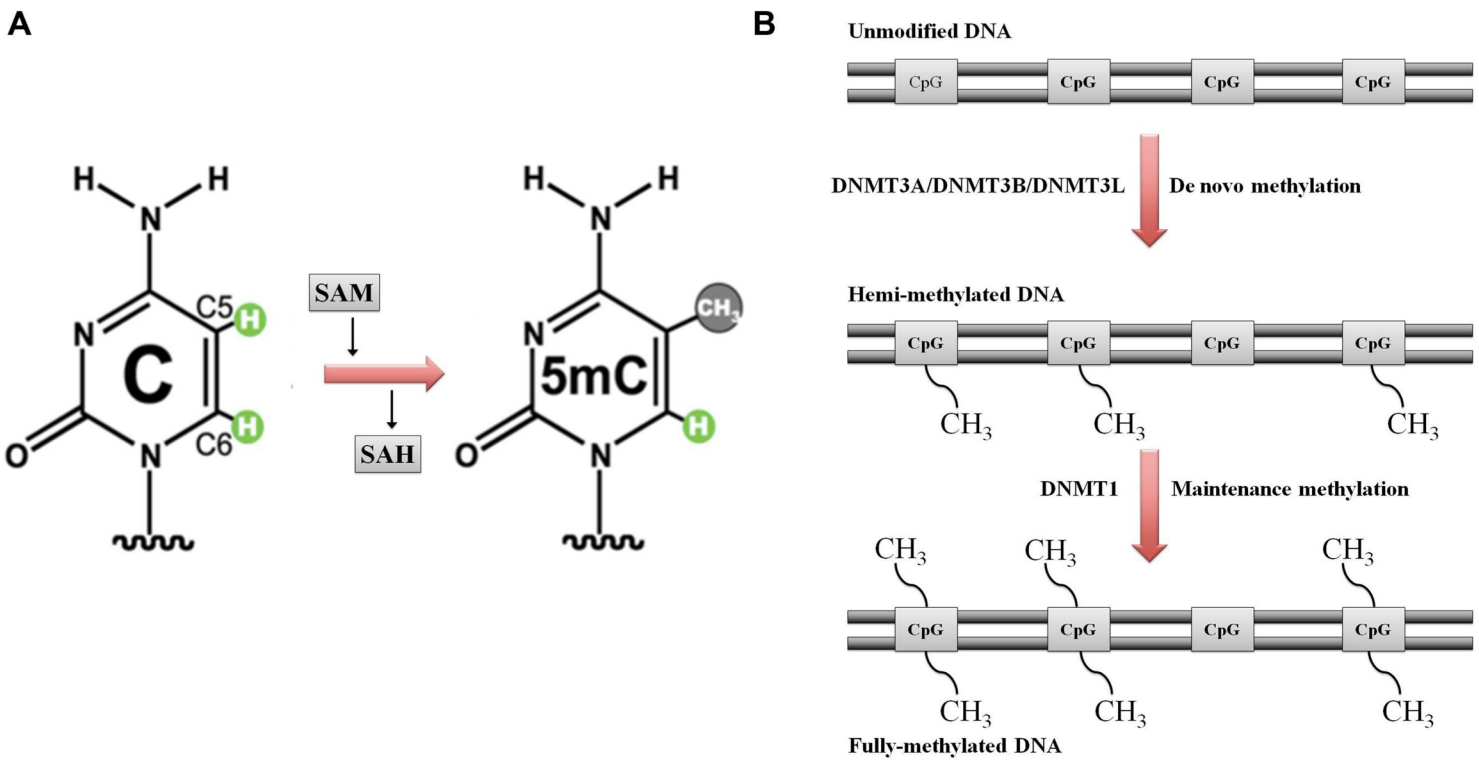
L4.3 – Transcriptional regulation (more on RNAP and Promoters)

AGENDA

1. Role of CpG methylation
2. Nucleosome positioning at promoters
3. RNA Polymerase pausing/stalling

DNA methylation



CpG-island distribution

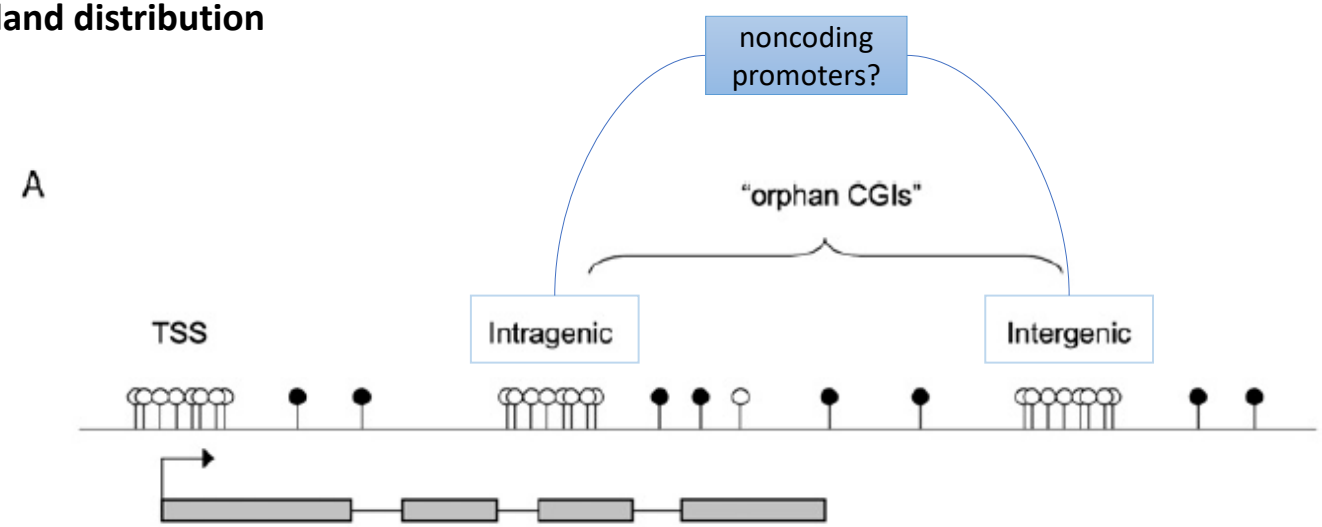


Figure 1. The genomic distribution of CGIs (CpG islands).
CGIs can be located at annotated TSSs, within gene bodies (Intragenic), or between annotated genes (Intergenic). Intragenic and intergenic CGIs of unknown function are classed as ‘‘orphan’’ CGIs. (Empty circles) Unmethylated CpG residues. (Filled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

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

Broad type promoters are usually GC-rich (I.e. they contains CpG-islands)

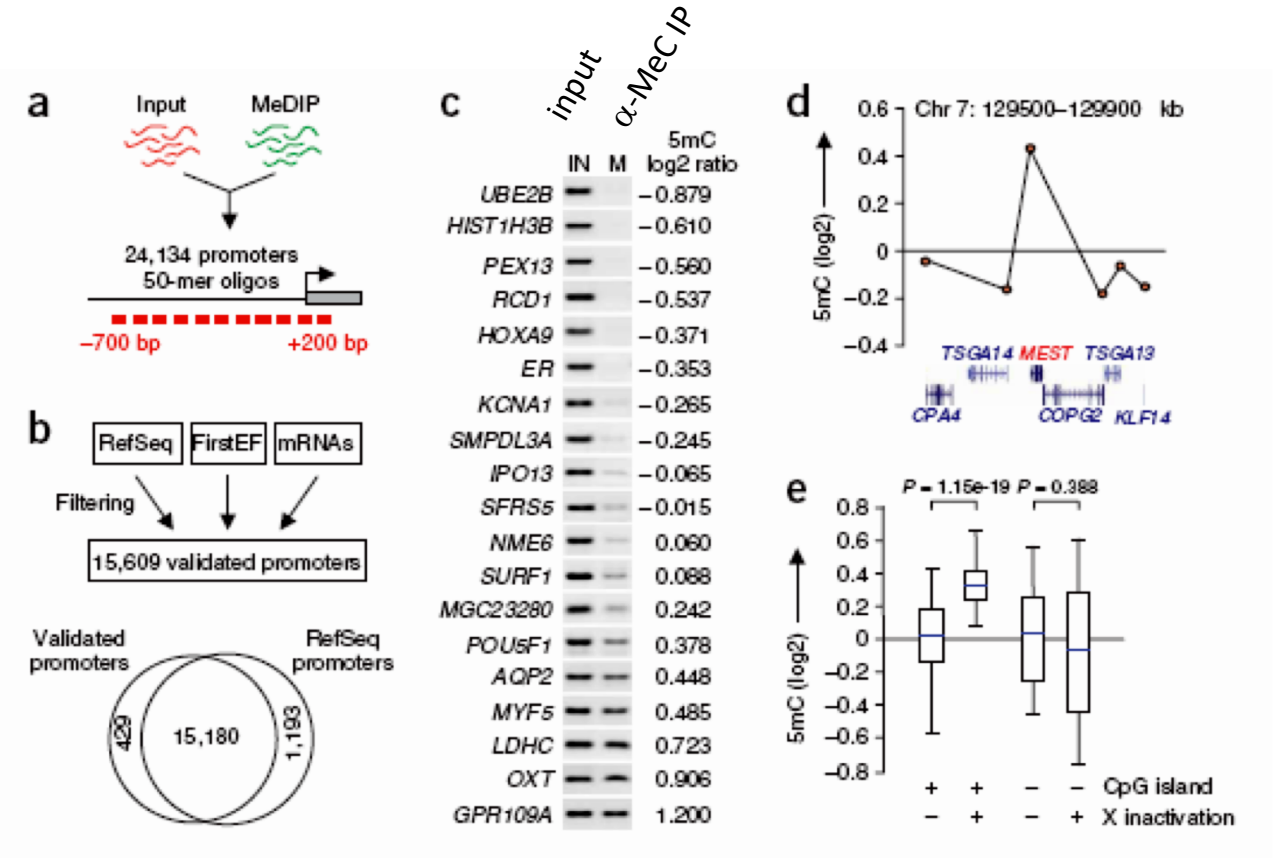
How is CpG methylation linked to function of these promoters ?

nature genetics

Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome

Michael Weber[1], Ines Hellmann[2,3], Michael B Stadler[1], Liliana Ramos[4], Svante Pääbo[2], Michael Rebhan[1] & Dirk Schübeler[1]

To gain insight into the function of DNA methylation at *cis*-regulatory regions and its impact on gene expression, we measured methylation, RNA polymerase occupancy and histone modifications at 16,000 promoters in primary human somatic and germline cells. We find CpG-poor promoters hypermethylated in somatic cells, which does not preclude their activity. This methylation is present in male gametes and results in evolutionary loss of CpG dinucleotides, as measured by divergence between humans and primates. In contrast, strong CpG island promoters are mostly unmethylated, even when inactive. Weak CpG island promoters are distinct, as they are preferential targets for *de novo* methylation in somatic cells. Notably, most germline-specific genes are methylated in somatic cells, suggesting additional functional selection. These results show that promoter sequence and gene function are major predictors of promoter methylation states. Moreover, we observe that inactive unmethylated CpG island promoters show elevated levels of dimethylation of Lys4 of histone H3, suggesting that this chromatin mark may protect DNA from methylation.

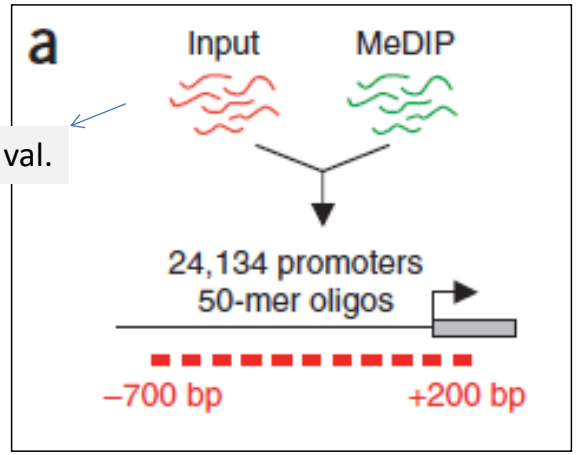


Validated
by PCR

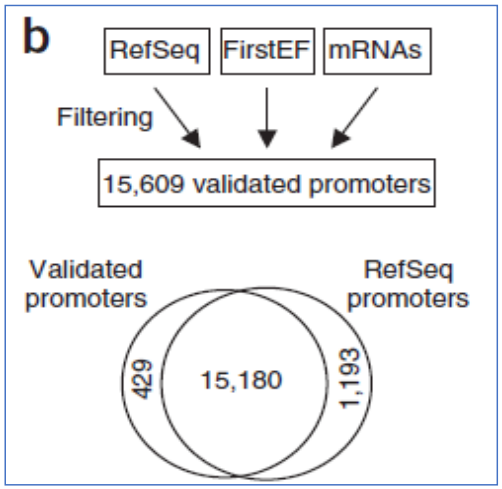
Human primary fibroblasts (ATCC WI38)/Mature spermatocytes (two healthy donors)

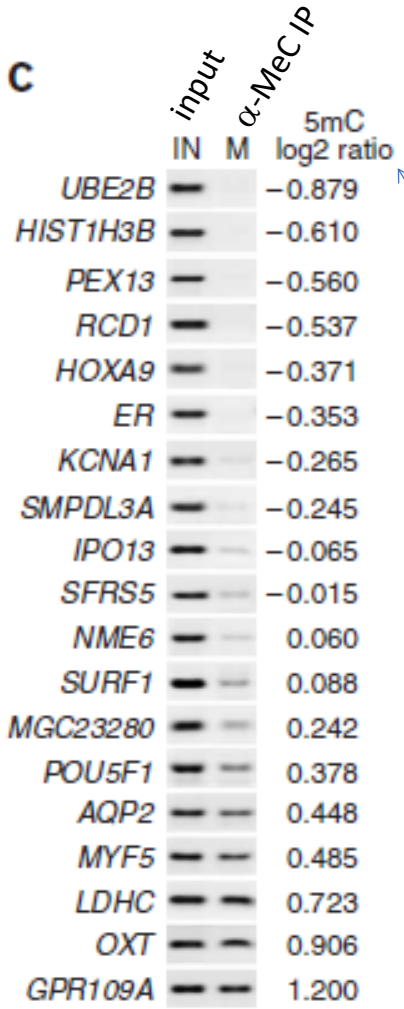
DNA extraction, purification. Sonicated to 300-1,000 bp

4μg fragmented DNA → anti-methyl-cytosine immunoprecipitation → labelling → promoter microarrays



15 probes from -700 to + 0.2 Kb

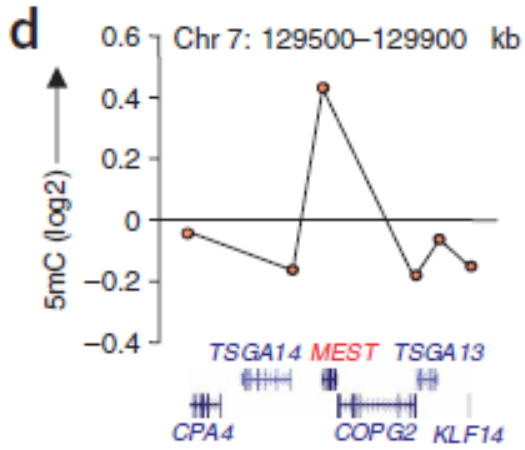


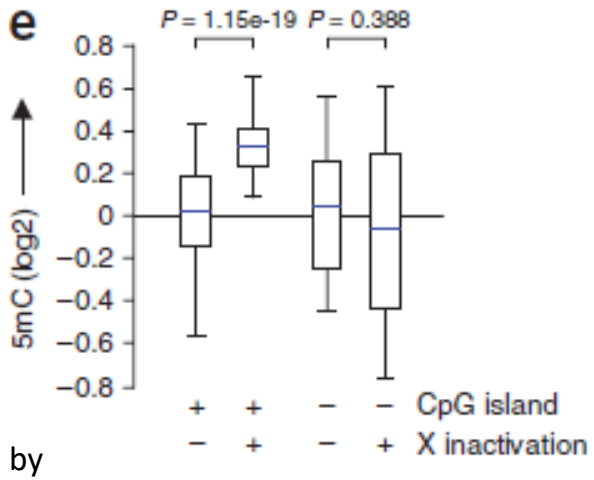


Technical validation by PCR

Values from microarray:
They are expressed aslog2

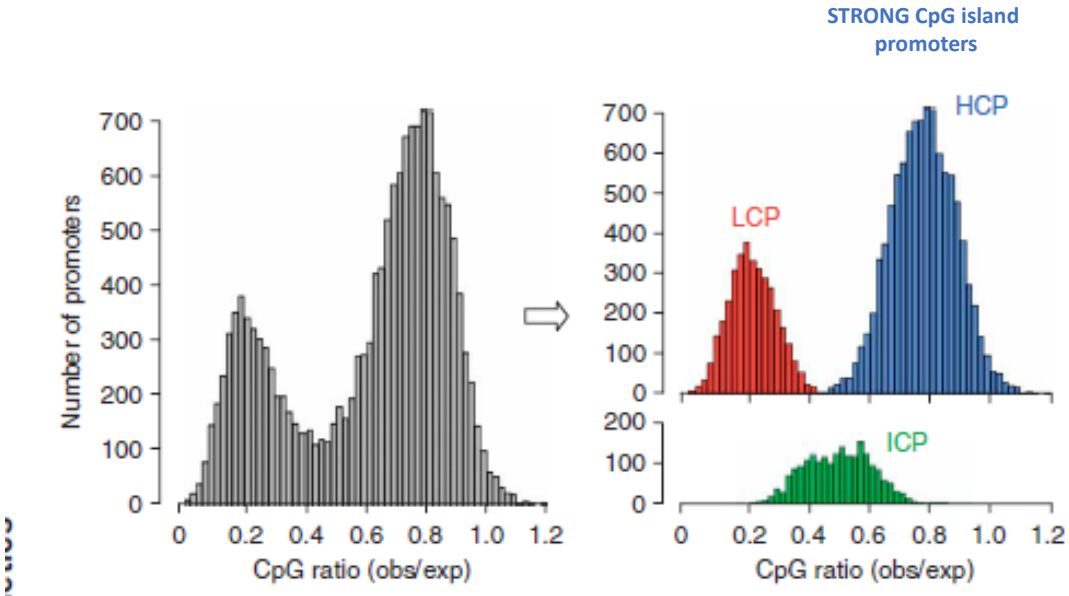
$$\log_2 \frac{\text{Intensity MeIP}}{\text{Intensity imput}}$$



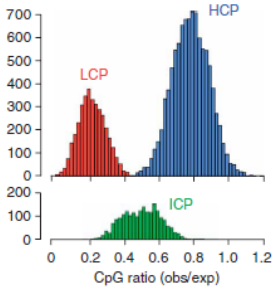


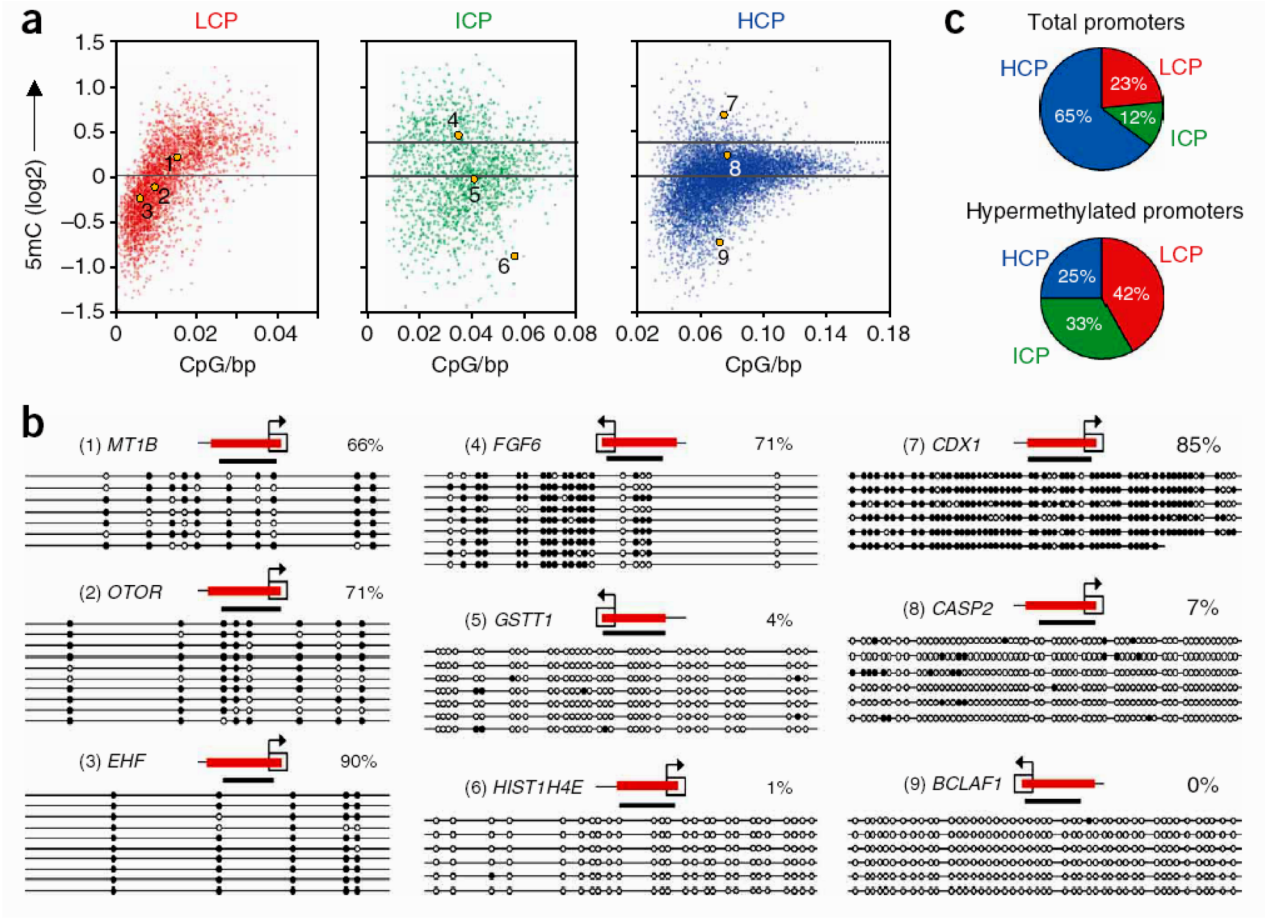
Validation by
Known genes

Promoter classification based on CpG density



$$\frac{\text{CpG observed \%}}{\text{CpG expected \%}}$$



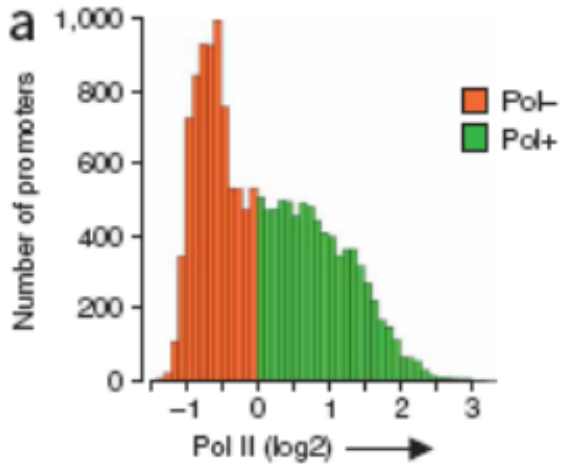


○ **unmethylated**

● **methylated**

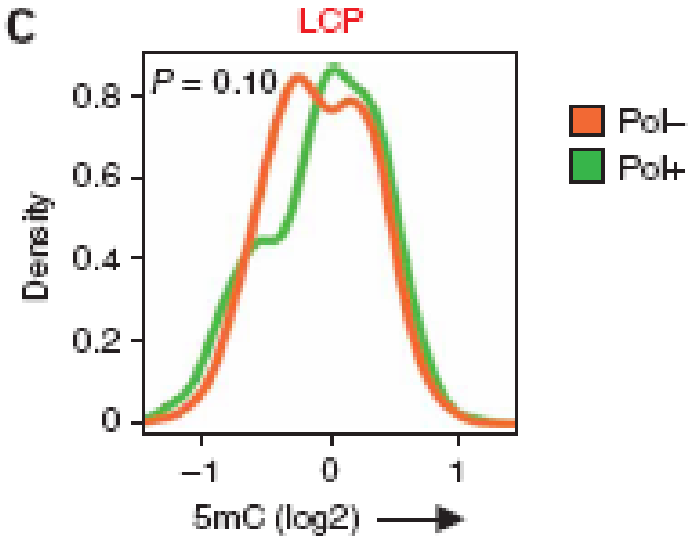
Does methylation correlate with activity?

ChIP using α-Pol II antibodies

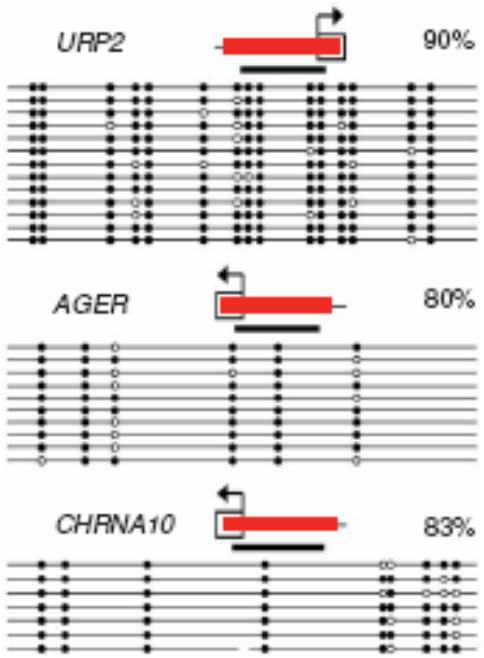


The frequency of activity varies between promoter classes, with 66% of HCPs being active compared with 41% of ICPs and 11% of LCPs. This reflects the enrichment of housekeeping genes in CpG island promoters and the higher abundance of rarely expressed tissue-specific genes in non-CpG island promoters

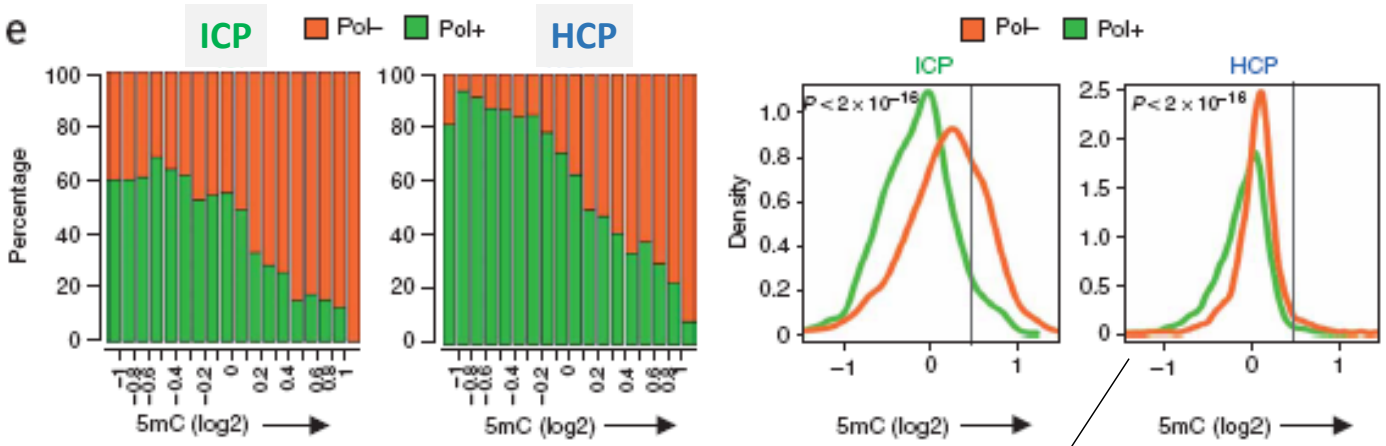
Low CpG Promoters



Examples of active LCP promoters:

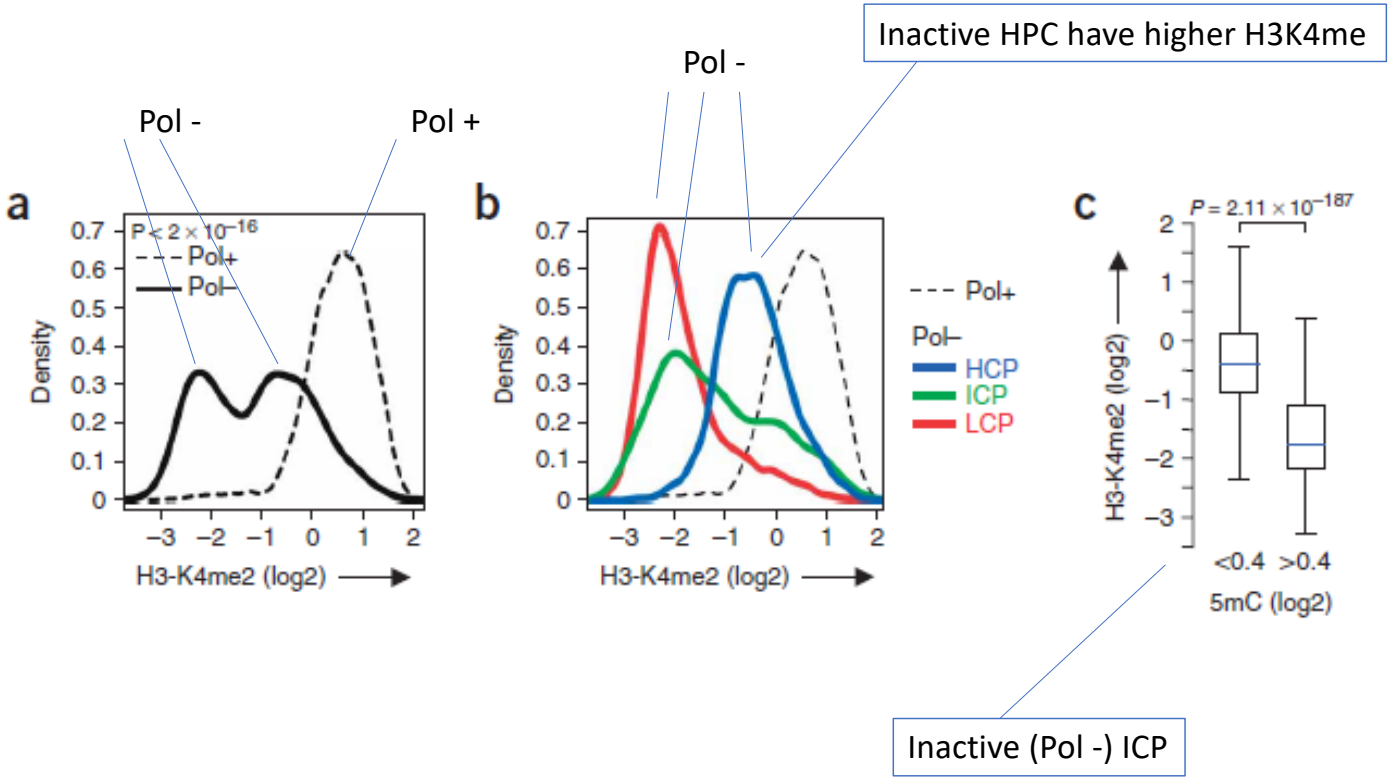


correlation between hypomethyation and activity in ICP and HCP



High proportion of inactive HCP are **hypomethylated**

α-H3K4me2 antibodies ChIP → same promoter microarrays



Conclusions:

Different promoters types are characterized by different CpG content and methylation effect

High CpG promoters are mostly unmethylated. They are enriched for active genes including most « housekeeper» gene promoters.

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

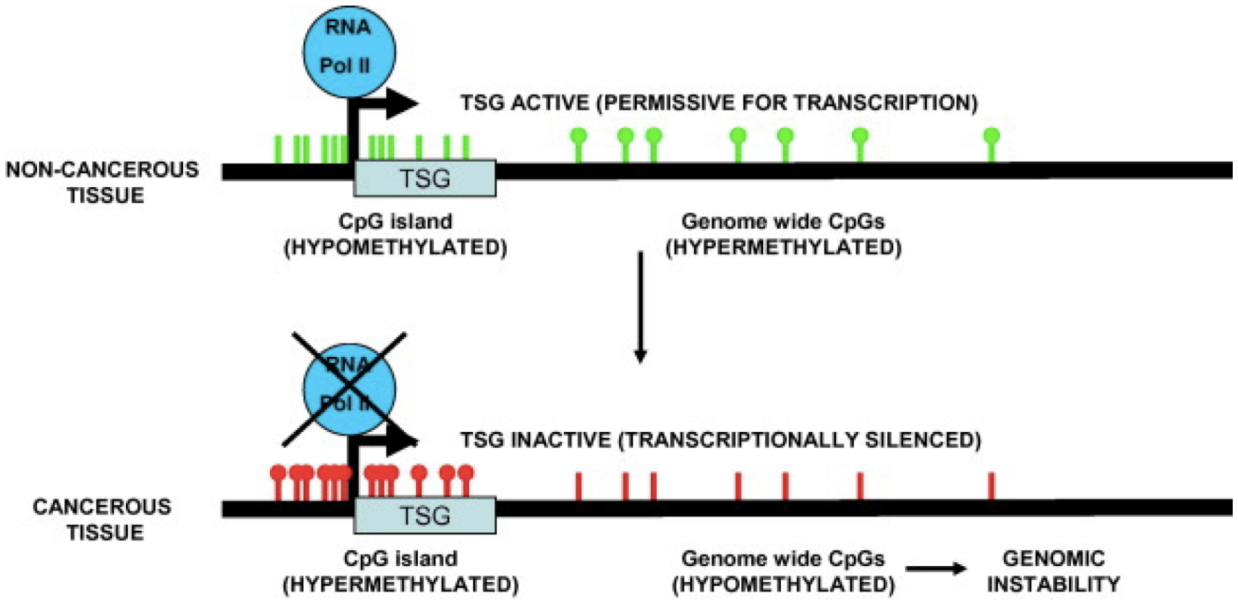
Intermediate CpG promoters (ICP) represent the class where methylation closely reflects RNA Pol occupancy, i.e. their transcription seems methylation dependent

Also, inactive HCP still have «active chromatin» mark H3K4me3

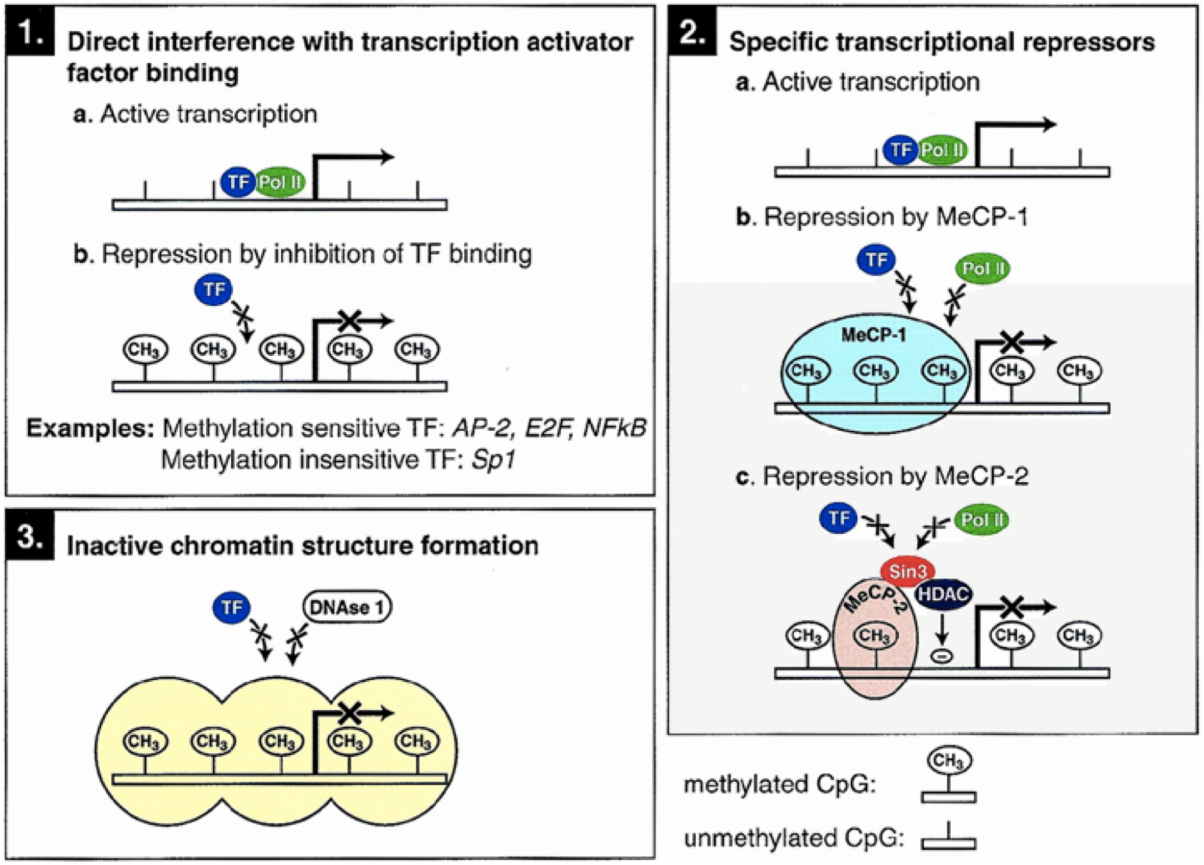
-> they might be «poised» to transcription

This suggests that the chromatin state can predict the DNA methylation state of inactive CpG-rich promoters and opens the possibility that chromatin structure is functionally involved in protecting CpG-rich promoters from DNA methylation.

Aberrant Methylation Patterns in Cancer



DNA Methylation and Transcriptional Repression

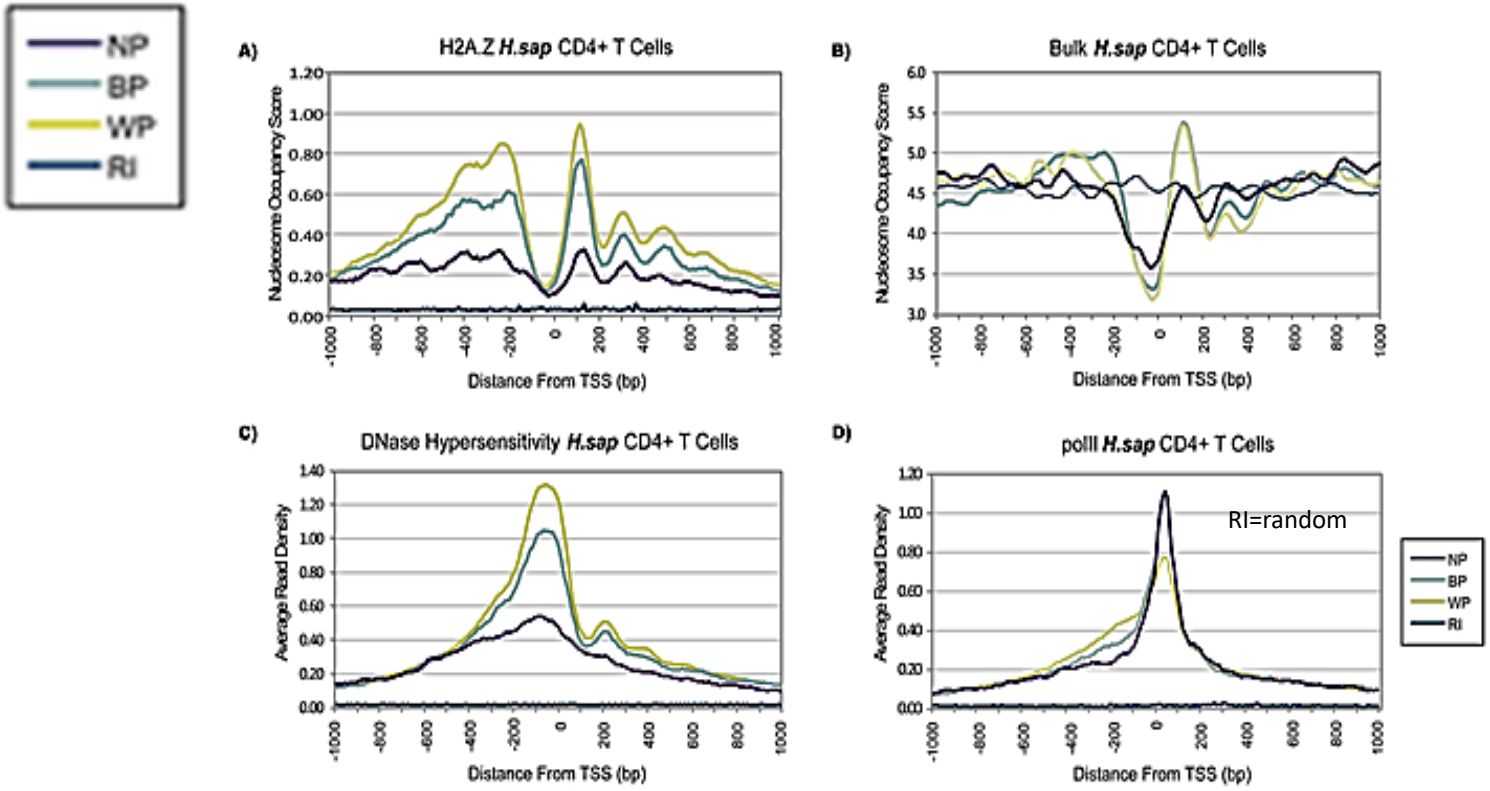


AGENDA

1. Role of CpG methylation
2. Nucleosome positioning at promoters
3. RNA Polymerase pausing/stalling

NP ≈ Sharp
WP ≈ broad
BP=Broad with peak

Figure 1. Promoter Classes Reflect Distinct Profiles of Nucleosome Organization.



From: Rach EA et al., 2011, PLOS Genet, e1001274

Profiles are based on promoters classified as Narrow Peak (NP), Broad with Peak (BP), and Weak Peak (WP), and show the region of –1 kb to +1 kb around the designated TSS. 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, and consistent spacing were observed for human BP and WP promoters compared to NP. 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).

Resource

Cell

Dynamic Regulation of Nucleosome Positioning in the Human Genome

Dustin E. Schones,[1,2] Kairong Cui,[1,2] Suresh Cuddapah,[1] Tae-Young Roh,[1] Artem Barski,[1] Zhibin Wang,[1] Gang Wei,[1] and Keji Zhao[1,*]
[1]Laboratory of Molecular Immunology, The National Heart, Lung and Blood Institute, NIH, Bethesda, MD 20892, USA
[2]These authors contributed equally to this work.
*Correspondence: zhaok@nhlbi.nih.gov
DOI 10.1016/j.cell.2008.02.022

SUMMARY

The positioning of nucleosomes with respect to DNA plays an important role in regulating transcription. However, nucleosome mapping has been performed for only limited genomic regions in humans. We have generated genome-wide maps of nucleosome positions in both resting and activated human $CD4^+$ T cells by direct sequencing of nucleosome ends using the Solexa high-throughput sequencing technique. We find that nucleosome phasing relative to the transcription start sites is directly correlated to RNA polymerase II (Pol II) binding. Furthermore, the first nucleosome downstream of a start site exhibits differential positioning in active and silent genes. TCR signaling induces extensive nucleosome reorganization in promoters and enhancers to allow transcriptional activation or repression. Our results suggest that H2A.Z-containing and modified nucleosomes are preferentially lost from the −1 nucleosome position. Our data provide a comprehensive view of the nucleosome landscape and its dynamic regulation in the human genome.

To prepare mononucleosomes, the resting and activated T cells were treated with MNase to generate approximately 80% mononucleosomes and 20% dinucleosomes.

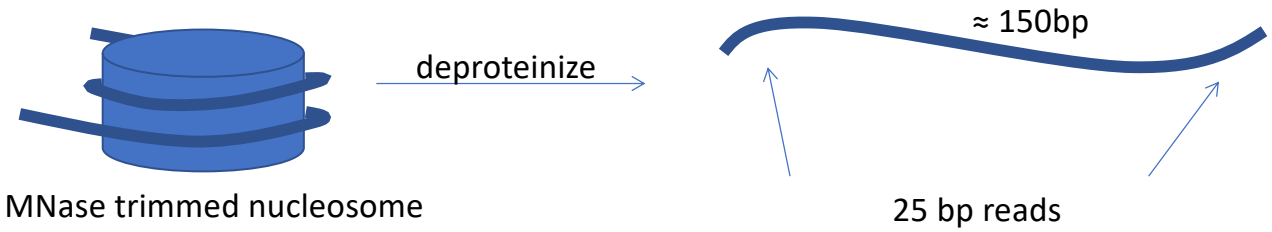
The DNA fragments of approximately 150 bp were isolated from agarose gel, blunt-ended, ligated to the Solexa adaptors, and sequenced using the Illumina 1G Genome Analyzer as described previously (Barski et al., 2007).

Cell *132*, 887–898, March 7, 2008 ©2008 Elsevier Inc.

MNase I

Micrococcal Nuclease II is a mixed endo-exonuclease.

MNase induces single-strand breaks and subsequently double stranded ones by cleaving the complementary strand in close proximity to the first break. MNase continues to digest the exposed DNA ends until it reaches an obstruction, such as a nucleosome or a very stably bound Transcription Factor.
In appropriate condition, then, MNase releases fragments of approximately one nucleosome length (~147 bp), which are typically selected for sequencing in MNase-Seq experiments.



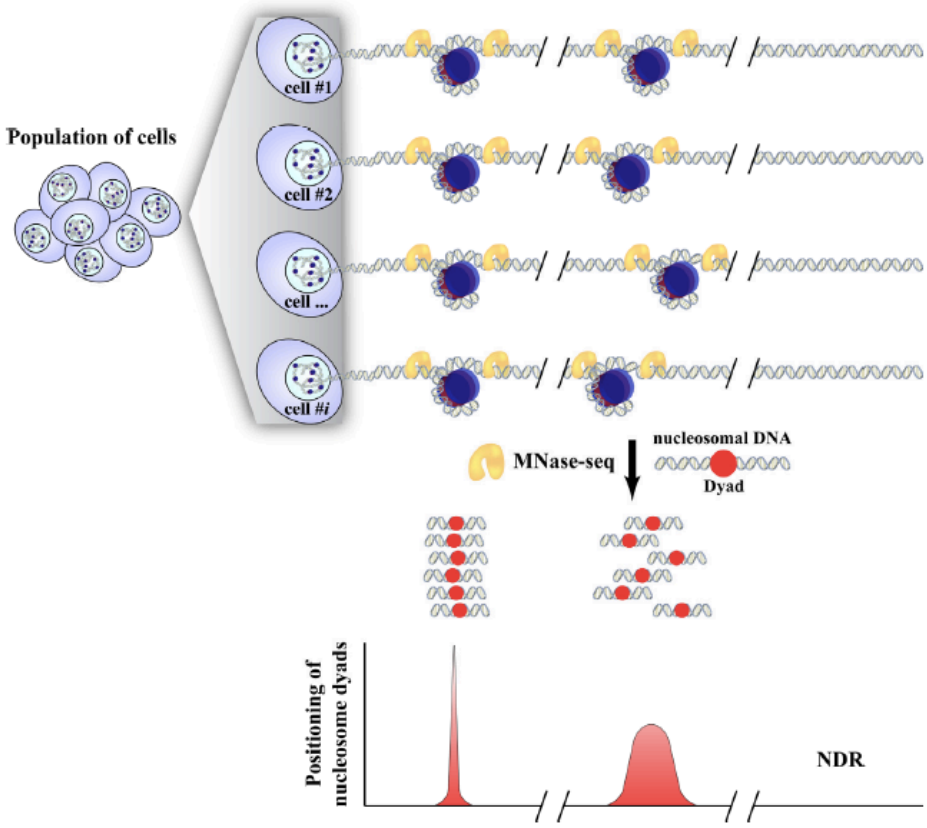
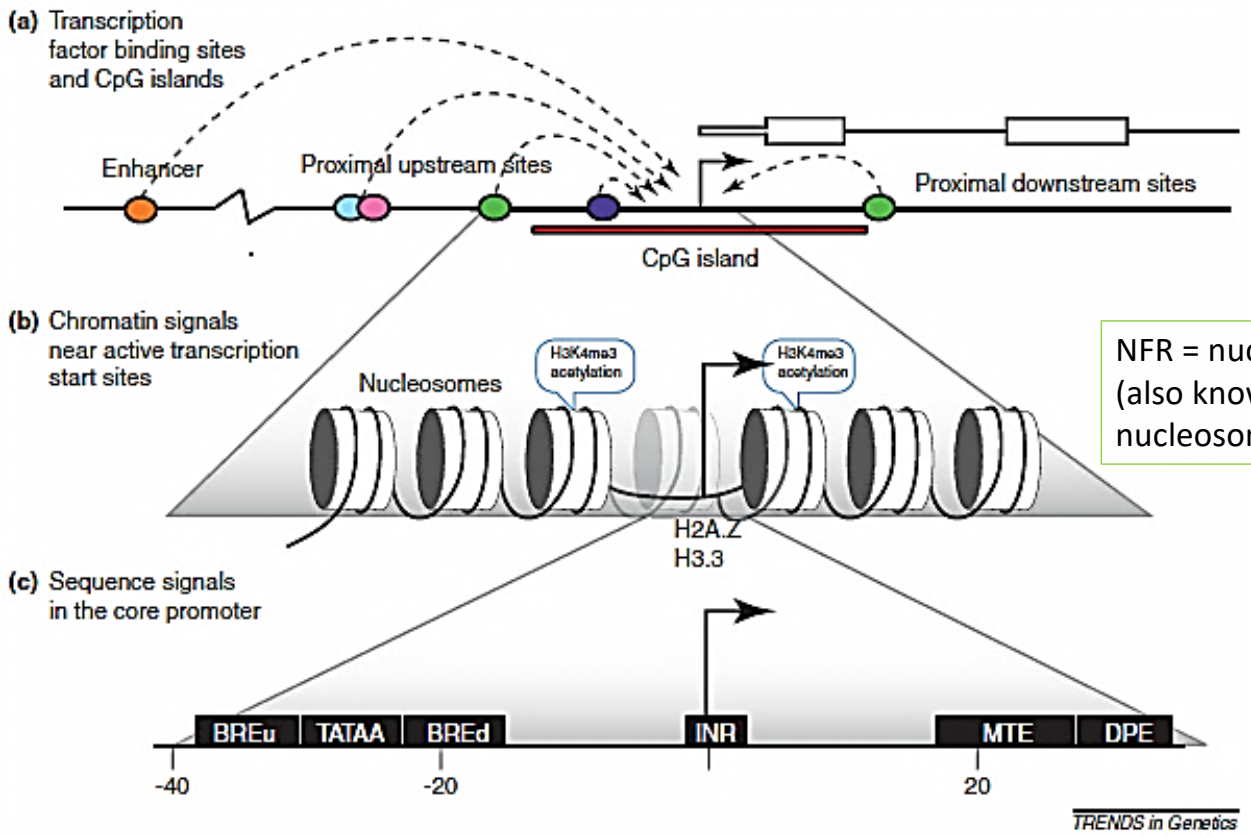


Figure 4. Mapping of nucleosome positioning by MNase-seq. MNase cuts linker DNA releasing free nucleosomes from chromatin. Alignment of the centres of paired-end sequenced nucleosomal DNA fragments to a reference genome points to nucleosome positioning (red peaks).

CONSTITUTIVE Genes



NFR = nucleosome-free region (also known as NDR or nucleosome-depleted region)

REGULATED Genes

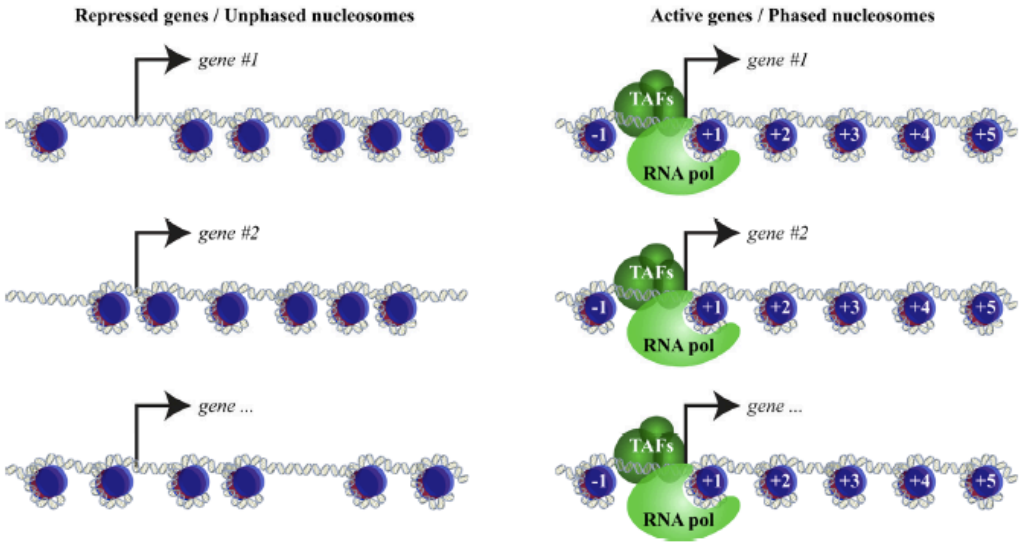
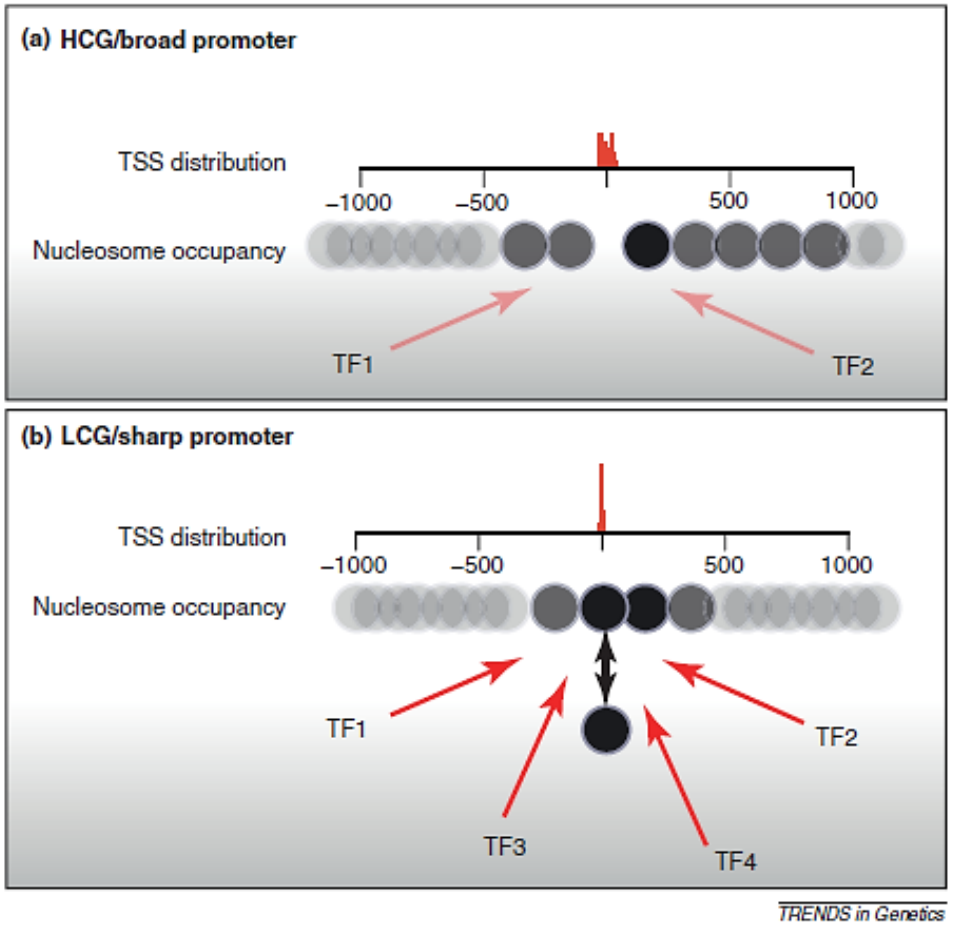


Figure 5. Nucleosome arrays of silenced (repressed) genes are out of phase with respect to TSS (left), whereas nucleosome arrays of active genes are phased off the TSS (right).

Mostly constitutive

Mostly regulated



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

NFR = nucleosome-free region
(also known as nucleosome-depleted region NDR)

HCG promoters

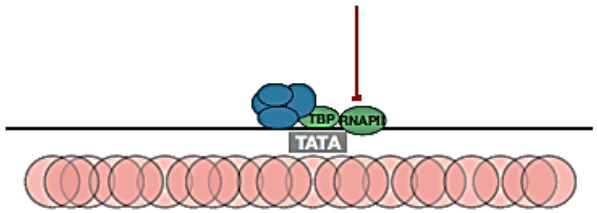
- «Broad» TSS – several TSS spread over 30-100bp
- CpG-island, undermethylated
- Functions: housekeepers, i.e. Ubiquitous expression
- Stable expression levels
- Bound mostly by ubiquitous Transcription Factors (TFBS over-represented close to TSS, e.g. Sp1)
- First nucleosome downstream TSS strongly positioned.
- Nucleosomes flanking NFR enriched for H2A.Z + H3.3 (also if «poised»)
- Most expression-predictive PTMs: H3K27ac and H4K20me1

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

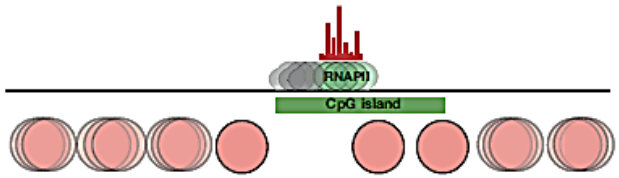
Adult tissue-specific

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



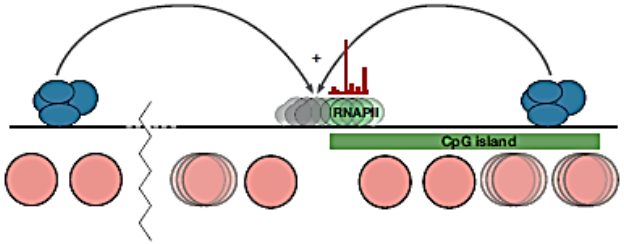
Ubiquitously expressed ("housekeeping")

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



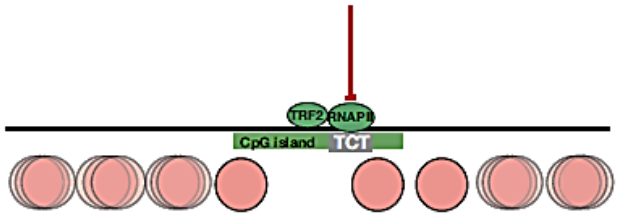
Developmentally regulated

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

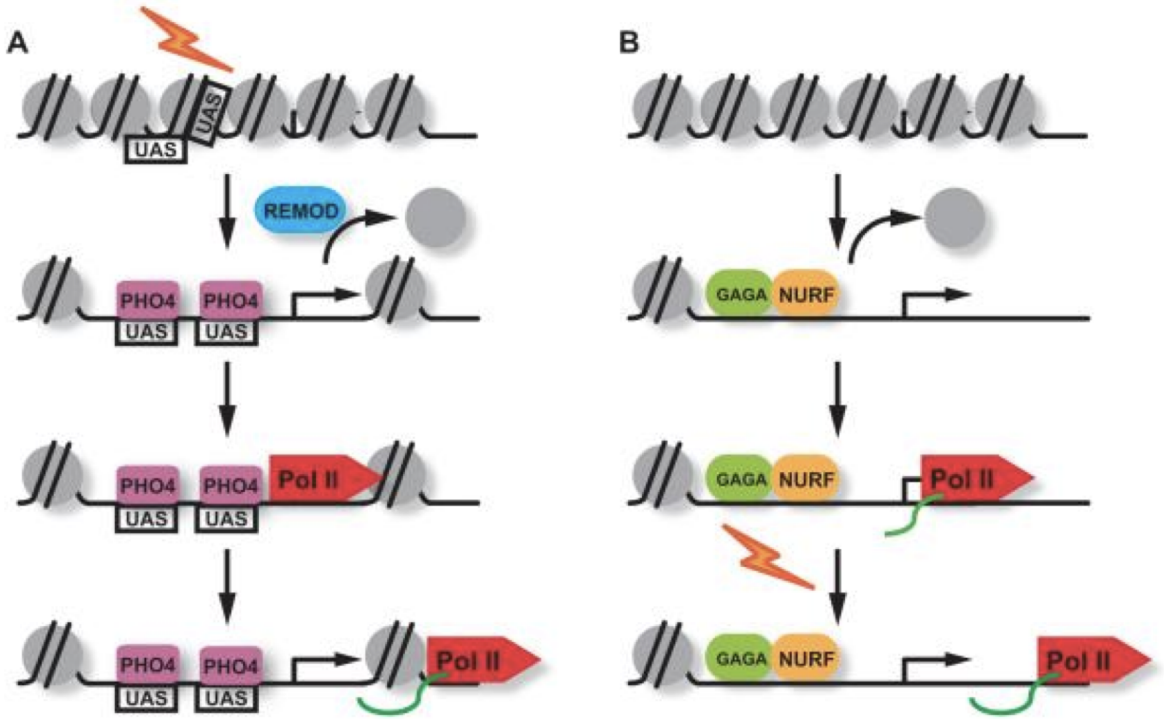


BREAK

AGENDA

1. Role of CpG methylation
2. Nucleosome positioning at promoters
3. RNA Polymerase pausing/stalling

Chromatin remodeling in the regulation of Inducible Gene Expression

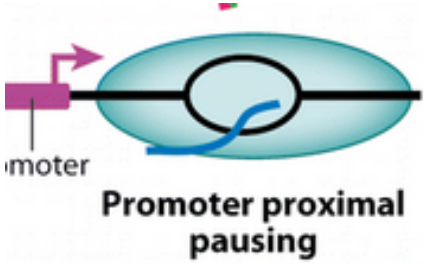


Gilchrist and Adelman, BBA 2013

Role of RNAP Pausing in transcriptional Regulation

Textbook

Adelman & Lis
Nature Reviews 2013

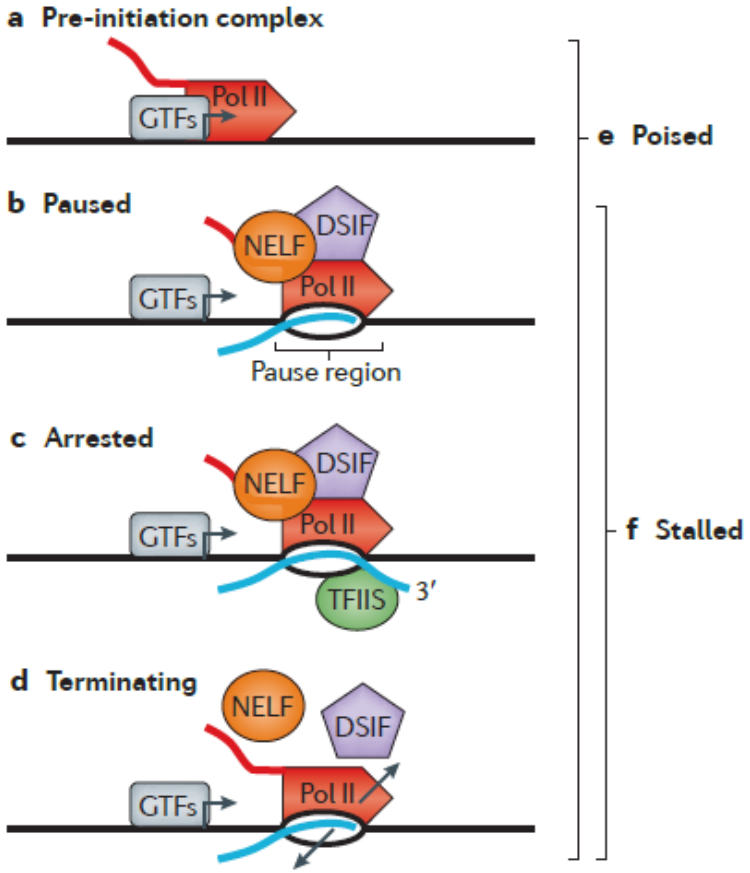


Many genes were found to contain «stalled» or paused RNA Polymerase II (PolII)

This behaviour was originally observed on stress response genes, but today appears to be quite widespread.

Textbook

NOMENCLATURE

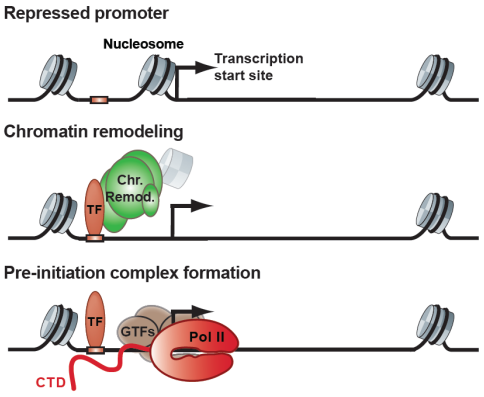


b | Paused: an early elongation complex that has transiently halted RNA synthesis. Paused polymerase is fully competent to resume elongation, remaining stably engaged and associated with the nascent RNA. Two protein complexes, DSIF and NELF, reduce the rate of elongation and facilitate the establishment of the stably paused state.
c | Arrested: a stably engaged elongation complex wherein the polymerase has backtracked along the DNA template, such that the RNA 3′ end is displaced from the active site. Restart of an arrested complex usually requires TFIIS, which induces Pol II to cleave the nascent RNA at the active site, creating a new 3′ end that is properly aligned with the Pol II active site and releasing a short (2–9-nucleotide) 3′ RNA.
d | Terminating: an unstable elongation complex that is in the process of dissociating from the DNA template and releasing the nascent RNA. The released Pol II could have the potential rapidly to reinitiate transcription and to 'recycle' at the promoter.

e | Poised: a generic term that simply indicates that Pol II is located near the TSS but does not specify anything about its transcriptional status. It can include any of the above complexes (**a**–**d**).
f | Stalled: a term that indicates Pol II is engaged in transcription but that makes no assumptions about its ability to resume synthesis. This term includes paused, arrested and terminating complexes

Metazoan gene expression is regulated by promoter-proximal pausing of Pol II

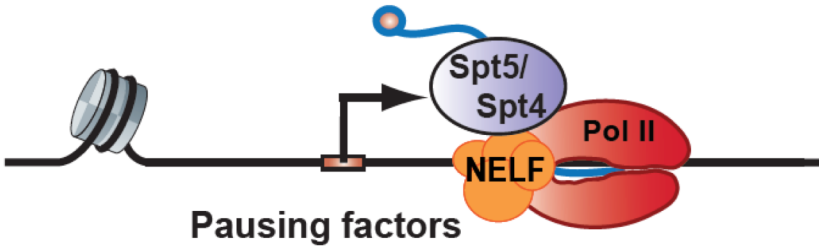
- Paused Pol II remains associated with nascent 25-60 nt RNA
- This complex can 'wait' patiently for pause release
- Kinase P-TEFb is required for productive elongation and mRNA synthesis



Henriques et al. (2013) Mol *Cell;* Jonkers et al. (2014) *eLife;* Chen et al. (2015) *Genes. Dev*

Textbook

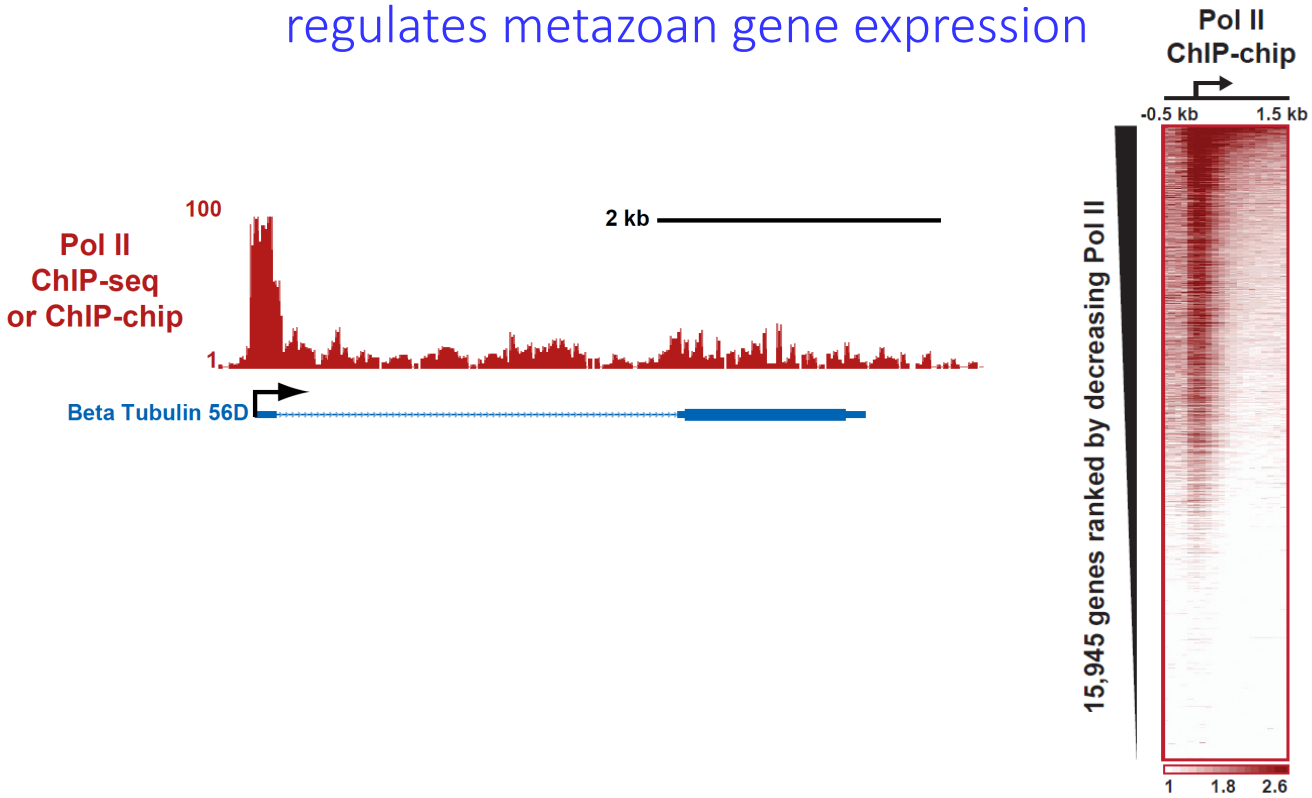
Pause-inducing factors NELF and DSIF (Spt4/Spt5)



- Spt5 subunit of DSIF is conserved from bacteria (NusG) to man
- Spt5 is an essential protein
- After pause release, Spt5 stays associated with Pol II and recruits chromatin and RNA processing factors

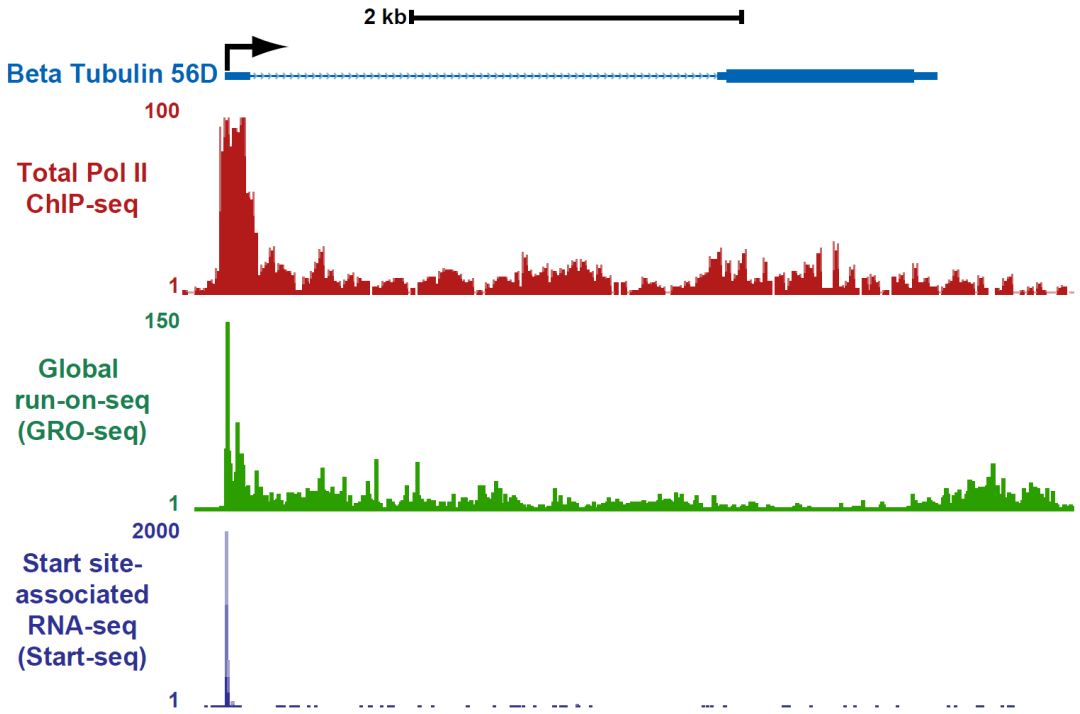
- NELF stabilizes paused Pol II and dissociates upon pause release
- NELF complex (4 proteins) is present in flies and mammals, but absent in yeast, worms and plants

Pausing of RNA polymerase II regulates metazoan gene expression



Muse et al. (2007) *Nat. Genet.*; Zeitlinger et al. (2007) *Nat. Genet.*; Guenther et al. (2007) *Cell*

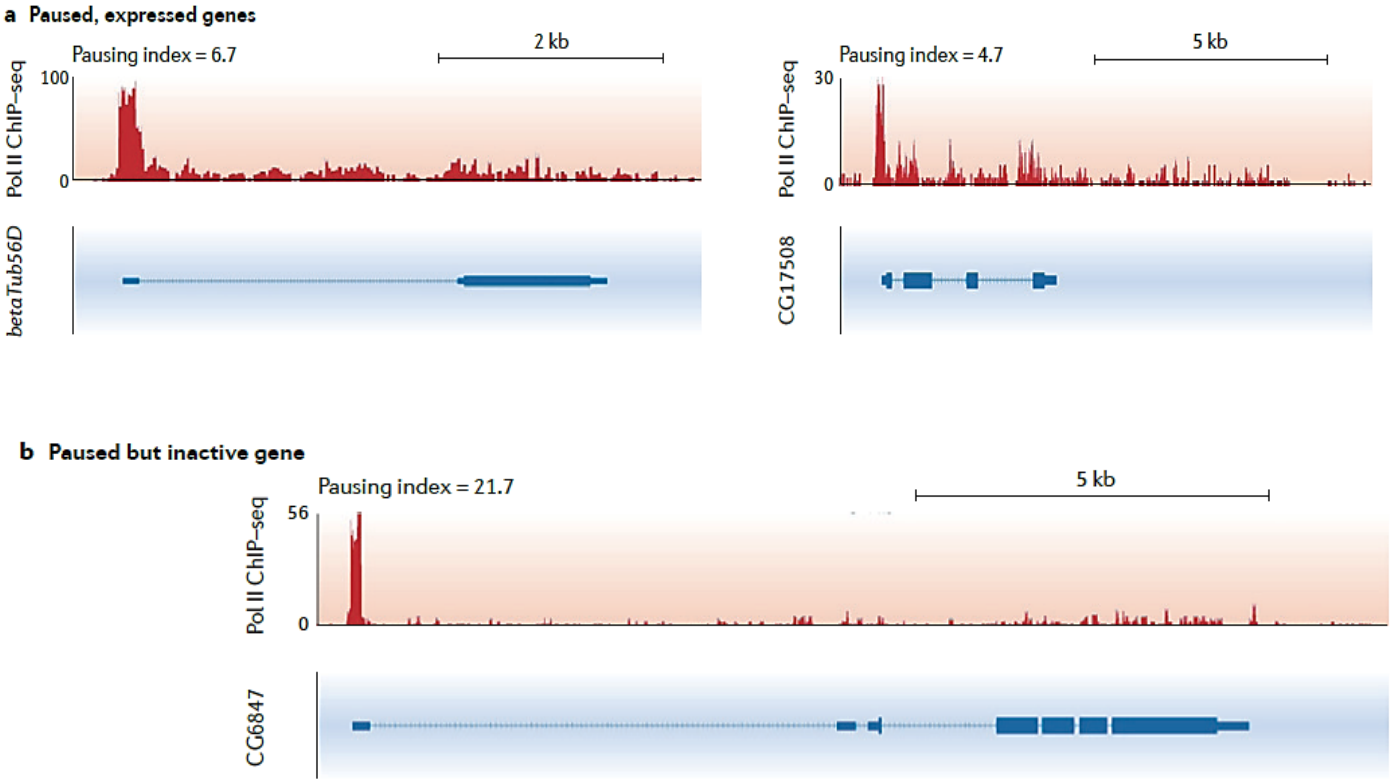
Analysis of nascent RNA confirms that promoter-associated Pol II is engaged in early elongation



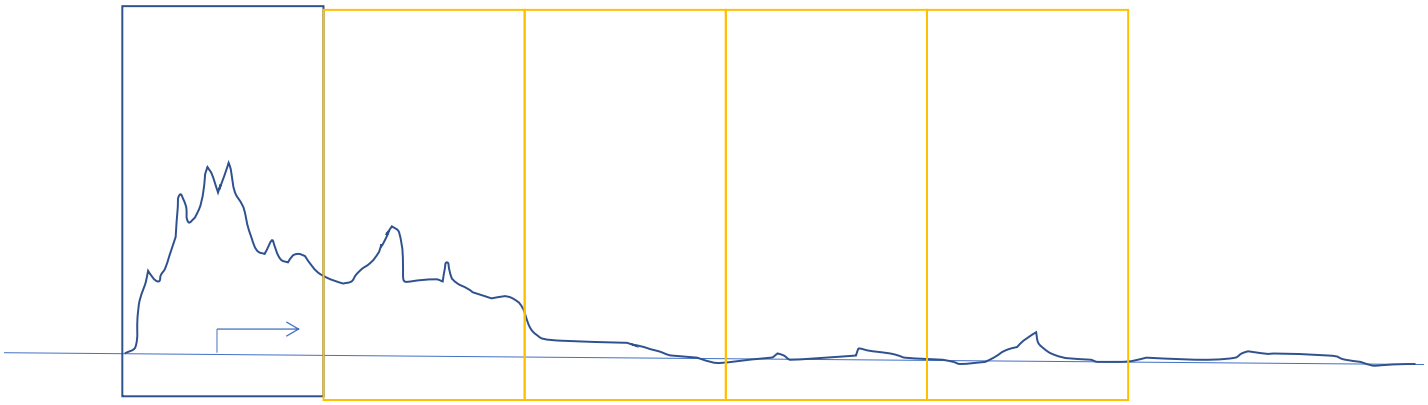
Core et al. (2008) *Science*; Nechaev et al. (2010) *Science*

Pol II ChIP-Seq

Textbook



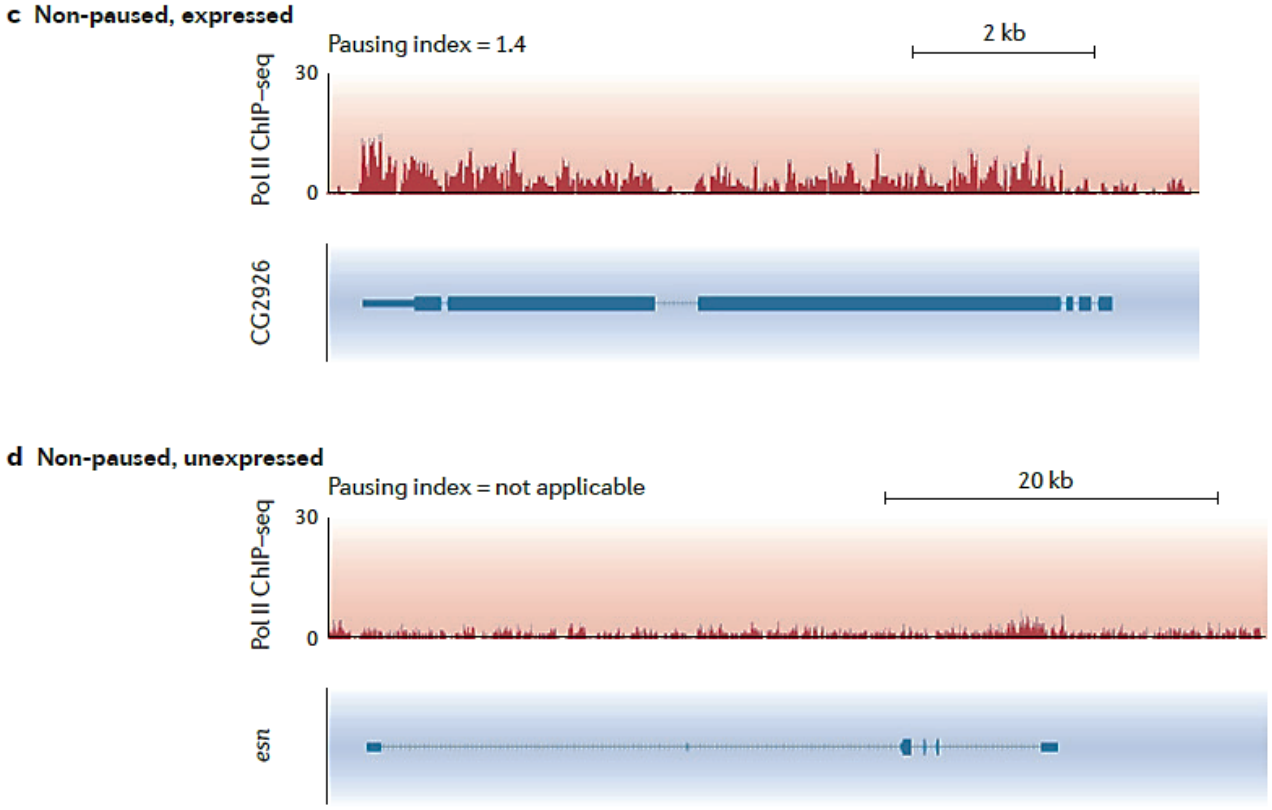
Pausing Index: ration of Pol2 promoter level over the average level of POL2 across the gene body



1 kb

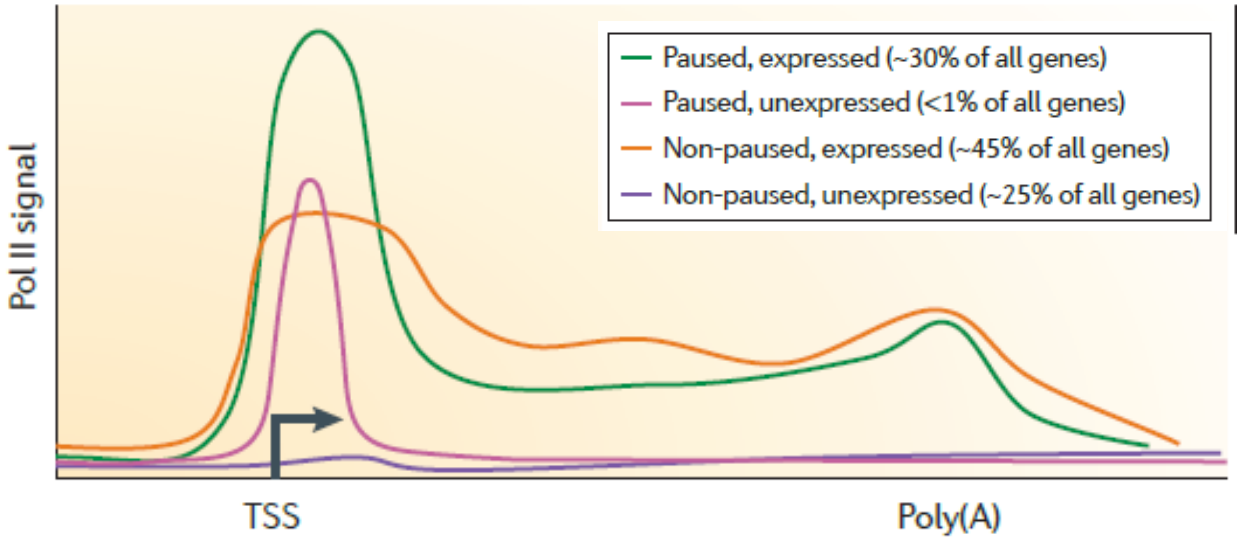
PolII occupancy at promoters (reads in blue box) relative to occupancy in the gene body (orange boxes)

Textbook


c Non-paused, expressed
Pausing index = 1.4
2 kb
30
0
Pol II ChIP–seq
CG2926
d Non-paused, unexpressed
Pausing index = not applicable
20 kb
30
0
Pol II ChIP–seq
esn

Textbook

e Summary



Textbook

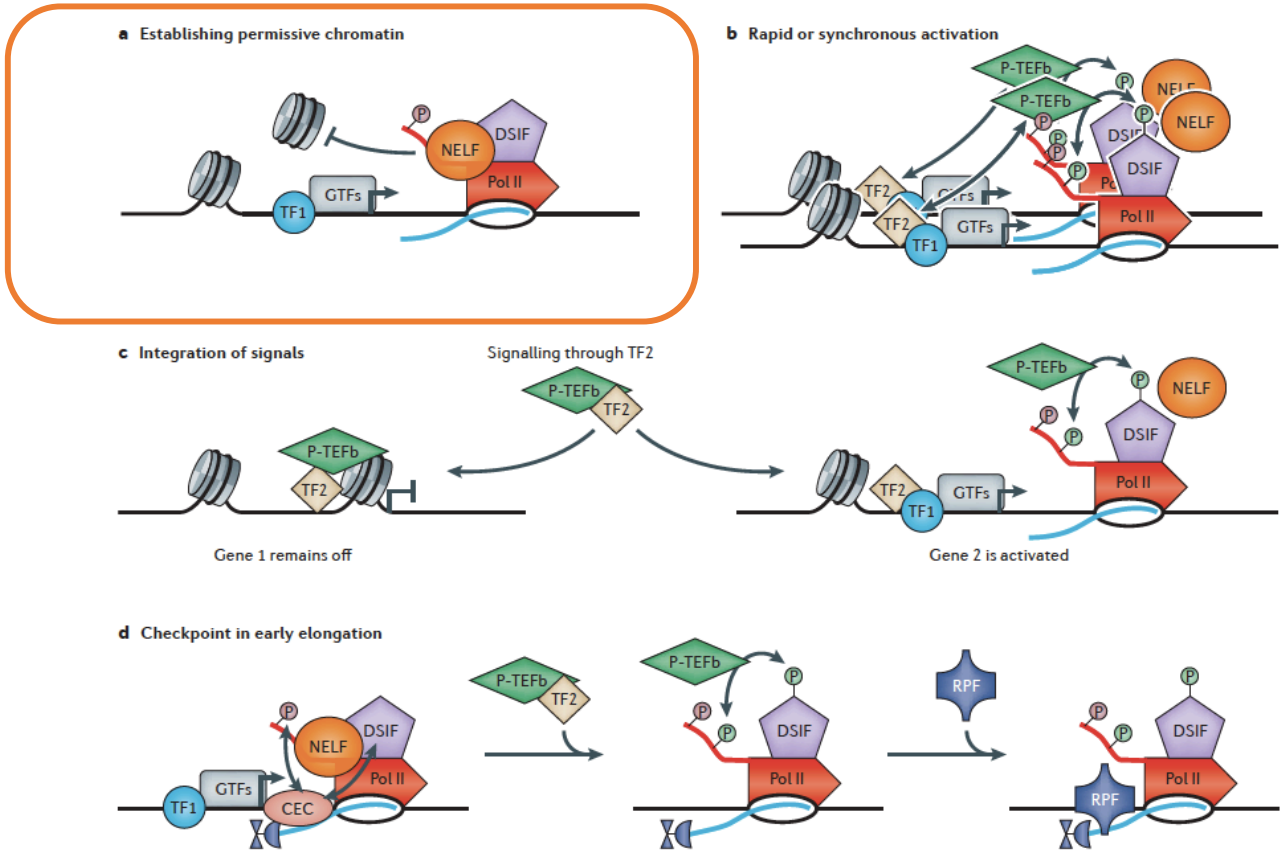
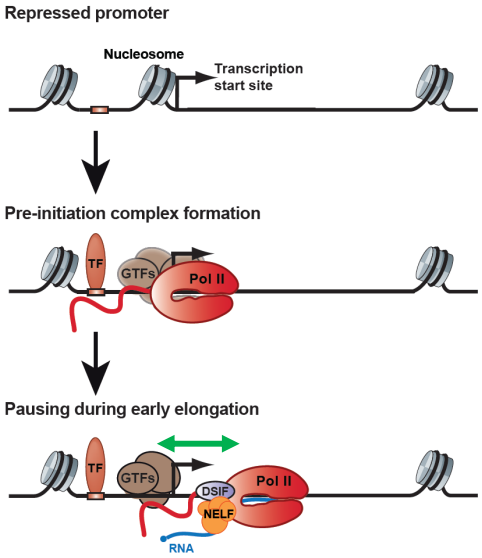


Figure 4 | **Illustrations of the main hypotheses for the functions of Pol II pausing.** The promoter region is depicted with the transcription start site (TSS) labelled with an arrow. Nucleosomes are shown in grey, and RNA polymerase II (Pol II) is illustrated as a red rocket. The general transcription factors (GTFs; grey oval) are shown centred at the TSS. The pause-inducing factors negative elongation factor (NELF; orange oval) and DRB-sensitivity-inducing factor (DSIF; purple pentagon) are shown. The nascent RNA transcript is shown in blue. **a** | Establishing permissive chromatin.

Stable Pol II occupancy of a region shapes chromatin structure

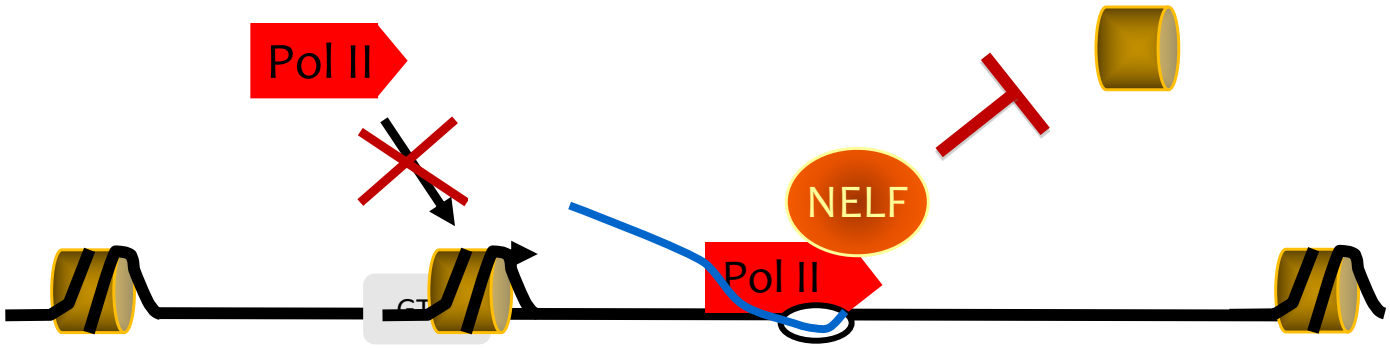


Paused Pol II

- Reduces nucleosome occupancy over promoters
- Stabilizes a scaffold of General Transcription Factors to stimulate multiple rounds of transcription

Adelman and Lis (2012) *Nat. Rev. Genetics*

NELF-mediated pausing promotes gene expression by blocking nucleosome assembly



Loss of NELF-mediated pausing causes:
- Lower levels of promoter Pol II
- Increased promoter histone occupancy
- Decreased gene expression
- Reduced gene inducibility

Gilchrist et al. (2008) *Genes & Dev.*
Gilchrist et al. (2010) *Cell*
Gilchrist et al. (2012) *Genes & Dev.*