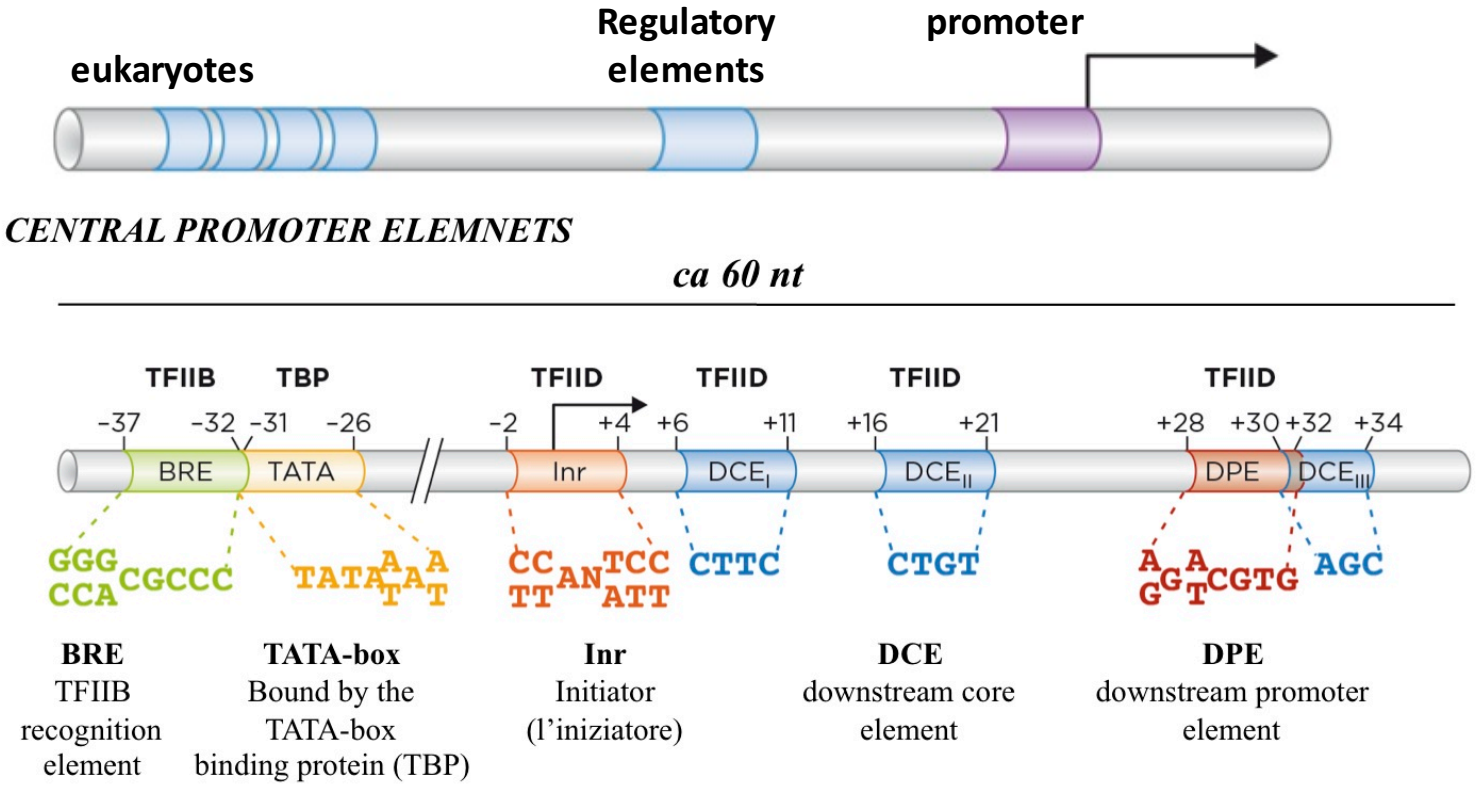
REGULATION OF GENE EXPRESSION BY ENHANCER RNAs - eRNAs

Central promoter elements control the activity of the basal apparatus of RNA Pol II at the promoter



BASAL TRANSCRIPTION FACTORS

BRE
TFIIB recognition element

TATA-box
Bound by the TATA-box binding protein (TBP)

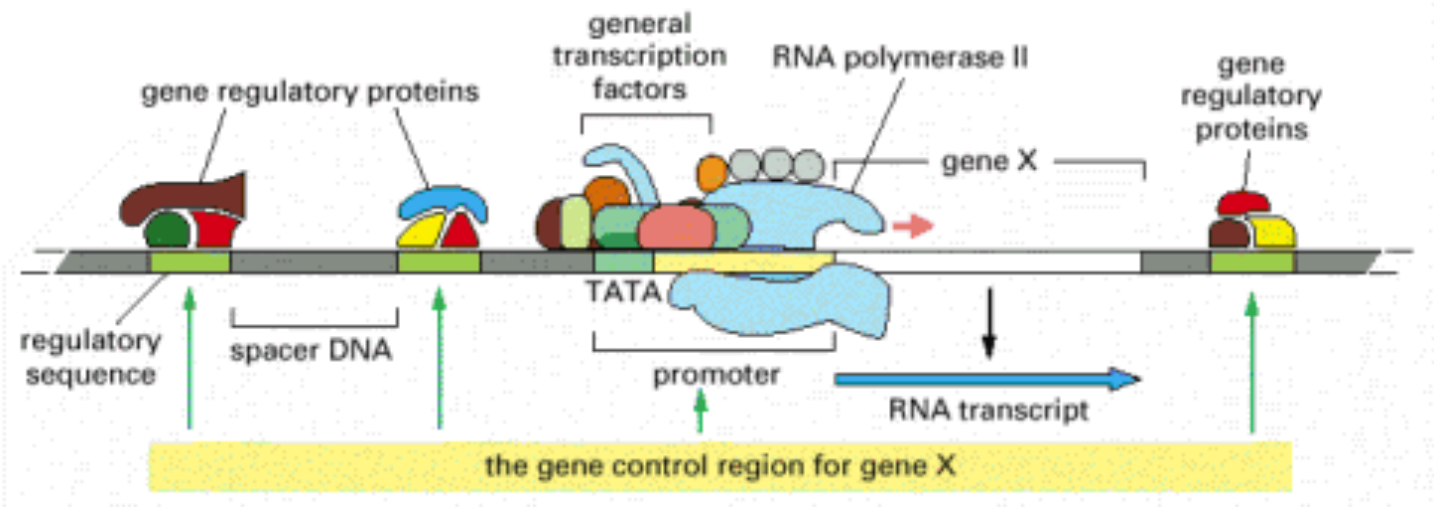
Inr
Initiator (l'iniziatore)

DCE
downstream core element

DPE
downstream promoter element

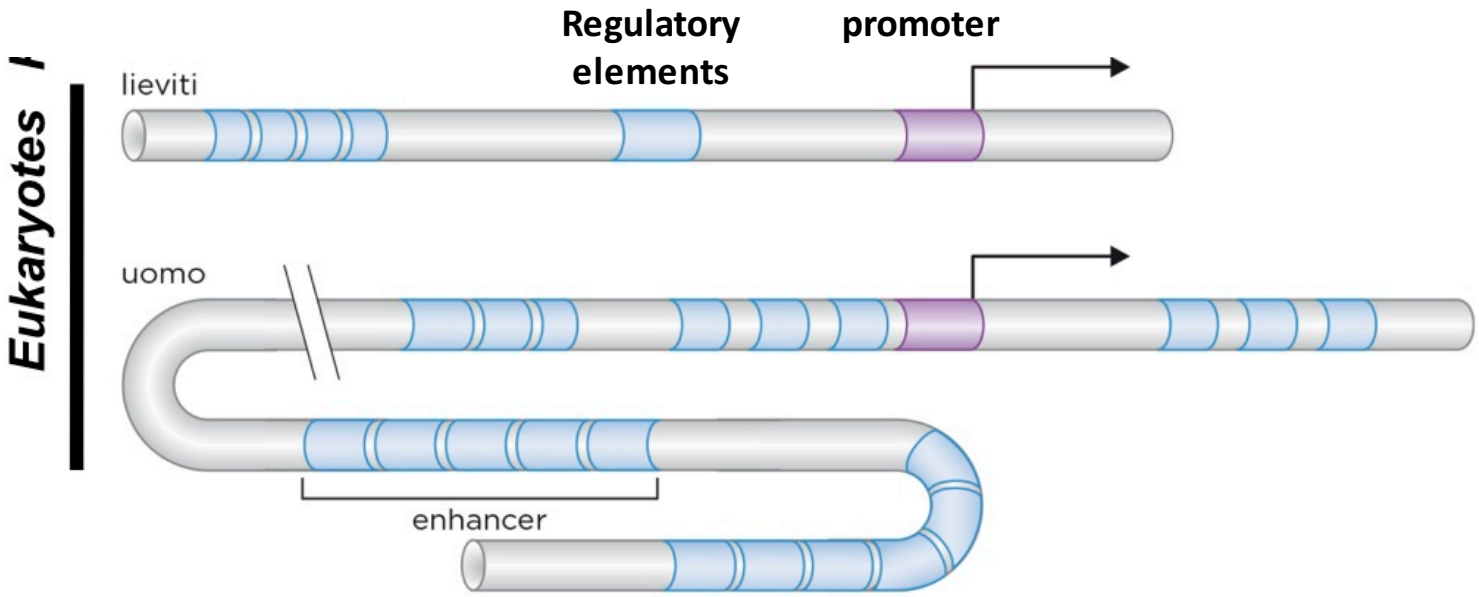
- Binding of **TFIID/TBP** to the TATA box is the first step in initiation.
- Other basal transcription factors bind to the complex in a **defined order**, extending the length of the protected region on DNA.
- When **RNA polymerase II** binds to the complex, it **initiates transcription**.

A complex interplay of regualtory sequences and transcription factors control the basal transcription complex



The gene control region of a typical eucaryotic gene. The ***promoter*** is the DNA sequence where the general transcription factors and the polymerase assemble. The ***regulatory sequences*** serve as **binding sites** for **gene regulatory proteins**, whose presence on the DNA affects the rate of transcription initiation. These sequences can be located **adjacent** to the promoter, far **upstream** of it, or even **within introns** or **downstream** of the gene. **DNA looping** is thought to allow gene regulatory proteins bound at any of these positions to interact with the proteins that assemble at the promoter. Whereas the **general transcription factors** that assemble at the promoter are **similar for all** polymerase II transcribed **genes**, the **gene regulatory proteins** and the **locations** of their binding sites relative to the promoter are **different for each gene**.

Additional regulatory elements increase complexity of gene regulation



Regulatory sequences also far away from the promoter (max 1 Mb)

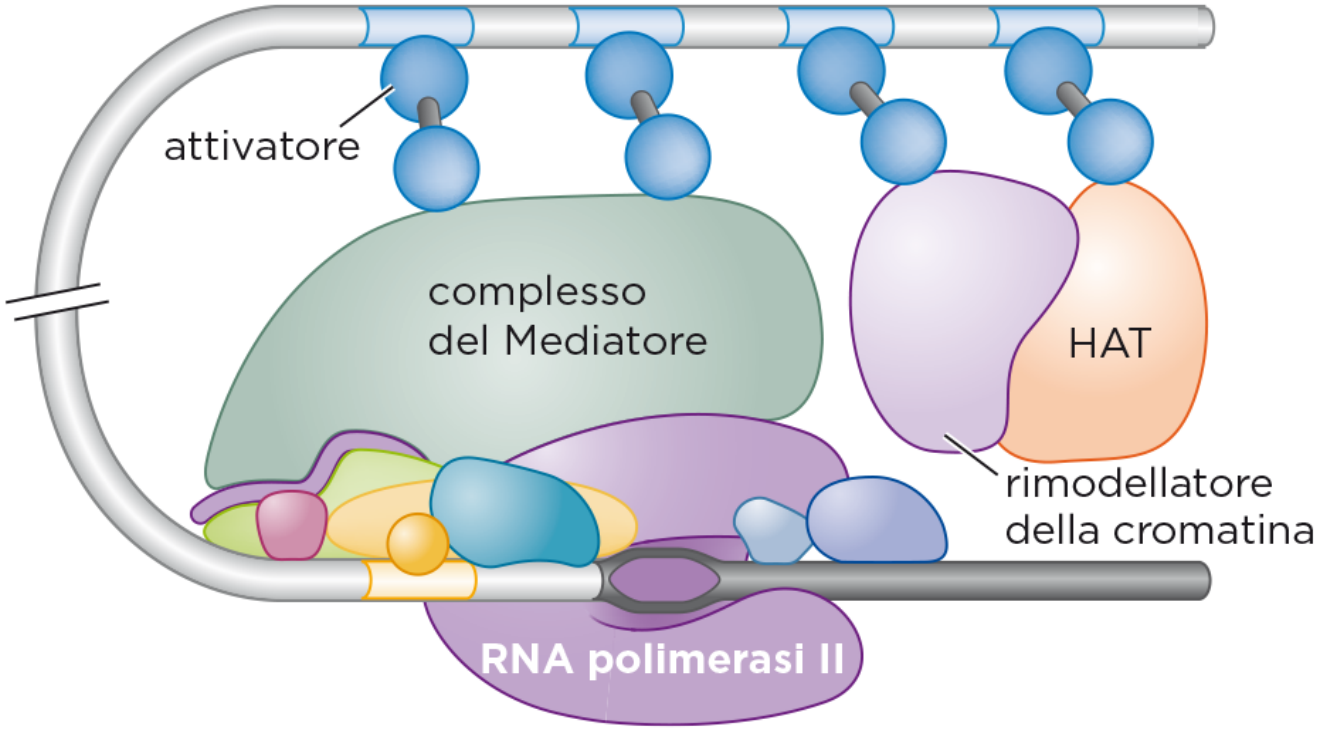
→ **Function of regulatory elements**

- **enhancer** (1 - X elements): regulate one gene in a particular moment and/or at different cell types and can respond to signals
- silencing elements: mediate the repression of a promoter
- insultors/bounday elements: sequenences that direct enhancer function to a particular gene

The Mediator complex

→ The mediator complex (<20 protein subunits) comunicates between **promoter and enhancer elements** (interconnects transcription factors)

→ **Essential for transcroptional activation**



LOOP FORMATION BRINGS ENHANCER ELEMENTS TO PROMOTER

→EFFICIENT ACTIVATOIN OF TRANSCRIPTION

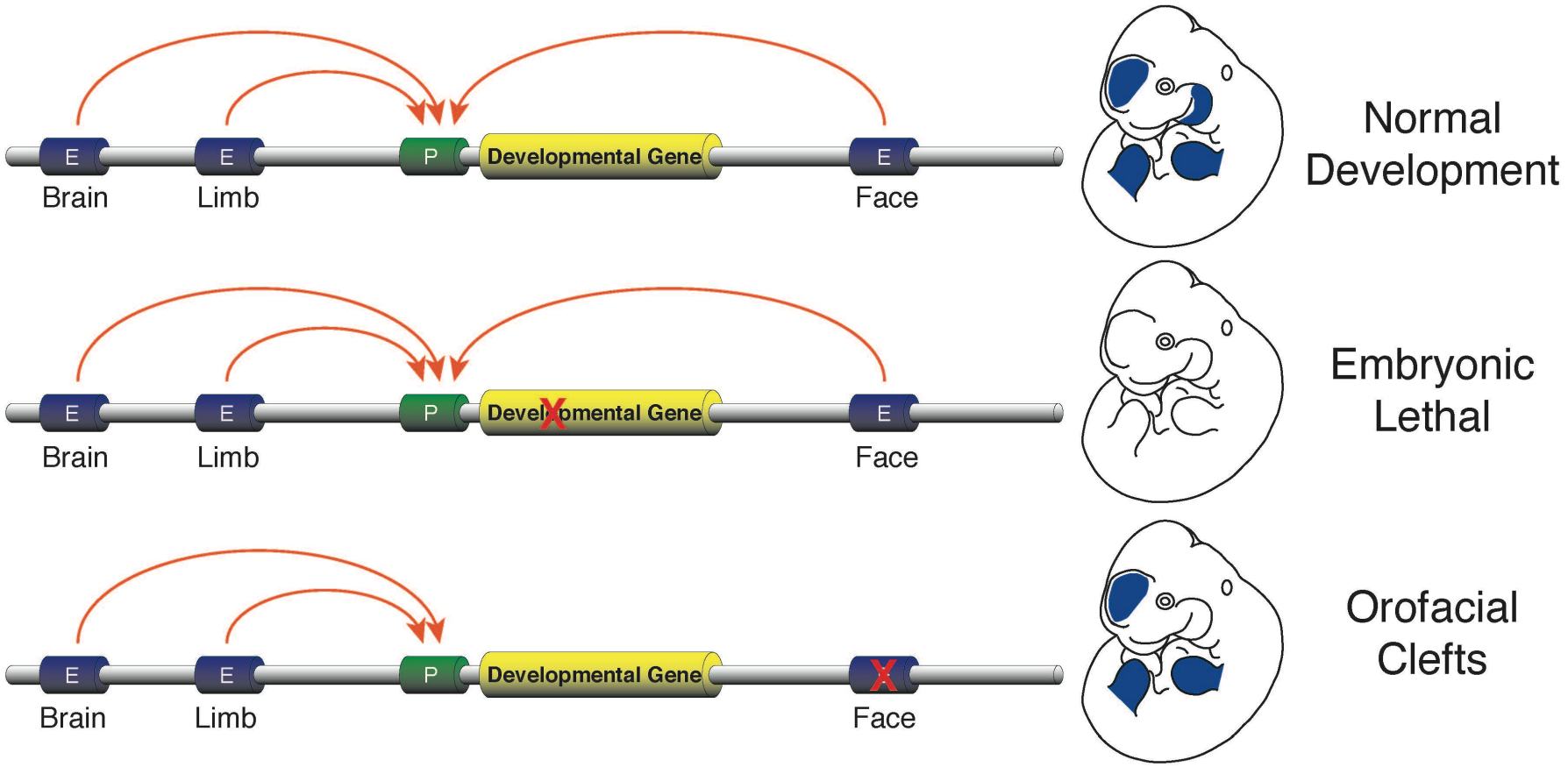
→LOOP IS FORMED BY COHESIN PROTEINS

ACTIVTY OF ENHANCER ELEMENTS IS REGULATED BY CHROMATIN STRUCTURE

Human Genome: 400.000 enhancers (up to 1.000.000)

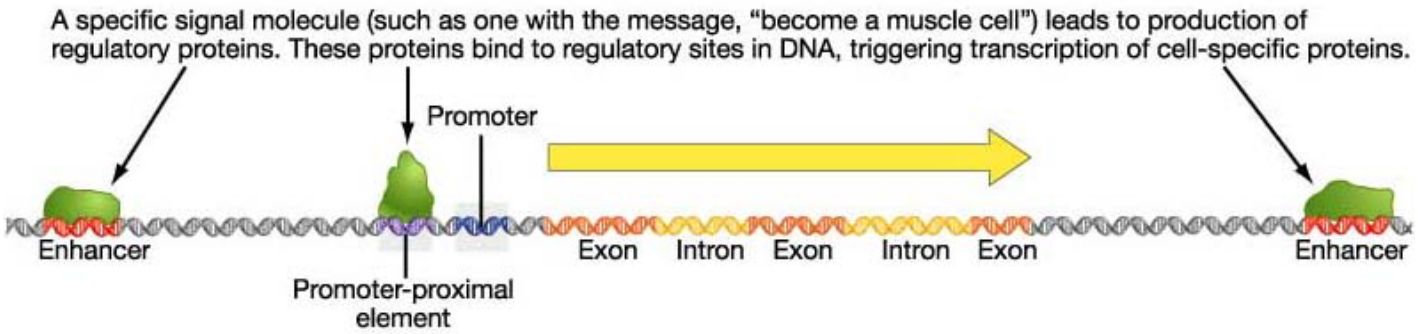
Distal cis-regulatory elements of transcription, such as enhancers, were first discovered in the simian virus SV40 genome in the early 1980s [6]. The inclusion of a 72bp SV40 DNA fragment upstream of the β-globin gene reporter increased its expression by 200-fold. Moreover, it was noticed that the viral enhancer remained active when placed in both sense and antisense orientation as well as upstream and downstream of the β-globin gene, thereby laying the foundation of basic enhancer features. Although it was found that the SV40 enhancer could influence the expression of β-globin over a distance of 10 kb, most enhancers in lower eukaryotes such as yeast were located within 100 – 200 bp from their target promoters [7]. Nevertheless, it seems that enhancers evolved their ability to mediate a long-range action in concomitance with the genome expansion of higher eukaryotes. Indeed, studies in drosophila and human cells have revealed that most enhancers are on average located at > 50 Kb away from their target regions [4,8]. For instance, it was shown in leukemia cells that Myc expression is regulated by a group of five enhancers located 1.7 Mb downstream of the proto-oncogene [9]. Recent development of high-throughput sequencing technologies in combination with techniques to probe the three-dimensional structure of the genome allowed the precise genome-wide mapping of enhancers and the initiation of global assessment of their functions and disease-related alterations.

Enhancers are central regulators of gene expression and development



Enhancers are bound by defined sets of transcription factors that are expressed in a tissue/cell specific manner

Enhancers



General features of enhancers:

- Positively drive target gene expression
- Functional independence of genomic distance and orientation relative to target gene promoter
- Hypersensitivity to DNase digestion
- De-compacted chromatin status
- Presence of specific sequences allowing the binding of transcription factors
- enriched binding of transcription co-factors
- Histone acetylation (p300)

Epigenetics of enhancers:

- **High ratio of H3K4me1/H3K4me3**
- **H3K27Ac**
- **Histone variants: H2AZ**
- **Transcription of small RNAs (eRNAs)**

Topology of DNA/Chromatin

- **Loop formation**

H3K4me1/me3 low, H3K27me3: INACTIVE ENHANCER

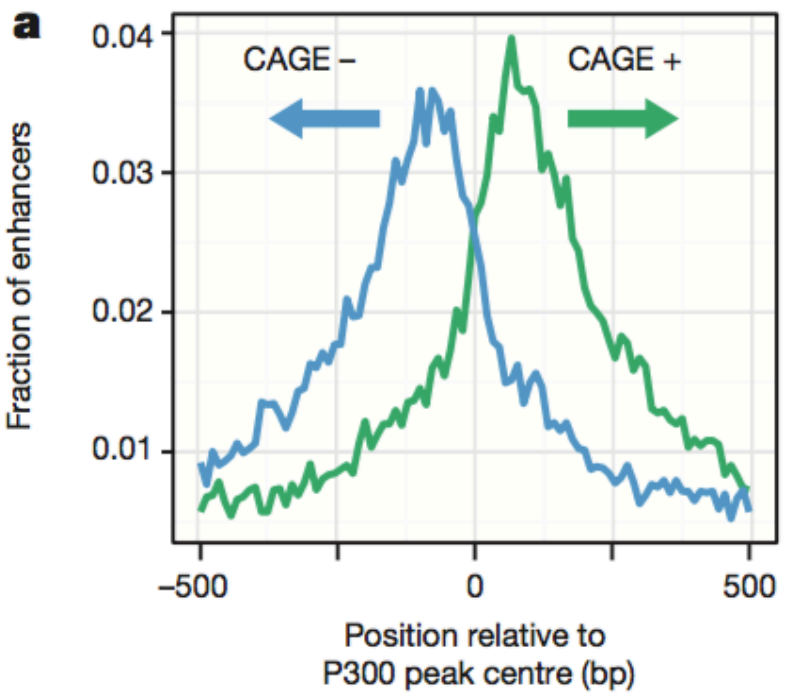
H3K4me1/me3 high, H3K27Ac, p300 HAT: ACTIVE ENHANCER

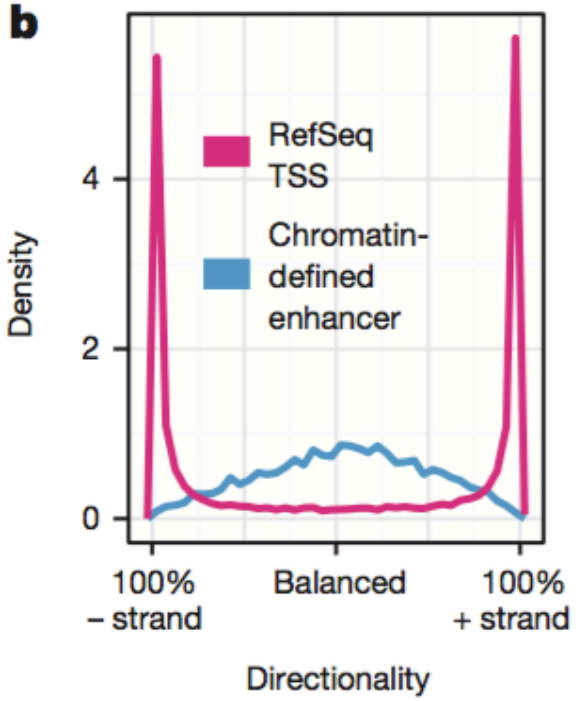
Mapping of transcripts on transcribed enhancers

- Selection of DNA elements that carry epigenetic code of enhancers (high H3K27Ac; high H3K4me1/me ratio; peak for p300 HAT

- CAGE sequencing of Hela cells

- Mapping CAGE tags in putative enhancer regions

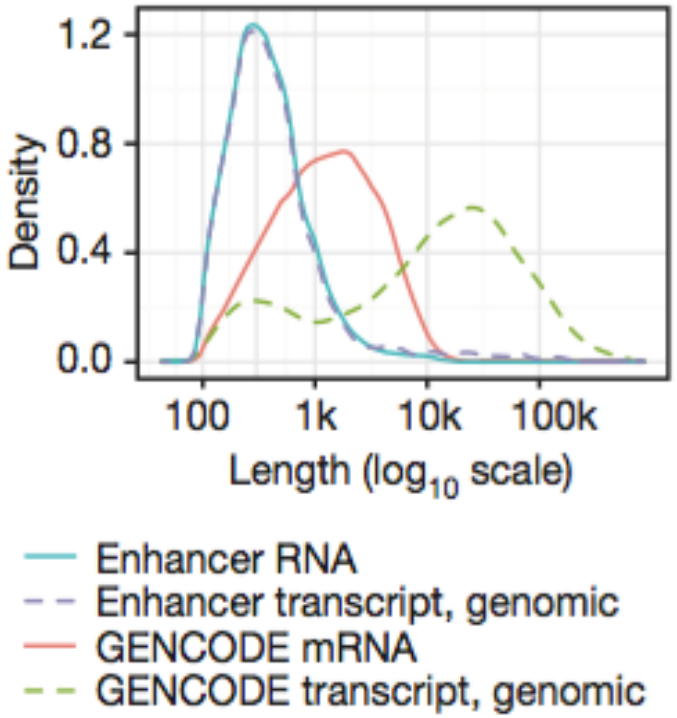




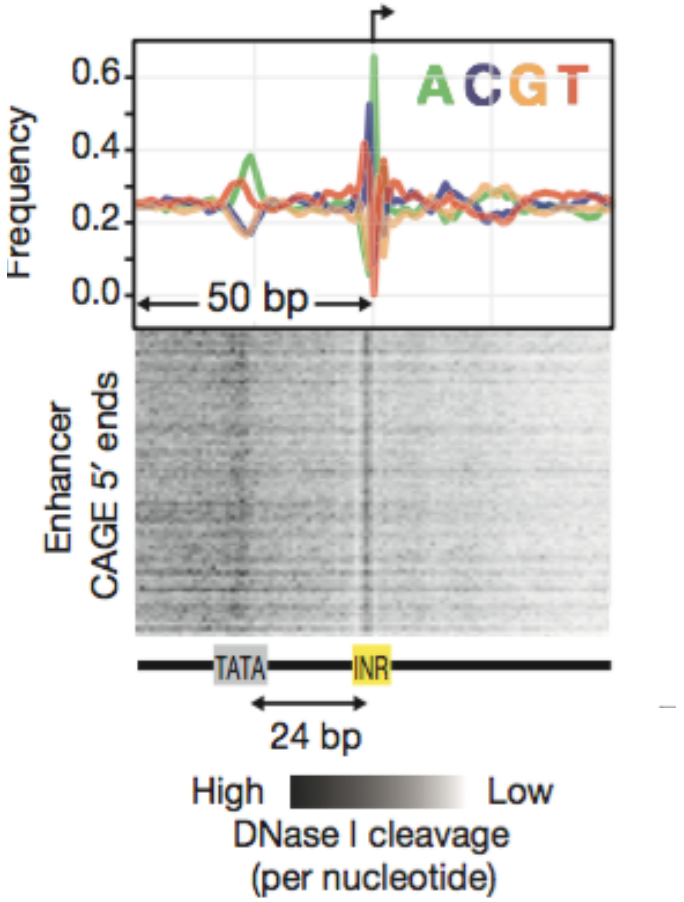
eRNA transcription:
can be directional, divergent
on + and - strand

eRNAs: balanced +/- strand usage
mRNAs: unbalanced -/+ strand usage

Mapping of transcribed enhancers



eRNAs are short (100-1000)



eRNAs transcriptional start site is well defined and TATA and INR site (initiation element) are sensitve to DNase digestion

→ ***Similar to classic Pol II promoter***

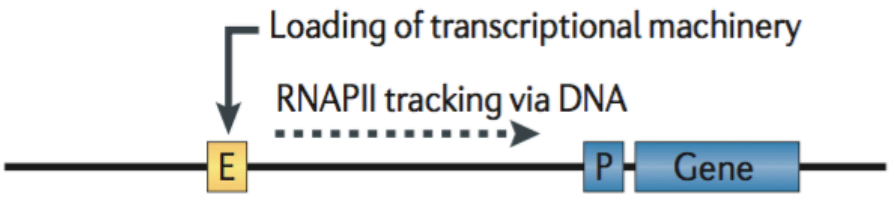
Mapping of transcribed enhancers

40.000 – 65.000 eRNAs in human cells

Features	eRNA	lncRNA	mRNA
DNase HS	Yes	Yes	Yes
H3K4me1	High	Medium	Low
H3K4me3	Low	Medium	High
H3K36me3	No	Yes	Yes/high
H3K27ac	High	High	High
RNA polymerase II (RNAPII)	Yes	Yes	Yes
RNAPII Tyr1p	High	Unclear	Low
RNAPII Ser2p	No	Yes	Yes/high
RNAPII Ser5p	Yes	Yes	Yes
RNAPII Ser7p	Yes	Unclear	Yes
CpG island	Low	Medium	High
Splicing	Rare	Common (2-exon bias)	Yes
Polyadenylation	Some	Mostly	Mostly
Stability	Low	Low to medium	High
Number*	~40,000–65,000	Several to tens of thousands	~23,000
Conservation	Low	Medium to high	High
Small RNAs	Yes	Unclear	Yes
Tissue specificity	Extremely high	High	Low
Preferential subcellular enrichment	Nuclear and chromatin-bound	Nuclear and chromatin-bound and cytoplasmic	Mostly cytoplasmic
Exosome targets	Yes	Partially yes	Mostly not

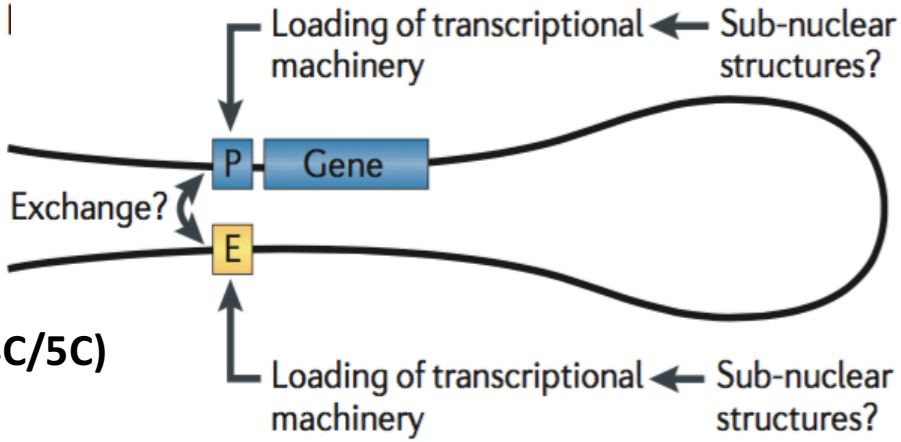
Models for enhancer function

Tracking model



Looping model

→ **Supported by chromatin conformation techinques (4C/5C)**



Common aim: Increase the concentration of transcriptional active machineries at cognate promoters

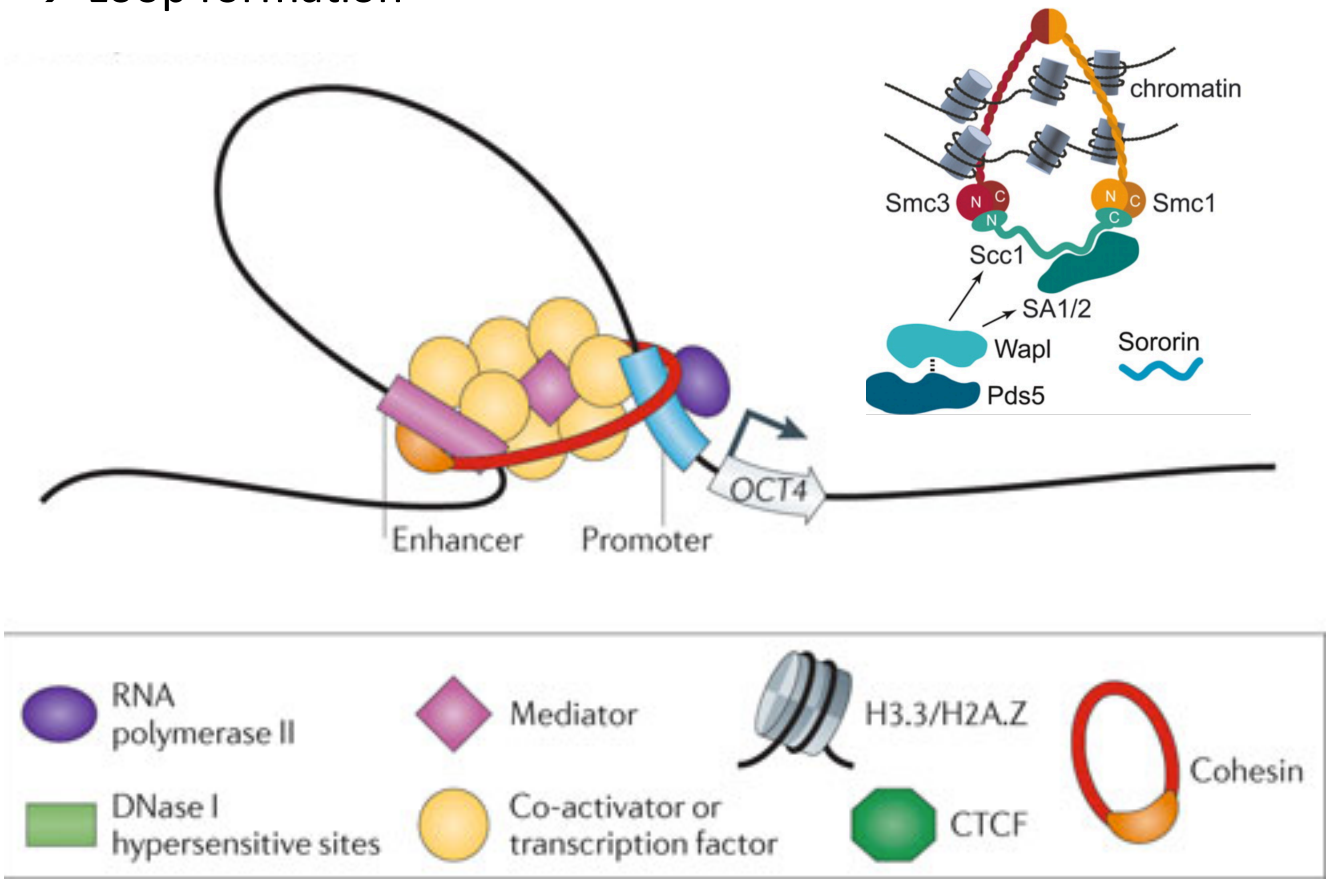
Tracking model: RNA polymerase II (RNAPII) and the associated transcriptional machinery track through the intervening DNA in-between enhancers and promoters

Looping model: Machinery is loaded at the enhancers and then reaches the promoter due to a physical interaction (that is, through looping)

4. Gene regulation by loop formation - enhancers and insulators

Interaction between Enhancer and Promoter

→ Loop formation



Association of enhancers and promoters at several genes in embryonic stem cells (ESCs) — for example, octamer-binding protein 4 (OCT4; also known as POU domain, class 5, transcription factor 1) is mediated by physical interactions between Mediator and cohesin complexes.

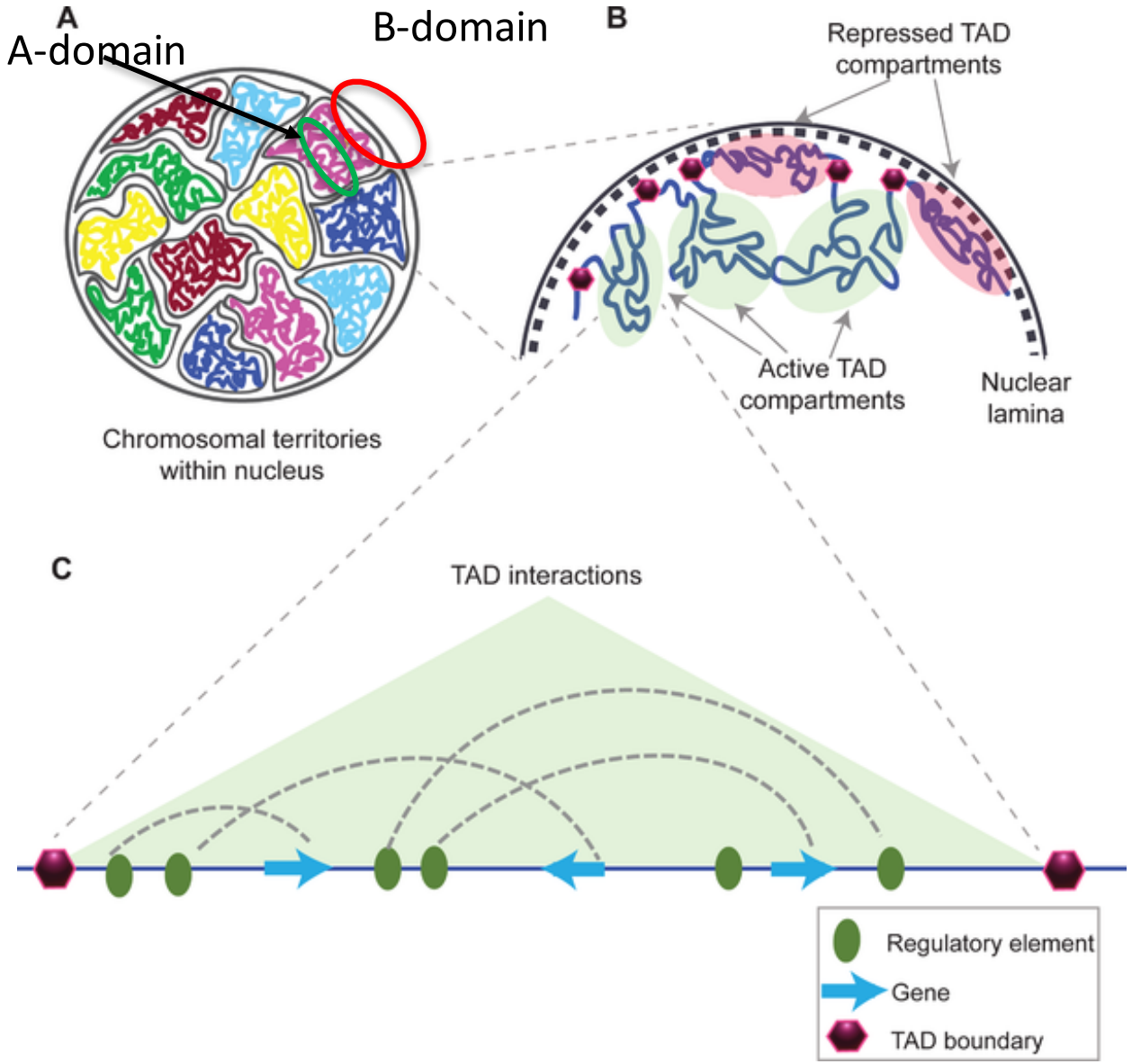
→cohesin can interact with the mediator complex
→Stable loop formation
→ activator proteins are localized at promoter/enhancer (note: in chromosome arms the major amount of cohesin is involved in intrachromosomic loops; only little amount holds sister chromatids together)
In contrast at centromeres high concentration of cohesin hold chromatids together

All chromatin interactions are created and maintained in a hierarchy of 3D chromatin architectures, including A/B domains, topologically associated domains. These findings indicate that any inter-relationships between enhancers and promoters, both as regulated and potentially regulatory transcription units, have to be considered in the context of a highly organized 3D genome.

4. Gene regulation by loop formation - enhancers - TADs

A topologically associating domain (TAD) is a self-interacting genomic region, meaning that DNA sequences within a TAD physically interact with each other more frequently than with sequences outside the TAD. These three-dimensional chromosome structures are present in animals as well as some plants, fungi, and bacteria. TADs can range in size from thousands to millions of DNA bases.The functions of TADs are not fully understood, but in some cases, disrupting TADs leads to disease because changing the 3D organization of the chromosome disrupts gene regulation. The mechanisms underlying TAD formation are also complex and not yet fully elucidated, though a number of protein complexes and DNA elements are associated with TAD boundaries.

The arrangement of chromosomes can determine their properties. Chromosomes are organised into two compartments labelled A ("active") and B ("inactive"), each with distinct properties. Moreover, entire chromosomes segregate into distinct regions called chromosome territories.



eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription

Carlos A. Melo,[1,2,8] Jamo Drost,[1,8] Patrick J. Wijchers,[3] Harmen van de Werken,[3] Elzo de Wit,[3] Joachim A.F. Oude Vrielink,[1] Ran Elkon,[1] Sónia A. Melo,[4] Nicolas Léveillé,[1] Raghu Kalluri,[5,6] Wouter de Laat,[3] and Reuven Agami[1,7,*]

[1]Division of Gene Regulation, the Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, the Netherlands
[2]Doctoral Programme in Biomedicine and Experimental Biology, Centre for Neuroscience and Cell Biology, 3004-517 Coimbra, Portugal
[3]Hubrecht Institute-KNAW and University Medical Centre Utrecht, Uppsalalaan 8, 3584CT Utrecht, the Netherlands
[4]Division of Matrix Biology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA
[5]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA
[6]Harvard-MIT Division of Health Sciences and Technology, Boston, MA 02215, USA
[7]Centre for Biomedical Genetics, 3584 CG, Utrecht, the Netherlands
[8]These authors contributed equally to this work
*Correspondence: r.agami@nki.nl
http://dx.doi.org/10.1016/j.molcel.2012.11.021

524 Molecular Cell 49, 524–535, February 7, 2013 ©2013 Elsevier Inc.

Does the tumorsuppressor p53 bind to active enhancers?

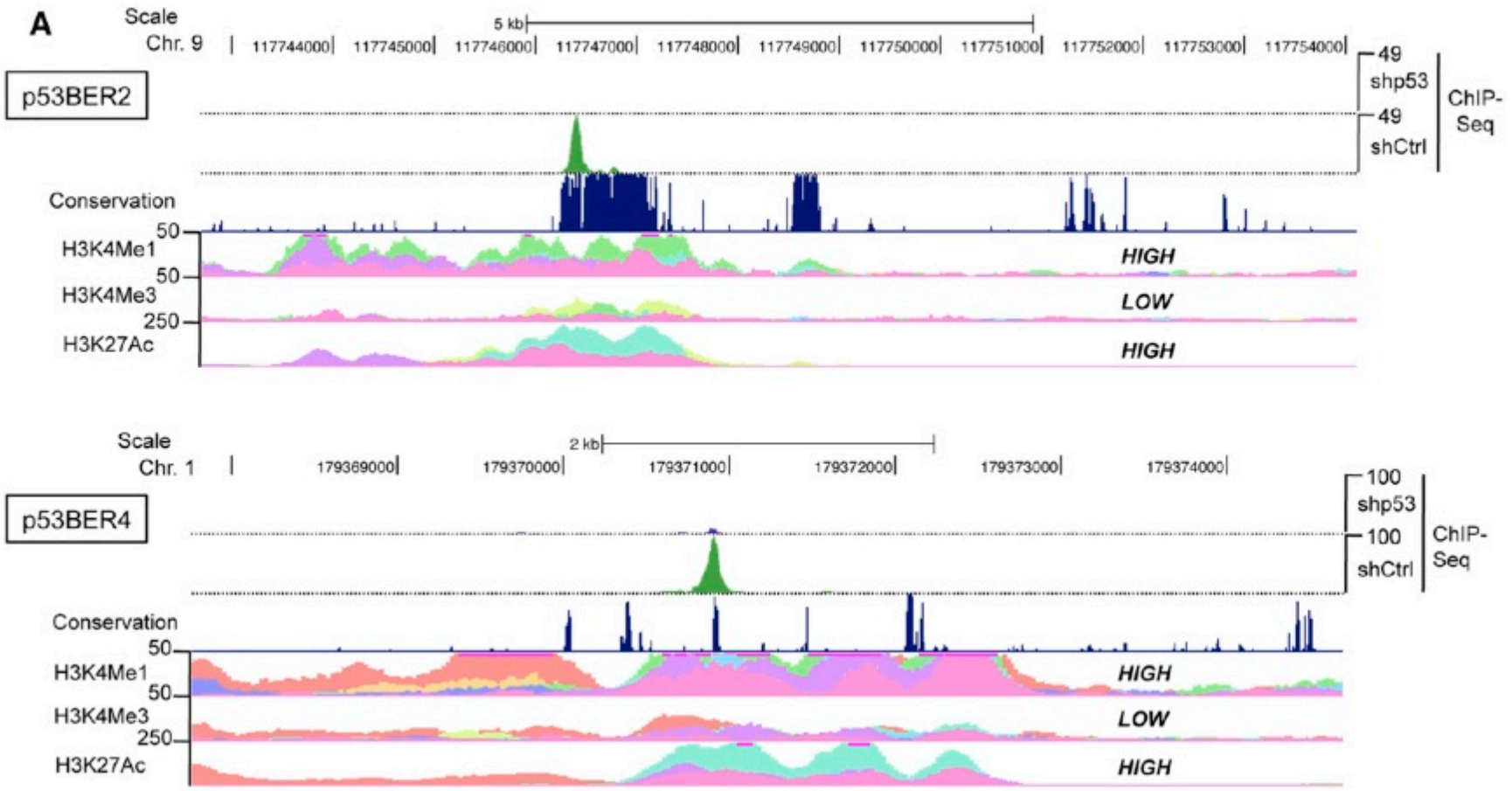
Step 1: Identify p53 binding sites (anti-p53 ChIP-seq)

Step 2: What of these sites are associated with marks of active enhancer histone modifications (ENCODE DATA): H3K4me1 high/ H3K4me3 low /high H3K27Ac

Step 3: How does p53 regulate the activity of enhancers

p53 binds to active enhancers in immortalized human cells

Representative enhancers, bound by p53 = p53BER (p53 binding enhancer regions)



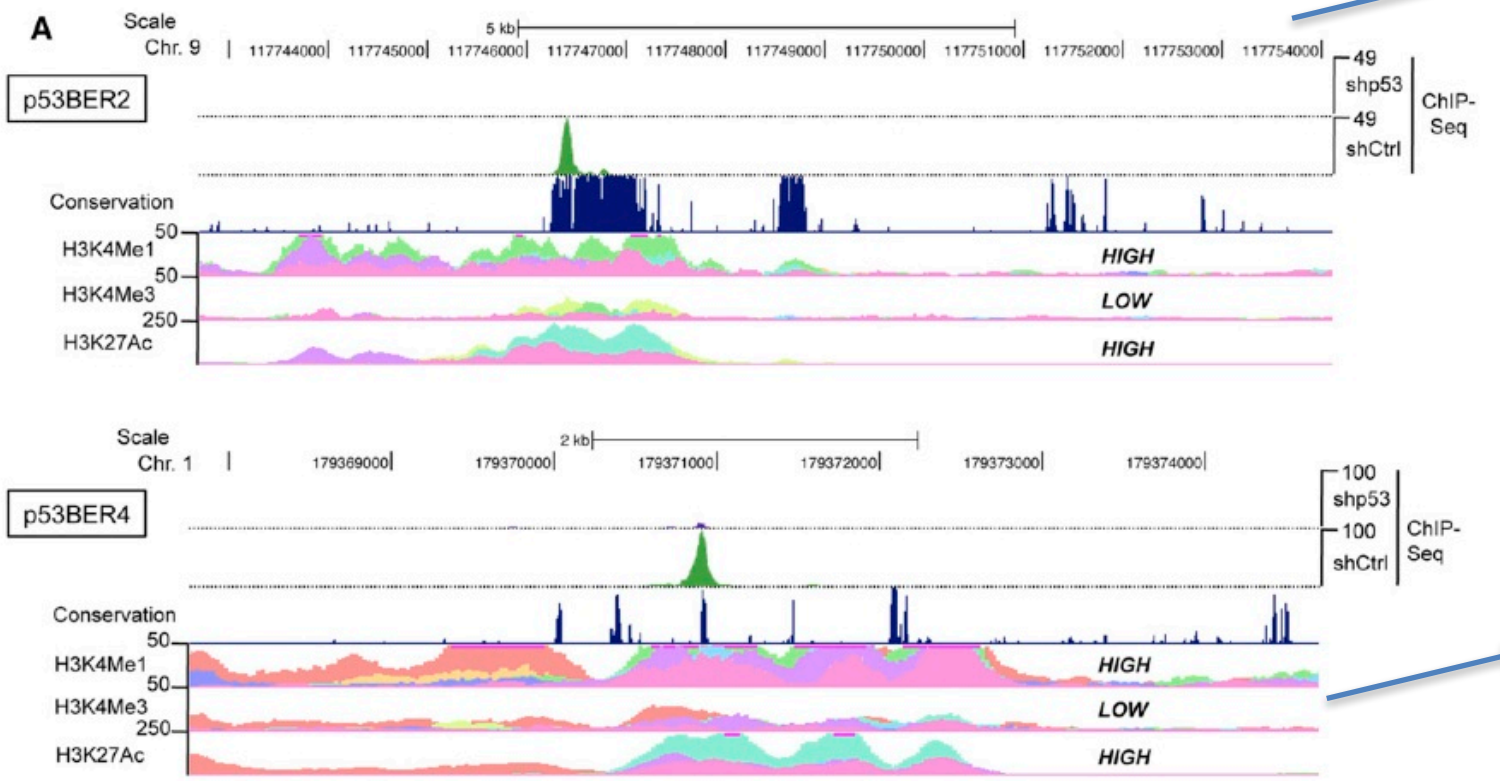
p53 binding site

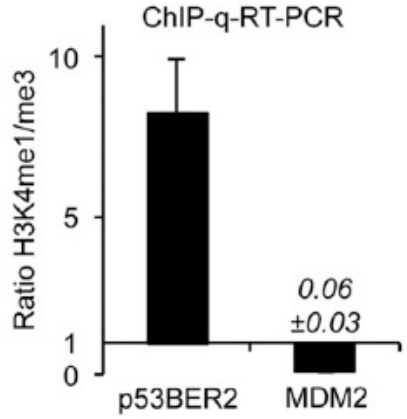
p53 binding site

ENHANCER: H3K4me1 high/ H3K4me3 low /high H3K27Ac

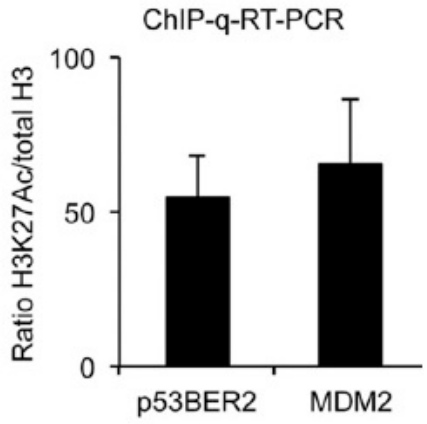
p53 binds to active enhancers in immortalized human cells

Validation of ChIP seq + ENCODE data by ChIP on precise regions of p53BER





C



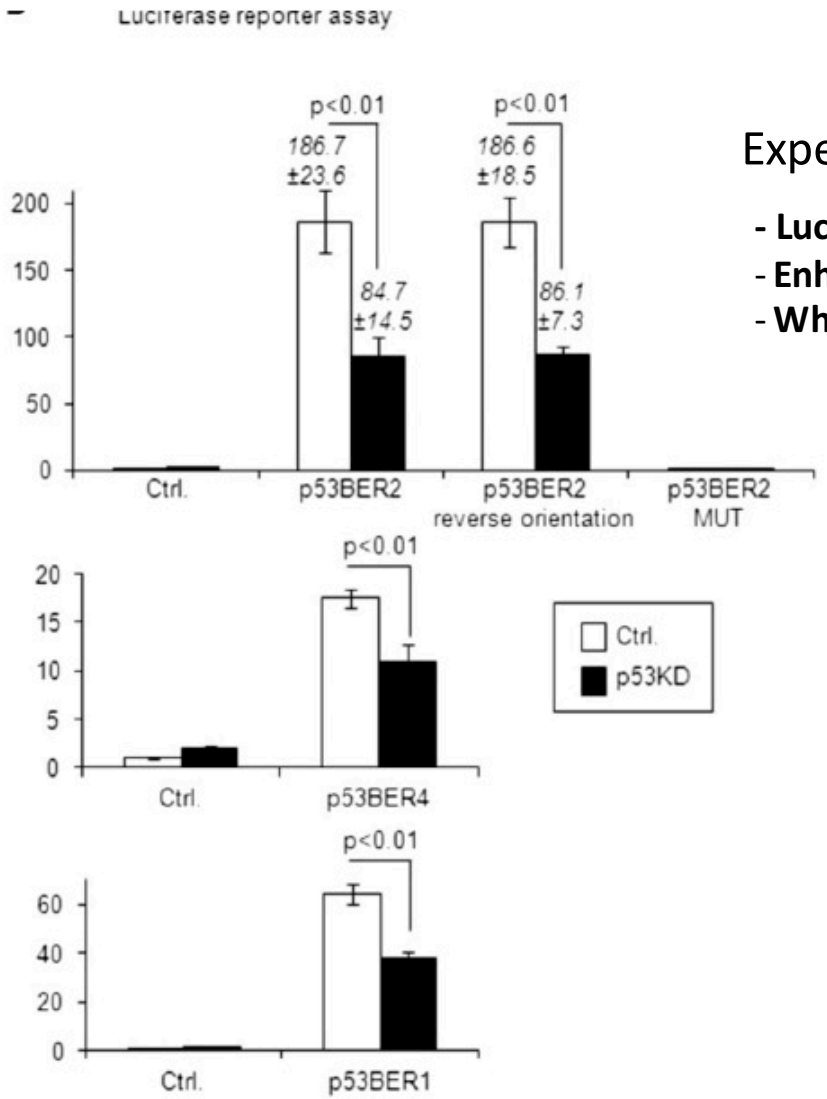
anti-p53 ChiP PCR amplification of p53BER2

q-RT-PCR: quantitative real-time PCR

p53 binds to active enhancers in immortalized human cells

Enhancer activity depends on p53 binding to the enhancer

luciferase reporter

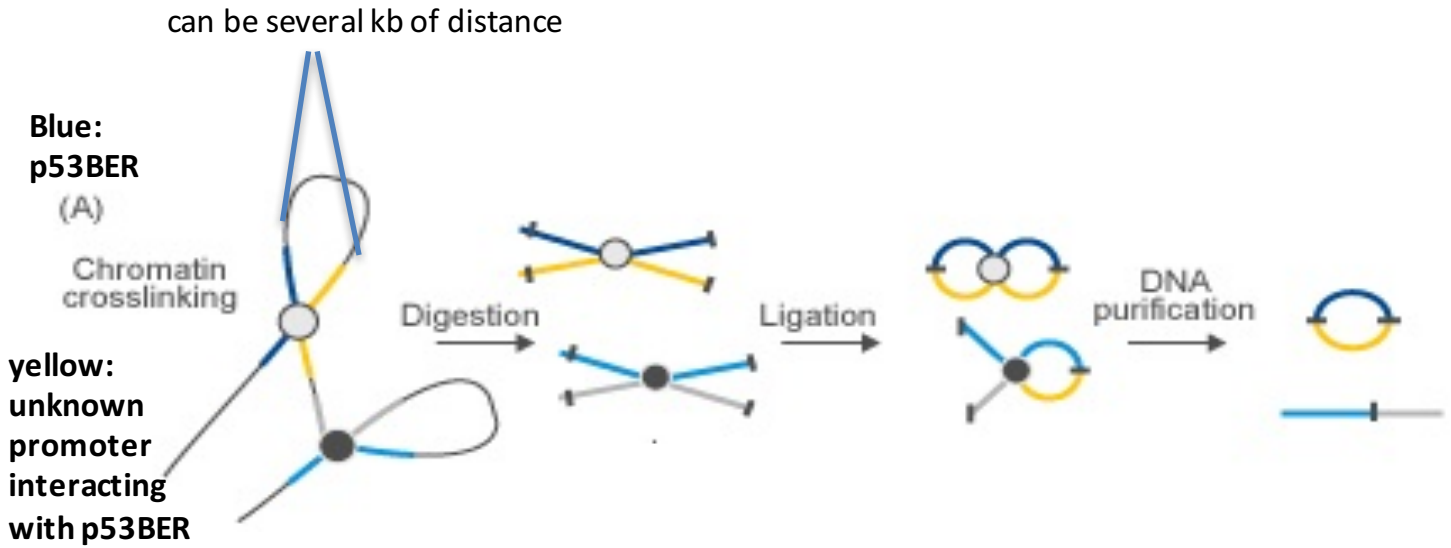


Experiments in cells with p53 expression

- **Luciferase reporter reduced when p53 RNAi (black bars)**
- **Enhancer works in both orientations**
- **When p53 binding site destroyed enhancer has reduced activity**

What are the genes regulated by p53 binding site containing enhancers ??

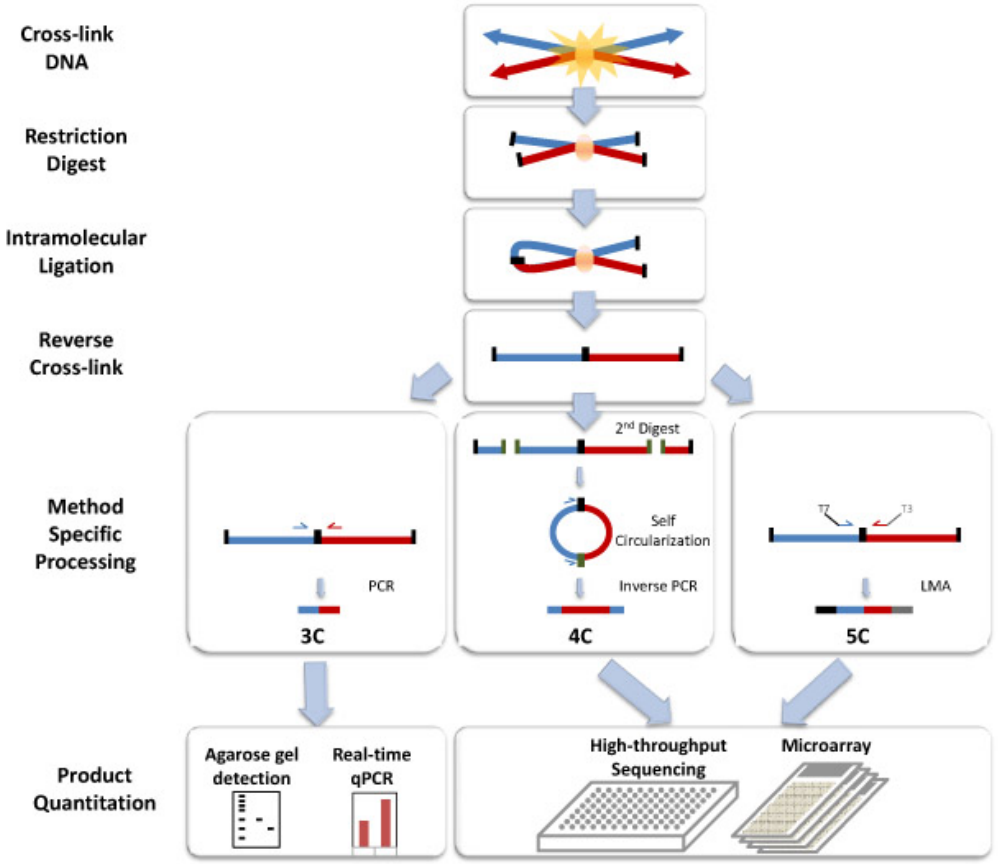
Circularized Chromosome Conformation Capture (4C)



3C analyzes excised DNA fragments generated from formaldehyde cross-linked, then restriction enzyme digested chromatin to find points where selected DNA regions are connected through a protein complex. The frequency and identity of these fragments are then determined by site specific PCR.

4C enables identification of previously unknown DNA regions that interact with a locus of interest, which makes 4C especially well suited to discover novel interactions with a specific region that is being investigated

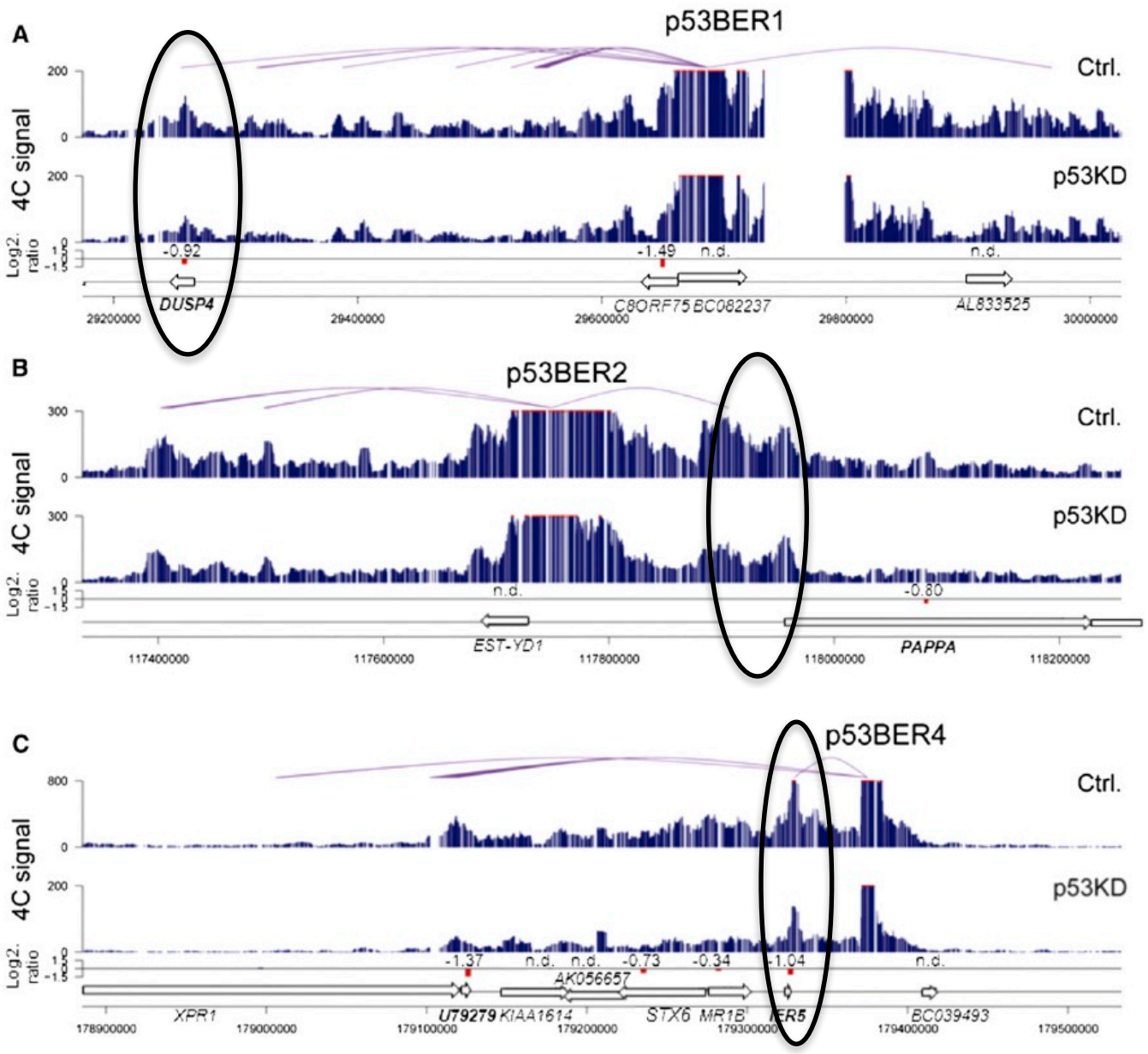
Circularized Chromosome Conformation Capture (4C)



3C (one-vs-one): The chromosome conformation capture (3C) experiment quantifies interactions between a single pair of genomic loci. For example, 3C can be used to test a candidate promoter-enhancer interaction. Ligated fragments are detected using PCR with known primers.

4C (one-vs-all): Chromosome conformation capture-on-chip (4C) captures interactions between one locus and all other genomic loci. It involves a second ligation step, to create self-circularized DNA fragments, which are used to perform inverse PCR. Inverse PCR allows the known sequence to be used to amplify the unknown sequence ligated to it. In contrast to 3C and 5C, the 4C technique does not require the prior knowledge of both interacting chromosomal regions. Results obtained using 4C are highly reproducible with most of the interactions that are detected between regions proximal to one another. On a single microarray, approximately a million interactions can be analyzed.

p53BER interact with several genes



NOTE:
Sequencing identifies unknown sequences ligated to the p53BER

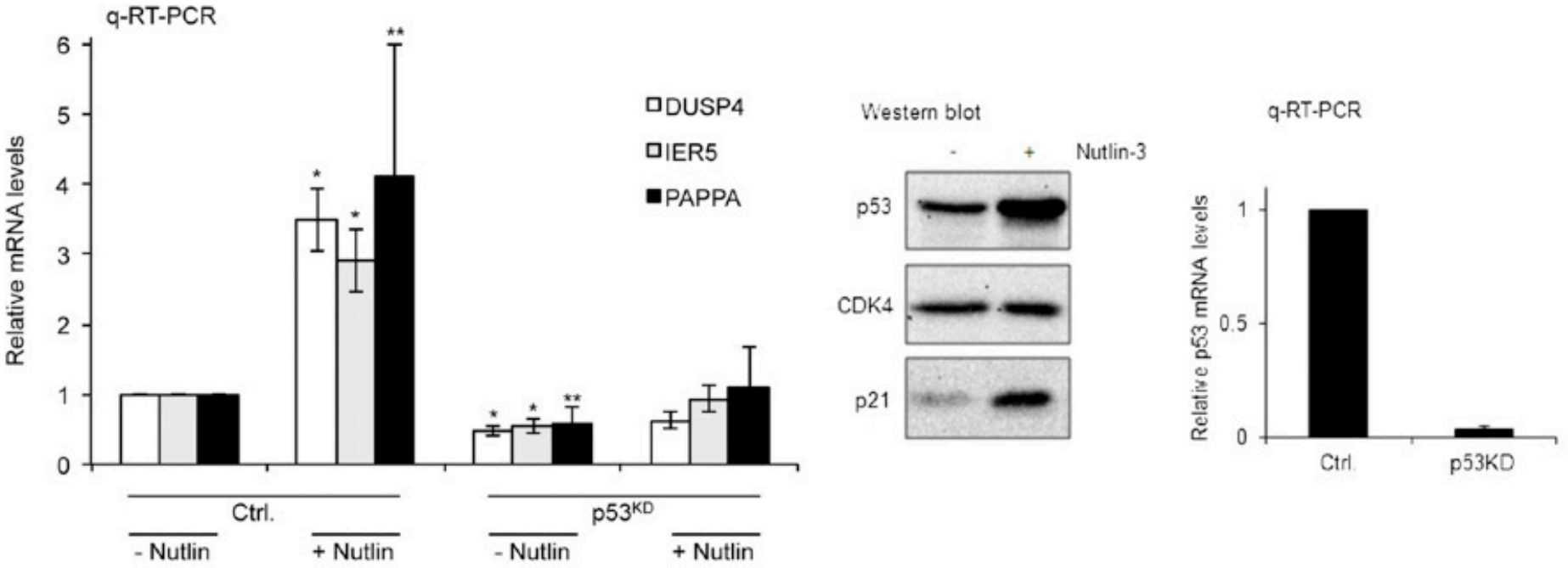
p53BER: shows highest Sequencing counts (Y-axis). Sequences that are ligated often with p53BER sequences show high sequence counts →in cells these sequences interact with the p53BER

IMPORTANT:
Interaction is reduced when p53 is knocked down!!
HOWEVER: interaction is still present!!!

p53 is important for activating genes that interact with p53BER

Nutlin is a compound that stabilizes p53 protein = p53 levels increases

Genes regulated by p53BERs: DUSP4, IER5, PAPPA



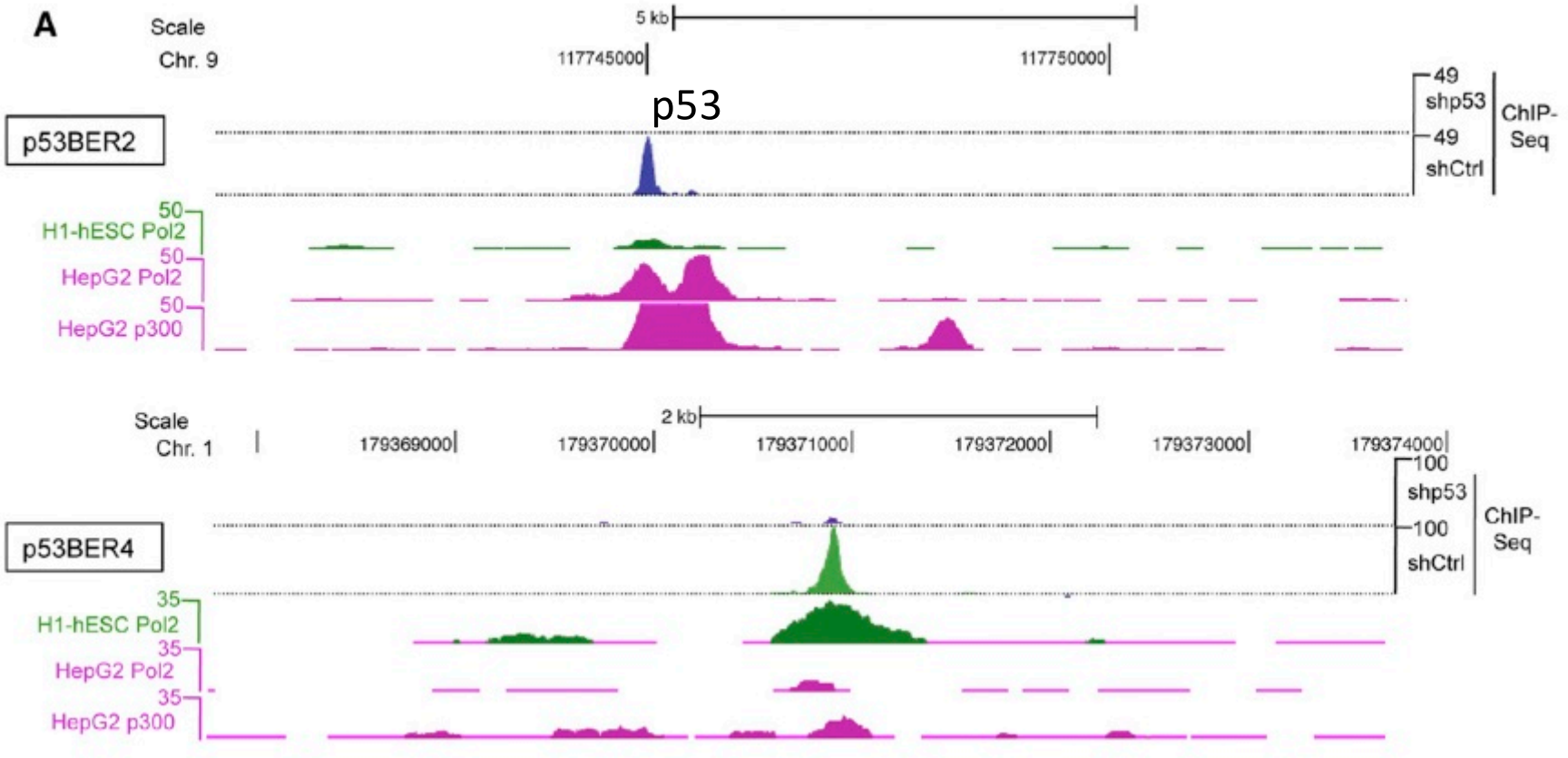
p53 increase results in a increase of genes that interact with p53BERs

p53BER associate with RNA Polymerase II

Question: does p53 cause the transcription of p53BERs = eRNAs ???

Enhancers with eRNAs:

- Max 2000 nt
- Transcribed by RNA Pol II
- Not polyadenylated
- High H3Kme1; low H3K4me3; high H3K27Ac
- eRNA loci bound by RNA Pol II and p300/CBP



p53 peaks overlap with p300 and RNA Pol II peaks (ENCODE project

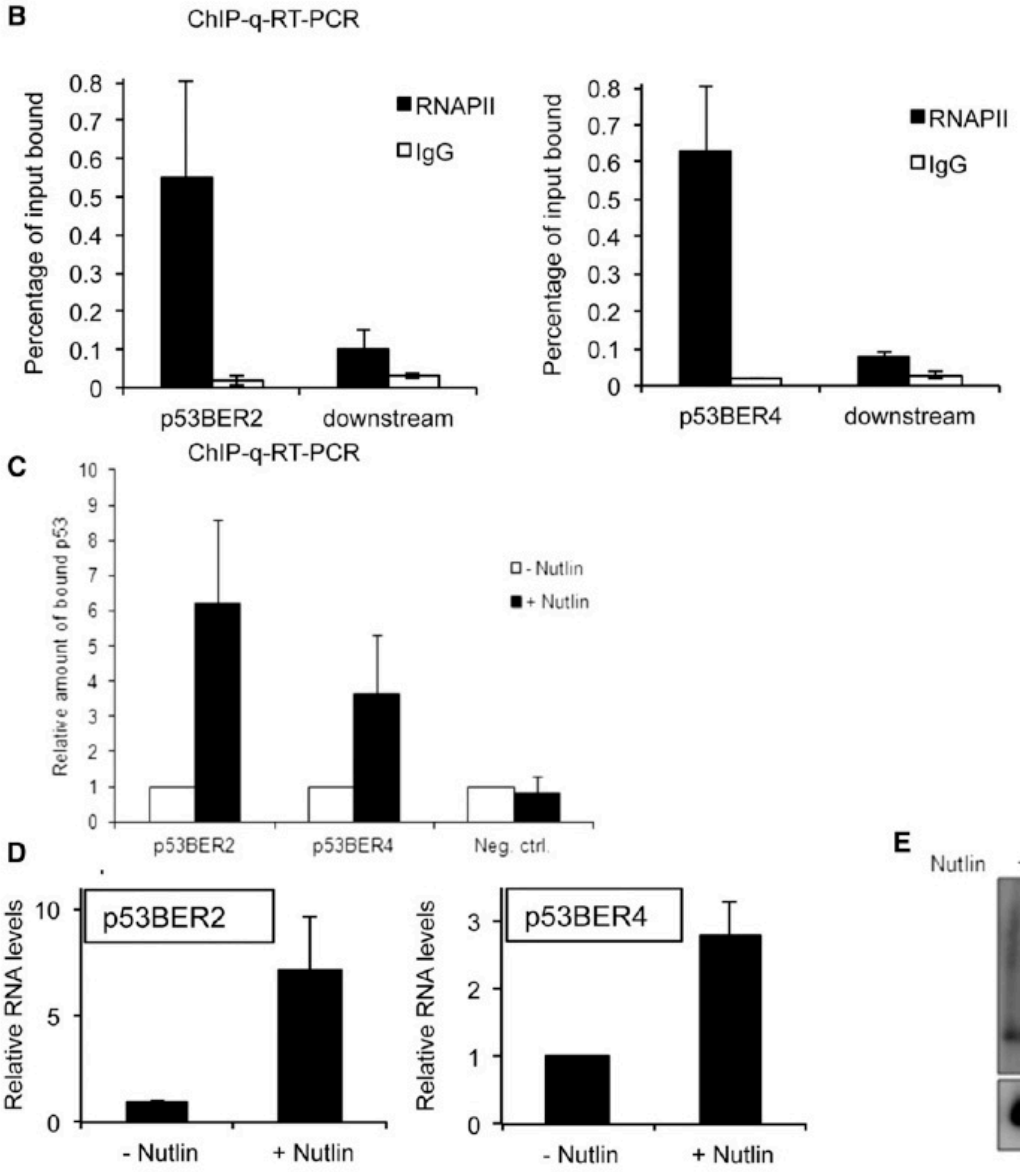
p53 peaks overlap with p300 and RNA Pol II peaks (ENCODE project

p53 stimulates eRNA expression at p53BERs

Anti-RNA Pol II ChIP confirms that RNA Pol II is located on p53BER

Anti-p53 ChIP
Increasing p53
by Nutlin treatment
increases p53
at p53BERs

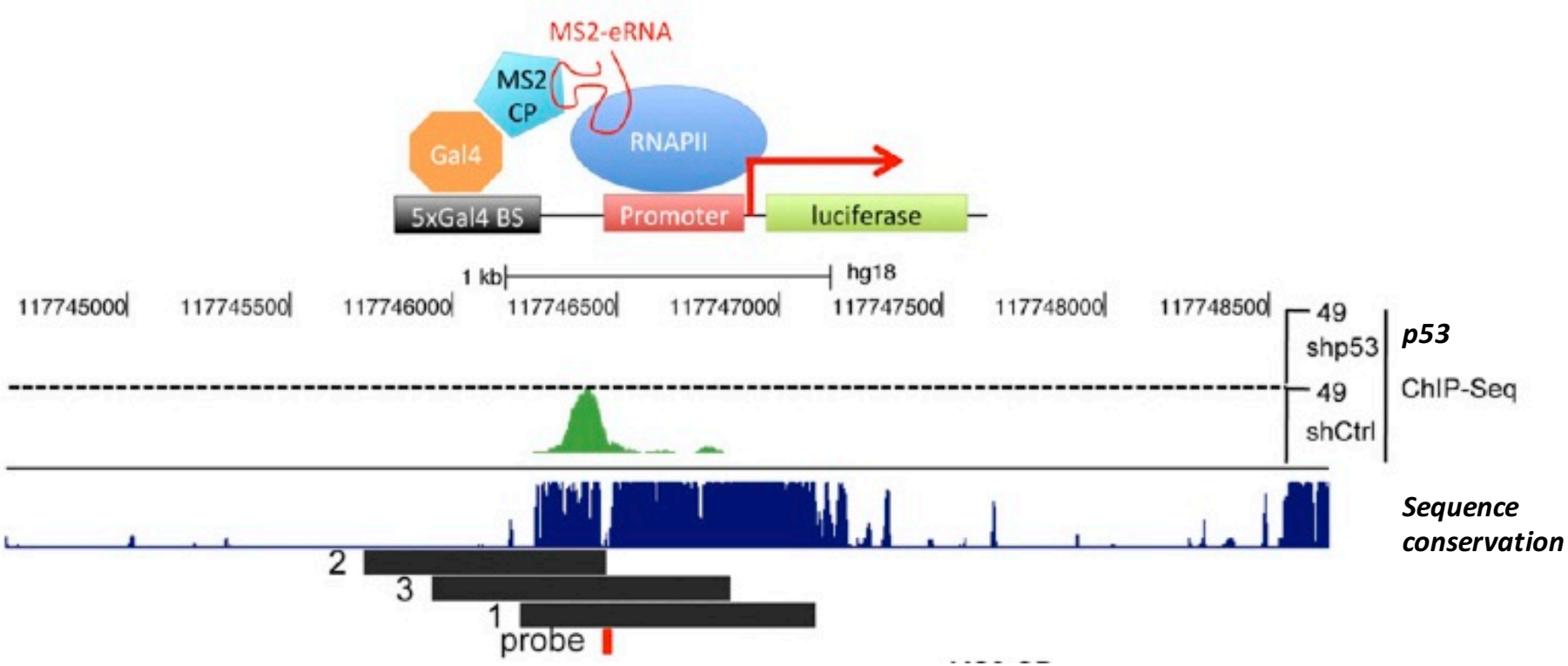
Quantitative RT-PCR:
Increasing p53
by Nutlin treatment
Increases RNA from
p53BERs !!!
→ eRNA production



Northern blot

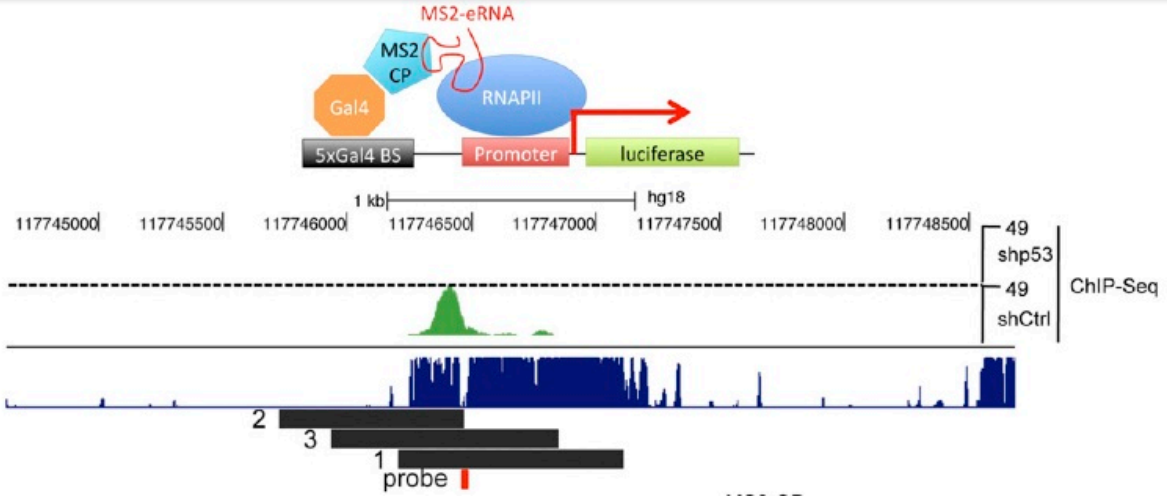
Re-building p53BER function in vitro

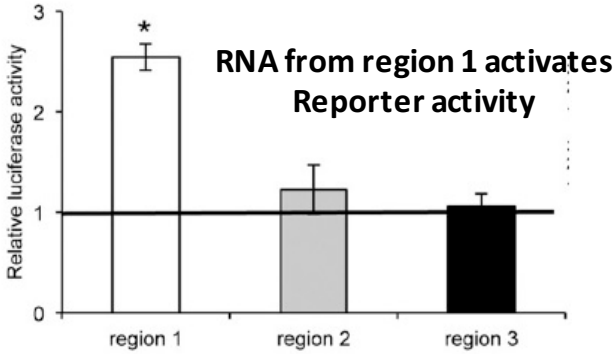
1. Create vectors that contains p53BER elements. This regions 1, 2, 3 are tagged to the MS2 RNA-stem loop motif
2. Express a MS2 coat protein that is fused with a GAL4 DNA binding domain
3. Make a luciferase reporter that carries 5 repeats of the DNA sequence bound by MS2-GAL4



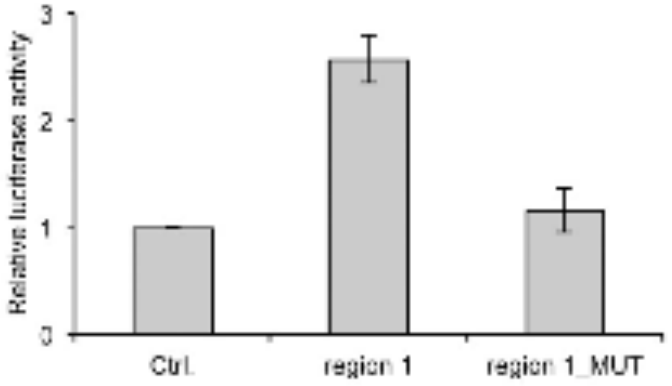
Idea: p53BER eRNA connects the GAL4 domain with the promoter and stimulates transcription of the reporter

Re-building p53BER function in vitro

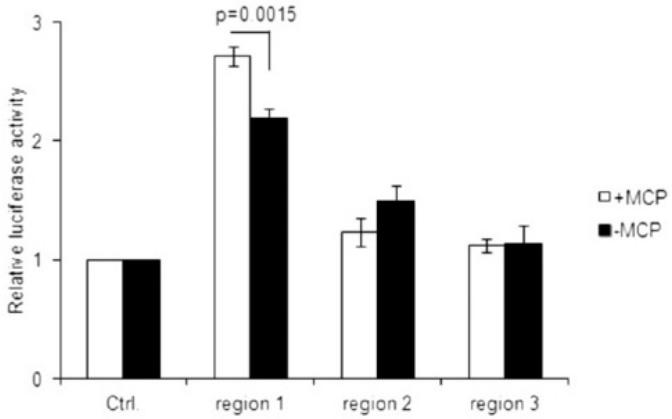




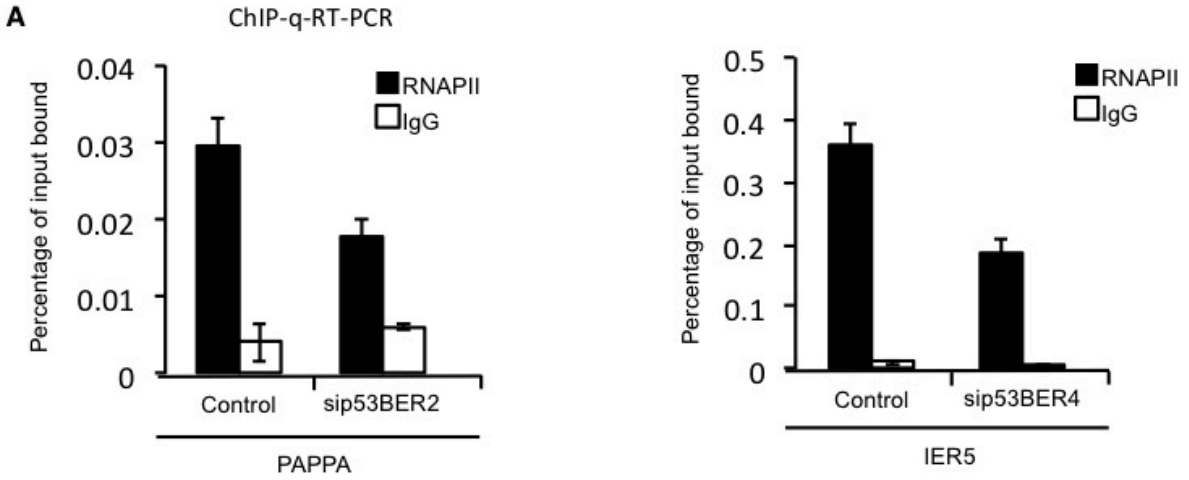
When p53 binding sites in Region 1 are mutated reporter activity is reduced
Why: Region 1 BER element is transcriptional inactive due to lack of p53 binding



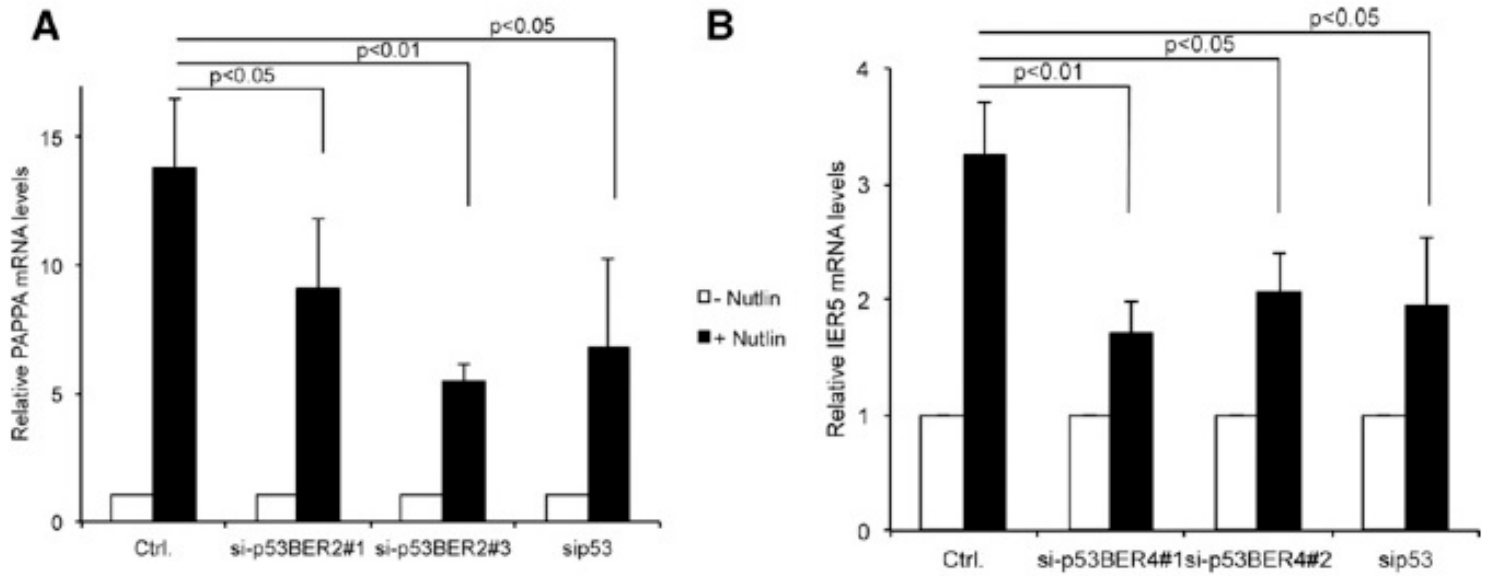
White: cells transfected with reporter, p53BER-MS2 RNA, MS2-GAL4
Black: cells transfected with reporter, MS2-GAL4
→→ p53BER-RNA component is required for the stimulation of the reporter



LOSS OF p53BER eRNA EXPERIMENTS



Anti-Pol II ChIP on Nutlin treated cells (more p53) → Knock down of p53BER eRNAs result in reduced abundance of RNA Pol II at genes that interact with the p53BER (PAPPA, IER5)



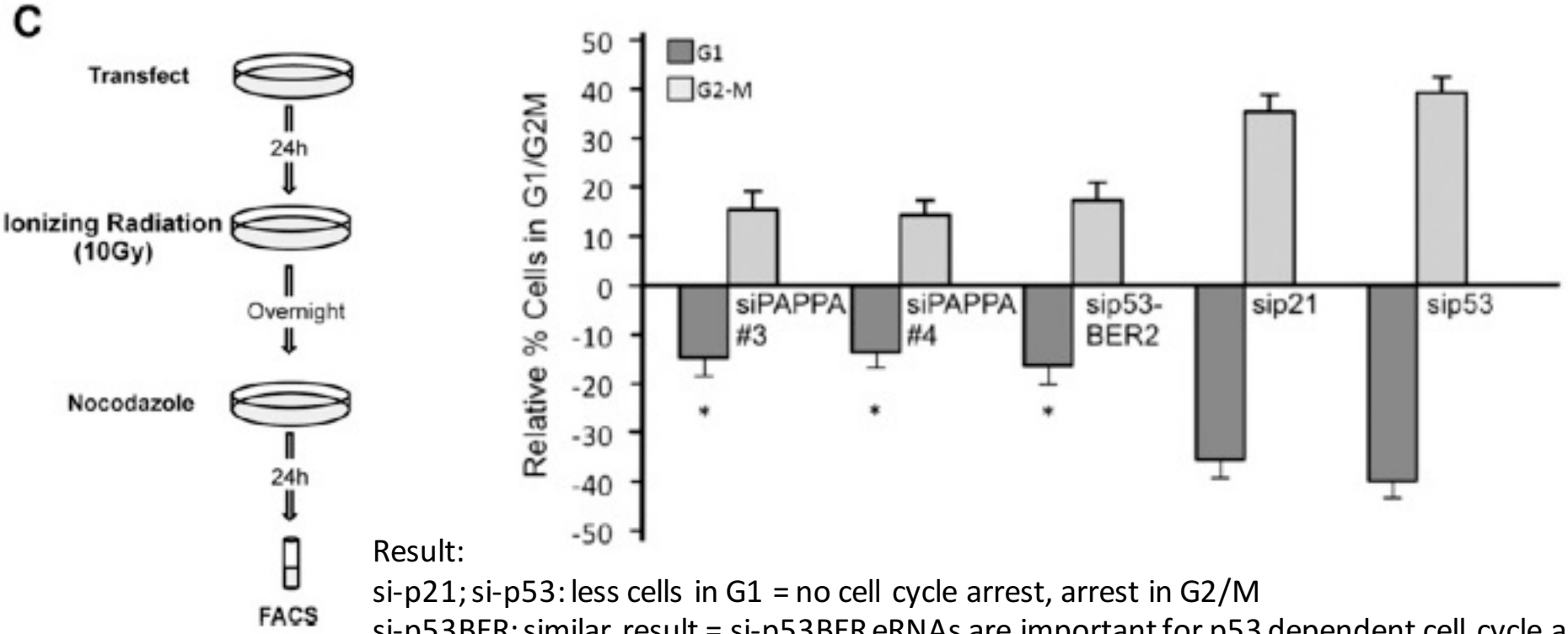
Knock down of p53BER eRNAs result in reduced expression of genes that interact with the p53BER (PAPPA, IER5)

PUTTING p53BER eRNA IN A PHYSIOLOGIC CONTEXT – p53 DEPENDENT DNA DAMAGE RESPONSE

Gamma irradiation → DNA damage → DNA damage signaling → Activation of **p53**, p21 → ***eRNA*** → **Cell cycle Arrest** at G1/S; G2/M

Experiment: si-p53BER eRNA → DNA damage → Nocodazole (microtubule inhibitor; arrest in M) –> count cells in G1/S/G2/M phase (=FACS)

Question: how many cells can still arrest at G1/S check point?



Result:

si-p21; si-p53: less cells in G1 = no cell cycle arrest, arrest in G2/M

si-p53BER: similar result = si-p53BER eRNAs are important for p53 dependent cell cycle arrest

si-PAPPA: has same effect like si.p53BER eRNA

FINAL MODEL

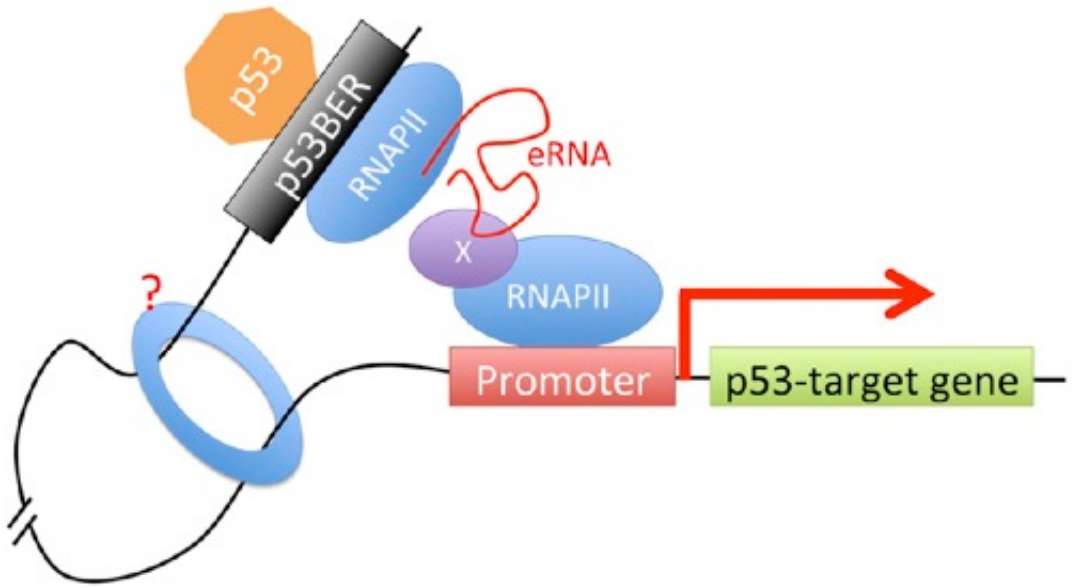


Figure 6. Schematic Model Depicting How p53BERs May Enhance Target Gene Transcription

p53-independent intrachromosomal interactions bring p53-bound p53BERs in close proximity of target genes, through yet unidentified protein complex (marked by ?). It seems likely that p53 is bound at p53BERs in a poised state and that upon p53 activation eRNAs are produced and transcription enhancement takes place. eRNAs are produced in a p53-dependent manner and affect transcription enhancement by a currently unknown mechanism. It is possible that eRNAs interact with proteins (X) that activate transcription.

Enhancer element and promoter of nearby genes physically interact (loop formation) → 4C data!!!

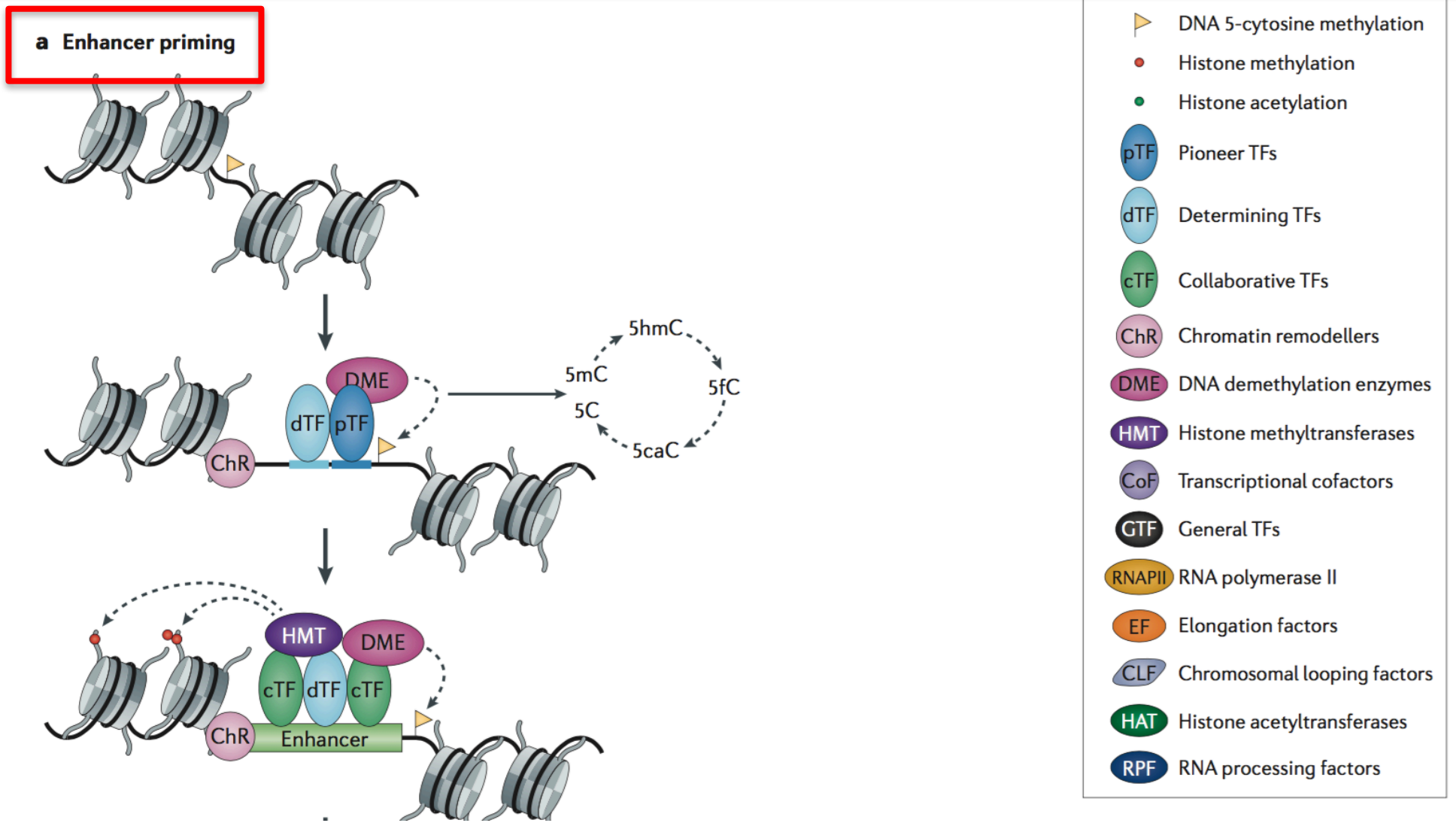
Transcription factor associated with enhancer element

Important: the direct interaction between enhancer and nearby promoters does not depend on p53 (transcription factor)

Transcription factors bound to enhancer elements stimulates eRNAs production by RNA Pol II.

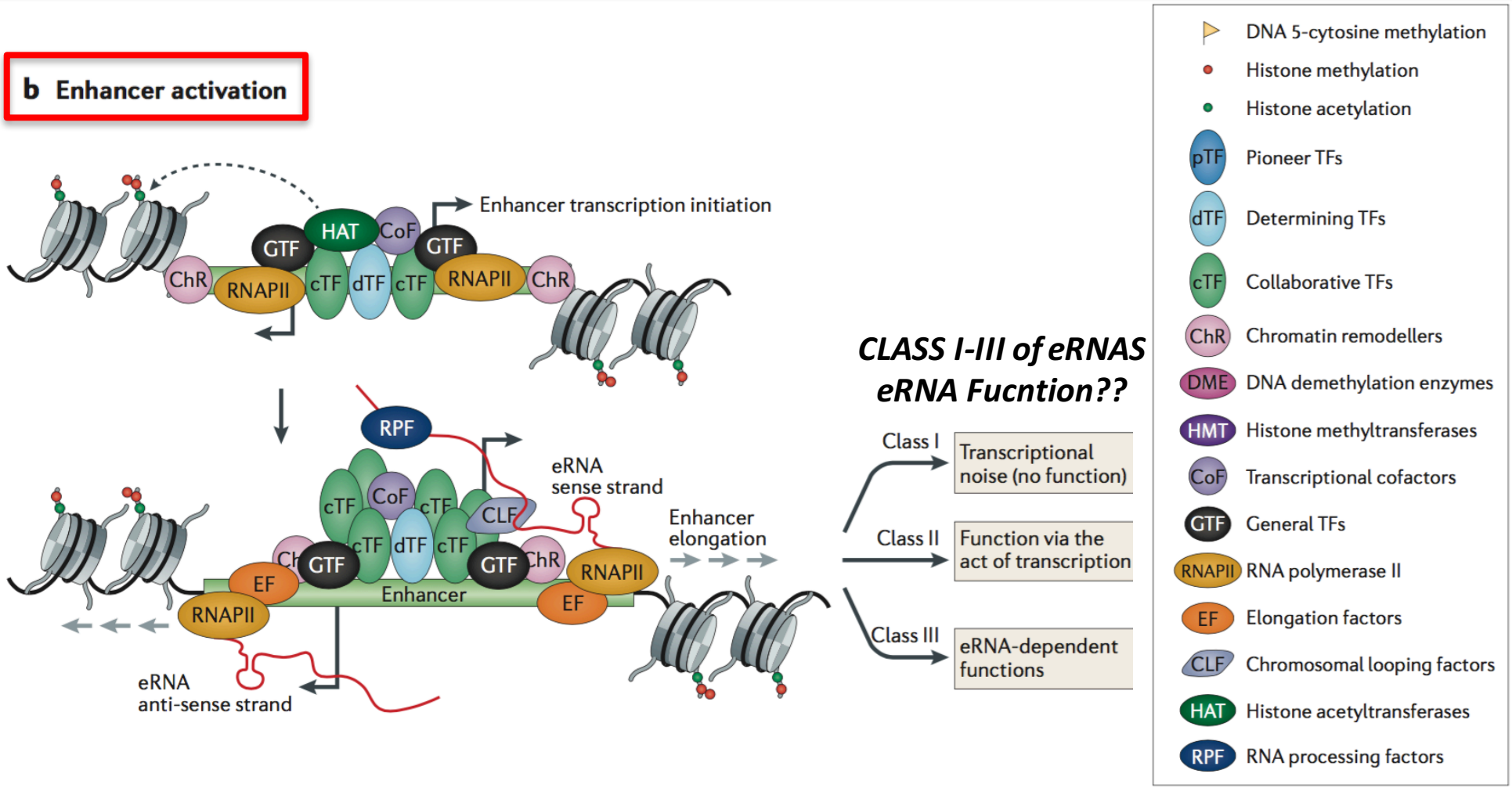
eRNAs activate the transcription of promoters attached to enhancer elements
→Pol II located on enhancer and also on promoter (in close vicinity)

Activation of an enhancer.



The assembly of the transcriptional apparatus at an enhancer is first initiated by the binding of pioneer transcription factors (pTFs)170, which bind the DNA in nucleosomes to generate open chromatin, allowing the recruitment of lineage-determining transcription factors (dTFs) to dictate the enhancer site for activation in a specific cell lineage. Collaborative transcription factors (cTFs) and important cofactors (CoFs) are further recruited, such as histone methyltransferases, which 'write' mono- and dimethylation on the H3 tail at lysine 4 (H3K4me1 and H3K4me2, respectively). Each red circle represents a methyl group.

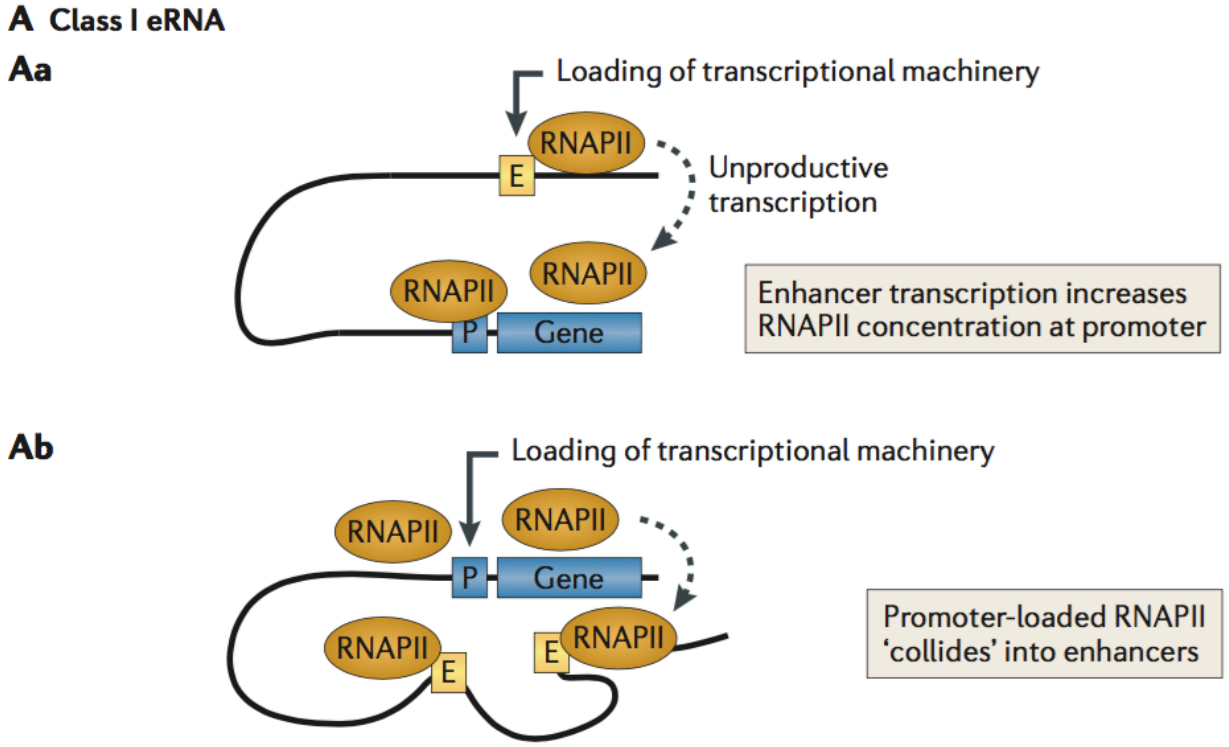
Activation of an enhancer



These previous steps prepare the enhancer for further recruitment of other CoFs, such as histone acetyltransferases (HATs; for example, CREB-binding protein and p300 (CBP/p300)) that deposit histone acetylation marks (green circles), as well as general transcription factors (GTFs) and **RNA polymerase** II (RNAPII) holoenzymes to initiate bidirectional transcription. The acetylated histone tail recruits additional CoFs, such as bromodomain-containing protein 4 (BRD4), which probably works together with positive transcription elongation factor-b complex (pTEFb) (not shown) to promote transcriptional elongation of enhancer RNAs (eRNAs). The recruitment of chromatin looping factors (CLFs) facilitates enhancer–promoter interactions. <u>However, the order of their recruitment and roles in enhancer transcription is not fully understood</u>

Classes of eRNAs

CLASS I: The transcription process and enhancer RNAs (eRNAs) are non-functional and are merely transcriptional noise



Aa)
Enhancer transcription merely increases the local concentration of transcriptional machinery to augment promoter activities.

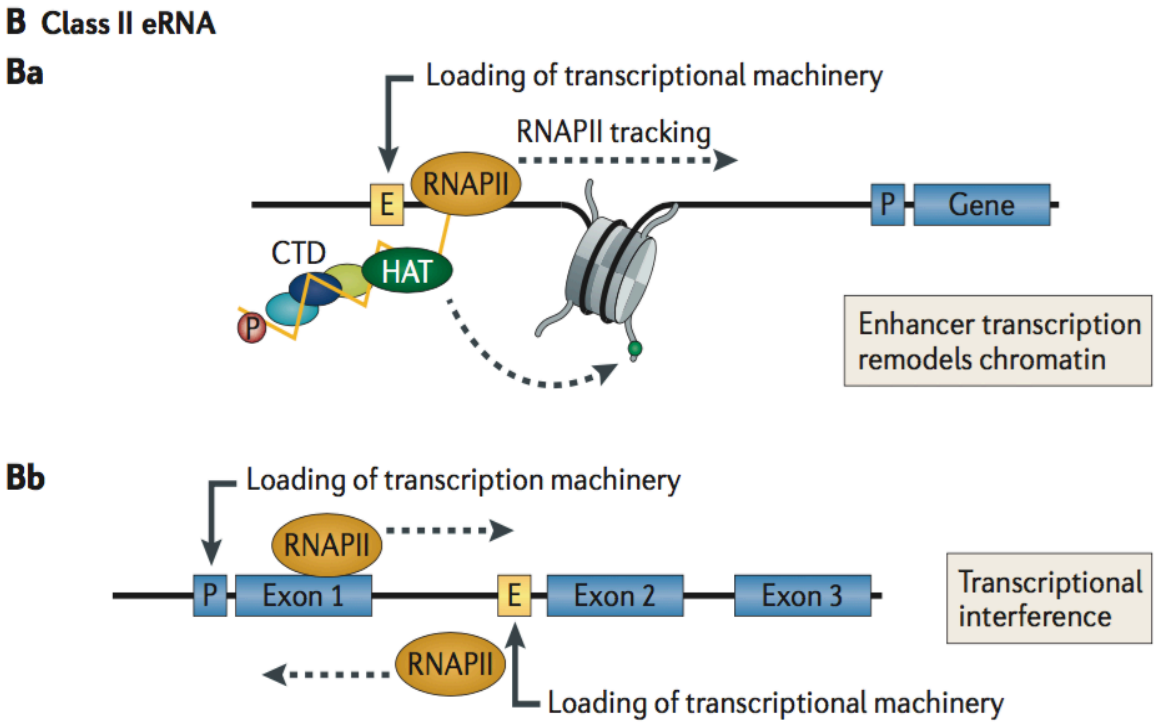
Ab)
Promoter- loaded machinery passively or randomly collides with enhancers due to their physical proximity.

RNA production at enhancer is rather a by-product of transcriptional noise (remember: virtually all parts of the genome are transcribed (at least at low levels))
Conservation of enhancers is low; chromatin rather open, low RNA levels → random transcription, eRNAs result from a random scanning by RNA PolII

- eRNA does not have a biological function; objective: load RNA Pol II to promter
- eRNAs are by product by RNA Pol II that might read Enhancers looped to promoter

Classes of eRNAs

CLASS II: The act of enhancer transcription mediates function



Ba)
The transcription of some enhancers by RNAPII could remodel the intervening chromatin between enhancers and promoters and activate target genes over a long range; transcribing RNAPII could carry histone modifiers such as histone acetyltransferases (HATs) or histone methyltransferases (not shown) to modify the enhancer region and the intervening DNA.

Bb)
For other enhancers, especially those in introns, their transcription may interfere with the overlapping gene transcription.

eRNAs are just a by-product.
The ACT of transcription by RNA Polymerase II is relevant

eRNAs are functionally relevant

Evidence from a series of experiments:

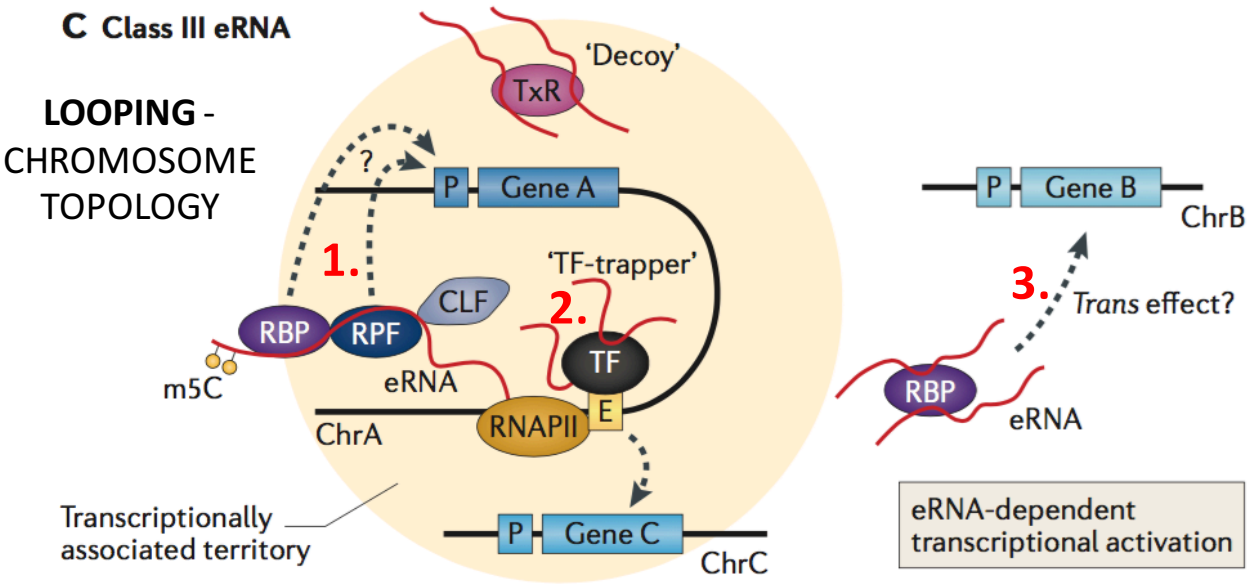
- **siRNA mediated knock-down of eRNAs: target promoter expression is lower**

- **RNA-tethering assays (MS2 system) → eRNAs are a structural/functional component mediating activation of promoters by enhancers**

- **Multiple studies show that eRNA production is important for enhancer activity (→ General concept of enhancer activity)**

Class III of eRNAs and eRNA function

CLASS III: Genes on the same chromatin fibre (cis), or potentially on other chromosomes (trans), are regulated by an eRNA (MAJOR TYPE)



Functional mechanisms of eRNAs include:

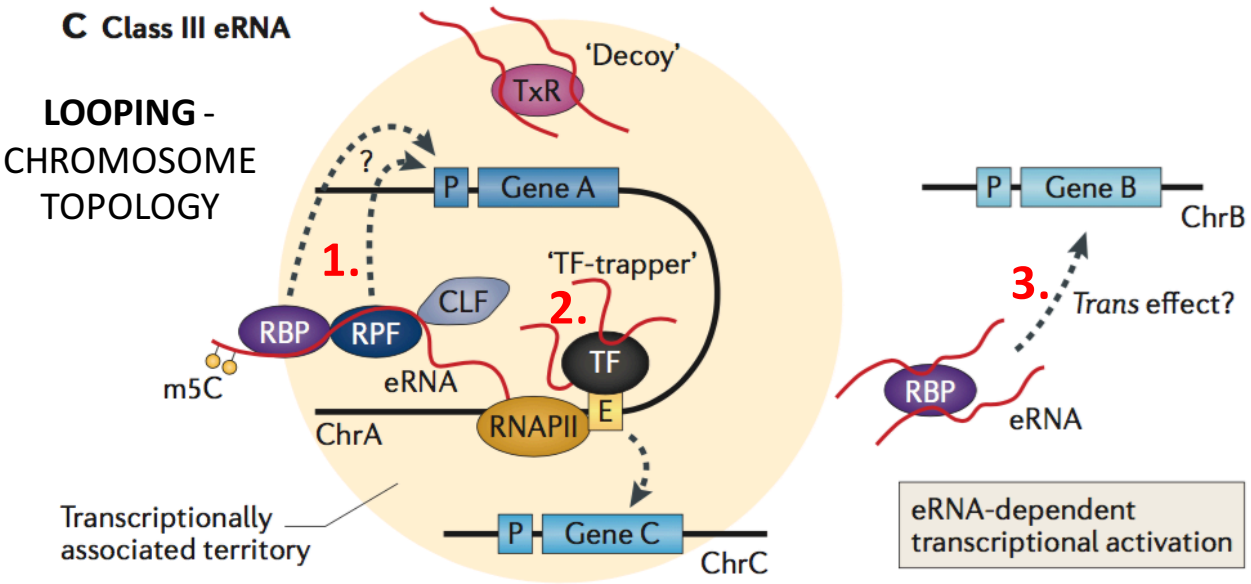
1. eRNAs interact with chromosomal looping factors (CLFs) to positively influence enhancer–promoter looping and gene transcription;

RNA binding protein (RBP)
RNA processing factor (RPF)
Trancription factor (TF)
Chromosomal looping factor (CLF)

Examples for a role of eRNAs in looping:
-eRNA knock-down results in impaired looping (in some cases)
-eRNAs were found to interact with cohesin
-eRNAs were found to interact with the mediator complex
Examples that claim that eRNAs to not support looping:
-Blocking the elongation of RNA transcription does not negatively impact on looping
-Knock-down of eRNAs did not impact looping (in some cases)

Class III of eRNAs and eRNA function

CLASS III: Genes on the same chromatin fibre (cis), or potentially on other chromosomes (trans), are regulated by an eRNA (MAJOR TYPE)



Functional mechanisms of eRNAs include:

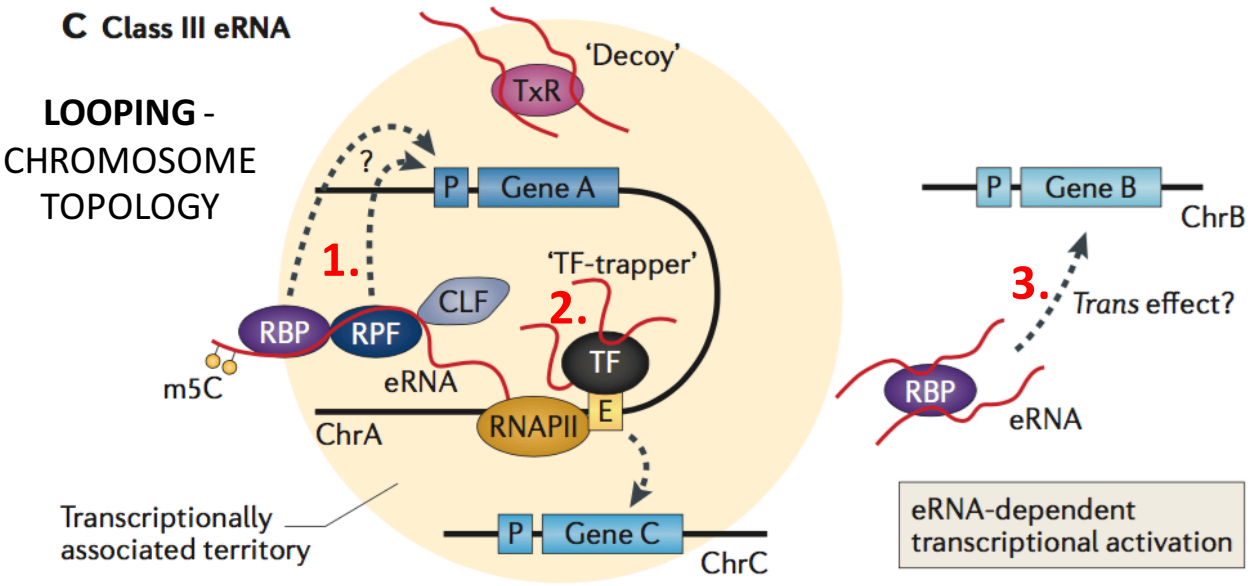
2. eRNAs bind transcription factors (TFs) to help 'trap' them at enhancers

RNA binding protein (RBP)
RNA processing factor (RPF)
Trancription factor (TF)
Chromosomal looping factor (CLF)

eRNAs trap transcription factors that enhance the transcription of enhancers (YY1 – eRNA interaction; YY1 binding is increased at enhancers

Class III of eRNAs and eRNA function

CLASS III: Genes on the same chromatin fibre (cis), or potentially on other chromosomes (trans), are regulated by an eRNA (MAJOR TYPE)



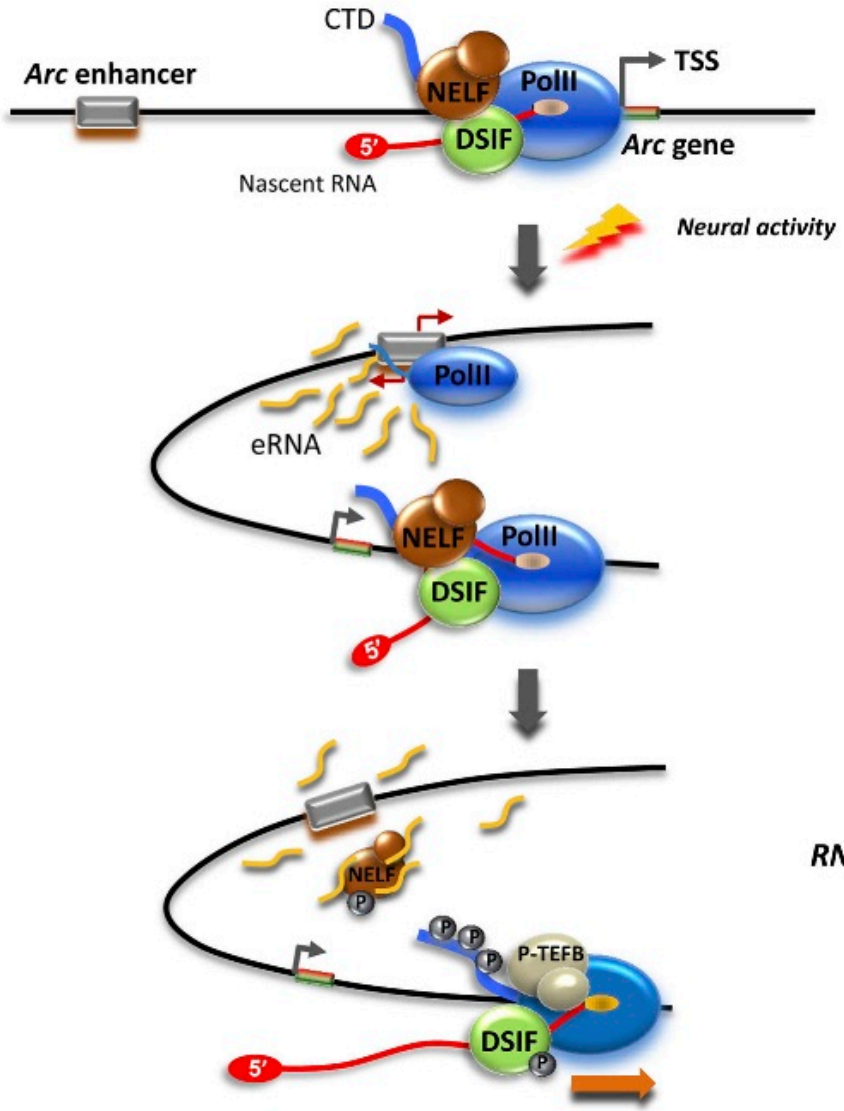
RNA binding protein (RBP)
RNA processing factor (RPF)
Trancription factor (TF)
Chromosomal looping factor (CLF)

Functional mechanisms of eRNAs include:

3. eRNAs act as a 'decoys' or 'repellents' to inhibit transcriptional repressors (TxRs). Trans roles could be achieved by eRNA translocation to distant sites (right side of panel) or proximity-based regulation in which eRNAs and target gene(s) reside in certain transcriptionally associated territory (light yellow area).

Class III of eRNAs and eRNA function

EVIDENCE THAT eRNAs CONTROL PAUSING OF RNA Pol II AT PROMOTERS



Paused RNA PolII

Quiescent

Local synthesis of eRNA

Decoying NELF

< 30 min activation

RNA PolII elongation and mRNA synthesis

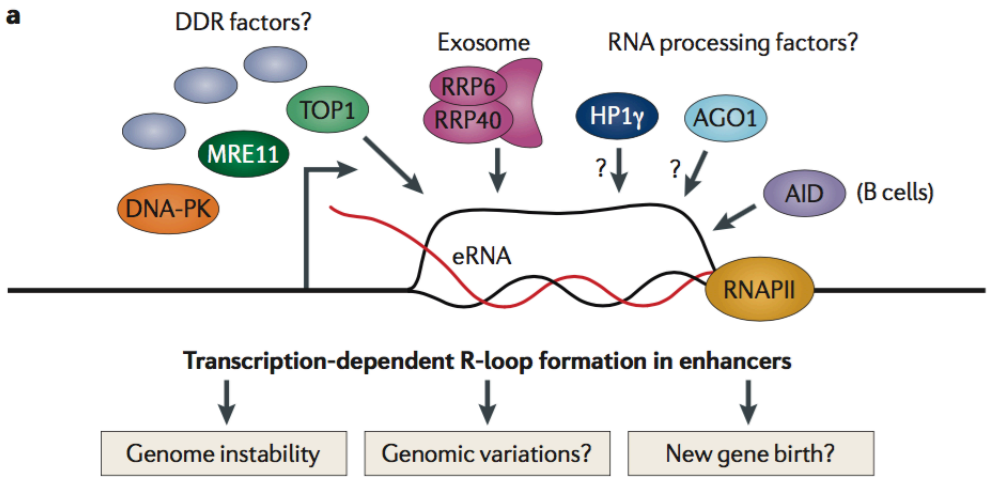
< 60 min activation

Arc gene is essential for brain development and function.
RNA Pol II and TFs are ready on promoter. After a cell stimulus the activation of the gene is very fast and efficient (=IEG: immediately early gene)
In the inactive status, RNA Pol II is located at the promoter but the protein NELF (negative elongation factor) pauses the activity of RNA Pol II

After stimulus, Arc Enhancer produces eRNAs That bind to NELF and cause the release of NELF from the promoter; P-TEFB phosphorylates RNA Pol CTD Ser-2

→ IMMEDIATE BURST OF TRANSCRIPTION

Genomic Variation of enhancers in disease



Transcription of enhancers that are characterized by low conservation my lead genomic instability that contributes to the formation of disease relevant mutations or the generation of new genes

Transcriptional activity at enhancers may result in nucleic acid structures including R-loops, although the prevalence of R-loops at enhancers still requires further genome-wide elucidation. Various factors, as shown here, are presumably involved in the process of R-loop resolution or enhancer transcription. Despite their binding to enhancers, most DNA damage response (DDR) factors are still mechanistically enigmatic in enhancer transcription or function. The additional factors in the figure include Argonaute 1 (AGO1), which has been shown to bind active enhancers depending on enhancer transcription178, and activation-induced cytidine deaminase (AID), the mis-targeting of which in the B cell genome has recently been associated with enhancer transcription. Question marks denote unclear functional roles or mechanisms. The transcription process of enhancers may be linked to several important biological, pathological or evolutionary functions, as shown in the diagram. Exosome component 10 (EXOSC10; also known as RRP6) and EXOSC3 (also known as RRP40) denote the two components of the RNA exosome complex.

ARTICLE

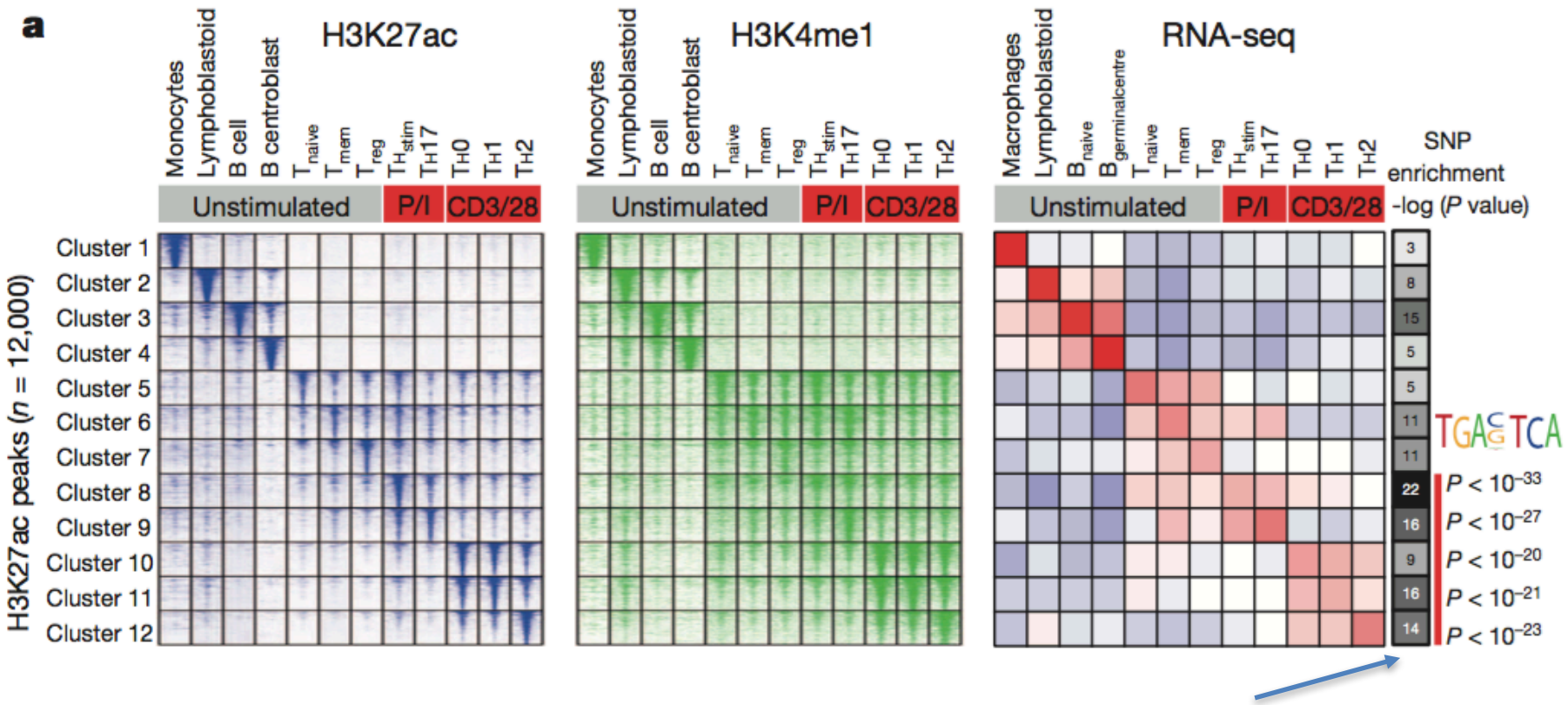
doi:10.1038/nature13835

Genetic and epigenetic fine mapping of causal autoimmune disease variants

Kyle Kai-How Farh[1,2]*, Alexander Marson[3]*, Jiang Zhu[1,4,5,6], Markus Kleinewietfeld[1,7]†, William J. Housley[7], Samantha Beik[1], Noam Shoresh[1], Holly Whitton[1], Russell J. H. Ryan[1,5], Alexander A. Shishkin[1,8], Meital Hatan[1], Marlene J. Carrasco-Alfonso[9], Dita Mayer[9], C. John Luckey[9], Nikolaos A. Patsopoulos[1,10,11], Philip L. De Jager[1,10,11], Vijay K. Kuchroo[12], Charles B. Epstein[1], Mark J. Daly[1,2], David A. Hafler[1,7]§ & Bradley E. Bernstein[1,4,5,6]§

Genome-wide association studies have identified loci underlying human diseases, but the causal nucleotide changes and mechanisms remain largely unknown. Here we developed a fine-mapping algorithm to identify candidate causal variants for 21 autoimmune diseases from genotyping data. We integrated these predictions with transcription and *cis*-regulatory element annotations, derived by mapping RNA and chromatin in primary immune cells, including resting and stimulated CD4$^+$ T-cell subsets, regulatory T cells, CD8$^+$ T cells, B cells, and monocytes. We find that ~90% of causal variants are non-coding, with ~60% mapping to immune-cell enhancers, many of which gain histone acetylation and transcribe enhancer-associated RNA upon immune stimulation. Causal variants tend to occur near binding sites for master regulators of immune differentiation and stimulus-dependent gene activation, but only 10–20% directly alter recognizable transcription factor binding motifs. Rather, most non-coding risk variants, including those that alter gene expression, affect non-canonical sequence determinants not well-explained by current gene regulatory models.

Genomic Variation of enhancers in disease

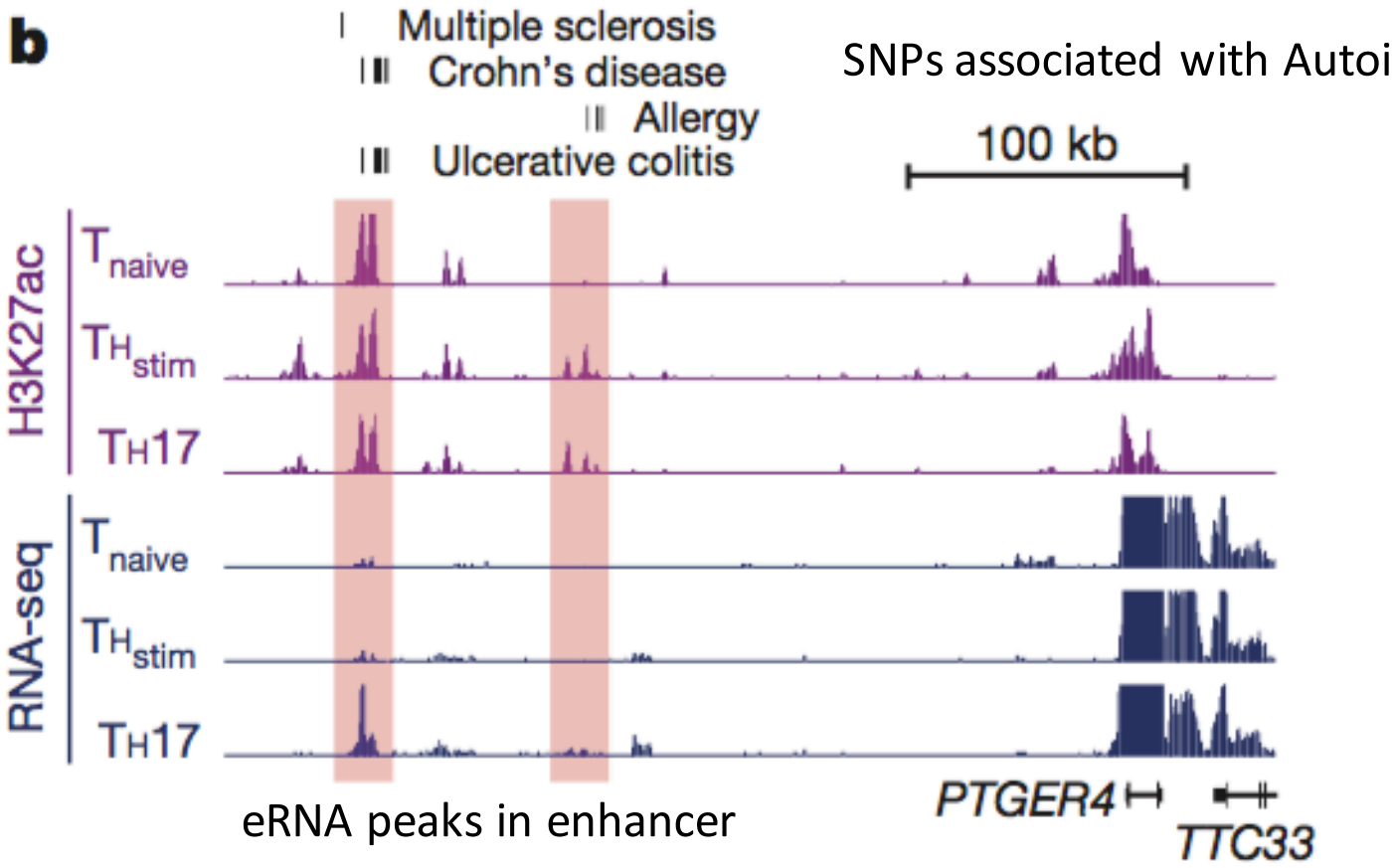


SNPs found in autoimmune genes were mapped to enhancer elements and the respective enrichment was determined. → SNPés are enriched in enhancers.

Heatmaps show H3K27ac and H3K4me1 signals for 1,000 candidate enhancers (rows) in 12 immune cell types (columns). Enhancers are clustered by the cell type-specificity of their H3K27ac signals. Adjacent heatmap shows average RNA-seq expression for the genes nearest to the enhancers in each cluster. Greyscale (right) depicts the enrichment of PICS autoimmunity SNPs in each enhancer cluster (hypergeometric P values calculated based on the number of PICS SNPs overlapping enhancers from each cluster, relative to random SNPs from the same loci). The AP-1 motif is over-represented in enhancers preferentially marked in stimulated T cells, compared to naive T cells.

PICS: Probabilistic Identification of Causal SNPs

Genomic Variation of enhancers in disease



SNPs associated with Autoimmune disease

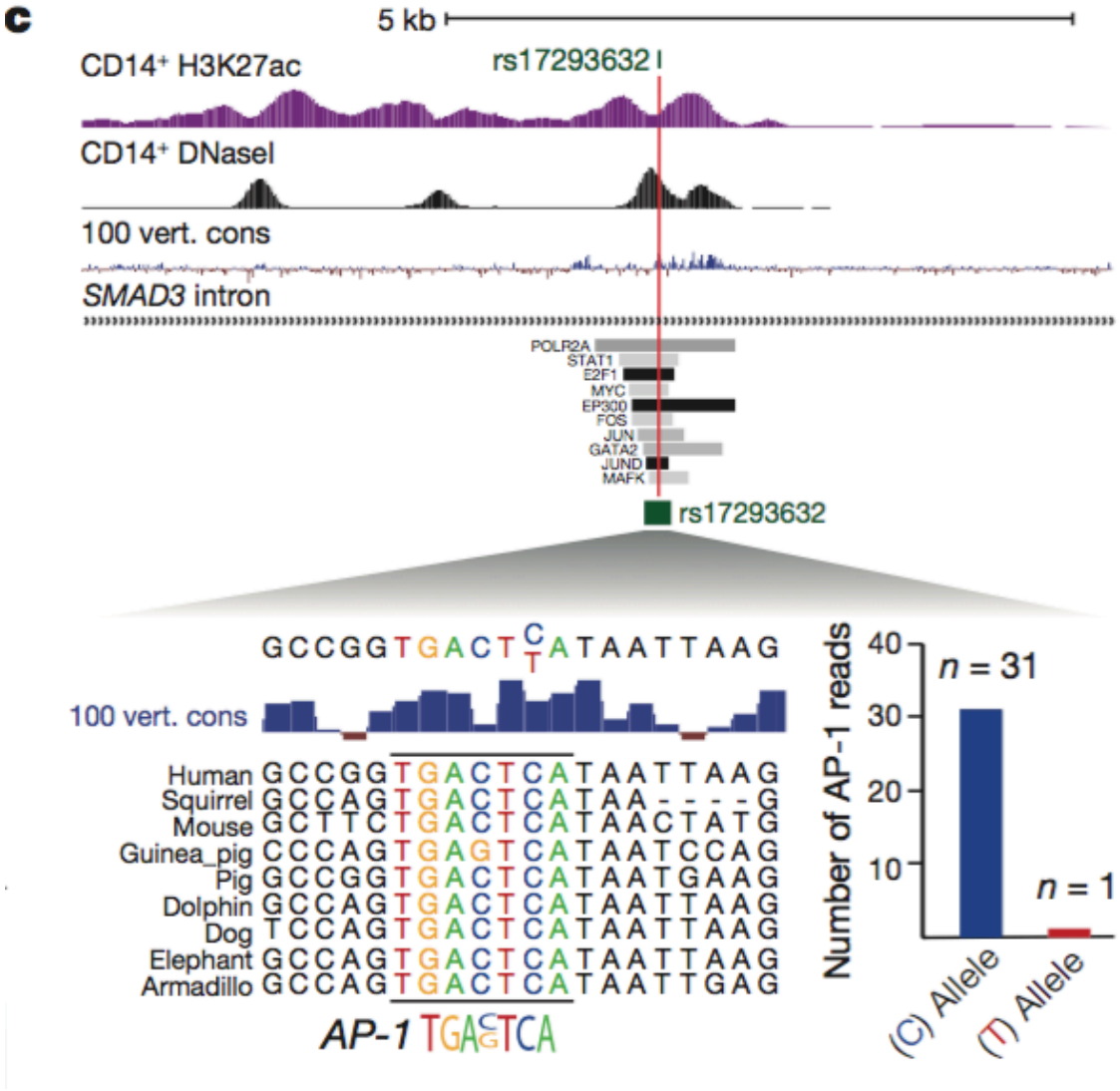
eRNA peaks in enhancer

expression of closest gene

In autoimmune disease it is estimated that immune enhancers overall account for 60% of candidate causal SNPs, whereas promoters account for another 8% of these variants

Candidate causal SNPs displayed along with H3K27ac and RNA-seq signals at the PTGER4 locus. A subset of enhancers with disease variants (shaded) shows evidence of stimulus-dependent eRNA transcription.

Genomic Variation of enhancers in disease



A SNP C-->Tin a enhancer of destroys the bindign site for a key transcription factor AP1 (heterodimer of Jun and Fos)

Note: the T variant is the respective SNP

H3K27ac, DNaseI26 and conservation signals, and selected transcription factor binding intervals are shown in a SMAD3 intronic locus. rs17293632, a non-coding candidate causal SNP for Crohn's disease, disrupts a conserved AP-1 binding motif in an enhancer marked by H3K27ac in CD141 monocytes. Summing of ChIP-seq reads overlapping the SNP in the heterozygous HeLa cell line shows that only the intact motif binds AP-1 transcription factors, Jun and Fos.