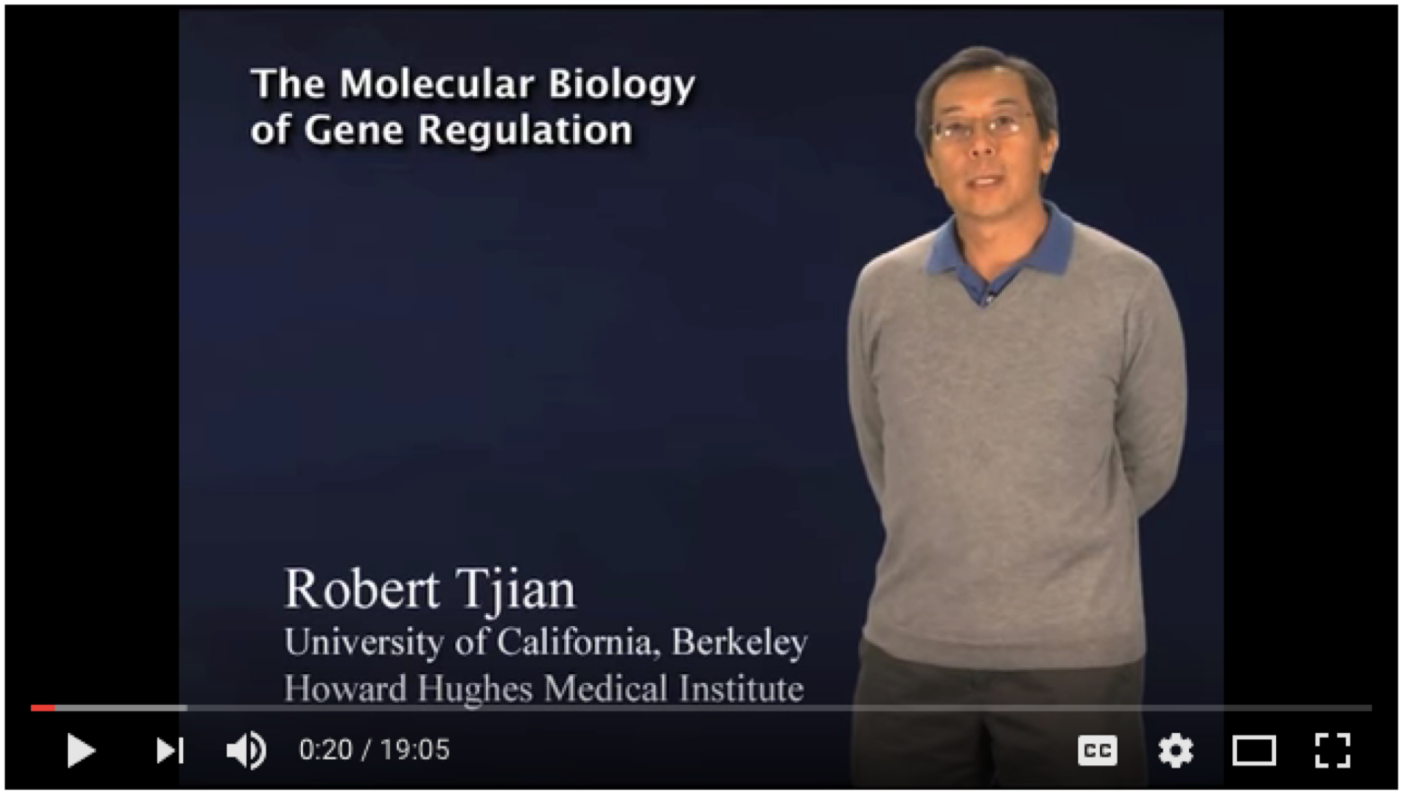
These studies demonstrated, first, that the "textbook promoter" is present at no more that 10-20% of mammalian genes (17% in human and mouse), which represent a group of inducible, tissue-specific genes.

Remaining transcription units have different structures, more often relying on CpG islands.

AGENDA

1. How to define and identify promoters
2. Different types of gene promoters
3. Transcriptional initiation: PIC and RNAPolII

Robert Tijan's lectures on Youtube



Introduction to Transcription - Robert Tjian (Berkeley/HHMI)

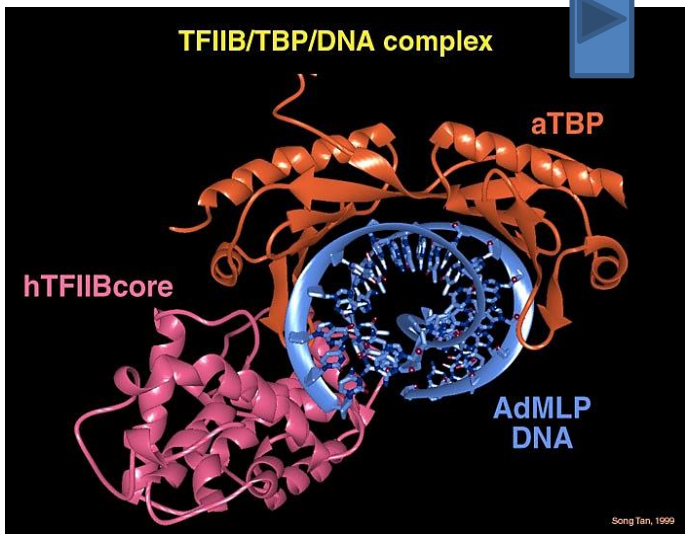
https://www.youtube.com/watch?v=ytOvnly6cgg

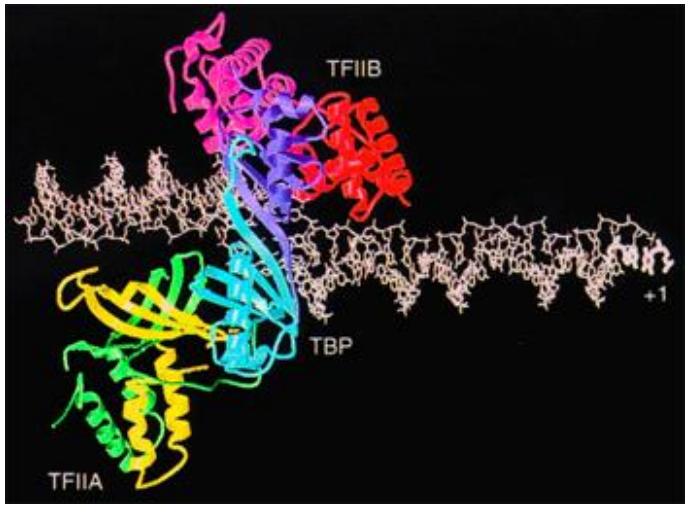
BREAK

From Thomas & Chiang, 2006

TABLE 1 **Components of the human general transcription machinery**

Factor	Protein composition	Function
TFIIA	p35 (α), p19 (β), and p12 (γ)	Antirepressor; stabilizes TBP-TATA complex; coactivator
TFIIB	p33	Start site selection; stabilize TBP-TATA complex; pol II/TFIIF recruitment
TFIID	TBP + TAFs (TAF1-TAF14)	Core promoter-binding factor Coactivator Protein kinase Ubiquitin-activating/conjugating activity Histone acetyltransferase
TFIIE	p56 (α) and p34 (β)	Recruits TFIIH Facilitates formation of an initiation-compentent pol II Involved in promoter clearance
TFIIF	RAP30 and RAP74	Binds pol II and facilitates pol II recruitment to the promoter Recruits TFIIE and TFIIH Functions with TFIIB and pol II in start site selection Facilitates pol II promoter escape Enhances the efficiency of pol II elongation
TFIIH	P89/XPB, p80/XPD, p62, p52, p44, p40/CDK7, p38/Cyclin H, p34, p32/MAT1, and p8/TFB5	ATPase activity for transcription initiation and promoter clearance Helicase activity for promoter opening Transcription-coupled nucleotide excision repair Kinase activity for phosphorylating pol II CTD E3 ubiquitin ligase activity
pol II	RPB1-RPB12	Transcription initiation, elongation, termination Recruitment of mRNA capping enzymes Transcription-coupled recruitment of splicing and 3′ end processing factors CTD phosphorylation, glycosylation, and ubiquitination





X-ray and E.M. was used to clarify the structure of RNA Pol II, as well as nearly all components of the PIC.

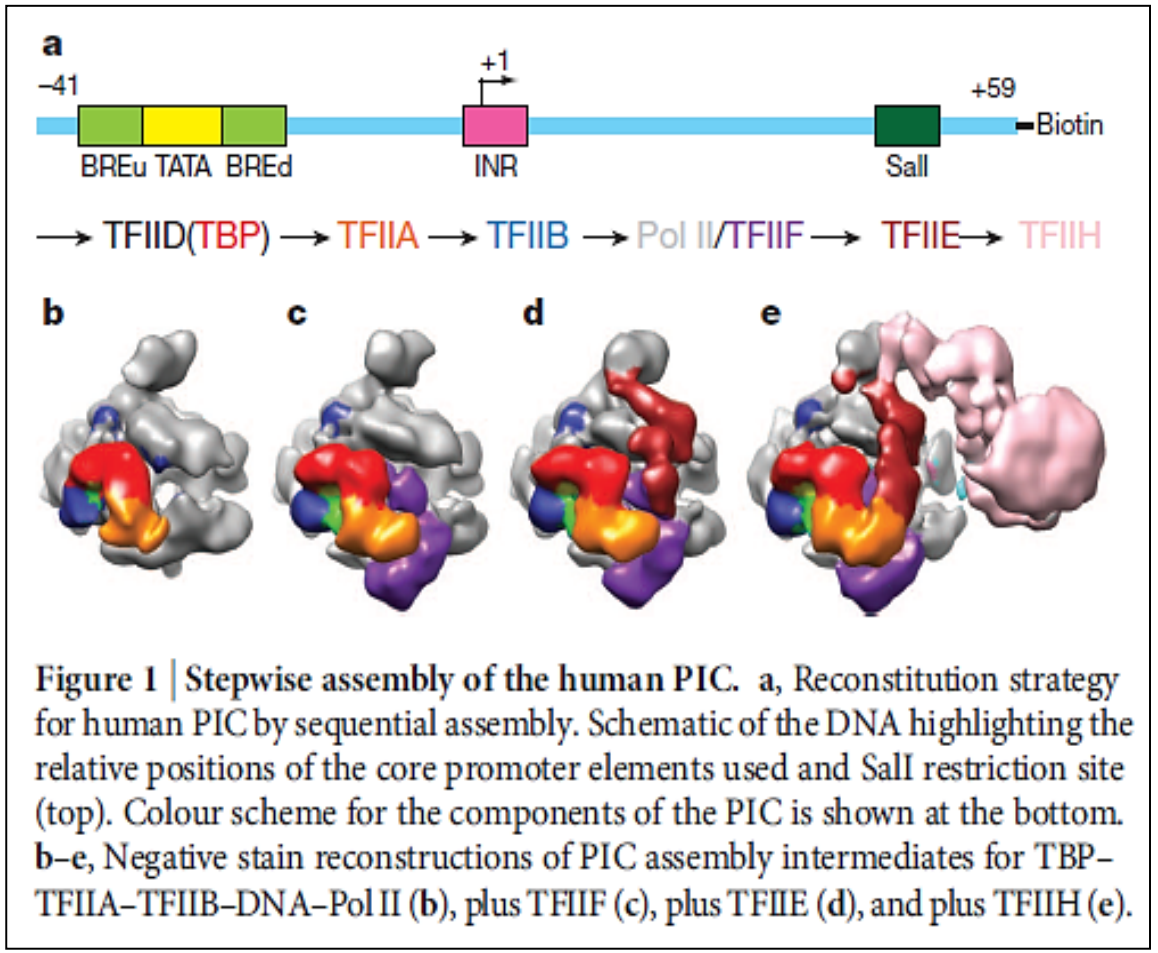
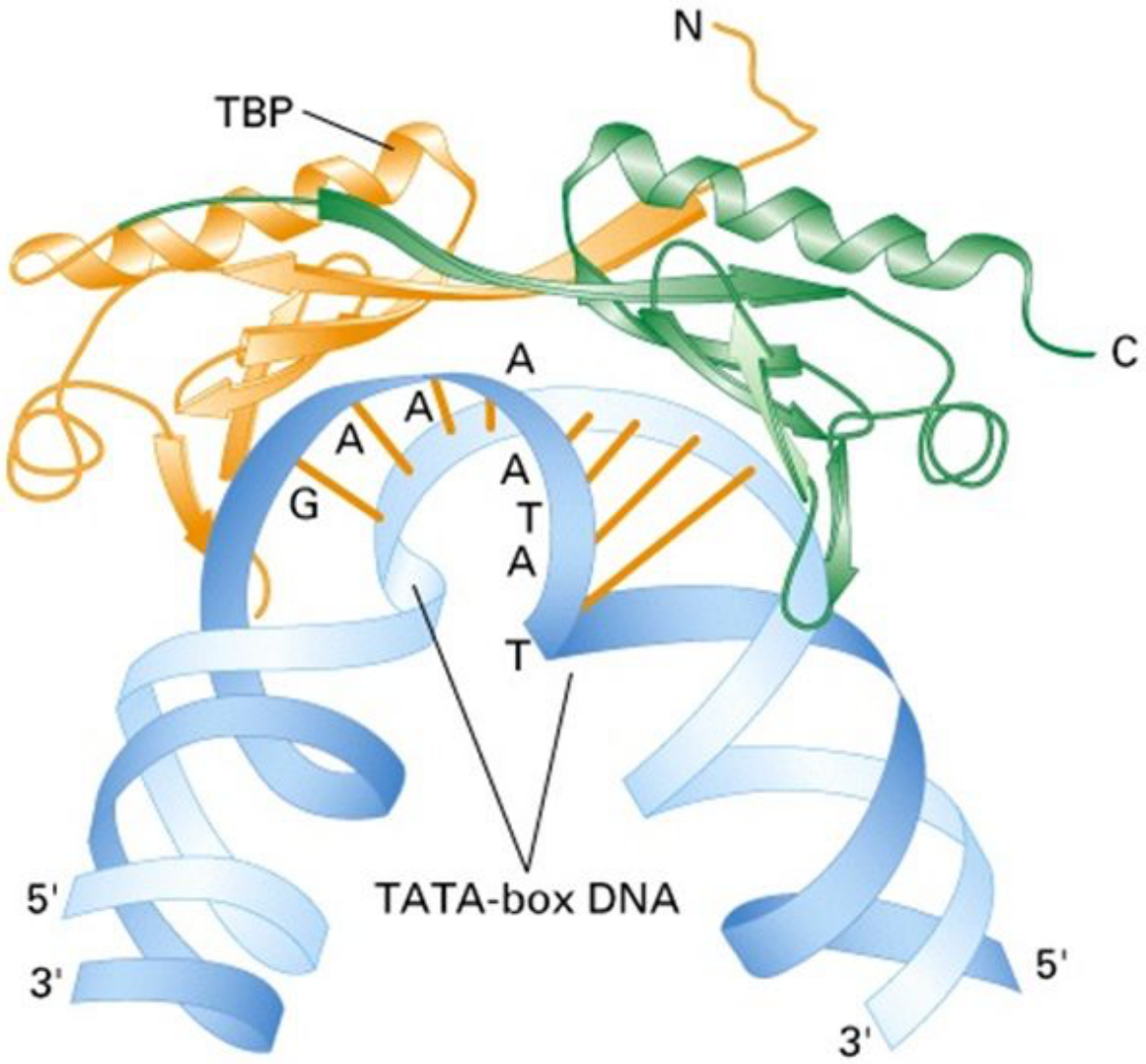


Figure 1 | Stepwise assembly of the human PIC. **a**, Reconstitution strategy for human PIC by sequential assembly. Schematic of the DNA highlighting the relative positions of the core promoter elements used and SalI restriction site (top). Colour scheme for the components of the PIC is shown at the bottom. **b–e**, Negative stain reconstructions of PIC assembly intermediates for TBP–TFIIA–TFIIB–DNA–Pol II (**b**), plus TFIIF (**c**), plus TFIIE (**d**), and plus TFIIH (**e**).

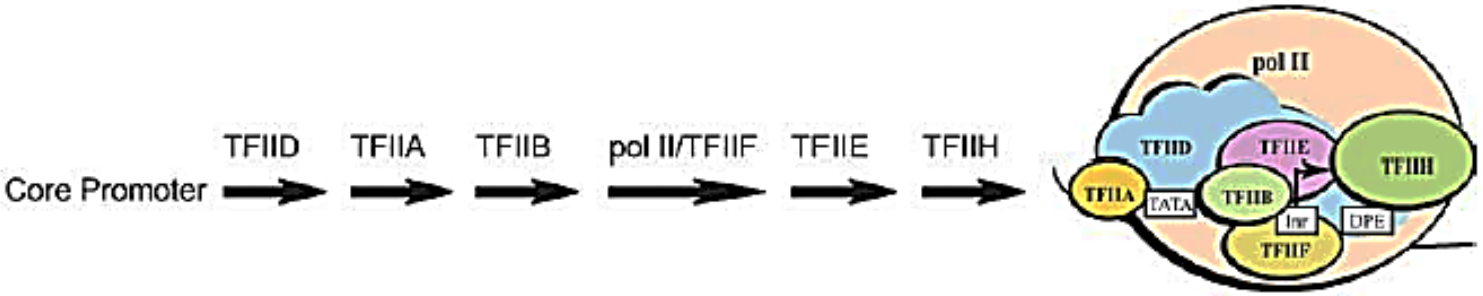
From He et al., 2013

TBP bends DNA ~80° and forces open the minor groove.



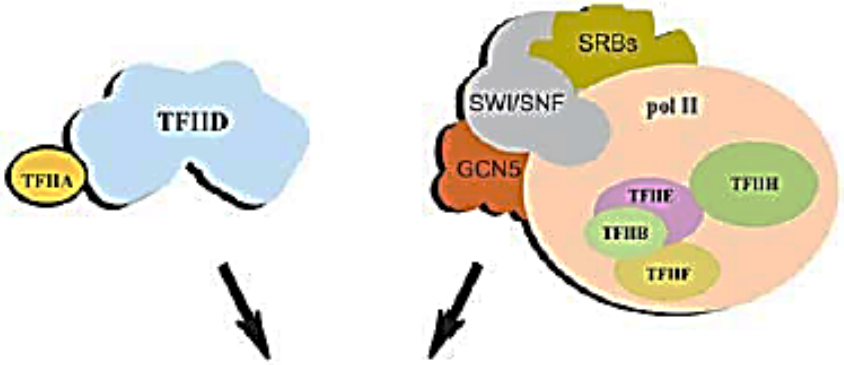
A

The Sequential Assembly Pathway



B

The RNA Polymerase II Holoenzyme Pathway (Two-Component)



cf.

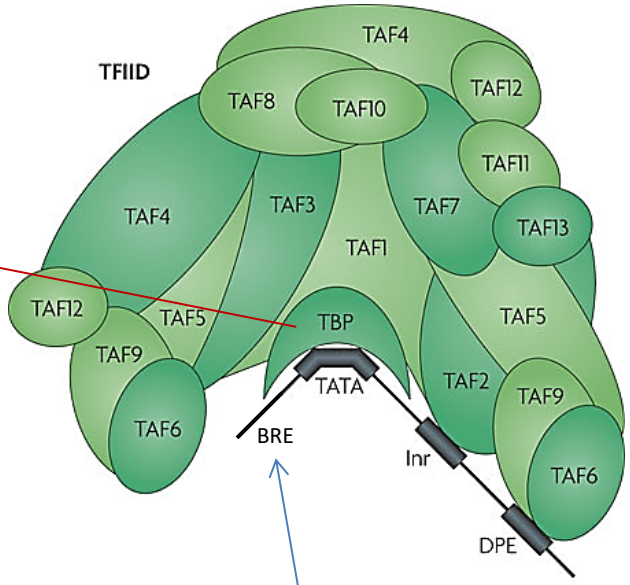
σ	$\alpha_2\beta\beta'$
specificity factor	core enzyme

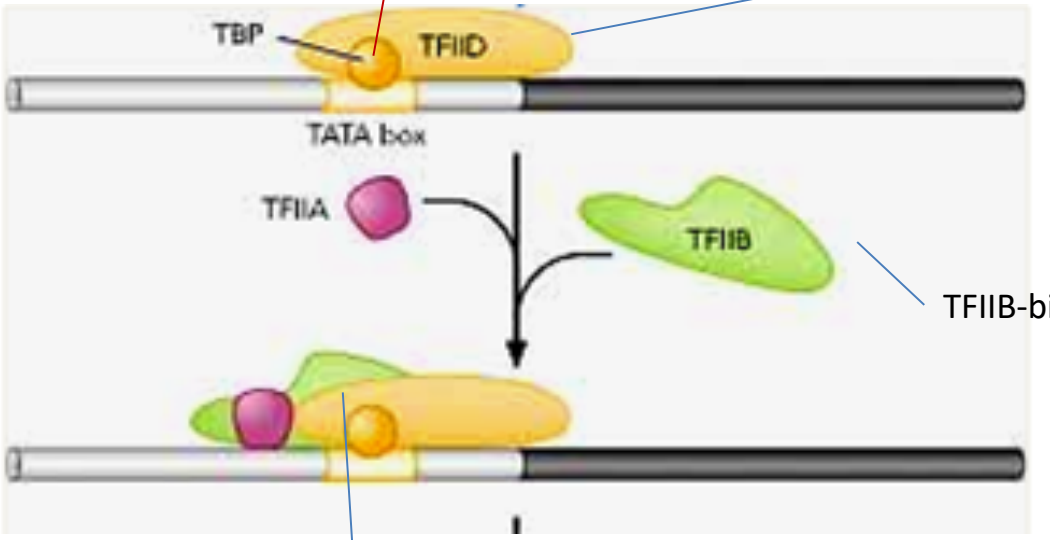
Prokaryotic RNA Polymerase Holoenzyme

From Thomas & Chiang, 2006

TAF=TBP-Associated Factors

TBP (TATA-binding protein) binds to the TATA element (if present) and bends DNA ~ 90°

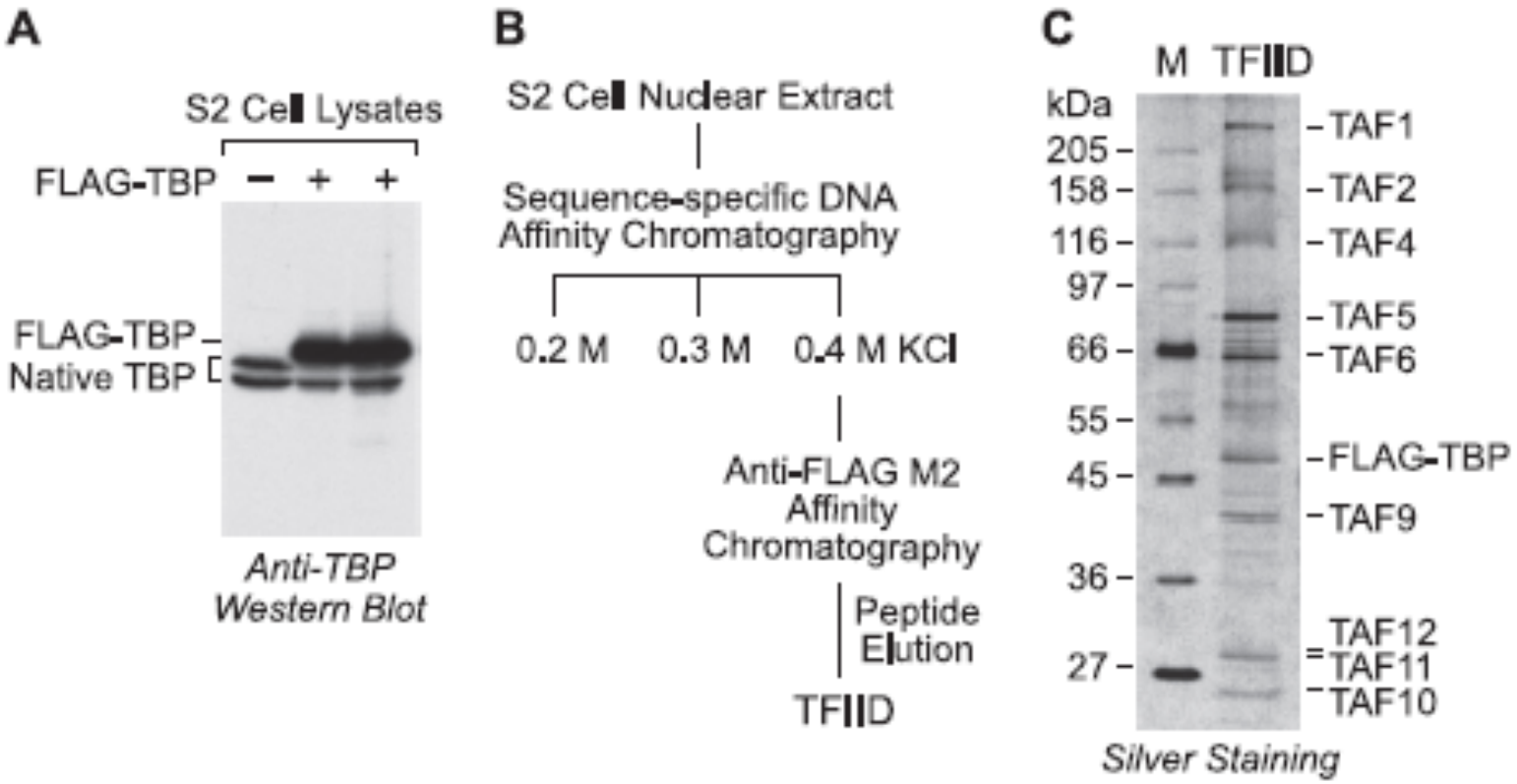




TFIIB-binds to TBP and to BRE

TFIIA stabilizes the complex

Examples of original data

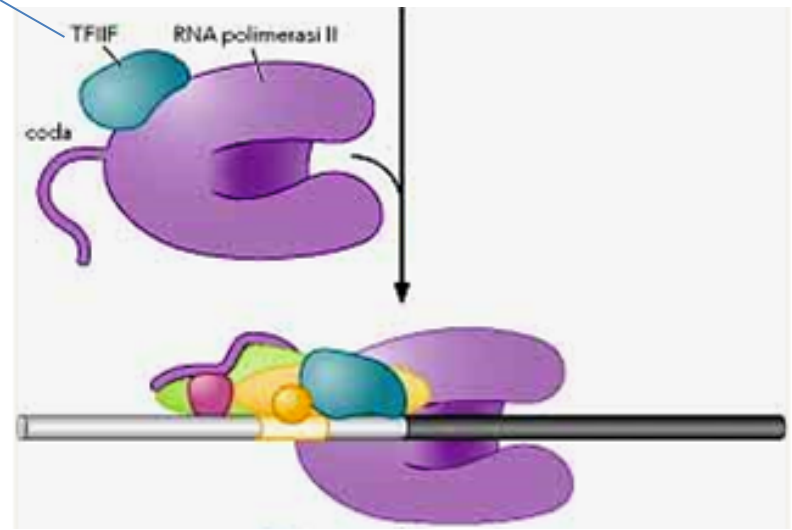


drosophila

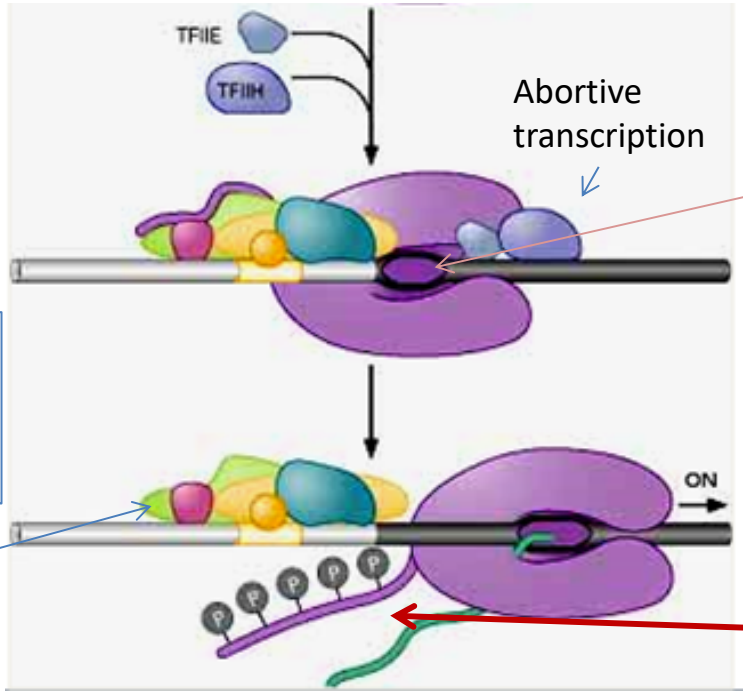
TAF proteins purified from S2 cells nuclear extract by anti-FLAG affinity chromatography

Theisen et al. 2010

$TF_{II}F$ facilitates RNA PolII recruitment to the complex



$TF_{II}E$ + $TF_{II}H$ join



Abortive transcription

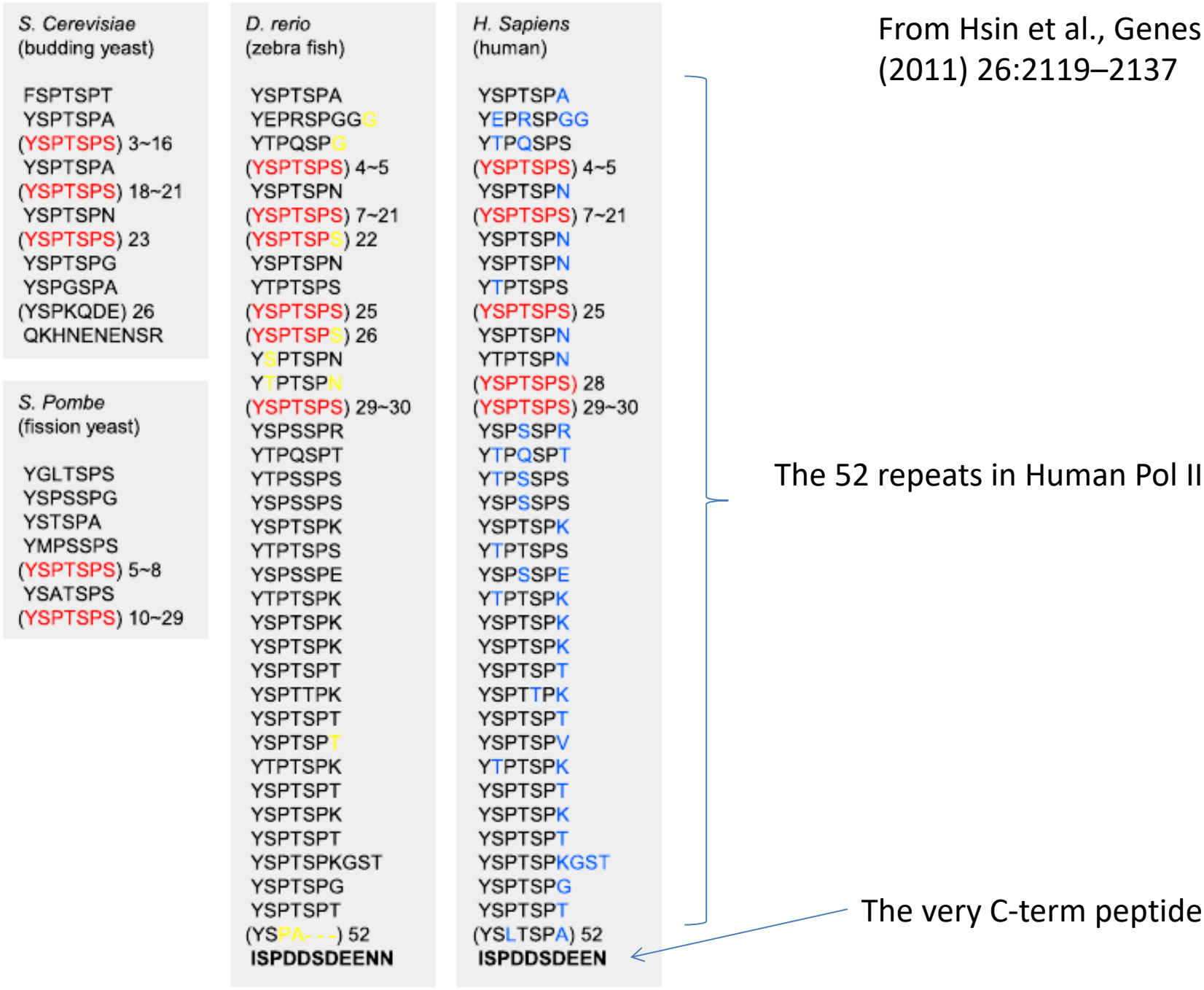
«open» complex

Elongation factor(s) $TF_{II}S$

RNA PolII CTD is phosphorylated and initiates transcription (promoter clearance)

Factors may be left behind to re-initiate

CTD


S. Cerevisiae
(budding yeast)
FSPTSPT
YSPTSPA
(YSPTSPS) 3~16
YSPTSPA
(YSPTSPS) 18~21
YSPTSPN
(YSPTSPS) 23
YSPTSPG
YSPGSPA
(YSPKQDE) 26
QKHNENENSR
S. Pombe
(fission yeast)
YGLTSPS
YSPSSPG
YSTSPA
YMPSSPS
(YSPTSPS) 5~8
YSATSPS
(YSPTSPS) 10~29
D. rerio
(zebra fish)
YSPTSPA
YEPRSPGGG
YTPQSPG
(YSPTSPS) 4~5
YSPTSPN
(YSPTSPS) 7~21
(YSPTSPS) 22
YSPTSPN
YTPTSPS
(YSPTSPS) 25
(YSPTSPS) 26
YSPTSPN
YTPTSPN
(YSPTSPS) 29~30
YSPSSPR
YTPQSPT
YTPSSPS
YSPSSPS
YSPTSPK
YTPTSPS
YSPSSPE
YTPTSPK
YSPTSPK
YSPTSPK
YSPTSPT
YSPTTPK
YSPTSPT
YSPTSPT
YTPTSPK
YSPTSPT
YSPTSPK
YSPTSPT
YSPTSPKGST
YSPTSPG
YSPTSPT
(YSPA- - -) 52
ISPDDSDEENN
H. Sapiens
(human)
YSPTSPA
YEPRSPGG
YTPQSPS
(YSPTSPS) 4~5
YSPTSPN
(YSPTSPS) 7~21
YSPTSPN
YSPTSPN
YTPTSPS
(YSPTSPS) 25
YSPTSPN
YTPTSPN
(YSPTSPS) 28
(YSPTSPS) 29~30
YSPSSPR
YTPQSPT
YTPSSPS
YSPSSPS
YSPTSPK
YTPTSPS
YSPSSPE
YTPTSPK
YSPTSPK
YSPTSPK
YSPTSPT
YSPTTPK
YSPTSPT
YSPTSPV
YTPTSPK
YSPTSPT
YSPTSPK
YSPTSPT
YSPTSPKGST
YSPTSPG
YSPTSPT
(YSLTSPA) 52
ISPDDSDEEN
From Hsin et al., Genes Dev (2011) 26:2119–2137
The 52 repeats in Human Pol II
The very C-term peptide

TRANSCRIPTION STEPS and POL2 STATUS

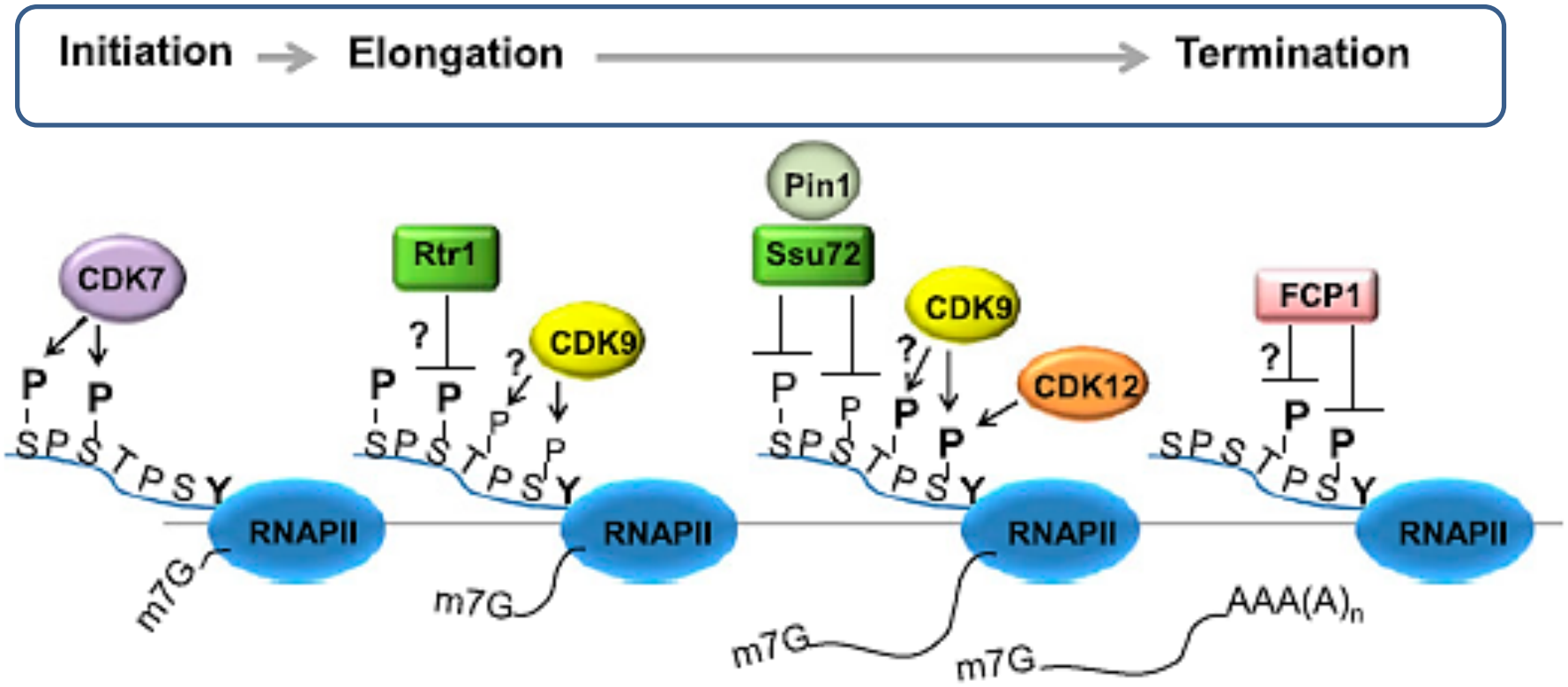


Figure 2. Dynamic modification of the CTD during the transcription cycle. At transcription initiation, CDK7 phosphorylates Ser5 and Ser7 residues. Later, during elongation, CDK9 phosphorylates Ser2 and perhaps Thr4, while the phosphate groups on Ser5 and Ser7 are gradually removed by phosphatases. (…)

From Hsin et al., Genes Dev (2011) 26:2119–2137

MORE on PIC complex functions and specificity:

Discussion question

Is this a core machinery and specificity comes from upstream regulatory regions or there are also Gene-specific/Stage-specific TAF functions?

Cell, Vol. 91, 71–83, October 3, 1997, Copyright ©1997 by Cell Press

Transcription Properties of a Cell Type–Specific TATA-Binding Protein, TRF

Stig K. Hansen,* Shinako Takada,*
Raymond H. Jacobson,* John T. Lis,†
and Robert Tjian*
*Department of Molecular and Cell Biology
Howard Hughes Medical Institute
University of California
Berkeley, California 94720-3204
†Section of Biochemistry
Molecular and Cell Biology Biotechnology Building
Cornell University
Ithaca, New York 14853-2801

Summary

Eukaryotic cells are thought to contain a single TATA-binding protein (TBP) that directs transcription by cellular RNA polymerases. Here we report a cell type–specific TBP-related factor (TRF) that can form a stable TRF/IIA/IIB TATA DNA complex and substitute for TBP in directing RNA polymerase II transcription in vitro. Transfection studies reveal that TRF can differentially mediate activation by some enhancer proteins but not others. Like TBP, TRF forms a stable complex containing multiple novel subunits, nTAFs. Antibody staining of embryos and polytene chromosomes reveals cell type–specific expression and gene-selective properties consistent with the shaker/male sterile phenotype of *trf* mutants. These findings suggest TRF is a homolog of TBP that functions to direct tissue- and gene-specific transcription.

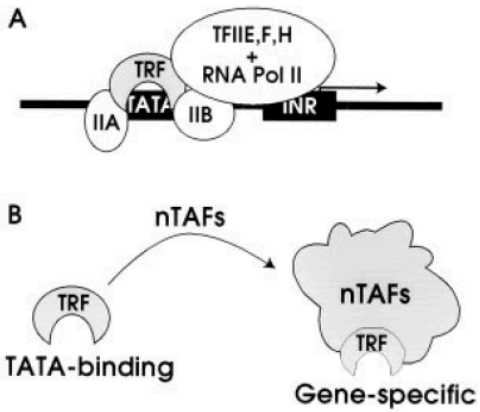
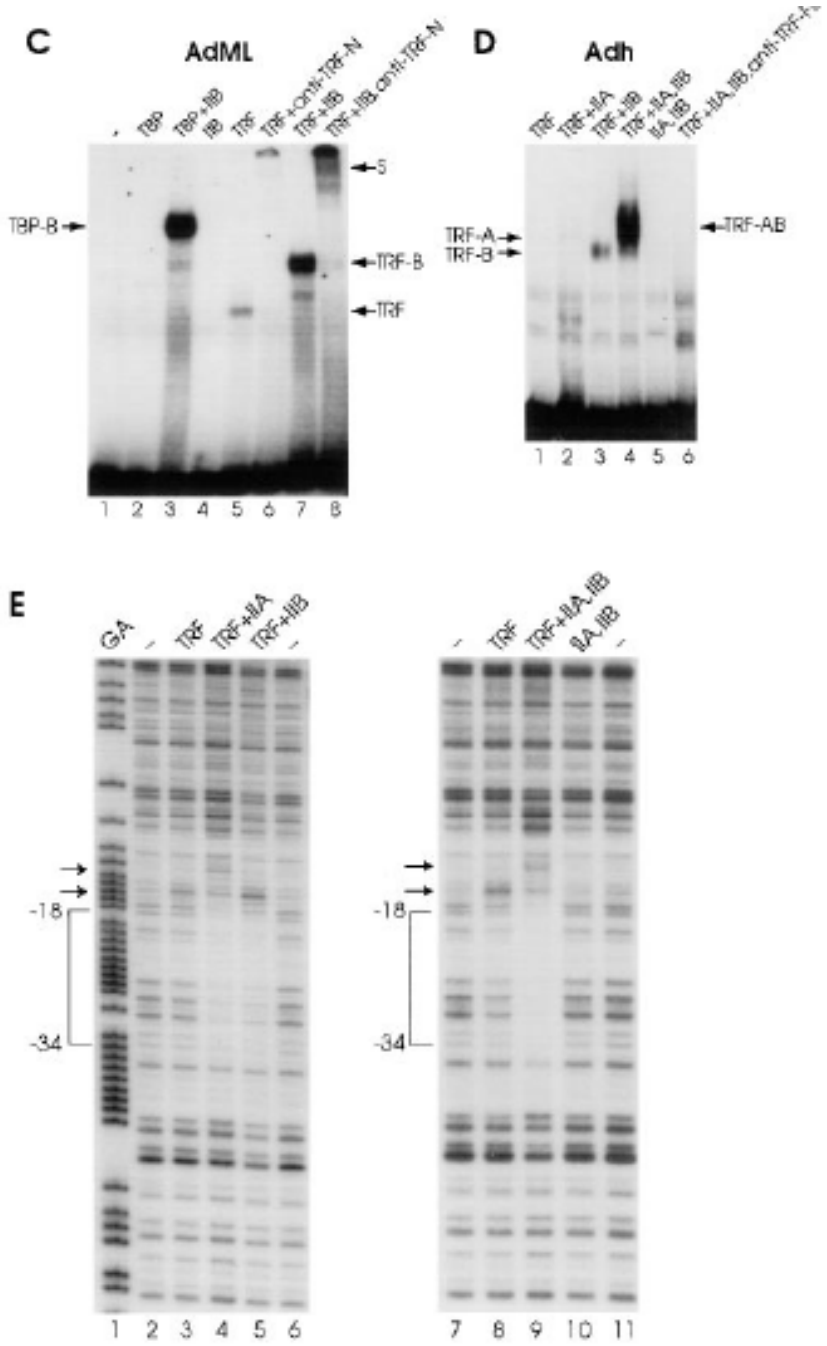


Figure 9. Model of TRF Function



(C) Band shift experiments using an oligonucleotide containing the AdML TATA box sequence. TBP, TRF, TFIIA, and TFIIB were expressed in bacteria and purified to homogeneity. The positions of the various shifts are indicated at the side; S indicates super shift caused by antibody binding. Labeled DNA was incubated with no protein (lane 1); 2 ng TBP (lane 2); 2 ng TBP and 50 ng TFIIB (lane 3); 50 ng TFIIB (lane 4); 50 ng TRF (lane 5); 50 ng TRF and 300 ng anti-TRF-N (lane 6); 50 ng TRF and 50 ng TFIIB (lane 7); and 50 ng TRF, 50 ng TFIIB, and 300 ng anti-TRF-N (lane 8).

(D) Band shift experiments using an oligonucleotide containing Drosophila *Adh* distal core promoter sequences. Protein amounts were 10 ng TRF (lane 1); 10 ng TRF and 5 ng TFIIA (lane 2); 10 ng TRF and 10 ng TFIIB (lane 3); 10 ng TRF, 5 ng TFIIA, and 10 ng TFIIB (lane 4); 5 ng TFIIA and 10 ng TFIIB (lane 5); and the same as lane 4 but in the presence of 1 μg anti-TRF-FL (lane 6).

(E) Footprint analysis of TRF, TFIIA, and TFIIB using the AdML promoter radiolabeled on the coding strand. Thirty nanograms of TRF protein was used in lanes 3–5, 8, and 9. Three hundred nanograms of TFIIA was used in lanes 4, 9, and 10, whereas 500 ng of TFIIB was included in lanes 5, 9, and 10. Boundaries of the TRF footprint and positions of hypersensitive sites are indicated.

Figure 1. TRF Interaction with TFIIA, TFIIB, and DNA

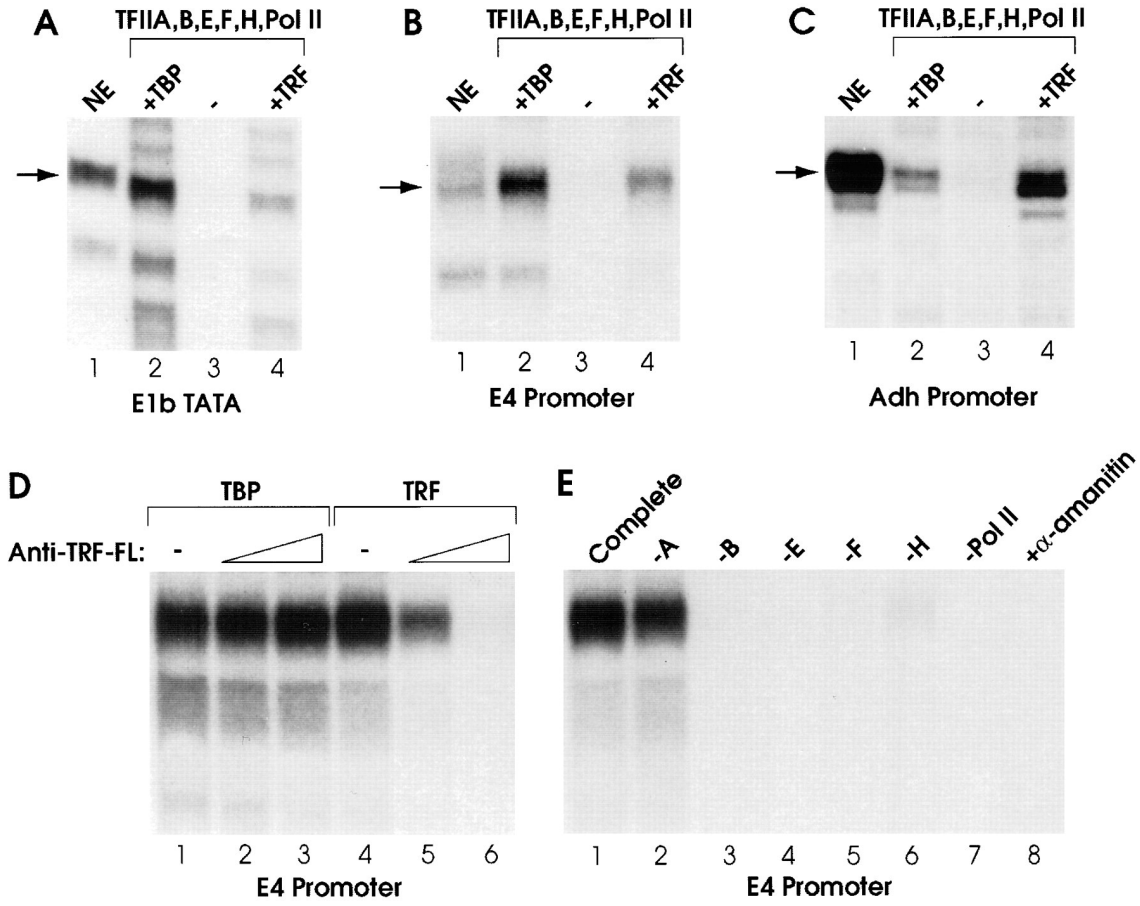


Figure 2. Reconstituted In Vitro Transcription Using TBP or TRF

In vitro transcription was performed with supercoiled templates using either Drosophila embryo NE or a transcription system reconstituted from purified components (see Experimental Procedures for details).

(A) The E1b TATA template is composed of a 10-mer oligonucleotide encompassing the TATA box sequence from the adenoviral E1b promoter inserted in the pUC polylinker. Transcription was reconstituted with TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, RNA polymerase II, and 5 ng TBP (lane 2), no TBP (lane 3), or 20 ng TRF (lane 4).

(B) Same as (A) except that the transcription template contained sequences from position −38 to +250 of the adenoviral E4 promoter.

(C) Same as (A) except that the transcription template contains the Drosophila *Adh* distal promoter from position −61 to position +325.

(D) Transcription reconstituted with purified GTFs and 5 ng TBP (lanes 1–3) or 20 ng TRF (lanes 4–6) in the absence (lanes 1 and 4) or presence of increasing amounts of anti-TRF-FL: 1 μg (lanes 2 and 5), 2.5 μg (lanes 3 and 6). The transcription template contained the E4 promoter.

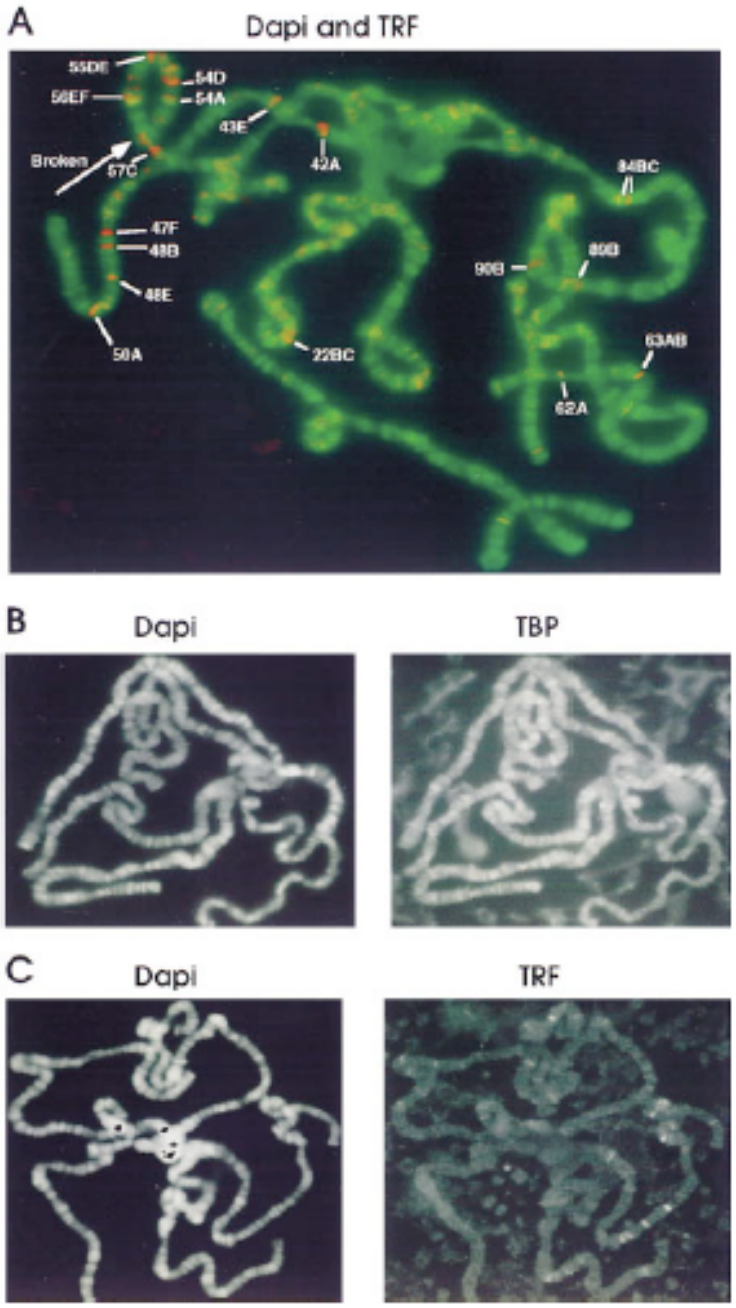
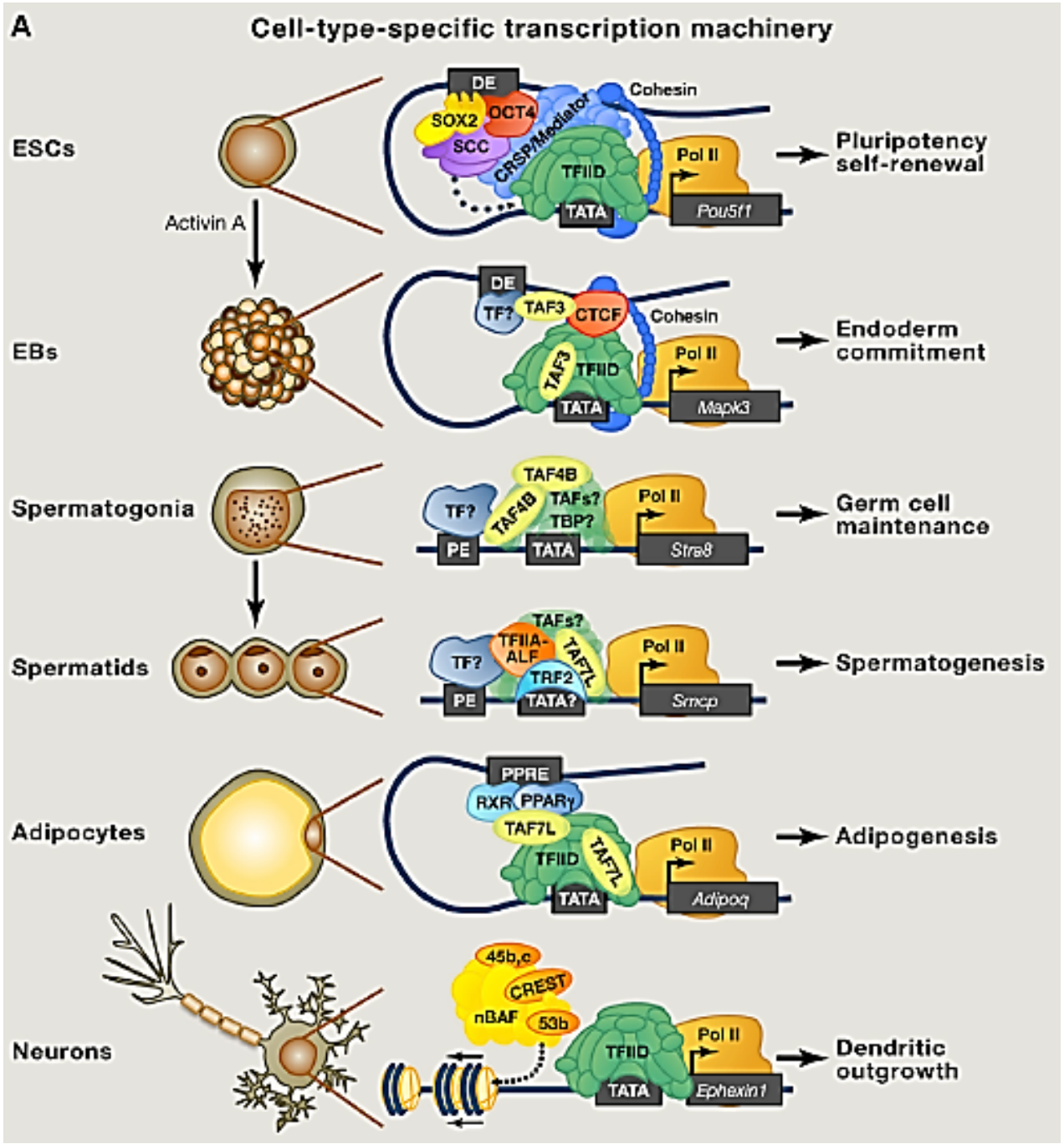


Figure 8. TRF and TBP Localization on Drosophila Polytene Chromosomes

sistently scored TRF at 17 out of the 300–600 resolvable sites, suggesting that TRF is highly gene-specific, at least in the salivary gland. In contrast to TRF, staining of polytene chromosomes with anti-TBP antibodies revealed a ubiquitous staining pattern with many strongly stained sites, suggesting that TBP is present at many genes throughout the genome (Figures 8B and 8C). These data suggest that TBP mediates transcription of many genes, while TRF, in addition to being tissue-specific, may function to direct transcription of a considerably smaller subset of genes.

Many of the identified TRF sites contain genes whose associated phenotypes or expression patterns correlate with those of *trf* (FlyBase–CytoSearch, Harvard Univer-



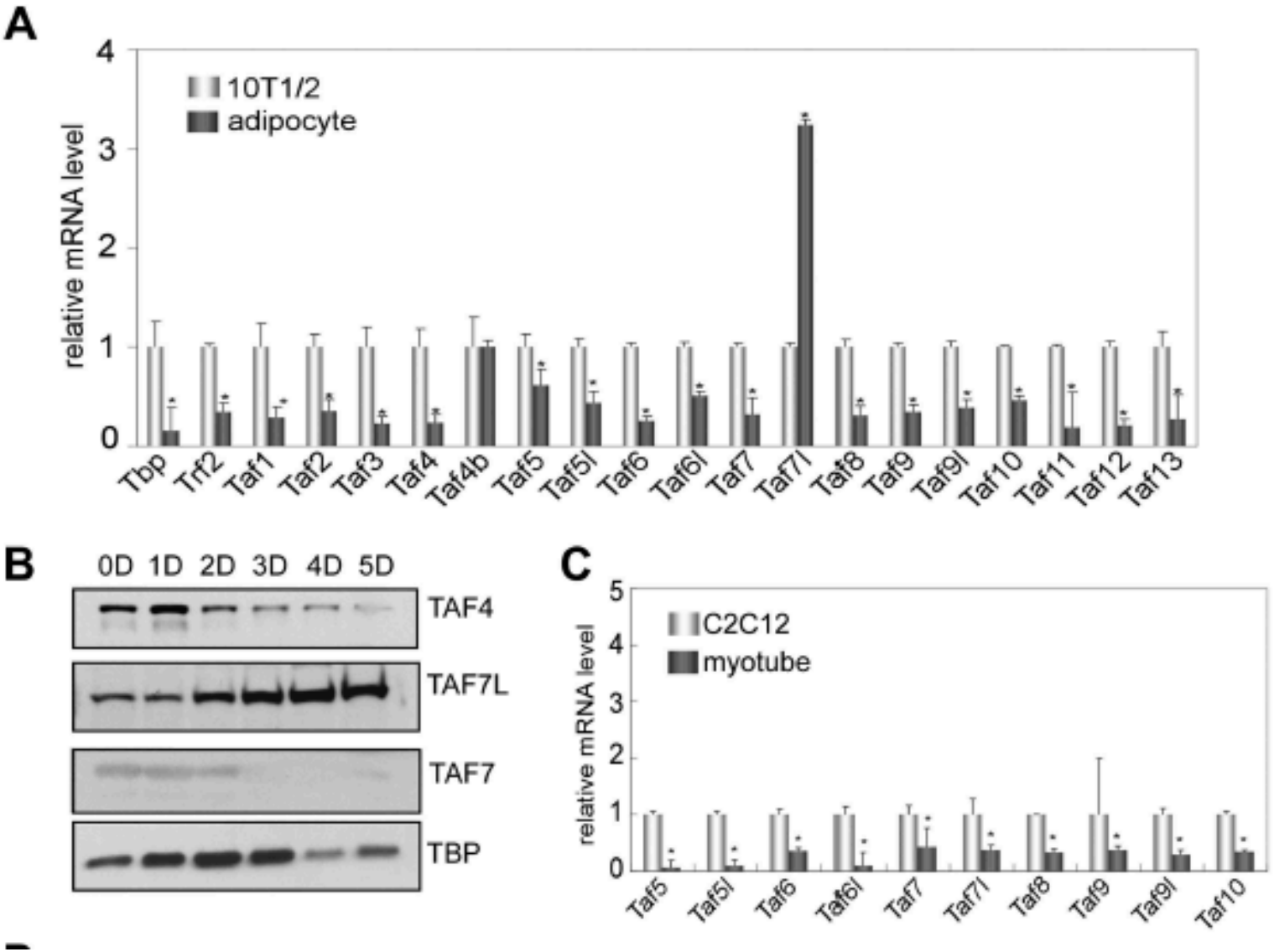
Levine et al. 2014

eLIFE
elife.elifesciences.org

Dual functions of TAF7L in adipocyte differentiation

Haiying Zhou[1,2], Tommy Kaplan[3,4], Yan Li[5], Ivan Grubisic[2,6], Zhengjian Zhang[5], P Jeremy Wang[7], Michael B Eisen[1,3], Robert Tjian[1,2]*

Abstract The diverse transcriptional mechanisms governing cellular differentiation and development of mammalian tissue remains poorly understood. Here we report that TAF7L, a paralogue of TFIID subunit TAF7, is enriched in adipocytes and white fat tissue (WAT) in mouse. Depletion of TAF7L reduced adipocyte-specific gene expression, compromised adipocyte differentiation, and WAT development as well. Ectopic expression of TAF7L in myoblasts reprograms these muscle precursors into adipocytes upon induction. Genome-wide mRNA-seq expression profiling and ChIP-seq binding studies confirmed that TAF7L is required for activating adipocyte-specific genes via a dual mechanism wherein it interacts with PPARγ at enhancers and TBP/Pol II at core promoters. In vitro binding studies confirmed that TAF7L forms complexes with both TBP and PPARγ. These findings suggest that TAF7L plays an integral role in adipocyte gene expression by targeting enhancers as a cofactor for PPARγ and promoters as a component of the core transcriptional machinery.



Zhou et al., eLIFE, 2013

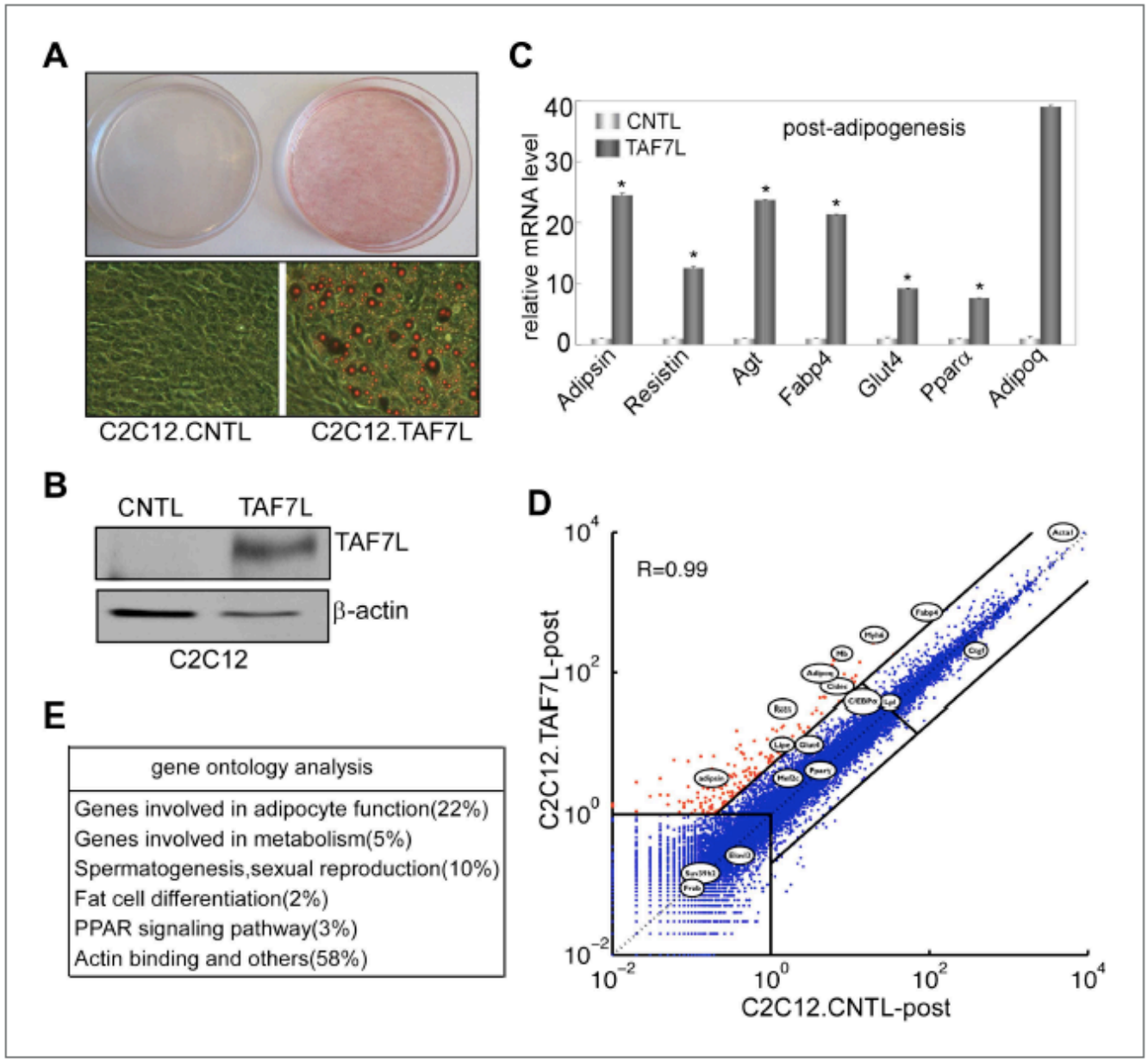


Figure 5. Ectopic expression of TAF7L transdifferentiates C2C12 myoblasts into adipocytes under adipogenic induction.

Zhou et al., eLIFE, 2013

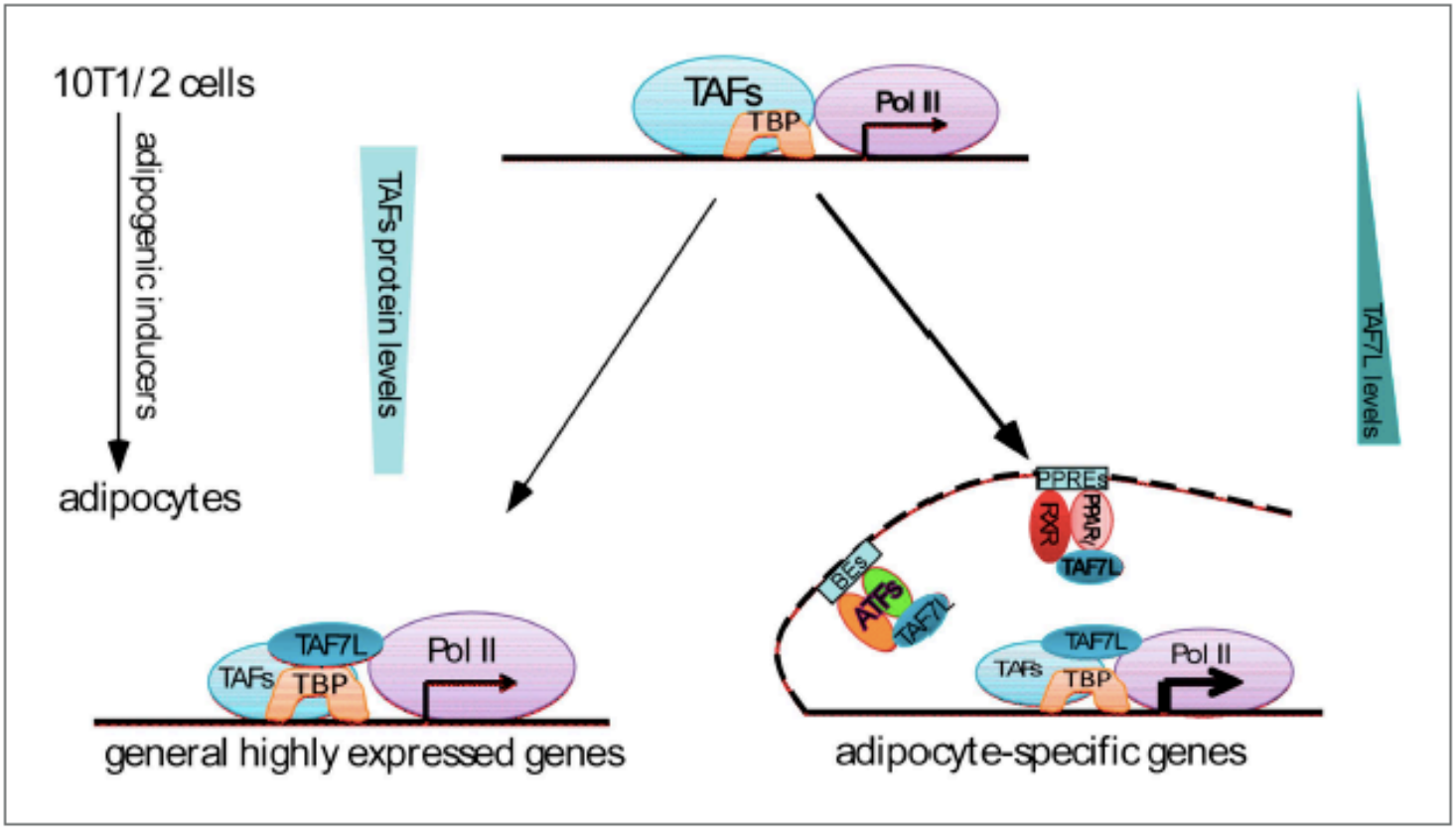


Figure 8. Dual functions of TAF7L in adipocyte differentiation. TAF7L expression is enriched during C3H10T1/2 MSCs adipocyte differentiation while other TFIID subunits (TAFs) decrease in expression. TAF7L regulates adipogenesis by associating with TBP as a component of adipocyte TFIID complex at promoters and with PPARγ or other adipocyte transcriptional factors (ATFs) as a cofactor at enhancers on adipocyte-specific genes, providing the mechanisms of its dual roles during differentiation. General highly-expressed genes are those with high expression before and after adipocyte differentiation include a portion of housekeeping genes; adipocyte-specific genes are those required for adipocyte differentiation and highly upregulated during adipocyte differentiation. TAFs,TBP-associated factors; ATFs, adipocyte transcriptional factors; BEs, binding elements.

Zhou et al., eLIFE, 2013