In Lesson 1

- **Functional genomics is a field of molecular biology based on genome-wide sequencing data.**
- **Genome-wide sequencing data describe genomic regulatory regions that control gene expression**
- **Gene expression disregulation may be linked to the disease**
- **Understanding molecular mechanisms of disease outcome opens the way to discovery drug and identify biomarkers**

How SNPs play a FUNCTIONAL role in disease:

Impact on transcription

- Changing consensus sequenses for transcription factors binding sites
- Changing interaction between for transcription factors
- Changing epigenetic profiling of specific genomic regions
- Changing long range interaction between two genomic regions

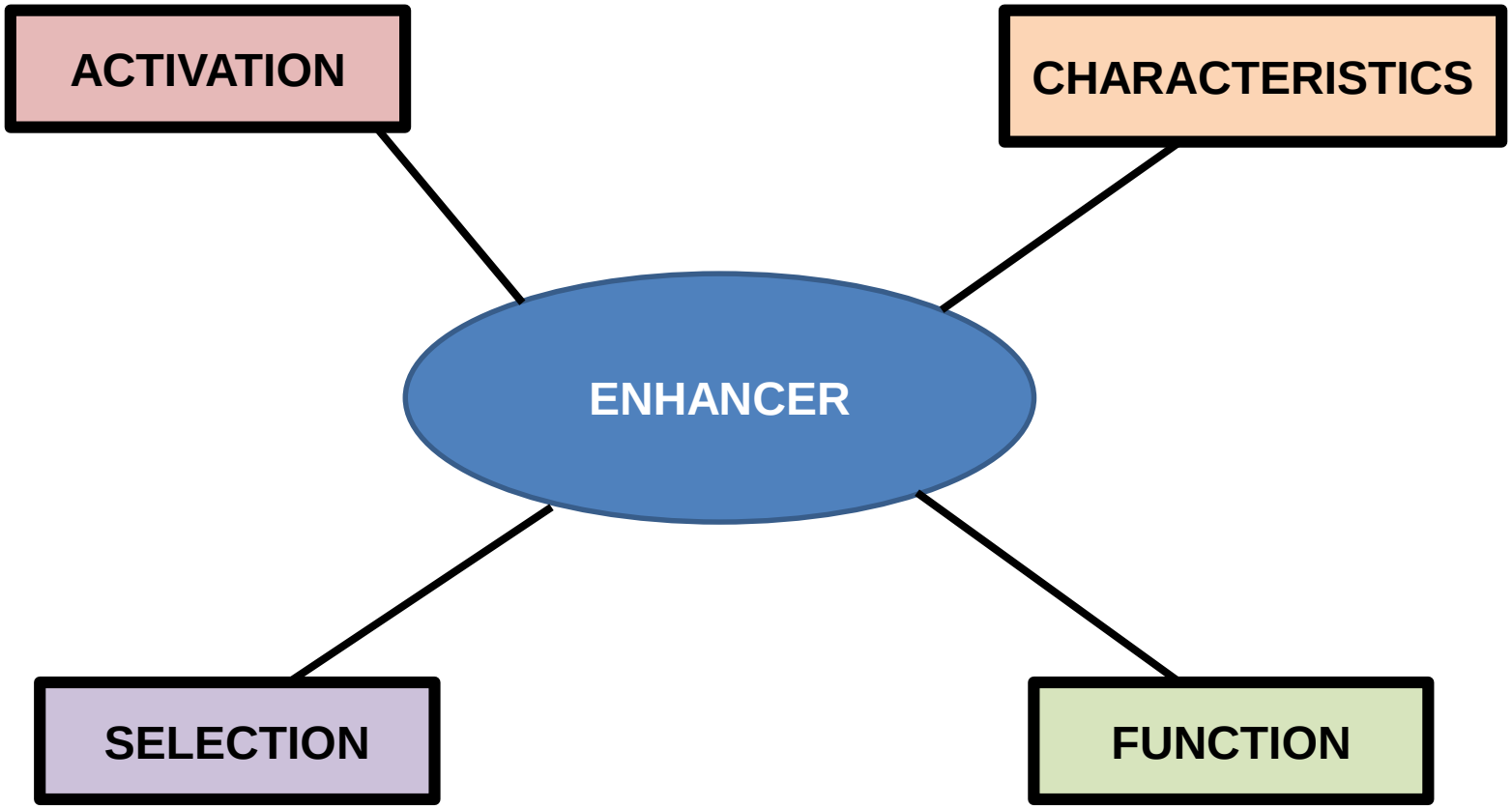
In this Lesson

- Enhancer Overview
- Genomic regulatory network to define cell identity
- Genetic variations meaning in cell identity

The selection and function of cell type-specific enhancers

Sven Heinz[1], Casey E. Romanoski[2], Christopher Benner[1] and Christopher K. Glass[2,3]

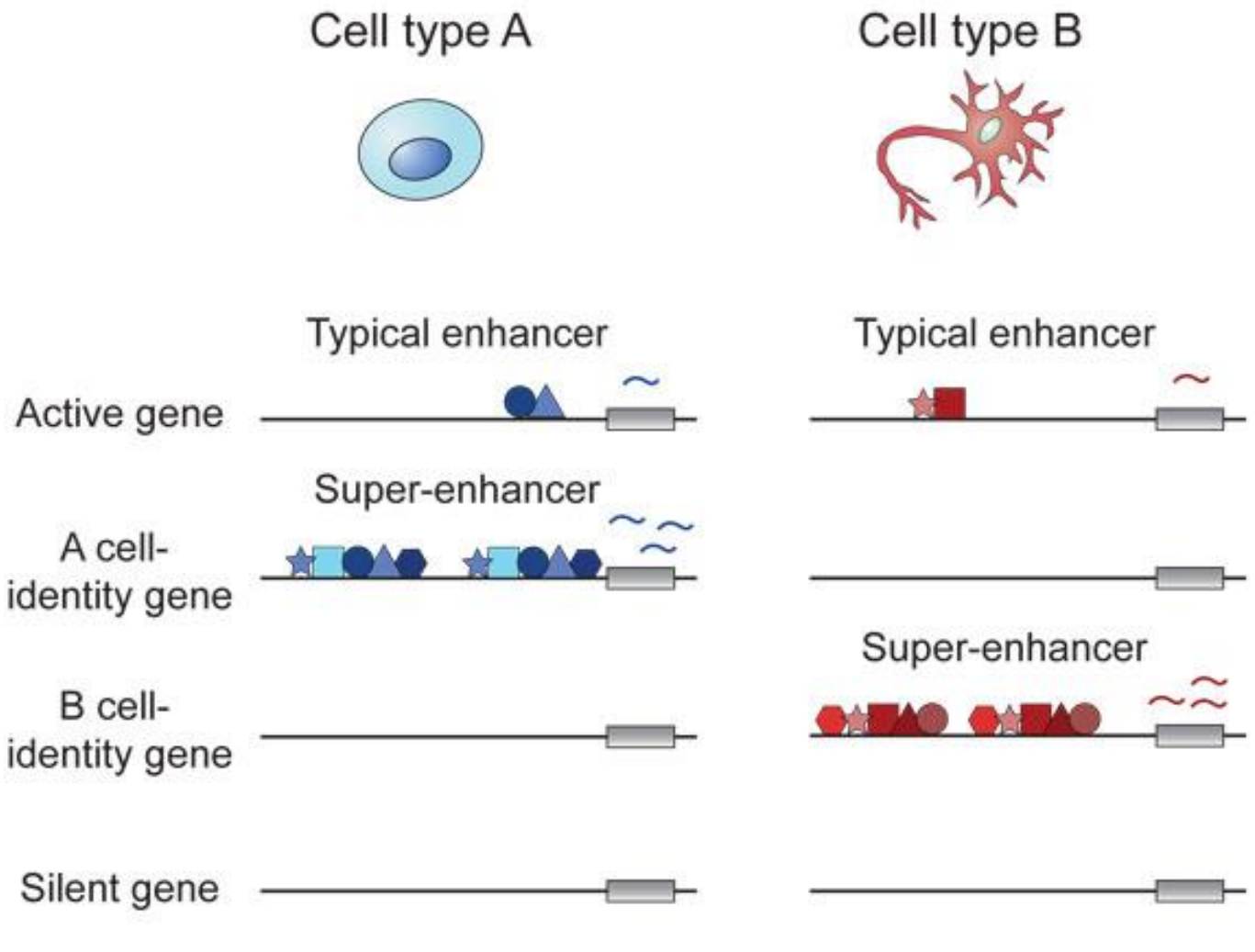
Abstract | The human body contains several hundred cell types, all of which share the same genome. In metazoans, much of the regulatory code that drives cell type-specific gene expression is located in distal elements called enhancers. Although mammalian genomes contain millions of potential enhancers, only a small subset of them is active in a given cell type. Cell type-specific enhancer selection involves the binding of lineage-determining transcription factors that prime enhancers. Signal-dependent transcription factors bind to primed enhancers, which enables these broadly expressed factors to regulate gene expression in a cell type-specific manner. The expression of genes that specify cell type identity and function is associated with densely spaced clusters of active enhancers known as super-enhancers. The functions of enhancers and super-enhancers are influenced by, and affect, higher-order genomic organization.


ACTIVATION
CHARACTERISTICS
ENHANCER
SELECTION
FUNCTION

Enhancer Characteristics

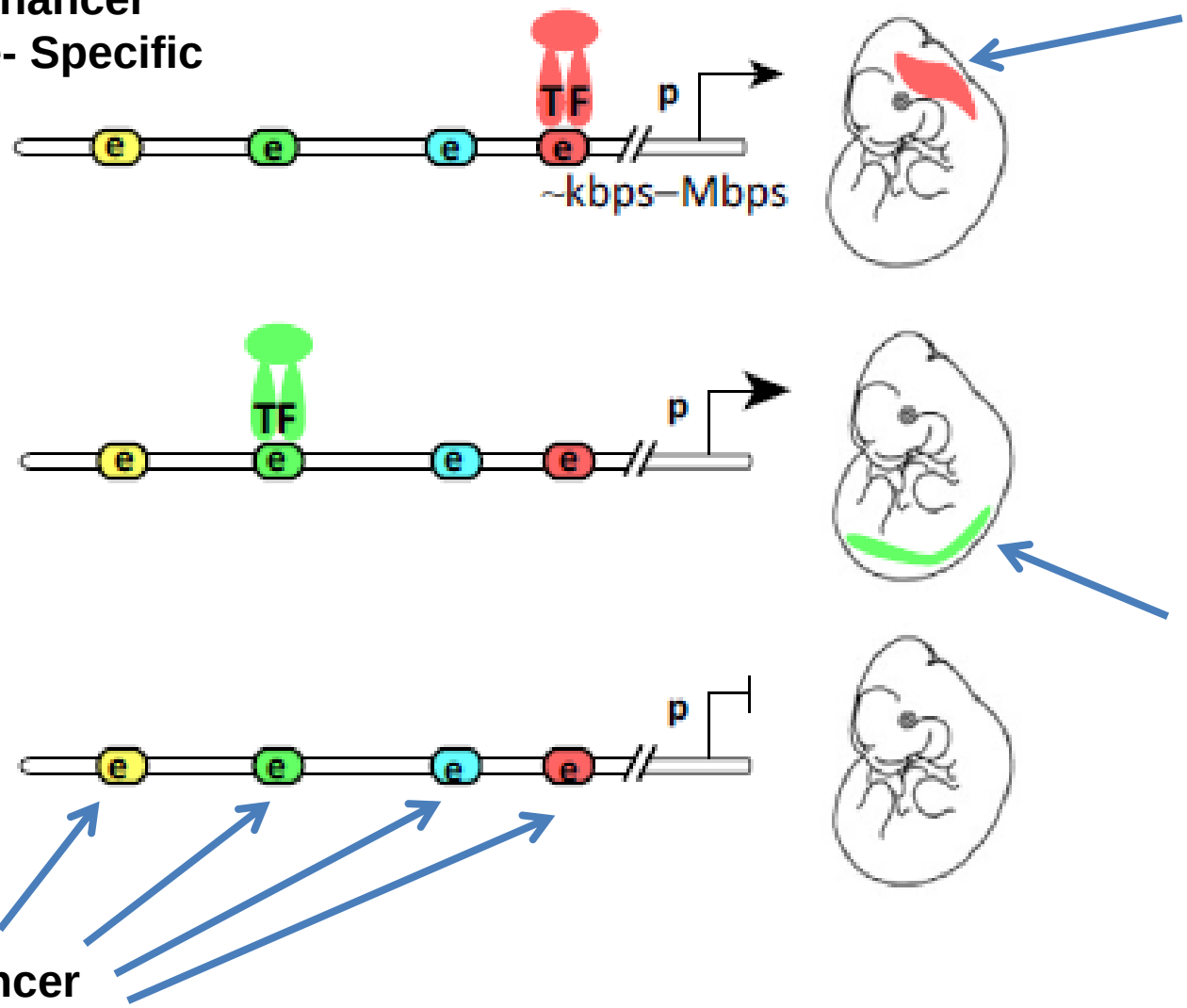
- Enhancers are regulatory elements in proximity of genes
- Each cell has a set of enhancers
- Enhancers have motifs for sequence-specific transcription factors
- Enhancers are marked with epigenetic modifications
- Enhancers are in different states of activation

MUCH OF THE REGULATORY CODE THAT DRIVES CELL-TYPE-SPECIFIC GENE EXPRESSION IS LOCATED IN DISTAL ELEMENTS CALLED ENHANCERS



CELL TYPE USE A SMALL SUBSET OF MILLIONS OF POTENTIAL ENHANCERS

Active Enhancer in Cell-Type- Specific



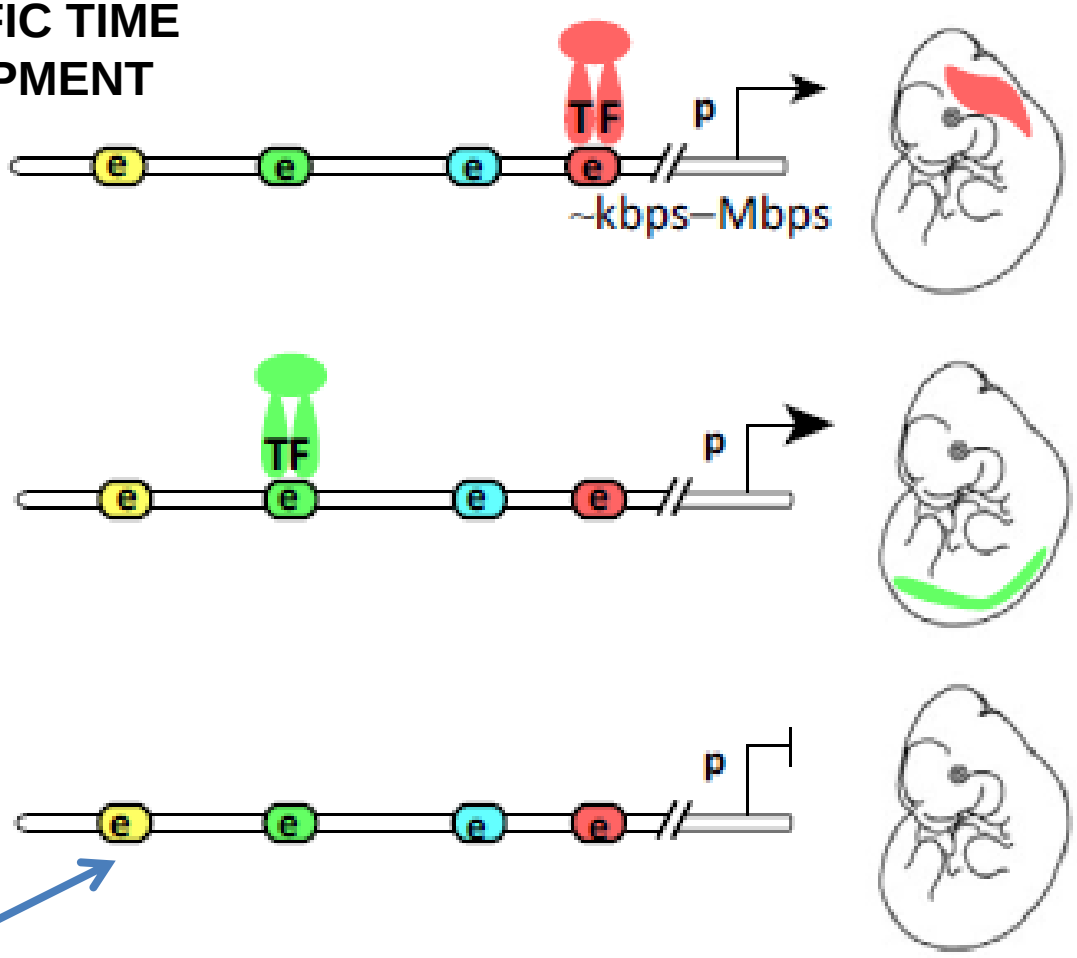
Potential enhancer

Enhancers in tissue/cell-specific gene expression

Trends in Biochemical Sciences April 2014, Vol. 39, No. 4

CELL TYPE USE A SMALL SUBSET OF MILLIONS OF POTENTIAL ENHANCERS

Active Enhancer during SPECIFIC TIME OF DEVELOPMENT

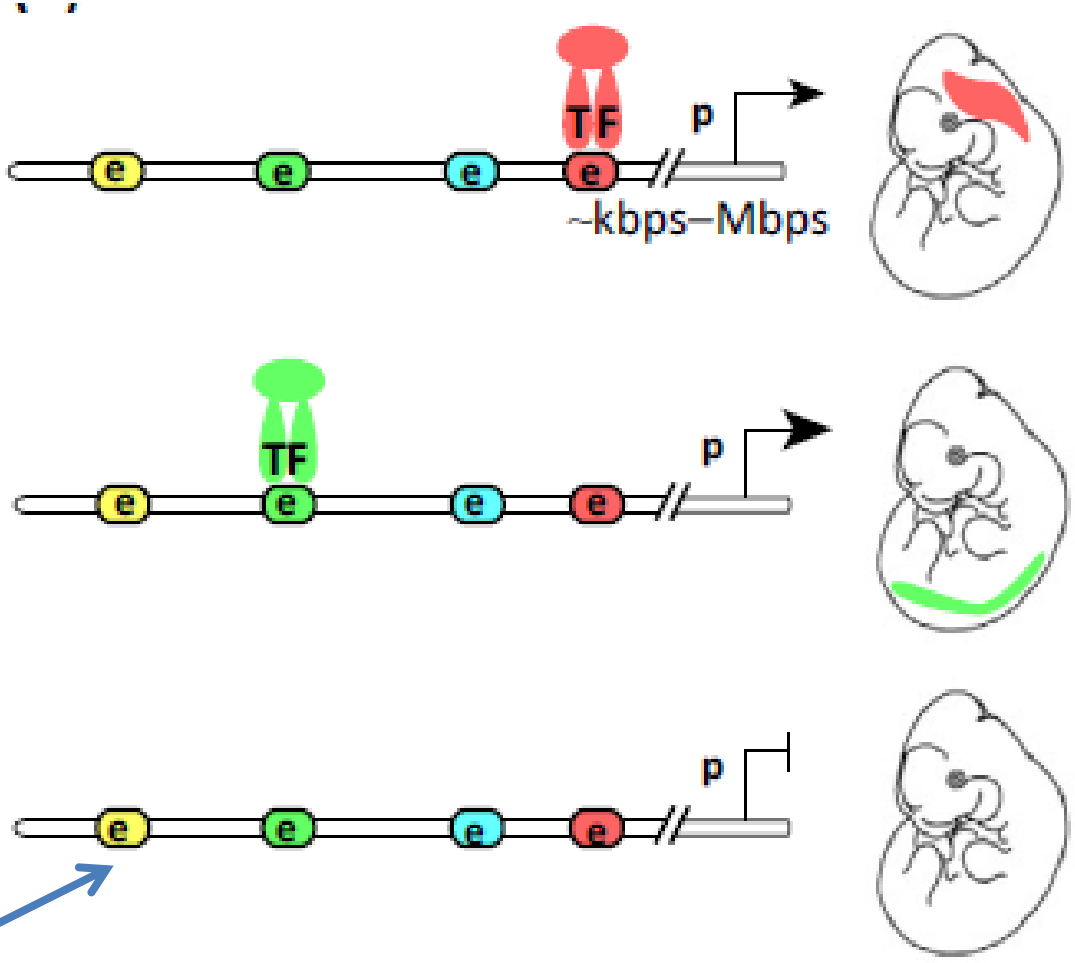


Potential enhancers

Enhancers in tissue/cell-specific gene expression

Trends in Biochemical Sciences April 2014, Vol. 39, No. 4

LINEAGE-DETERMINING TRANSCRIPTION FACTORS BIND AT CELL-TYPE SPECIFIC ENHANCERS



TF RED

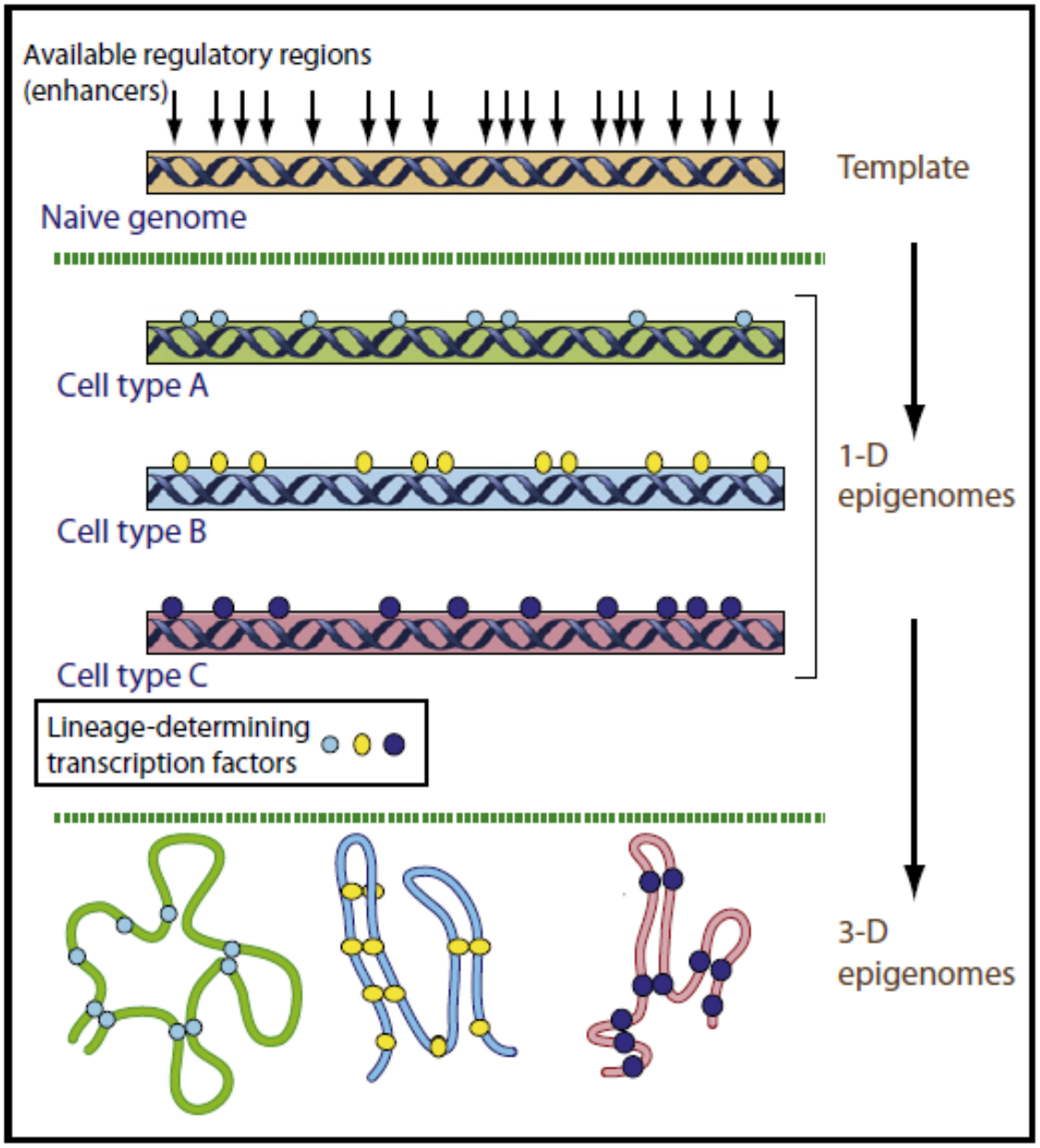
TF GREEN

Potential enhancers

Enhancers in tissue/cell-specific gene expression

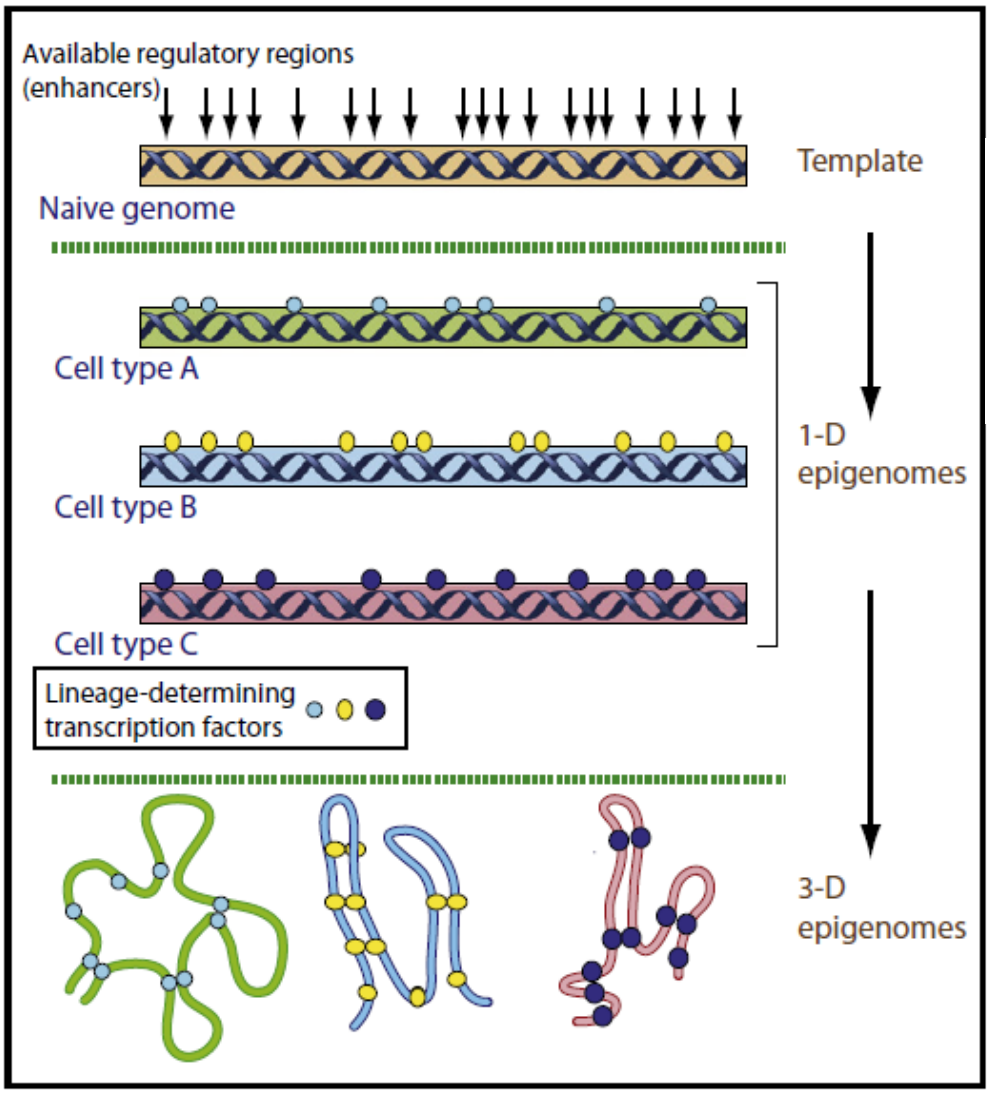
Trends in Biochemical Sciences April 2014, Vol. 39, No. 4

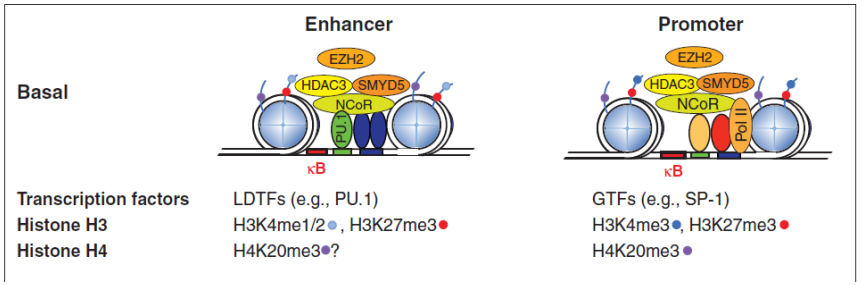
EACH CELL HAS ACTIVE ENHANCERS

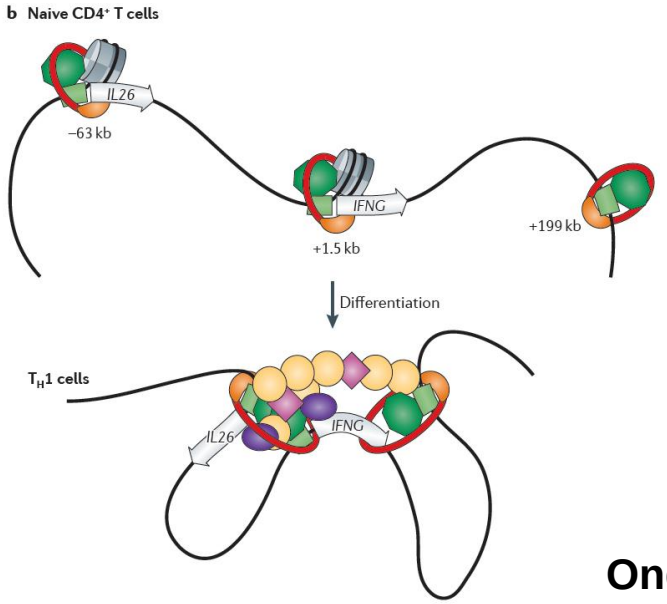


Maintaining Cell Identity through Global Control of Genomic Organization

Gioacchino Natoli[1,*]





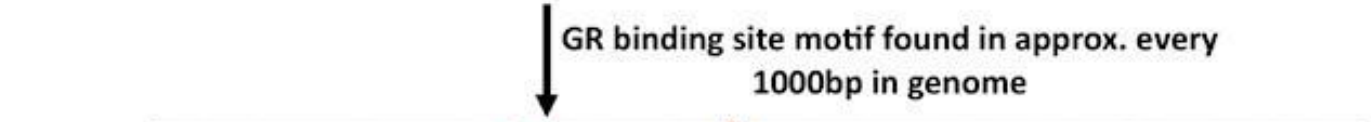


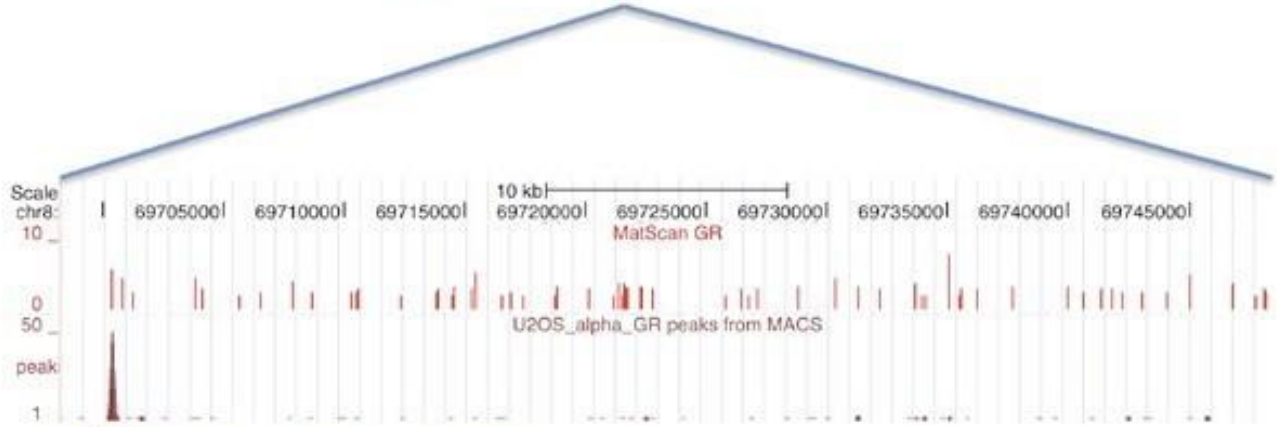
Ong 2011

Immunity 33, July 23, 2010

TRANSCRIPTION FACTOR BINDS SPECIFIC CONSENSUS SEQUENCE IN ACTIVE ENHANCER

GR binding site motif found in approx. every 1000bp in genome

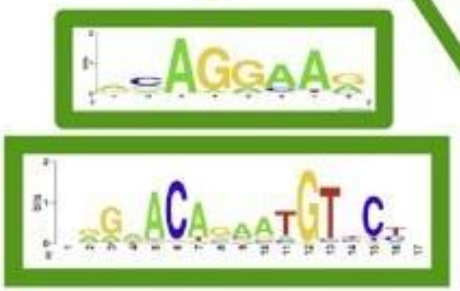


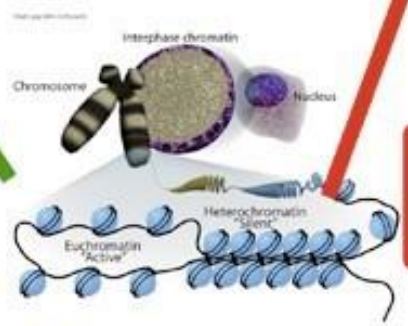


Red line marks GR binding motif:

Peak reflects actual GR Binding event:

Genetic and epigenetic inputs directing binding





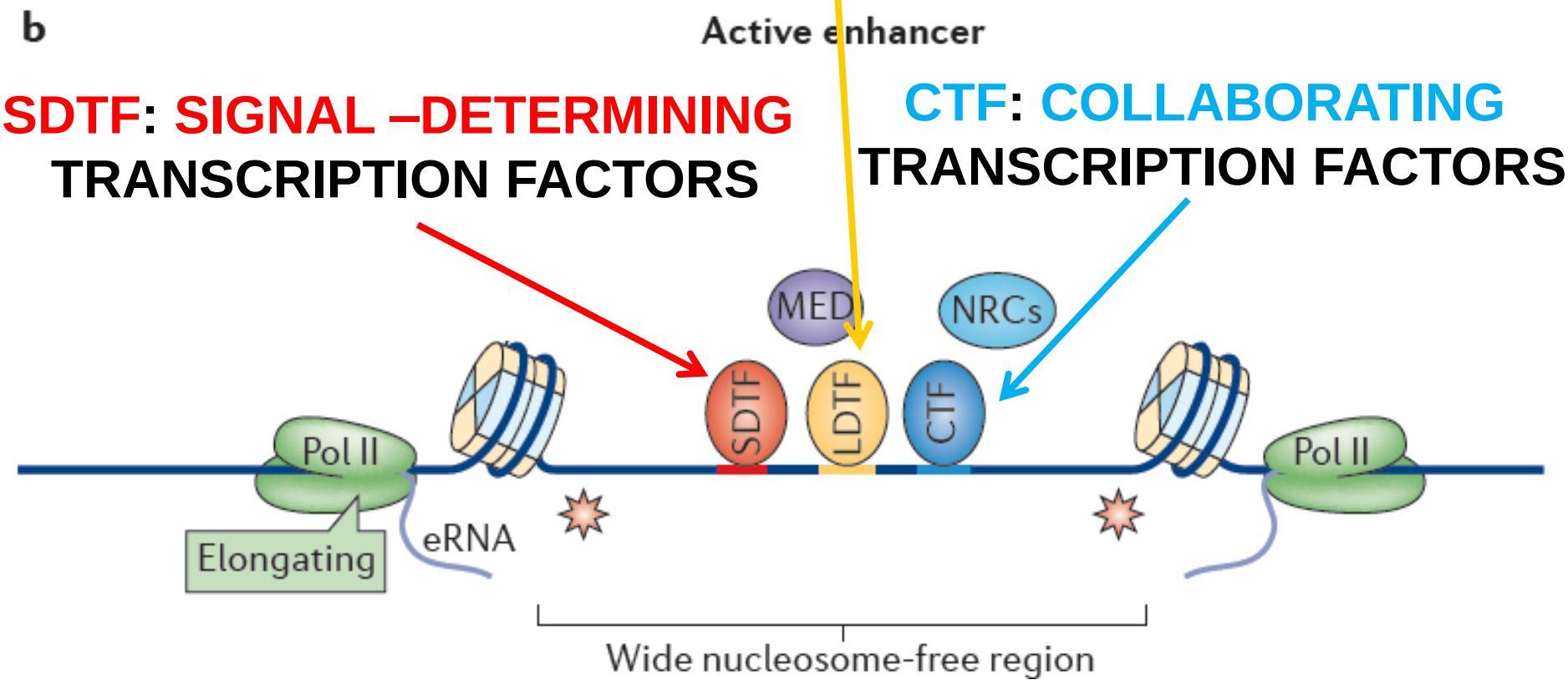
Genetic and epigenetic inputs that restrict binding

TRANSCRIPTION FACTORS THAT BIND ENHANCERS

LDTF: LINEAGE –DETERMINING TRANSCRIPTION FACTORS

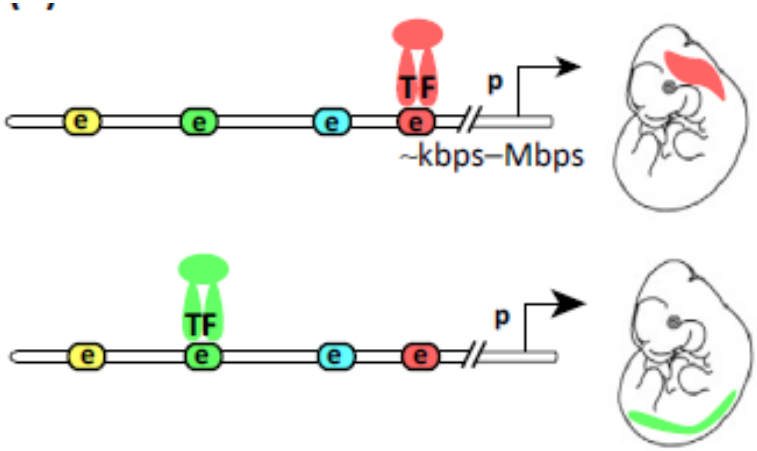
SDTF: SIGNAL –DETERMINING TRANSCRIPTION FACTORS

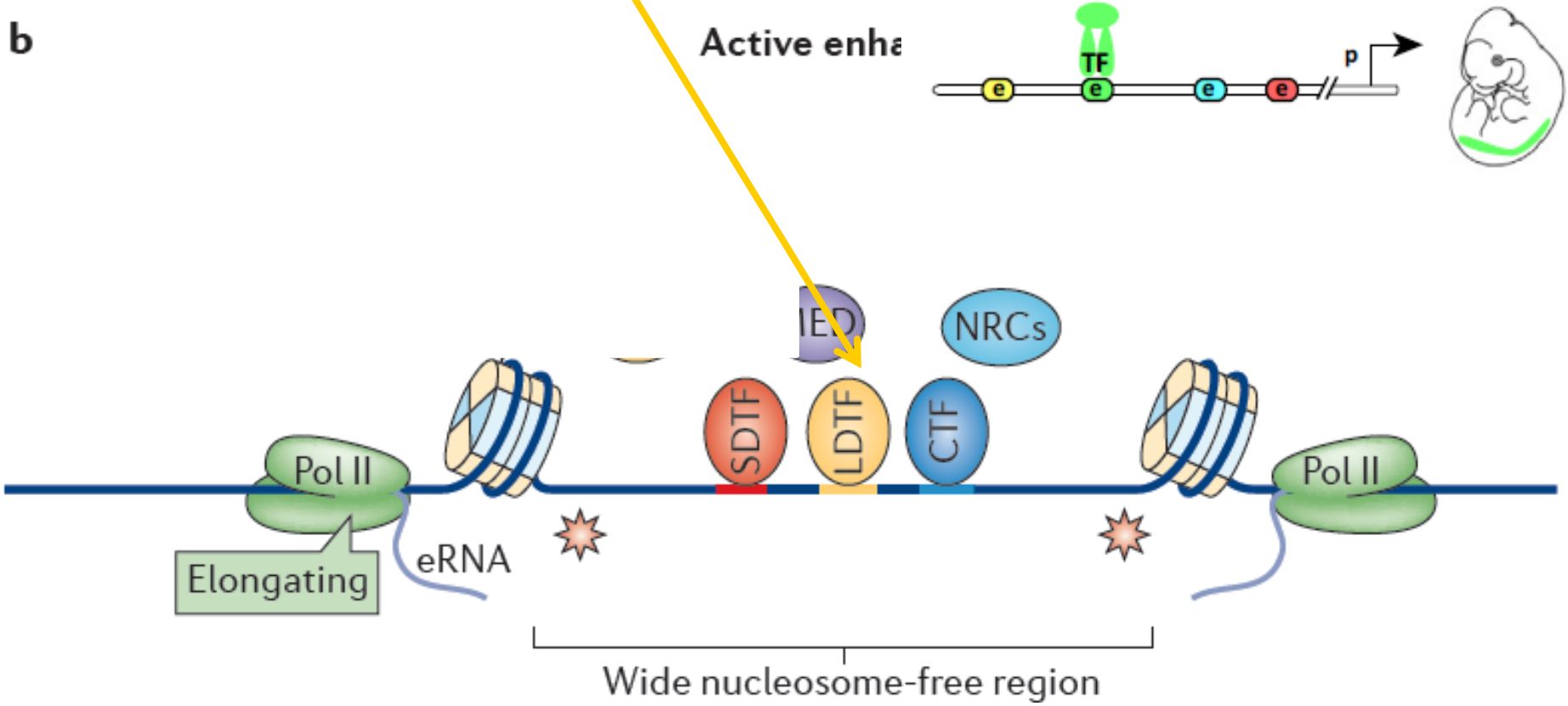
CTF: COLLABORATING TRANSCRIPTION FACTORS



TRANSCRIPTION FACTORS THAT BIND ENHANCERS

LDTF: LINEAGE –DETERMINING TRANSCRIPTION FACTORS



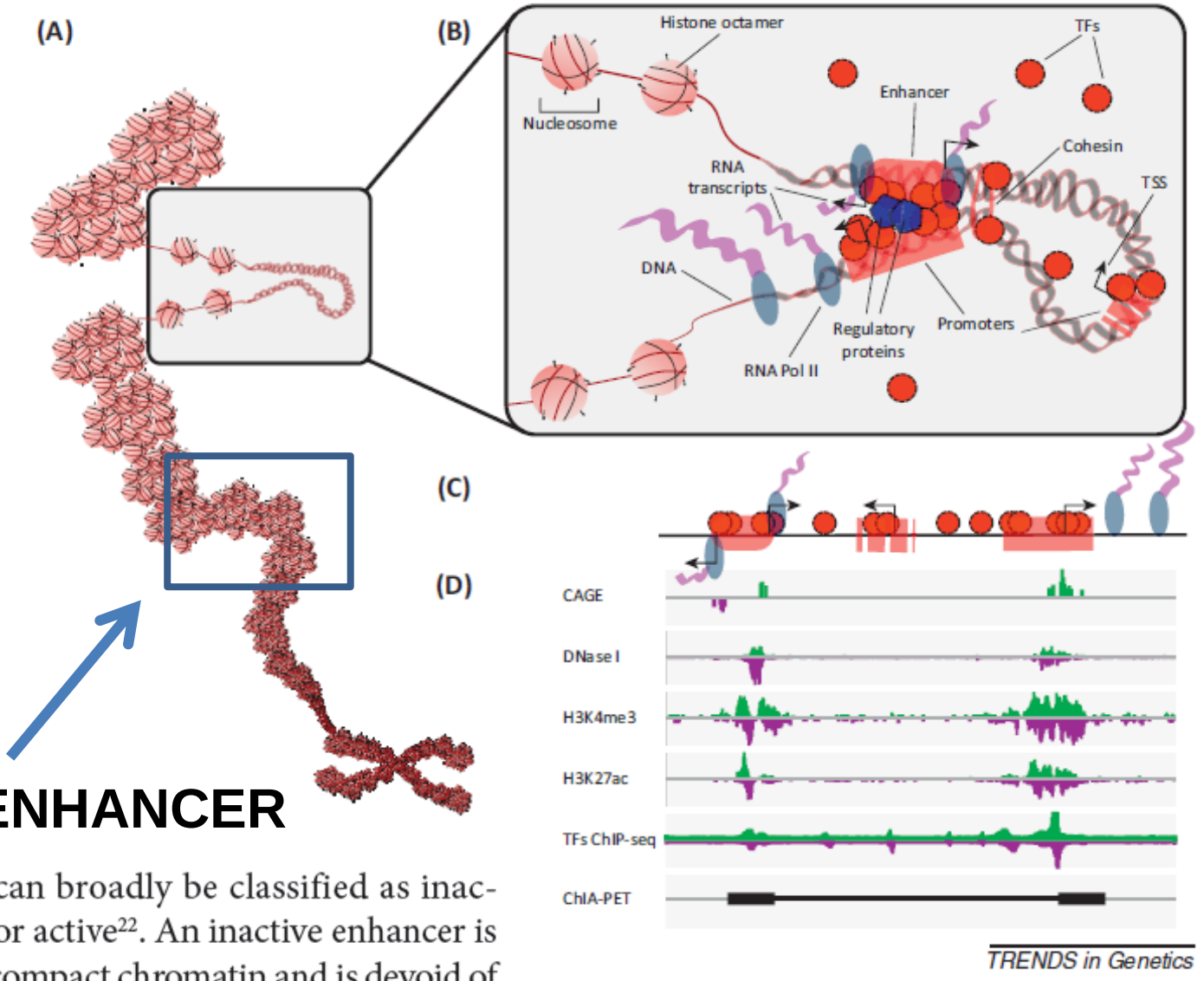


Enhancer Activation

- **Transcription Factors bind specific genomic regions and allow access to other proteins remodelling chromatin**

- **Differentiation states and external stimuli induce enhancers activation**

ACTIVE ENHANCER



INACTIVE ENHANCER

Enhancer states can broadly be classified as inactive, primed, poised or active[22]. An inactive enhancer is essentially buried in compact chromatin and is devoid of transcription factor binding and histone modifications.

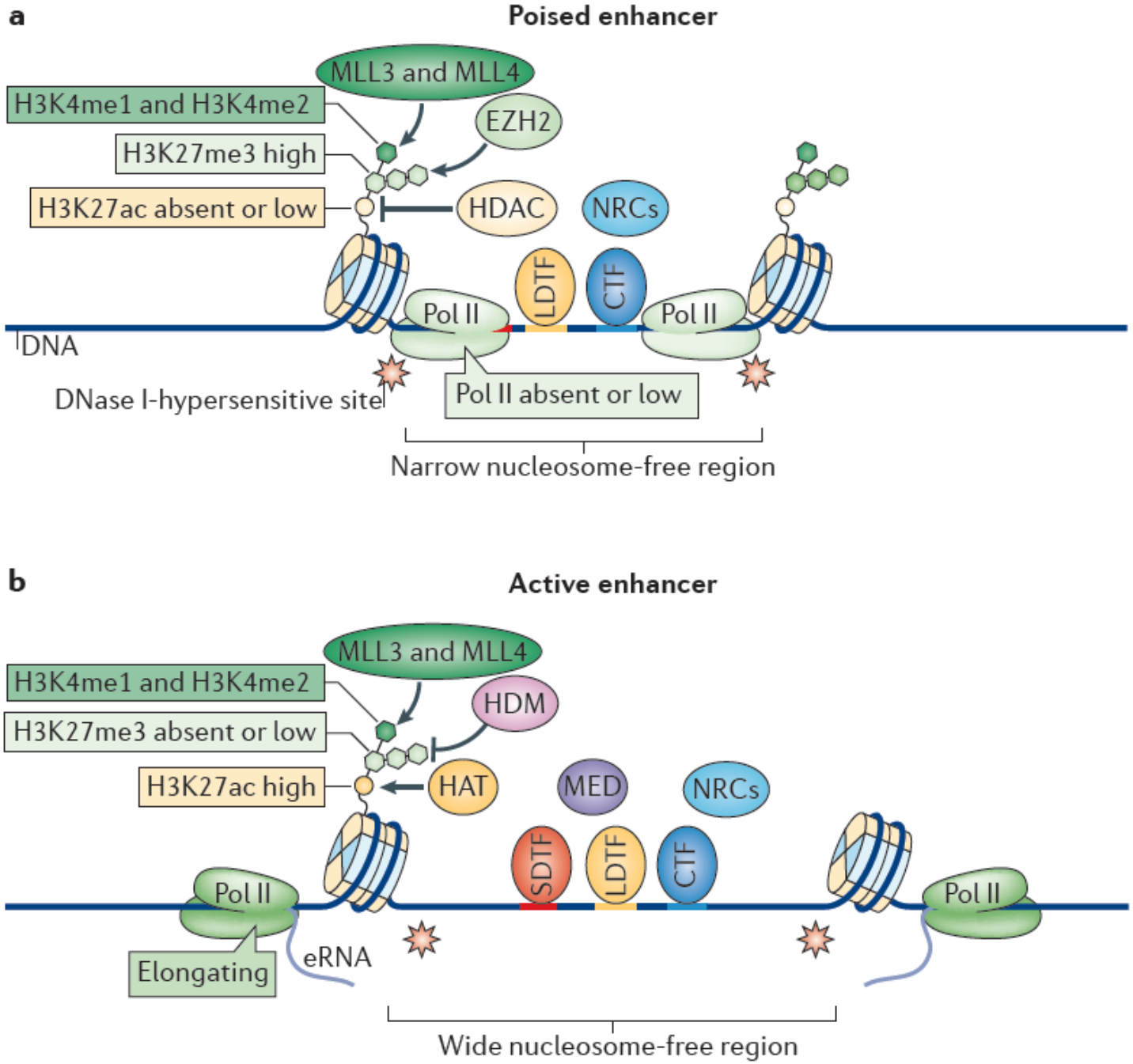
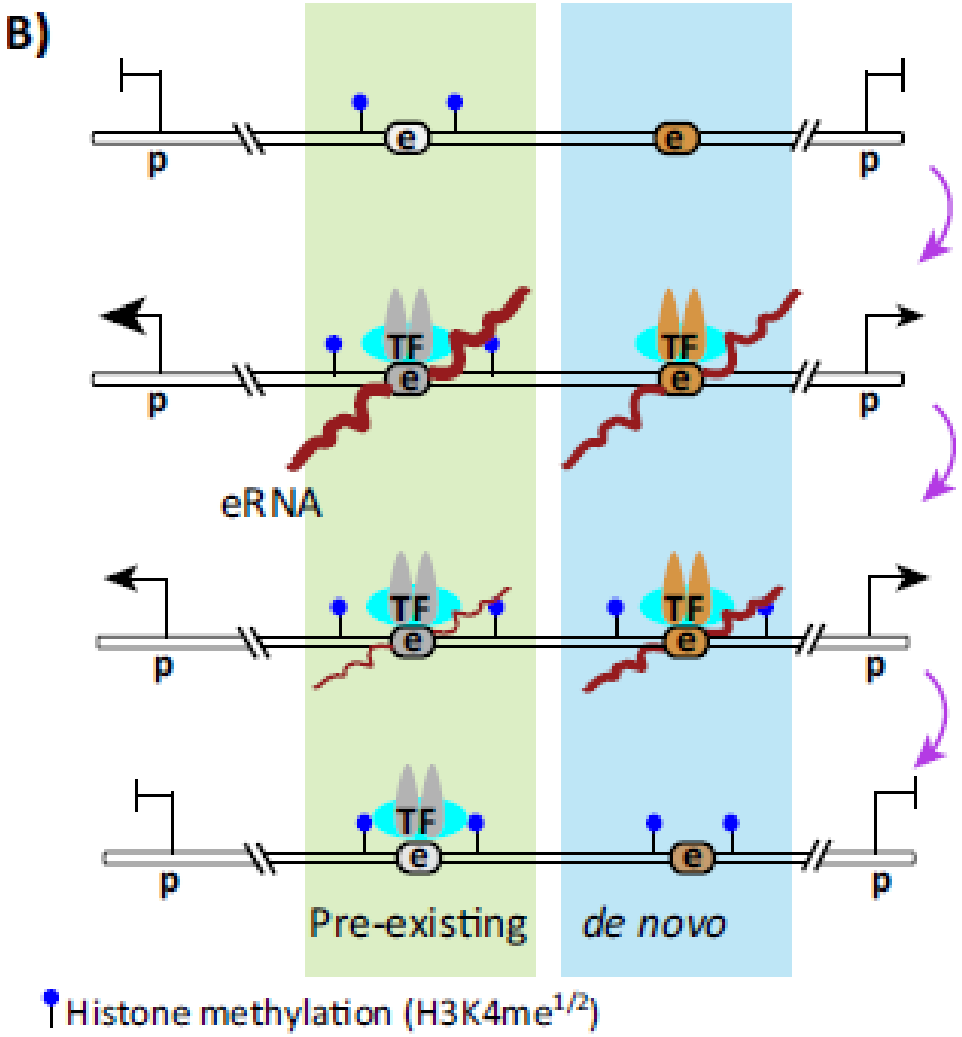


Figure 1 | **The anatomies of poised and active enhancers.** The characteristic features

LDTF: LINEAGE –DETERMINING TRANSCRIPTION FACTORS



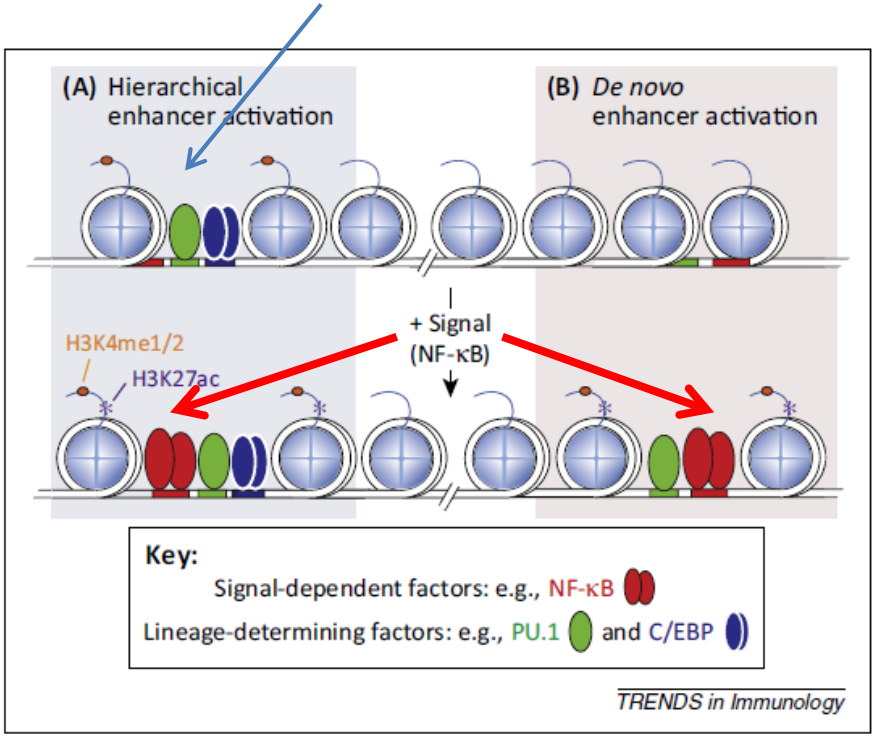
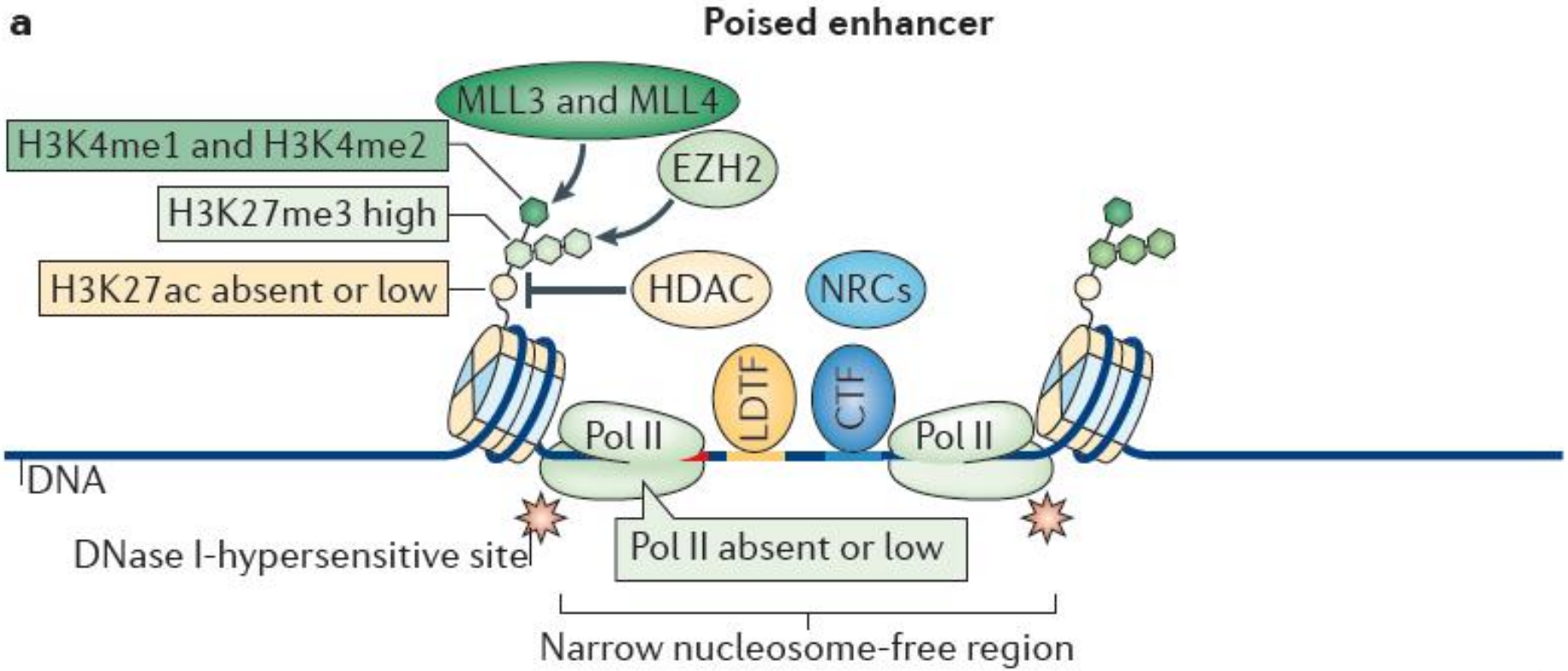


Figure 3. Chromatin transitions to active enhancers involve interactions between cell lineage-determining transcription factors and signal-dependent factors. **(A)** Enhancers primed by lineage-determining factors frequently require signal-dependent transcription factor binding to gain H3K27ac and become active. **(B)** Active enhancers can also be selected by interactions between signal-dependent factors and lineage-determining factors. Abbreviaions: C/EBP, CCAAT/enhancer binding protein; NF-κB, nuclear factor-κB; PU.1, transcription factor originally named spleen focus forming virus (SFFV) proviral integration oncogene.

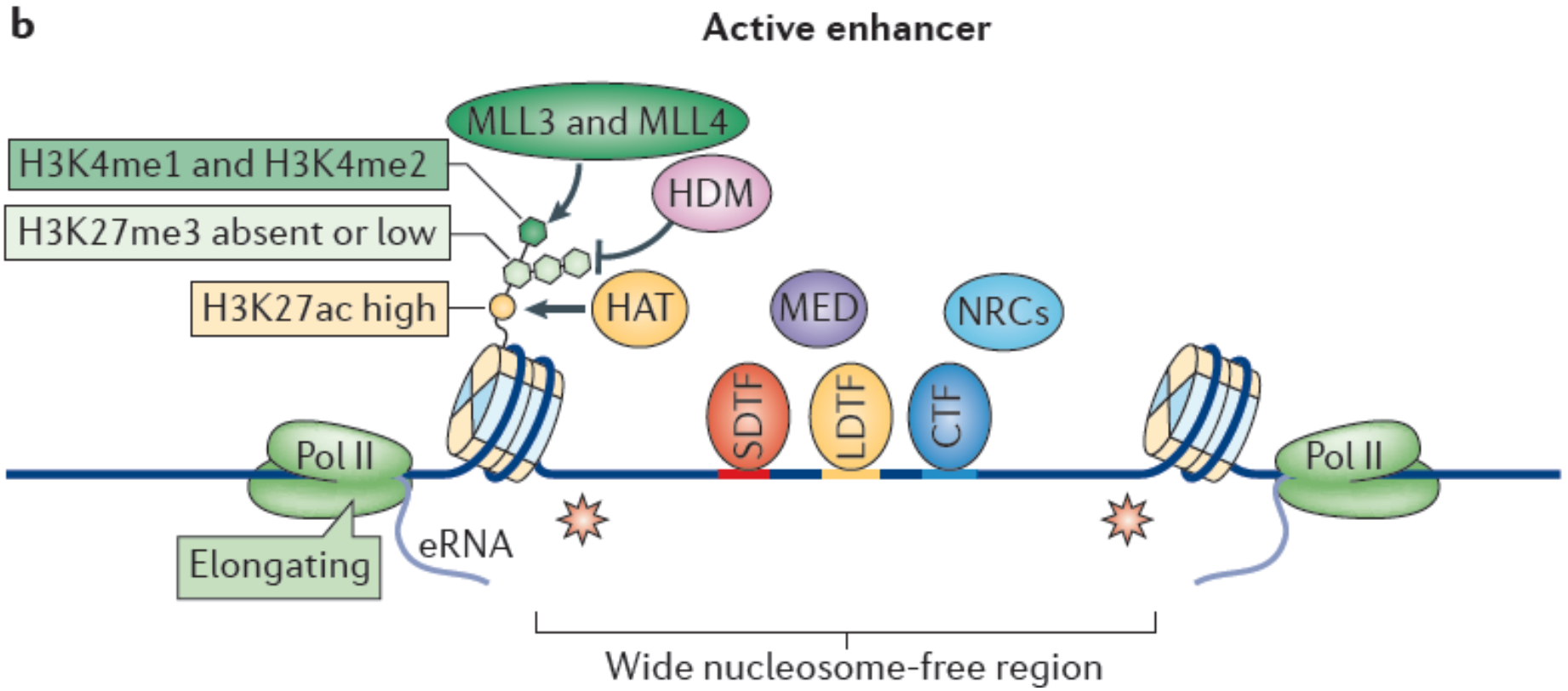
Enhancers in stimulus-induced gene activation

H3K4me1, H3K4me2, lack histone acetylation and Pol II, high H3K27me3

mark POISED ENHANCERS

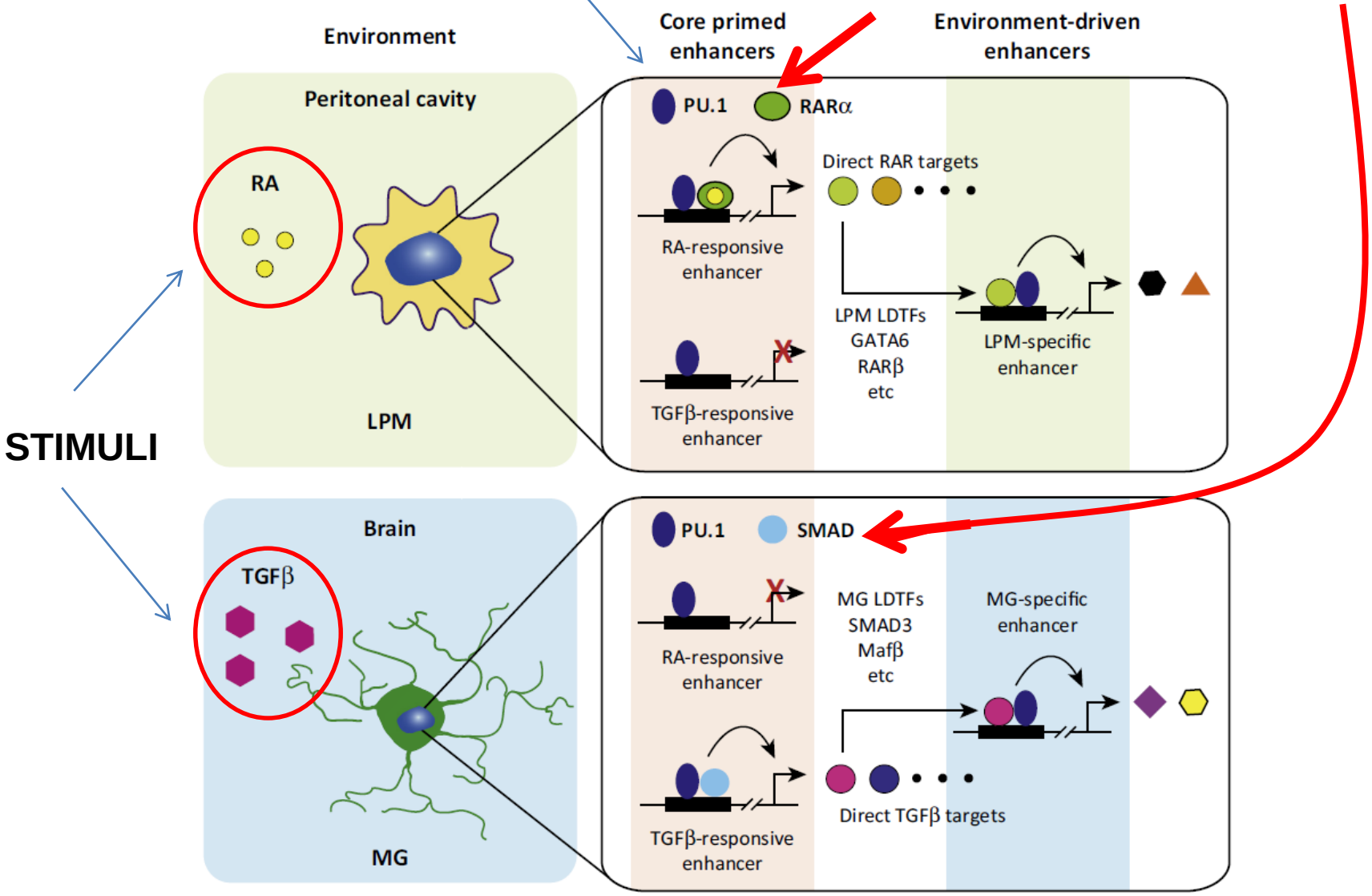


H3K4me1, H3K4me2, high H3K27Ac, lack H3K27me3, presence of Pol II and RNA transcript mark ACTIVE ENHANCERS



LDTF: LINEAGE –DETERMINING TRANSCRIPTION FACTORS

SDTF: SIGNAL –DETERMINING TRANSCRIPTION FACTORS



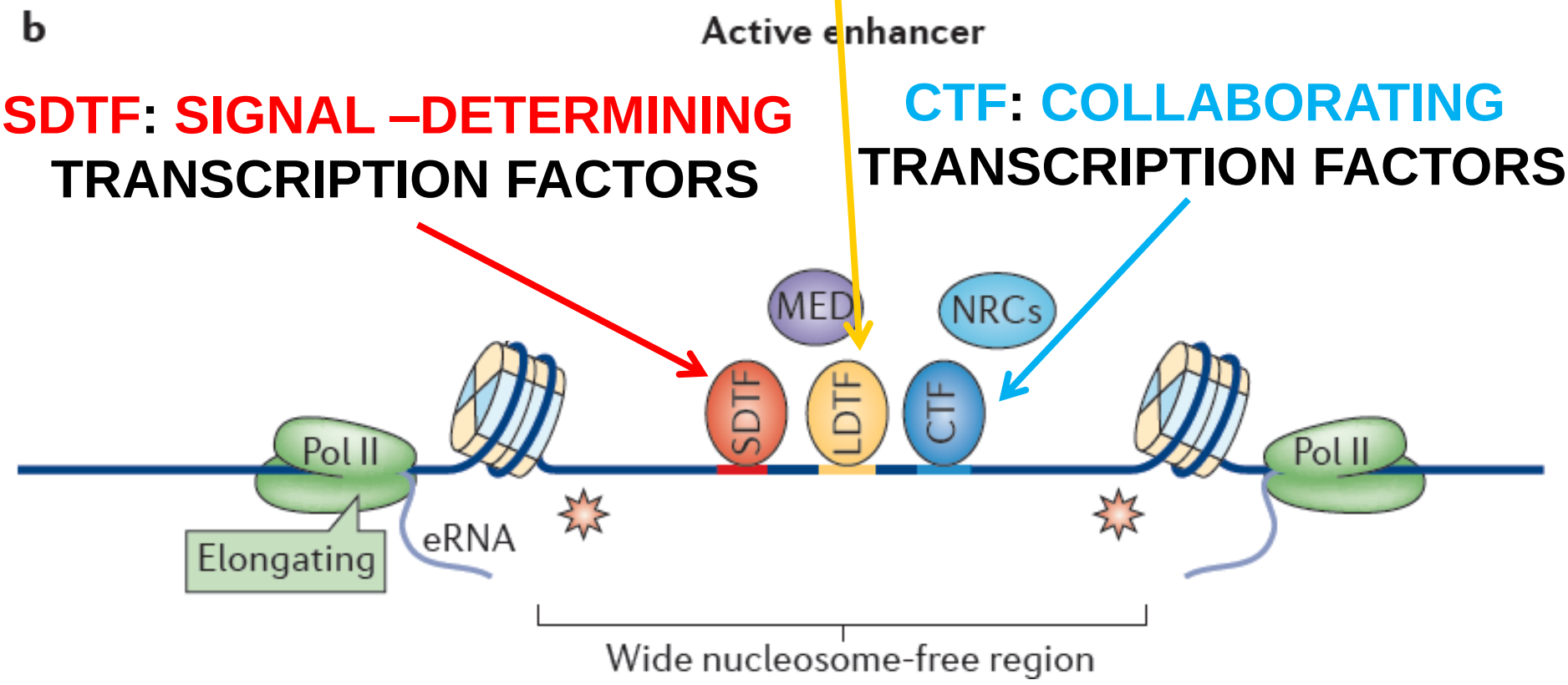
Romanoski et al., 2015

Trends in Immunology September 2015, Vol. 36, No. 9

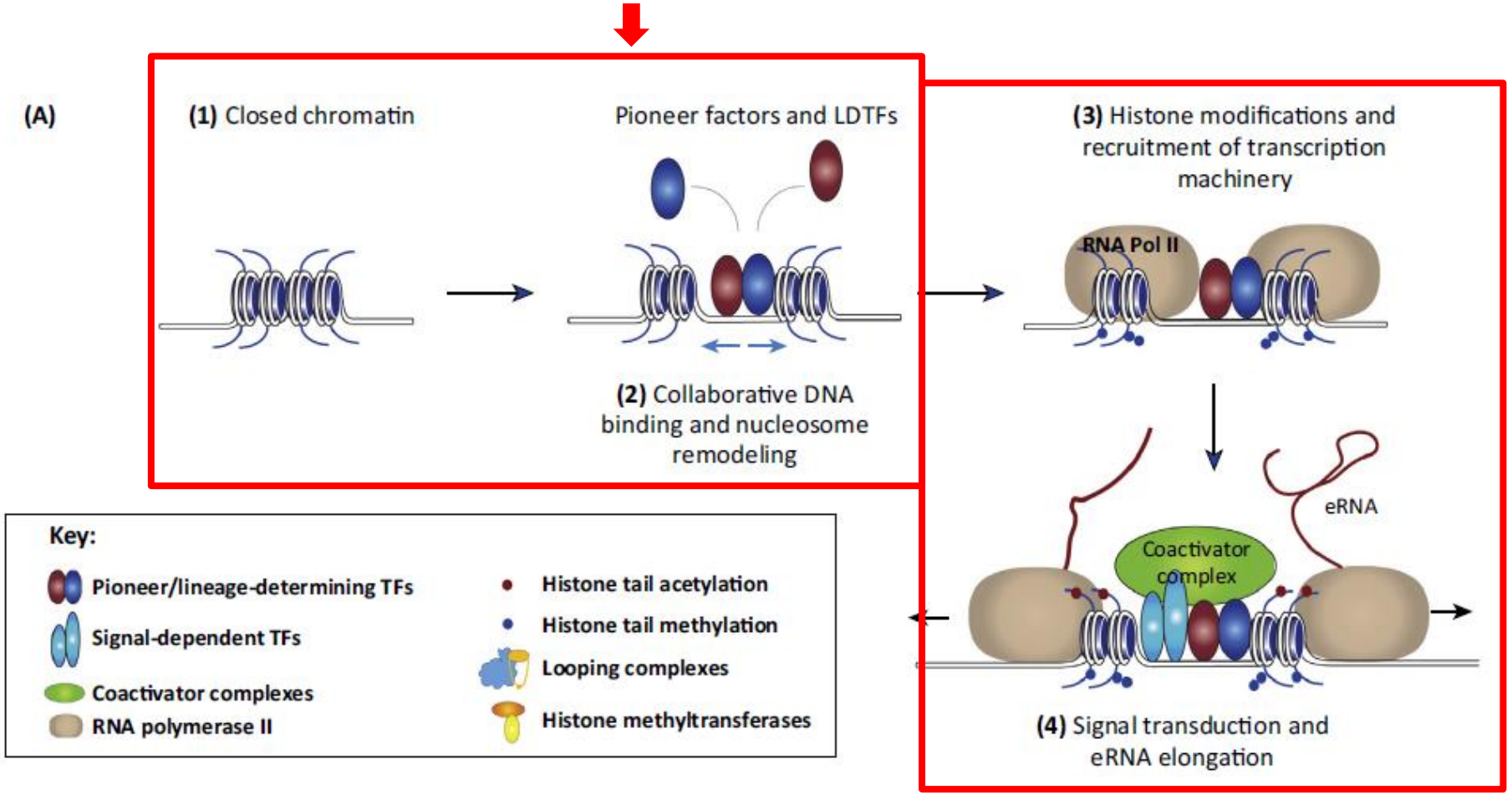
Enhancer Selection

- **The role of lineage-determining transcription factors.**
- **The role of signal-dependent transcription factors.**

LDTF: LINEAGE –DETERMINING TRANSCRIPTION FACTORS



Pioneer Factors and Lineage-determining Transcription Factors leads to nucleosome remodeling and increased chromatin accessibility



Pioneer Factors and Lineage-determining Transcription Factors leads to histone modifications and basal transcription machinery

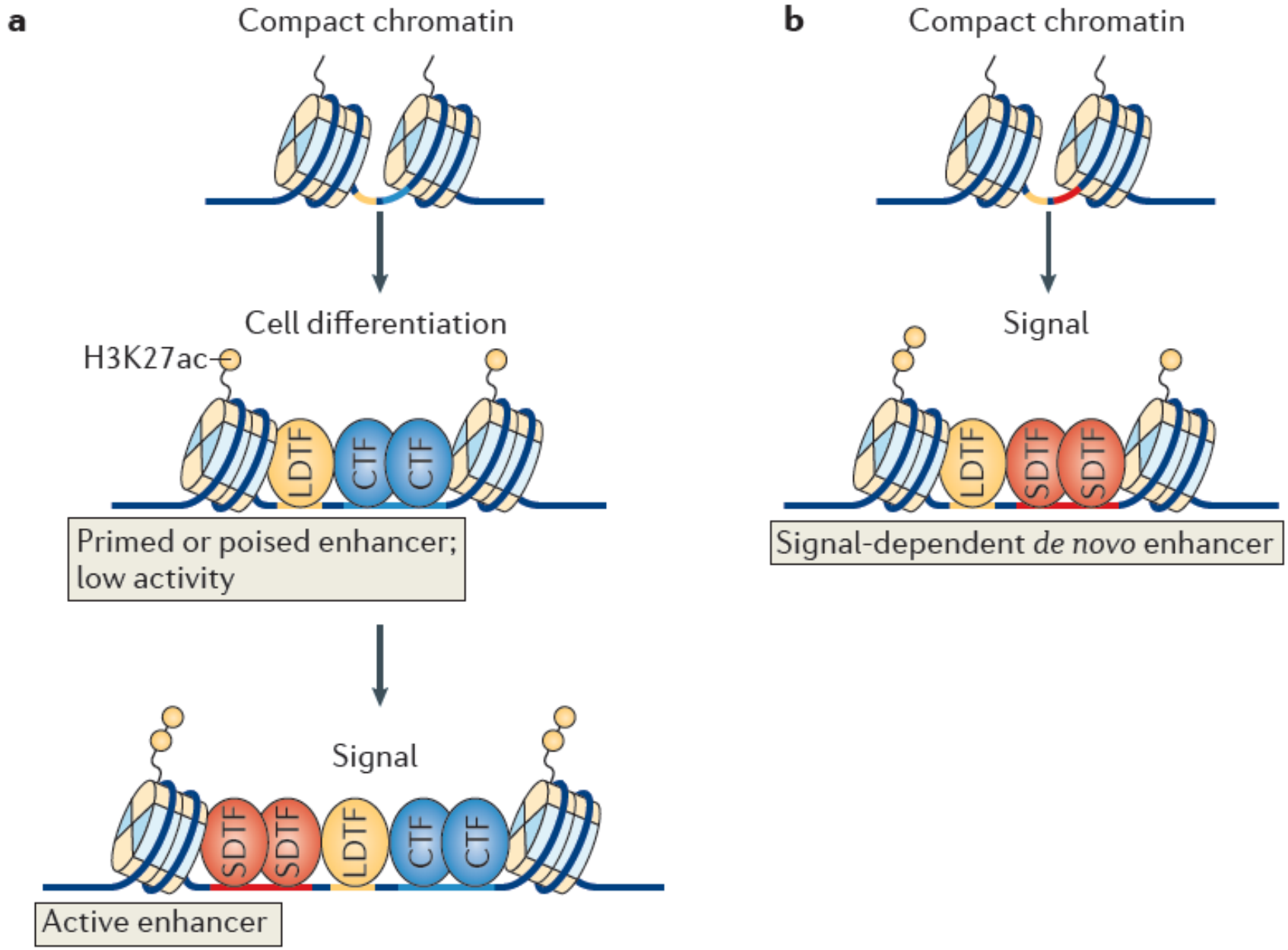
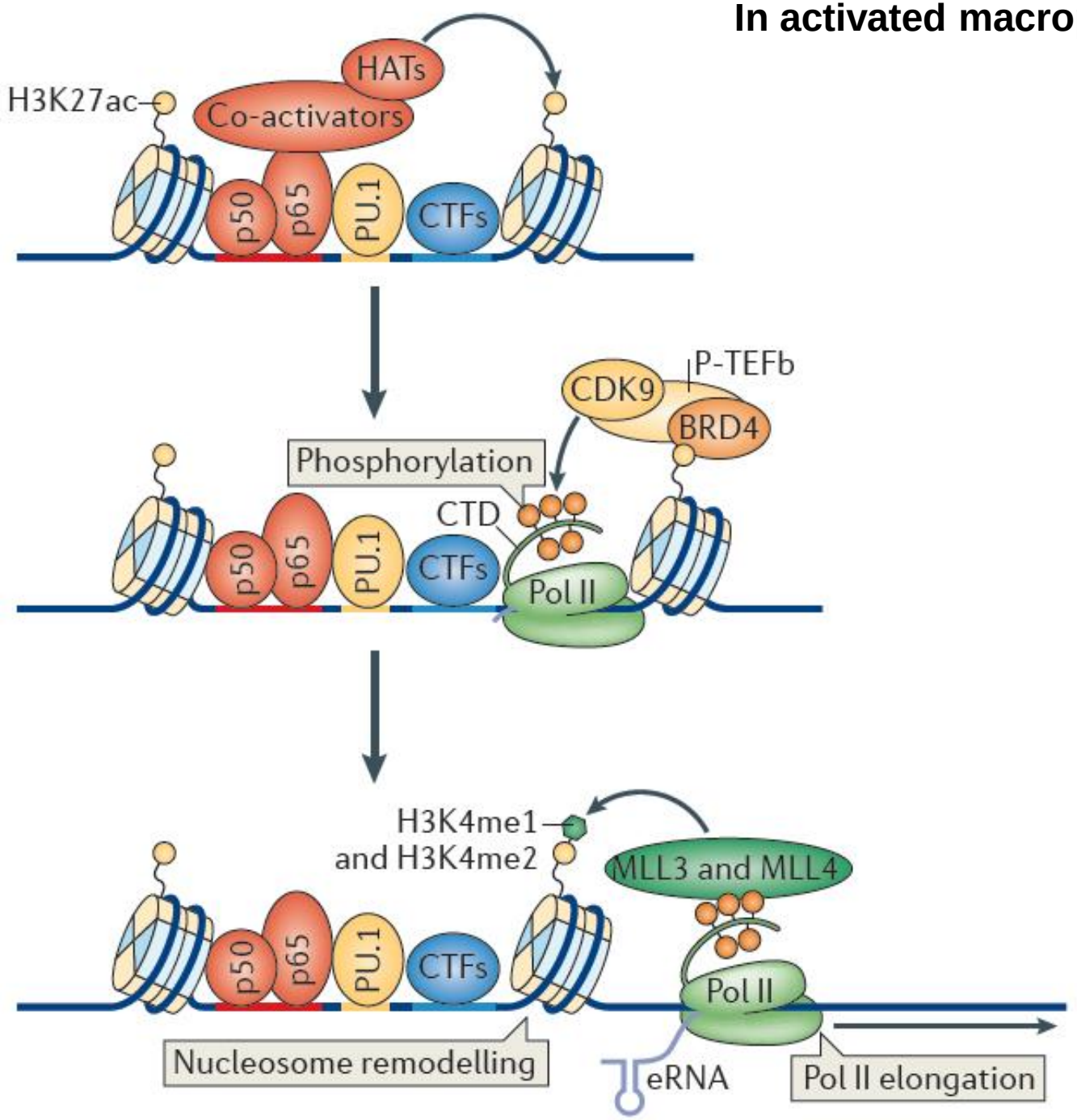


Figure 3 | **Cell type-specific enhancer selection and activation.** **a** | Collaborative

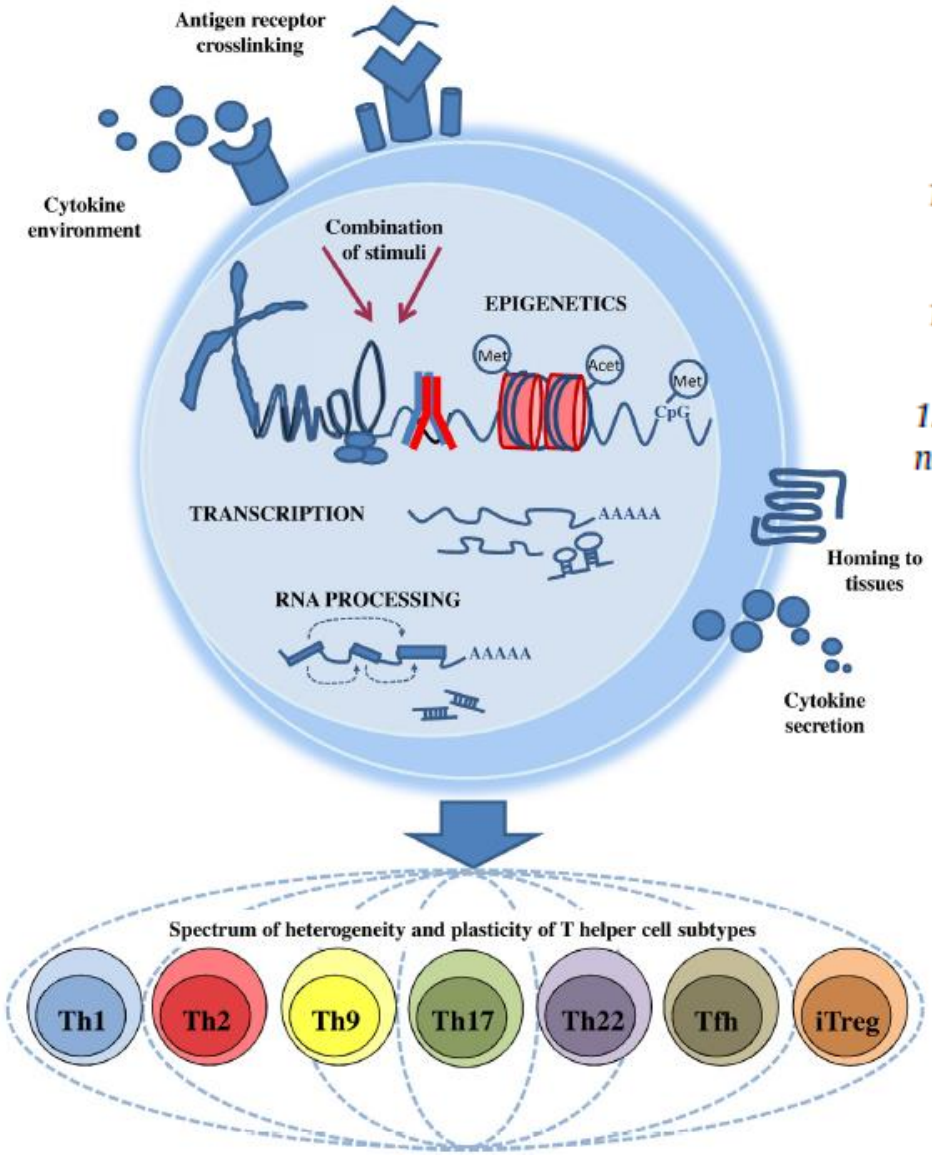
In activated macrophages

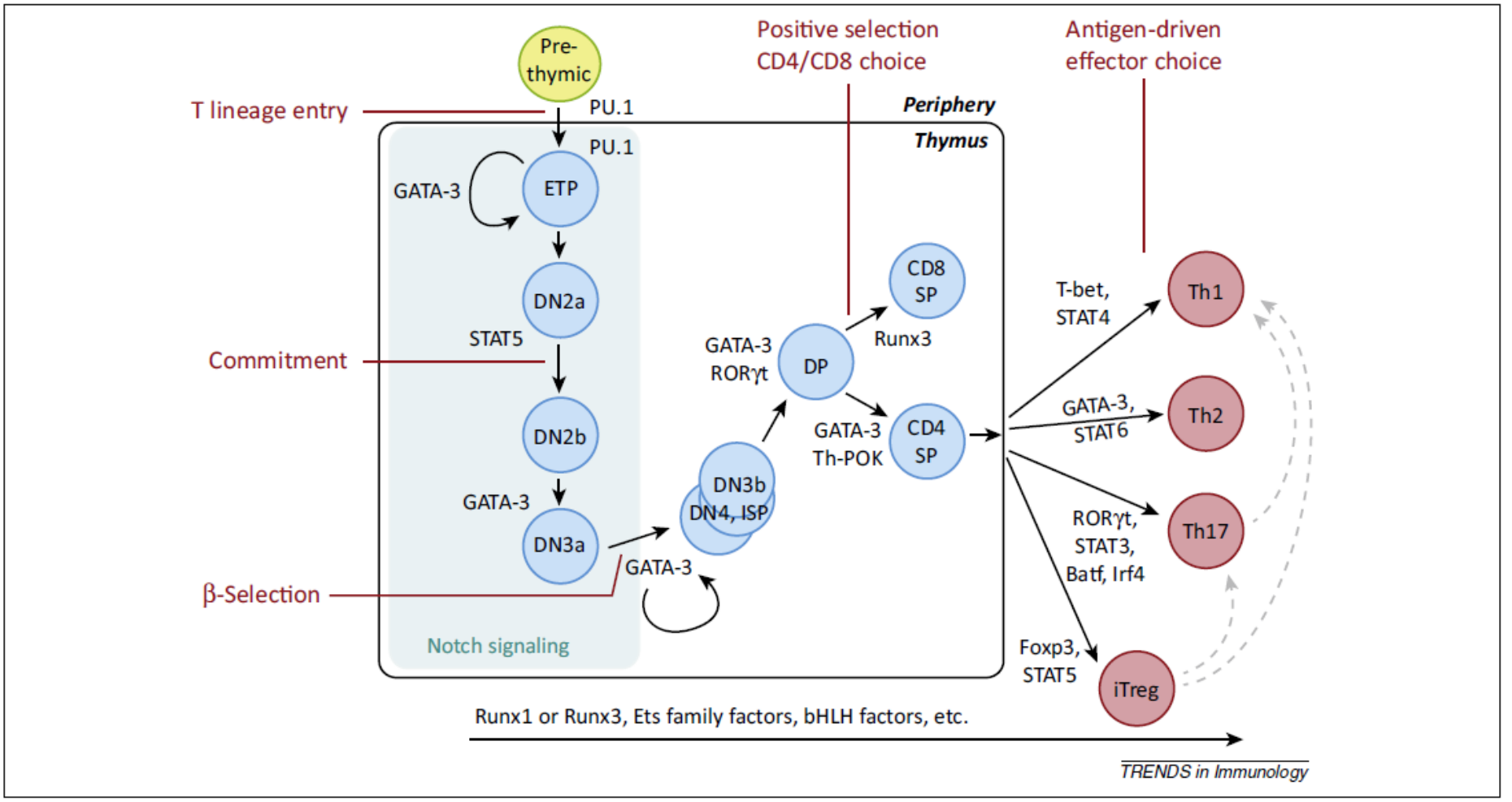


Review

Early T helper cell programming of gene expression in human

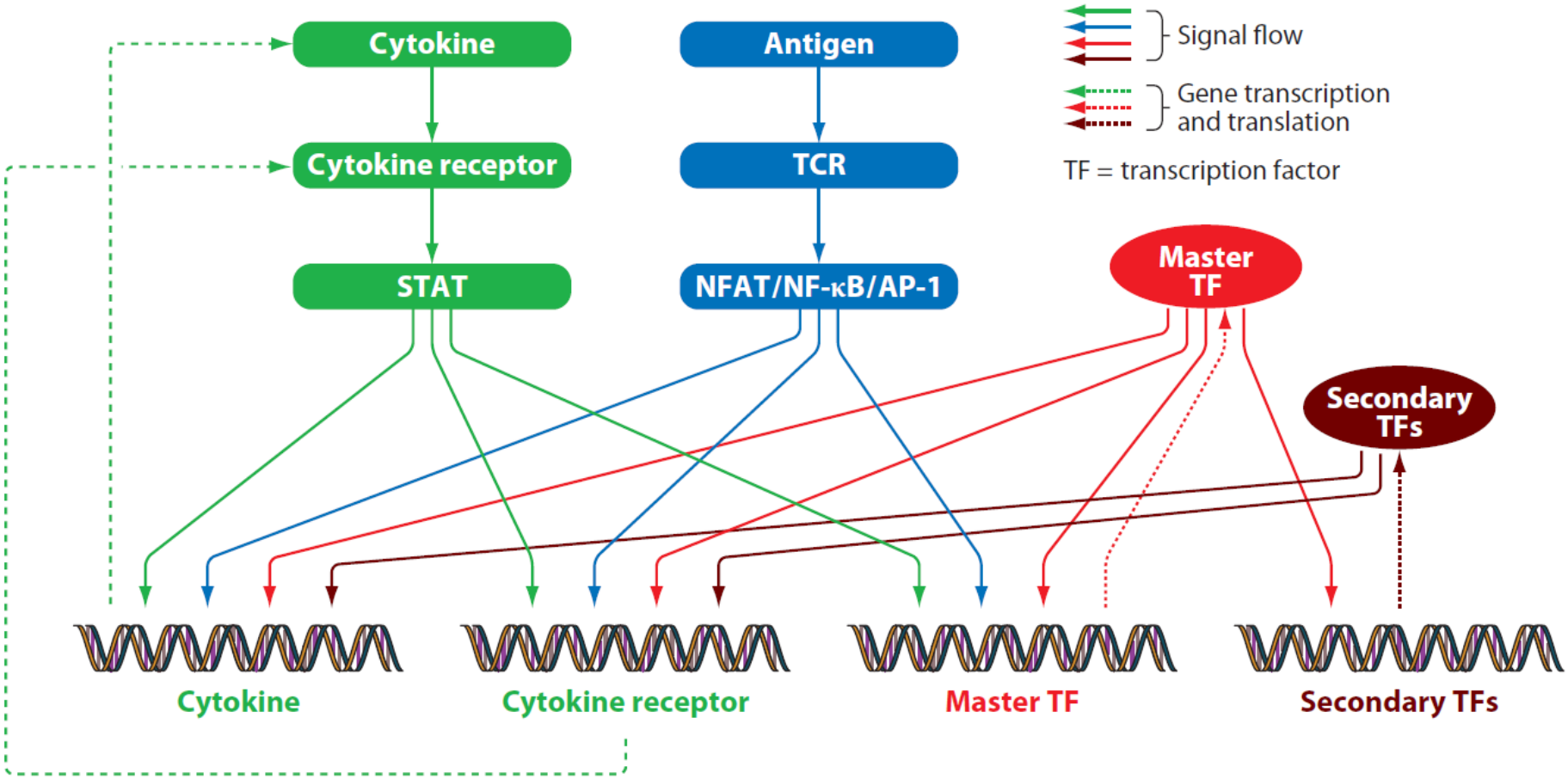
Soile Tuomela, Riitta Lahesmaa*




Pre-thymic
T lineage entry
PU.1
Positive selection
CD4/CD8 choice
Antigen-driven
effector choice
Periphery
Thymus
PU.1
GATA-3
ETP
DN2a
STAT5
Commitment
DN2b
GATA-3
DN3a
β-Selection
DN3b
DN4, ISP
GATA-3
Notch signaling
GATA-3
RORγt
DP
CD8
SP
Runx3
GATA-3
Th-POK
CD4
SP
T-bet,
STAT4
Th1
GATA-3,
STAT6
Th2
RORγt,
STAT3,
Batf, Irf4
Th17
Foxp3,
STAT5
iTreg
Runx1 or Runx3, Ets family factors, bHLH factors, etc.
TRENDS in Immunology

Differentiation of Effector CD4 T Cell Populations*

Jinfang Zhu, Hidehiro Yamane, and William E. Paul



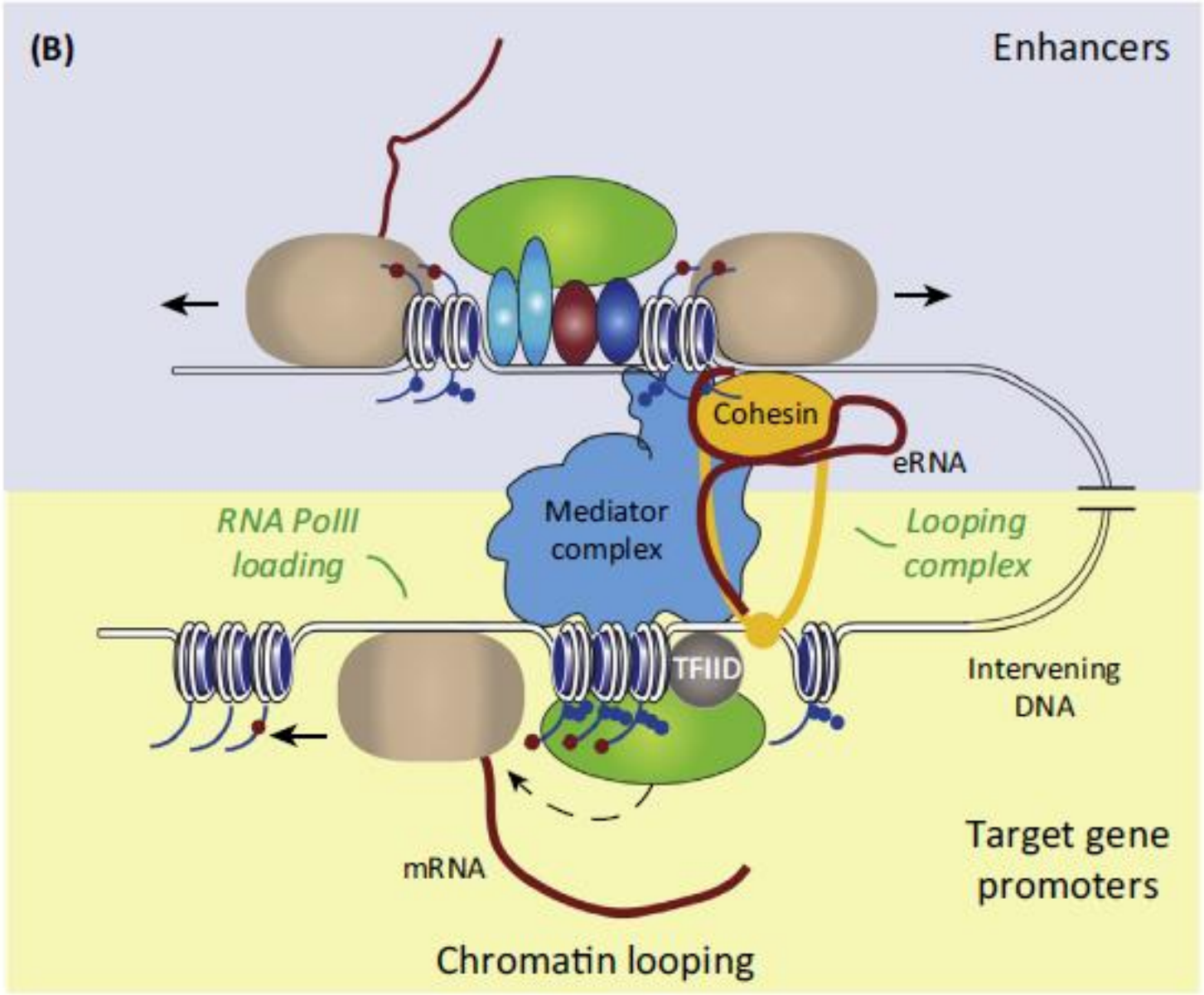
Annu. Rev. Immunol. 2010. 28:445–89

Enhancer RNAs and regulated transcriptional programs

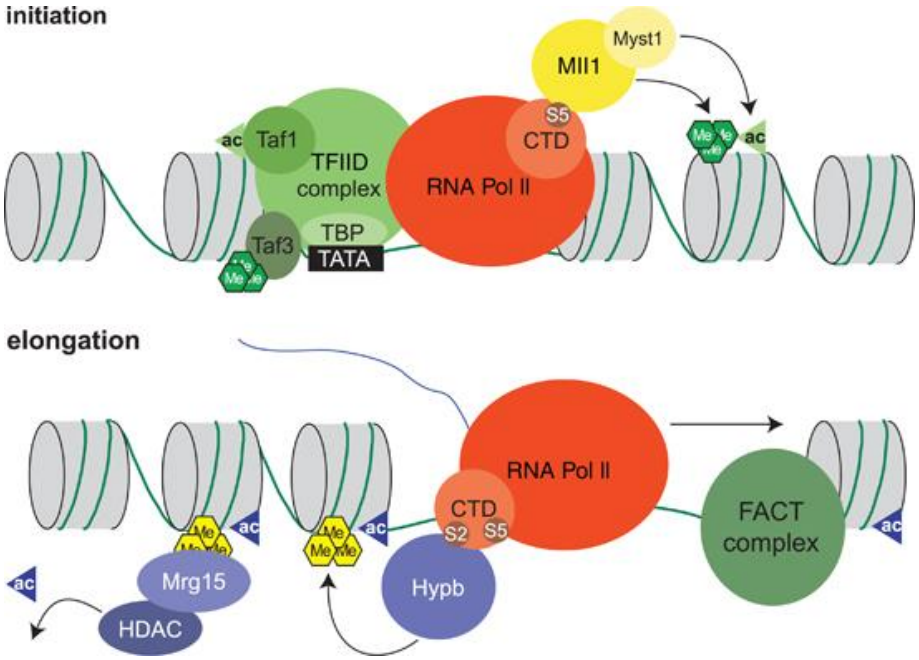
Michael T.Y. Lam[1], Wenbo Li[2], Michael G. Rosenfeld[2], and Christopher K. Glass[1,2]

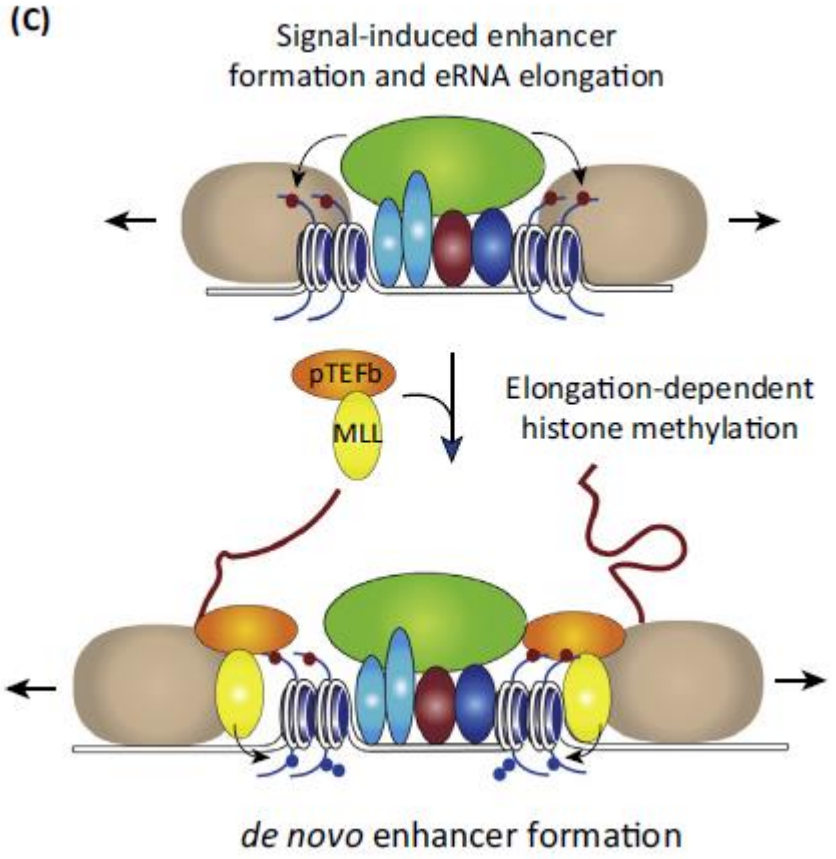
Trends in Biochemical Sciences April 2014, Vol. 39, No. 4

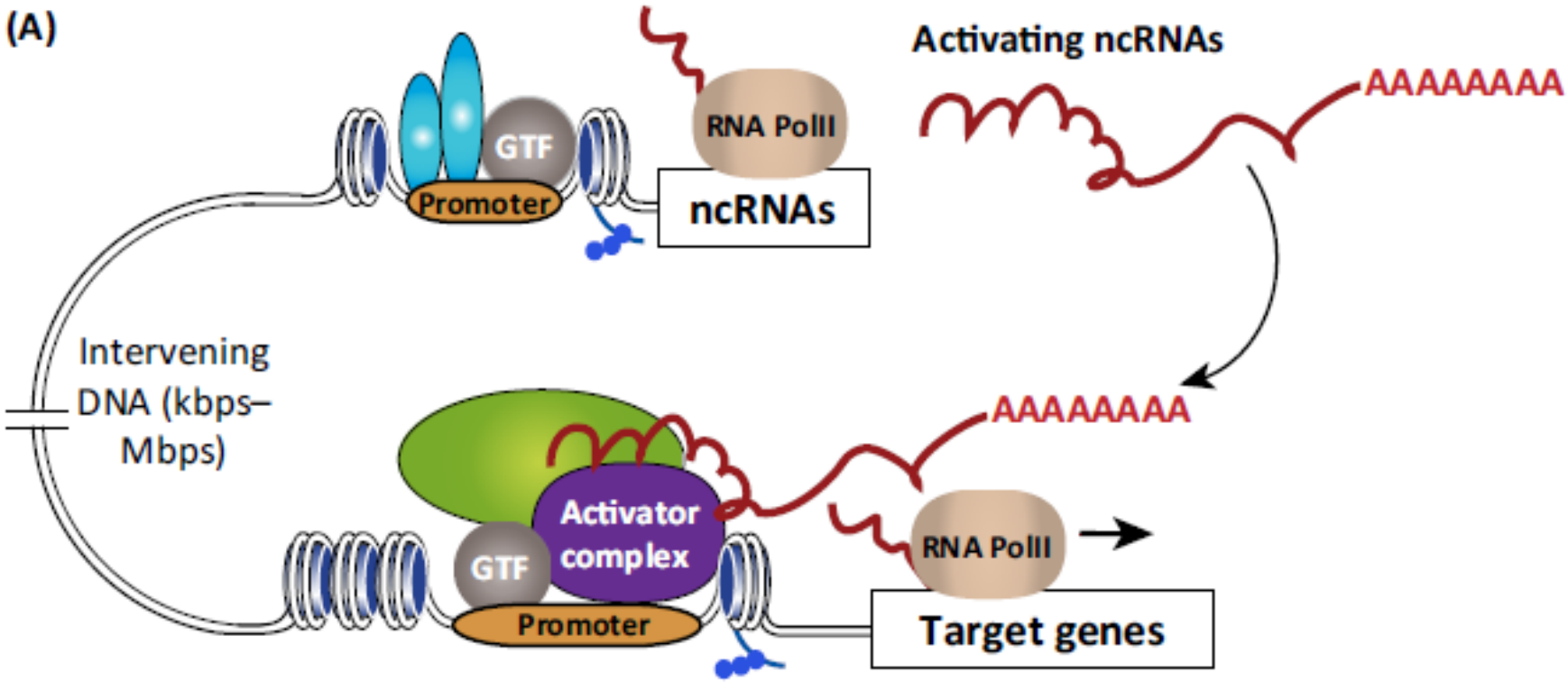
eRNA mediates the long interactions



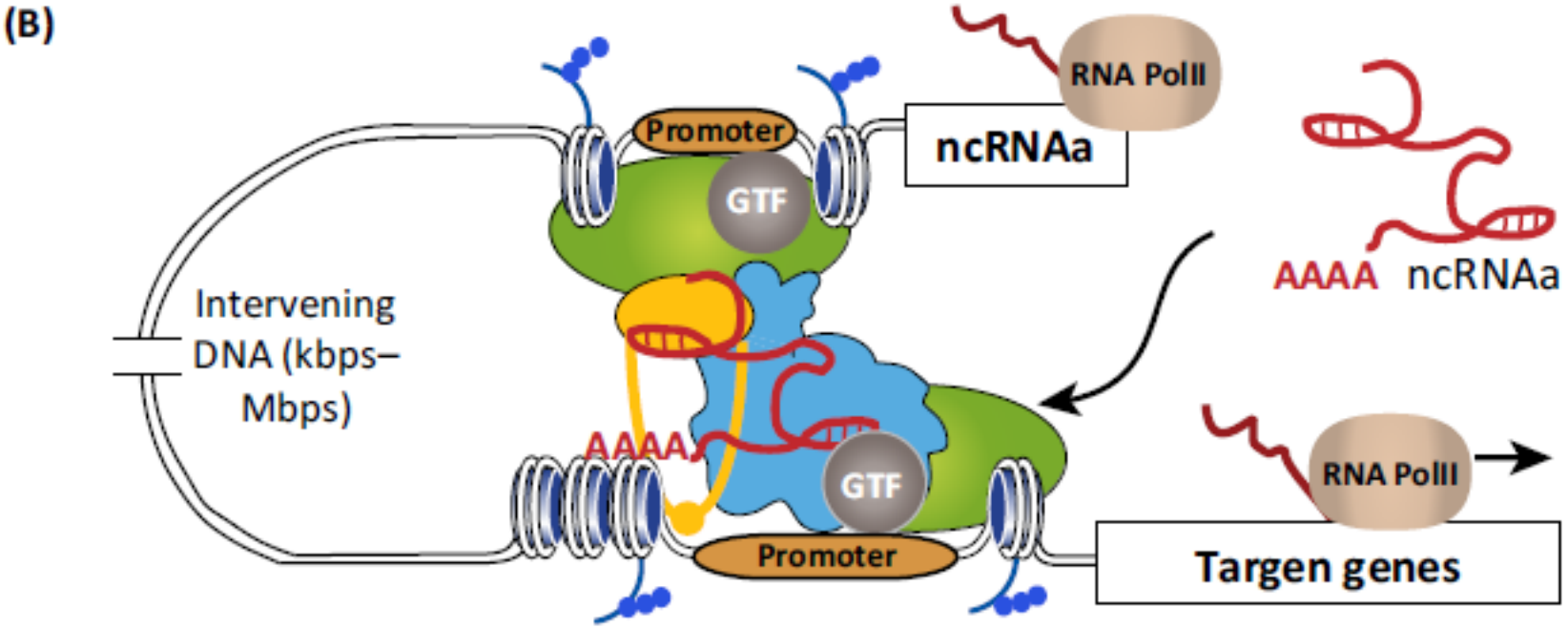
Molecular mechanisms that underline enhancer activation



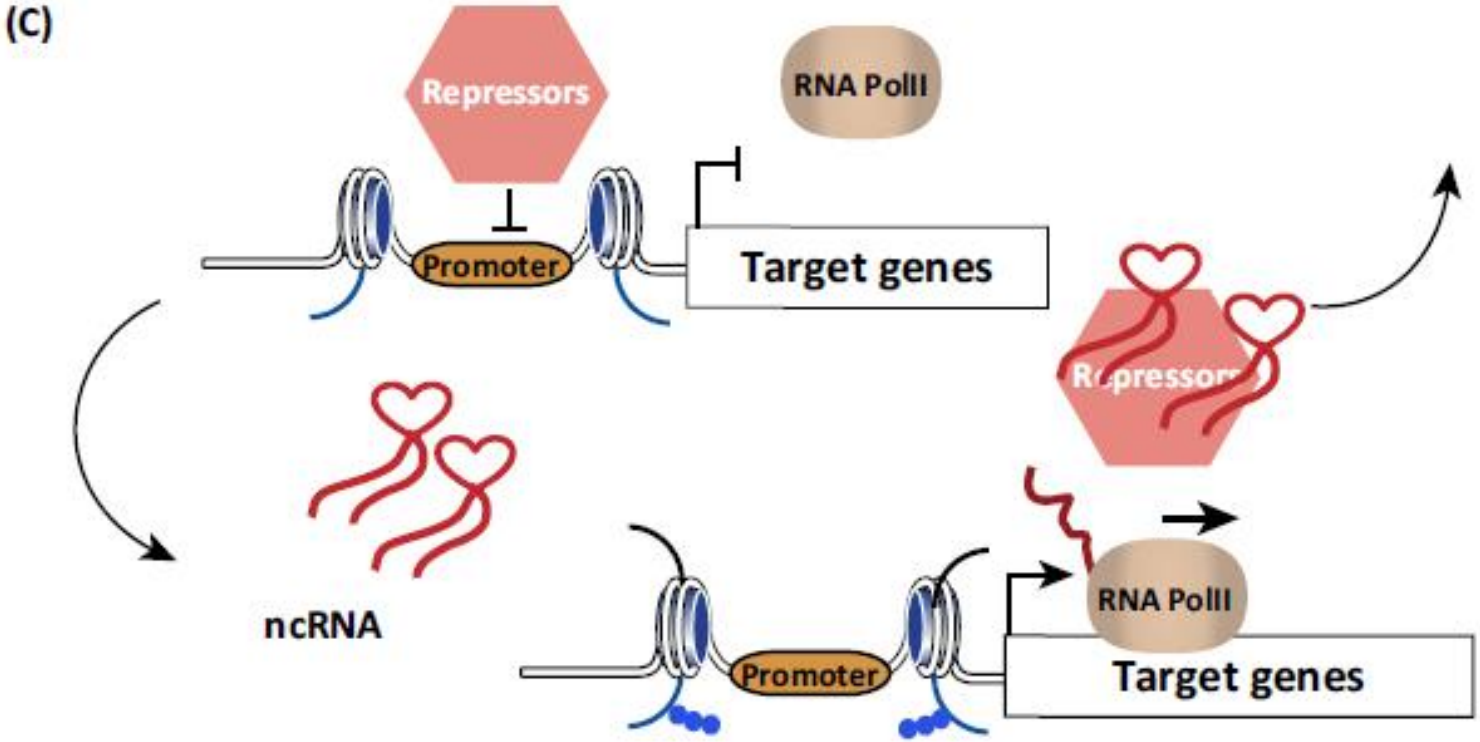




Model 1: ncRNAs collaborate with transcriptional activators



Model 2: ncRNAs modulate chromatin loops



Model 3: ncRNAs evict transcriptional repressors

Key:

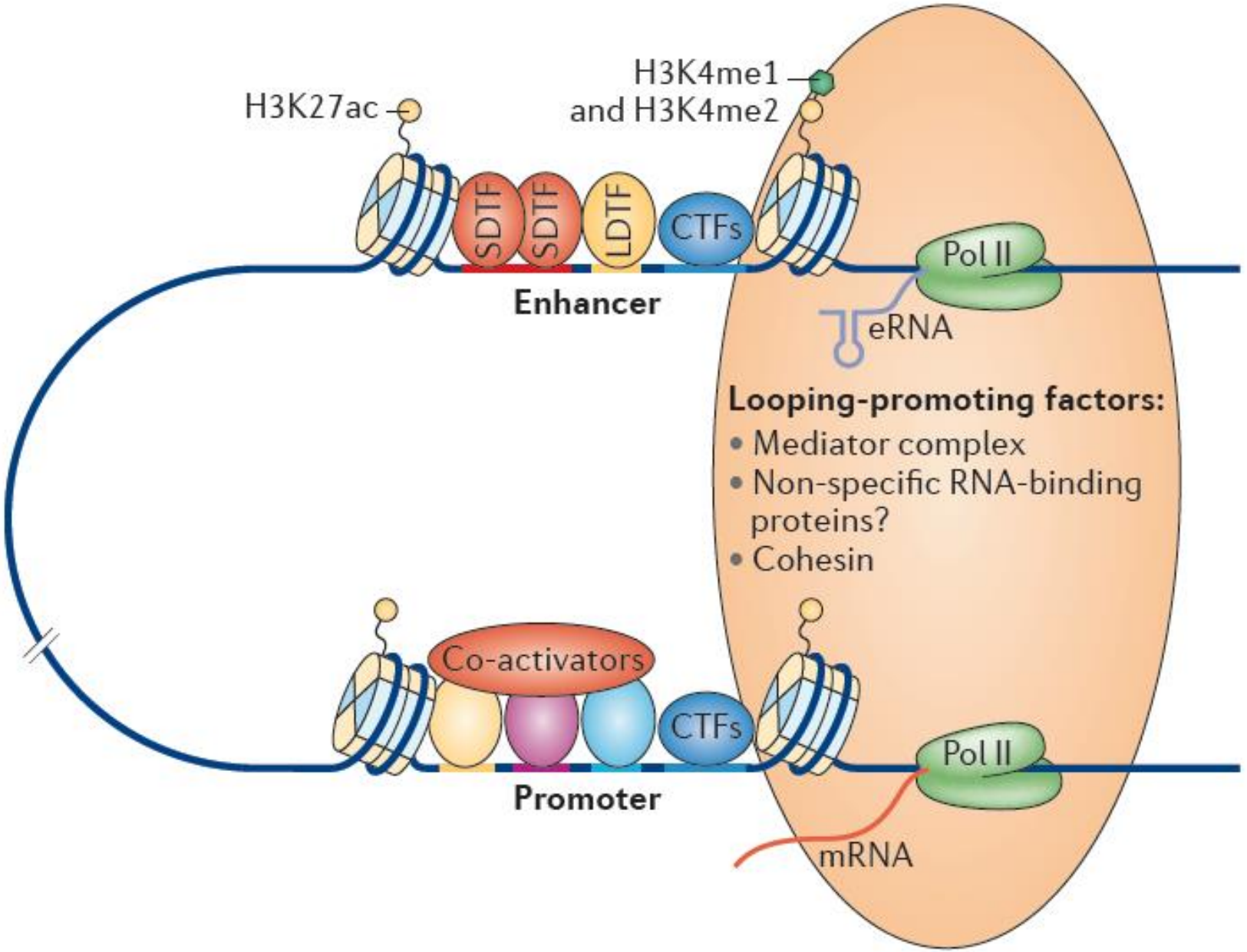
- Histone tail methylation
- Signal dependent TFs
- Looping complexes
- Coactivator complexes
- General TFs
- RNA polymerase II
- Transcriptional repressors

Enhancer Function

- **Chromatin looping**

- **Super-enhancers, cluster of enhancers, key player in the cell identity and differentiation**

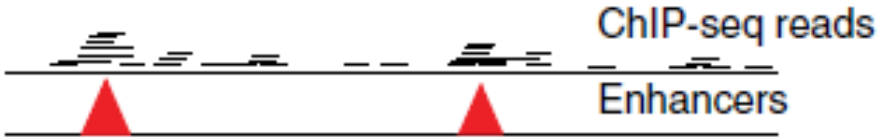
CHROMATIN LOOPING



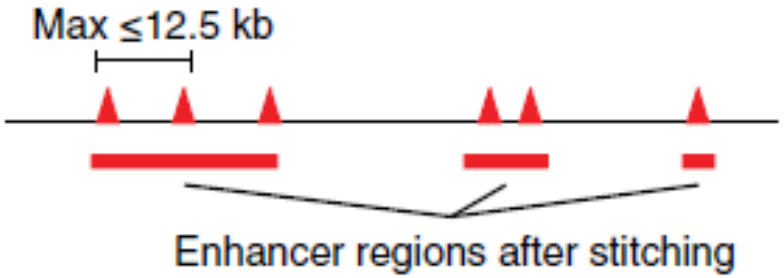
Super-enhancers.

a

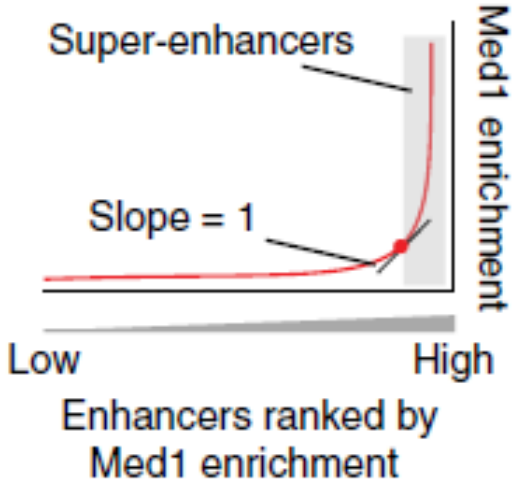
Step 1. Identification of enhancer locations



Step 2. Clustering of enhancers



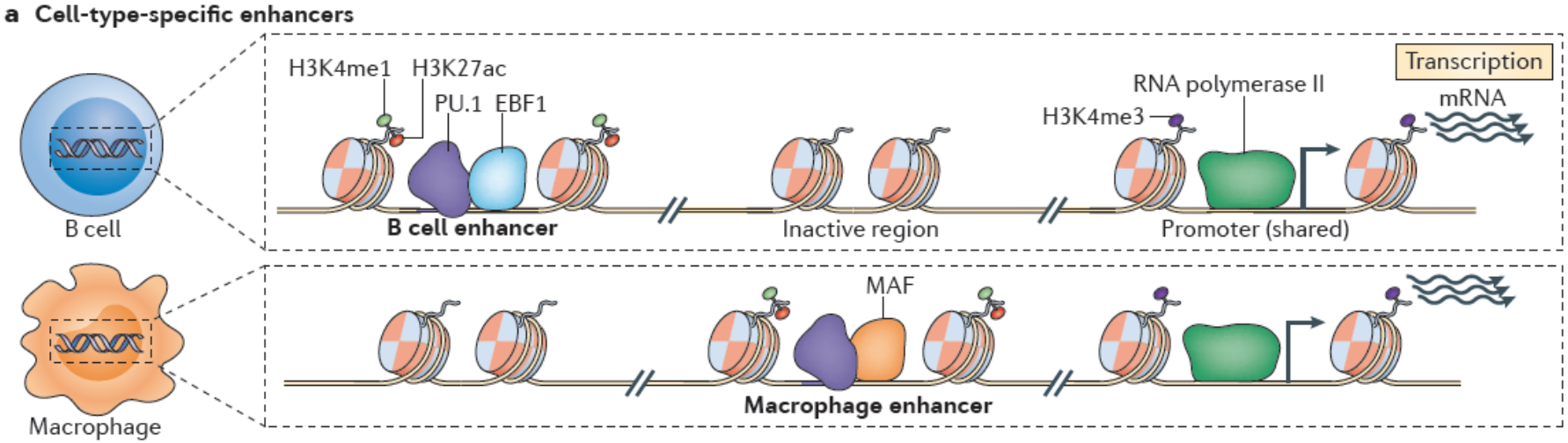
Step 3. Identify super-enhancers



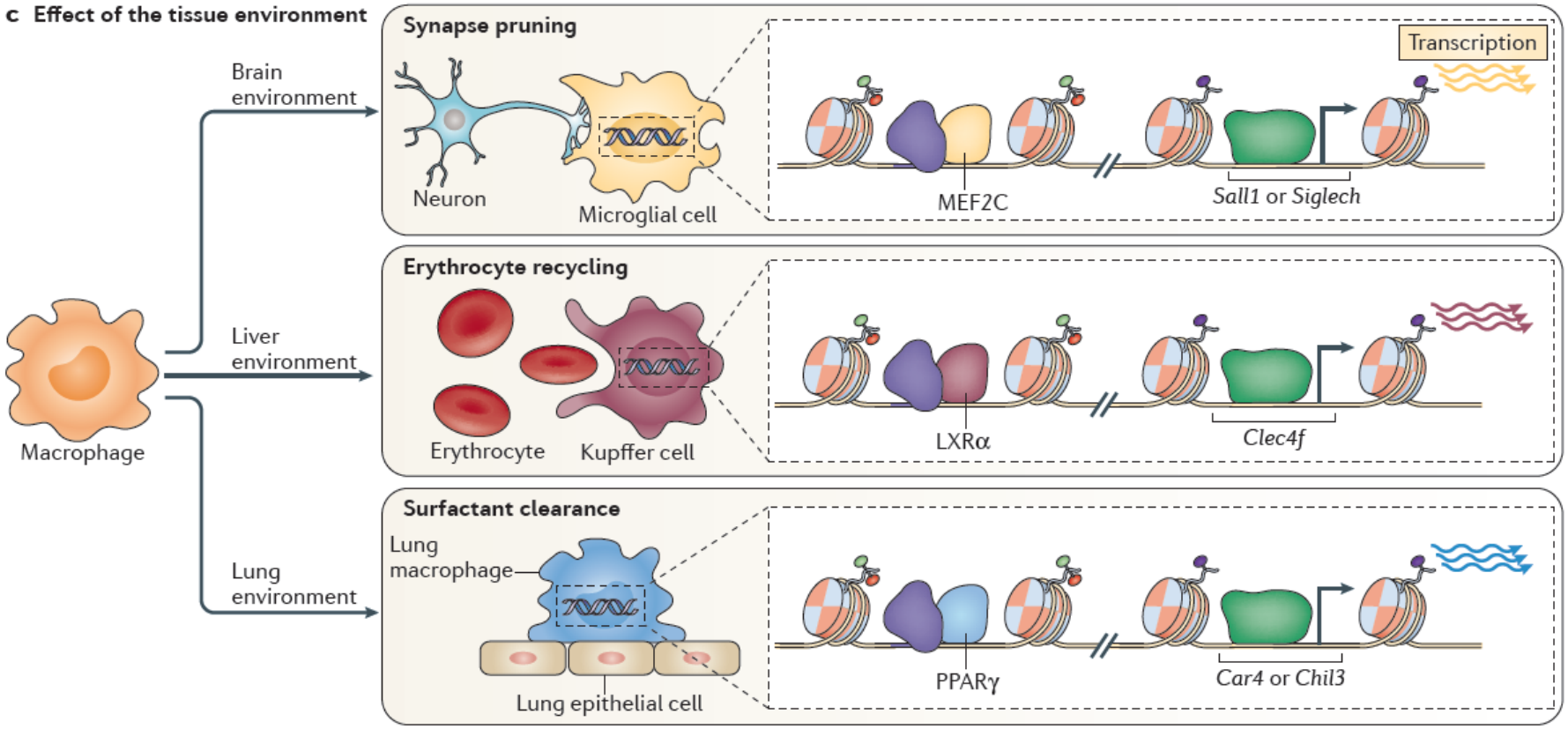
b

Factor used for step 1	Factor used for step 3	Reference
Oct4 + Sox2 + Nanog, Pu.1	Med1	Whyte *et al.*
MyoD, T-bet, C/EBPα	MyoD, T-bet, C/EBPα	Whyte *et al.*
H3K27ac	H3K27ac	Hnisz *et al.*
Med1	Med1	Loven *et al.*

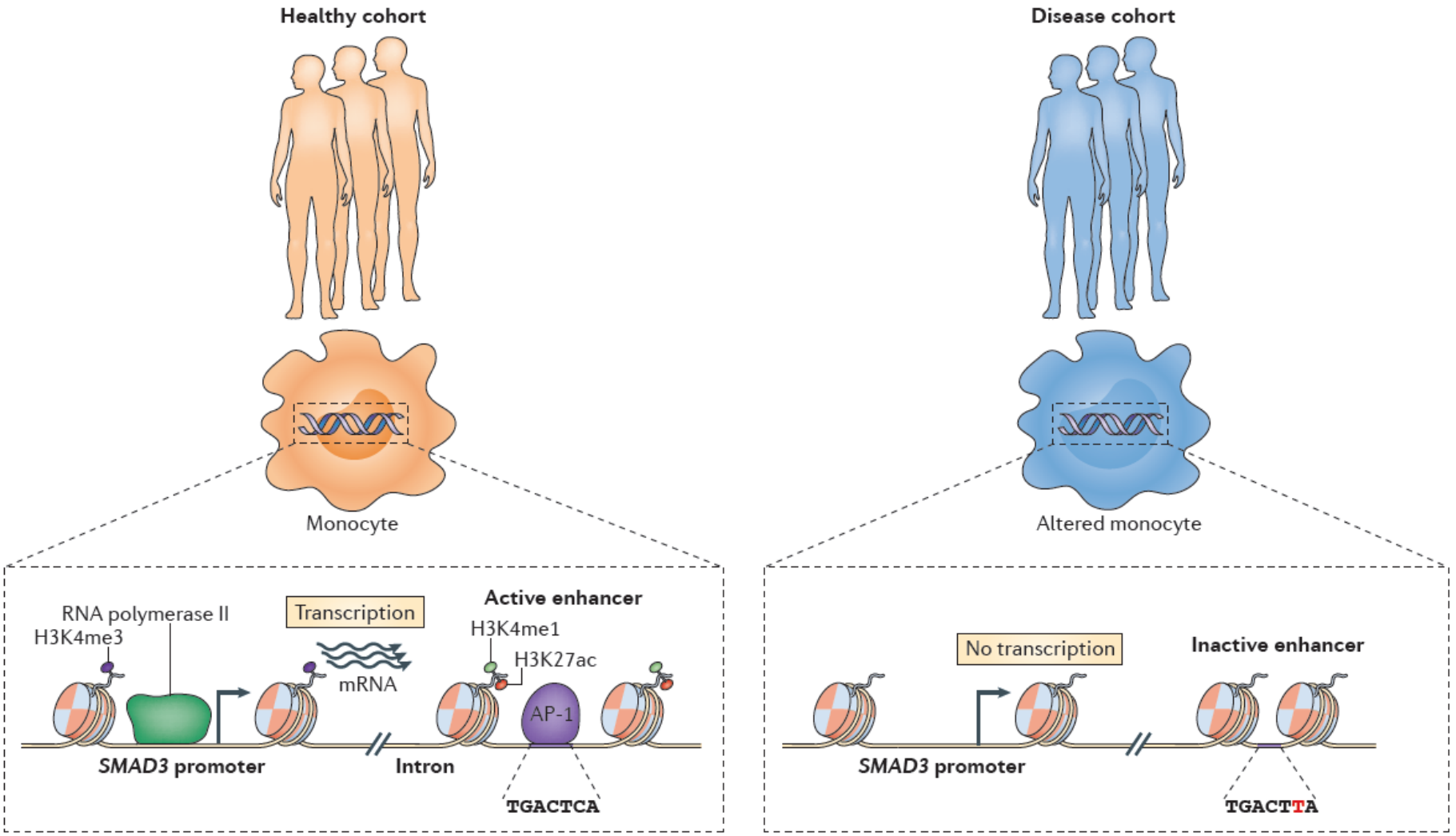
Cell-type-specific enhancers to regulate same genes



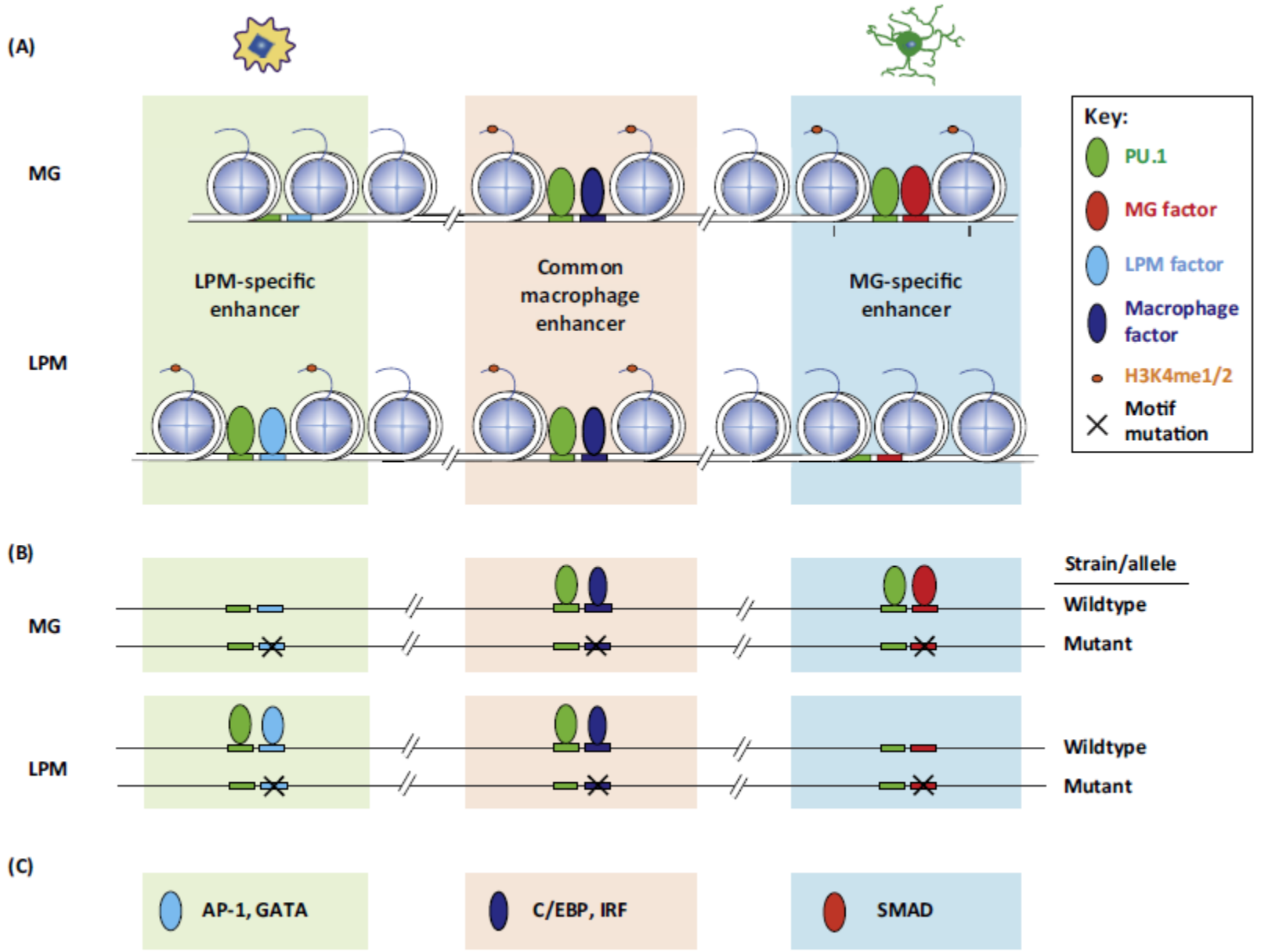
Effect of the tissue environment



Association of human chromatin data and susceptibility to immune disease



NATURAL GENETIC VARIATION IS ASSOCIATED WITH TF BINDING



Romanoski et al., 2015

TRENDS in Immunology

SNPs in the genomic regulatory regions may affect:

- **Enhancer Activation: loss of TFs interaction or TFs recruitment.**
- **Enhancer Selection: loss or association of LTDF**
- **Alteration of timing or specific tissues activation**
- **Long range interaction between genomic regulatory regions**

How are SNPs studying in genome-wide manner?

Super-Enhancers in the Control of Cell Identity and Disease

Denes Hnisz,[1,3] Brian J. Abraham,[1,3] Tong Ihn Lee,[1,3] Ashley Lau,[1,2] Violaine Saint-André,[1] Alla A. Sigova,[1] Heather A. Hoke,[1,2] and Richard A. Young[1,2,*]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]These authors contributed equally to this work

*Correspondence: young@wi.mit.edu

http://dx.doi.org/10.1016/j.cell.2013.09.053

SUMMARY

Super-enhancers are large clusters of transcriptional enhancers that drive expression of genes that define cell identity. Improved understanding of the roles that super-enhancers play in biology would be afforded by knowing the constellation of factors that constitute these domains and by identifying super-enhancers across the spectrum of human cell types. We describe here the population of transcription factors, cofactors, chromatin regulators, and transcription apparatus occupying super-enhancers in embryonic stem cells and evidence that super-enhancers are highly transcribed. We produce a catalog of super-enhancers in a broad range of human cell types and find that super-enhancers associate with genes that control and define the biology of these cells. Interestingly, disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types. Furthermore, we find that cancer cells generate super-enhancers at oncogenes and other genes important in tumor pathogenesis. Thus, super-enhancers play key roles in human cell identity in health and in disease.

SUMMARY

Super-enhancers are large clusters of transcriptional enhancers that drive expression of genes that define cell identity. Improved understanding of the roles that super-enhancers play in biology would be afforded by knowing the constellation of factors that constitute these domains and by identifying super-enhancers across the spectrum of human cell types. We describe here the population of transcription factors, cofactors, chromatin regulators, and transcription apparatus occupying super-enhancers in embryonic stem cells and evidence that super-enhancers are highly transcribed. We produce a catalog of super-enhancers in a broad range of human cell types and find that super-enhancers associate with genes that control and define the biology of these cells. Interestingly, disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types. Furthermore, we find that cancer cells generate super-enhancers at oncogenes and other genes important in tumor pathogenesis. Thus, super-enhancers play key roles in human cell identity in health and in disease.

DEFINITION

AIM

1) Protein complexes

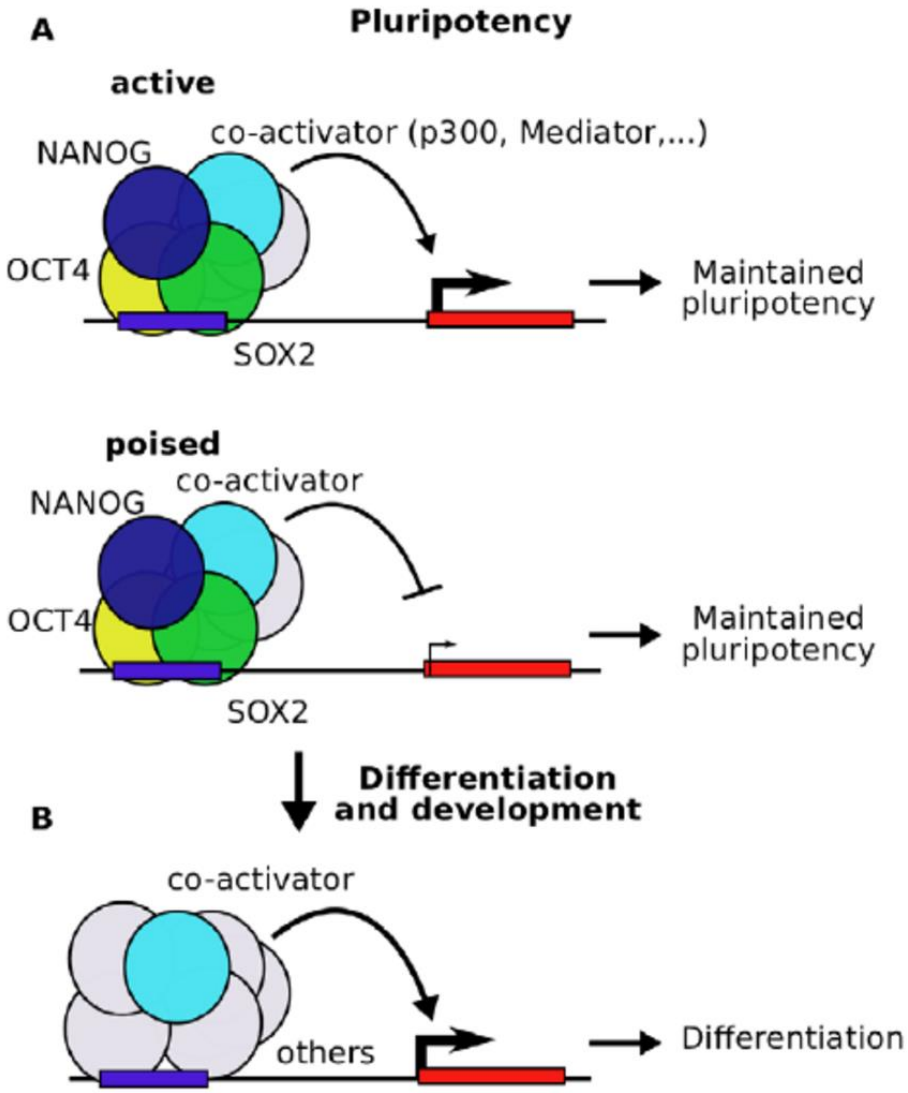
2) SE cell type-specific

3) SNPs linked to disease in SE

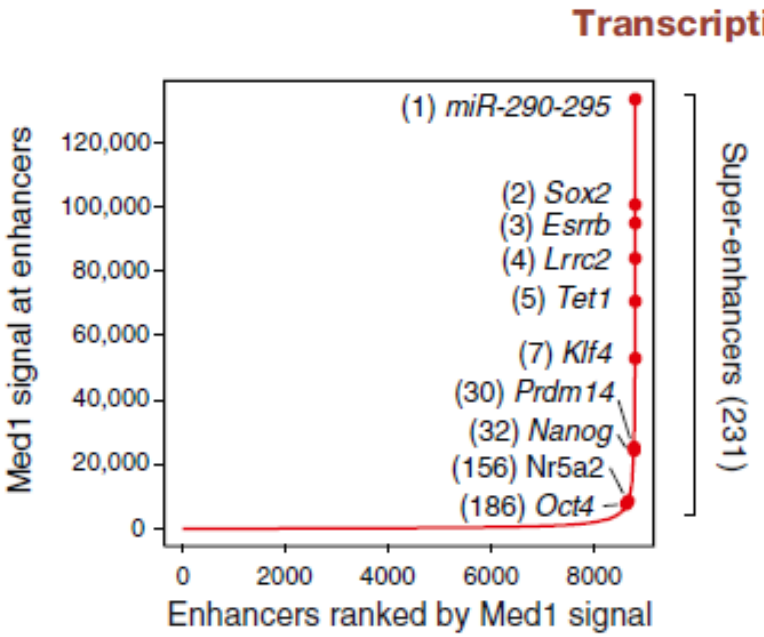
CONCLUSION

Murine Embrionic stem cells (ESC)

SOX2, Nanog, OCT4: transcription factors that bind SE controlling genes for mantained pluripotency



Mediator Coactivator Complexes and Master TFs are bound at Super-enhancers



Super-enhancers are clusters of enhancers—formed by binding of high levels of master transcription factors and Mediator coactivator—that drive high-level expression of genes encoding key regulators of cell identity (Figure 1A) (Whyte et al., 2013). Five ESC transcription factors were previously shown to occupy super-enhancers (Oct4, Sox2, Nanog, Klf4, and Esrrb) (Whyte et al., 2013), but there are many additional transcription factors that contribute to the control of ESCs (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). We compiled ChIP-seq data for 15 additional transcription factors in ESCs, for which high-quality ChIP-seq data were available, and investigated whether they occupy enhancers defined by Oct4, Sox2, and Nanog (OSN) co-occupancy (Whyte et al., 2013) (Table S1 avail-

(A) Distribution of Med1 ChIP-seq signal at enhancers reveals two classes of enhancers in ESCs. Enhancer regions are plotted in an increasing order based on their input-normalized Med1 ChIP-seq signal. Super-enhancers are defined as the population of enhancers above the inflection point of the curve. Example super-enhancers are highlighted along with their respective ranks and their associated genes.

Bioinformatic analysis for the definition of SE:

- Signal in proximity of the gene

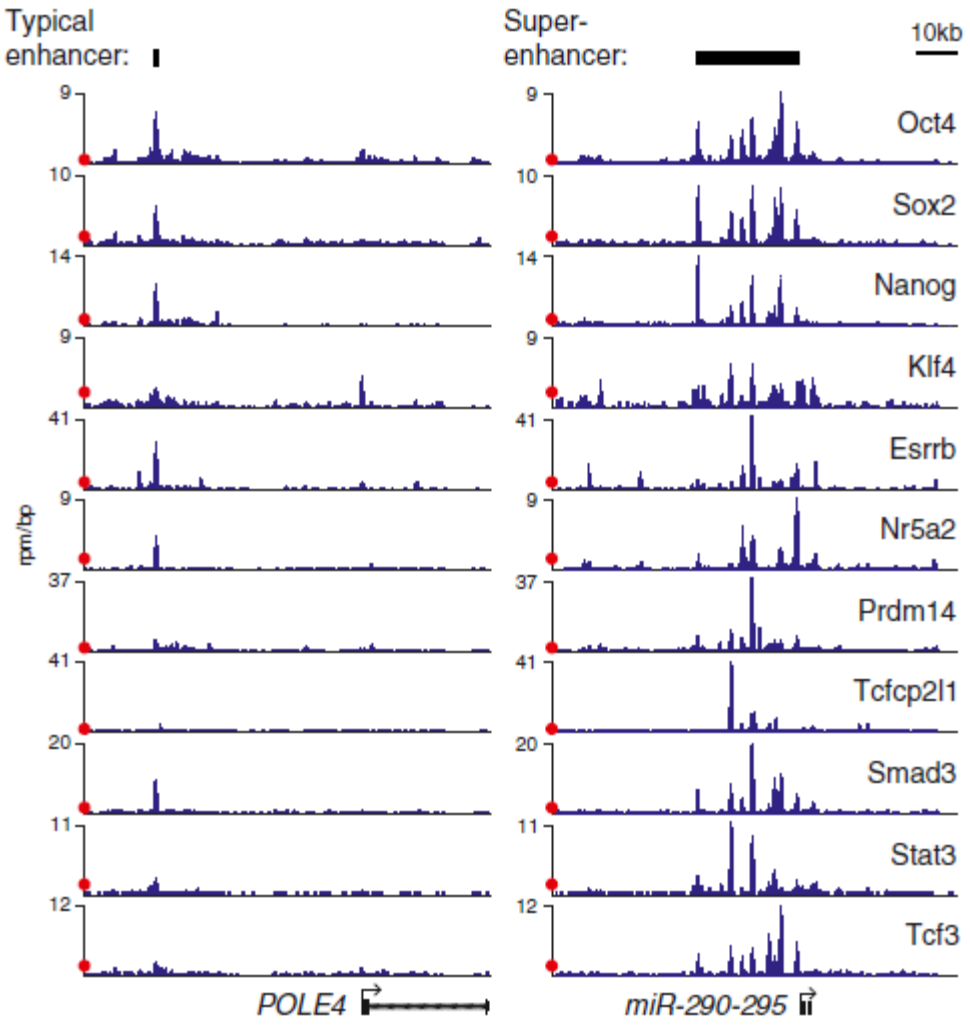
- signal extended in the genomic regions that identify SE

- increased numbers of reads into SE respect to costituent, single enhancer in the SE

- increased signal into SE respect to typical enhancer

Mediator Coactivator Complexes and Master TFs are bound at Super-enhancers

Chromatin Immunoprepitation Binding Profiles at target genes

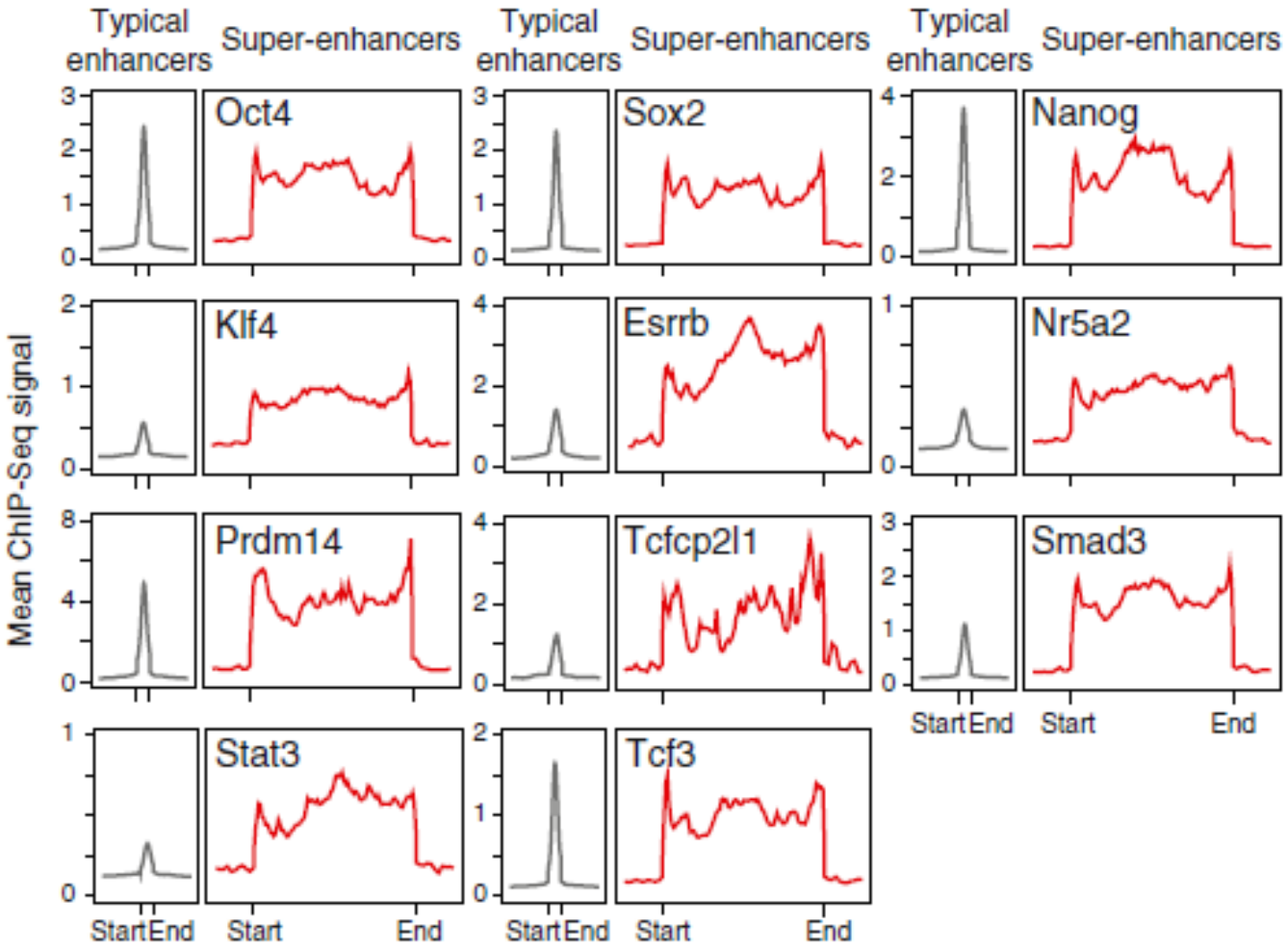


Transcription Factors

Specific Loci

(B) ChIP-seq binding profiles for the indicated transcription factors at the *POLE4* and *miR-290–295* loci in ESCs. Red dots indicate the median enrichment of all bound regions in the respective ChIP-seq data sets and are positioned at maximum 20% of the axis height. rpm/bp, reads per million per base pair.

ChIP-seq signal across SE domains

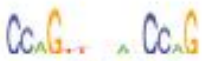


(C) Metagene representations of the mean ChIP-seq signal for the indicated transcription factors across typical enhancers and super-enhancer domains. Metagenes are centered on the enhancer region, and the length of the enhancer reflects the difference in median lengths (703 bp for typical enhancers, 8,667 bp for super-enhancers). Additional 3 kb surrounding each enhancer region is also shown.

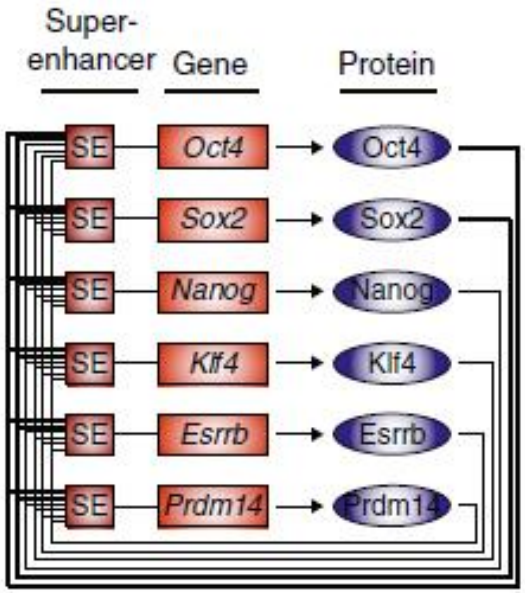
TFs motif enrichment are used to associate gene target

F

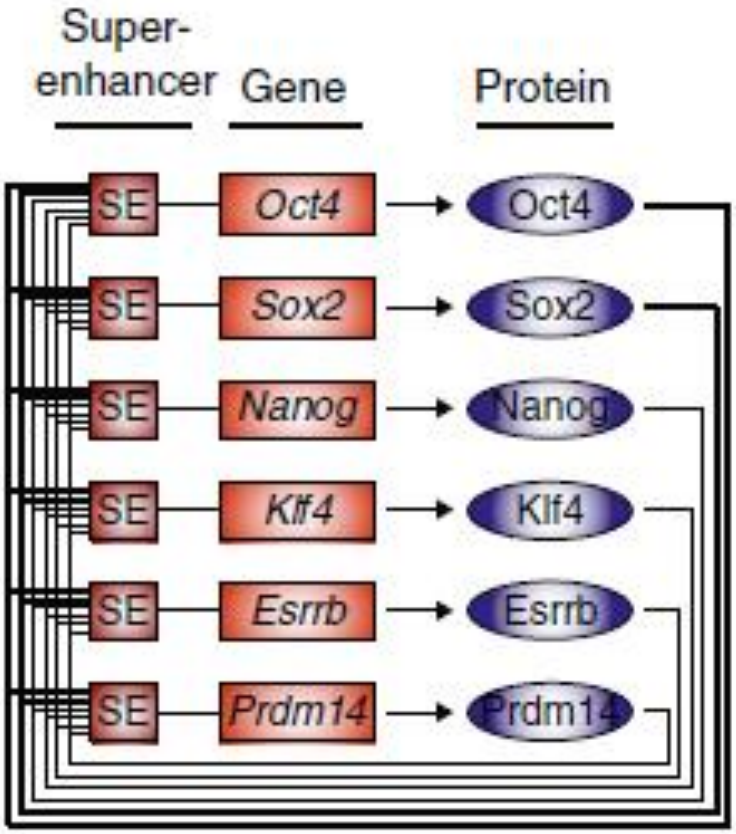
Transcription factor	Motif	P-value
Oct4		$9.19*10^{-64}$
Sox2		$3.01*10^{-67}$
Nanog		$9.46*10^{-17}$
Klf4		$4.33*10^{-6}$
Esrrb		$2.55*10^{-84}$
Nr5a2	n.a.	n.a.

Transcription factor	Motif	P-value
Prdm14	n.a.	n.a.
Tcfcp2l1		$6.83*10^{-11}$
Smad3		$9.31*10^{-11}$
Stat3		$2.90*10^{-10}$
Tcf3		$5.46*10^{-27}$

G

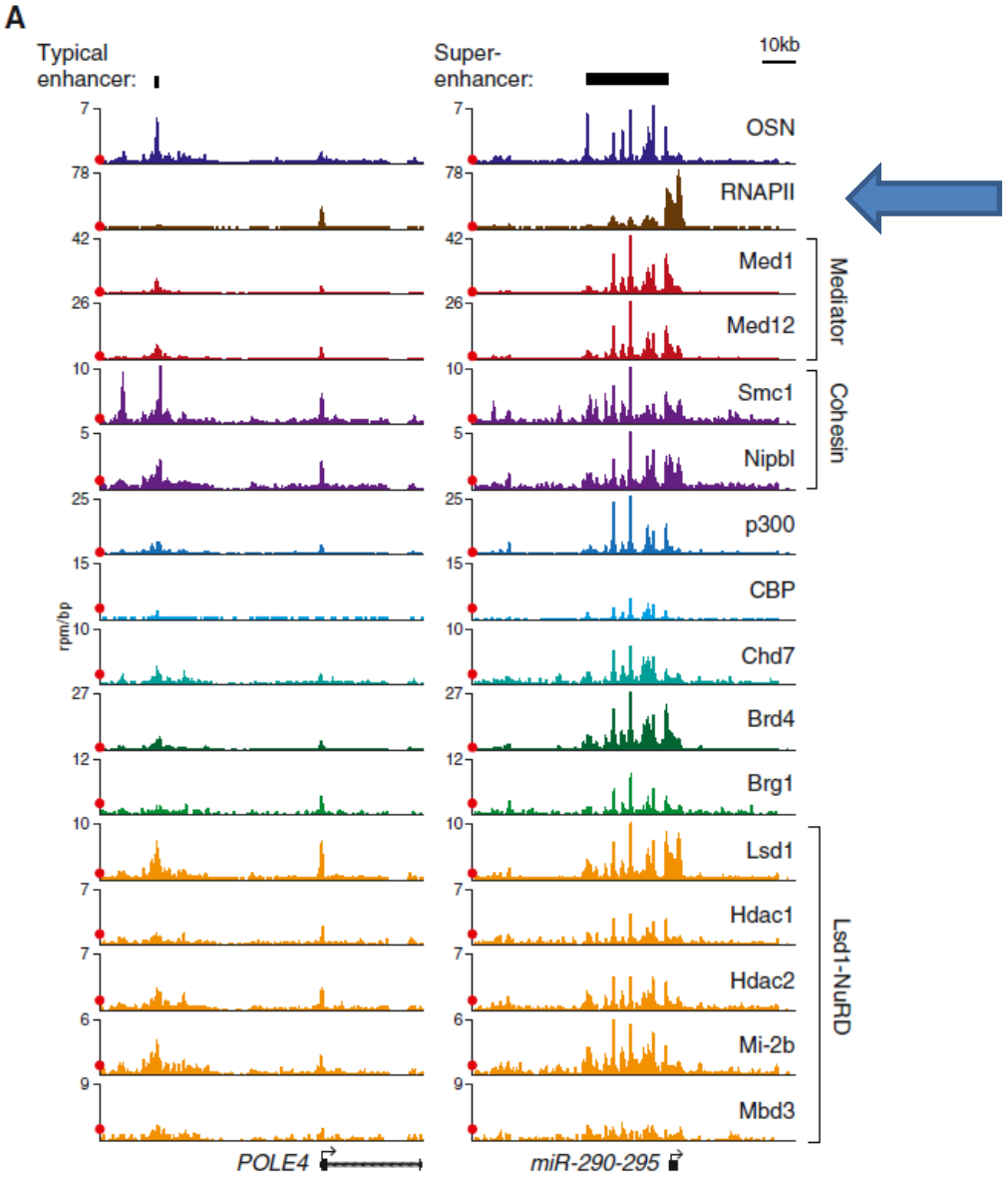


Core Transcriptional Regulatory Circuit of ESCs

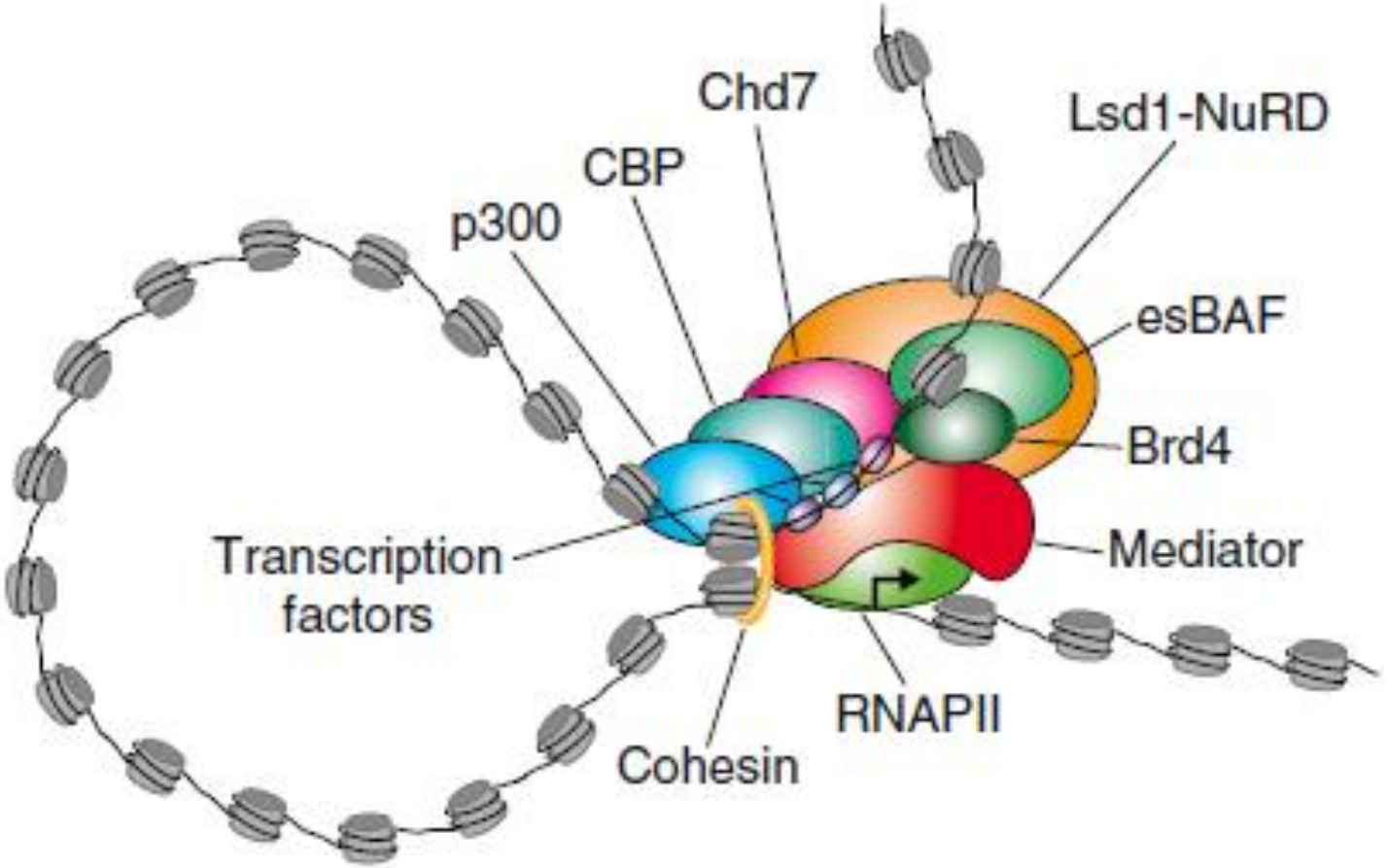


(G) Revised model of the core transcriptional regulatory circuitry of ESCs. The model contains an interconnected autoregulatory loop consisting of transcription factors that meet three criteria: (1) their genes are driven by super-enhancers, (2) they co-occupy their own super-enhancers as well as those of the other transcription factor genes in the circuit, and (3) they play essential roles in regulation of ESC state and iPSC reprogramming. The layout of the circuit model was adapted from Whyte et al. (2013).

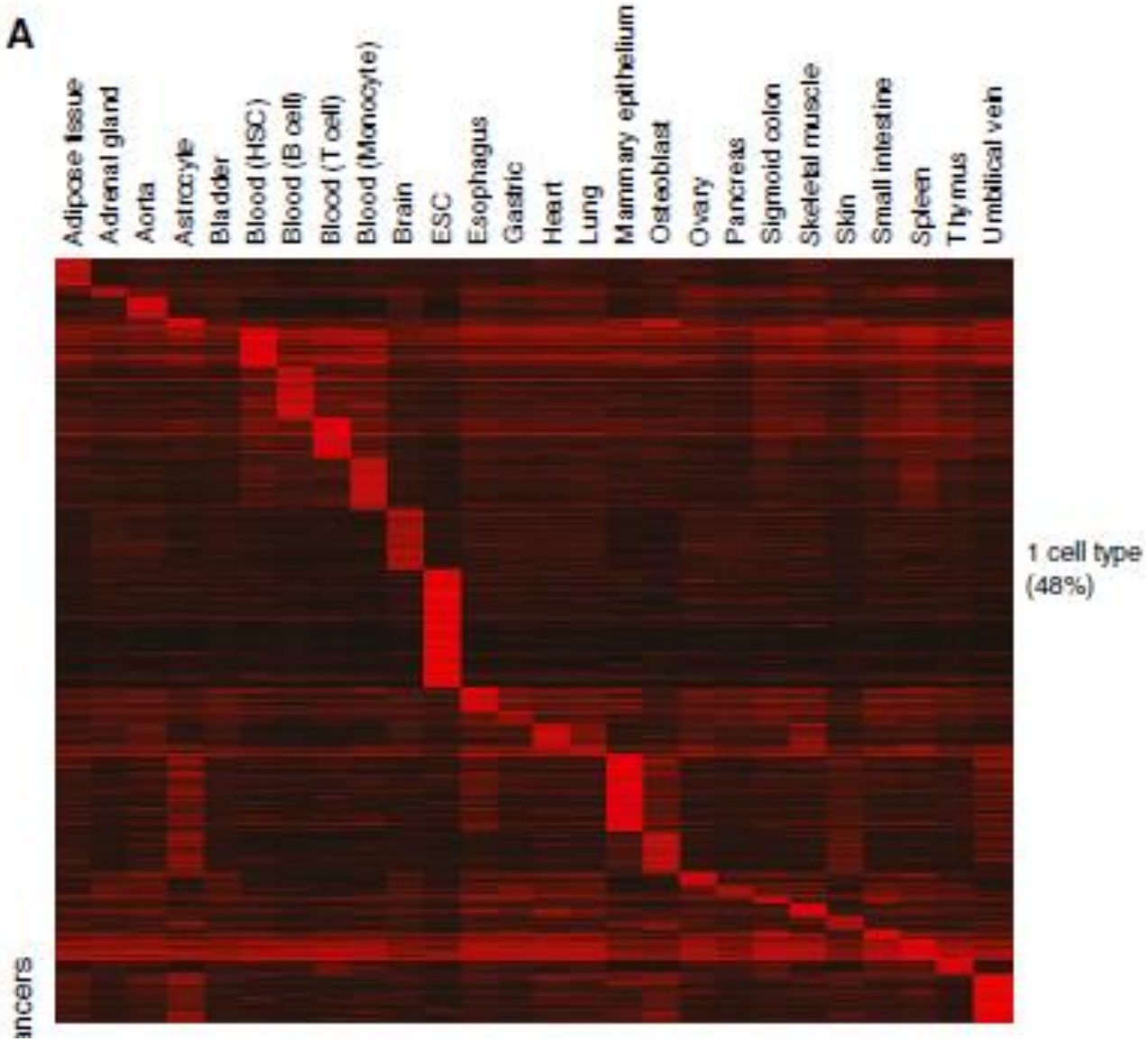
Super-enhancers are occupied by a large portion of the enhancer-associated RNA polymerase II



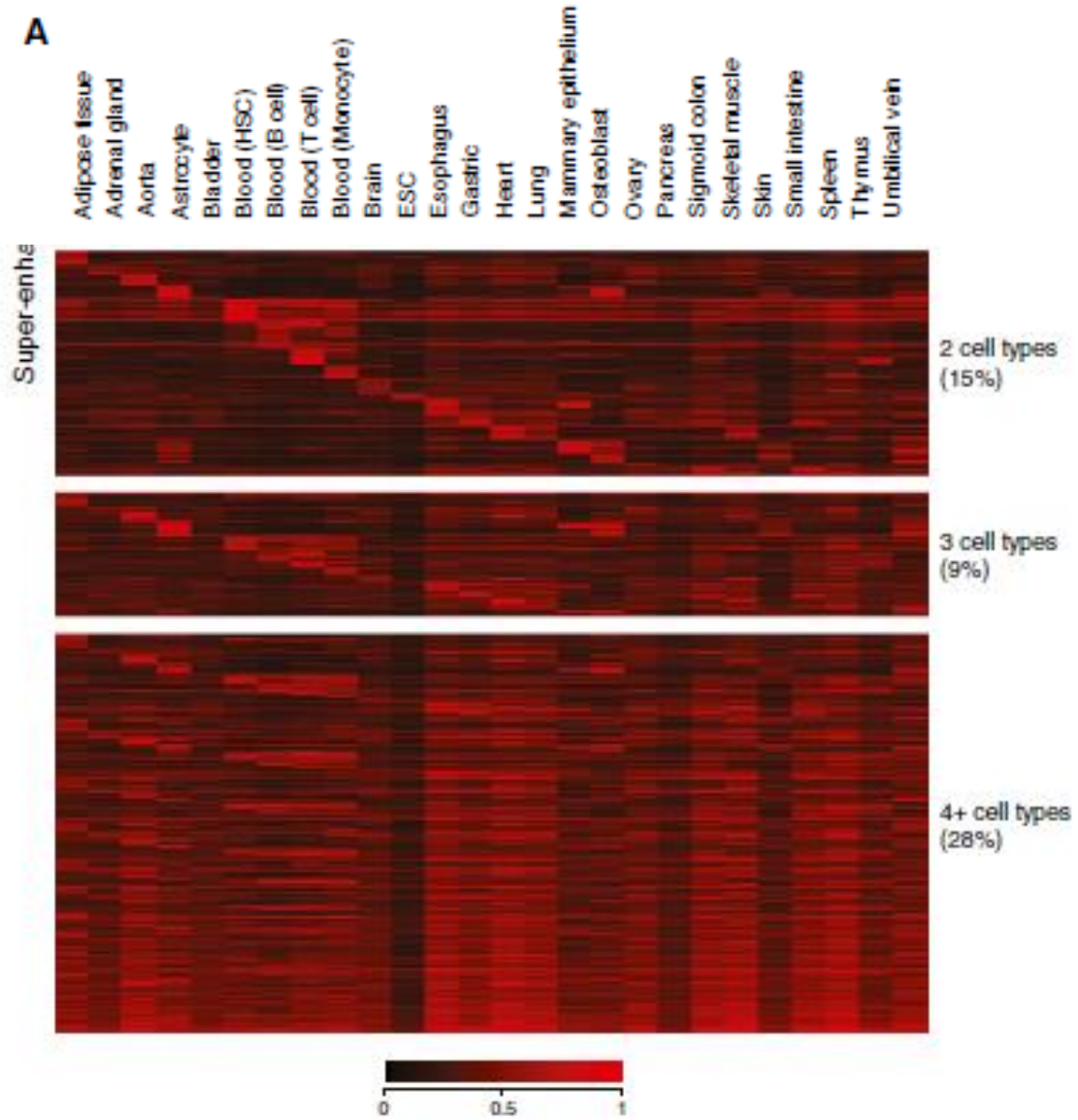
Model showing **RNAPII, transcriptional cofactors, and chromatin regulators that are found in ESC super-enhancers**. The indicated proteins are responsible for diverse enhancer-related functions, such as enhancer looping, gene activation, nucleosome remodeling, and histone modification.



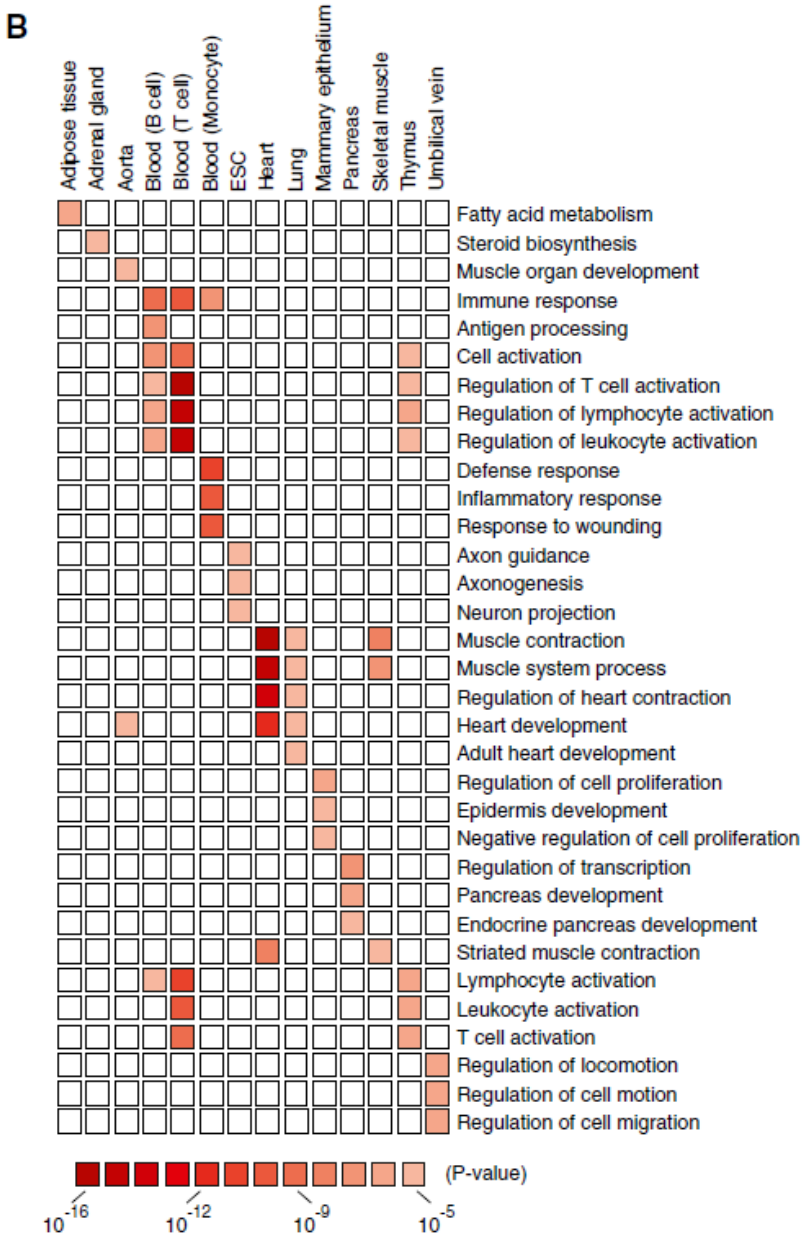
SUPERENHANCER SHARED BETWEEN SEVERAL CELL TYPES



SUPERENHANCER SHARED BETWEEN SEVERAL CELL TYPES



GENE ASSOCIATED TO SUPERENHANCER IN SEVERAL CELL TYPES: GENE ONTOLOGY



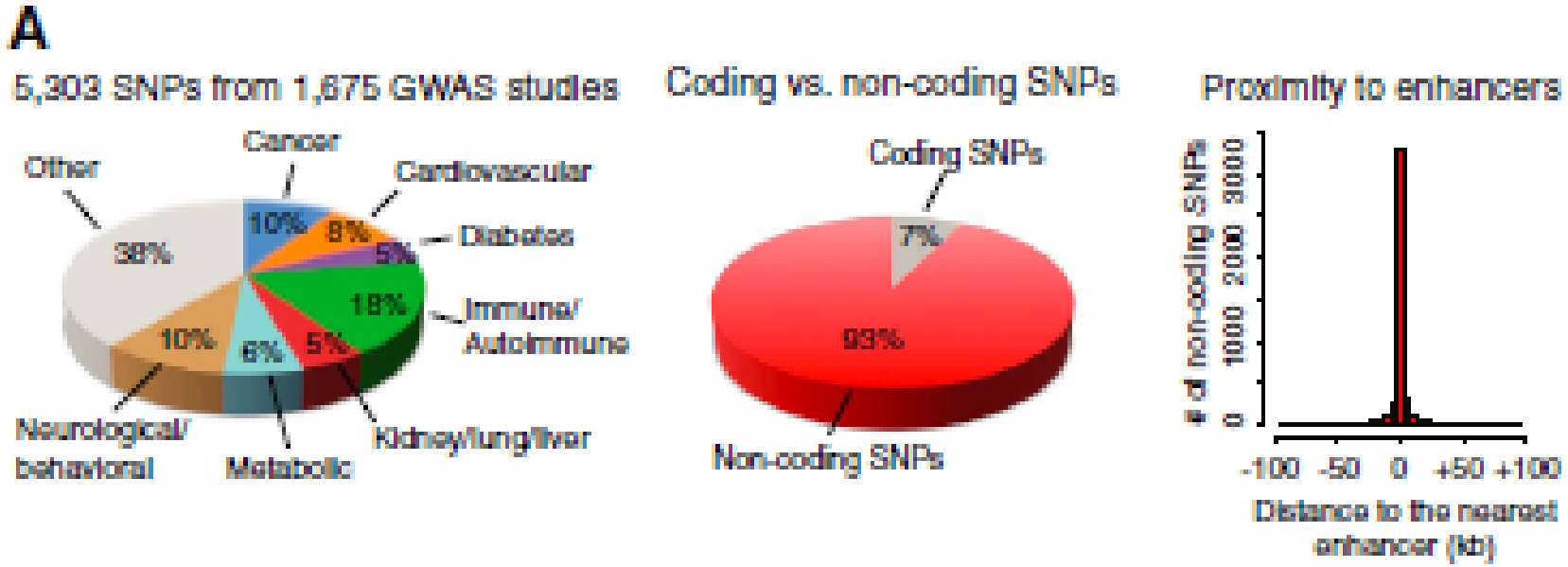
MASTER TRANSCRIPTION FACTORS IN SIX CELL TYPES

C

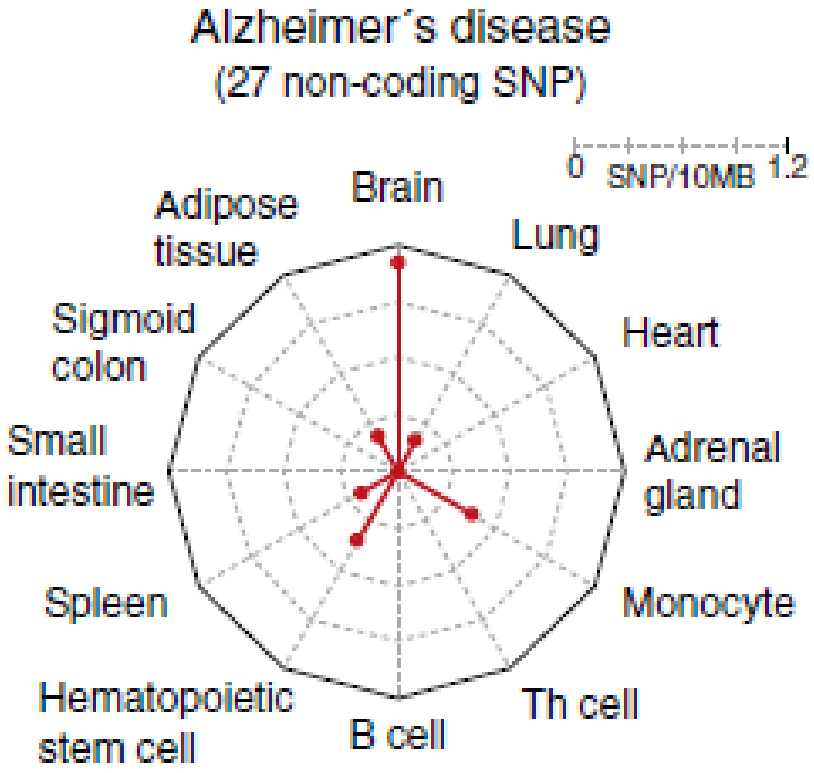
Brain	Heart	Skeletal muscle	Lung	Adipose tissue	B cell
NKX2-2	TBX20	MYOD1	NFIB	PPARG	IKZF3
OLIG1	TBX5	PITX2	TBX5	CEBPB	PAX5
BRN2	MEF2A	SIX1	CEBPA	CEBPD	BACH2
SOX10	NKX2-5	TEAD4	TBX2	CREB1	OCT2
SOX2	GATA4		TBX3		IKZF1
					IRF8

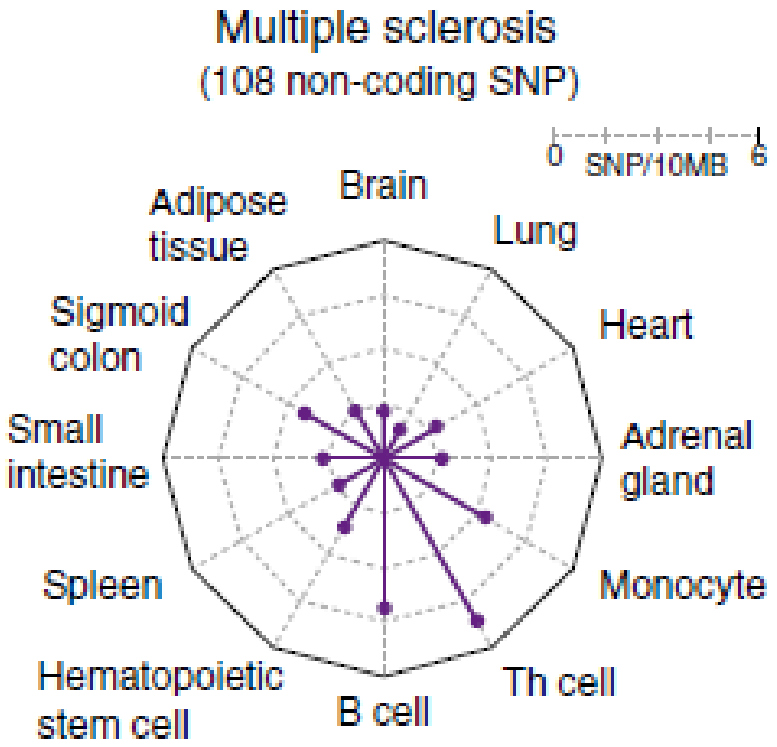
(C) Candidate master transcription factors identified in six cell types. All of these transcription factors were previously demonstrated to play key roles in the biology of the respective cell type or facilitate reprogramming to the respective cell type.

SINGLE NUCLEOTIDE MUTATIONS LINKED TO DISEASE (GWAS) ASSOCIATED TO SUPERENHANCERS



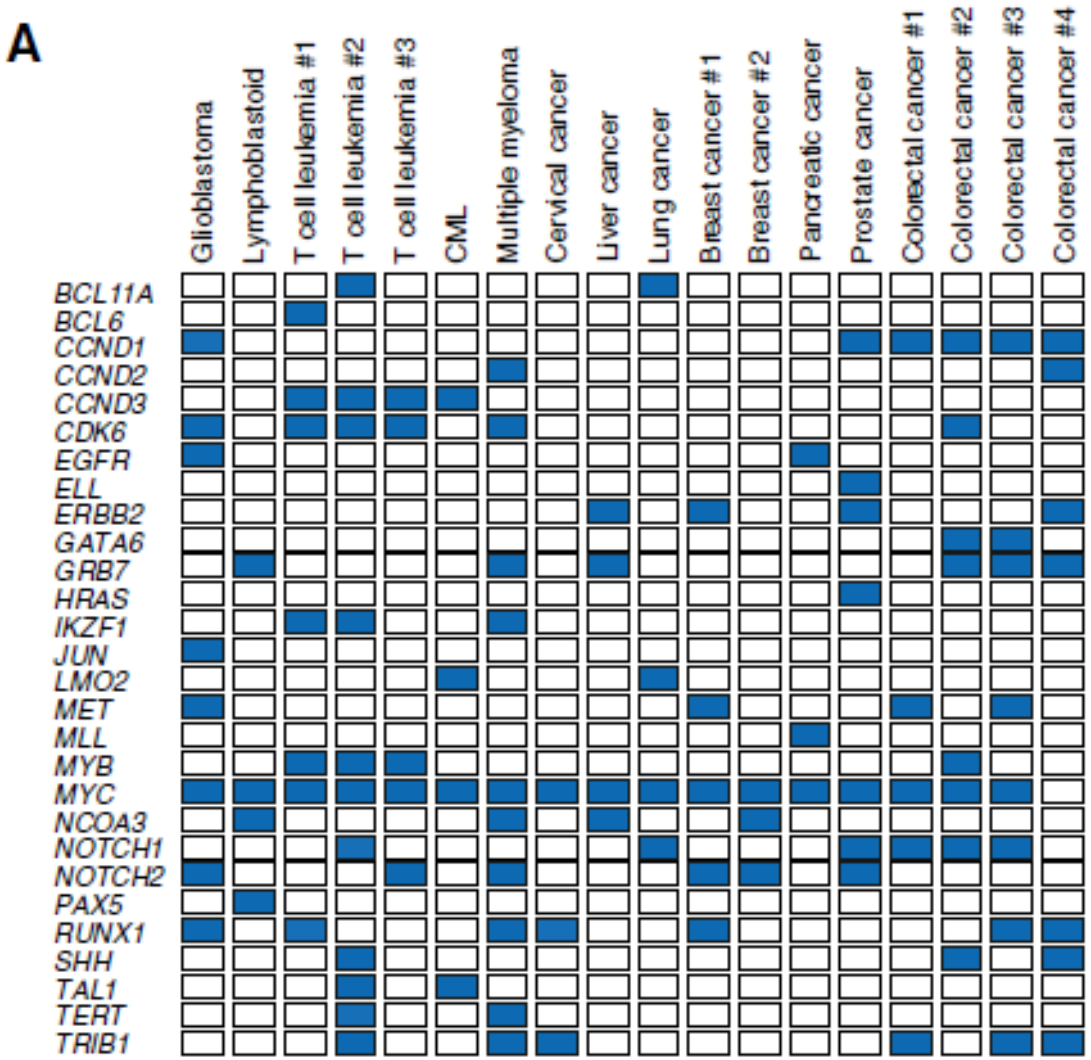
SINGLE NUCLEOTIDE MUTATIONS LINKED TO DISEASE (GWAS) ASSOCIATED TO SUPERENHANCERS



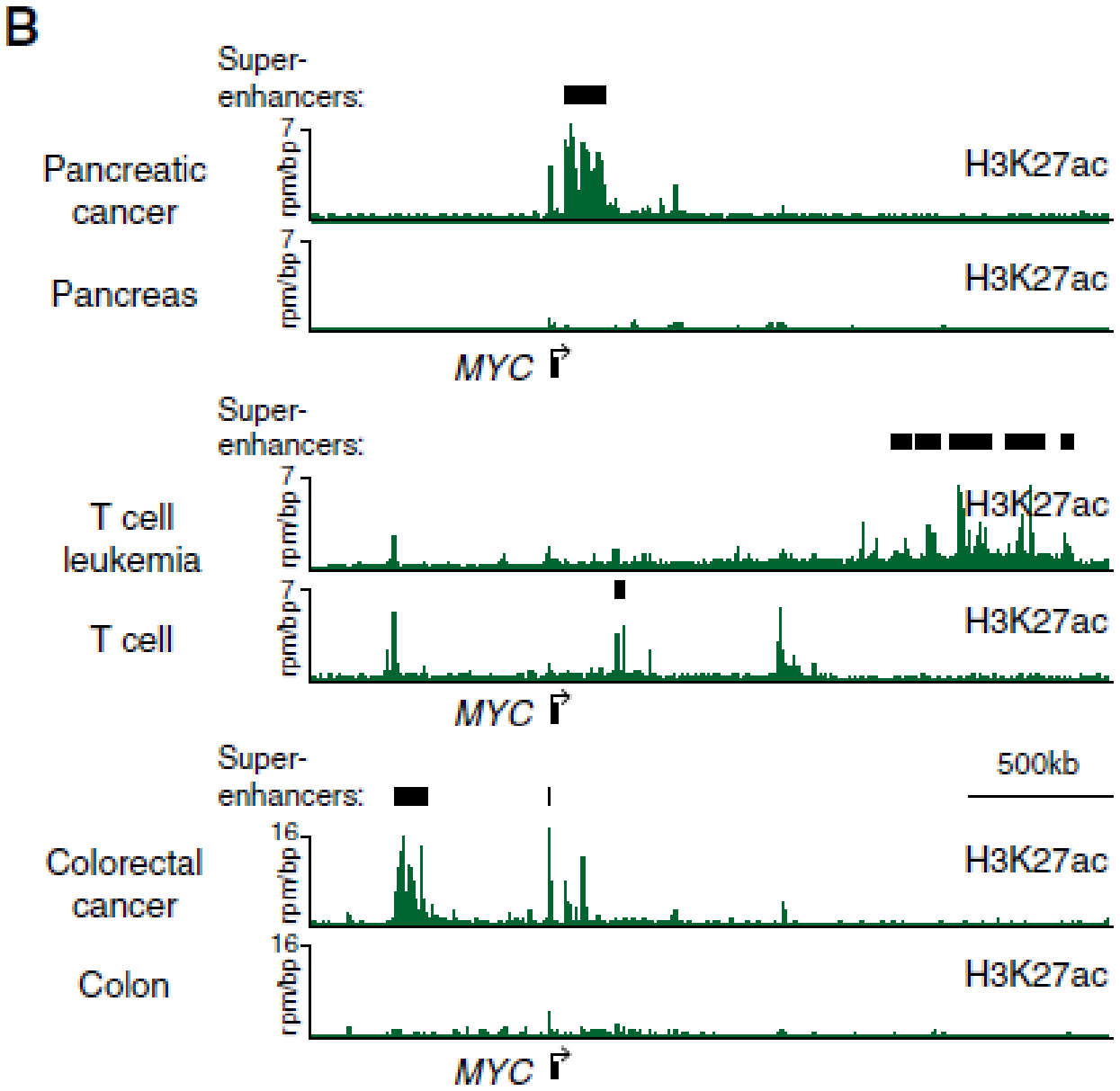


Super-enhancers in Cancer

Genes associated with SE and involved in cancer progression



The super-enhancers formed in the MYC locus were tumor type specific



Super-enhancers are associated with genes that act as hallmarks in colonrectal cancer