L3.5

Promoters

L3.5 - Agenda

1. Basics (BM book)
2. Mapping of TSS
3. Chromatin marks – CpG methylaion
4. More than one type of gene promoter
5. Modes of transcriptional activation

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.

First, let us quickly revise basics on basal transcription complexes for RNA Polymerase II (the «PIC»)

NOTE: basics on

RNA Polymerase function, steps in initiation, elongation, termination

as well as

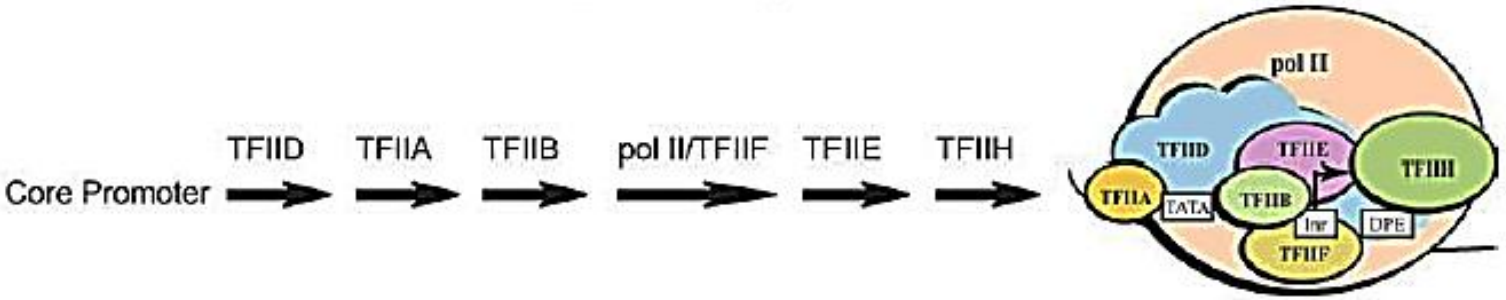
PIC composition and functions TFII proteins, PolII-CTD phosphorylation, etc.

were required for admission to this Master and revising ut is **up to you**!

Quick revision

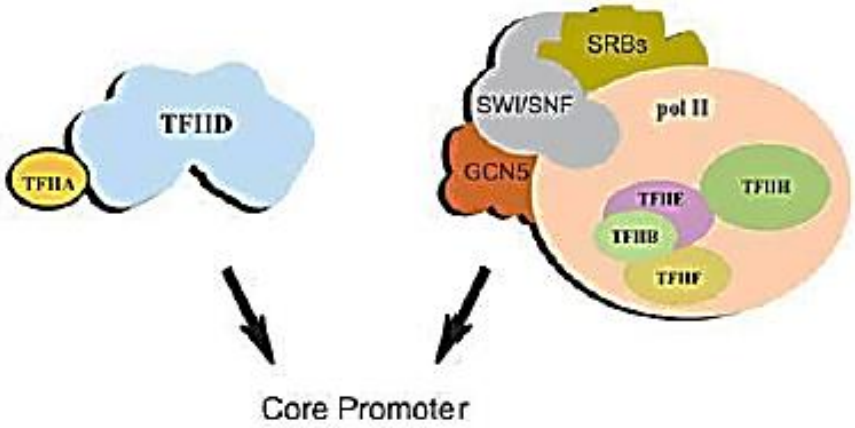
A

The Sequential Assembly Pathway



B

The RNA Polymerase II Holoenzyme Pathway (Two-Component)

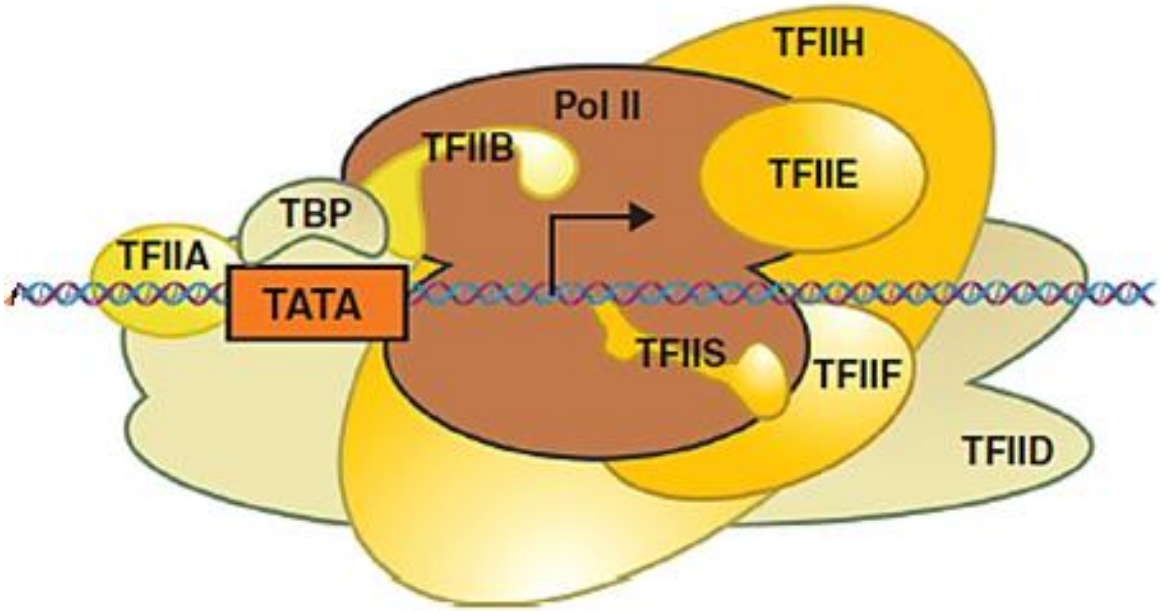


cf.

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

Prokaryotic RNA Polymerase Holoenzyme

From Thomas & Chiang, 2006



General Pol II machinery

Current Opinion in Cell Biology

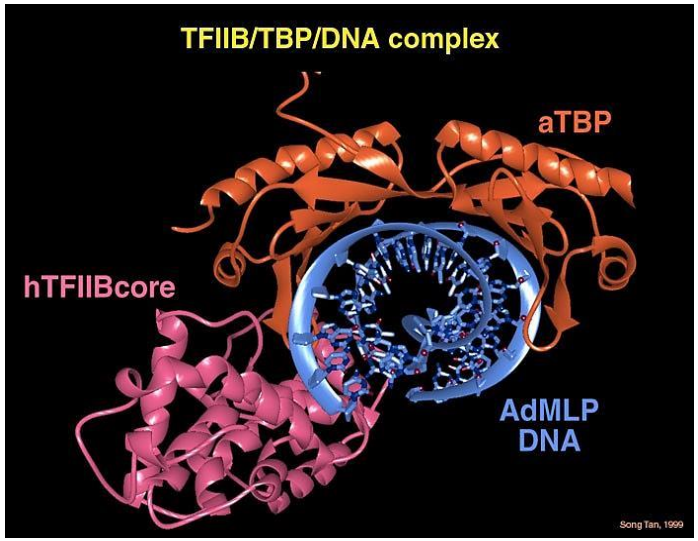
From Thomas & Chiang, 2006

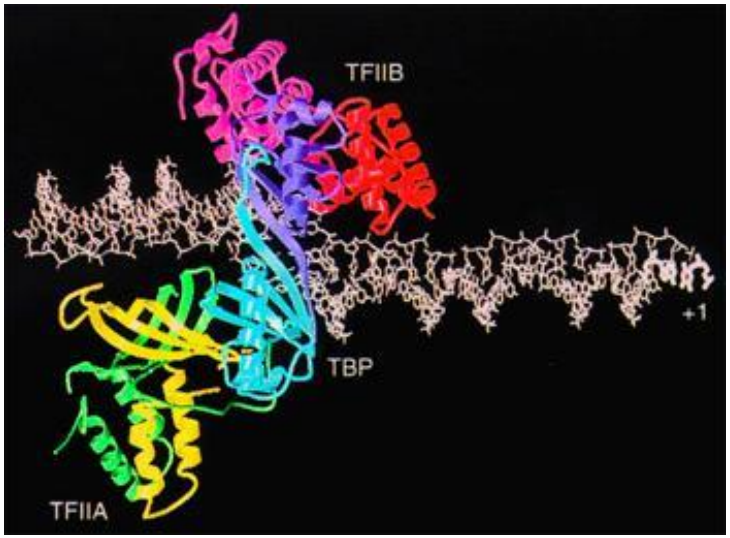
TABLE 1 **Components of the human general transcription machinery**

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.

Structural models available in the Protein Data Bank
https://www.rcsb.org/

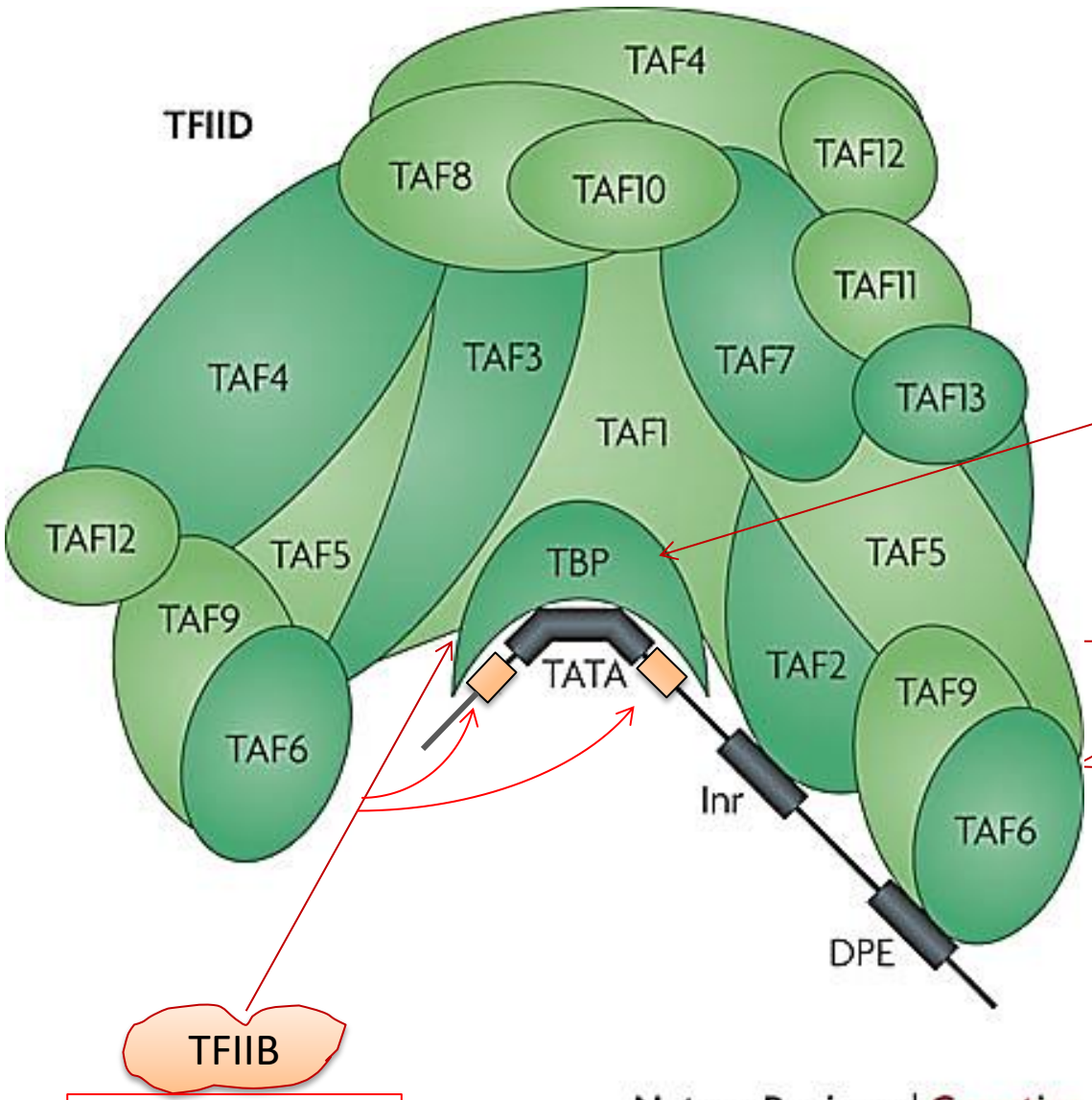




Primary role of TFIID in promoter recognition

TFIID=TBP+TAFs

TAF=TBP-Associated Factors



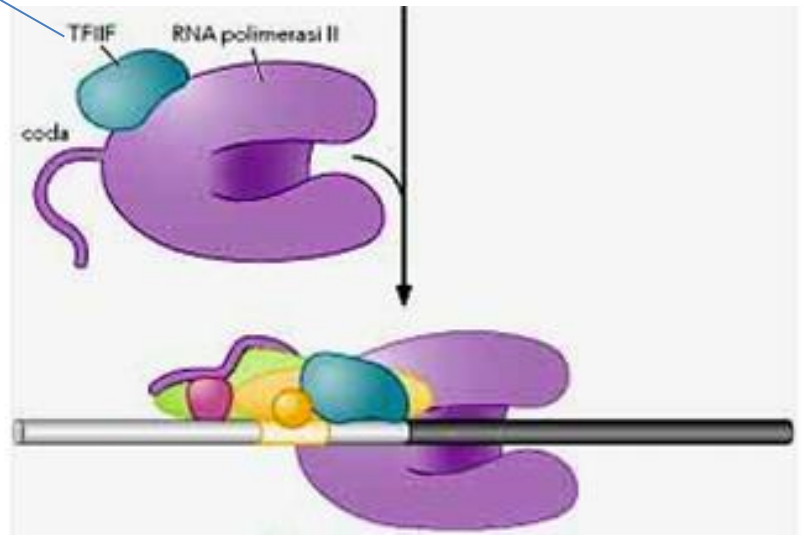
TBP (TATA-binding protein) binds to the TATA element (if present) and bends DNA ~ 90°

TAF2+TAF6 recognize Inr and DPE (if present)

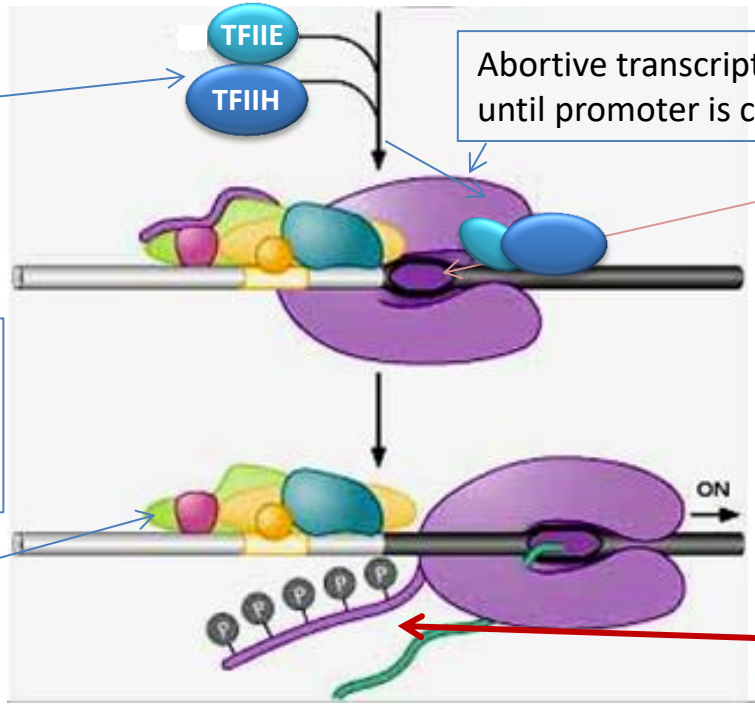
TFIIB-binds to TBP and to BRE

Nature Reviews | Genetics

$TF_{II}F$ allows RNA PolII joining the complex



$TF_{II}E$ + $TF_{II}H$ needed for promoter clearance



Abortive transcription until promoter is cleared

«open» complex

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

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

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

Different steps are mainly regulated by phosphorylation of the RNA PolII CTD (Carboxy-terminal domain)

The 52 repeats in Human Pol II

The very C-term peptide

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

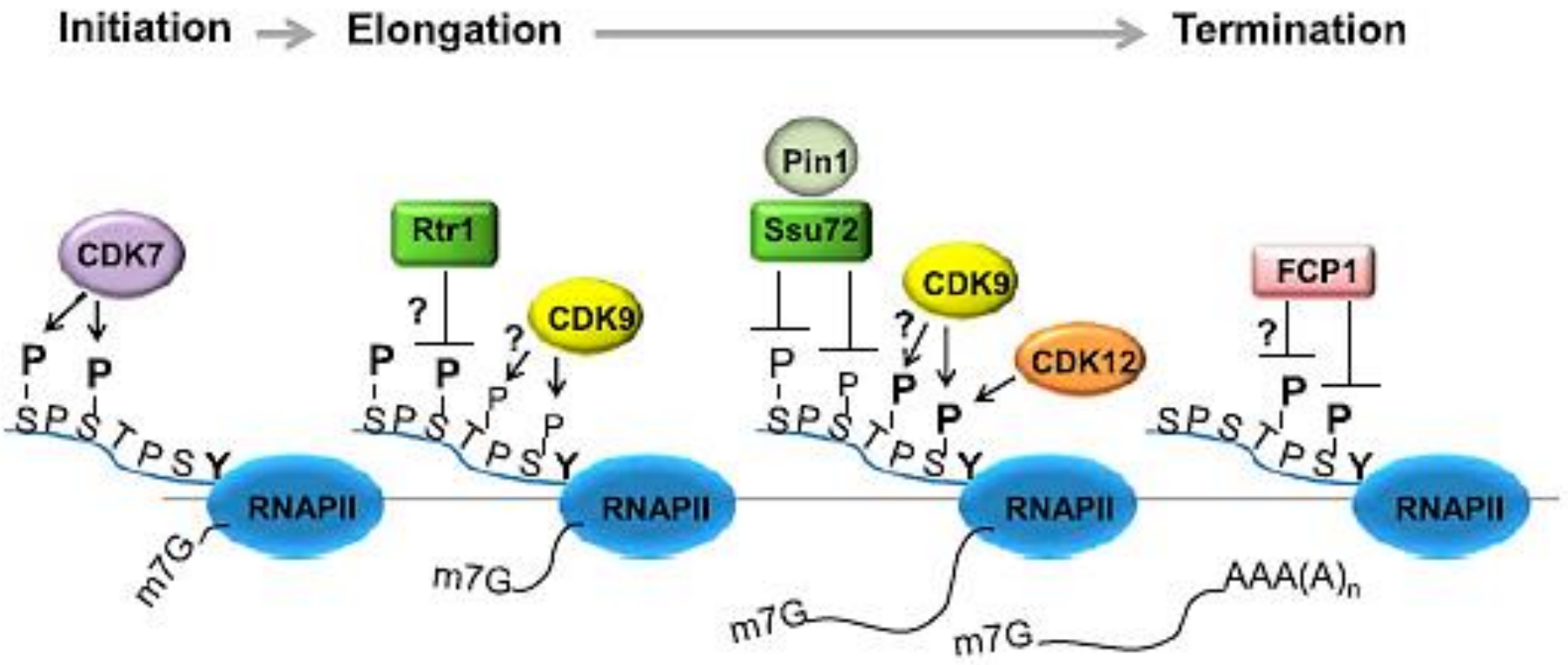
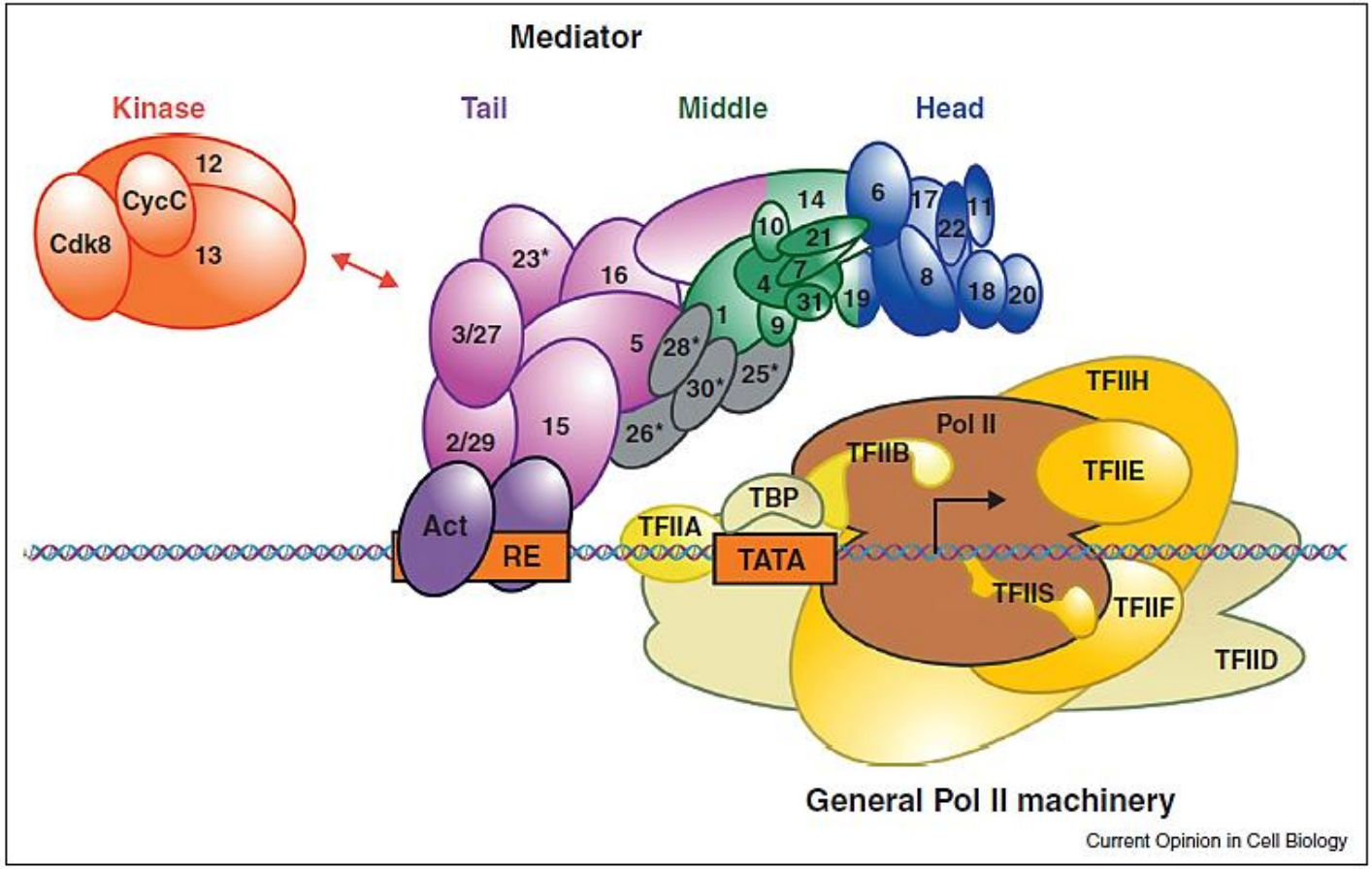


Figure 2. Dynamic modification of the CTD during the transcription cycle. At transcription initiation, CDK7 phosphorylates Ser5 and Ser7 residues. Later, during elongation, CDK9 phosphorylates Ser2 and perhaps Thr4, while the phosphate groups on Ser5 and Ser7 are gradually removed by phosphatases. (…)

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



From:
Lariviere et al.
Curr Op Cell Biol.
2012, 24:305–313

RNA Polymerase II pre-initiation complex comprising Mediator and the general Pol II machinery at the promoter. Mediator bridges between gene specific activators (Act) bound to regulatory DNA elements (RE) and the general transcription machinery comprising Pol II and the general transcription factors TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIF, and TFIIH, and the factor TFIIS. The transcription start site is indicated with a black arrow. The Mediator modules head, middle, tail, and kinase are colored blue, green, purple, and orange, respectively. Mediator subunits Med14 and Med19 are probably bridging between modules and are therefore shown in two colors. Subunits that are not assigned to any module are colored gray. Yeast Mediator subunits Med2 and Med3 are identical to human Mediator subunits Med29 and Med27, respectively. Subunits present only in higher eukaryotes are marked with an asterisk.

Promoter functional validation:

- Reporter vectors, using progressive deletions & mutations
- In vitro transcription, using promoter-template constructs with progressive deletions & mutations

Go to any Molecular Biology book for more

also the Watson e-book in the auxiliary in Moodle.

First essential issue:
the identification of Promoters genome-wide

Promoter identification

To identify promoters (genome-wide) we need essentially to have clear definition of TSS. This is not easy since:

- classical methods (i.e. cDNA sysnthesis using RT enzyme) rarely reach the very 5'-end due to secondary RNA structures.
- normal RNA-seq which is based on random fragmentation shows very poor enrichment of terminal fragments.

People have used different methods to map promoters genome-wide:

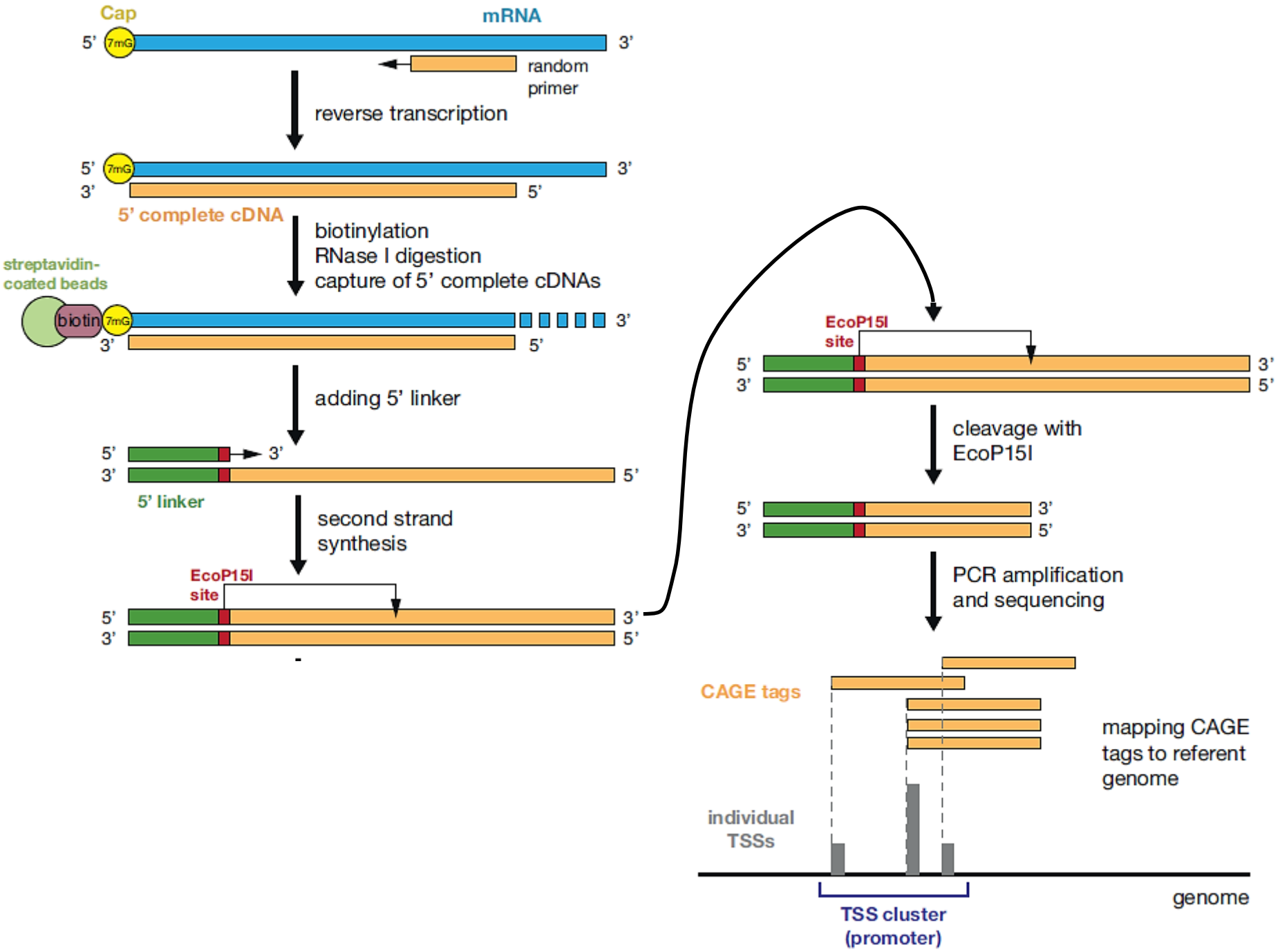
- CAGE (with both classical sequencing and later using NGS)
- 5'-SAGE (same)
- mapping PIC-component by ChIP-Seq
- mapping Histone PTMs/variants by ChIP-Seq
- Bioinformatics (prediction of basal promoter elements)

SAGE (Serial Analysis of Gene Expression) and CAGE (Capped RNA Analysis) were originally based on making short fragments from either 3'- or 5'-end of mRNAs, concatamerize and sequence by Sanger.

Most important:

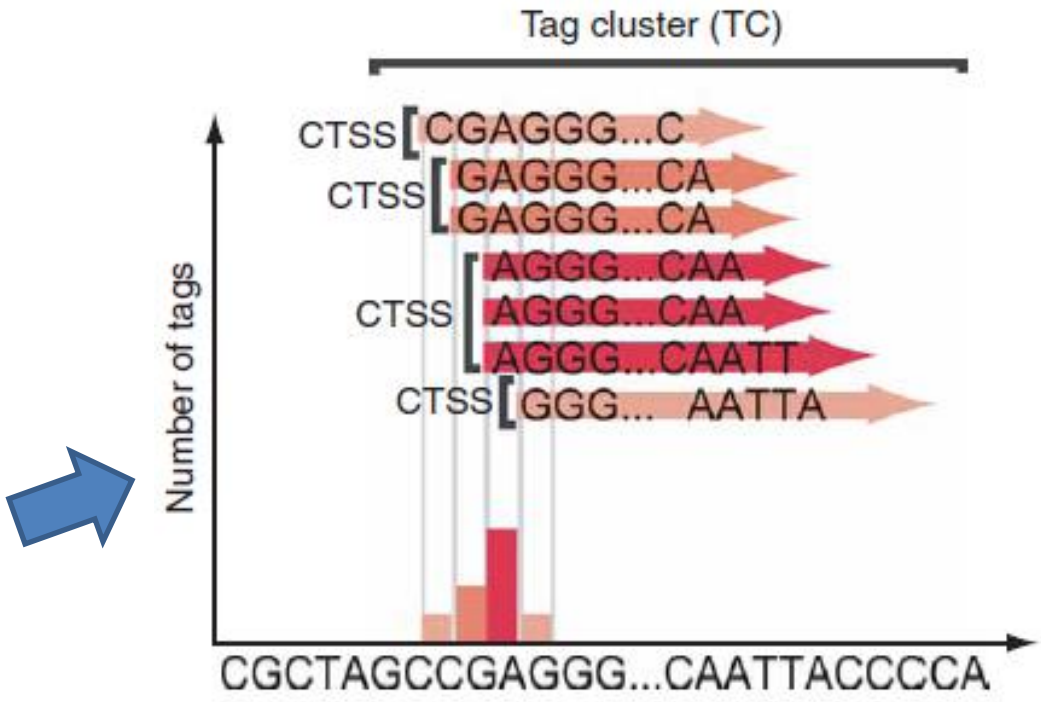
CAGE was developped specifically for 5'-end definition and is based on chemical modification of the the RNA Cap, allowing enrichment of correctly extended cDNAs.

Today, both methods were adapted to using NGS sequencing of the fragments obtained (e.g. CAGE-seq).


Cap
mRNA
5'
7mG
3'
random primer
reverse transcription
5' complete cDNA
biotinylation
RNase I digestion
capture of 5' complete cDNAs
streptavidin-coated beads
biotin
adding 5' linker
5' linker
second strand synthesis
EcoP15I site
cleavage with EcoP15I
PCR amplification and sequencing
CAGE tags
mapping CAGE tags to referent genome
individual TSSs
genome
TSS cluster (promoter)

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).

The FANTOM project used these methods to study a nuber of cell lines and tissues from Mouse and Human origin

These studies identified unprecedented numbers of TSSs therefore allowing intensive re-examination of Promoters features

It was clearly seen that, depending on the shape and dispersion of TSSs, Promoters could be grouped in (at least) four different groups, as exemplified in the following figures:

1. Single peak class (SP) (a single nucleotide or with few alternatives around it)
2. Broad TSS (various nucleotides in a range)
3. Bi- or multi-modal (some dominant peaks within broader initiation sites)
4. Broad with dominant peak (much like mixing type 1 and 2)

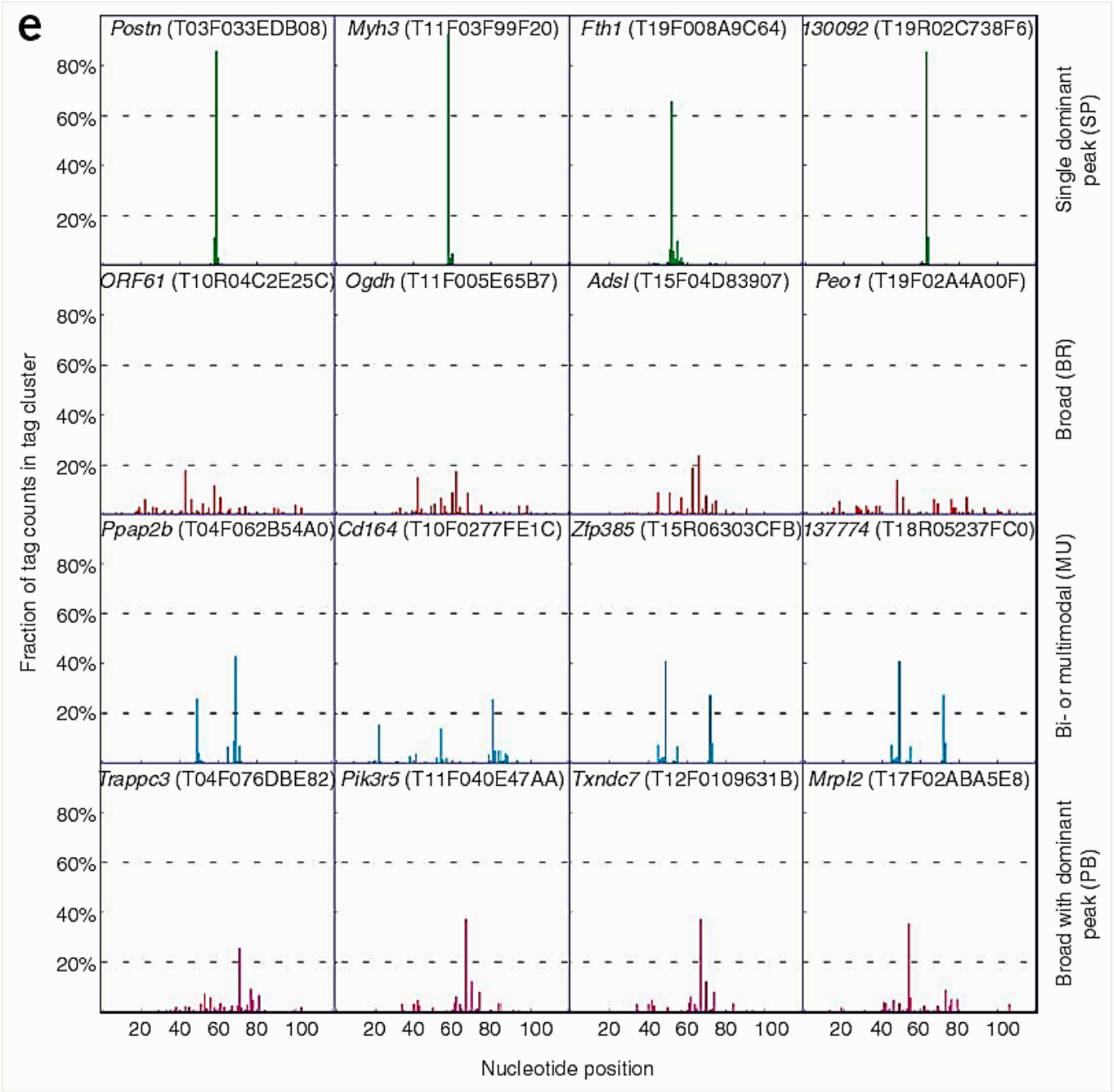
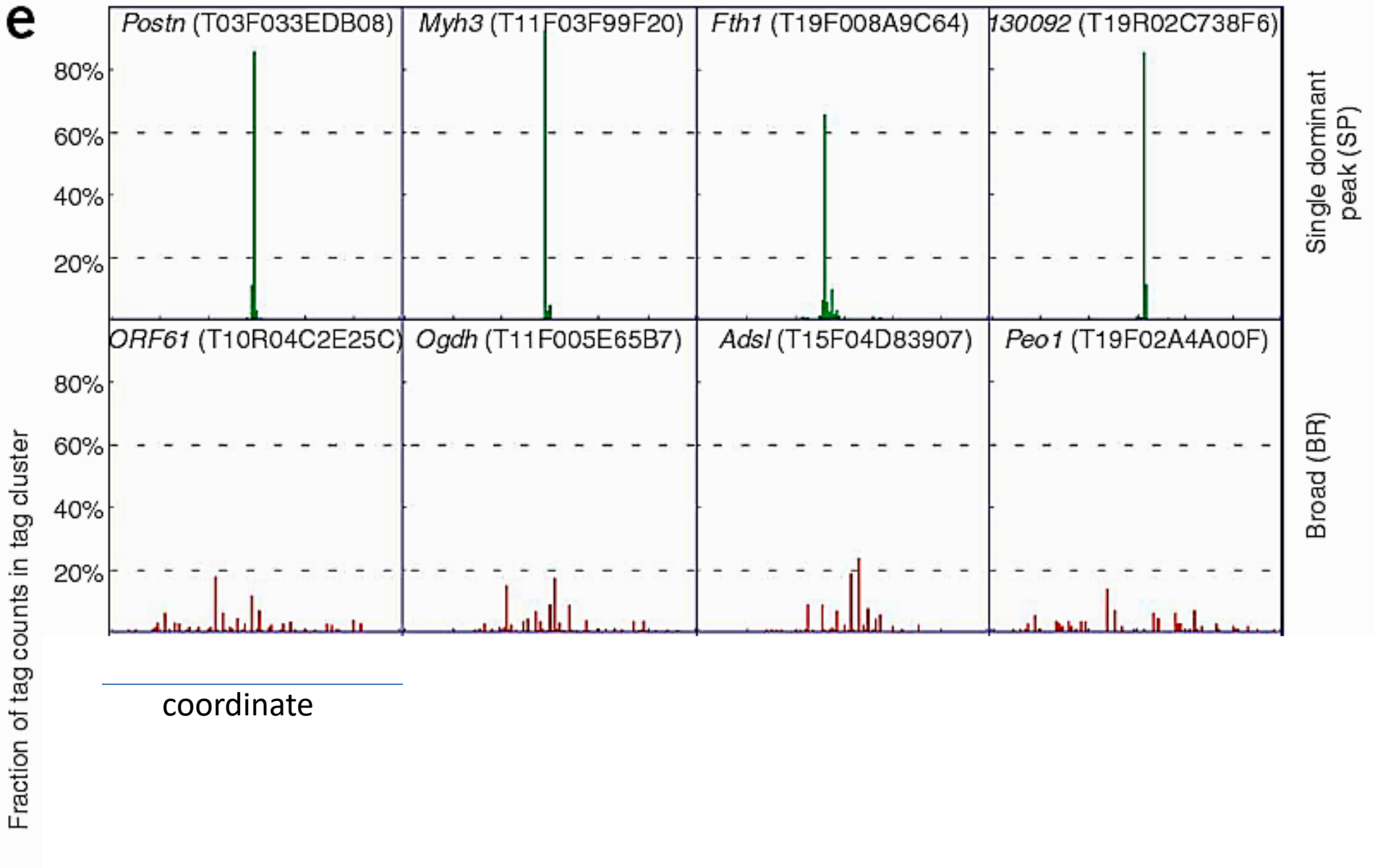


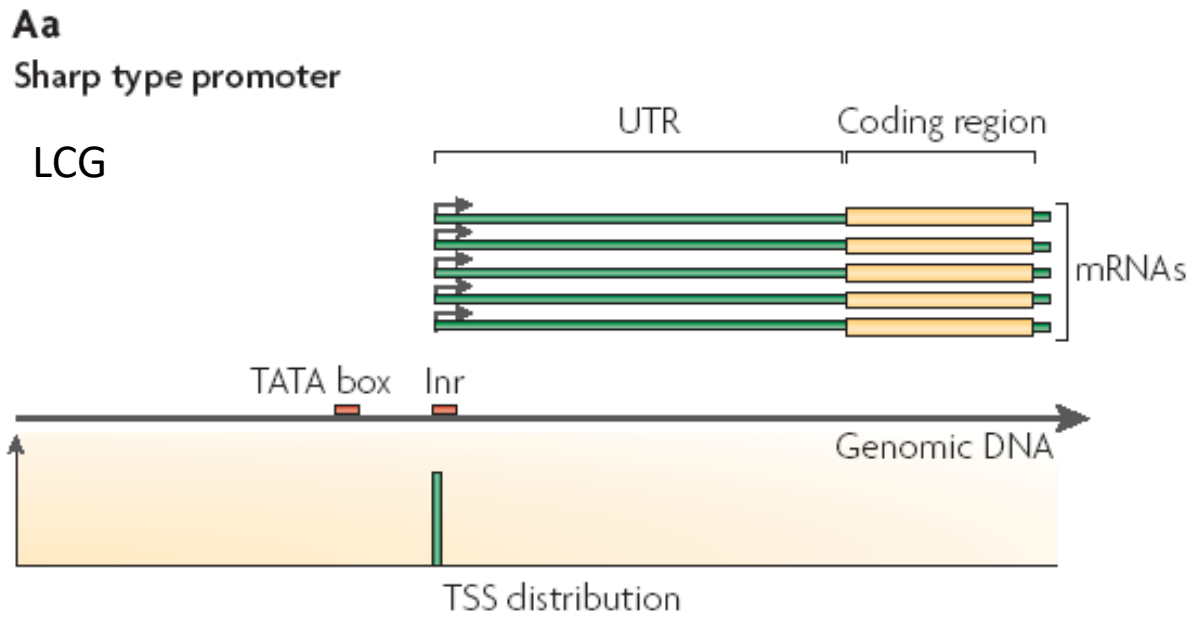
Figure 1. (e) Arrays of representative tag clusters for different shape classes. Histograms indicate the fraction of tags in the tag cluster mapping into each position in a 120-bp window centered on the tag cluster. The single peak (SP) class is characterized by a sharp peak, indicative of a single, 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


e
Postn (T03F033EDB08)
Myh3 (T11F03F99F20)
Fth1 (T19F008A9C64)
130092 (T19R02C738F6)
ORF61 (T10R04C2E25C)
Ogdh (T11F005E65B7)
Adsl (T15F04D83907)
Peo1 (T19F02A4A00F)
80%
60%
40%
20%
Single dominant peak (SP)
Broad (BR)
Fraction of tag counts in tag cluster
coordinate

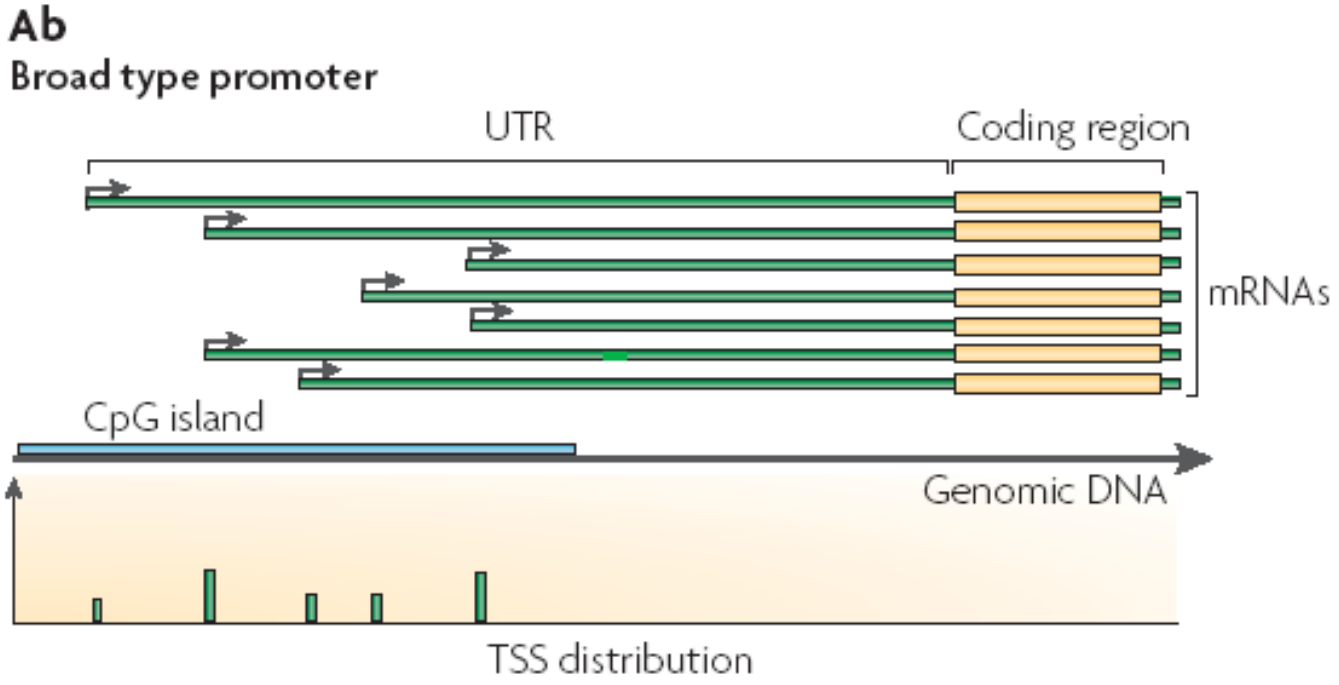
For common usage, we classify today in only two classes,
as
«sharp-type promoters»
and
«broad-type promoters»

intended that many situations in between exist.



The two extreme models

HCG



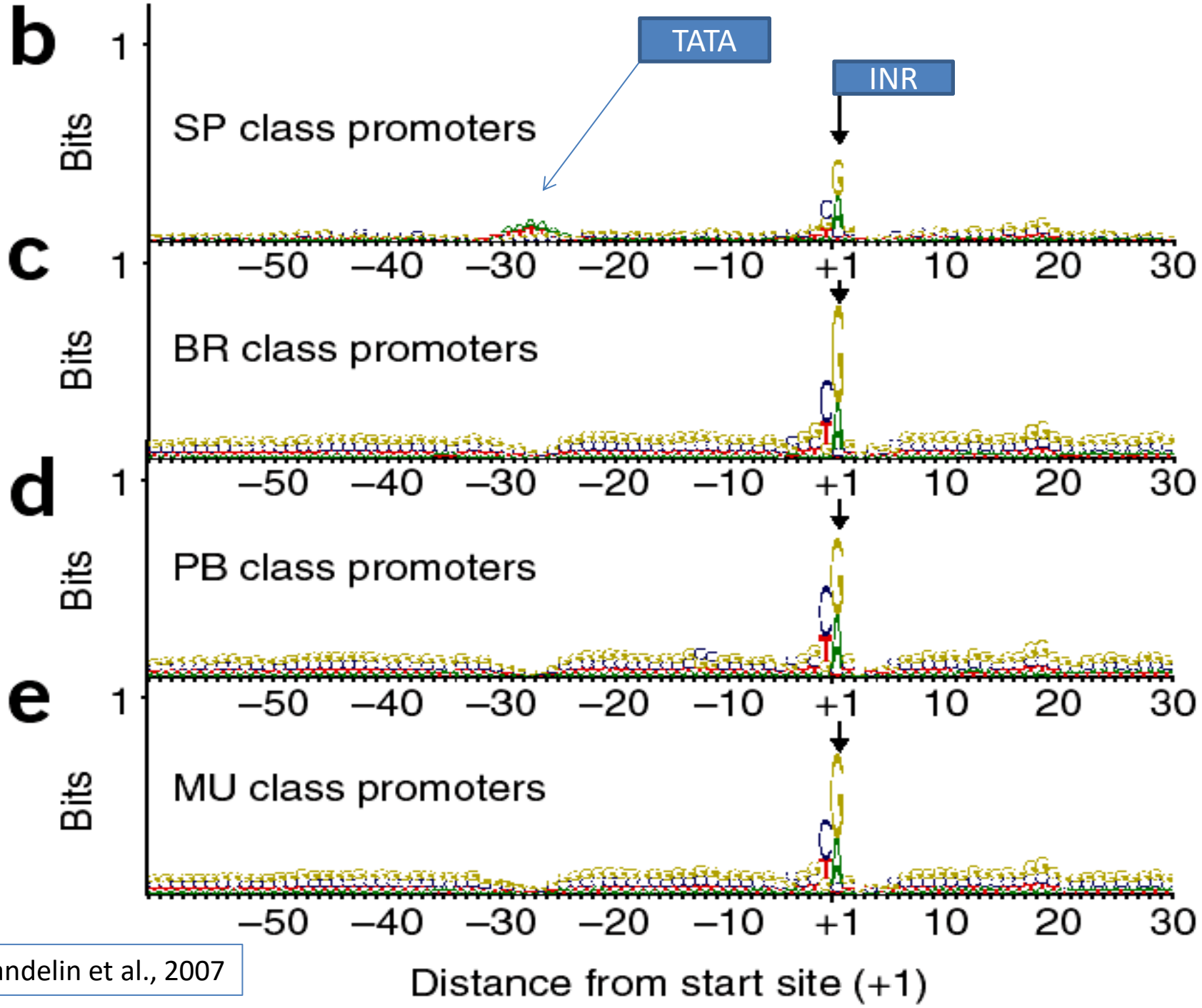
Sandelin et al., 2007

Genomic studies have partially changed our knowledge of promoters.

These studies demonstrated, first, that the “textbook promoter” with a clear TATA-box and a single TSS, is present at no more that 10-20% of mammalian genes (<15% in human and mouse), which represent a group of tissue-specific genes.

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

Alignment of thousands Promoters has allowed appreciation of strict geometrical constraint in the position of Promoter elements, like TATA-box, Initiator (INR), downstream promoter element (DPE).

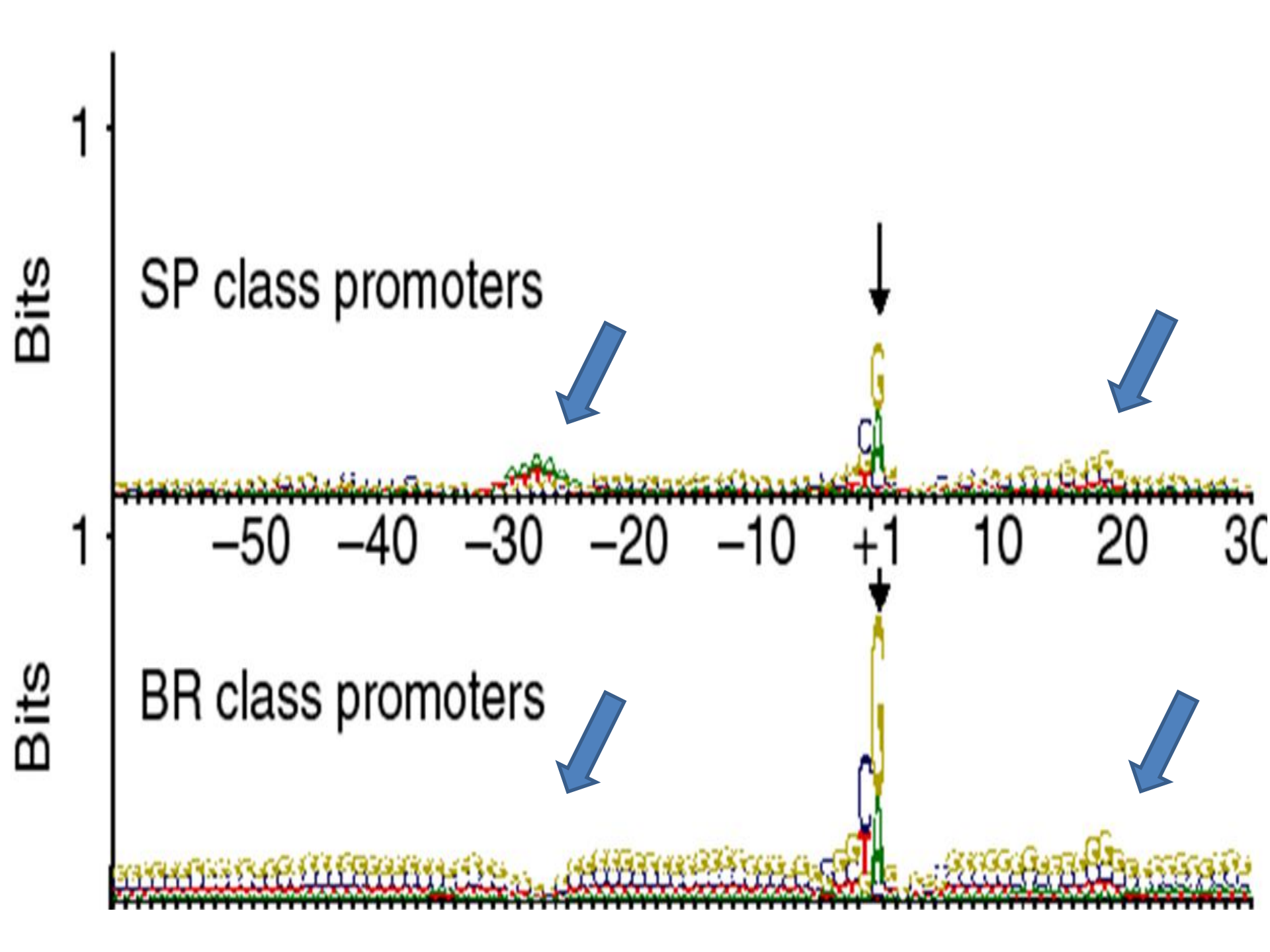


Sandelin et al., 2007

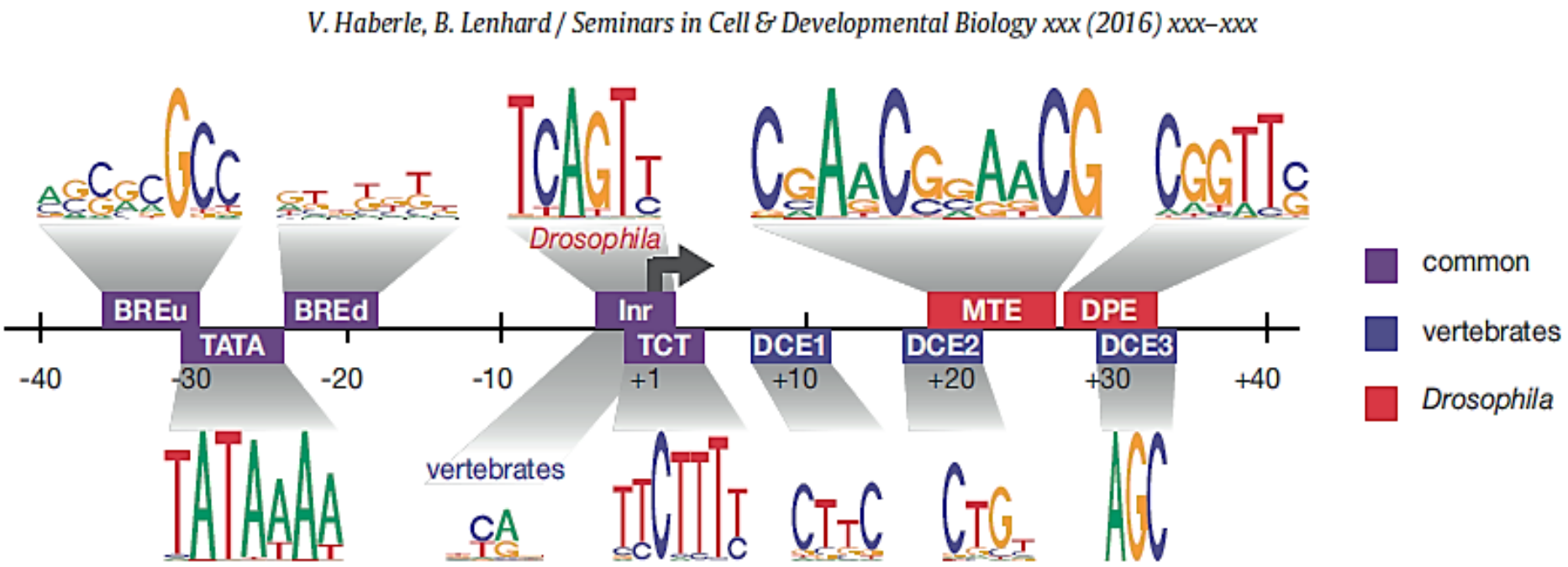
← Figure legend

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.


Bits
1
SP class promoters
1
−50
−40
−30
−20
−10
+1
10
20
30
Bits
BR class promoters

Basic promoter elements in insects and Vertebrates are positionally constrained



Associated sequence logos are based on motifs from described for Drosophila and motifs from the JASPAR database for Vertebrates.
The initiator motif (Inr). 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.

These elements are assorted in various combinations in vertebrate promoters, with some rules (e.g. INR associates with either a TATA or a DPE, very rarely all together)

BRE (TFIIB response element) was identified essentially by cristallography, very weak consensus

Different promoters exist that are recognized by sets of different proteins (e.g. TFR instead of TBP)(see Levine)

Some promoters have «mixed sequences» that are recognized by different sets of proteins (on a developmental or tissue-specific basis)(see Levine).

CpG islands and Cytosine Methylation at Promoters

Broad type promoters are usually CG-rich and lack a clear TATA-box

I.e. they contains CpG-islands

How is CpG methylation linked to function of these promoters ?

CpG-island distribution

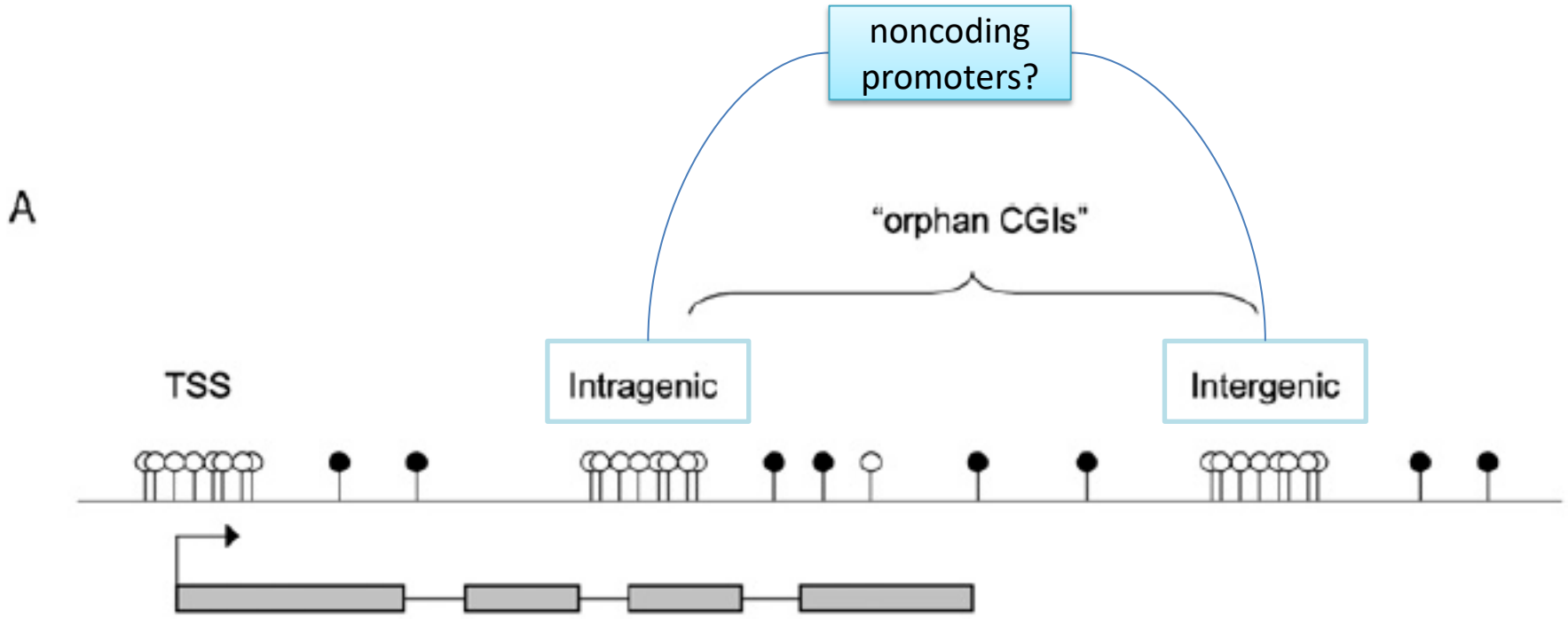


Figure 1. The genomic distribution of CGIs (CpG islands).
(A) CGIs can be located at annotated TSSs, within gene bodies (Intragenic), or between annotated genes (Intergenic).
Intragenic and intergenic CGIs of unknown function are classed as ‘‘orphan’’ CGIs.
(Empty circles) Unmethylated CpG residues. (Filled circles) Methylated CpG residues.

Deaton & Bird, 2011. Genes Dev 25:1010–1022

CpG methylation has been studied genome wide by several laboratories. The first exhaustive study was published by Weber et al. in 2007, using Methyl-DNA immunoprecipitation and analysis on tiling promoter arrays.

Later other groups have used bisulfite-NGS sequencing, and results were confirmed and extended.

Weber et al. first divided the Promoters in three classes, on the basis of the number of CpG in the sequence (CpG/base pair): LCP=low; ICP=intermediate; and HCP=High.

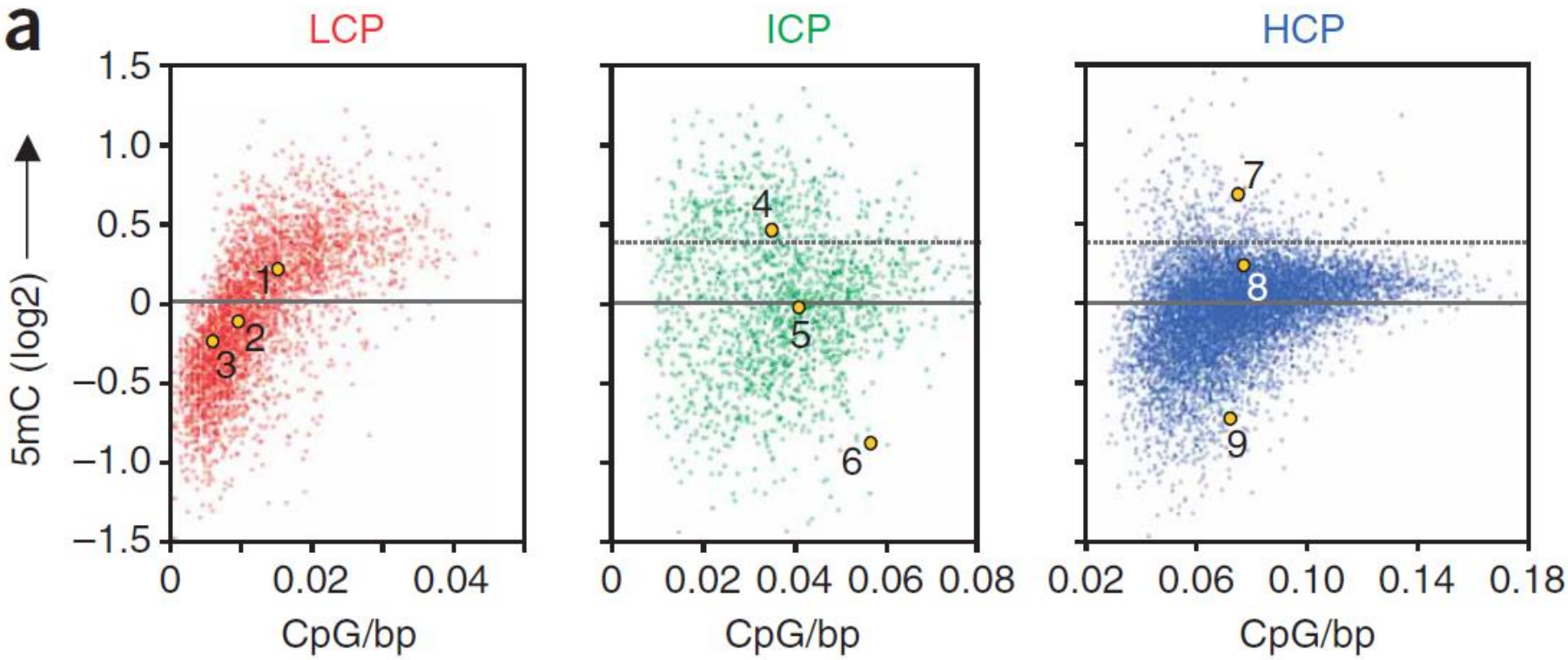
Next, they measured the level of CpG methylation and expressed it in comparison to the density of CpG in each promoter (a window of -700bp to +200bp respect to the TSS).

Finally, they correlated the %-methylation with RNA PolII and H3K4me3 occupancy (by ChIP) as detectors of the activity status.

Enrichment by MeDIP, as an indirect measure of the amount of Cytosine methylation, *versus* the density of CpG in each promoter, in each class.

In LCP, there is a almost linear correlation, i.e. more CpG you have, more meC you get. In HCP, is seems like a «saturation curve» i.e. inreasing most are unmethylated independently on the CpG density.

Numbers indicate single promoters that were studied with bisulfite (next)



Weber et al, (2007) Nat Genet 39:457.

Conclusions:

1) different kind of promoters are also characterized by different CpG content and methylation effect

2) High CpG promoters are mostly « housekeeper» gene promoters.

3) HCP with no RNA PolII ChIP signal still have «active chromatin» mark H3K4me3, i.e. they are «poised» to transcription

Low CpG (LCP) promoters are mainly tissue-specific inducible promoters and are unlinked to methylation degree

Intermediate CpG promoters (ICP) represent the class where methylation most closely reflects RNA PolII occupancy, i.e. their transcription appears methylation-dependent

By associating these studies to CAGE-defined promoters, we can also conclude that

HCP promoters, house-keepers, are Broad-type

while

LCP promoters, tissue-specific and inducible, are sharp-type and feature prevalent TATA-box initiation

Nucleosome positioning at promoters

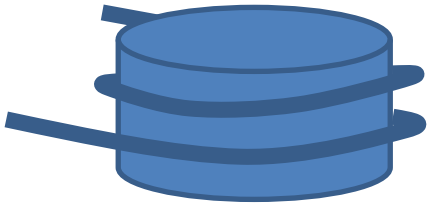
MNase-seq is a method where chromatin is digested using Micrococcal Nuclease (mixed exo- endo-), them the remaining «nucleosomal» DNA is sent to library preparation and NGS sequencing.

While in most of the genome nucleosomes are not «positioned» (i.e. they are random positioned), in functional positions like Enhancers and Promoters we see a very strong positioning.

By parallel ChIP-seq experiments, it was observed that these nucleosomes also feature histone variants (H3.3 and H2A.z) and the typical PTMs.

MNase-Seq workflow

- Treat Nuclei with MNaseI
- Nucleosome-protected DNA is isolated by deproteinization
- DNA fragements ~150 bp purified
- Ligated to adaptors → Library
- Sequenced by NGS technology



MNase trimmed nucleosome

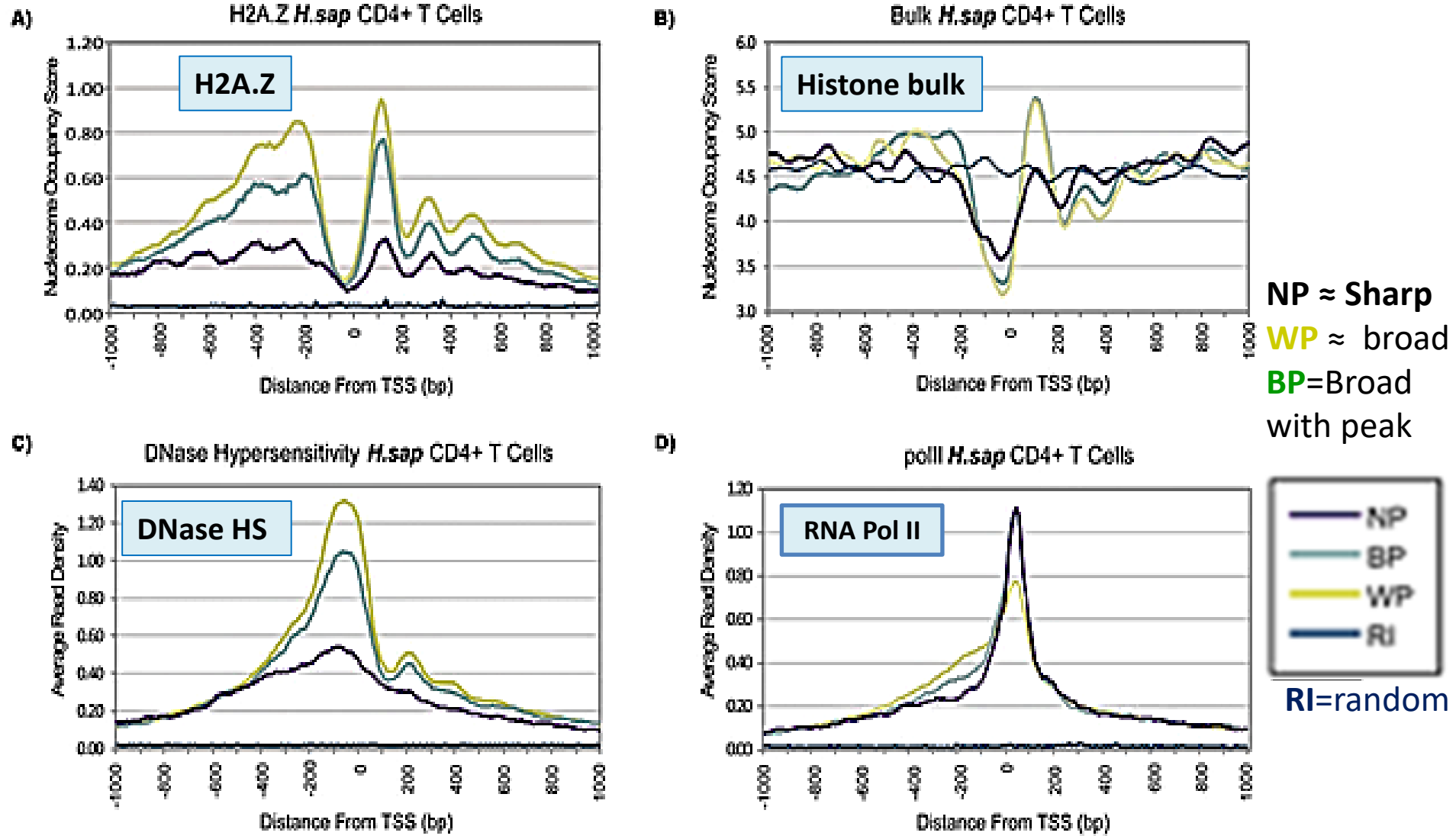
deproteinize

~150 bp

Adapter, library prep. → NGS

After mapping to reference genome, highest peaks signal most «positioned» nucleosomes

Figure 1. Promoter Classes Reflect Distinct Profiles of Nucleosome Organization. Rach 2011 Plos Genet e1001274



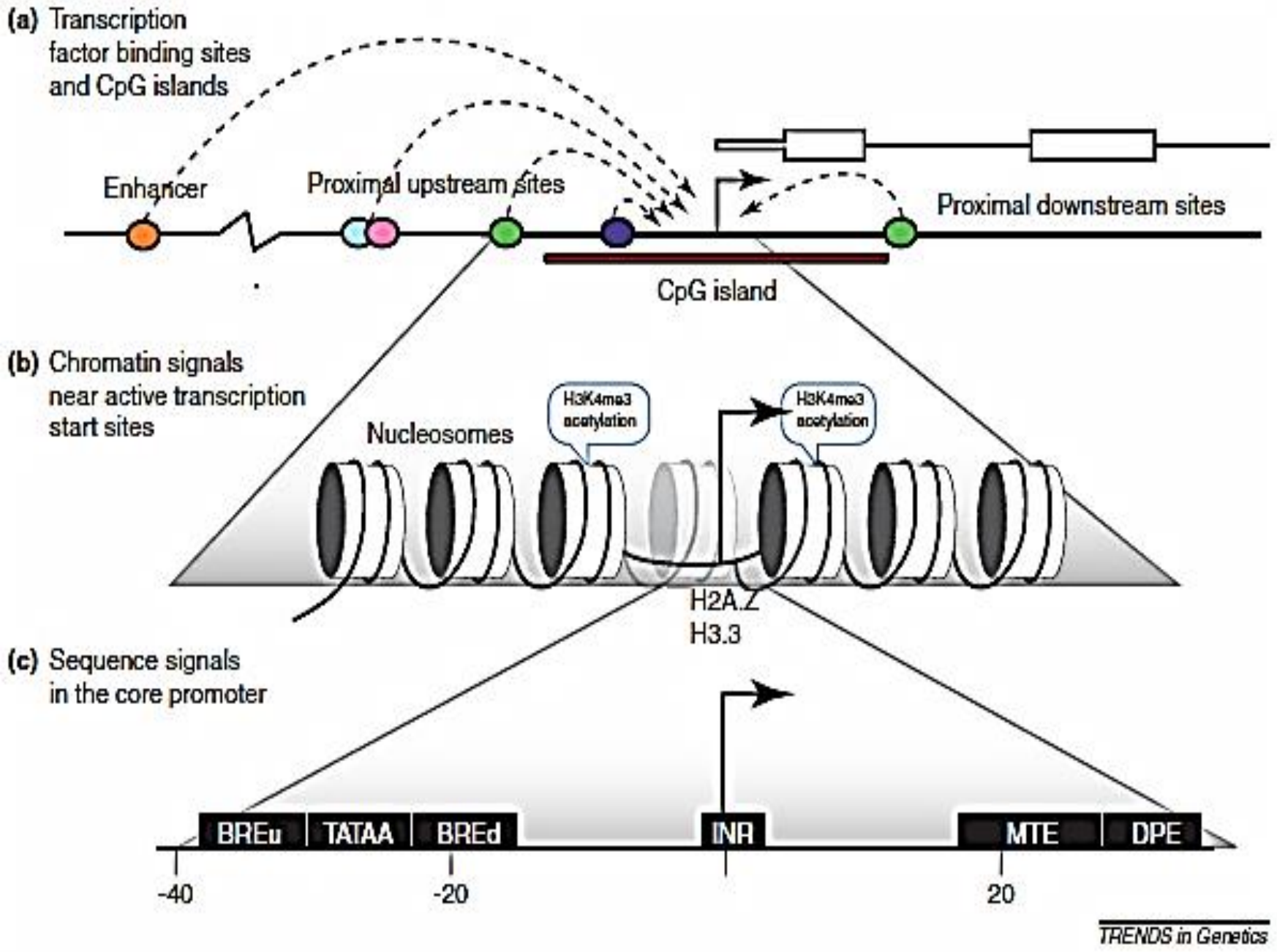
NP ≈ Sharp

WP ≈ broad

BP=Broad with peak

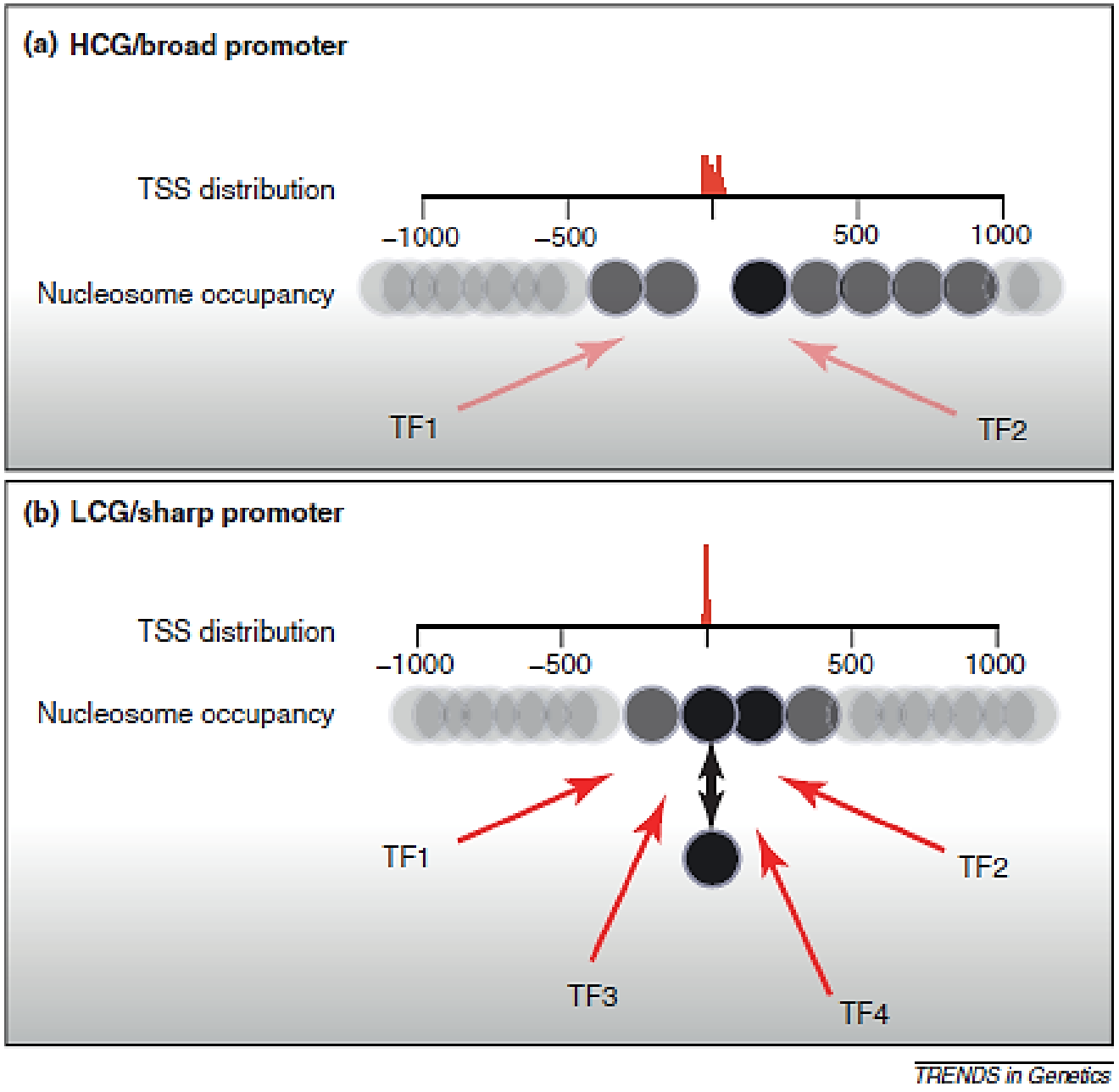
RI=random

Authors classified Promoters as Narrow Peak (NP), Broad with Peak (BP), and Weak Peak (WP). RI refers to average levels at random intergenic sites, which is used as a baseline. (A) Increased H2A.Z levels ($p<10E-36$), (B) increased bulk levels. DNase hypersensitive sites revealed a more accessible nucleosome-free region at BP and WP but not at NP promoters (C), yet pol II levels were higher at NP promoters (D).


(a) Transcription factor binding sites and CpG islands
Enhancer
Proximal upstream sites
Proximal downstream sites
CpG island
(b) Chromatin signals near active transcription start sites
Nucleosomes
H3K4me3 acetylation
H3K4me3 acetylation
H2A.Z
H3.3
(c) Sequence signals in the core promoter
BREu
TATAA
BREd
INR
MTE
DPE
-40
-20
20
TRENDS in Genetics

Mostly constitutive

Mostly regulated



From Valen & Sandelin 20011, Trends in Genetics 27:475

HCG promoters

- «Broad» TSS – several TSS spread over 30-100bp
- CpG-island, undermethylated
- Functions: housekeepers, i.e. Ubiquitous expression
- Stable expression levels
- Bound mostly by ubiquitous Transcription Factors (TFBS over-represented close to TSS, e.g. Sp1)
- NFR (NDR) region more evident at CGI (nucleosome is unstable: H2A.Z and H3.3 present)
- First nucleosome downstream TSS strongly positioned.
- Nucleosomes flanking NFR enriched for H2A.Z + H3.3 (also if «poised»)
- Most expression-predictive PTMs: H3K27ac and H4K20me1
- CGIs «intrinsically» promoters (CpG recognized by Cfp1/Set1 H3K4 methyltransferase)

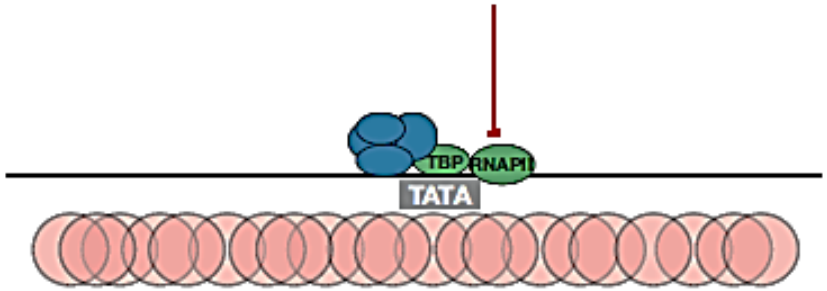
LCG promoters

- «Sharp» TSS – One single TSS at 1-3 adjacent nucleotides
- Clear TATA sequence present (30%)
- Function: Tissue-specific, inducible
- Large variability in expression level
- Bound mainly by tissue-specific and inducible Transcription Factors
- «covered», i.e. Nucleosome-occupied (nucleosome is stable), NFR less evident
- Nucleosome positioning and PTMs average less evident since part active and part inactive
- Most expression-predictive marks: H3K4me3 and H3K79me1
- «Intrinsically» repressed, require TF and chromatin remodelers to be freed and activated

Promoter types

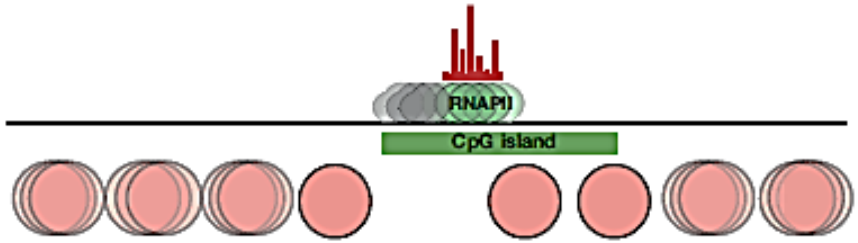
Adult tissue-specific

1) sharp transcription initiation pattern
2) context-specific regulatory input close to TSS
3) TATA-box ~30 bp upstream of TSS
4) disordered nucleosomes



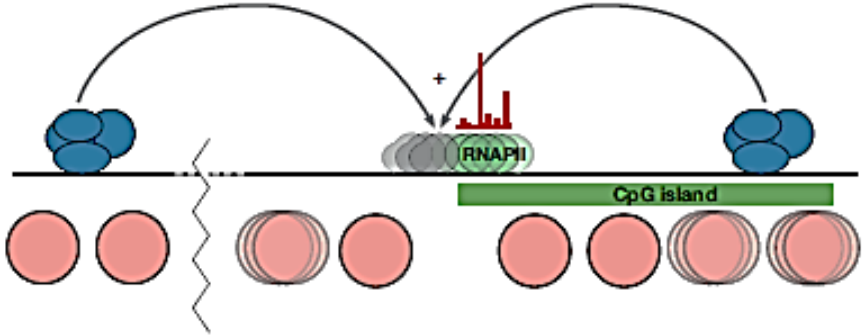
Ubiquitously expressed ("housekeeping")

1) broad transcription initiation pattern
2) no context-specific regulatory input
3) CpG island around TSS
4) nucleosome-free region with precisely positioned -1 and +1 nucleosomes



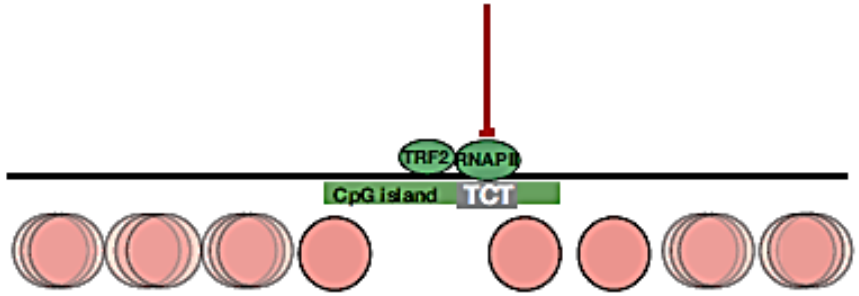
Developmentally regulated

1) broad transcription initiation pattern
2) context-specific regulatory input from distal enhancers
3) long CpG island(s) into gene body
4) nucleosome-free region with precisely positioned -1 and +1 nucleosomes



Translational machinery-specific

1) sharp transcription initiation pattern
2) no context-specific regulatory input
3) TCT initiator and CpG island
4) nucleosome-free region with precisely positioned -1 and +1 nucleosomes



Alternative promoters and overlapping codes

There is evidence that a number of genes present alternative TSS that depend on alternative promoters, which are used in a differential way.

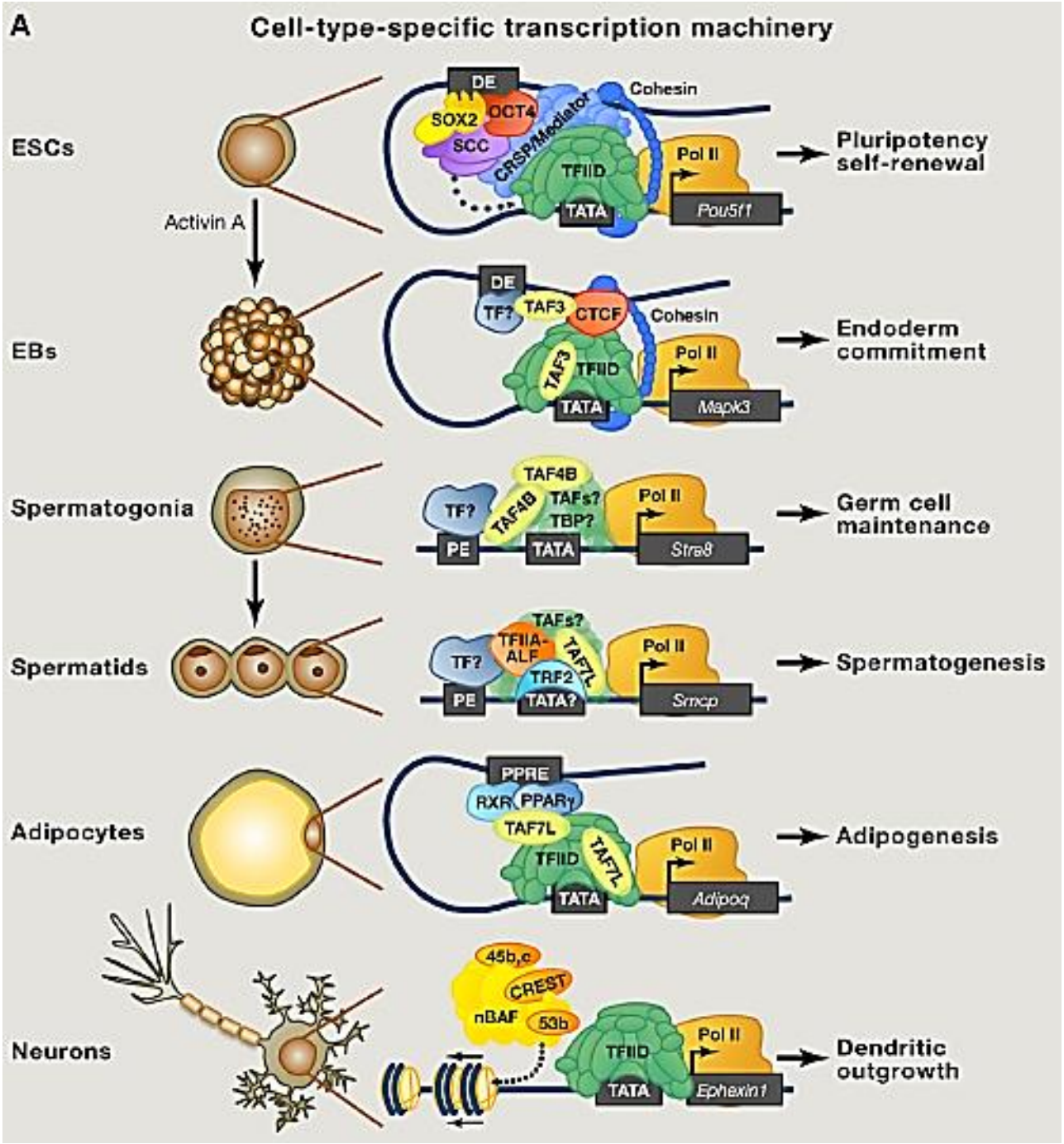
Another story concerns the fact that, in Broad-type Promoters, there is evidence that the utilization of close but different TSS may depend upon «overlapping» codes. In other words, a group of promoter elements is superimposed to a second group of elements.

Important example comes from oocyte: hosekeeper genes here are trancribed using a different machinery and a promoter «code» that is different form those used by somatic cells at the same genes. Often the two codes overlap.

In your Textbook «G» by Levine et al. 2014, a discussion is made of some unusual featurs concerning promoters and basal factors.

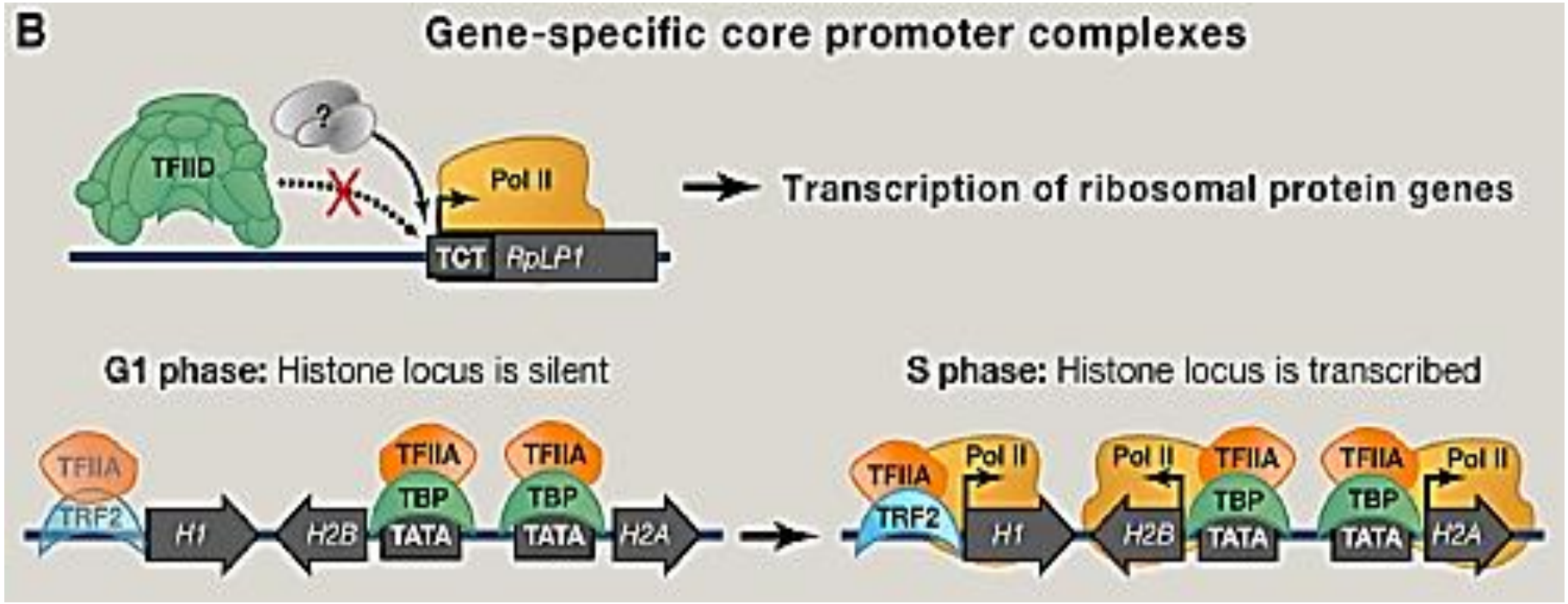
This is much like the same story: different basal transcriptional machineries exist in different tissues, which take care of specific and common genes.

This include a set of Promoters that in spermatids require the TBP-related TRF2 factor instead of the common TBP.



Levine et al. 2014

Finally, please take note that a number of special Promoters have aso been described, were the core promoter complex is different from the classical one, as in the case of Histone gene promoters.



Levine et al. 2014

Activation modes

Different categories of Promoters are regulated following different modalities. Aside from those that are mainlyregulated by Enhancer contacts, we observe two general modes of activation:

TATA+, LCG are usually covered by nucleosomes when silent. When an inductive stimulus arrives, nucleosomes are remodeled and basal apparatus can bind. We will refer to these as «on-off» genes.

HCG on the contrary are basally transcribed, show a permanent NFR, but their rate of transcription can be «modulated».

HCG promoters

- Many HCG promoter undermethylated even if not transcribed («poised»)
- High density of CG-containing TFBS (transcription factor binding sites) for constitutive TFs (Sp1, Nrf-1, E2F, ETS and other)
- «intrinsic» promoter activity due to reduced nucleosome occupancy
- Do **not** require ATP-dependent remodellers for activity

Nonetheless, the level of transcription can be «modulated».

Let’s see an example of this....

OPEN ACCESS Freely available online PLoS GENETICS

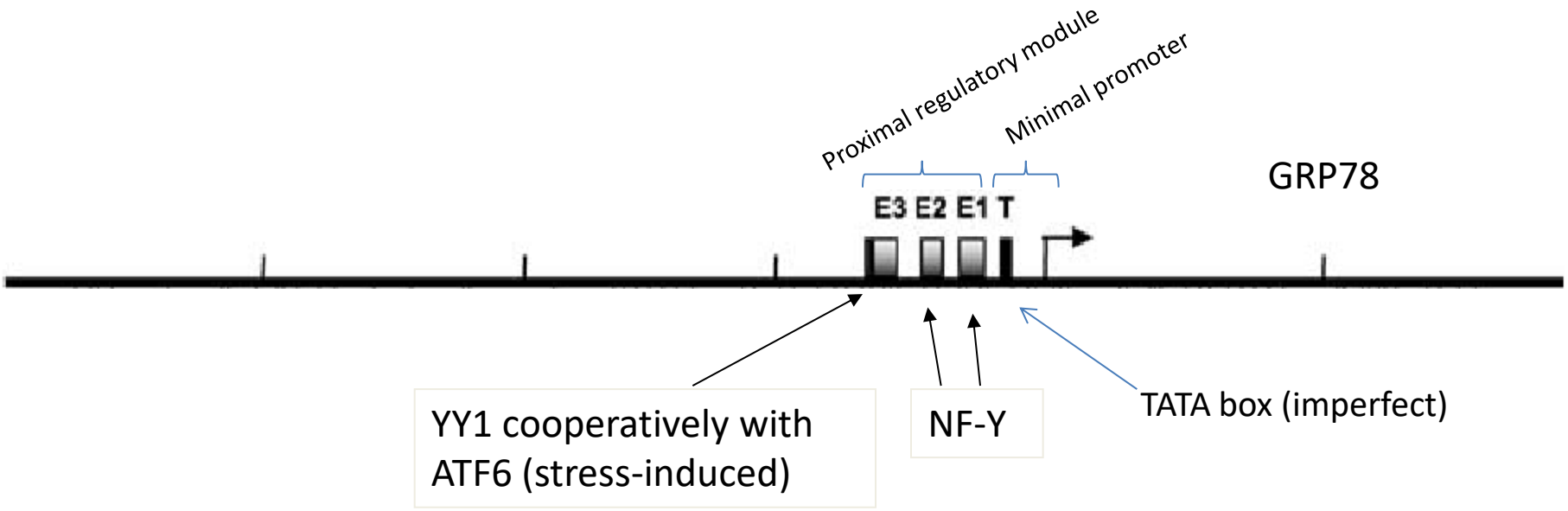
Constitutive Nucleosome Depletion and Ordered Factor Assembly at the *GRP78* Promoter Revealed by Single Molecule Footprinting

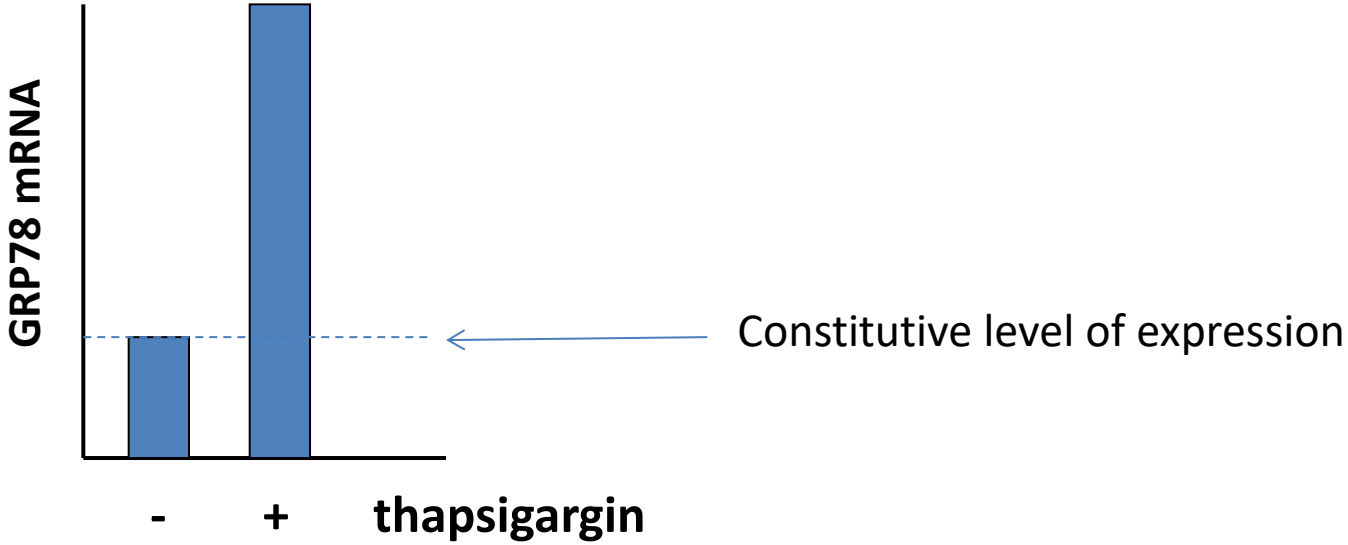
Einav Nili Gal-Yam[1,2], Shinwu Jeong[1,2], Amos Tanay[3], Gerda Egger[1,2], Amy S. Lee[2], Peter A. Jones[1,2*]

1 Department of Urology, USC/Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **2** Department of Biochemistry and Molecular Biology, USC/Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **3** Center for Studies in Physics and Biology, Rockefeller University, New York, New York, United States of America

Chromatin organization and transcriptional regulation are interrelated processes. A shortcoming of current experimental approaches to these complex events is the lack of methods that can capture the activation process on single promoters. We have recently described a method that combines methyltransferase M.SssI treatment of intact nuclei and bisulfite sequencing allowing the representation of replicas of single promoters in terms of protected and unprotected footprint modules. Here we combine this method with computational analysis to study single molecule dynamics of transcriptional activation in the stress inducible *GRP78* promoter. We show that a 350–base pair region upstream of the transcription initiation site is constitutively depleted of nucleosomes, regardless of the induction state of the promoter, providing one of the first examples for such a promoter in mammals. The 350–base pair nucleosome-free region can be dissected into modules, identifying transcription factor binding sites and their combinatorial organization during endoplasmic reticulum stress. The interaction of the transcriptional machinery with the *GRP78* core promoter is highly organized, represented by six major combinatorial states. We show that the TATA box is frequently occupied in the noninduced state, that stress induction results in sequential loading of the endoplasmic reticulum stress response elements, and that a substantial portion of these elements is no longer occupied following recruitment of factors to the transcription initiation site. Studying the positioning of nucleosomes and transcription factors at the single promoter level provides a powerful tool to gain novel insights into the transcriptional process in eukaryotes.

E = ERSE = Endoplasmic Reticulum Stress response Element





Authors studied the position of nucleosomes, PIC and Transcription Factors
at basal status
after stress induction
in the promoter + proximal regulatory module of this gene

Method: methylation protection

Ex-vivo methylation protection assay

1) Intact nuclei are treated with **M.SssI**
2) DNA extraction, bisulfite conversion of the DNA (converts C→U)
3) PCR amplification of the studied regions
4) PCR products are cloned and single clones are sequenced (ante-NGS)
5) C→T mutations provide protection patterns for single promoter molecules

Description

Back to Top

The CpG Methyltransferase, M.SssI, methylates all cytosine residues (C^5) within the double-stranded dinucleotide recognition sequence 5'...CG...3' (1).

Product Source

The CpG Methyltransferase(M.SssI) is isolated from a strain of *E. coli.* which contains the Methyltransferase gene from *Spiroplasma* sp. strain MQ1 (2,3).

This region studied in induction

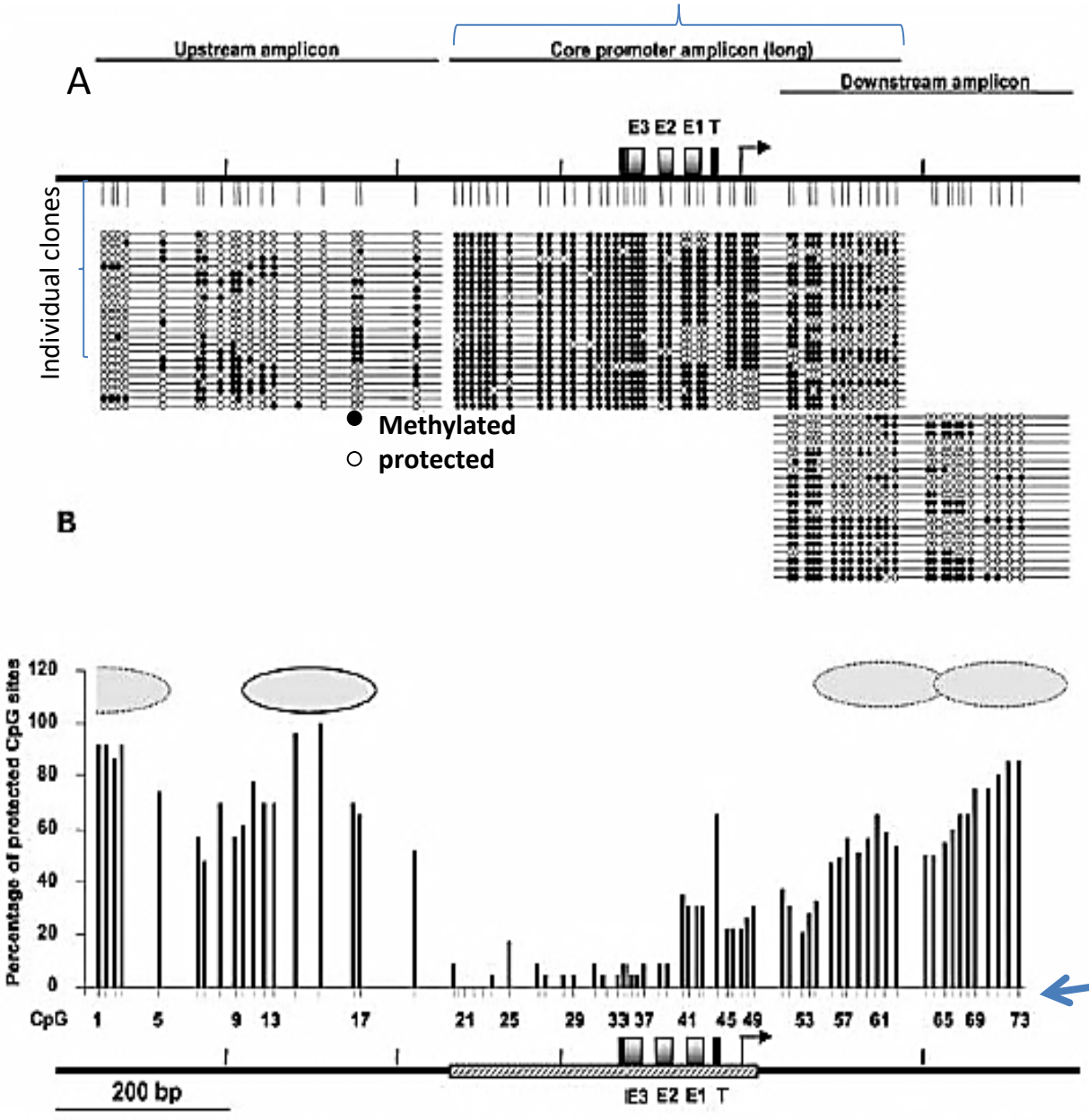


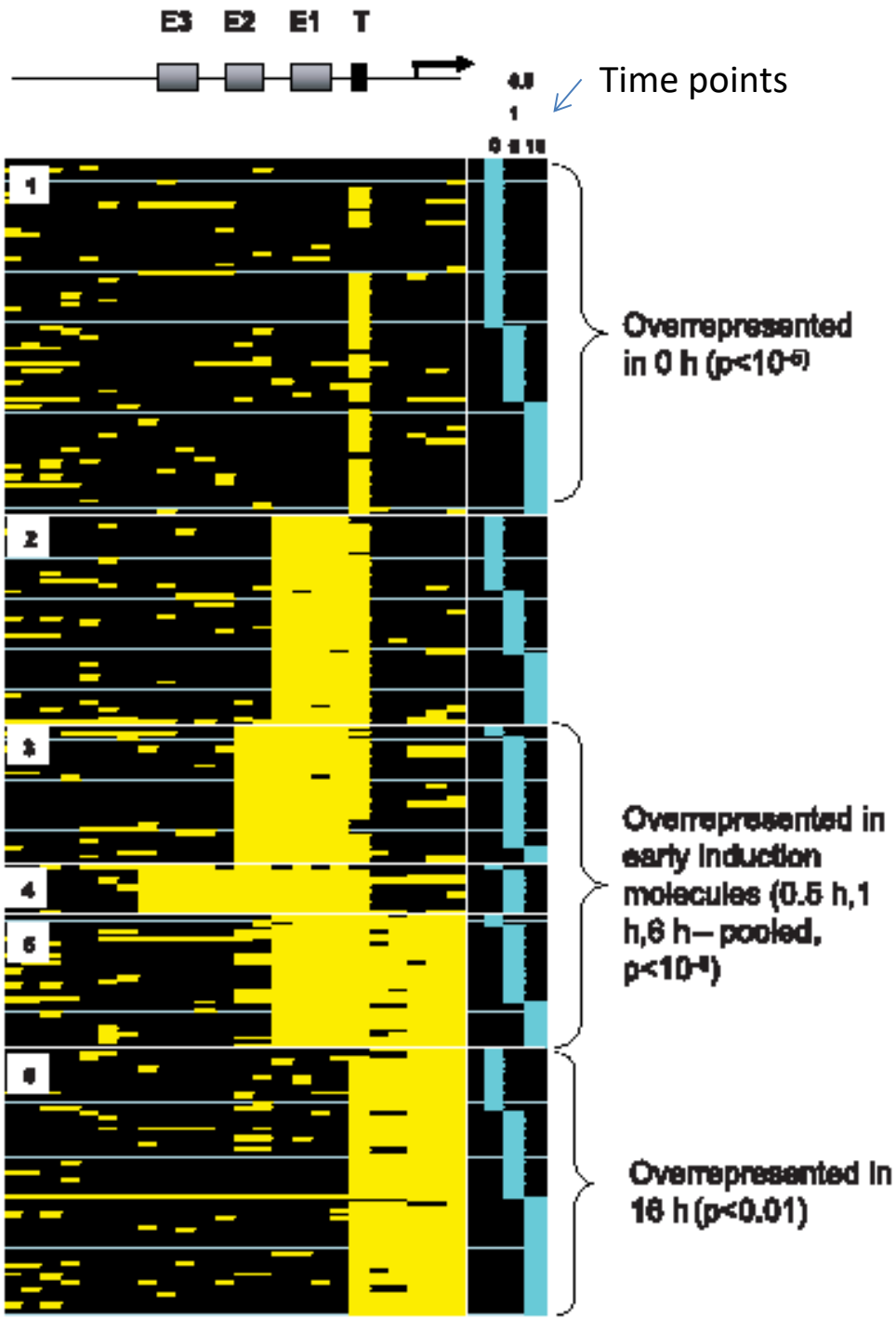
Figure 2. The Nucleosome-Free Region on the *GRP78* Core Promoter Is Minimally 350 bp Long

basal status analysis

A-Map of the promoter; E1, E2, E3 regulatory elements. T=TATA

Each bar is a single allele, showing accessible (methylated) and covered (unmethylated) CpGs.

Each small bar (I) ia a CpG position. CpG are numbered 1 to 73

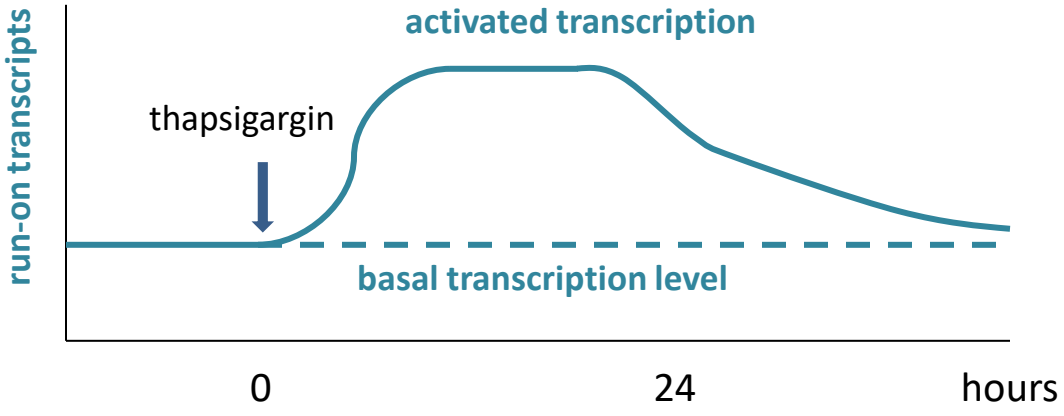


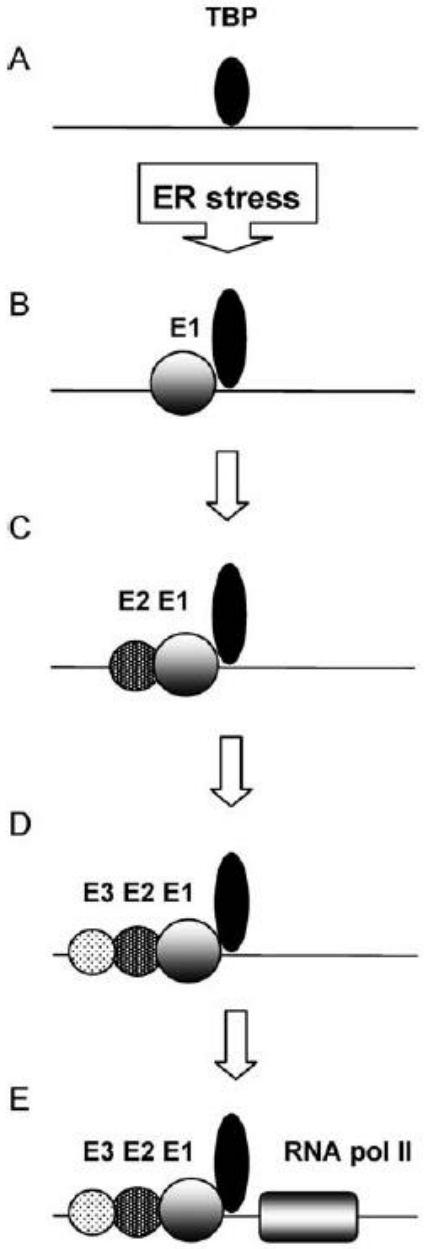
Map of protected region at different time points after Thapsigargin (Yellow dots). Each line is one allele studied.

Conclusions. The GRP78 promoter is always nucleosome-free. Short-time after induction (stimulation by E.R. stress) there is occupancy of the three E1 E2 & E3 elements, while at late time points only the region where transcription initiates is occupied.

This is the paradigm of a «always transcribed», but modulated promoter.

Concluding…
GRP78 is an Housekeeping gene that is always basally activated (A in figure) and transcribed at low level. TBP and basal factors are always bound to this promoter, and rare alleles show the presence of RNA Pol II molecules traveling on the gene. The promoter is always nucleosome-free.
After induction by thapsigargin, a progressively more extended region gets protected from methylation in vitro. These are Transcription Factors binding to regulatory elements (B, C, D). After a while, a consistent RNA Pol II footprint is seen in the transcribed region, testifying frequent initiation and increased transcription (E). This "activated status" is quite stable even when factors are no longer present and is self-sustaining.

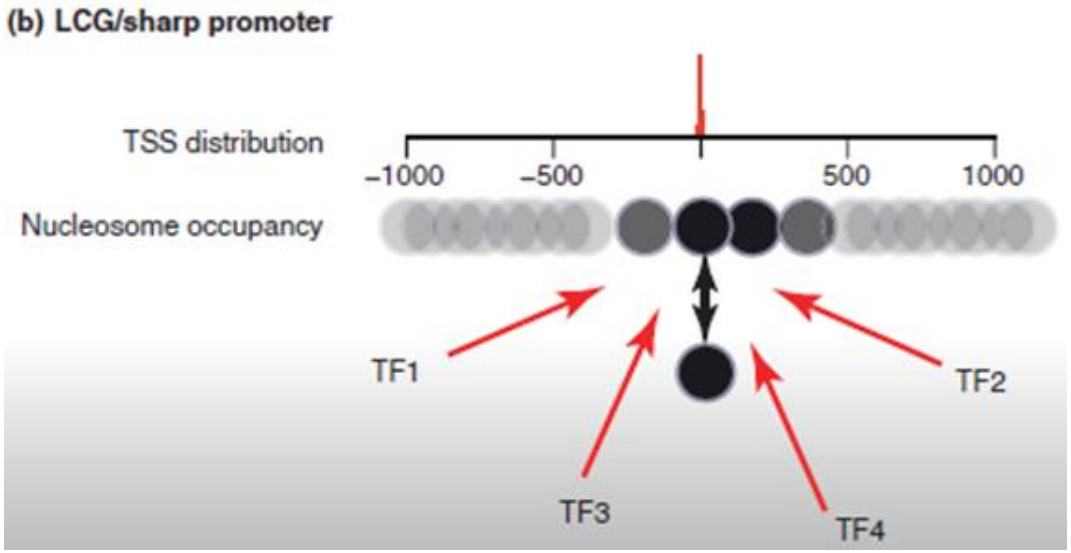




On-off genes

Usually LCG, TATA-plus

They possess intrinsic tendency to be «covered» i.e. repressed by nucleosome positioning

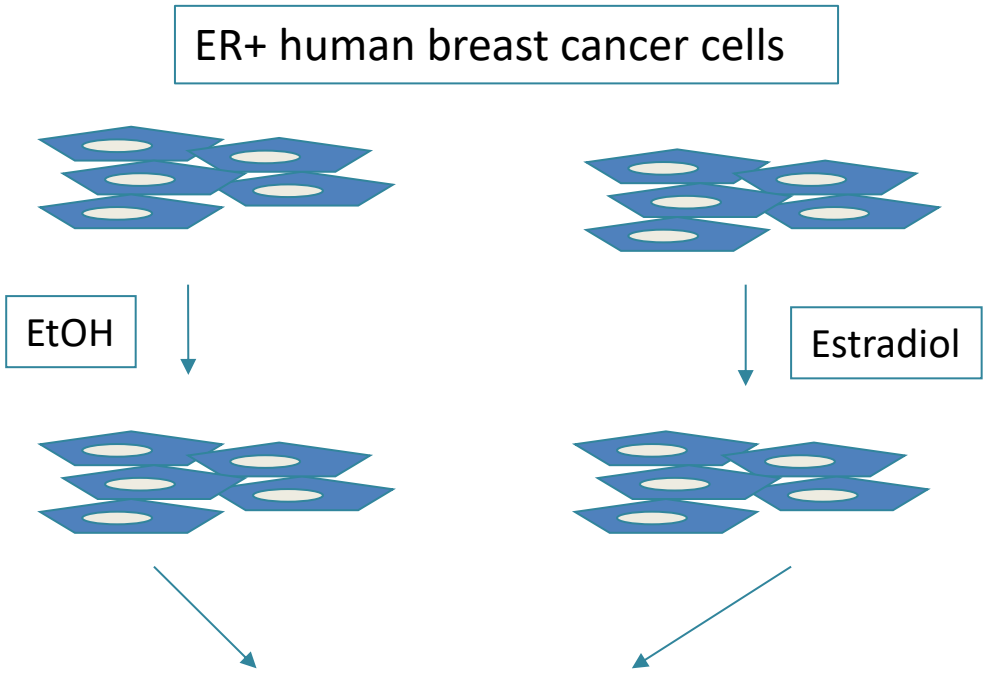


This system **requires ATP**, since activation requires ATP-dependent chromatin remodelers.

Many ATP-dependent chromatin remodelling enzyme families in Mammals:

SWI/SNF – ISWI – CHD - INO80 - SWR1

How is an «on-off» gene activated ?



Expression analysis (microarrays, RNA-Seq)
ER binding analysis (ChIP-chip, ChIP-Seq)

The pS2/TTF1 gene is estrogen-responsive and is one of the few genes that contains an ER-binding site close to the promoter. (From Métivier et al., 2003)

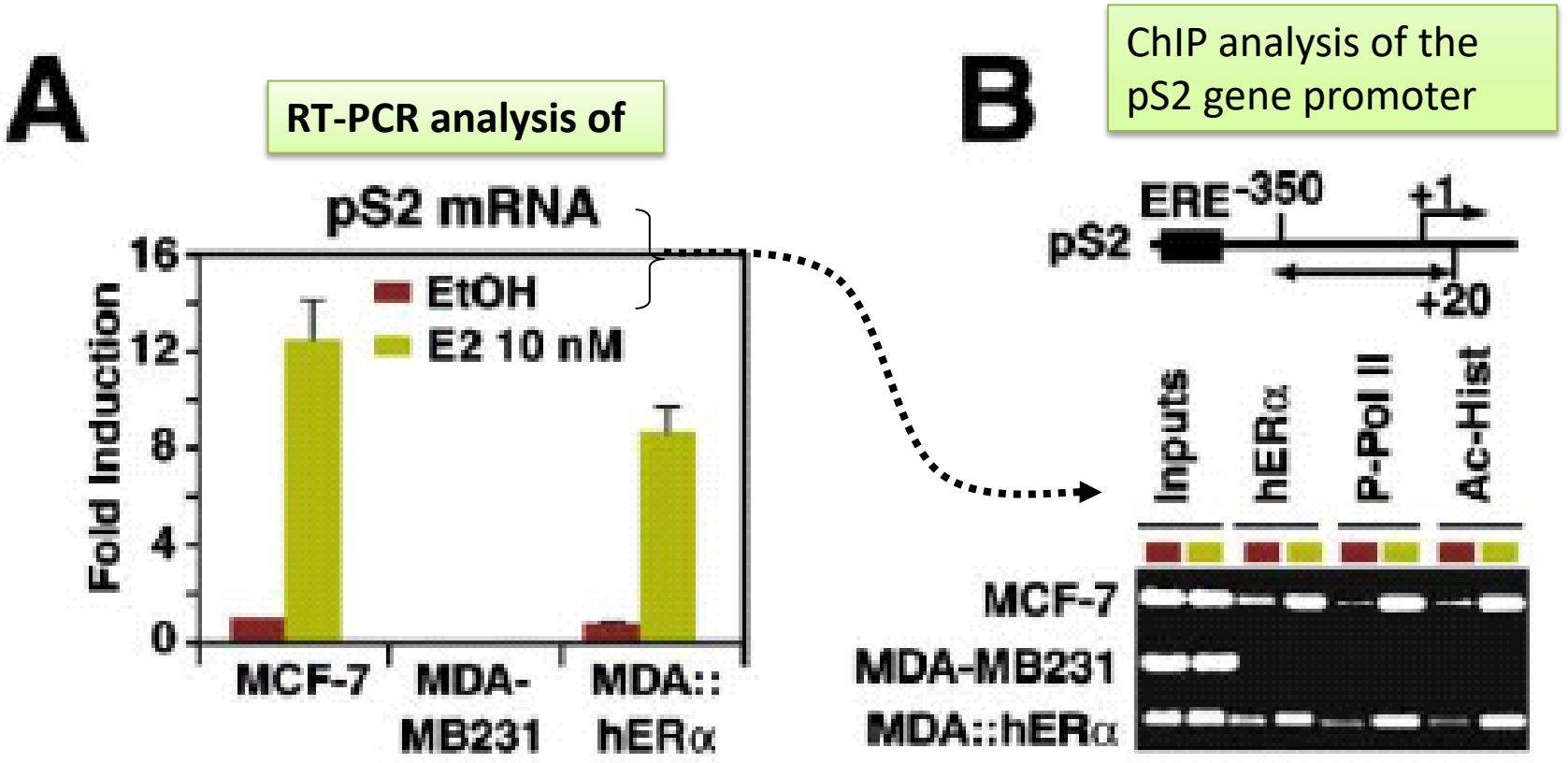


Figure 1. In Vivo Identification of the Transcription Factors Involved in pS2 Gene Activity

(A) Cells were treated for **3 hr** with 10 nM Estradiol (E2) or ethanol (EtOH) as vehicle control. The pS2 mRNA levels were normalized against invariant GAPDH mRNA, as measured by real-time PCRs.

(B) (B) Chromatin immunoprecipitations (ChIP) determining the recruitment of hER, Phosphorylated Pol II (P-Pol II) and acetylated histones (Ac-Hist) to the pS2 promoter after 3 hr treatment with 10 nM E2 or EtOH.

Cell, Vol. 115, 751–763, December 12, 2003, Copyright ©2003 by Cell Press

Estrogen Receptor-α Directs Ordered, Cyclical, and Combinatorial Recruitment of Cofactors on a Natural Target Promoter

Raphaël Métivier,[1,2,4,*] Graziella Penot,[1,2]
Michael R. Hübner,[1] George Reid,[1]
Heike Brand,[1] Martin Koš,[1,3] and Frank Gannon[1,*]
[1]European Molecular Biology Laboratory (EMBL)
Meyerhofstrasse 1
D-69117 Heidelberg
Germany

Summary

Transcriptional activation of a gene involves an orchestrated recruitment of components of the basal transcription machinery and intermediate factors, concomitant with an alteration in local chromatin structure generated by posttranslational modifications of histone tails and nucleosome remodeling. We provide here a comprehensive picture of events resulting in transcriptional activation of a gene, through evaluating the estrogen receptor-α (NR3A1) target pS2 gene promoter in MCF-7 cells. This description integrates chromatin remodeling with a kinetic evaluation of cyclical networks of association of 46 transcription factors with the promoter, as determined by chromatin immunoprecipitation assays. We define the concept of a "transcriptional clock" that directs and achieves the sequential and combinatorial assembly of a transcriptionally productive complex on a promoter. Furthermore, the unanticipated findings of key roles for histone deacetylases and nucleosome-remodeling complexes in limiting transcription implies that transcriptional activation is a cyclical process that requires both activating and repressive epigenetic processes.

- MCF-7 cells are starved of estrogen for several days
- 2 hours before treatment, they are added of **α-amanitin** (blocks transcription)
- Cells are then washed and treated with estradiol
- ChIP analysis for several factors is run at **5 minute intervals** on the pS2/TFF1 gene promoter

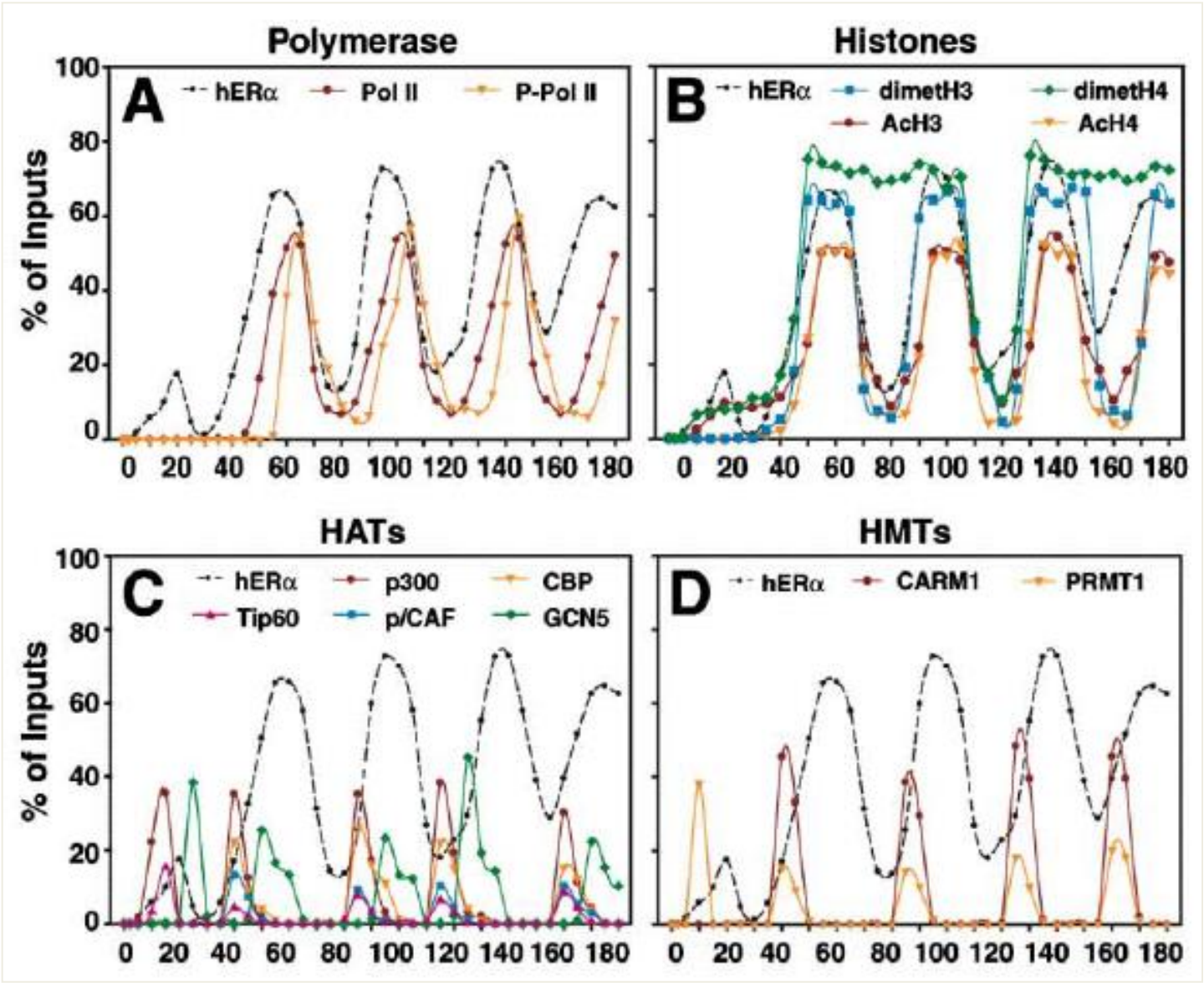


Figure 3. Dynamics of Cofactor Recruitment Directed by E2-Liganded hER on the pS2 Promoter. Kinetic ChIP experiments were performed using specified antibodies as shown within the images. Chromatin was prepared on sampled cells at 5 minutes intervals. The amount of immunoprecipitated pS2 promoter was quantified by real-time PCR. Values, expressed as % of the inputs, are the mean of three separate experiments, and have a SD 2%. All ChIP were performed from a single chromatin preparation for each time point.

Important to note that this dynamic is dramatically inhibited by proteasome inhibitors.

The role of Ubiquitin has been demonstrated in this dynamics.

In other words, the on-off kinetics seen in these experiments does not mena simply association-dissociation, but active displacement of components from the complexes, mediated by the proteasome.

Conclusion

Micro-chromatin environments are extremely dynamic, contrary to constitutive HC or constitutive “on” loci

These promoters tend to return quickly to «repressed» status.

This model is more or less demonstrated for a number of genes and regulators showing «transient» immediate transcriptional response.