



HAL
open science

DISSECTION OF THE COORDINATED INTERFERON RESPONSE BY PROMOTER ELEMENTS WITH ENHANCER ACTIVITY

David de Santiago Algarra

► **To cite this version:**

David de Santiago Algarra. DISSECTION OF THE COORDINATED INTERFERON RESPONSE BY PROMOTER ELEMENTS WITH ENHANCER ACTIVITY. Quantitative Methods [q-bio.QM]. Aix-Marseille Université, 2020. English. NNT : 2020AIXM0430 . tel-04420220

HAL Id: tel-04420220

<https://hal.science/tel-04420220v1>

Submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AIX-MARSEILLE UNIVERSITÉ

FACULTÉ DES SCIENCE

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

THÉORIES ET APPROCHES DE LA COMPLEXITÉ GENOMIQUE (TAGC)

THÈSE DE DOCTORAT
GÉNOMIQUE ET BIOINFORMATIQUE

**DISSECTION DE LA REPONSE COORDONNEE DE
L'INTERFERON PAR DES ELEMENTS PROMOTEURS AVEC UNE
ACTIVITE ENHANCER**

Par

David de Santiago Algarra

Soutenue le 10/12/2020 devant le jury :

Dr. Jean-Christophe Andrau	Président
Dr. Susanne Mandrup	Rapporteur
Dr. Gioacchino Natoli	Rapporteur
Dr. Sandrine Sarrazin	Examineur
Dr. Salvatore Spicuglia	Directeur de thèse

AIX-MARSEILLE UNIVERSITY
FACULTY OF SCIENCE
DOCTORAL SCHOOL OF LIFE SCIENCES AND HEALTH
THEORIES AND APPROACHES OF GENOMIC COMPLEXITY (TAGC)

DOCTORAL THESIS
GENOMICS AND BIOINFORMATICS

**DISSECTION OF THE COORDINATED INTERFERON RESPONSE
BY PROMOTER ELEMENTS WITH ENHANCER ACTIVITY**

By

David de Santiago Algarra

Defended the 10/12/2020 in front of the jury:

Dr. Jean-Christophe Andrau	President
Dr. Susanne Mandrup	Reporter
Dr. Gioacchino Natoli	Reporter
Dr. Sandrine Sarrazin	Examinator
Dr. Salvatore Spicuglia	Supervisor

Résumé

L'expression génique est contrôlée par l'implication des éléments régulateurs du gène proximal (promoteurs) et distal (enhancer). Des résultats antérieurs ont démontré qu'un sous-ensemble de promoteurs de gènes, également appelés Epromoteurs, fonctionne comme des enhanceurs de bona fide et régule l'expression distale des gènes. Nous avons émis l'hypothèse que les Epromoteurs pourraient jouer un rôle clé dans la coordination de l'induction rapide des gènes dans la réponse au stress, en particulier pendant l'inflammation. En utilisant un test de reporter à haut débit, nous avons exploré la fonction d'Epromoter en réponse à l'interféron de type I. Nous avons constaté que les facteurs de transcription STAT1/STAT2 et IRF se lient préférentiellement aux Epromoteurs induits par l'IFN α et régulent de manière distale l'activation des gènes de réponse à l'interféron. Des observations similaires ont été faites dans d'autres types de réponse inflammatoire. Nos résultats suggèrent que les Epromoteurs pourraient fonctionner comme une plaque tournante pour recruter les TF clés nécessaires à la régulation coordonnée des grappes de gènes pendant la réponse inflammatoire, et plus généralement sur la réponse cellulaire aux signaux intra et extracellulaires.

Abstract

Gene expression is controlled by the involvement of gene-proximal (promoters) and distal (enhancers) regulatory elements. Previous results have demonstrated that a subset of gene promoters also termed Epromoters, works as bona fide enhancers and regulates distal gene expression. We hypothesized that Epromoters might play a key role in the coordination of rapid gene induction in the stress response, in particular during inflammation. Using a high-throughput reporter assay we explored the function of Epromoter in response to type I interferon. We found that STAT1/2 and IRF transcription factors preferentially bind to IFN α -induced Epromoters and distally regulate the activation of interferon-response genes. Similar observations were made in other types of inflammatory response. Our findings suggest that Epromoters might function as a hub to recruit key TFs required for coordinated regulation of gene clusters during the inflammatory response, and more generally upon cellular response to intra- and extra-cellular signals.

Acknowledgment

First and foremost, I'd like to acknowledge the jury member **Susanne Mandrup, Gioacchino Natoli, Sandrine Sarrazin**, and **Jean-Christophe Andrau** who dedicate their time for critical reading, proof, and assessing my dissertation. I am grateful for their valuable input and advice that help me enlarge my vision of science.

I would like to show my special thanks to my supervisor **Salvatore Spicuglia** for his guidance, advice, relentless support, and feedback throughout this time. I am extremely grateful for giving me the opportunity to work on such an amazing project in collaboration with his wonderful team. His personal and academic support has been invaluable.

I would like to thank the institutions that support me during my Ph.D.: to the **Aix-Marseille University** and the **Doctoral school ed.62** for accepting me as part of this great institution and for providing academic formations and administrative support; and to the **CONACYT Mexico**, for granting me the fellowship to pursue my ambition to do a Ph.D. in France.

I want to extend my special thanks to my former lab mate and collaborator **Lan T. M. Dao**, our bioinformatician **Himanshu Singh**, and our post-doc **Charbel Souaid** for their valuable help in this project.

Special thanks to our collaborators **Jaime Castro Mondragon, Denis Puthier, Lydie Pradel, Guillaume Charbonnier, Nori Sadouni, Lucia Ramirez, Alejandra Medina, Alexandre España, Charlotte Andrieu-Soler**, and **Eric Soler** who have been involved in this project for their invaluable contributions.

My sincere gratitude to the past and present members of the team **Jose David Abad, Iris Manosalva, Saadat Hussain, Eve-lyne Mathieu, Ariel Galindo, Juliette Malfait, Wiam Saadi, Yasmina Kermezli, Jaafar Al-Rekaabi**, and **Quentin Ferré** for their constant support.

Thanks to all the members of the **TAGC** and **TGML** for your invaluable help and the great work environment.

I cannot begin to express my deepest thanks to **my Mexican and French family**, especially to **my parents** and my wife, **Audrey Stephan**, who believed in me and my abilities, and to my grandfather, **Jose de Santiago Ortiz** who supported me and all the family until his last breath.

Finally, I would like to thank my dear friends in Mexico and Europe for your constant support, friendship, and kindness.

Table of contents

CHAPTER 1. TRANSCRIPTION REGULATION IN MAMMALS AND REGULATORY ELEMENTS	1
I. General introduction	1
II. DNA packing and modification	2
1. Chromatin structure	2
2. DNA methylation	2
II. Post-translational Histone Modifications	4
1. Histone acetylation	5
2. Histone methylation	5
3. Histone phosphorylation	5
IV. Transcription factors	7
III. Cis-regulatory elements	8
1. Promoters	8
<i>Histone configuration and modifications</i>	9
<i>Transcription elongation and directionality</i>	10
2. Enhancers	11
3. Silencers	13
4. Insulators	14
5. Long non-coding RNAs	14
6. Similarities between enhancers and promoters	15
7. Epromoter	16
8. Long-range chromatin interactions	16
CHAPTER 2. IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF TRANSCRIPTIONAL REGULATORY ELEMENTS	19
I. Identification of transcriptional regulatory elements	19
1. Identification of histone modifications and non-histone proteins	19
2. Chromatin accessibility	20
<i>DNase-seq</i>	21
<i>ATAC-seq</i>	21
<i>MNase-seq</i>	22
<i>NOMe-seq</i>	22
<i>FAIRE-seq</i>	23
3. TSS identification techniques	24
<i>Steady-state RNA analysis</i>	24
<i>Nascent-RNA analysis</i>	25
<i>Concluding Remarks</i>	26
II. Methods for long-range chromatin interactions	26
1. Chromatin conformation capture (3C)	26
2. One versus all, chromatin conformation capture (4C)	27
3. Chromosome conformation capture carbon copy (5C)	27
4. Hi-C	28

5. Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET)	28
6. Capture-C	28
7. Concluding Remarks	29
III. Functional characterization of transcriptional regulatory elements	30
1. MPRA	31
2. STARR-seq	32
3. Derivative approaches.	32
<i>Reduce complexity</i>	32
<i>CapSTARR-seq</i>	32
<i>ATAC-STARR-seq</i>	32
<i>FAIRE-STARR-seq</i>	32
<i>BAC</i>	33
<i>ChIP-STARR-seq</i>	33
<i>Assessing genetic variation</i>	33
<i>Massive oligo synthesis</i>	34
<i>Pool DNA</i>	34
<i>Pool genome and SuRE</i>	34
<i>Integrative</i>	34
<i>Lenti-based assays</i>	34
<i>Locus integration</i>	35
<i>Silencers</i>	35
<i>DNA methylation</i>	35
<i>Methyl-STARR-seq</i>	35
4. Concluding Remarks	35
5. Technical considerations	36
IV. CRISPR-based techniques	37
1. Discovery	37
2. CRISPR-Cas9 applications	38
<i>CRISPR-mediated genome editing</i>	38
<i>CRISPR interference</i>	39
<i>CRISPR activation</i>	39
3. Concluding Remarks	39
CHAPTER 3. INTERFERON INFLAMMATORY RESPONSE	42
I. Types of interferon	42
II. Induction of IFN expression	42
III. Receptor and signal transduction	44
IV. Interferon regulatory factors	46
V. Chromatin Remodeling	47
VI. Interferonopathies and autoimmune diseases	48
VII. Concluding remarks	49
CHAPTER 4. ROLE OF THE EPROMOTERS IN THE TYPE I INTERFERON RESPONSE	50

I. Objectives	50
II. Main results	50
III. Contributions	51
IV. Manuscript	52
V. Complementary results	90
CHAPTER 5. GENERAL DISCUSSION AND PERSPECTIVES	93
I. Epromoters activity in the type I interferon response	93
II. Epromoters role in the formation of 3D interactions	94
III. Genetic variants in the Epromoters	95
IV. Epromoters in other cellular stress models	98
V. Molecular bases of promoter versus enhancer activity of Epromoters	101
VI. Concluding remarks	103
BIBLIOGRAPHY	105
ANNEX 1	128
ANNEX 2	146

Chapter 1. Transcription regulation in mammals and regulatory elements

I. General introduction

The human genomic information is encoded in the DNA included in each cell of the body; 3 billion letters are packaged into 23 chromosomes. Over 200 different cell types interpret the same information very differently and perform a specific task to keep us alive. The gene expression consists of two major steps: the transcription of the DNA into mRNA and the translation of the mRNA into proteins. This constitutes the central dogma of biology.

The gene expression is regulated at several stages including changes in the chromatin and access to the DNA information, transcription initiation and elongation, RNA processing, transportation, translation y post-translational modifications. In each stage, many proteins, factors, and cofactors are involved.

The initial control of the gene expression is regulated in time and space during the transcription, mainly by (i) the chromatin structure and chemical DNA modifications, (ii) the post-transcriptional modifications in the chromatin, (iii) by the transcription factors present in the milieu, and (iv) by regions of the DNA called cis-regulatory elements (Fig. 1.1). This control is critical for cellular differentiation, embryogenesis, defense against pathogens, and in general to the homeostasis of a multicellular organism. Dysregulation of these gene expression programs can cause a broad range of diseases ranging from birth defects, cancer, susceptibility to infections, autoimmune diseases, and others.

Among the cis-regulatory elements, the most studied are the promoters and the enhancers and the relationship between the two mediated by 3D interactions. However other elements like the novel Epromoters are also being described in important roles.

This chapter is dedicated to giving an overview of the factors and mechanisms that contributes to transcriptional regulation.

The available literature describing the transcriptional regulation in mammals is huge. I choose to focus my bibliography and this introduction on describing the basic mechanisms, whereas some points are broadly explained in our published review (annex 1).

II. DNA packing and modification

1. Chromatin structure

The eukaryote genome is organized by the chromatin, which manages to package almost three meters of DNA in the nucleus (about 10 μm) and is perhaps one of the first transcription regulatory mechanisms. The chromatin is composed primarily of DNA and proteins and its fundamental unit is the nucleosome: 147 bp of DNA wraps around eight histone proteins supported by electrostatic interactions. This octamer consists of two copies of each four core histone proteins (H2A, H2B, H3, and H4). Also, the fifth histone protein (H1) might serve as a stabilizer of the nucleosome, binding at the entry/exit of the DNA. The chromatin organization depends on the chemical modifications of the DNA (cytosine methylation) along with the histone variants and modifications (Fig. 1.1).

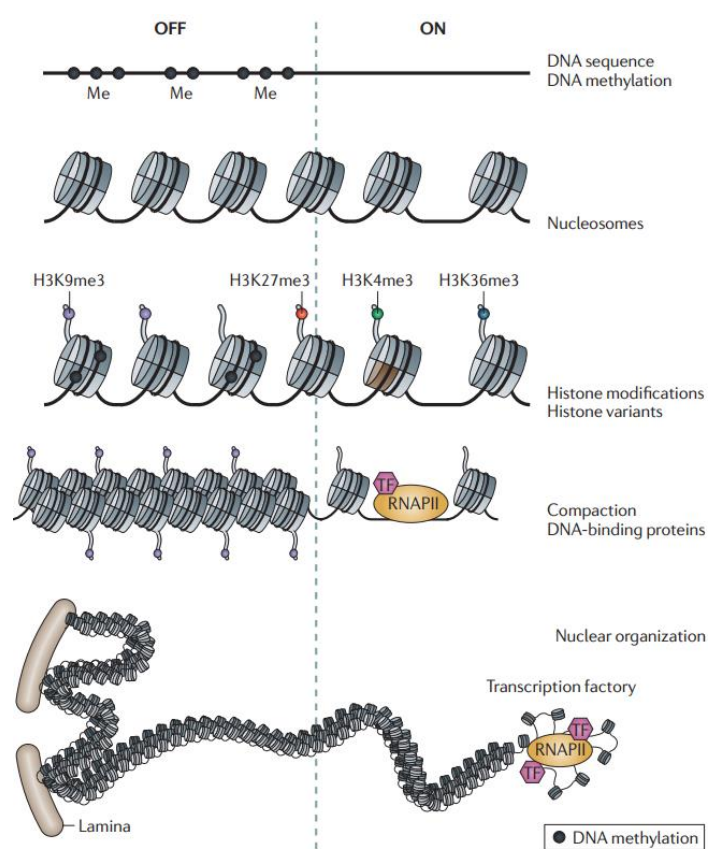


Figure 1.1 Layers of chromatin organization in the mammalian cell nucleus. Features at different levels of chromatin organization are generally associated with inactive (off) or active (on) transcription. From the top, genomic DNA is methylated (Me) on cytosine bases and is packaged into nucleosomes, which vary in histone composition and histone translational modifications. Histone variants such as H2A.Z are in brown. DNA in chromatin may remain accessible to DNA-binding proteins such as transcription factors (TFs) and RNA polymerase II (RNAPII) or may be further compacted. Chromatin can also organize into higher-order structures such as nuclear lamina-associated domains and transcription factories. Each layer of organization reflects aspects of gene and genome regulation. Adapted from (Zhou et al., 2011)

2. DNA methylation

The DNA methylation consists of the covalent union of a methyl group at the 5' position of the cytosine to form 5-methylcytosine (5mC) in the dinucleotide sequence CG, called CpG. This process is regulated through the action of DNA methyltransferase (DNMT) and is involved in the silencing of the gene expression (Keshet et al., 1986), embryo development (Bird, 1992; Li et al., 1992), cancer (Kulis and Esteller, 2010), parental imprinting (Tucci et al., 2019) and X-chromosome inactivation (da Rocha and Gendrel, 2019).

The CpG dinucleotides are less frequent than expected (based on nucleotide frequencies) but some regions contain a larger number of CpG repeats, called CpG islands. These CpG islands are generally unmethylated and they are frequently located upstream of promoter regions of genes, while the CpGs outside the CpG islands are methylated (Miranda and Jones, 2007; Nafee et al., 2008).

The human genome encodes five DNMTs: DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. DNMT2 and DNMT3L are non-canonical family members since they don't possess the catalytic DNMT activity, whereas DNMT3A and DNMT3B catalyze de novo DNA methylation (Ooi and Bestor, 2008) and DNMT1 maintain a fully methylated state of DNA after replication through copying mechanism (Goll and Bestor, 2005).

DNA methylation in the gene promoter is associated with gene silencing either through modification of transcription factors binding sites, blocking de binding and preventing the initiation of the transcription (Zhu et al., 2016) or through proteins containing methyl-CpG-binding domain (MBD) that recognize 5mC, recruit co-regulators, and modify the chromatin compaction (Du et al., 2015).

DNA methylation in CpGs can be removed either by passive demethylation process due to failure of DNMT1 to maintain DNA methylation patterns during cell divisions or by active demethylation mechanisms through enzymatic oxidation by ten-eleven translocation (TET) family [reviewed in (Wu and Zhang, 2017)] (fig. 1.2). 5-hydroxymethylcytosine (5hmC) is the most abundant product of 5mC and can be further oxidized to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC). Genome-wide mapping for 5mC oxidation products indicates that 5hmC is enriched in promoters regions containing low to intermediate CpG density, and correlates with gene expression in mouse and human ESCs (Pastor et al., 2011; Szulwach et al., 2011).

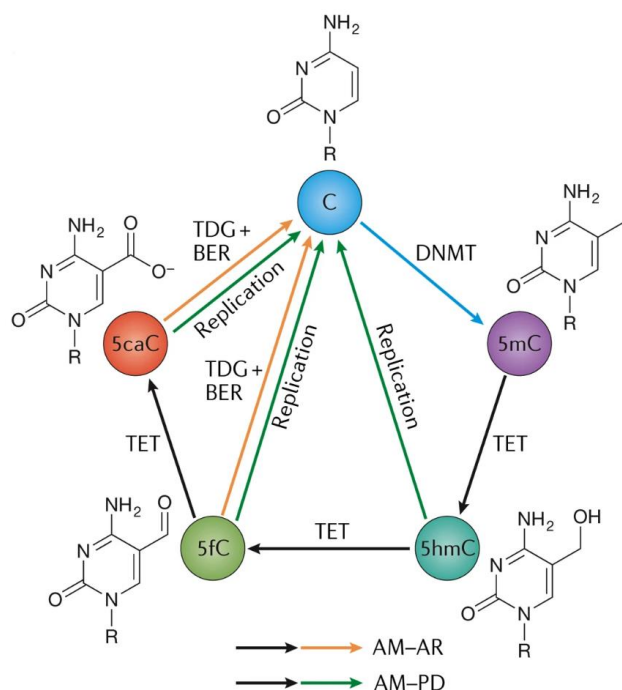


Figure 1.2 The cycle of active DNA demethylation. Thymine DNA glycosylase, TDG; base excision repair, BER; Active modification–active removal, AM–AR, active modification–passive dilution, AM–PD. Modified from (Wu and Zhang, 2017).

Abnormal changes in methylation patterns such as hypomethylation and hypermethylation have been found in cell aging and diverse types of cancer, respectively (Smith et al., 2017; Zhou et al., 2018b).

While 5mC is a well-studied epigenetic modification in gene promoters, the functional consequence of this mark in certain genomic contexts remains enigmatic, therefore, the use of methylome editing tools such as cleavage deficient (dCas9)-TET (Liu et al., 2016) and dCAS9-DNMT3A fusion (Pflueger et al., 2018) might help to know the impact of 5mC or 5hmC on gene transcription and their relationship with other epigenetic modifications.

II. Post-translational Histone Modifications

As said before, nucleosomes are formed by two copies of each histone protein, and each histone contains an exposed N-terminal tail. These tails can be modified and affect the nucleosome structure and function; chromatin compaction and therefore, affect the DNA availability to the transcriptional machinery. The amino acid residues susceptible to modification are lysine (K), arginine (R), serine (S), threonine (T), tyrosine (Y), and proline (P) (Fig. 1.3.). The first report of histone posttranslational modification was made by Vincent Allfrey in 1964 (Allfrey et al., 1964). The best characterized post-translational histone modifications are methylation, acetylation, and phosphorylation; others include ubiquitination, sumoylation, histone clipping, ADP ribosylation, and deamination. Depending on their effect on transcription, histone modifications are considered activating or repressing. Table 1.4 summarizes various forms of histone modifications and their effect on gene transcription (Lawrence et al., 2016).

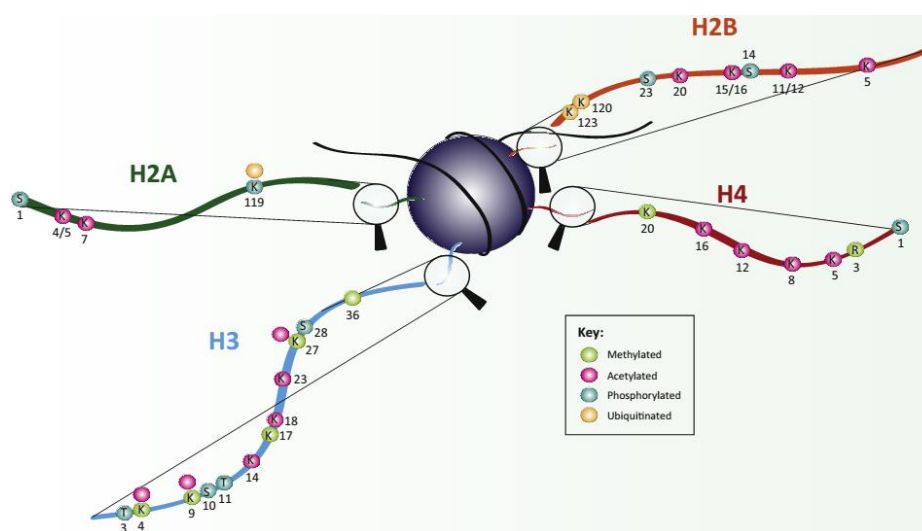


Figure 1.3 Schematic showing the post-translational modification and the amino acid residue in each histone tail. Adapted from Lawrence et al., 2016.

The study of the post-translational histone modifications was possible thanks to the development of the chromatin immunoprecipitation technique (ChIP). This technique uses antibodies to isolate any protein or histones modification of interest along with the DNA bound to it and serves as a basis for several other methods to identify protein-gene interaction.

1. Histone acetylation

Histone acetylation is regulated by two enzymes, histone acetyltransferases (HATs or KATs, for K-acetyltransferase) and histone deacetylases (HDACs). The HATs transfer the acetyl group from acetyl-Co-A to the ϵ -amino lysine residue in the histone tail. This process removes lysine's positive charge and weakens the electrostatic interaction between histones and the (negative charged) DNA, loosening the chromatin and thus, accessible to the transcriptional machinery.

There are two major classes of HATs: type-A and type-B. Type-B is mostly found in the cytoplasm and acetylates free histones but not those deposited in chromatin. The type-A can be classified into at least three separate groups depending on amino-acid sequence homology and conformational structure: GNAT, MYST, and CBP/p300 families (Hodawadekar and Marmorstein, 2007).

HDAC can be recruited as a component in multi-protein complexes, by interacting with DNA binding factors or other epigenetic modifiers, leading to transcriptional repression. The HDAC superfamily includes 18 enzymes classified into four classes (I, IIa, IIb, III, and IV). Class I are ubiquitous in the nucleus in all tissues, IIb is present in the nucleus and cytoplasm, while IIa is mainly cytosolic. Class III and IV might localize in the nucleus, cytoplasm, or mitochondria (Hull et al., 2016). HDACs are important to cell-cycle progression and apoptosis. Class I and class II HDACs can be inhibited by trichostatin A (TSA), a classical HDAC inhibitor; while class III requires the coenzyme NAD⁺ as a cofactor.

2. Histone methylation

The histone methylation is mediated by histone methyltransferases (HMTs), including lysine methyltransferases (KMTs) and arginine methyltransferases (PRMTs), while histone demethylases remove the methyl group. This enzymatic reaction transfers up to three methyl group from S-adenosyl-L-methionine (SAM) to lysine, or up to two to arginine. As a result, lysine can be mono-, di- or trimethylated, whereas arginine can be mono- or di- (symmetric or asymmetric) methylated (Kaniskan et al., 2018; Morera et al., 2016). This histone modification does not affect this electrostatic bond, but instead indirectly influences the recruitment and binding of different regulatory proteins to chromatin (Kaniskan et al., 2018; Morera et al., 2016). Contrary to the histone acetylation, the histone methylation, in turn, has either a transcriptional permissive or repressive character, depending on the position of the modified residue within the histone tail and/or the number of modifying methyl groups added (Bannister and Kouzarides, 2011; Swygert and Peterson, 2014). Thus, methylation on H3K4, K36, and K79 is associated with activation, whereas H3K9, K27, and H4K20 are associated with repression (Kouzarides, 2007).

3. Histone phosphorylation

Histone phosphorylation is highly dynamic and is controlled by two types of enzymes with opposing modes of action. While kinases add phosphate groups, phosphatases remove the phosphates (Bannister and Kouzarides, 2011; Rossetto et al., 2012). This histone modification increases the negative charge on the histone and alters the structure of the nucleosome, and it's important for DNA damage repair, control of chromatin compaction in mitosis and meiosis, and, similar to histone acetylation, the regulation of transcriptional activity. Phosphorylation can

work in conjunction with other histone modifications. For example, phosphorylation of histone H3S10 can affect the acetylation in H3K9 and H3K14 (Edmondson et al., 2002; Lo et al., 2000) and can interact with H4K16ac and induce transcription elongation (Zippo et al., 2009).

Table 1.1 List of histone modifications in eukaryotes and their role. Modified from Lawrence et al., 2016.

Histone	Modification	Role
H2A	H2AS1ph	Mitosis
	H2AK4/5ac	Chromatin assembly
	H2AK7ac	Transcriptional activation
	H2AK119ph	Spermatogenesis
	H2AK119uq	Transcriptional repression
H2B	H2BS14ph	Apoptosis
	H2BS33ph	Transcriptional activation
	H2BK5ac	Transcriptional activation
	H2BK11/12ac	Transcriptional activation
	H2BK15/16ac	Transcriptional activation
	H2BK20ac	Transcriptional activation
	H2BK120uq	Spermatogenesis/meiosis
H2BK123uq	Transcriptional activation	
H3	H3K4me2	Permissive chromatin
	H3K4me3	Transcriptional elongation
	H3K9me3	Transcriptional repression; imprinting; DNA methylation
	H3R17me	Transcriptional activation
	H3K27me3	Transcriptional silencing; X-inactivation; gene poising
	H3K36me3	Transcriptional elongation
	H3K4ac	Transcriptional activation
	H3K9ac	Histone deposition; transcriptional activation
	H3K14ac	Transcriptional activation; DNA repair
	H3K18ac	Transcriptional activation; DNA repair; DNA replication
	H3K23ac	Transcriptional activation; DNA repair
	H3K27ac	Transcriptional activation
	H3T3ph	Mitosis
	H3S10ph	Mitosis; meiosis; transcriptional activation
	H3T11/S28ph	Mitosis
H4	H4R3me	Transcriptional activation
	H4K20me1	Transcriptional silencing
	H4K20me3	Heterochromatin
	H4K5ac	Histone deposition; transcriptional activation; DNA repair
	H4K8ac	Transcriptional activation; DNA repair
	H4K12ac	Histone deposition; telomeric silencing; transcriptional activation; DNA repair
	H4K16ac	Transcriptional activation; DNA repair
	H4S1ph	Mitosis

IV. Transcription factors

Transcription factors (TFs) are proteins capable of bind DNA in a sequence-specific manner to regions called binding sites and regulates transcription (Vaquerizas et al., 2009). All TFs contain a DNA binding domain (DBD) that recognizes specific sequences, and their activity determines how cells function and respond to cellular environments. Once the TFs interact with their binding sites can recruit transcriptional cofactors to alter the chromatin environment. Also, these TFs bind to regulatory regions (especially promoters) and facilitate the assembly of a pre-initiation complex (PIC), which is composed of general transcription factors (GTFs) and RNA polymerase II (RNAPII) (Venters and Pugh, 2009). The PIC, containing the GTFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIF), the RNAPII, and the mediator proteins assembles the transcription apparatus (Orphanides et al., 1996). The PIC assembly is a multistep process, reviewed in (Lee and Young, 2000) and (Sainsbury et al., 2015).

There are methods to characterize TF-DNA binding preferences *in vivo* e *in vitro*. *In vivo* approaches can reveal TF binding events in a particular biological condition (e.g., cell type, treatment, kinetics), while *in vitro* methods are well suited for large-scale characterization of intrinsic TF binding sequence preferences (Inukai et al., 2017). The combination of *in vivo* and *in vitro* approaches has shed some light on features that influence TF-DNA association (Fig. 1.4)

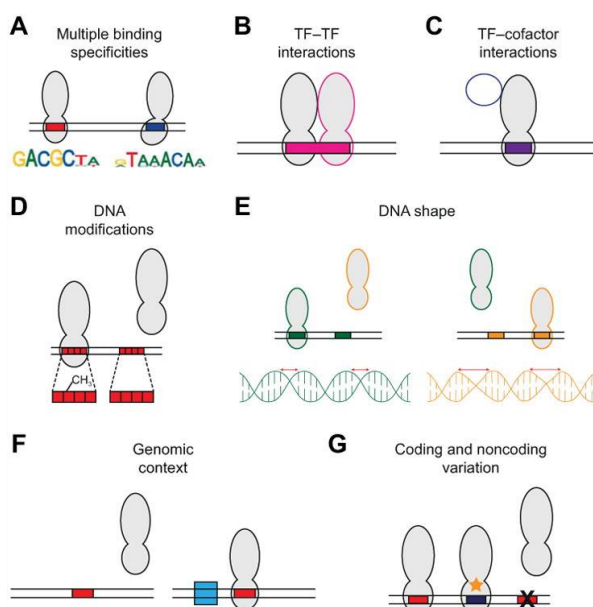


Figure 1.4 Features of TF/DNA binding sites. *TF-level features*: (A) Several TFs can display binding specificity for multiple, distinct nucleotide sequence motifs. Interactions between (B) TFs and (C) TFs and non-DNA-binding cofactors can specify distinct binding site motifs from the monomeric TF motif. *DNA-level features*: (D) DNA modifications, such as 5mC (left), can modulate TF binding. (E) Numerous TFs use DNA shape readout, such as minor groove width (depicted by red arrows), and rotational parameters such as helix twist, propeller twist, and roll, as part of TF–DNA recognition. (F) Sequences and features outside of the binding site motif (depicted by blue box), such as GC content and / or DNA shape, can modulate TF–DNA binding. (G) Genetic variation in either the TF protein sequence (depicted by orange star, middle) or the DNA binding site (depicted by X, right) can alter TF–DNA binding. Modified from (Inukai et al., 2017)

Some TFs can contain multiple, independent DBDs, which allows them to recognize different DNA sequences (Siggers and Gordan, 2014). However, most TFs demonstrate singular binding specificities. All these intrinsic DNA binding specificities are frequently summarized as “motifs”: models representing a short (6 to 12 base pair) consensus sequence of a TF binding sequence (Stormo, 2013). These motifs can be used to scan sequences of interest in genomic regions and predict TF binding sites.

III. *Cis*-regulatory elements

The *Cis*-regulatory elements (CRE) are DNA sequences (generally, non-coding DNA) that contain all the regulatory elements (TF, histone modifications, etc.) to activate and sustain transcription (Ong and Corces, 2011). There are two major CRE, named to their relative position from a gene transcription start site (TSS): TSS-proximal regulatory element (i.e. promoter) and TSS-distal regulatory element. The TSS-distal regulatory elements can positively affect gene expression (i.e. enhancer) or negatively (silencer) (Montgomery et al., 1990), while others act as boundary elements (insulators), preventing genes to be affected by the transcriptional activity of neighboring domains. One CRE can regulate several genes, and one gene can be regulated by multiple CREs (Moorthy et al., 2017).

Active CRE is closely related to chromatin states that allow TF binding. Some of these histone modifications include, but are not limited to H3K27ac for active enhancers, H3K4me3 for the active promoter, and the combinatory of H3K27me3 and H3K4me1 for poised enhancer. Techniques like DNase-seq (DNase I hypersensitive sites sequencing), ATAC-seq (assay of transposase-accessible chromatin using sequencing), MNase-seq (micrococcal nuclease digestion sequencing), and FAIRE-seq (formaldehyde-assisted isolation of regulatory elements) allow the identification of those open chromatin states (Buenrostro et al., 2013; Crawford et al., 2006b; Giresi et al., 2007; Schones et al., 2008). However, these techniques will not discern between epigenetic features.

1. Promoters

The promoters are *cis*-regulatory elements that define where transcription of a gene by RNA polymerase begins. Promoter sequences are typically located directly upstream or at the 5' end of the transcription initiation site. The promoter region is integrated by two elements: the core promoter region, which contains the TSS, and the proximal promoter region, which contains specific TF binding sites. The core promoter is defined as the minimal region of DNA around the TSS that can induce transcription (± 50 bp). The core promoter acts as a binding platform for the RNAPII and the GTFs to form the PIC and initiate transcription. Although core promoters are sufficient to start transcription, its basal activity is low. Several core promoter motifs have been identified: TATA-box, CpG island, initiator (Inr), TFIIB recognition elements (BRE), motif ten element (MTE) and downstream core promoter element (DPE) (Deng and Roberts, 2005; Juven-Gershon and Kadonaga, 2010; Kadonaga, 2012; Lagrange et al., 1998; Lim et al., 2004; Sandelin et al., 2007). These motifs, their relative position from the TSS, and the protein binding to them are summarized in Table 1.2 (Haberle and Stark, 2018).

Due to the low activity of the core promoter, they require proximal regions upstream core promoter generally limited to a few hundred base pairs. This proximal region is where TF and activator proteins bind to increase the rate of transcription by interacting with the GTFs and is associated with tissue-specific expression (Chen et al., 1997; Lewis et al., 2005; Smith et al., 2006).

Table 1.2 Core promoter motif and the TFs that binds them. Adapted from (Haberle and Stark, 2018).

Element	Relative position to TSS	Consensus sequence	Binding protein
TATA-box	-31 to -24	TATAAWWR	TBP
Inr	-1 to +1	YR	NA
	-3 to -3	BBCABW	
DRE	-100 to -1	WATCGATW	Dref
BREu	-38 to -32	SSRCGCC	TFIIB
BREd	-23 to -17	RTDKKKK	TFIIB
TCT	-2 to -6	YYCTTTY	NA
MTE	+18 to +29	CSARCSSAACGS	TAF1 and TAF2
DPE	+28 to +34	RGWCGTG	TAF6 and TAF9
	+28 to +33	RGWYVT	
	+24 to +32	GCGWKCGGTT	

The classical term of ‘promoter’ is an arbitrarily sized region that contains both the core promoter region and the proximal promoter, even though only part of this region may be important for the regulation of transcription initiation (Andersson and Sandelin, 2020).

Histone configuration and modifications

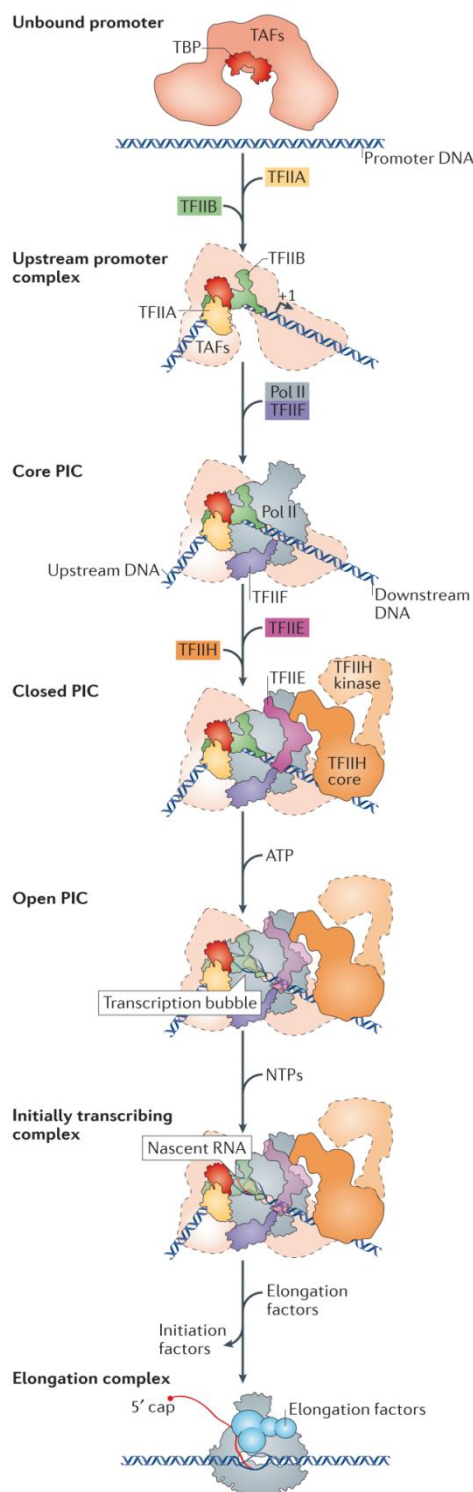
Active transcription occurs in regions that are nucleosome –depleted (NDR) and flanked by two well-positioned nucleosomes (upstream and downstream the TSS) with specific histone variations and modifications (Jiang and Pugh, 2009). For example, the histone variations observed in the nucleosomes adjacent to the NDR of promoters and enhancers, are enriched with histone variants H3.3 and H2A.Z (Jin et al., 2009), while the nucleosome located at the downstream edge of the NDR (+1) contains trimethylation in the histone 3 in its lysine 4 (H3K4me3) (Lauberth et al., 2013).

Actively transcribed genes are characterized by a high level of H3K4me3 and H3K27ac. Moreover, the promoter regions with high CpG content are enriched with H4k20me1 and H2K27ac, whereas promoters with low CpG content are enriched in H3K4me3 and H3K79me1 (Karlic et al., 2010). Not all histone modification regulates positively. The histone modification H3K27me3 is associated with inactive promoters and avoids the transcription elongation, while H3K36me3 acts as a mark for HDACs to deacetylate the histone, preventing the transcription (Carrozza et al., 2005).

Transcription elongation and directionality

As mentioned before, the transcription initiation requires the assembly of the PIC, composed of the GTFs and the RNAPII (Sainsbury et al., 2015). PIC assembly is followed by dsDNA dissociation and the formation of the open PIC that allows the incorporation of the first nucleotides to the nascent transcript. Finally, the RNAPII can continue transcribing by releasing the core promoter and the GTFs binding it (Fig. 1.5). The core promoter is sufficient to initiate transcription but generally has a low basal activity which can be further suppressed by histone

modifications or activated by often more distally located enhancers (Banerji et al., 1981; Haberle and Stark, 2018).



Once initiated, promoters can produce long, stable mRNAs from a gene core promoter in the downstream direction of the gene and short, unstable upstream antisense RNAs (uaRNAs/PROMPTs) from the upstream edge of the nucleosome-depleted proximal promoter, which contains TF binding sites (Andersson et al., 2015a; Core et al., 2014; Core et al., 2008; Preker et al., 2008; Seila et al., 2008). These short non-coding RNAs are transcribed by separated RNAPII complexes in a process called divergent transcription (Haberle and Stark, 2018) and is observed at 50-80% of promoters from human protein-coding genes (Core et al., 2014; Duttke et al., 2015).

Antisense transcription will face premature termination leaving a PROMPTs between 500 and 2500 bp, carrying a 3' polyadenylated tail (poly-A) and 5' cap structure and eliminated by the exosome complex (Preker et al., 2011). The bidirectional transcription has also been observed in active enhancers with the production of enhancer RNA (eRNA), but in contrast with the PROMPTs, in general, they lack 5' cap and 3' poly-A (Fig. 1.6).

There are two major types of methods to identify gene TSSs that allow mapping the core promoter regions across the genome: (i) Steady-state RNAs analysis allows the identification of the TSS and the abundance

Figure 1.5 Schematic of Pol II transcription initiation. GTF (various colors), RNAPII (grey), transcription elongation factors (blue). NTP, nucleoside triphosphate, TATA box-binding protein (TBP), and TAF, TBP-associated factor. Adapted from (Sainsbury et al., 2015)

of the produced RNAs on full-length cDNA sequencing (flcDNA-seq) (Kawai et al., 2001) or high-throughput sequencing for the first 20-50 nucleotides (TSS-seq or cap analysis of gene expression, CAGE) (Takahashi et al., 2012; Yamashita et al., 2011), and (ii) Nascent RNA sequencing techniques measure the rate of transcription rather than the RNA levels. These techniques are based on the incorporation of labeled nucleotides in the cells and sequencing either 3' end (Global run-on sequencing, GRO-seq; or nuclear run-on sequencing, PRO-seq) (Core et al., 2008; Kwak et al., 2013), the nascent 5' RNA (GRO-cap, PRO-cap) (Kruesi et al., 2013) or the RNAs bound to the RNAPII (native elongation transcript sequencing, NET-seq, mammalian NET-seq, mNET-seq) (Mayer et al., 2015; Nojima et al., 2015).

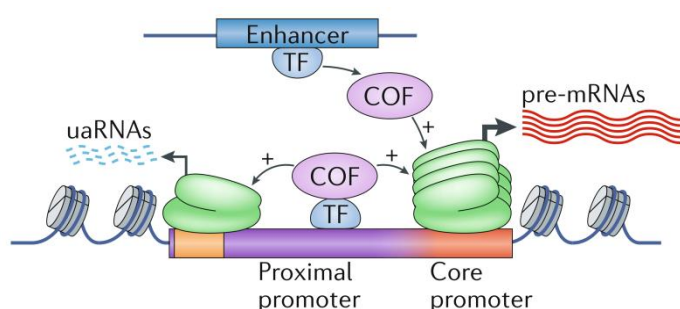


Figure 1.6 Functional model of transcription initiation in promoters. Adapted from (Haberle and Stark, 2018)

2. Enhancers

Enhancers are cis-regulatory sequences located distally to the TSS and play a major role in the regulation of the gene expression. The first enhancer identified was a tandem repeat (72 bp) from the simian vacuolating virus 40 (SV40) (Banerji et al., 1981). Later, it was demonstrated that this sequence could interact with a heterologous promoter of the β -globin gene, increasing the transcription by 200-fold over a long distance in an orientation-independent manner (de Villiers and Schaffner, 1981). This observation was confirmed in other viruses, but the first endogenous enhancer in mammals was found in the intronic region of immunoglobulin genes coding for the heavy and light chains (Banerji et al., 1983; Gillies et al., 1983; Queen and Baltimore, 1983). Since then, several enhancers in different cells, tissues, and conditions have been reported and studied. Nowadays the Encyclopedia of DNA Elements (ENCODE) Consortium, has identified hundreds of thousands of putative enhancer loci in the human genome, using genomic data sets (Consortium, 2012). The data suggest that multiple enhancers can regulate a single gene and that multiple genes can be regulated by a single enhancer (Medina-Rivera et al., 2018). Enhancers can act over its target promoter as far as 1 Mb, regardless of its relative sequence orientation, and that they do not necessarily regulate the closest genes (Santiago-Algarra et al., 2017).

The enhancer sequence harbors a high density of DNA-binding TFs that will vary from cell to cell, and are normally lineage-specific; making the enhancer function cell-dependent (de Villiers et al., 1982). However, when the cell responds to stimuli, this lineage TFs requires the cooperation of signal-dependent TF in response to the cellular environment.

The chromatin patterns in enhancers are more variable and cell type-specific than the promoters. Typically, the active enhancers are depleted of nucleosomes with enrichment in

H3K27ac (Creyghton et al., 2010) and a high H3K4me1 to H3K4me3 ratio (Zhou et al., 2014), often in conjunction with DNase I hypersensitivity (Ernst et al., 2016). These two histone modifications became the state-of-the-art in the search algorithms for putative enhancers in ENCODE and Roadmap Epigenomics (Ernst et al., 2011; Hoffman et al., 2012). The presence of high acetylation in enhancers is consistent with the binding of p300 and related acetyltransferases serving as another way to locate putative enhancers (Hatzis and Talianidis, 2002).

In addition to the histone modifications H3K27ac and the high H3K4me1/H3K4me3 ratio, integrative analysis also suggests that enhancers are enriched in H2BK5me1, H3K4me1, H3K27me1, and H3K36me1 suggesting the redundancy in the histone modifications in enhancers (Hon et al., 2009). Recent studies, however, have found also enrichment of H3K4me3 on active enhancers, supporting further this idea (Pekowska et al., 2011). Nonetheless, the enhancers can be further categorized into various generally accepted states: neutral/intermediate (marked by H3K4me1), poised (H3K4me1 and H3K27me3), and active (H3K4me1 and H3K27ac) (Creyghton et al., 2010; Natoli and Andrau, 2012; Rada-Iglesias et al., 2011; Zentner et al., 2011).

Enhancer not only recruit TF and co-factors like p300 but also recruit the RNAPII and can transcribe bidirectional short non-coding RNAs, termed enhancer RNAs (eRNAs) (Kim et al., 2010; Koch et al., 2008; Natoli, 2010). This fortuitous transcription is the result of high activator concentrations, the presence of degenerate Inr and TATA-box motifs at TSSs within enhancers (Andersson et al., 2014; Core et al., 2014; Scruggs et al., 2015), and the bidirectional initiation pattern at enhancers (Core et al., 2014). In any case, the eRNA transcription levels correlate with the proximal gene follows the activity of the enhancer is inducible (Kim et al., 2010) and cell-type specific (Andersson et al., 2014) (Fig. 1.7). Thanks to this correlation, the production of eRNA can be utilized to identify the location of active enhancers and as a marker of regulatory activity (Melgar et al., 2011). Using the CAGE methodology the FAMTOM5 (Functional Annotation of the Mouse/Mammalian Genome) consortium identified active enhancers and validate 75% of their predictions (Andersson et al., 2014). This validation rate is comparable with the 75% obtained when enhancers are defined by their histone modification (Ernst et al., 2011). However, fall low when compared with the 87% validation rate when active enhancers are defined by the binding of p300 (Visel et al., 2009).

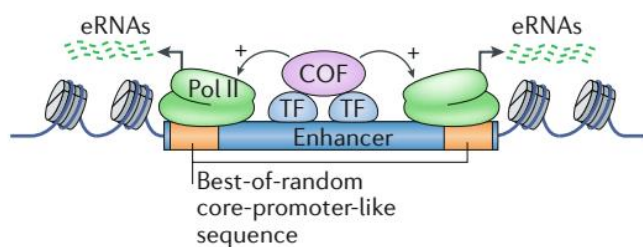


Figure 1.7 Functional model of transcription initiation at genomic enhancers. Adapted from (Haberle and Stark, 2018)

Transcriptional regulation requires some sort of communication between enhancers and promoters. Two models tried to explain how enhancers and promoters communicate (Vernimmen and Bickmore, 2015). The tracking model was the mechanism in which all the enhancer-bound proteins (TFs, co-factors, RNAPII, etc.) move progressively towards the promoter, forming a progressive loop that grows in size until it reaches the target promoter and forms a stable conformation (Hatzis and Talianidis, 2002; Wang et al., 2005). On the other hand, the looping model is broadly more accepted; this model implies the direct interaction between enhancers and promoters facilitated by mediator proteins that help to the recruitment and the stability of the different PIC components [reviewed in (Soutourina, 2018)].

Another class of enhancers has been described as a super-enhancer. This regulatory element shared many typical features of typical enhancers but on a larger scale: on average they are 15-fold larger (over 19 Kb), recruits more mediator protein 1 (18-fold) and BDR4 (16-fold), and have higher levels of H3K27ac (26-fold) and H3K4me1 (10-fold) (Loven et al., 2013; Whyte et al., 2013). Super-enhancers are densely occupied by master transcription factors and Mediator complex. These master TFs establish auto-regulatory networks not evenly distribute but form dense clusters of transcription factors termed 'hotspots' (Siersbaek et al., 2014). Super-enhancer plays a role in the re-reorganization of promoter interactions in cellular differentiation (Madsen et al., 2020; Novo et al., 2018; Siersbaek et al., 2014), circadian gene expression (Kim et al., 2018), and stress response (Schmidt et al., 2015)

Diseases-related cells acquire SEs through various mechanisms, including (i) mutations and genomic alterations like deletions, duplications, translocations, insertions, inversions; (ii) single-nucleotide polymorphisms (SNPs); (iii) chromosomal rearrangements; and (iv) 3D genome structural changes (Qu et al., 2020).

3. Silencers

Silencer elements are yet another type of cis-regulatory element with distal activity, but contrary to enhancers, silencers constitute a transcriptional repressor. They were first identified in 1985 when the yeast mating-type loci revealed that distal silencer elements could control gene expression from afar (Brand et al., 1985). Shortly after, a silencer element was identified upstream of the rat insulin gene (Laimins et al., 1986).

There are only a handful of identified silencers (Goodbourn et al., 1985; Henson et al., 2014; Li et al., 1993; Petrykowska et al., 2008) and their characteristics are yet to be described. However, some models have been proposed for its repression activity. (i) Silencers can block activator elements or (ii) they can directly block the PIC assembly.

Conversely, in a recent study in *Drosophila*, the silencers identified were previously identified in a different tissue as enhancers (Gisselbrecht et al., 2020). These bifunctional enhancer/silencer elements suggest other models for gene transcription repression, where this bifunctional element can act as a silencer or as an enhancer depending on the tissue, or these regions contain separable but adjacent silences and enhancers (Halfon, 2020) (Fig. 1.8).

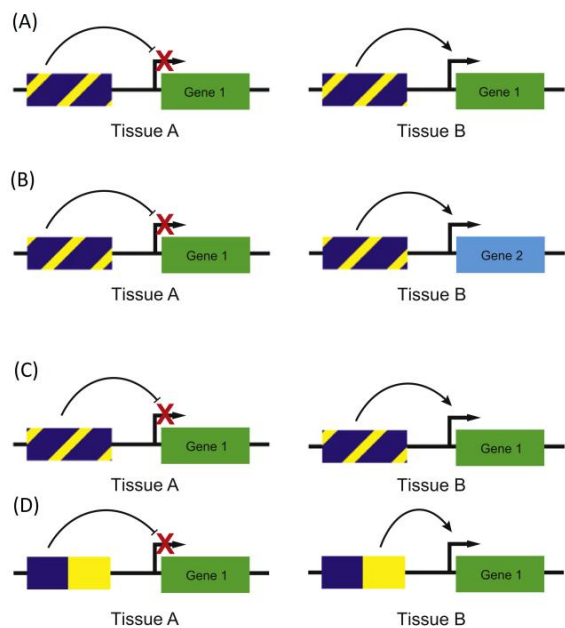


Figure 1.8 Transcriptional Silencers as Bifunctional Regulatory Elements.

In (A), the bifunctional element regulates a single gene, silencing 'gene 1' in tissue A but activating 'gene 1' in tissue B. In (B), the bifunctional element is a silencer with respect to 'gene 1' in tissue A, but an enhancer with respect to 'gene 2' in tissue B. In (C), silencer functions (blue) and enhancer functions (yellow) are mixed within a single bifunctional sequence. An alternative is shown in (D), where silencer and enhancer functions reside in separable but adjacent sequence elements. Adapted from (Halfon, 2020)

4. Insulators

Chromatin insulator sequences or boundary elements are the names given to a *cis*-regulatory element that possesses the ability to protect genes from inappropriate signals emanating from their surrounding environment (blocking promoter-enhancer interaction) and/or serve as barriers against heterochromatin silencing effects (West et al., 2002).

The enhancer-blocking activity of an insulator depends on the positioning of the insulator, as it must be located between the enhancer and the promoter. One key element in its insulating function is the insulator protein CTCF (CCCTC-binding protein). The CTCF not only block enhancer activity but also mediates the 3D conformation of the chromatin required for the enhancer-promoter interaction. The CTCF also mediates the boundary effect in insulators by restricting the spread, from one domain to the others, of histone modification associated with active genes (H3K27ac) or inactive (H3K27me3).

5. Long non-coding RNAs

Less than 2% of the human genome encodes proteins, while around two-thirds of all genome is being pervasively transcribed into many non-coding RNAs (ncRNAs). Coding transcripts are six times more frequent in the cytosol, whereas non-coding transcripts are five-times more frequent in the nucleus (Djebali et al., 2012). Among all these ncRNAs, those who exceed 200 nucleotides long are named long non-coding RNAs (lncRNAs), this includes the eRNAs and the antisense transcripts (Fig. 1.9) (Derrien et al., 2012; Mathieu et al., 2014). On average the lncRNAs are shorter than mRNA (592 bp vs 2453 bp) (Derrien et al., 2012).

The lncRNAs are less conservative than the mRNA, but the conservation in sequence does not always correlate with their function (Derrien et al., 2012). There have been proposed several mechanisms in how the lncRNAs can regulate transcription either in *cis* or *trans* (Kopp and Mendell, 2018). The proposed mechanisms in *cis* include (i) the interaction sequence-specific

between the mature lncRNA with a coding-gene, (ii) the recruitment of the spliceosome machinery, or (iii) the promoter of the lncRNA act as an enhancer and interact with the promoter region of the protein-coding gene, independently of the lncRNA. In addition to the regulation in cis, the mechanisms described for trans-regulation are categorized into three groups: (i) lncRNAs that can reorganize the chromatin state in regions afar from their own, (ii) lncRNAs that associate with mRNA or proteins affecting its interaction, and (iii) lncRNAs that modify the nuclear structure.

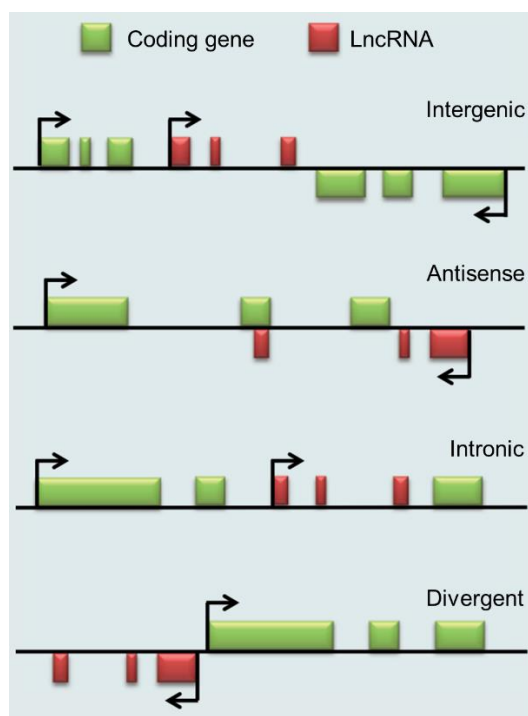


Figure 1.9 Illustration scheme for different lncRNA classes. Adapted from (Mathieu et al., 2014)

Whether the lncRNAs can regulate gene expression at transcriptional (Ng et al., 2012; Rinn et al., 2007; Zhao et al., 2008) or post-transcriptional level (Cesana et al., 2011; Gong and Maquat, 2011; Yoon et al., 2012), it is now clear that they can contribute to the development of human diseases (reviewed in (DiStefano, 2018; Sparber et al., 2019)). For example, one of the first described lncRNAs, H19 (Bartolomei et al., 1991), is currently used as a biomarker in the diagnosis of Silver-Russel and Bechwith-Wiedemann syndromes (Wakeling, 2011). Some reports suggest that H19 can regulate the expression of several genes necessary for embryonic growth and development (Gabory et al., 2009; Monnier et al., 2013). However this effect may not require the specific lncRNA transcripts themselves and instead involved general processes associated with their production, including enhancer-like activity of gene promoters, the process of transcription, and the splicing of the transcript (Engreitz et al., 2016).

6. Similarities between enhancers and promoters

The focus of several studies is centered on two major classes: enhancers and promoters. However, as described in Annex 1, this definition is vague and dichotomic. Improvements in the DNA sequencing with high-throughput sequencing and the reporter assays help the identification of active regulatory elements on a genomic scale. These tools led us to realize that enhancers and promoters share many properties; for example, the chromatin architectures are

very similar, both can recruit RNAPII and GTFs, enhancers can drive divergent local transcription and promoters can have enhancer activity. Several features associated with promoters and enhancers are approached in our published review [Table 1 in Annex 1 (Medina-Rivera et al., 2018)].

7. Epromoter

The Epromoters are a novel class of cis-regulatory element defined as a promoter element that displays enhancer activity in a functional experimental setting and might regulate the expression of a distant gene. In table 2 of annex 1, we summarize the early evidence of promoters with enhancer activity. These promoter-promoter interactions were first identified by 3C-based methods, suggesting that this regulatory network is common in mammalian cells (Pancaldi et al., 2016). Using various powerful techniques that incorporate high-throughput sequencing into reporter assays have enabled quantitative and straightforward measurements of enhancer activity. Epromoters has been identified in some studies *in vitro* in *Drosophila* (Zabidi et al., 2015), mouse (Nguyen et al., 2016b) and human, and some have validated its functional activity *in vivo* (Dao et al., 2017; Diao et al., 2017b; Engreitz et al., 2016) (Fig. 1.10). A list of experimentally validated promoters with enhancer-like Activity in their natural context is summarized in table 3 in annex 1.

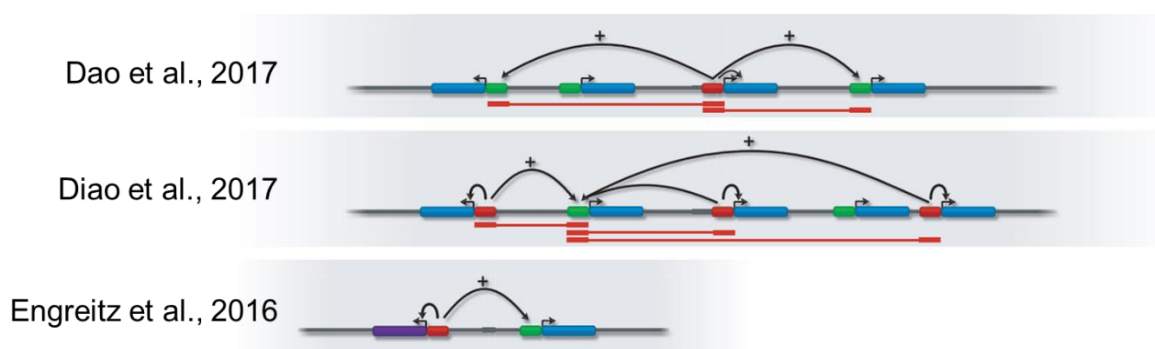


Figure 1.10 Epromoter function *in vivo*. One or multiple Epromoters (red) can regulate the expression of one or more distal genes. Modified from (Medina-Rivera et al., 2018).

8. Long-range chromatin interactions

As mentioned before, the DNA is wrapped around nucleosomes and forms the chromatin. Depending on the histone modification, histone variants, and DNA methylation, the chromatin can be transcriptionally active or inactive. The chromatin can form local loops that can be further folded into larger domains and compartments. Each chromosome occupies a specific space in the nucleus called the chromatin territories (CT) (Yamamoto and Saitoh, 2019) (Fig. 1.11). These chromatin territories are generally accepted as a model of non-random but variable chromosomal organization.

Each chromosome is comprised of many chromatin domains and they are defined depending on their nuclear association, including the topologically associating domains (TADs), lamina associating domains (LADs), and the nucleolus-associated domain (NADs).

The TADs size varies but it ranged from 215 Kb to 1.2 Mb and represents the functionally isolated units of the genome. This means that the genes within TADs can be co-regulated and that transcriptional activity does not spread between neighbor TADs (Pope et al., 2014). In mammals, the TADs are strongly conserved (Dixon et al., 2012) and their boundaries are enriched with CTCF and cohesin binding sites, histone marks of active chromatin, and housekeeping genes (Ea et al., 2015).

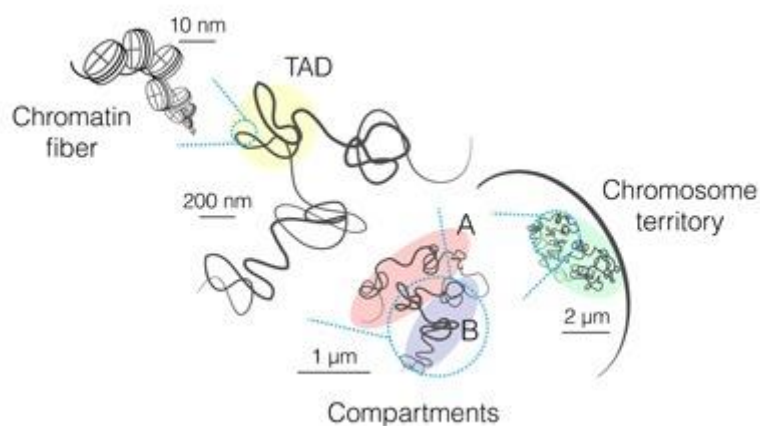


Figure 1.11 Schematic view of chromosome folding inside the nucleus. Adapted from (Szabo et al., 2019).

One of the key methodologies to better understand the chromosome organization within the nucleus has been the chromosome conformation capture (3C) assay in combination with next-generation sequencing (NGS). These techniques, especially the Hi-C methodology, allows unbiased identification of chromatin interactions across the entire genome (Lieberman-Aiden et al., 2009). The corner peaks in the Hi-C maps between TADs are enriched in CTCF binding sites, suggesting the formation of the loops between them (Fig. 1.12). These loops form between CTCF sites positioned in a convergent orientation and changes in the orientation of a single CTCF can affect the positioning of a TAD boundary (Lupianez et al., 2015; Sanborn et al., 2015)

The current mechanism for the CTCF-looping formation of TADs is referred to as the “loop extrusion” model. According to this model, the cohesin complex binds to the chromatin and starts to extrude the loop until it encounters the two convergent and bound CTCF sites (supplemental movie in (Fudenberg et al., 2017)).

The CTCF plays a crucial role in brain development and neural function (Beagan et al., 2017; Sams et al., 2016), cardiovascular (Gomez-Velazquez et al., 2017), limb development (Andrey et al., 2017), and is crucial to maintain the established insulated neighborhoods that guaranty proper gene expression. This is particularly true in embryonic stem cells (ESCs). These cells can self-renew and differentiate thanks to the chromatin landscape influenced by CTCF. When CTCF is deleted genome-wide from murine ESCs (mESCs) it results in changes in gene expression of 370 and 4996 genes after 24 and 96 hours of depletion, respectively. Also, this results in the loss of the chromatin loop, leading to an interaction between domains normally separated by CTCF (Nora et al., 2017). Also, somatic mutations at the CTCF binding motif can lead to the

development of cancer, including colorectal cancer (Katainen et al., 2015), leukemia (Hnisz et al., 2016), and breast cancer (Oh et al., 2017).

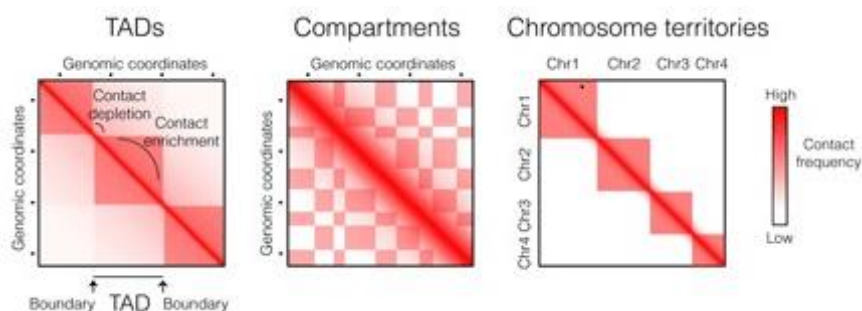


Figure 1.12 Schematic representation of Hi-C maps at different genomic scales, reflecting the different layers of higher order chromosome folding. Adapted from (Szabo et al., 2019).

Moreover, the genome is also organized into separated membraneless sub-compartments involved in ribosome biogenesis (nucleolus), transcriptional factories associated with active RNAPII (Cook, 2010), Polycomb bodies, paraspeckles, viral replication compartments, and constitutive heterochromatin compartments (fig. 1.13) (A and Weber, 2019). Different biophysical concepts are currently used to explain the formation of “chromatin bodies” in a self-organizing manner and without consuming energy. The most accepted models are the liquid-liquid phase separation (LLPS) and the polymer-polymer phase separation (PPPS). The LLPS mechanism is driven by multivalent interactions among soluble components, while the PPPS mechanism is based on bridging interactions that tend to compact the chromatin fiber (Erdel and Rippe, 2018).

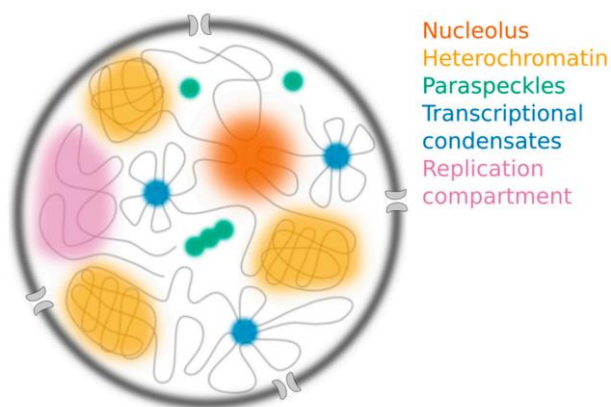


Figure 1.12 Scheme of nuclear compartments. Modified from A and Weber, 2019.

Chapter 2. Identification and functional characterization of transcriptional regulatory elements

I. Identification of transcriptional regulatory elements

The sequence of the human genome was first published 19 years ago (Lander et al., 2001; Venter et al., 2001) and marked one of the biggest breakthroughs in biomedical history. Today, we are still deciphering our own “instruction manual”. We know that the transcriptional regulatory elements are key elements in the usage of the genome information during development, responses to stimuli, cell homeostasis, and disease. Therefore, identify and understand the function of these regulatory elements is challenging. The systematic recognition of regulatory elements has relied on the identification of their epigenomic signature, including histone modifications and DNA accessibility.

1. Identification of histone modifications and non-histone proteins

The chromatin immunoprecipitation (ChIP) is an antibody-based technique that allows us to identify where in the genome the histone modifications and non-histone proteins (e.g. transcription factors and cofactors) are. The technique is quite simple, use a specific antibody to isolate a histone modification or a protein and the chromatin binding to it.

Depending on the starting material there are two techniques: native ChIP (N-ChIP) and crosslinked ChIP (X-ChIP). For the N-ChIP there is no fixing agent to fix proteins to the chromatin. The chromatin is isolated and digested with nucleases. This method only allows for the detection of histone modifications. The X-ChIP is the most common and requires a chemical fixator like formaldehyde to crosslink the proteins. The chromatin can then be fragmented by nucleases digestion or sonication. This method permits the isolation of histone and non-histone proteins (Fig 2.1)

Once the chromatin is isolated and the DNA purified, is possible to do single gene analysis, by PCR, called ChIP-PCR; whole-genome analysis by microarrays, called ChIP-chip; or by whole genome analysis using high-throughput sequencing, also called ChIP-seq.

Before the surge of the Next Generation Sequencing (NGS), the microarrays were used to identify by hybridization DNA fragments obtained by ChIP (Blat and Kleckner, 1999; Ren et al., 2000). Now in the era of NGS, the ChIP-seq is one of the favourite methods to map epigenetic marks and DNA-binding proteins genome-wide (Barski et al., 2007; Johnson et al., 2007; Robertson et al., 2007). There are multiple public databases available to download ChIP-seq data for histones and non-histone proteins, like the ENCODE project.

Although is one of the most used methods to identify epigenomic marks, there are several important issues to consider: (i) excessive crosslink can mask the epitope, therefore reducing the antigen-antibody efficiency; (ii) antibodies should be validated for binding an exposed epitope in the native protein; (iii) cell number can vary for each particular target and requites

optimization; and (iv) the amount of data generated requires an analysis that, until date, is not “user-friendly”.

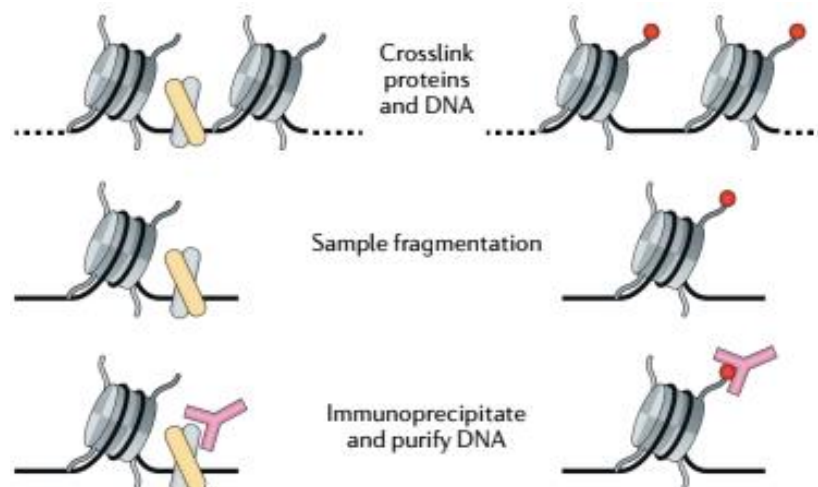


Figure 2.1 Basic schematic for chromatin immunoprecipitation targeting DNA-binding proteins (left) or histone modifications (right)

Finally, with the NGS technology in continuum optimization now is possible to do high-throughput sequencing in single cells, and recently several methods have been proposed for single-cell ChIP-seq (scChIP-seq) in native and fixed cells. This methodology will allow us to analyse complex tissues and tumours [reviewed in (Nakato and Sakata, 2020)].

Nowadays ChIP-seq is widely used to map where a certain protein binds in the DNA; however, to overcome the background noise it requires a large amount of input material. To overcome this, the novel strategy CUT&RUN (Cleavage Under Targets and Release Using Nuclease) has been developed (Skene and Henikoff, 2017). This method consists of unfixed permeabilized cells, incubated with antibody and followed by binding of a protein A-Micrococcal Nuclease (pA/MNase) fusion protein that cuts the DNA (Meers et al., 2019). The advantage of this technique is that it requires a lower number of cells (between 100 and 1000 cells), provides base-pair resolution, and does not require chromatin fixation.

2. Chromatin accessibility

Chromatin accessibility is the degree to which DNA-binding proteins (TFs and transcription machinery) can find their targets in the DNA and it is determined by the topological organization of nucleosomes (Fig. 2.2). The basic principle of the chromatin accessibility methods relies on either enzymatic methylation or the cleavage of the chromatin. These methods do not distinguish between accessibility due to TFs binding to chromatin or any epigenetic marks. In 1973, studies of chromatin fragmentation with DNA endonuclease showed that nucleosomes confer periodic hypersensitivity (Hewish and Burgoyne, 1973), providing the earliest evidence of nucleosome phasing.

The regulatory elements located in highly accessible regions normally have high transcriptional activity (Mattioli et al., 2019), therefore the identification and functional characterization are of great importance.

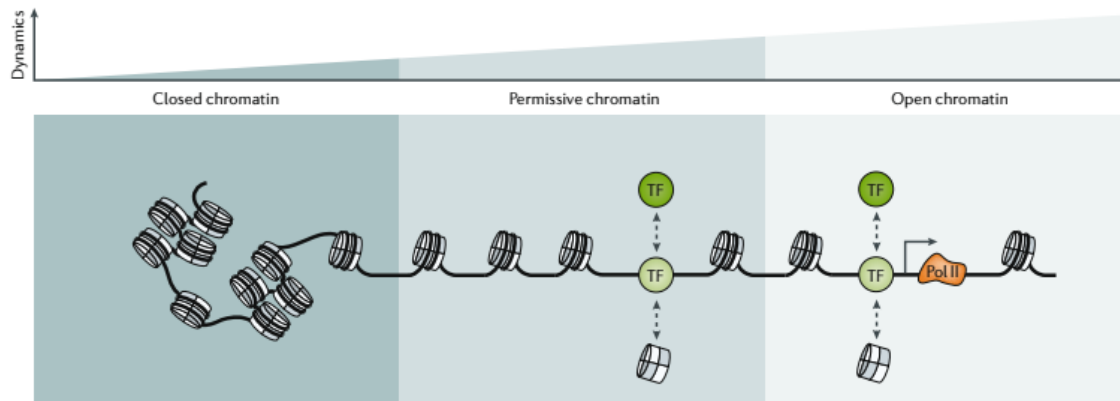


Figure 2.2 In contrast to closed chromatin, permissive chromatin is sufficiently dynamic for transcription factors to initiate sequence-specific accessibility remodelling and establish an open chromatin conformation. Pol II, RNA polymerase II; TF, transcription factor. Adapted from (Klemm et al., 2019)

DNase-seq

This method uses traditional DNase I footprinting to selectively digest nucleosome-depleted regions, also called DNase I hypersensitive sites (DHSs). The first genome-scale analysis for open DNase-sensitive chromatin was back in 2006 and consisted of hybridized DNase I-cleaved fragments from native chromatin onto microarrays spanning approximately 30 Mb of the human genome (DNase-chip) (Crawford et al., 2006a; Sabo et al., 2006). Now, these DNase I-cleaved fragments can be mapped across the genome using short-read sequencing (DNase-seq). There are two major protocols: use a type II restriction enzyme to isolate and subsequently barcode each DNase cut site (single cut) (Boyle et al., 2008), or strict size selection to enrich for sequenceable fragments coming from paired cleavage fragments within DHSs (double cut) (Hesselberth et al., 2009) (Fig. 2.3a).

The DHS sites make up approximately 2% of the genome and about 14% of them are cell-specific (Neph et al., 2012; Thurman et al., 2012), and can be classified based on their genomic locations as transcriptional start site (TSS) DHSs, gene body DHSs, and intergenic DHS (Natarajan et al., 2012).

ATAC-seq

Assay for transposase-accessible chromatin using sequencing is an alternative method for chromatin accessibility. In ATAC, the chromatin is exposed to a hyperactive Tn5 transposase that fragments the DNA and tags the target sequence with sequencing adaptors, in a process called tagmentation (Fig. 2.3b) (Adey et al., 2010). ATAC-seq selectively amplifies double-cleavage events and can be used in native chromatin to profile open chromatin, TF binding, and nucleosome position (Buenrostro et al., 2013). The open chromatin measurements by ATAC are comparable with those in both double-cut and single-cut DNase-seq assays. There are two major advantages in the use of ATAC-seq over the DNase-seq: (i) complex libraries can be generated

with as few as 500 cells, and (ii) libraries can be generated in less than 2 hours, while the DNase is a multiday protocol (Corces et al., 2017). This is a robust method that simplifies protocol time and is done with primary tissue and limited clinical tissues. ATAC methodology can also be coupled with single-cell sequencing (scATAC-seq) allowing the study of open chromatin at single-cell resolution.

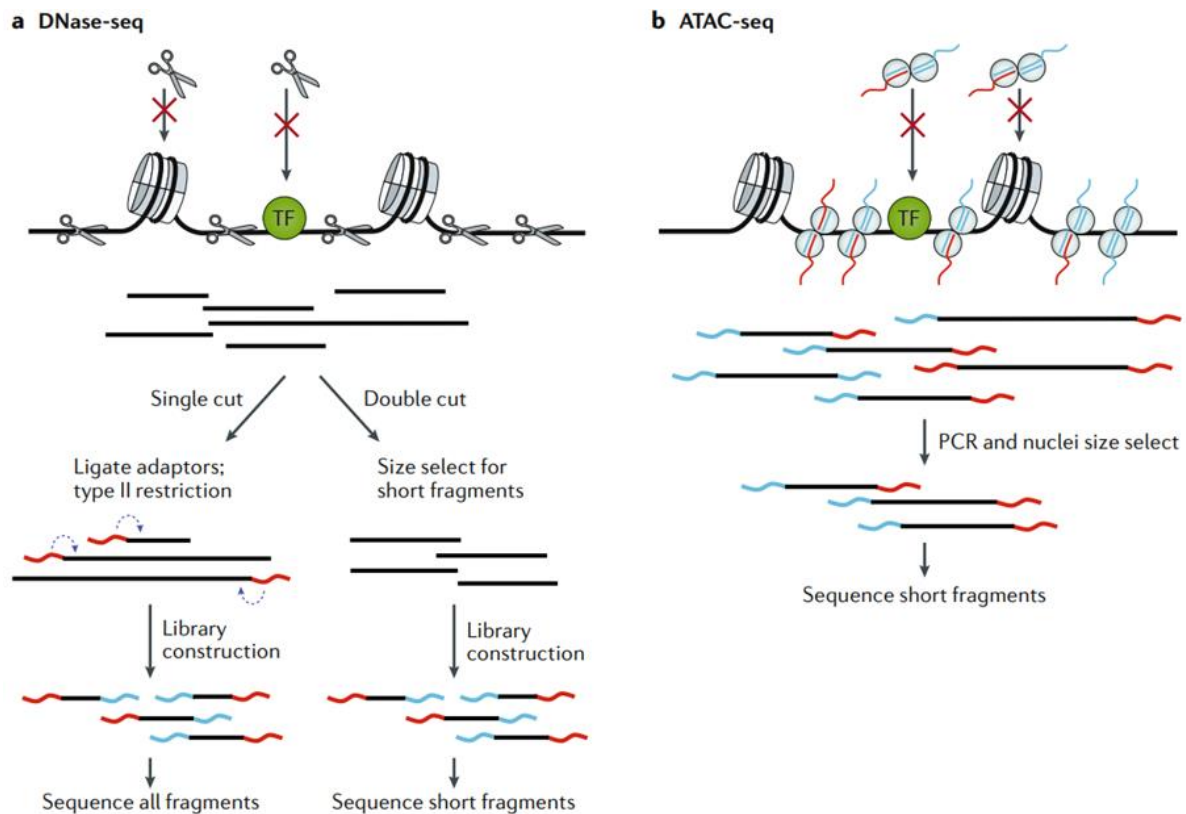


Figure 2.3 Principal methods for measuring chromatin accessibility. (a) DNA-seq. (b) ATAC-seq. Adapted from (Klemm et al., 2019)

MNase-seq

Nucleosome positioning regulates the chromatin accessibility and can be measured by micrococcal nuclease sequencing (MNase-seq). MNase has an endonuclease activity that cleaves internucleosomal DNA and yields mainly mono-nucleosomes, and an exo-nuclease activity that digests DNA not protected by proteins (fig.2.4a).

Undigested DNA is protected by the nucleosomes and subjected to high-throughput sequencing. However, MNase has a sequence bias, and it cleaves DNA about 30 times faster upstream of an A or T than it does of a G or C (Horz and Altenburger, 1981). This result in differences in nucleosome occupancy is dependent on different levels of digestions (Bryant et al., 2008).

NOMe-seq

Nucleosome occupancy and methylome sequencing (NOMe-seq) is based on the native chromatin treatment with GpC methyltransferase M.CviPI, which methylates GpC dinucleotide sites that are not bound by nucleosomes, to create a digital footprint of nucleosome positioning. GpCm does not occur in the human genome, therefore there is no endogenous background of

GpCm (Fig. 2.4b). After, the treated chromatin goes under bisulfate treatment that converts unmethylated Cs to Ts, and therefore, allows the distinction of GpC from GpCm and the CpG from CmpG after sequencing. Thus, this method probes chromatin accessibility and also the methylation status of DNA (Kelly et al., 2012). DNA trinucleotides GCG cannot be used to distinguish between the enforced GpC methylation and the endogenous CpG methylation on the other strand.

NOMe-seq is not an enrichment method, therefore requires a large number of sequencing reads to obtain sufficient depth to determine accessibility levels over the whole genome. However, the absence of enrichment bias and the single molecule character of this technique creates a more quantitative view of the chromatin accessibility landscape than DNase-seq, ATAC-seq, or MNase-seq, as the relative accessibility level of each genomic locus can be directly determined (Klemm et al., 2019; Krebs et al., 2017).

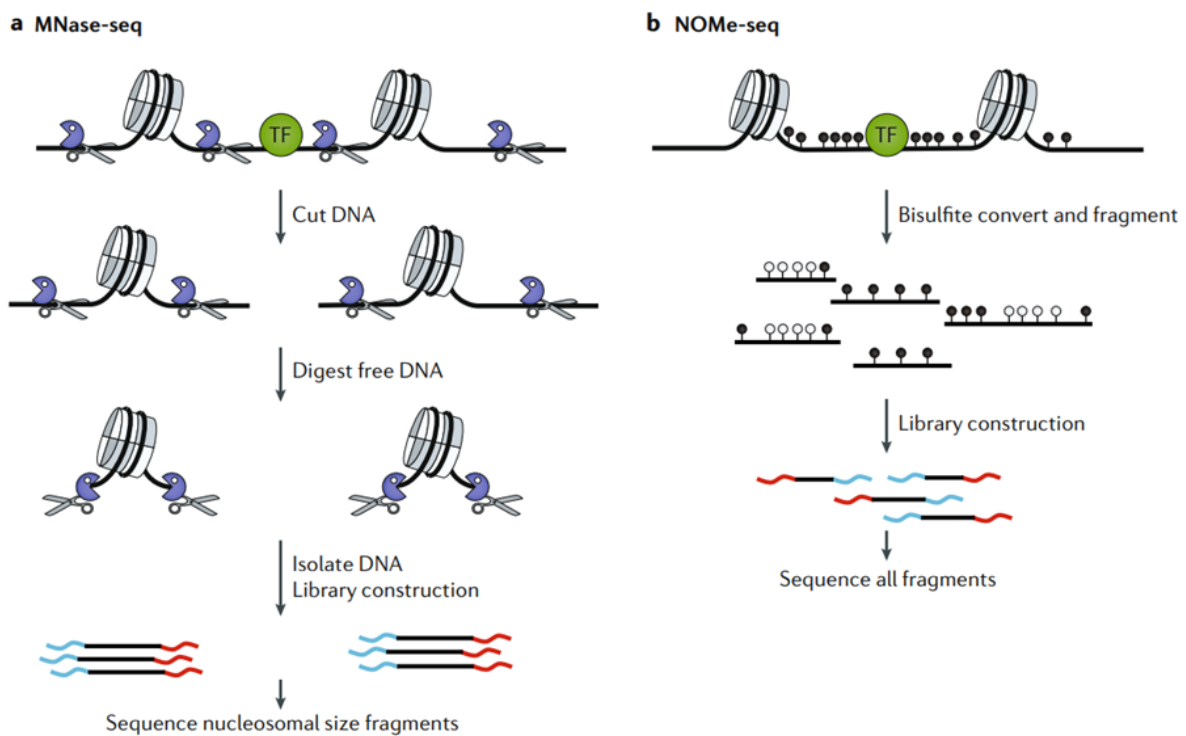


Figure 2.4 Principal methods for measuring chromatin accessibility. (a) MNase-seq. (b) NOMe-seq. Adapted from (Klemm et al., 2019)

FAIRE-seq

The method of chromatin accessibility by formaldehyde-assisted isolation of regulatory elements and sequencing (FAIRE-seq) consists of the fixation of the cell with formaldehyde, fragmentation by sonication, and the purification of free DNA (Fig. 2.5) (Giresi et al., 2007). While this technique can be noisier than DNase-seq, the FAIRE-seq derived mapping highly correlates with DNase-seq open regions (Song et al., 2011). FAIRE libraries can be amplified by PCR (FAIRE-qPCR), by hybridization to a tiling DNA microarray (FAIRE-chip), or by NGS (FAIRE-seq) (Gaulton et al., 2010). This technique offers the advantage that is an antibody- and an enzyme-free method does not require input control and can be applied to tissue samples (Simon

et al., 2012). However, this method can provide low signal close to noise, does not allow specific TF footprinting and the fixation efficiency varies from different tissues (Giresi et al., 2007).

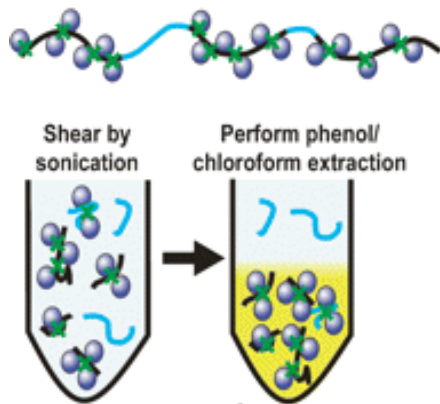


Figure 2.5 Schematic for formaldehyde-assisted isolation of regulatory elements and sequencing, FAIRE-seq Adapted from (Giresi et al., 2007)

3. TSS identification techniques

As mention in Chapter 1, the transcription is a complex process with multilevel control. The knowledge of the exact position of the transcriptional start site (TSS) of any RNA molecule is crucial to discover the regulatory elements that govern it. Two major mechanisms allow the identification of TSS: by analysing the steady-state RNAs or the nascent RNAs.

Steady-state RNA analysis

Once the transcription is initiated, the TSS hallmark is the incorporation of an inverted 7-methyl guanosine cap structure linked via a 5'-5' Triphosphate Bridge to the first nucleotide transcribed by RNAPII. This cap structure has an important effect on the maturation, translations, and stability of the mRNA (Furuichi, 2015). The cap contains a 2',3'-diol structure susceptible to chemical oxidation followed by biotinylation. This reaction permits the selection of capped RNAs by immunoprecipitation with streptavidin (Cap-trapping). The enriched capped RNAs then are converted into full-length cDNA copies.

These full-length cDNA can be sequenced (f1cDNA-seq) (Fig. 2.6a), which is typically based on large-scale Sanger sequencing (Kawai et al., 2001). This technique allows the identification of: first, the exact position of the mRNA transcription start site of individual genes; second, the sequence of the 5' untranslated region (UTR); third, the entire protein-coding region; and last, the sequence of the 3'UTR (Suzuki and Sugano, 2003).

The cap analysis of gene expression (CAGE) allows high-throughput identification of sequence tags corresponding to 5' ends of mRNA at the cap sites and the identification of the TSS (Shiraki et al., 2003). After the cap-trapping, the cDNA is treated with tobacco acid pyrophosphatase that removes the cap and leaves a 5'-monophosphate that is used to ligate a specific oligo linker. This is followed by the cleavage of the first 27 base pairs by a class II restriction enzyme (Fig. 2.6a) (Takahashi et al., 2012). Using this technique, FANTOM and ENCODE consortium produced high-quality annotations of the human and mouse promoter regions (Birney et al., 2007; Forrest et al., 2014).

These methods have the advantage of not only identify the TSS location but also estimate the abundance of produced RNAs. However, one of the limitations of CAGE methodology is the selective removal of all non-capped RNAs transcribed by RNAPI and RNAPIII (Takahashi et al., 2012).

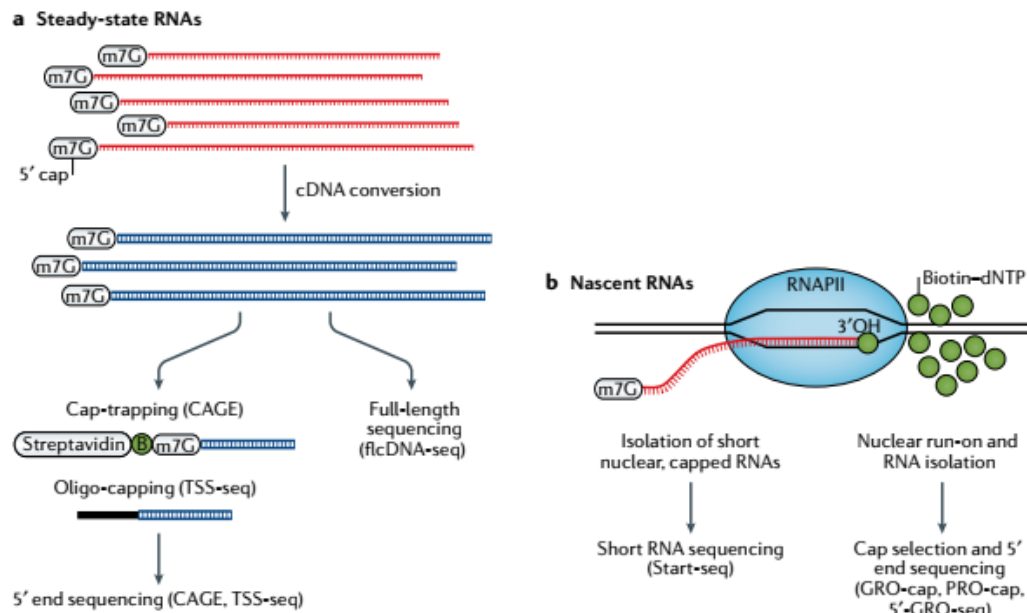


Figure 2.6 Schematic for (a) steady-state RNAs and (b) nascent RNAs identification. B, biotin labeled dNTP; m7G, 7-methylguanosine cap. Adapted from (Andersson and Sandelin, 2020)

Nascent-RNA analysis

In general, the nascent-RNA techniques aimed to capture the RNA as it is being transcribed, providing transcription information not only of multiple genes, but also long non-coding RNAs, microRNAs, and enhancer RNAs, without interference from splicing, capping, and post-transcriptional stabilization.

The method to assess real-time transcription from engaged RNA polymerase is the nuclear run-on assay. This method consists of isolation of the nuclei from mammalian cells and places them in ice to stop RNAPII from transcribing. Then, the nuclei were permeabilized and washed to remove free nucleotides. Traditionally, endogenous nucleotides were replaced by radio-labeled dNTPs and detected through hybridization to complementary DNA sequences (Smale, 2009). In global run-on sequencing (GRO-seq) methodology de incorporated dNTPs contain 5'-bromouridine-5'-triphosphate (BrUTP) to enriched the newly formed RNA (Core et al., 2008), while the precision nuclear run-on sequencing (PRO-seq) method adds biotin-labeled dNTPs (Kwak et al., 2013). To resume the transcription in vitro, anionic detergent sarkosyl is used to remove proteins from chromatin, including pausing factors (Gariglio et al., 1974).

These methods can be modified to enrich only the capped RNAs via cap-trapping and sequenced from the 5' end, just like the CAGE method: GRO-cap (Core et al., 2014) and PRO-cap (Mahat et al., 2016) (Fig. 2.6b).

The methods that sequence 5' ends can map the location of active TSSs genome-wide, while those that sequence the 3' ends only detect the active RNAPII sites within gene bodies. However, promoter regions can be inferred as peaks in the reads from these methods, due to the RNAPII pausing after transcription initiation.

Even though these methods were not designed to specifically detect active enhancers, it is possible to detect the small non-coding RNAs that are bidirectionally transcribed from active enhancers (miRNAs), that otherwise cannot be detected by traditional transcriptome profiling like RNA-seq (Lam et al., 2014).

Concluding Remarks

The annotation of the precise promoter is important to know the 5' UTR usage and the impact of genetic variation in non-coding regions. For example, a single variation between the alpha-globin genes and their regulatory region was shown to create a new TSS, affecting the normal activation of all downstream alpha-like globin genes, thereby causing thalassemia (De Gobbi et al., 2006). In many cases, the differential TSS usage is important for the gene function and diseases [reviewed in (Davuluri et al., 2008)]. Levin and cols. compared several methods to identify sample-specific TSS and report that the CAGE method performed better when combined with DNase-seq (Adiconis et al., 2018).

II. Methods for long-range chromatin interactions

The cell nucleus contains all the heritable information coded inside the DNA and the epigenetic components associated with it. But the reduced size of the nucleus would not be enough to contain the DNA fully stretched. Therefore, correct DNA packaging is essential for the correct cellular function.

However, each cell needs to have a dynamic equilibrium between DNA packaging and the access to the gene information, DNA replication sites accessibility, proof-reading repair, and recombination events; varying during different phases of the cell cycle and among different cell types (Sati and Cavalli, 2017).

Nowadays, is clear that the chromatin organization within the three-dimensional (3D) nuclear space constitutes another factor affecting gene regulation and changes in the conformation can cause oncogenic transformation (as discussed in Chapter 1).

The chromosome conformation capture (3C) was the first assay to study the chromatin organization in 3D space (Dekker et al., 2002), and since then it laid the foundations for developing new methods to infer the structures. Most of the 3C-derived techniques share the same initial steps: formaldehyde cross-linking, restriction digestion, and proximity ligation.

1. Chromatin conformation capture (3C)

First described in 2002, this assay estimates the contact between two genomic loci in the cell (one versus one). The first step of the method consists of the formaldehyde crosslink of the cell population nucleus, followed by digestion with a restriction enzyme (RE) and the ligation of the restriction fragments (Fig. 2.7). During the nuclei fixation, the amount of formaldehyde can vary

from different cell types; however, 1% formaldehyde for 10 minutes is preferred. For fine mapping of the genomic interactions, four base cutters REs are much frequently used (for example, HindIII and DpnII). On the other hand, the ligation step is advised to be performed under dilute conditions to facilitate the intermolecular ligation, instead of the intramolecular. The final product is a linear template representing the 3D interaction of two known loci. This template can be analysed by PCR to describe these interactions only in a limited range. Also, the “one versus one” nature of this method limits its application to genome-wide sequencing. To overcome these limitations, several 3C-derived methods have been developed.

2. One versus all, chromatin conformation capture (4C)

The next generation on the “C” technologies was the 4C. This methodology evaluates the genome-wide interactions made by a single locus (one versus all) and was developed independently by four groups. These 4C methods contrast in the significance of the acronym and a few steps; however, they are based on the same molecular principles (Lomvardas et al., 2006; Simonis et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006) (Fig 2.7).

The most popular variant of the 4C family is the 3C on-chip. This method uses another round of RE (usually 4-cutter RE, different from the first) on the 3C library, followed by ligation. This creates smaller circular fragments that can be amplified using PCR with specific primers for our locus of interest (called bait or viewpoint). The amplified fragments can be evaluated either using microarrays or high-throughput sequencing (4C-seq) (van de Werken et al., 2012).

Several studies have used the 4C-seq technology to identify regulatory interactions (van de Werken et al., 2012), chromatin loops (de Wit et al., 2015), TAD modifications in structural variants (Lupianez et al., 2015), and the identification of sub-structures inside TADs (Wijchers et al., 2016).

3. Chromosome conformation capture carbon copy (5C)

Both the 3C and the 4C method fail to see interactions outside the pre-selected fragment. Chromosome conformation capture carbon copy (5C) allowing the simultaneous detection of millions of interactions between multiple regions in a single assay (many versus many).

The initial steps of the 5C are the same in 3C. The 3C library is then incubated with a complex mix of oligos (containing universal primers) designed to anneal the restriction sites in a ligation-mediated amplification (Dostie et al., 2006). The product can then amplify by PCR using the universal sequence incorporated in the primers (Fig 2.7).

The 5C was developed to analyse the human β -globin locus (Dostie et al., 2006), and later to study the human α -globin locus (Brandt et al., 2010), the human HOXA-D gene cluster (Ferraiuolo et al., 2010; Fraser et al., 2009; Wang et al., 2011), and the TADs landscape in the X chromosome (Nora et al., 2012). This methodology is not truly genome-wide since each locus must be designed in the regions of interest.

4. Hi-C

The Hi-C assay confers the advantage of interrogating “all versus all” interactions and maps them genome-wide. In the first step, much like in the 3C method, nuclei are cross-linked with formaldehyde, and chromatin is digested. From here the Hi-C the procedure is slightly different since must use a RE that leaves a 5' overhang. These sticky ends are filled with biotin-labeled nucleotides followed by blunt-end ligation. After the Hi-C library is sheared and purified by pulling-down biotin, enriching the biotinylated ligation junctions for high-throughput sequencing (Lieberman-Aiden et al., 2009) (Fig 2.7).

The first Hi-C maps were low resolution (1 Mb) but could identify the compartmentalization of the human genome in open and active (A compartment), and closed and inactive (B compartment) compartments (Lieberman-Aiden et al., 2009). A modified protocol in *Drosophila* using the size selection of long products identified different chromatin organizations overlapping with epigenomic domains (Sexton et al., 2012). They classified the observed TADS into four types, one represented by active chromatin, and three represented by silent chromatin (polycomb enriched regions, heterochromatin, or without any specific mark).

5. Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET)

This method is the combination of 3C technology with ChIP. The key feature of this method is that the chromatin interactions are enriched by ChIP, using a specific antibody after RE digestion. Then tethered DNA fragments in each of the chromatin complexes are connected with DNA linkers via proximity ligation and Paired-End Tags (PETs) are extracted for sequencing. The resulting ChIA-PET sequences are mapped to reference genomes to reveal relationships between remote chromosomal regions brought together in close spatial proximity by protein factors (Fullwood et al., 2009).

Using this method, it was possible to identify the importance of CTCF, ZNF143, and two members of the cohesion complex (SMC3 and RAD21) for the formation of the 3D chromatin structure (Heidari et al., 2014).

A new method called HiChIP combines the antibody capture of the ChIP, the biotin enrichment of the Hi-C, and Tn5-mediated on-bead library construction (Mumbach et al., 2016). This method improves the yield of conformation-informative reads by over 10-fold and lowers input requirement over 100-fold relative to ChIA-PET.

6. Capture-C

This method combines 3C with oligonucleotide capture and NGS. Once the 3C library is generated, is sonicated (200 bp size) and sequencing adaptors are added. This library is enriched by hybridization to biotinylated capture probes and purify by streptavidin pull-down. Finally, the enriched material is amplified by PCR and sequenced (Fig 2.7). This strategy generates genome-wide contact of the selected loci (Hughes et al., 2014).

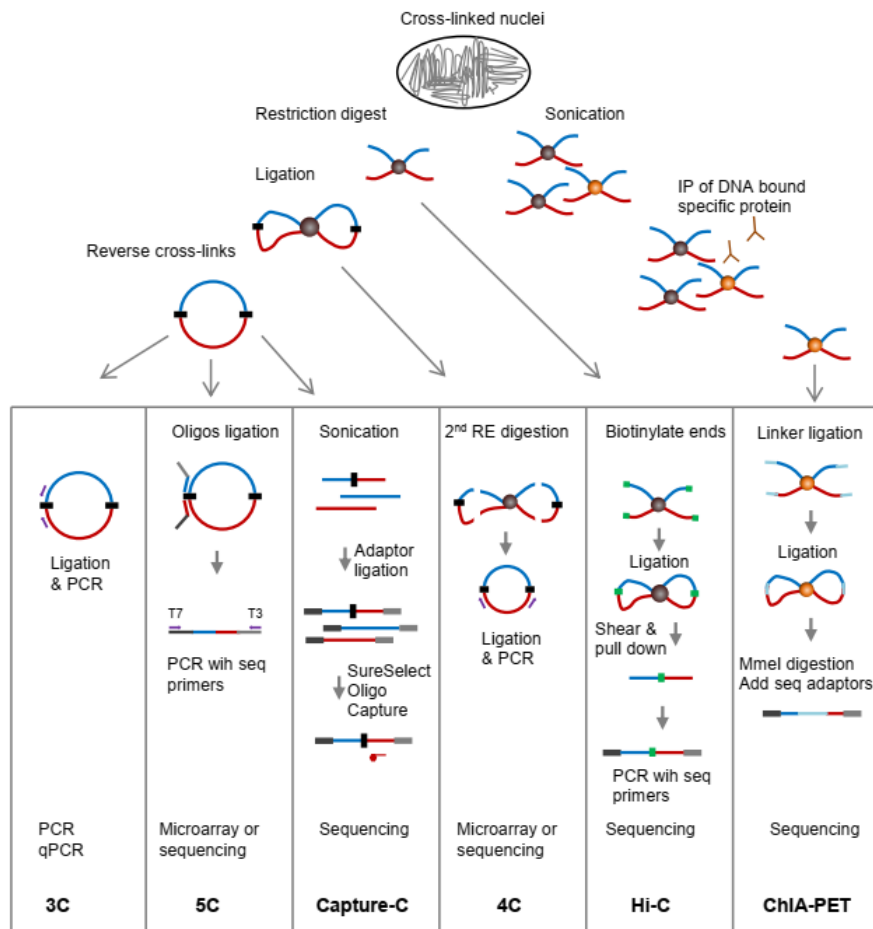


Figure 2.7 Method summaries for 3C and 3C-based technologies

7. Concluding Remarks

The 3C and the 3C-based technologies have helped to uncover the intrinsic mechanisms of genome architecture at the genome-scale. Researchers have been working to improve the 3C technology and overcoming the technical limitations by creating new tools or using the tools already existing.

The first limitation in most of the 3C-based methods is the selection of the first RE. This step determines the resolution of the 3C library since only the contacts between DNA fragments can be detected. Another limiting factor is the fact that the RE cutting sites are not evenly distributed in the genome, resulting in different resolutions across the genome. This seems to be overcome by the substitution of the first RE with MNase (Hsieh et al., 2015) or DNase (Ma et al., 2015).

Integrative data analysis using data set coming from different 3C-based techniques and chip-seq data have led to the composition organization of super-enhancers (Huang et al., 2018; Ing-Simmons et al., 2015), dynamics of promoter-anchored loops during differentiation (Siersbaek et al., 2017), and disease prediction (Wang et al., 2018a).

One of its main limitations of the Hi-C is the low resolution, therefore is not well suited for probing short-distance interactions such as intra-TAD interactions. This could be fixed by changing the RE to a higher digestion frequency, but this will increase the cost in the sequencing. To circumvent this, researchers use the Capture-C strategy to enrich Hi-C libraries, a new technique known as Capture Hi-C (CHi-C) (Mifsud et al., 2015). This method enables deep sequencing of target fragments and reduces the background. With this method, several groups have been able to interrogate enhancer-promoter interactions in cancer risk loci (Baxter et al., 2018; Dryden et al., 2014; Jager et al., 2015; Martin et al., 2015), to detect the promoter-interacting regions with promoter-capture Hi-C (PCHi-C) (Orlando et al., 2018; Schoenfelder et al., 2018; Siersbaek et al., 2017), and to characterize the enhancer interactome during cell differentiation with enhancer-capture Hi-C (EChi-C) (Madsen et al., 2020).

The continuous improvements of the C-technologies and the high-throughput sequencing will be led to a better understanding of how the 3D structure of the genome can control gene expression.

III. Functional characterization of transcriptional regulatory elements

Several points concerning the massively parallel reporter assays are approached in our published review (Santiago-Algarra et al., 2017). A shorter version of this review was also published in a French journal (España et al., 2017).

The control of the gene expression requires multiple levels of control, from the chromatin accessibility, histone modification, and DNA methylation. All these elements lead to the proper communication of two of the major DNA *cis*-regulatory elements: promoters and enhancers. The identification of enhancers is challenging since this DNA does not code for a protein (like 98% of the genome), its location relative to the gene(s) they control can vary (upstream, downstream or intergenic) and their sequence can be different among other enhancers. Using the tools previously described, like open chromatin regions (DNase-seq, ATAC-seq, MNase, etc.) and histone modifications (H3K4me1/H3K27ac ratio), is possible to identify putative enhancer (Chatterjee and Ahituv, 2017). However, these techniques do not provide a functional readout for the sequences identified in different cell types and with different nucleotide variants.

The standard enhancer and promoter-reporter assays are functional and sensitive methods to test the activity of one candidate region to express a reporter gene (LacZ, fluorescent protein, and luciferase). The candidate region is cloned in an episomal plasmid upstream of the reporter gene (for a promoter assay) or downstream a minimal promoter and the reporter gene (for an enhancer assay). Then, the construct is introduced in the cells (primary cells, cell lines, embryos) of interest. If under the controlled conditions the candidate is a promoter or an enhancer, the expression of the reporter will happen. The expression of the reporter gene can be observed and quantified; however, this method only allows the testing of a single candidate per construct and is used as confirmation of the high-throughput assays.

There exist two major high-throughput assays to test the enhancer activity of thousands of putative regulatory sequences in a single experiment. These are the massively parallel reporter assays (MPRA) and the self-transcribing active regulatory region sequencing (STARR-seq).

1. MPRA

The MPRA method was originally developed to test the promoter activity and effect of single-based mutagenesis in three bacteriophage promoters *in vitro* (Patwardhan et al., 2009), and test the enhancer activity of three mammalian enhancers and its variants *in vivo* (Patwardhan et al., 2012).

This assay uses a microarray-synthesized DNA construct that contains the regulatory element and a unique sequence tag or “barcode” (Fig. 2.8A). These DNA-synthesized sequences are cloned onto the plasmid backbone. Then, a minimal promoter and a reporter gene are inserted between the tested regions and the tag, leaving the tag in the 3' UTR. The enhancer activity of any sequence will transcribe the tag associated with it.

Once the MPRA library is finished, is transfected into cells and RNA sequencing of the transcripts/tags is performed. The enhancer activity of any sequence is directly proportional to the tag enrichment.

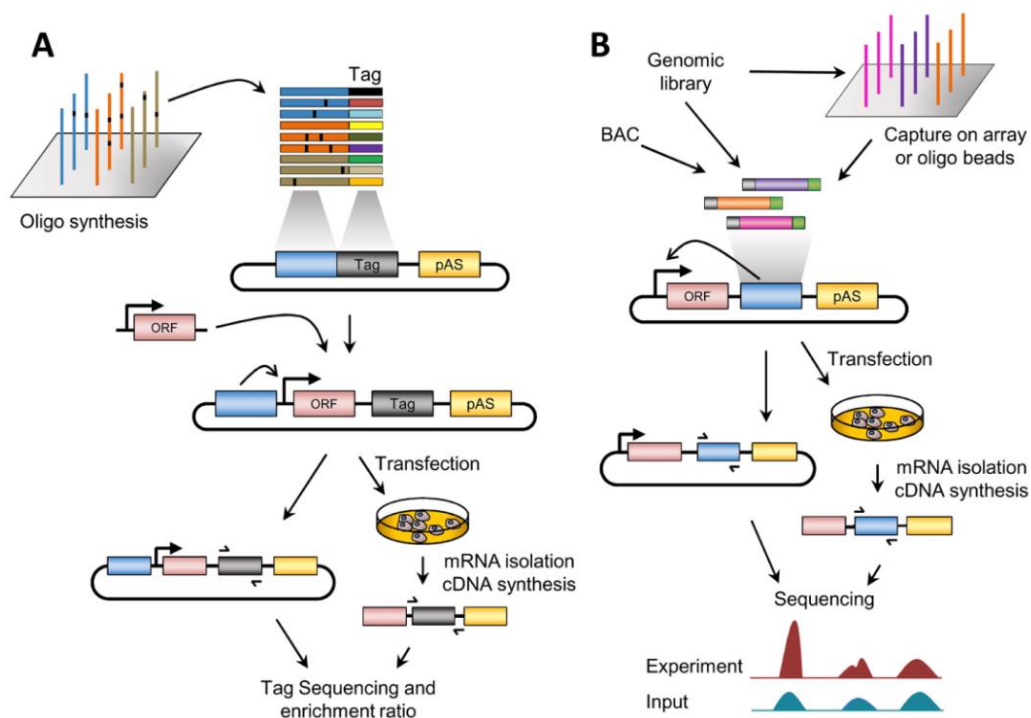


Figure 2.8 Methods for functional characterization of gene regulatory elements. (A) massively parallel reporter assays, MPRA. (B) self-transcribing active regulatory region sequencing, STARR-seq.

2. STARR-seq

An alternative method to MPRA is the STARR-seq, a high-throughput assay introduced by Stark and cols. to identify transcriptional enhancers based on their activity in the entire *Drosophila* genome and to assess their activity quantitatively (Arnold et al., 2013).

This method does not require synthesized “barcodes” since the DNA sequences are cloned into the 3’ UTR of the reporter gene. The active enhancer will transcribe the reporter gene and themselves, becoming part of the reporter transcript. Thus, the enhancer activity of any enhancer constitutes an enrichment of the “barcode” of itself. These transcripts can be isolated and detected by high-throughput sequencing (Fig. 2.8B).

Stark’s lab used the STARR-seq approaches to ask several basic mechanistic questions of enhancer biology in *Drosophila*. This method can be applied to mammals (Muerdter et al., 2018; Peng et al., 2020); however, the size of the genome can increase the cost in experimental and sequencing deep.

3. Derivative approaches.

Reduce complexity

To reduce the complexity and the prohibited cost of doing whole-genome STARR-seq in mammals, several groups developed strategies to enrich the genomic regions of interest (Fig. 2.10). In general, these methods overcome the synthesis limitations in MPRA methods and reduce the complexity of the genomic library.

CapSTARR-seq

Spicuglia, and cols. developed a capture-based approach (CapSTARR-seq) to assess a subset of mouse DNase I hypersensitive sites (DHS) (Vanhille et al., 2015a) and the human protein-coding promoters repertoire (Dao et al., 2017). After capture enrichment, the DNA fragments are cloned into the backbone of the STARR-seq (3’ UTR of the reporter gene) and transfected into de cells. This method provides a fast and cost-effective approach to assess the activity of potential enhancers for a given cell type and will help decrypt transcription regulation mechanisms (Vanhille et al., 2015a)

ATAC-STARR-seq

Kellis and Claussnitzer developed a strategy called HiDRA (High-resolution Dissection of Regulatory Activity) (Wang et al., 2018b). This method combines ATAC-seq with STARR-seq and offers the advantage to test the activity of fragments derived from open chromatin. With this method, they identified 65,000 regions showing enhancer function in GM12878 lymphoblastoid cells. Moreover, they were able to identify the driver elements in those enhancers (Wang et al., 2018b).

FAIRE-STARR-seq

Singh and cols. coupled FAIRE-seq with STARR-seq to assess a complex set of accessible chromatin regions for enhancer activity in LPS-activated B cells (Chaudhri et al., 2020). This

technique allows them to identify the numerous STARR-positive sequences located in the proximal promoter, intergenic and intragenic regions.

BAC

Another strategy to eliminate arbitrary sources of DNA is the use of bacterial artificial chromosomes (BACs). Within these BACs, DNA targeted regions can be cloned in the STARR-seq vector and tested for the enhancer function. This technique helped to identify enhancer specificity to a core promoter (either housekeeping promoter or developmental promoter) and reported that many core promoter regions have enhancer activity with the housekeeping promoter (Zabidi et al., 2015). Another study used BACs to interrogate GWAS-associated loci to atrial fibrillation in humans and identify 55 regulatory elements with variants associated with atrial fibrillation (van Ouwerkerk et al., 2020).

ChIP-STARR-seq

Enhancer function requires transcription factor binding and open chromatin marked by histone modifications. Therefore, the use of ChIP-STARR-seq, strategy incorporates the chromatin immunoprecipitation with STARR-seq, could potentially identify functional enhancers. Using this method, Chambers and cols. identify that only a minority of regions marked by NANOG, OCT4, H3K27ac, and H3K4me1 function as an enhancer in embryonic stem cells, indicating that no individual TF, histone mark, or a combination thereof could unequivocally predict enhancer activity (Barakat et al., 2018)

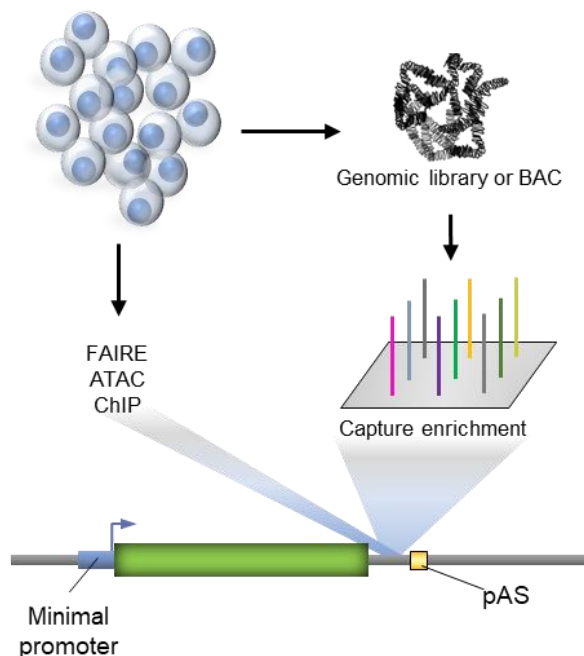


Figure 2.10 Methods to reduce the library complexity for STARR-seq assays

Assessing genetic variation

The STARR-seq and the MPRA can be used to study the effect in promoter or enhancer function of sequence variations from selected populations or between DNA of healthy versus disease tissue. Nowadays, with genome-wide association studies (GWAS) we can identify genetic

variants associated with a disease. Moreover, we can associate those genetic variants with gene expression variability.

Massive oligo synthesis

Custom made oligonucleotides can be synthesized on microarray-based synthesizers that can produce oligonucleotides up to 200 bp in length. The sequence synthesized to contain the regulatory regions and the SNP of interest within the first 150 bp, a 20 bp barcode (for MPRA assays), and universal primers to amplify the material. This method allows us to interrogate the effect of thousands of variants in the transcriptional regulation. In particular, this method has been used to test the effect of 81 expression quantitative trait loci (eQTL) SNP associated with platelet *CD36* expression (Madan et al., 2019), the functional SNPs at prostate cancer-risk loci (Zhang et al., 2018), atrial fibrillation (van Ouwerkerk et al., 2020), and variants with allele-specific expression (Kalita et al., 2018).

Pool DNA

It is also possible to test the naturally occurring population genetic heterogeneity and test the regulatory variants by pooling the genomic DNA of individuals. Using capture arrays is possible to enrich the cancer risk-associated SNPs of interest. This variant library can then be cloned in the STARR-seq backbone and test their regulatory activity. With this strategy Wang, White, and cols. were able to find two regulatory variants associated with breast cancer and leukaemia (Liu et al., 2017a).

Pool genome and SuRE

Survey of Regulatory Elements (SuRE) is a method that allows us to test more than 10^8 barcoded DNA genomic fragments. Instead of short synthetic promoter sequences, SuRE queries random genomic fragments in the size range of 0.2–2kb, which is long enough to include most elements that constitute fully functional promoters (van Arensbergen et al., 2017). SuRE also allows us to measure the promoter and the enhancer activity of these elements. With this method, van Steensel and cols., were able to identify 30,000 SNP that alter the activity of putative regulatory elements from four lymphoblast cell lines (van Arensbergen et al., 2019).

Integrative

One of the major concerns about reporter assays is that they are mainly implemented on episomes, which are thought to lack the natural physiological chromatin effect. One of the most popular tools to circumvent this issue is the use of lentiviruses or adenovirus to integrate the reporter assay into the chromatin.

Lenti-based assays

The advantages of using a lentivirus to integrate the reporter assay in the genome are (i) the stable integration into the host genome, (ii) can be used in cultured primary cells and non-dividing cells, and (iii) can deliver up to 8 Kb with a relatively low immune response (Zheng et al., 2018). The use of a lentivirus-based massively parallel reporter assay (lenti-MPRA), designed

by Schendure and cols. (Inoue et al., 2017), have help to identify the transcription factors and regulatory sequences involved in neural differentiation (Inoue et al., 2019).

Locus integration

One of the major disadvantages of the lentivirus integration is that it occurs in random locations along the genome and thus, susceptible to the effects of local chromatin environment and interaction with neighbouring enhancers. Segal and cols. design a new high-throughput method for accurately measuring designed sequences from a fixed and predefined locus in the human genome. This locus-specific integration in the human genome is achieved using site-specific integration into the “safe harbour” *AAVS1* site (Weingarten-Gabbay et al., 2019).

Silencers

So far, the high-throughput methods have been used to assess the functional activity of promoters and enhancers, however, it is possible to redesign these assays to identify silencer elements (Doni Jayavelu et al., 2020). Using a sense tiling MPRA constructs, Kellis and cols. identified sequences with repressive motifs at 5 bp resolution (Ernst et al., 2016).

DNA methylation

Epigenome-wide association studies (EWAS) have identified thousands of correlations between DNA methylation levels at individual CpG sites and age, cancer, diabetes, and Alzheimer’s disease. Differential DNA methylation is mechanistically important in trait variation, it should also have downstream consequences for gene regulation. However, while a functional relationship between differential methylation and gene expression levels is often assumed, experimental studies have shown that it does not always hold (Lea et al., 2018).

Methyl-STARR-seq

This novel strategy can test the relationship between DNA methylation and the regulatory activity within a cellular environment. To accomplish this, first, they eliminated all the CpG dinucleotide in the backbone of the STARR-seq plasmid, then combines high-throughput cloning of hundreds of thousands of query fragments that contain millions of CpG sites. These constructs can then go under enzymatic manipulation of DNA methylation using the CpG methyltransferase M.Ssl. Unmethylated and methylated conditions are then tested. This strategy can interrogate which DNA methylation is important for gene regulation (Lea et al., 2018).

4. Concluding Remarks

These techniques allow the characterization of cis-regulatory sequences on a large scale and different biological contexts. MPRA and STARR-seq present different advantages over one another. For example, MPRA has been used to test the impact of thousands of single nucleotide polymorphisms (SNPs) in the regulatory elements, whereas STARR-seq can be used to test natural variants coming from PCR amplifications.

Although the basic principle of MPRA and STARR-seq is as an episomal reporter assay, MPRA can be modified to be cloned in a lentiviral vector (Lenti-MPRA) to study the function of genome-integrated regulatory elements (Maricque et al., 2017).

In particular, these methodologies shed light on the study of the Epromoters, an annotated promoter that exhibit promoter and enhancer activity in vitro and in vivo (Dao et al., 2017; Nguyen et al., 2016b; Vanhille et al., 2015a; Zabidi et al., 2015).

5. Technical considerations

One of the major problems with STARR-seq was reported by the group of Stark, affirming that while in *Drosophila* the transcription starts in the core promoter of the vector, in mammals the transcription started predominantly in the bacterial origin of replication (ORI) (Muerdter et al., 2018). Preferential transcription in one of the two core promoters (ORI and SCP1) could create a bias in the enhancer-activity measurements, suggesting a new setup using the ORI as a core promoter and removing the SCP1 promoter (“promoterless” STARR-seq). Moreover, they also reported that during the DNA transfection the cell can also trigger an innate antiviral response,

Table 2.1 Recent examples of high-throughput functional assays of mammalian enhancers. This table complement and update table 1 in annex 2

Technique	DNA origin	Application	Specificity	No. of regions ¹	Size (bp)	Cell lines	Promoter	Species	Ref.
Lenti-MPRA	Synthetic	Identification of regulatory dynamics during neural differentiation	Centered on ChIP	2,463	171	hPSCs differentiation	Minimal promoter	Human	(Inoue et al., 2019)
STARR-seq	Genomic	Use of STARR plasmid in mammalian cells	Whole-genome	71,968	1000-1500	HeLa-S3	Minimal promoter and ORI	Human	(Muerdter et al., 2018)
MPRA	Synthetic	Analysis of promoter variants in lncRNAs that affect the expression	-80 to +34 bp from TSS of lncRNAs	2,078	114	HeLa, HepG2 and K562	Minimal promoter	Human	(Mattioli et al., 2019)
MPRA	Synthetic	Test of 81 eQTL SNPs associated with CD36	Centered on the SNP	1	150	K562	Minimal promoter	Human	(Madan et al., 2019)
STARR-seq	Genomic	Analysis of variants in cancer risk loci.	Genomic capture centered on the SNP	996	500	HEK293T	SCP1 ²	Human	(Liu et al., 2017a)
ChIP-STARR-seq	Genomic	Identification of enhancers in human embryonic cells	Genomic capture using ChIP	361,737	600	hESCs	Minimal promoter	Human	(Barakat et al., 2018)
STARR-seq	Synthetic	Characterization of transcriptional regulation of cancer risk loci.	Centered on the SNP	374	21	LNCaP	SCP1 ²	Human	(Zhang et al., 2018)
ATAC-STARR-seq	Genomic	Genome-wide testing of putative regulatory regions	ATAC accessible regions	7x10 ⁶	150-500	GM12878	SCP1 ²	Human	(Wang et al., 2018b)
BiT-STARR-seq	Synthetic	Characterize variants with allele-specific effects in regulatory regions	Centered on the SNP	43,500	200	GM18507	Minimal promoter	Human	(Kalita et al., 2018)
STARR-seq	Genomic	Genome-wide testing of putative regulatory regions	FAIRE accessible regions	55,133	100-400	Splenic B cells	Minimal promoter	Mouse	(Chaudhri et al., 2020)
STARR-seq	Genomic	Genome-wide characterization of enhancers in two pluripotent states	Genomic DNA sonicated	48,311	700-1200	mESCs	SCP1 ²	Mouse	(Peng et al., 2020)
Methyl STARR-seq	Genomic	Assess the causal effect of mDNA on the regulatory activity of millions of CpG sites.	<i>MspI</i> -digested DNA	262,829	300-700	K562	CpG-free EF1	Human	(Lea et al., 2018)
STARR-seq	Genomic	Characterization of variants associated with atrial fibrillation	GWAS loci cloned in BACs	12	450-900	Myocytes	SCP1 ²	Rat	(van Ouwkerk et al., 2020)

¹ Number of targeted DNA sequences; not necessarily the number of unique fragments that are tested.

² Synthetic Super Core Promoter-1

inducing the production of type-I interferon and altering the gene expression and the enhancer activity. No reports of the induction of type-I interferon in MPRA-transfected cells have been found. Also, the same group suggests the use of unique molecular identifiers to count individual reporter transcript molecules, enabling their precise quantification (Neumayr et al., 2019).

In MPRA the size of the target sequences is limited due to synthesis limitations (approx. 170 bp). Also, the sequences that contain the restriction site used for cloning cannot be tested by MPRA. The advantage of the STARR-seq is the elimination of the tags since the target sequences can come from the fragmented genome, DNase I accessible sites, ATAC accessible regions, FAIRE accessible regions, ChIP, or specific genomic enrichment using a capture microarray of PCR amplification. Table 1 included in annex 2 contains the high-throughput assays for enhancer activity until 2017 when it was published. Whereas in table 2.1, I summarize the latest reports.

Due to the amount of data generated by MPRA and STARR-seq, several groups try to integrate the data with multiple epigenomic marks (coming from ChIP-seq, DHS, ATAC-seq, etc.) to predict the regulatory function (Kreimer et al., 2019; Liu et al., 2017c; Shigaki et al., 2019).

Recently, the groups of Shendure and Ahituv made a systematic evaluation of several strategies using the STARR-seq and MPRA (episomal and integrative) (Klein et al., 2020). Their results show high inter-assay correlations between the classical reporter assay, the promoterless STARR-seq, and both episomal and integrated MPRA. They also suggest that enhancer assays should test larger fragments since they found that larger sizes (600 bp) add more biologically relevant signals than shorter fragments (190 bp). Finally, they confirmed that enhancer activity is largely, but not completely, independent of orientation.

IV. CRISPR-based techniques

1. Discovery

Clustered regularly interspaced short palindromic repeat (CRISPR) is part of the adaptive prokaryotic immune system and was first detected in *Escherichia coli* in 1987 (Ishino et al., 1987) and later in 1993 in the archaea *Haloferax mediterranei* (Mojica et al., 1993). This cluster consisted of a series of repeat palindromic units separated by spacer virus-derived sequences (Mojica et al., 2005).

The identification of the CRISPR regions led to the discovery of four conserved genes present next to the CRISPR regions (CRISPR-associated system, Cas) (Jansen et al., 2002). In 2006, Makarova et al. suggested that the CRISPR-Cas system could be exploited to silence genes in organisms encoding Cas proteins (Makarova et al., 2006). The function of the CRISPR-Cas system as a prokaryotic immune system was proven in *Streptococcus thermophilus* (Barrangou et al., 2007) and could be transferred to *E. Coli*, providing heterologous protection (Sapranaukas et al., 2011). Soon after, it was proven that the purified Cas9-CRISPR RNA (crRNA) complex is capable of cleavage target DNA in vitro (Jinek et al., 2012).

The CRISPR system was further classified into six types. The first three are the most studied, each distinguishable by the unique signature protein: Cas3 for type I, Cas9 for type II, and Cas10

for type III. Up to date, the most used and studied is Cas9, a multidomain effector protein that can bind and cleave the target DNA.

The Cas9 system forms a base-paired structure between a trans-activating crRNA (tracrRNA) and the targeting crRNA to cleave target dsDNA. Site-specific cleavage occurs at locations determined by both base-pairing complementarity between crRNA and the target dsDNA and a short motif, referred to as the protospacer adjacent motif (PAM), juxtaposed to the complementary region in the target DNA (Jinek et al., 2012). For Cas9, the PAM recognition sequence is 5'-NGG-3', while other Cas systems differ in the PAM recognition and the cleavage pattern.

Nowadays, a simpler engineered CRISPR-Cas9 system consists of a single guide RNA (sgRNA) and the SpCas9 nuclease. The sgRNA is a chimeric non-coding RNA formed by the crRNA and the tracrRNA. In this complex, the tracrRNA will work as a nuclease-recruiting sequence, while the crRNA will guide the complex to the targeting sequence. This artificial linkage does not affect function.

Experimentally, there exist different strategies for genome editing using Cas9: (i) plasmid encoding the Cas9 protein and the sgRNA (episomal or integrative); (ii) direct delivery of the Cas9 mRNA and the sgRNA; (iii) direct delivery of the Cas9 protein and the sgRNA combination. Depending on the strategy to use, the delivery systems might change [reviewed in (Karimian et al., 2019)]. Once the Cas9-sgRNA complex recognizes the target DNA, the nuclease activity of the Cas9 can perform a cut in the DNA, triggering the cellular DNA repair pathways. This constitutes the principle of genome editing via CRISPR-Cas9.

2. CRISPR-Cas9 applications

CRISPR-mediated genome editing

One of the classical applications for CRISPR-Cas9 technology is the gene knock-out. The result of the Cas9-sgRNA is a double-strand break in the target DNA typically repaired by the non-homologous end-joining (NHEJ) pathway, resulting in a “scar” in the DNA sequence (small deletions, insertions, or indels); or by homology-directed repair (HDR) pathway, enabling precise base modifications using a DNA template (dsDNA or ssDNA) (Fig. 2.11a). Mutations in one of the nuclease activity domains (HNH or RuvC1) of the Cas9 has shown to produce single-strand DNA breaks (nicks), instead of the original double-strand blunt break (DSB) (Ran et al., 2013a). The main difference between the nicks and the DSB is that nicks are repaired by the high-fidelity base excision repair pathway (Dianov and Hubscher, 2013).

Genomic deletions can be obtained by CRISPR-Cas9 dual-targeting using two sgRNA. The NHEJ will repair the two double-strand breaks by deletion of the intervening segment. This technique has been used to create deletion ranging from one kilobase (Horii et al., 2013) to over 30 megabases (Essletzbichler et al., 2014).

Currently is possible to interrogate more than one region at a time and apply the CRISPR-Cas9 methodology to a genome-wide scale. This method, called CRISPR screening, requires the use of

lentiviral libraries that could include the sgRNAs and Cas9 in the same vector or as a two-vector system (sgRNAs and Cas9 on separated plasmids) (Sanjana et al., 2014).

One of the biggest worries about the CRISPR-Cas9 technology is the possibility to have genomic modifications at unintended sites, also called off-target modifications (Wu et al., 2014). However, the use of a Cas9 nickase can help minimize the off-target effect (Ran et al., 2013b).

CRISPR interference

The main function of the Cas9 is endonuclease; however, mutations in the Cas9 catalytic domains, HNH and RuvC1, generate a catalytically dead Cas9 (dCas9). The approach is known as CRISPR interference (CRISPRi) inhibits the gene expression of any gene by targeting the dCas9 to a region near the TSS using a sgRNA (Fig. 2.11b). This creates a blockage for the initiation of the transcription or the elongation with up to 1,000-fold repression (Qi et al., 2013).

In addition to the basal Cas9 interference, it is possible to fuse the dCas9 with a repressive chromatin modifier domain to further reduce the expression by increasing the chromatin compaction, therefore, silencing the gene expression. The most used repressor is the Kruppel associated box (KRAB) that mainly induces deacetylation and methylation of histones (Gilbert et al., 2013) and a recent report suggests that the ZIM3 KRAB domain is the most efficient to induce gene silencing (Alerasool et al., 2020).

This tool has helped to identify 500 functional long non-coding RNAs in human cells important for robust cellular growth (Liu et al., 2017b). Moreover, the use of dCas9-KRAB as a specific epigenomic tool can help to reveal the connections between regulatory elements and gene expression genome-wide, with possible applications to gene therapy and cellular reprogramming.

CRISPR activation

Contrary to CRISPRi, fusing different transcriptional activators to dCas9 upregulates gene expression at targeted loci in a technique termed CRISPR activation (CRISPRa) (Fig. 2.11c). Early methods used the dCas9 fused with the transcriptional activator VP64 (four tandem copies of VP16) with modest activation of the targeted gene (Gilbert et al., 2013). Activation can be increased by using several sgRNAs to the same locus; however, other approaches are the Cas9 fusion with VPR (a tripartite activator comprised of VP64-p65-Rta) with a single sgRNA (Chavez et al., 2015) or the Cas9 fusion with the core p300 domain (Hilton et al., 2015).

In combination, CRISPRi and CRISPRa can control the transcriptional level of endogenous genes in human cells and provide complementary information for mapping complex pathways (Gilbert et al., 2014).

3. Concluding Remarks

CRISPR-Cas9 is a powerful tool that can help edit genomes and have revolutionized the life sciences. It is a simple and low-cost method for genome editing and provides more advantages than the previous genomic editing tools (zinc finger nuclease, ZNF; and transcription activator-like effector nucleases). The amount of literature around this technology is vast and diverse,

with over 3,000 papers released annually for the past three years (2017-2020) in different cellular contexts.

The use of CRISPR screening has helped to identify gene regulatory networks in their natural context (Diao et al., 2017b; Gilbert et al., 2014; Joung et al., 2017; McCarty et al., 2020; Parsi et al., 2017), in stem cells reprogramming (Yang et al., 2019) and to detect drug resistance (Ding et al., 2018; Kurata et al., 2018; Liu et al., 2019; Saber et al., 2020). In particular, Diao et al. identify 45 cis-regulatory elements of the POU5F1 gene, including 17 annotated promoters, supporting the existence of the Epromoters (Diao et al., 2017b).

Moreover, the CRISPR-Cas9 system has immense therapeutic potential for treating different diseases with a genetic background or to study the disease in animal or cell-based models. So far, it has been used to inactivate viral DNA that causes the disease (Lin et al., 2014; Sakuma et al., 2016; Zhen et al., 2014) or mutate the receptor avoiding the virus binding to the cell (Kang et al., 2015; Wang et al., 2014). Moreover, the precise genome editing can be used to reproduce disease mutations in animal models (SNPs, chromosomal rearrangements, etc), like kidney (Miyoshi et al., 2020), brain (Fernando et al., 2020), retina (Fuster-Garcia et al., 2020), and cardiac diseases (Vermersch et al., 2020).

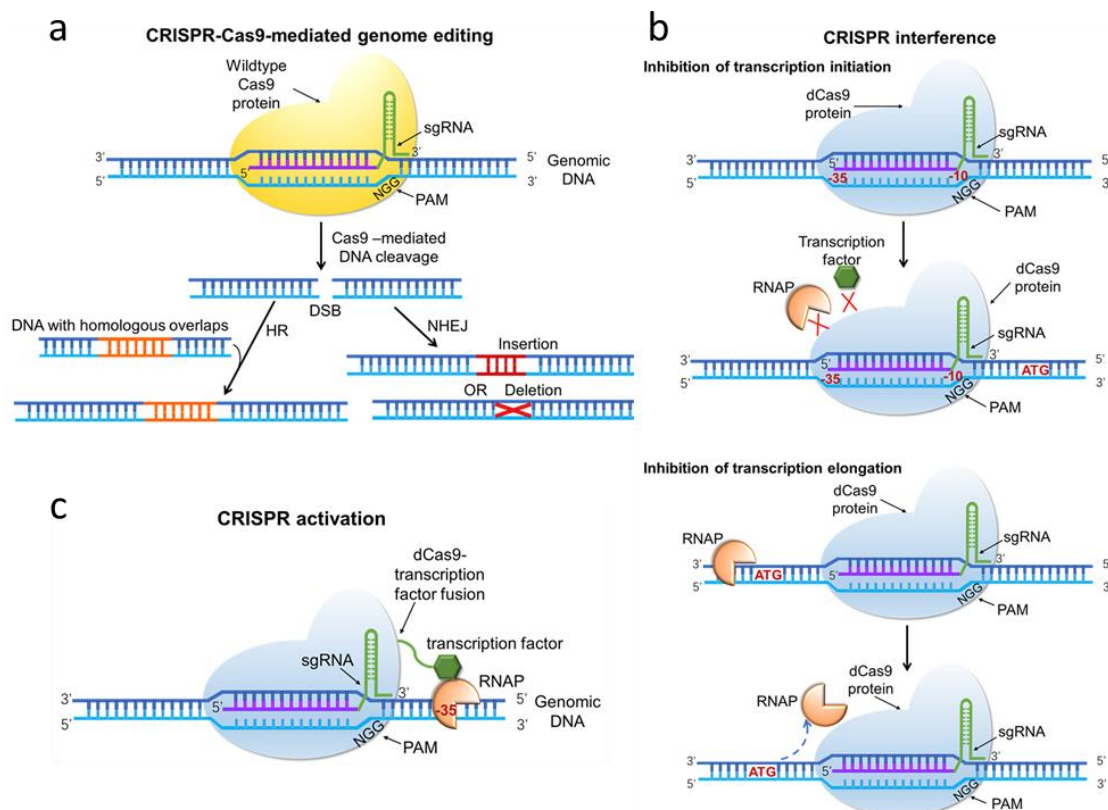


Figure 2.11 CRISPR tools for (a) genome editing, (b) CRISPR interference and (c) CRISPR activation. Adapted from (Tian et al., 2017).

The CRISPR-Cas9 can also impact food production, transforming traditional agriculture by making plants tolerant to biotic and abiotic stresses and improving their nutritional value and yield (El-Mounadi et al., 2020; Manghwar et al., 2019; Tyagi et al., 2021).

The rapid development and use of CRISPR-Cas9 technology have led to new bioethical, social, and legal issues in medicine, agriculture, livestock, and the environment. This has been ignited by the lack of regulation in the do-it-yourself (DIY) independent “labs”, and the 2018 scandal of the Chinese CRISPR babies lead by He Jiankui.

Chapter 3. Interferon inflammatory response

The innate immune response is involved in various inflammatory processes, in bacterial and viral infections. During a viral infection, the cells induced the production of a protein that interferes with viral replication and therefore, blocking the spread of viruses to uninfected cells. These antiviral proteins were called interferons (IFNs) and are imperative to regulate the activation and functions of various immune cell populations of the innate and adaptive immune systems.

I. Types of interferon

Type I interferon (IFNs) was first described as the responsible factor for viral interference, the ability of infection with a virus to induce resistance to infection with a different virus (Isaacs and Lindenmann, 1957). There are three types of interferons (IFNs), types I, II, and III. Each has numerous roles in innate and adaptive immunity. The most abundant and well-studied type I IFNs are IFN α and IFN β , while other types I IFN are less well characterized. IFN α gene is composed of 13 subfamilies in humans, while IFN β exists in a single copy. All the human type I IFNs are clustered in chromosome 9 and are expressed and signal in every cell. Type II IFN is composed only of a single copy of IFN γ and is more restrictively expressed by immune cells and is structurally unrelated to the other types of IFNs. The IFN γ is secreted predominantly by activated cells of the immune system, such as T cells and natural killer (Corthay et al., 2005; Flaishon et al., 2000; Matsushita et al., 2015). Under physiological conditions, type I and II are expressed basally and contribute to tissue homeostasis without systemic effects (Chen et al., 2009; Tovey et al., 1987). The type III IFN includes IFN λ 1 (IL-29), IFN λ 2 (IL-28A), IFN λ 3 (IL-28B), and IFN λ 4. These genes have an important role in protecting the epithelial surfaces from viral infection (Zhou et al., 2018a).

II. Induction of IFN expression

The IFNs are polypeptides that are secreted by infected cells and have three major functions. First, they induce cell-intrinsic antimicrobial states in infected and neighboring cells that limit the spread of infectious agents, particularly viral pathogens. Second, they modulate innate immune responses in a balanced manner that promotes antigen presentation and natural killer cell functions while restraining pro-inflammatory pathways and cytokine production. Third, they activate the adaptive immune system, thus promoting the development of high-affinity antigen-specific T and B cell responses and immunological memory (Ivashkiv and Donlin, 2014).

Type I IFNs can be induced by almost any cell type in the body in response to microbial pathogen-associated (PAMPs) and damage-associated molecular patterns (DAMPs). These processes not only trigger immediate host defensive responses such as inflammation but also prime and orchestrate antigen-specific adaptive immune responses. The key inducers of type I IFN are the recognition of viral nucleic acid, predominantly by intercellular pattern recognition receptors (PRRs), that leads to the production of pro-inflammatory cytokines, including large amounts of type I IFNs (fig. 3.1) (Barrat et al., 2016). Mammals have several distinct classes of PRRs including Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), Nod-like receptors (NLRs), AIM2-like receptors (ALRs), C-type lectin receptors (CLRs), and intracellular DNA

sensors such as cGAS (Kawasaki and Kawai, 2014). The receptors responsible for triggering the type I IFN responses are the cytosolic nucleic acid sensors cGAS, RIG-I, and MDA5 (melanoma differentiation-associated gene 5) and four predominantly endosomal TLRs: TLR3 (double-stranded RNA), TLR7 and TLR8 (single-stranded RNA) and TLR9 (double-stranded DNA) (Barrat et al., 2016).

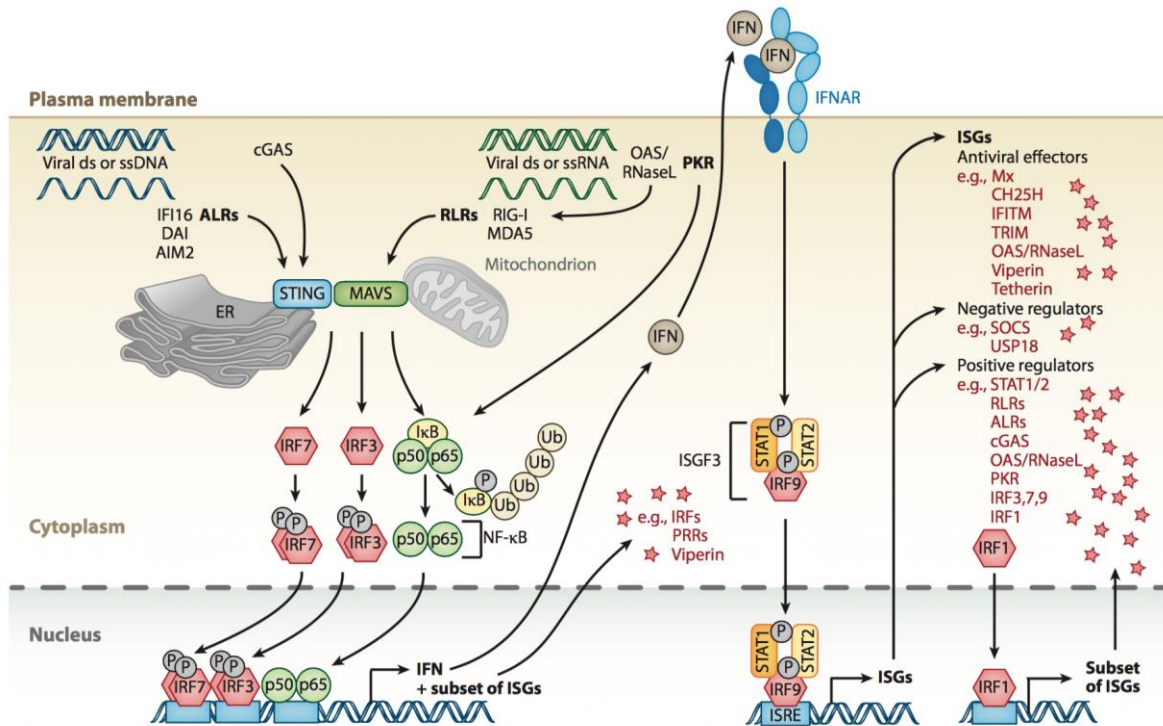


Figure 3.1 Cytosolic nucleic acid pattern recognition and activation of ISGs.

Diverse pathways downstream of these receptors transduce signals that converge on a few key molecules, such as the IFN-regulatory factor (IRF) family of transcription factors that activate the transcription of genes encoding IFN α/β . In most cases, IRF3 and IRF7 phosphorylation are fundamental to induce IFN α/β gene transcription. While the initial production of IFN α/β relies on IRF3, IRF7 mediates a positive feedback loop, leading to the induction of a second wave of gene transcription, including additional IFN α -encoding genes (Guiducci et al., 2008; Kawai and Akira, 2011; McNab et al., 2015).

A model for the induction of the *IFN β* gene was proposed by Lomvardas and Thanos (Lomvardas and Thanos, 2001): The nucleosome that obstructs transcription begins 5 bp downstream of the TATA box and extends over the start site of transcription. The enhanceosome, which forms upon virus infection, counteracts this repressive effect by instructing a recruitment program beginning with HATs, followed by recruitment of SWI/SNF via CBP. Then, SWI/SNF modifies the histone-DNA contacts, causing changes in the super helicity of the DNA around the histone core, without changing the relative position of the histone core relative to DNA. However, these changes suffice for the binding of TBP to the nearby TATA box, thus completing the assembly of the basal machinery. Simultaneously, the radical DNA bend induced by TBP on DNA causes

repositioning of the nucleosome to a new location 36 bp downstream, thus fully exposing the start site of transcription (Fig. 3.2). Although this promoter region controls the expression of the *IFN β* gene, it also functions as an inducible enhancer suggesting that the *IFN β* gene promoter is indeed an Epromoter (Goodbourn et al., 1985).

Cells from the innate immune response, NK (natural killer), and NKT (natural killer T) cells constitutively express IFN γ , allowing a rapid induction of secretion after stimulation. While the adaptive immune cells CD4⁺ and CD8⁺ naïve T cells produce only small quantities of IFN γ right after their initial activation. However, naïve CD4 and CD8 T cells can gain the ability to efficiently transcribe IFN- γ over several days. This process depends on their proliferation, differentiation, and upregulation of IFN- γ -promoting transcription factors (Schoenborn and Wilson, 2007). Type I IFNA, IL-12, IL-15, and IL-18 are capable of inducing IFN γ production in NK cells (Pegram et al., 2011).

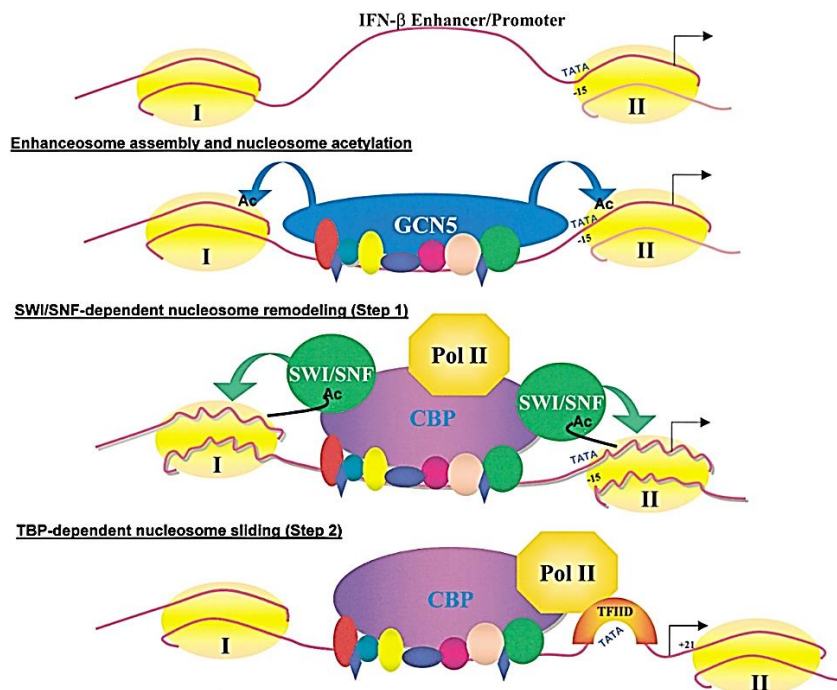


Figure 3.2 Model depicting the chromatin remodelling at the IFN β promoter. Adapted from (Lomvardas and Thanos, 2001)

III. Receptor and signal transduction

IFN β and all the IFN α subtypes bind to and signal through the IFN α receptor subunits 1 (IFNAR1) and IFNAR2, a heterodimeric transmembrane receptor. Ligation of IFNAR activates the receptor-associated protein tyrosine kinases Janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2). In the canonical pathway, the activated JAK1 and TYK2 phosphorylate the signal transducer and activator of transcription 1 (STAT1) and STAT2 molecules that are present in the cytosol, leading to the dimerization, nuclear translocation, and binding of these molecules to IRF9 to form the interferon-stimulated gene factor 3 (ISGF3) complex. This complex then binds to IFN-stimulated response elements (ISRE, consensus sequence: A/GNGAAANNGAAACT) in the

promoter regions of IFNs or IFN-stimulated gene (ISG), leading to the activation of ISG transcription (Fig 3.3). In this manner, IFN α/β induces the expression of several hundred ISGs, many of which function to induce an antiviral state within the cell (McNab et al., 2015).

The type III IFNs signal through a shared heterodimeric receptor, IFNLR, comprised of IFNLR1 (also termed IL28R α) and IL10R β (Lazear et al., 2019). Just like type I IFNs, the IFNLR activates the JAK-STAT signaling, recruiting IRF9, and formation of active ISGF3 complex (Fig. 3.3).

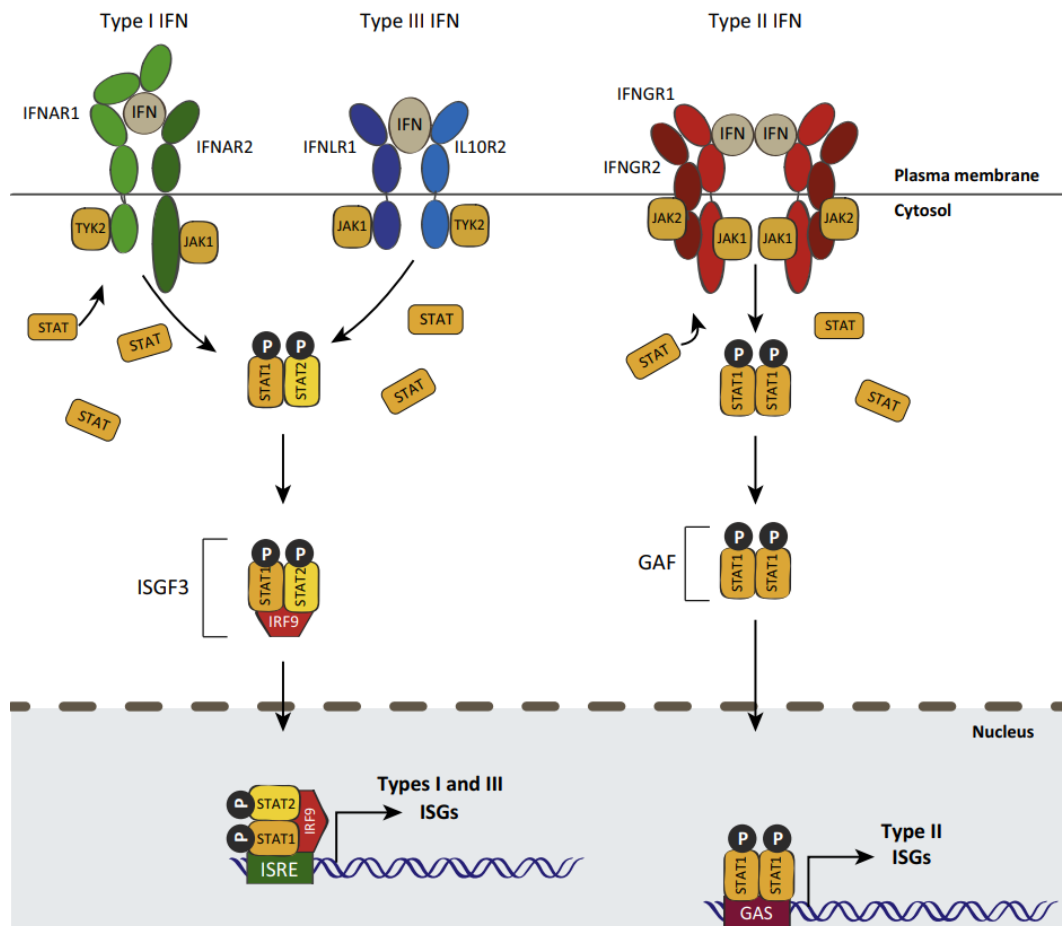


Figure 3.3 Canonical IFN signalling. Modified from (Hoffmann et al., 2015)

While the canonical IFN signaling is defined by the STAT1/STAT2 heterodimerization, the activation of IFNAR can also trigger the formation of homodimers of phosphorylated STAT1. STAT1 homodimers bind to IFN γ response elements, leading to the induction of the IFN γ -induced genes (Platanias, 2005). Also, IFN-induced STAT6 can form a complex with STAT2 and promote the antiproliferative activities of IFN via Sp1 and BCL6 (Hsu et al., 2016). In addition to the JAK/STAT pathways, other STAT-independent signaling pathways in different immune cells have been described, including c-Jun N-terminal kinase (JNK), p38 mitogen-activated protein kinase (MAPK), the extracellular signal-regulated kinase (ERK), and the mammalian target of rapamycin (mTOR) (Platanias, 2005).

The STAT family contains seven members (STAT1-5A, 5B, and 6) and shared the same domain architecture an N-terminal domain, a coiled-coil domain, a DNA-binding domain, a linker domain, an Src homology 2 domain, and a transactivation domain (Steen and Gamero, 2013). In the ISGF3 complex, STAT2 contributes with a potent transactivation domain, since has been shown to bind and recruit important transcriptional co-activators, such as p300/CPB (Bhattacharya et al., 1996), whereas STAT1 stabilizes the complex by providing additional DNA contacts. Finally, a dominant nuclear export signal present in the C-terminus of STAT2 returns the STAT2-IRF9 complex to the cytoplasm (Banninger and Reich, 2004).

The IFN γ signals through the IFN γ receptor (IFNGR) composed of IFNGR1 and IFNGR2 subunits. The canonical signaling pathway, ligation of IFN γ to the IFNGR leads to activation of JAK1 and TYK2, is resulting in phosphorylation and homodimerization of STAT1 to form the gamma interferon activation factor (GAF) (Gotthardt and Sexl, 2016; Lee and Ashkar, 2018). The GAF complex can bind to the gamma interferon activation site elements (GAS, consensus sequence: TTCCNGGAA) for the rapid transcriptional induction of genes in response to IFN γ (Fig. 3.3). However, like type I IFN, IFN- γ has also been shown to signal through alternative pathways, including STAT4, Erk1/2, Pyk2, and CrkL, among others (Lee and Ashkar, 2018).

IV. Interferon regulatory factors

The mammalian interferon regulatory factor (IRF) family of transcription factor comprises nine members (IRF1-9) that possess a conserved amino-terminal DNA-binding domain characterized by a series of five well-conserved tryptophan-rich repeats. This DNA binding domain is a helix-turn-helix domain that recognizes sequences like the ISRE (A/GNGAAANNGAAACT) (Zhao et al., 2015). The carboxy-terminal domain is more diverse amongst family members and confers their unique function via regulating their ability to interact with each other and proteins outside of the IRF family. IRF1, IRF3, IRF7, and IRF9 are considered positive regulators or activators of IFN transcription. In particular, IRF9 participates in primary type I and II IFNs responses, by activating the transcription of ISRE-containing ISGs, including IRF1. Thus, IRF1 participates in the secondary response to IFN-I and -II (fig. 3.4) (Michalska et al., 2018). IRF5 and IRF8 are frequently grouped as their transcriptional activities can be positive or negative depending on the context. IRF2 and IRF4 are generally considered inhibitory IRFs. On the other hand, IRF4 antagonizes IRF5 largely through binding to similar sequences on target genes as well as by competing for MyD88 interaction. IRF4 regulates both CD4 and CD8 T cell differentiation, as well as B cell maturation (Jefferies, 2019).

Although the primary function of the IRFs was originally thought to be restricted to the immune response (innate and adaptive immunity), they play a critical role in other biological processes like metabolism (Zhao et al., 2015).

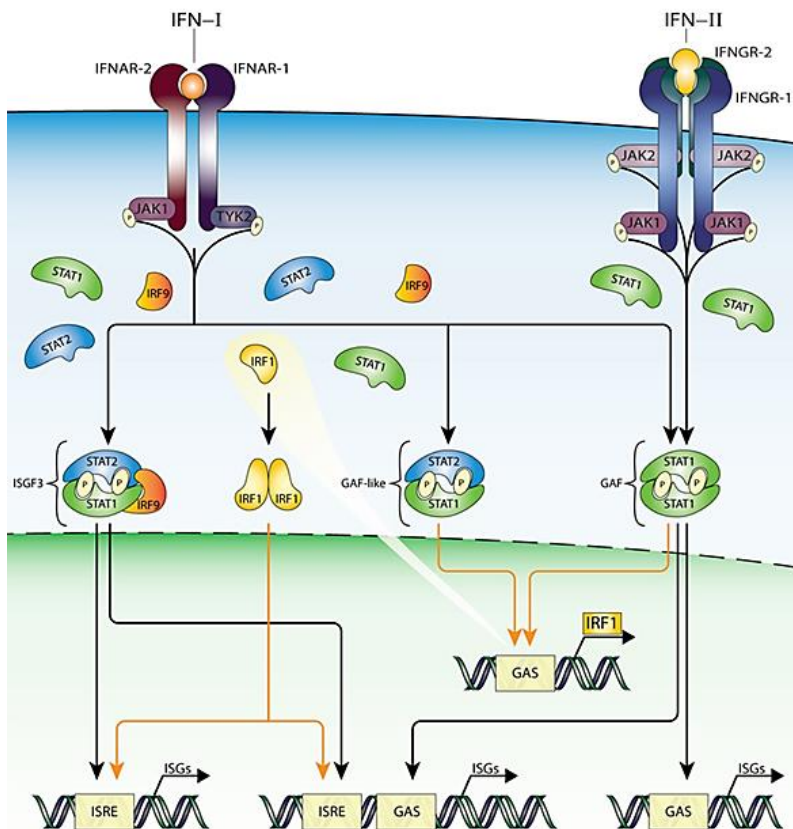


Figure 3.4 Primary (black arrow) and secondary (orange arrow) type I and II IFN response. Modified from (Michalska et al., 2018)

V. Chromatin Remodeling

Like other transcription factors, ISGF3 interacts with co-activator or co-repressor proteins that mediate interactions with the general transcriptional machinery and also facilitate the recruitment of enzymes that modify chromatin: for example, HATs, HDACs, and histone methyltransferases.

STATs and ISGF3 can remodel chromatin to activate regulatory elements. During cell differentiation, lineage-determining transcription factors create a cell-type-specific regulatory landscape by allowing access to certain regulatory elements (promoters and enhancers); in turn, the regulatory landscape of accessible chromatin determines the binding pattern of signal-activated factors, such as STATs (Ivashkiv and Donlin, 2014). STAT1 primes and/or activates enhancers genome-wide during the differentiation or polarization of T cells and macrophages and can contribute to the creation of new enhancers in differentiated cells (Fig. 3.5) (Ostuni et al., 2013). Thus, STAT1 can alter the epigenetic landscape, suggesting that IFNs have the potential to alter cell responses to subsequent stimuli. Type I IFN-induced ISG expression is also controlled quantitatively in a cell lineage-specific manner by the negative histone mark H3K9me2, which attenuates gene expression (Ivashkiv and Donlin, 2014; Sacconi and Natoli, 2002).

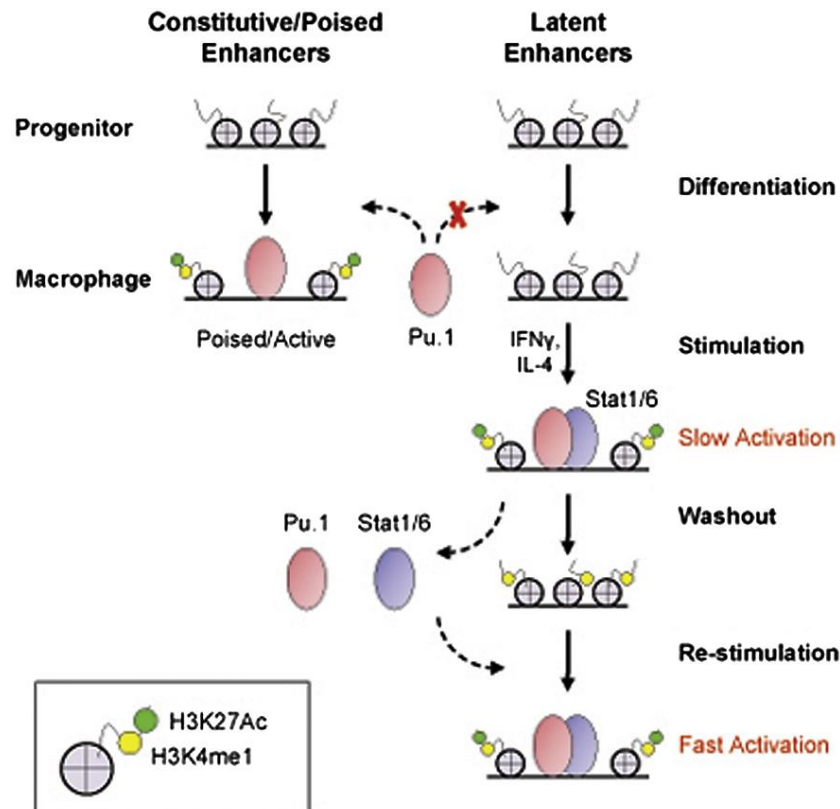


Figure 3.5 Activation of latent enhancers during differentiation mediated by STAT6 and Pu.1. Adapted from (Ostuni et al., 2013)

VI. Interferonopathies and autoimmune diseases

The type I and II IFNs can control the innate immune response; however, IFNs also have been implicated in several non-infectious pathological conditions, making their pathways a target for pharmacotherapies.

Interferonopathies are a genetically and phenotypically heterogeneous group of disorders of the innate immune system caused by constitutive activation of type I and II IFNs characterized by autoinflammation and varying degrees of autoimmunity or immunodeficiency (Bienias et al., 2018). The interferonopathies are caused by monogenic Mendelian mutations, however, some non-Mendelian disorders like systemic lupus erythematosus (SLE) are also classified as interferonopathies [reviewed in (Lee-Kirsch, 2017; Rodero and Crow, 2016; Volpi et al., 2016)].

Also, perturbations in the IRFs can result in dysregulated immune responses and potentially autoimmune diseases (Jefferies, 2019). For example, IRF9 deficiency disrupts both the IFN-I amplification loop and the ability to control viral replication, leading to the establishment of a chronic infection and T cell exhaustion. Also, polymorphisms within IRF5 increase both susceptibility to and pathogenesis of hepatitis B virus (HBV) infection.

Nowadays, the IFN is considered as a biomarker in autoimmune diseases and the differential gene expression of the IFN-induced gene is known as “IFN signature”. This signature has been

used to predict the evolution of autoimmune diseases (Castaneda-Delgado et al., 2017; Lin et al., 2019) and the response to treatment (Cooles et al., 2018).

VII. Concluding remarks

The IFNs play an essential role in the antiviral response and the cell maturation from the innate and adaptive immune response. Normally the ISGF3 complex requires the phosphorylation of STAT1 and STAT2, however unphosphorylated ISGF3 and unphosphorylated STAT2/IRF9 complex have been found in the nucleus of senescent cells under physiological conditions. This leads to the induction of the transcription of ISGs independent of JAK1-mediated phosphorylation (Yamagami et al., 2018). STAT1 and STAT2 might be related to aging and/or age-related pathologies and could be a therapeutic target.

With the surge of the coronavirus disease 2019 (COVID-19) pandemic, researchers have been looking for a treatment for the 5-10% of infected people that progress to a severe or critical disease (Salje et al., 2020). The exacerbated inflammatory response has been associated with a massive influx of innate immune cells, namely neutrophils and monocytes, which may aggravate lung injury and precipitate acute respiratory distress syndrome. These severe patients displayed an excessive NF- κ B-driven inflammatory response associated with increased TNF- α and IL-6 and an impaired type I IFN production associated with lower viral clearance (Hadjadj et al., 2020). The late suggest that IFN could be used as a severity biomarker and might provide therapy in combination with other therapeutic agents (Andreacos and Tsiodras, 2020; Lu et al., 2020; Zhou et al., 2020).

Overall, IFN homeostasis is important to guaranty a balance between immunity and immune tolerance of healthy self-tissues. Therefore, is important to understand the precise epigenetic mechanisms involved in the rapid and precise activation of IFN-stimulated genes.

Chapter 4. Role of the Epromoters in the type I interferon response

I. Objectives

The Epromoters are a novel class of regulatory elements that possess dual promoter and enhancer functions. Using the enhancer assay CapSTARR-seq, our team previously showed that 3% and 2.4% of the total promoter tested in K562 and HeLa, respectively, displayed enhancer activity *in vitro*. These Epromoters were frequently associated with genes related to stress response, and particularly in HeLa they were related to IFN I stress response. This suggests a link between the Epromoter activity and the stress response to type I interferon (Dao et al., 2017). Therefore, to characterize the role of Epromoters in the response to type I interferon (IFNa), we performed the CapSTARR-seq with all the Ref-seq-defined promoter. In parallel, we also performed RNA-seq. These experiments were done in the ENCODE cell line K562 with and without IFNa stimulation.

The identification of the IFNa induced Epromoters and its functional characterization has become the central focus of my study. This chapter is given in a manuscript format ready to be submitted. There are some outstanding questions that I will address in this chapter, including:

1. Can the enhancer activity of promoters be induced by the IFN response?
2. What are the distinct features of the Epromoters induced by the IFN response?
3. Is it possible to dissociate the enhancer and the promoter functions from the Epromoters?

II. Main results

Using the Ref-seq-defined promoter library in the CapSTARR-seq technique and the RNA-seq in IFNa-stimulated and non-stimulated K562 cells we identified a set of IFNa-stimulated Epromoters. After pairing the data coming from RNA-seq and CapSTARR in both conditions we identify three subsets of promoters: (i) promoters associated with IFNa-induced genes but without induced Epromoter activity, (ii) promoters with induced Epromoter activity and associated with IFNa-induced genes, and (iii) promoters associated with IFNa-induced Epromoter activity and associated with non-induced genes. We showed that the IFNa-induced Epromoters with the induced gene is enriched in the IFNa key transcription factors *in vivo* (ChIP-seq) and *in silico* (binding site discovery). We also have shown that some IFNa-induced Epromoters cluster with IFNa-induced genes and that the deletion of the Epromoter can impact the expression of the clustered genes. We illustrate this effect with the Epromoters of *ISG15* and *OAS3*. Moreover, we dissect the Epromoter of the *OAS3* gene to try to understand what transcription factors binding sites are essential for the enhancer and the promoter activity of this Epromoter. These results support the model that under stress conditions, the Epromoters constitute a key element in the transcriptional regulation of clusters of stress-response genes.

III. Contributions

Experimental contribution:

To carry out this project, my supervisor and I routinely discussed and conceptualized the experimental designs. I designed and performed almost all experimental works, except the experiment presented in Figure 6 of the article, including:

- Analysis of the CapSTARR-seq and RNA-seq in K562 cell line with and without IFN α
- CHIP-seq
- Design and perform CRISPR/Cas9-mediated knock-out
- qPCR analysis, gene expression analysis
- Cloning, Luciferase assay, mutagenesis

Manuscript contribution:

I contributed to the writing of the manuscript and editing figures.

IV. Manuscript

Epromoters function as a hub to recruit key transcription factors required for the inflammatory response

Authors: David Santiago-Algarra^{1,2,&}, Charbel Souaid^{1,2,&}, Himanshu Singh^{1,2,&}, Lan Dao^{1,2,3}, Alejandra Medina⁴, Lucia Ramirez⁴, Jaime A Castro-Mondragon^{1,5}, Nori Sadouni^{1,2}, Guillaume Charbonnier^{1,2}, Salvatore Spicuglia^{1,2,*}

¹Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France.

²Equipe Labéllisée Ligue Contre le Cancer

³Vinmec Research Institute of Stem cell and Gene technology (VRISG), Hanoi, Vietnam

⁴Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Mexico

⁵Present address: Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo, 0318 Oslo, Norway

&Equal contributions

*Correspondence: salvatore.spicuglia@inserm.fr

Abstract

Gene expression is controlled by the involvement of gene-proximal (promoters) and distal (enhancers) regulatory elements. Our previous results have demonstrated that a subset of gene promoters, termed Epromoters, works as bona fide enhancers and regulates distal gene expression. Here, we hypothesized that Epromoters play a key role in the coordination of rapid gene induction in the stress response, particularly during inflammation. Using a high-throughput reporter assay we explored the function of Epromoters in response to type I interferon. We found that clusters of IFN α -induced genes are frequently associated with Epromoters and that these regulatory elements preferentially recruit the STAT1/2 and IRF transcription factors and distally regulate the activation of interferon-response genes. Consistently, we identified and validated the involvement of Epromoter-containing clusters in the regulation of LPS-stimulated macrophages. Our findings suggest that Epromoters function as a hub recruiting the key TFs required for coordinated regulation of gene clusters during the inflammatory response, and more generally upon cellular response to intra- and extra-cellular signals.

Keywords: Epromoter, interferon alpha, inflammatory response, gene regulation, high-throughput reporter assay

Introduction

Regulation of gene transcription in higher eukaryotes is accomplished through the involvement of transcription start site (TSS)-proximal (promoters) and -distal (enhancers) regulatory elements. It is now well acknowledged that enhancer elements play an essential role during development and cell differentiation, while genetic alterations in these elements are a major cause of human diseases (Chatterjee and Ahituv, 2017). The classical definition of enhancers implies the property to activate gene expression at a distance, while promoters induce local gene expression. However, this basic dichotomy has been challenged by broad similarities between promoters and enhancers (Andersson and Sandelin, 2020; Medina-Rivera et al., 2018).

While epigenomic studies allow genome-wide identification of putative enhancers, they do not provide direct proof of enhancer function or activity. To tackle this problem, several high-throughput reporter assays have been developed and implemented to study enhancer activity in different cellular contexts (Santiago-Algarra et al., 2017). Using high-throughput reporter assays in different cellular contexts from drosophila to humans, it was found that a subset of gene-promoters, also termed Epromoters, displays enhancer activity when tested *in vitro* (Arnold et al., 2013; Corrales et al., 2017; Dao et al., 2017; Nguyen et al., 2016a; Zabidi et al., 2015). Importantly, several concomitant studies provided evidence that some core promoters indeed control distal gene expression in their natural context (Dao et al., 2017; Diao et al., 2017a; Engreitz et al., 2016; Rajagopal et al., 2016), and that human genetic variation within Epromoters influences distal gene expression, as assessed by expression Quantitative Trait Loci (eQTL) (Dao et al., 2017; Jung et al., 2019). Overall, the finding that a subset of promoters works also as *bona fide* enhancers have significant implications for the understanding of complex gene regulation in normal development and open the intriguing possibility that physiological traits or disease-associated variants lying within a subset of (E)promoters might also directly impact distal gene expression (Medina-Rivera et al., 2018).

Previous studies have suggested a link between the Epromoter function and the stress responses, particularly the regulation of interferon-response genes (Dao et al., 2017; Dao and Spicuglia, 2018; Muerdter et al., 2018). The inflammatory response requires the activation of a complex transcriptional program that is both cell-type- and stimulus-specific and involves the dynamic regulation of hundreds of genes (Smale and Natoli, 2014). In the context of inflamed tissue, extensive changes in gene expression occur in both parenchymal cells and infiltrating cells of the immune system. Transcriptional regulation of interferon (IFN)-response genes is one of the most intensively studied regulatory processes. In general, type I (IFN α /IFN β) and II (IFN γ) interferons are responsible for regulating and activating the immune response. Expression of type I interferons can be induced in virtually all cell types upon recognition of viral components, especially nucleic acids, whereas type II interferon is induced by cytokines such as IL-12, and its expression is restricted to immune cells. While most transcription factors (TFs) involved in type I and II response are well characterized (e.g. STAT1/2 and IRFs for type I), there are still open questions concerning the precise epigenetic mechanisms involved in the rapid and precise activation of IFN α -stimulated genes. For instance, not all promoters of IFN α -stimulated genes are

bound by the aforementioned factors and underlying mechanisms still await further clarification (Platanias, 2005).

We hypothesize that Epromoters might play an important role in the coordination of rapid gene induction after IFN α stimulation, and more generally during the inflammatory response. By combining high-throughput reporter assays and gene expression analyses, we found that a significant subset of IFN α -induced genes was associated with Epromoters and regulates the induction of other neighbor genes. Consistently, we found that IFN α -induced Epromoters have the ability to specifically and efficiently recruit the interferon-response transcription factors. In fact, within a typical cluster of IFN α -response genes, the interferon response factors were found to exclusively bind to the Epromoter and to be required for the proper induction of other genes within the cluster. Predictions based on gene expression dynamics and TF binding profiles allow us to identify and validate Epromoter-regulated clusters in primary LPS-stimulated macrophages. Thus, Epromoters have a broad role in the induction of co-regulated genes during the mammalian inflammatory response.

Results

Epromoters are involved in type I interferon response

To initially explore the link between Epromoters and type I interferon response, we analyzed active enhancers in HeLa cells in the presence or absence of Interferon type I response inhibitors, using published whole-genome enhancer screen from Self-Transcribing Active Regulatory Region Sequencing (STARR-seq) experiments (Muerdter et al., 2018). Noted that, HeLa cells have active interferon signaling in the absence of exogenous stimulation (Dao et al., 2017; Muerdter et al., 2018). The genomic distribution of active enhancers shows that the interferon-dependent enhancers (active without inhibitors) are significantly closer to the TSS (P val. = 2.2×10^{-16} ; KS test) as compared to interferon-independent enhancers (actives in the presence of inhibitors) (**Figure S1A**). Indeed, interferon-dependent enhancers are preferentially located within a 1 kb window to the TSS, with a median distance of 458.5 bp (**Figure S1B**). Therefore, we separated HeLa active enhancers into proximal (<1 kb from the closest TSS; Epromoters) and distal (>1 kb from any TSS). Interferon-dependent TSS-proximal enhancers (Epromoters) were enriched in Interferon-Stimulated Response Elements (ISRE), including binding sites for TFs of the IRF family (**Figure S1C**). In contrast, TSS-proximal enhancers in the presence of the interferon inhibitors, as well as, the TSS-distal enhancers were enriched in binding sites for developmental TFs (**Figure S1D**). This suggests that TSS-proximal enhancers (i.e. Epromoters) are preferentially activated by the type I interferon signaling in HeLa cells.

To directly assess the contribution of Epromoters to the regulation of type I interferon response, we performed paralleled experiments to assess gene expression (RNA-seq) and enhancer activity of gene promoters (CapSTARR-seq) (**Figure 1A; Table S1**) in the K562 cell line with and without IFN α stimulation. CapSTARR-seq, is a high-throughput reporter assay, coupling capture of regions of interest to STARR-seq (Dao et al., 2015; Vanhille et al., 2015b) and was previously

used to identify Epromoters in unstimulated K562 cells (Dao et al., 2017). Notably, the K562 cells do not express type I interferon response genes in the absence of stimulation and does not induce an interferon response after DNA transfection (Dao et al., 2017; Muerdter et al., 2018), making this cell line an appropriated model to study type I interferon stimulation. To quantify the enhancer activity of gene promoters, we used a capture-based library containing 17,941 promoters (-200 bp to +50 bp from the TSS), corresponding to 14,188 RefSeq-defined coding genes as previously described (Dao et al., 2017). Analyses of RNA-seq after IFN α stimulation of K562 cells resulted in 426 induced and 436 repressed genes (P val. <0.001; **Figure 1B**). However, induced genes generally reached higher significance as compared to repressed genes, consistent with previous findings (Duncan et al., 2015). Highly IFN α Induced genes included classical type I interferon response genes such as the *MX*, *OAS*, and *IFIT* family of genes (**Table S1**).

The enhancer activity was calculated as the fold change of the CapSTARR-seq signal over the input library. Gene promoters were defined as Epromoter when the enhancer activity was above the inflection point of the ranked signal of all captured regions, as previously described (Dao et al., 2017). Induced Epromoters were defined as Epromoters that gained enhancer activity by a fold change of 2 between IFN α stimulated and non-stimulated K562 cells. We identified 440 constitutive Epromoters, as well as 70 induced and 7 repressed Epromoters (**Figure 1C**). Combining the RNA-seq and CapSTARR-seq experiments allowed us to define three sets of IFN α -response loci in K562 cells (**Figure 1D-E; Table S2**): (i) 498 promoters associated with 394 IFN α -induced genes but without induced Epromoter activity (induced gene only), (ii) 44 promoters with induced Epromoter activity and associated with 38 IFN α -induced genes (induced gene & Epromoter), and (iii) 26 promoters associated with IFN α -induced Epromoter activity and associated with 25 non induced genes (induced Epromoter only). Note that the set of “induced gene & Epromoter” displayed stronger induction than the set of “induced gene only” (**Figures 1D and S2A**). Examples of “induced gene & Epromoter” (*SP100*) and “induced gene only” (*GBP4*) loci are shown in Figure 1F. Taken together these results suggest a potential involvement of Epromoters in the type I interferon response.

IFN α -induced Epromoters specifically recruit the interferon response factors

To better understand the relevance of the three sets of IFN α -stimulated loci, we analyzed the functional enrichment of the associated genes. As expected, the set of “induced gene only” was primarily linked to type I interferon response and related pathways. However, the set of “induced gene & Epromoter” achieved higher enrichment in those pathways even though they represent a minority (9%) of IFN α -response genes (**Figure 2A; Table S3**). The set of “induced Epromoter only” was not associated with any biological process, likely due to the low number of genes.

Classical type I interferon response requires binding to the ISRE by the regulatory complex formed by phosphorylated STAT1 and STAT2 TFs with the constitutive IRF9 TF also referred to as the ISGF3 complex (Chen et al., 2017; Smale and Natoli, 2014). Subsequent activation can be

also mediated by the inducible IRF1 TF. To assess the binding of these TFs to the three sets of IFN α -response promoters, we analyzed previously generated ChIP-seq data in IFN α -stimulated K562 cells for STAT1, STAT2, and IRF1 (**Table S4**) and generated ChIP-seq for IRF9. Strikingly, all four TFs preferentially bound to the set of “induced gene & Epromoter”, while the set of promoters associated with “induced gene only” and “induced Epromoter only” displayed binding levels close to the background (**Figure 2B**). Recruitment of these TFs was specific to IFN α -induced Epromoters as no binding was observed in non-induced Epromoters (**Figure S2B**).

To gain insight into the epigenetic dynamic of these promoters we performed ChIP-seq for the histone modifications H3K4me3 and H3K27ac in non-stimulated and IFN α -stimulated K562 cells (**Figure 2C, Table S4**). The strongest gain of histone modifications with H3K4me3 and H3K27ac was observed for the set of “induced gene & Epromoter”. This suggests that the set of “induced gene & Epromoter” has the ability to recruit the key TFs associated with the interferon response and acquires a highly active chromatin state, likely associated with their enhancer activity.

To determine whether the preferential binding of IFN-response TFs to the induced Epromoters was due to an intrinsic feature of the DNA sequence we analyzed the enrichment in TFBS in the three groups of promoters. *De novo* discovery of DNA motifs (**Figure 2D**) as well as global enrichment analyses of known motifs (**Figure 2E**), demonstrated that promoters associated with the “induced gene only” and “induced gene & Epromoter” sets, but not with the “induced Epromoter only” set, were similarly enriched in STAT and IRF binding sites and generally contain the consensus ISRE. Therefore, the presence of STAT/IRF binding sites did not explain *per se* the higher efficiency of TF recruitment observed at induced Epromoters.

We observed that 54.55% of the promoters associated with the “induced gene & Epromoter” set harbor a consensus ISRE site, while only 10.24% of the promoters associated with the “induced gene only” set contain the motif (**Figure 2D**). Consistently, a differential enrichment analysis shows that the STAT and IRF motifs were significantly enriched in the “induced gene & Epromoter” set as compared with the “induced gene only” set (**Figure 2F**). To assess whether the difference in the density of TFs could be different between the three promoter sets, we first computed the number of sites per promoter found for each IRF factor (**Figure S3**). We observed that the density of IRF sites per promoter was significantly higher in the “induced gene & Epromoter” set as compared with the two other sets.

Next, we computed the number of non-redundant ISRE sites *per* promoter by combining all STAT1-2 and IRF1-9 sites (**Figure 2G, top panels**). While the “induced gene & Epromoter” set harbor a majority of promoters with 2 or more ISRE sites (55%), the sets of “induced gene only” and “induced Epromoter only” were associated with a majority of promoters with no ISRE sites (60% and 58 %, respectively). Moreover, the same analysis performed with high confidence sites shows that the majority of the sites found in the promoters of the “induced gene only” set are of lower affinity (**Figure 2G, bottom panels**). This suggests that both the high density and the quality of ISRE sites likely contribute to the efficient recruitment of the STAT1/2 and IRFs complexes and that this is a specific property of the IFN α -stimulated Epromoters.

Clusters of IFN α -induced genes are regulated by Epromoters

The above results showed that the majority of IFN α -responding genes does not efficiently recruit the key IFN α -response TFs. One potential explanation is that the induction of these genes requires the action of distal regulatory elements, which might involve either typical enhancers or Epromoters. We reasoned that for induction to take place, a given locus needs to be associated with a regulatory element recruiting the ISGF3 complex (i.e., STAT1/STAT2/IRF9). The IFN-response element might be located at the promoter of the same gene or at another regulatory element, which might overlap (Epromoter-like) or not (TSS-distal or “typical” enhancers) the promoter of another gene. To address this issue, we determined the distance of the closest binding of the ISGF3 complex (i.e. overlapping STAT1/STAT2/IRF9 ChIP-seq peaks) to the promoter of all “induced genes only” set and classified it as located within the same promoter, in a TSS-distal region or in another promoter (**Figure 3A**). As expected only a minority of the promoters associated with the “induced genes only” loci (17.30 %) recruited the ISGF3 complex. For 48.60 %, the closest ISG3 binding was located in a TSS-distal region. However, for more than one-third of induced genes loci (34.10 %), the closest ISGF3 peak was found in another promoter (**Figure 3A**). Importantly, the binding of ISGF3 complex to a distal promoter was much more frequent than expected by chance (**Figure 3B**).

We hypothesize that many induced genes might be located in clusters, likely sharing the same IFN α -stimulated regulatory elements. Therefore, we analyzed the genomic distance between IFN α -induced genes. We found that induced genes were located closer to each other than expected by chance (**Figure 3C**; P val. 10^{-322} ; KS test), and frequently to less than 100 kb from a constitutive or induced Epromoter (**Figure S4A**), suggesting that many IFN α -induced genes are located in clusters and might be co-regulated by Epromoters.

Based on the above observations, we aimed to identify clusters of induced genes and Epromoters for which the corresponding TSS are located less than 100 kb from each other (**Figure 3D**). We identified a total of 49 clusters encompassing 121 IFN α -induced loci (**Table S5**). Of these, 21 (42%) were associated with either constitutive or induced Epromoters (**Figure 3E**) and were generally located within the same Topological Associated Domain (TAD) in K562 cells. Among these clusters, we found typical IFN α -response gene families, such as *OAS*, *IFIT*, *MX*, and *TRIM* (**Figure 3E**). The lack of association with Epromoters for some clusters might be due to our relatively stringent criteria to select induced genes and Epromoters. For instance, the *APOBEC3* locus was found to contain three induced genes (*APOBEC3D*, *APOBEC3F*, and *APOBEC3G*), however, it also contains the *APOBEC3C* gene which has an Epromoter, but the differential expression was slightly under the applied threshold (**Figure S4B**). Overall, our finding raises the possibility that a subset of induced genes might be coregulated by Epromoters located in the same cluster.

Consistent with the average patterns of TF binding, visual inspection of several clusters indicated that IFN-response TFs bind preferentially to the induced Epromoters (**Figures 4 and S4B-5A**). This is the case of the *ISG15/AGRN/HES4* cluster (**Figure 4A**). In this cluster, all three genes are induced by IFN α stimulation, but only *ISG15* is associated with an inducible

Epromoter. *ISG15* encodes for an IFN-induced ubiquitin-like protein and plays a central role in the host antiviral response (Perng and Lenschow, 2018). *HES4* and *AGRN* are also classified as interferon-stimulated genes (ISG) in the Interferome database (Rusinova et al., 2013), while *HES4*, encoding for a bHLH (basic helix loop helix) TF, have been suggested to play a role in the IFN response (Jin et al., 2010; Kikuchi et al., 2019b). As shown in Figure 4A, the induced *ISG15* Epromoter, recruits the TFs STAT1, STAT2, IRF1, and IRF9 in IFN α -stimulated K562 cells, while the other induced genes located in the same cluster, *AGRN*, and *HES4*, do not. The three genes are located in the same TAD and no other regions, apart the *ISG15* Epromoter, was found to bind the TFs within the TAD (**Figure S5**, upper panel). This suggested that the *ISG15* Epromoter might harbor the necessary TFs to induce the transcription of all three genes in an Epromoter-dependent manner. To test this hypothesis, we deleted the *ISG15* Epromoter in K562 cells and analyzed the gene expression of associated genes after different times after IFN α stimulation (**Figure 4B**). As expected, the expression of the *ISG5* gene was severely impaired. Although the *AGRN* and *HES4* genes have distinct stimulatory kinetics, the induction of both genes was significantly impaired in Δ Ep*ISG15* cells as compared to WT K562 cells. Thus, the ISG15 Epromoter is required for the accurate induction of the two neighbor IFN α -responsive genes.

We observed that some of the IFN α -induced clusters contained more than one gene associated with an Epromoter (**Figure 3D**). One example is provided by the *IFIT* locus, harboring a family of genes encoding for IFN-induced proteins with tetratricopeptide repeats and have broad-spectrum activity against replication, spread, and disease pathogenesis of a range of human viruses (Diamond and Farzan, 2013). The cluster contains four IFN α -induced genes; *IFIT3*, *IFIT1*, and *IFIT5* have inducible Epromoters that recruit all four TFs, while *IFIT2* does not have an Epromoter and only recruit the IRF1 factor (**Figure S6A**). All five genes were located in the same TAD (**Figure S5**, middle panel). Interestingly, the deletion of one of the Epromoter of *IFIT3* resulted in a specific impairment of *IFIT2* expression at the earliest (2h) and the latest (8h) kinetics point (**Figure S6B**). Thus, proper induction of *IFIT2* requires the function of at least one induced Epromoter within the cluster. One reason to explain the moderate impact of *IFIT3* Epromoter on *IFIT2* expression would be the synergistic interaction with the other Epromoters on the locus.

Another noticeable example is seen in the *OAS* locus, harboring three related genes encoding oligoadenylate synthase (OAS) family proteins, and playing a crucial antiviral function (Hornung et al., 2014). In this cluster, all three IFN α -induced genes were associated with inducible Epromoters (**Figure 4C**). The cluster was divided into two TADs in non-stimulated K562 cells (**Figure S5**, lower panel). However, visual inspection of the Hi-C matrix suggested that the insulation at the TAD boundary within the *OAS* cluster in non-stimulated cells is very weak. We noticed that the *OAS3*-associated Epromoter displayed a higher binding of the four interferon-response TFs, as compared with the *OAS1* and *OAS2* Epromoters. Therefore, we asked whether the specific binding of IFN-response factors was a predictor of the distal regulatory activity of the Epromoter. To explore the specific contribution of each Epromoter towards the regulation of the entire cluster, we deleted each of the *OAS* Epromoters in K562 cells. The deletion of *OAS1* and *OAS2* Epromoters did not affect the induction of *OAS3* expression (**Figure S7A**). However,

the deletion of the *OAS3* Epromoter resulted in a dramatic reduction of *OAS1* and *OAS2* induction after IFN α stimulation (**Figure 4D**). Overall, these results suggested that within a cluster of IFN-induced genes only the promoters that efficiently recruit the key IFN-response factors have the capability to provide a distal regulatory function.

Promoter versus enhancer activity of the *OAS3* Epromoter

A key question is whether enhancer and promoter activities are dictated by the same or distinct regulatory sequences. To start addressing this issue we analyzed the TF binding sites IFN α -inducible *OAS3* Epromoter. The *OAS3* Epromoter harbors two ISRE binding sites (hereafter ISRE¹ and ISRE²) as well as a RELA binding site (NF κ b) in proximity to ISRE1 (**Figure 5A**). Notably, ISRE¹ corresponds to a canonical IRF9 motif and is closer to the edge of the STAT1/2 and IRF1/9 binding peaks (**Figure 5A-B**). We set up a luciferase reporter assay to assess the contribution of each binding site to the enhancer and promoter activities. As expected, the *OAS3* Epromoter displayed IFN α -inducible promoter and enhancer activities in a luciferase assay (**Figure 5C**). Mutation of the ISRE¹ site, or replacement of ISRE¹ by ISRE², impaired enhancer activity by five-fold and promoter activity by two-fold, while mutation of the ISRE² or the NF κ b sites had no significant effect (**Figures 5C and S7B**), suggesting that enhancer activity of the *OAS3* Epromoter is more dependent on the presence of a “consensus” ISRE. The presence of the two ISRE binding sites, however, was absolutely required for the IFN α -dependent activity. In conclusion, the enhancer and promoter activities of the *OAS3* Epromoter rely on the same regulatory motifs but display different sensitivities with respect to the similarity of the sequences to the consensus ISRE motif.

Epromoters are involved in the regulation of an LPS-induced cluster in macrophages

Based on the above results, we reasoned that within a cluster of induced genes in response to extracellular signaling, the promoter that preferentially recruits the key TFs can be predicted as having an Epromoter-like function. We used this prediction to explore the function of Epromoters in other inflammatory responses, namely the primary immune response of macrophages. For this purpose, we analyzed a dataset including stimulation of mouse primary macrophages by Lipopolysaccharides (LPS) (Mancino et al., 2015). The dataset consisted of RNA-seq as well as ChIP-seq for the IRF1, IRF8, and STAT2 TFs before and after LPS stimulation for 4 hours. Briefly, we identified clusters of co-induced genes (for which the TSS are separated by less than 100 kb). Then, determined how many promoters from the cluster recruit a given TF (**Figure 6A**; see Methods). Clusters harboring only one promoter binding at least one of the three aforementioned TFs, while the other promoters were devoid of binding of any of the TFs, were considered as potentially regulated by Epromoters (i.e. Epromoter-clusters). From 252 LPS-induced genes, we identified a total of 21 induced clusters (consisting of 60 genes) (**Figure 6A**; **Table S6**). Of these, 8 clusters were classified as Epromoter-like clusters (**Figure 6B**). For

instances, we identified the *ISG15-AGRN* cluster (**Figure S8A**), which was also validated as an Epromoter dependent cluster in the IFN α response (**Figure S8**).

To further validate the approach, we selected the *Il15ra/Il2ra* cluster (**Figure 6C**) coding for the receptors of the IL15 and IL2 cytokines, respectively. IL15RA (also known as CD215) plays a major role in the modulation of pro-inflammatory response of LPS-stimulated macrophages (Alleva et al., 1997) and in supporting the homeostatic proliferation of CD8 $^+$ T lymphocytes (Mortier et al., 2009). IL2RA (also known as CD25) is mainly known for its role in T cell differentiation and activation (Malek, 2008). Previous studies, however, have observed induction of *IL2RA* and responsiveness to IL2 in LPS-stimulated macrophages (Valitutti et al., 1989; Wammers et al., 2018). This suggests a role of IL2RA in the macrophage-mediated immune response. We observed that the IRF1 and IRF8 factors were bound to the *Il15ra* promoter upon LPS stimulation, but no binding was observed at the *Il2ra* promoter (**Figure 6C**). We predicted that the *Il15ra* promoter might work as an Epromoter to coordinate the induction of both *Il15ra* and *Il2ra* genes after LPS-stimulation of macrophages.

To experimentally validate our hypothesis, we used PMA-induced macrophage differentiation of the human THP-1 monocyte cell line, a classical model to study macrophage function (**Figure S8B-C**). We observed that both *IL15RA* and *IL2RA* genes were induced by LPS stimulation of THP-1 macrophages with similar kinetics and reached their maximum induction after 4h, in line with the coregulation observed in mouse macrophages (**Figure S8D-E**). Next, we examined the 3D chromatin organization of the human IL15RA/IL2RA locus. CTCF binding and Hi-C data in macrophage-differentiated THP-1 cells (Phanstiel et al., 2017) revealed that both genes are in the same TAD within a regulatory subdomain flanked by CTCF (**Figure 6D**). Moreover, analysis of Promoter Capture Hi-C interactions (Javierre et al., 2016) centered on the *IL15RA* promoter showed that the contact frequency (ChICAGO score) with the *IL2RA*-proximal CTCF site increase from monocyte to macrophage differentiation (M0) to reach the highest contact frequency in LPS stimulated macrophages (M1), but not in Il13-stimulated macrophages (M2) (**Figure 6D, bottom panels**). Deletion of the *IL15RA* promoter in THP-1 cells resulted in significant impairment of both *IL15RA* and *IL2RA* induction after 4h of LPS stimulation of *in vitro* differentiated THP-1 macrophages (**Figure 6E**). Therefore, the *IL15RA* promoter function as a bona fide Epromoter to regulate both *IL15RA* and *IL2RA* induction in LPS stimulated macrophages as predicted by the binding profiles of involved TFs and the dynamic of 3D interactions. More generally, we show that it is possible to predict inducible Epromoter elements based on the binding profile of key TFs and that Epromoter-like regulatory elements play a role not only in type I interferon response, but also in other inflammatory responses, such as macrophage activation.

Discussion

By systematically assessing gene expression and enhancer activity of coding-gene promoters in response to type I interferon stimulation we found that a subset of IFN α response genes was associated with Epromoters. IFN α -induced Epromoters were found to preferentially recruit the

key interferon response factors (STAT1/2, IRF1/9) and to be required for the efficient induction of neighbor genes within the same cluster. Functional studies in LPS-stimulated macrophages suggested that Epromoters plays an important role in other inflammatory responses as well.

Previous studies suggested that Epromoters might be involved in stress response and, in particular, the interferon response (Dao et al., 2017; Dao and Spicuglia, 2018; Muerdter et al., 2018). In our previous study, we found that Epromoters actives in HeLa cells, which have an active interferon response, are significantly associated with stress and interferon response (Dao et al., 2017). Indeed, inhibition of type I interferon in HeLa cells, preferentially affect the activity of Epromoters as compared with distal enhancers ((Dao and Spicuglia, 2018; Muerdter et al., 2018); **Figure S1**). Moreover, distal genes interacting with Epromoters in both HeLa and K562 cells are significantly enriched in the interferon response (Dao et al., 2017). In the present study, we formally demonstrated that Epromoters were preferentially activated by the type I interferon stimulation. Induced Epromoters are associated with highly induced genes and correlate with a higher representation of *bona fide* interferon-response genes. Moreover, IFN α -response Epromoters can be found in the proximity of other induced genes and ensure coordinated regulation of interferon response clusters. Notably, IFN α -induced genes with no Epromoters were significantly enriched for interferon related GO terms arguing that they are also *bona fide* interferon response genes. We observed that IFN α -induced clusters were generally contained within a single TAD (**Figure 3E**), consistent with their potential coregulation. Overall, we suggest that distal regulation by interferon response Epromoters might play an important role in the coordinated response to IFN α -mediated signaling.

A general question about the Epromoter function is related to the mechanistic bases leading to their enhancer activity (Andersson and Sandelin, 2020; Catarino et al., 2017). In a previous study, we showed that Epromoters bind a higher number of TFs and harbor a more complex combination of TF binding sites as compared with classical promoters (Dao et al., 2017), suggesting that one of the potential mechanisms mediating enhancer function might be the efficient recruitment of key TFs. Here, we made the striking observation that essential type I interferon-response factors (STAT1/2, IRF1/9) are preferentially recruited by induced Epromoters, which represents a minority (~8%) of the IFN α -induced promoters in K562 cells. In other words, the vast majority of promoters associated with IFN α -induced genes do not recruit the TFs that are essential for their activation. This is surprising given that it is commonly acknowledged that ISG promoters harbor the ISRE sites and that binding of the ISG3 complex is required for the induction of ISGs (Smale and Natoli, 2014). Moreover, IFN α -induced activation of Epromoters was also associated with a preferential gain of activating histone modifications, H3K4me3 and H3K27ac, suggesting that they represent the main IFN α -response elements.

While IFN α -response can be mediated by distal enhancers, unbiased analysis of ISG3 complex binding demonstrated that for a significant subset of IFN α -induced genes, the closest binding of ISG3 was observed within a promoter of another gene. Indeed, we were able to identify several IFN α -response clusters that are associated with at least one Epromoter. Visual inspection of the epigenetic and TF binding profiles at typical type I IFN response clusters, such as *IFIT* and *OAS*, suggests that no potential distal “intergenic” enhancers are found within the TADs containing

these clusters (**Figure S5**). Experimental validation by CRISPR/Cas9 genome editing at three independent clusters demonstrated that the Epromoters were required for accurate induction of the ISGs within the same cluster.

What can then explain the favored recruitment of STAT-IRF complexes to the induced Epromoters? Although promoters of induced genes are generally enriched in ISRE, consistent with previous results (Smale and Natoli, 2014), the induced Epromoters were found to contain a higher number of ISRE binding sites. This can lead to increased efficiency of TF recruitment and/or stabilization required for enhancer function. Features defining the enhancer versus promoter activity of regulatory elements are a fundamental question in the gene regulation field and a focus of extensive research (Catarino et al., 2017; Core et al., 2014; Field and Adelman, 2020; Henriques et al., 2018; Mikhaylichenko et al., 2018; Nguyen et al., 2016b; Rennie et al., 2018). Our study suggests that it is both the higher density and the better quality of key TF binding sites that allow the efficient activation of the Epromoters and mediate its enhancer function. Indeed, mutation of one out of two ISRE in the *OAS3* Epromoter resulted in a more marked effect on its enhancer as compared to its promoter activity. Overall, our results show that for a subset of IFN α -regulated clusters, an Epromoter is required for the efficient recruitment of the key TFs and coordinated induction of the neighbor ISGs.

The use of high-throughput reporter assays provides the ability to test thousands of sequences in a certain nuclear environment, and pinpoint candidates with enhancer/promoter activity with or without any specific cellular stimulation. However, these activities do not necessarily reflect the actual function of any given sequence in their natural context, which might be influenced by several additional factors, including chromatin composition, histone modifications, DNA modifications, and the presence of other cis-regulatory elements (Santiago-Algarra et al., 2017). As highlighted by the results obtained with the *OAS1-3* cluster, it is the preferential *in vivo* binding of the IFN α -response factors to one of the promoters of the co-regulated genes, within a given cluster, that better predict the enhancer activity in the endogenous locus. Indeed, while all three promoters of the *OAS1-3* cluster displayed enhancer activity when assessed by CapSTARR-seq, only the *OAS3* promoter, which efficiently recruits the STAT1/2 and IRF1/9 factors, was required for the proper induction of the other *OAS* genes. Based on the analyses of RNA-seq and ChIP-seq of STAT2, IRF1, and IRF8 in primary mouse macrophages stimulated by LPS, we identified several LPS-response clusters potentially regulated by Epromoters, including one also identified in K562 cells after IFN α stimulation. Experimental validation of one of these loci (*IL15RA-IL2RA*) demonstrated the predictive value of this approach and suggests that the involvement of inducible Epromoters might also apply to other types of inflammatory responses. This is also consistent with previous findings that, promoters highly induced during the immune challenge of macrophages are characterized by enhancer-like features (Henriques et al., 2018; Scruggs et al., 2015).

Aside from our current evidence for the role of Epromoters in the interferon and immune response, it is likely that this type of regulatory elements could be more generally involved in the rapid response of genes to cellular stress. Noteworthy, many of the early characterized enhancers are located close to, or overlapping with, the promoter region of inducible genes, such

as metallothioneins, histones of early cleavage stages, viral immediate-early genes (from some papovaviruses, cytomegaloviruses and retroviruses), heat-shock genes and the antiviral interferon genes (Medina-Rivera et al., 2018; Schaffner, 2015). A common characteristic of most of the aforementioned promoters is that they are associated with inducible genes that have to quickly respond to environmental stress. We hypothesize that, within clusters of co-regulated genes, enhancer-like promoters play an essential role in the coordination of rapid gene induction during the inflammatory response, and likely upon other cellular responses to intra- and extra-cellular stress signaling. This idea fits well with the notion of inducible transcription factories defining discrete membrane-less sub-nuclear compartments containing a high concentration of RNA polymerase and key TFs and where efficient transcription can be triggered (Cook and Marenduzzo, 2018; Feuerborn and Cook, 2015). The enhancer-like promoters could either facilitate the assembly or maintenance of the transcription factories by tightening promoter-promoter interactions or bringing specific transcriptional regulators required for the regulation of the neighbor genes. This would be particularly relevant in the case of rapid and coordinated regulation of gene expression in response to environmental or intrinsic cellular stimuli. Reminiscent of this concept, a similar model has been reported for NFkB-regulated genes in response to TNF alpha stimulation. Experimental removal of a gene from the NFkB-dependent multi-gene complex was shown to directly affect the transcription of its interacting genes, suggesting that co-association of co-regulated genes might contribute to a hierarchy of gene expression control (Fanucchi et al., 2013). Our work provides a framework for future studies aiming to address the contribution of enhancer-like promoters in the formation or stabilization of transcription factories in response to different signaling. In particular, the development of a bioinformatic pipeline based on the analyses of RNA-seq and ChIP-seq data in different stimulatory conditions might allow us to identify clusters of stimulus-regulated genes where only one promoter recruits the key TFs and potentially plays enhancer-like functions. Our work has general implications for the understanding of genome organization and gene regulation in normal and pathological contexts. First, the coordinated regulation by Epromoters has physiological relevance for the co-regulation of inducible genes as a single regulatory element might ensure the synchronized gene expression and fine-tune the response to extracellular signaling. For instance, in the case of the interferon response, complexity can arise from differences in the endogenous levels of signaling pathways components and IFN properties (Schneider et al., 2014). These differences may be integrated by Epromoters in order to coordinate the variations in the nature and number of genes that are transcriptionally up and down-regulated and as a result, lead to distinct biological outcomes.

Secondly, as Epromoters potentially regulate several genes at the same time and have the ability to efficiently recruit essential TFs, mutations in these regulatory elements are expected to have a pleiotropic impact in disease, namely inflammatory diseases. Indeed, several examples point toward the relation between disease-associated variants and disrupted Epromoters (Medina-Rivera et al., 2018). Previous studies suggested that Single Nucleotide Polymorphisms (SNPs) affecting distal gene expression are significantly enriched within Epromoters (Dao et al., 2017; Jung et al., 2019; Li et al., 2019b; Mitchelmore et al., 2020; Wang et al., 2018b) and that human genetic variation within Epromoters influences distal gene expression (Dao et al., 2017; Sams et

al., 2016). Based on our current results, it is expected that Epromoters might have a role in the etiology of inflammatory diseases. Inflammation is viewed as the driving factor in many diseases, including atherosclerosis, cancer, autoimmunity, and infections, and it is a major contributor to age-related conditions (Netea et al., 2017). Although IFN signatures are induced by environmental cues, their amplitudes, time courses, and patterns of gene expression are modulated by genetic factors. Allelic variants that regulate gene expression have been associated with complex multigenic autoimmune diseases as well as monogenic interferonopathy, characterized by a type I IFN signature (Barrat et al., 2019). We forecast that dissecting the consequences of the genetic contributions to the risk of developing a systemic autoimmune disease can be more complex than previously anticipated, given that a variant associated with an Epromoter can have a complex contribution to the interferon response by influencing the expression of different genes involved in the interferon signalling pathway. For instance, a recent study integrating 3D interactions and GWAS found an association between the interferon gene *IFNA2* and a genetic variant lying within the promoter of *CDKN2B* in coronary artery disease (Li et al., 2019b). More generally, it is plausible that genetic variants within Epromoters might differentially impact enhancer versus promoter activity. For instance, two studies working on a promoter-overlapping SNP associated with prostate cancer demonstrated that the alternative variant increases the enhancer activity of the promoter leading to increased expression of two distal transcripts directly involved in cancer progression (Gao et al., 2018; Hua et al., 2018).

Finally, our findings have significant implications for the understanding of the evolution of regulatory elements and interferon response loci. On the one hand, recent works suggested that repurposing of promoters and enhancers facilitated regulatory innovation and the origination of new genes during evolution (Andersson et al., 2015b; Arenas-Mena, 2017; Carelli et al., 2018; Majic and Payne, 2020; Mikhaylichenko et al., 2018; Wu and Sharp, 2013; Xie et al., 2012). On the other hand, gene expansion has been shown to significantly contribute to the evolution of the IFN system and suggested to confer a selective advantage to the host species (Shaw et al., 2017). It is possible that during duplication of ISG, such as in the case of IFIT or OAS genes, ancestral promoters' elements have acquired enhancer functions to coordinate the interferon response of the newly appeared ISG within the cluster. Similarly, the proximity to an Epromoter-associated ISG locus might provide a rapid co-option for neighbor genes to acquire new functions in the interferon response. In this context, it will be interesting to re-examine the dynamic evolution of the interferon responses across different species (Shaw et al., 2017).

Overall, we show that Epromoters work as a hub for recruiting key TFs required for the activation of type I interferon response genes and are required for the accurate induction of interferon-regulated clusters. Our study suggests that Epromoters play an essential role in the coordination of rapid gene induction in the inflammatory response, and likely upon other cellular response to intra- and extra-cellular signals, and to establish connections with other distal response genes.

Materials and methods

Cell culture

Cell line K562 (CCL-243), a chronic myelogenous leukaemia cell line, was obtained from the ATCC (American Type Culture Collection) and maintained in RPMI 1640 media (Thermo Fischer Scientific) supplemented with 10% FBS (Thermo Fischer Scientific) at 37°C and 5% CO₂. Cells were passaged every 3 days at 2×10^5 cells/mL and routinely tested for mycoplasma contamination. Interferon alpha (Sigma Aldrich, SRP4594) was used to induce the IFN type I response. 1×10^6 wild-type and mutant cells were induced using a concentration of 50 ng/mL of IFN α in 2 mL culture for six hours. For each group, three independent stimulations were made.

THP-1, a human acute monocytic leukaemia cell line, was obtained from DSMZ (ACC 16). Cells were grown in RPMI 1640 media (Thermo Fischer Scientific) supplemented with 10% FBS (Thermo Fischer Scientific) at 37°C and 5% CO₂. Cells were passaged every 3 days at 10^6 cells/mL and routinely tested for mycoplasma contamination. To induce macrophage differentiation, THP-1 cells were firstly plated on 6-well plates (2×10^6 cells/well), in media containing 10 ng/ml phorbol 12-myristate 13-acetate (PMA; Sigma-Aldrich, P1585) for 48 hours. After 48 hours of incubation, the PMA containing media was replaced with fresh media, and cells were incubated for an additional 24 hours. THP-1 in-vitro differentiated macrophages were then stimulated by replacing media with 2 mL of fresh media containing 100 ng/mL of LPS (Sigma-Aldrich, L9143) for 4 hours, upon which the extraction of RNA was performed. For each group, three independent stimulations were made.

Human promoter CapSTARR-seq

The CapSTARR-seq promoter library used in this study has been described previously (Dao et al., 2017). We transfected 50 million of K562 cells with 1.25 mg of the CapSTARR-seq promoter library using the Neon Transfection System (Thermo Fisher Scientific; pulse voltage 1,450 V, pulse width 10, pulse number 3) and after 18 hours we treated the cells with IFN α (50 ng/mL) for six hours. After 24 hours we processed the cells according to the Starr-seq protocol (Dao et al., 2015). Transfected and non-transfected (input) libraries were single-end sequenced on the Illumina NextSeq 500 sequencer and mapped to the hg19 reference genome (**Table S4**). Experiments were performed in duplicate. Each promoter was normalized with the ratio of the CapSTARR-seq signal and the input (fold change over input). Promoter regions with enhancer activity were defined using the inflection point of the ranked fold change as a threshold. Promoter regions with an FPKM < 1 in the input library were removed (421 promoters). Finally, we obtained the enhancer activity of 17,941 promoters regulating the transcription of 14,188 genes (**Table S1**). IFN α -induced Epromoters were defined as those Epromoters with a ratio between IFN α stimulated and non-stimulated greater than two and the list is provided in Table S2.

RNA-seq

Poly(A) RNA was isolated from three replicates of K562 cells non-stimulated or stimulated with IFN α for 6 hours and used for RNA-seq library preparation, following the TruSeq RNA Library

Prep Kit v2 (Illumina). Libraries were paired-end sequenced on an Illumina NextSeq 500 sequencer (Table S4). Reads were aligned using STAR aligner (v2.4.2a) with arguments “outFilterMismatchNoverLmax” and “outFilterMultimapNmax” set to 0.08 and 1, respectively. Differential gene expression was performed using DESeq2 (v1.6.3) Bioconductor package (Love et al., 2014) implemented in the R statistical environment, and transcripts associated with promoters analysed by CapSTARR-seq were retrieved. Genes with changes in expression with a log₂ fold change >1 and adjusted P-value <0.001 were considered as significant (Tables S1 and S2). To create bigwig files reads from Watson and Crick strands were selected using SAM tools (v0.1.9) and provided to the bam2wig.py script from the RSeQC program suite (v2.6.4).

ChIP-seq

Chromatin Immunoprecipitation for the IRF9 TF was performed with 5x10⁶ K562 cells stimulated with IFN α for 6 hours and 5 μ g of IRF9 Rabbit mAB (Cell signal, D2T8M). Chromatin Immunoprecipitation for the histone modifications was performed with 5x10⁵ of K562 cells non-stimulated and stimulated with IFN α for 6 hours and 1 μ g of mAB against H3K27ac or H3K4me3 (Diagenode, C15410016, and C15410003, respectively). ChIP-seq libraries were generated with the MicroPlex Library Preparation Kit (Diagenode), according to the manufacturer instructions. The libraries were sequenced in paired-end 50/30nt mode using the NextSeq[®] 500/550 (Illumina), according to manufacturer’s instructions (**Table S4**). Reads were mapped to the hg19 reference genome using standard procedures. ChIP-seq data for the IRF1, STAT1, and STAT2 TFs in K562 cells stimulated with IFN α were obtained from the ENCODE Consortium (**Table S4**). Median average profiles were generated using the SeqPlots tool (Stempor and Ahringer, 2016) for the 2 Kb and 10 Kb regions centered at the TSS of the selected genes for the TFs (IRF9, IRF1, STAT1, and STAT2) and the histone modifications (H3K27ac and H3K4me3), respectively. ChIP-seq profiles were visualized using the IGV genome browser (Thorvaldsdottir et al., 2013).

Functional enrichment

The GO enrichment for biological functions was performed using the webtool GREAT (McLean et al., 2010) using the hg19 genome and default settings. The background regions used were all the captured promoters. Only the top 10 most significant terms are shown. The full list is provided in Table S3.

CRISPR/Cas9 genome editing

For the deletion of EpOAS3, EpISG15, EpIFIT3, EpOAS1, and EpOAS2 Epromoters we used the web tool CRISPRdirect (Naito et al., 2015) to design two guide RNAs flanking the Epromoter region and clone them into the gRNA vector (Addgene #41824), as previously described (Mali et al., 2013). Two primers were designed flanking the target region that allows us to identify the wild-type and the mutant alleles. For the deletion of EpOAS3, EpISG15, and EpIFIT3 the transfection was made with 1x10⁶ K562 cells mixed with 2 μ g of hCas9 vector (Addgene #41815) and 1 μ g of each gRNA vector using the Neon Transfection System (Thermo Fisher Scientific; pulse voltage 1,450 V, pulse width 10, pulse number 3) and cultured in 5 mL. For the deletion of EpOAS2 and EpOAS1 1x10⁶ K562 cells were transfected with 0.5 μ g of each gRNA and

1 µg of Cas9 (Addgene, 41815) in 20 µl transfection solution using the 4D-Nucleofactor X Unit (Lonza), P1 Primary Cell 4D-Nucleofactor X kit S, program FF-120. For the deletion of the *IL15RA* promoter in THP-1 cells, gRNAs were designed using CrispRGold (Chu et al., 2016). We used the Alt-R CRISPR-Cas9 System from IDT (Integrated DNA Technology) where Ribonucleoprotein (RNP) complexes were assembled *in-vitro* and then transfected by electroporation to the cells, accordingly to the manufactured instructions. Three days after transfection, cells were cultured in 3x96-well plates at limit dilution (0.5 cells/100 µL/well). After 2-4 weeks the clones were screened for homologous deletion using the kit Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific, F170L). Clones with homologous deletion were those showing a mutant band of the expected size and no wild-type band. Primers and gRNAs are listed in Table S7.

Gene expression analysis

RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) according to the manufacture instruction. One µg of total RNA was reverse transcribed using the SuperScript II (Thermo Fisher Scientific). qPCR reactions were made using the SYBR green Master Mix (Thermo Fisher Scientific) on the QuantStudio 6 instrument (Thermo Fisher Scientific). 1:10 dilution of cDNA was used for the qPCR. The relative expression was analysed using the relative standard curve method and the *GAPDH* gene was used for normalizing samples. For each group, three independent RNA and cDNA preparations were made. For the analyses of *IL15RA* and *IL2RA* genes in the THP-1 cell line, reverse transcription of 2,5 µg of RNA was done using Master Mix SuperScript™ VILO™ (Thermo Fisher Scientific, 11755250). 1:50 dilution of cDNA was used for the qPCR. qPCR reactions were made using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, A25742) and the measurement was made using the Applied Biosystems QuantStudio 6 Flex Real-Time PCR System. The relative expression was calculated using the relative standard curve method from the mean of two technical replicates values according to the *GAPDH* gene expression. The mean value and standard deviations were calculated between the relative expression values of the three biological replicates, and values were afterward normalized by the value of the unstimulated WT. Primers used are listed in Table S3.

Luciferase reporter assay

The promoter of human *OAS3* (500bp, Chr12:113375900-113376399) was cloned into the pGL3-Basic vector (Promega, E1751) upstream the luciferase gene at the BglII and HindIII restriction sites (pOAS3-luc), and into the pGL3-Promoter vector (Promega, E176; containing the SV40 promoter) downstream the luciferase gene at the BamHI and SalI restriction sites (pSV40-luc-pOAS3). Site-specific mutagenesis of the pOAS3 was done using the Q5 site-directed Mutagenesis Kit (NEB, E0554S) using a set of primers listed in Table S3. For cell transfection, 1×10^6 K562 cells were mixed with 1 µg of each construct and 200 ng of Renilla vector using the Neon Transfection System (Thermo Fisher Scientific; pulse voltage 1,450 V, pulse width 10, pulse number 3) and cultured in 2 mL in 12-well plates. After 18 hours, half of the cell population was treated with 100 ng of IFN α (Sigma Aldrich, 50ng/mL) for 6 hours. Data were normalized to Renilla values and represented as the fold change of relative light units over the

wild-type pOAS3-luc or pSV40-luc-pOAS3 vector from non-stimulated K652 cells. Experiments were performed in triplicate.

Genomic analyses of interferon-dependent enhancers in HeLa cells

Interferon-dependent and independent enhancers in HeLa-S3 cells were retrieved from the Muerdter et al. study (Muerdter et al., 2018). The dataset comprises a genome-wide STARR-seq enhancer screening in HeLa-S3 cells treated or not with a combination of two inhibitors of type I interferon response. The annotation of enhancer regions with respect to the closest RefSeq TSS was performed using HOMER's annotatePeaks.pl tool (Heinz et al., 2010). Based on the distance of active enhancers from the TSS, the genomic density distribution of active enhancers in inhibitor-treated and untreated conditions was calculated. Kolmogorov-Smirnov (KS) test was performed to compare the genomic distribution of active enhancers in the two conditions. Based on the distance to the nearest TSS, the enhancer regions were further divided into two categories: (i) TSS-proximal (≤ 1 kb from the closest TSS) and (ii) TSS-distal (> 1 kb from the closest TSS) enhancer regions. Based on the distance of active enhancers from the TSS, the genomic density distribution of active enhancers in inhibitor-treated and non-treated conditions was calculated. Kolmogorov-Smirnov (KS) test was performed to compare the distribution of active enhancers in the different conditions.

Motif enrichment analyses

Known motif enrichment and de novo motif discovery analyses for TFs binding sites were performed using the findMotifsGenome.pl program in HOMER suit (Heinz et al., 2010) using the default motif lengths and 200 nt long sequences centered on each enhancer region as input. The ten enriched known motifs with the lowest adjusted P-value from each condition and dataset were extracted. For Muerdter et al. dataset, all proximal and distal enhancers for each condition was used as a background. A heatmap was plotted using the $-\log_{10}$ (adjusted P-value). Lower P-value corresponds to higher enrichment in dark color. For motif enrichment in K562 promoters, all promoters analyzed by CapSTARR-seq was used as a background. The top de novo motif with the lowest P-value from each promoter category was demarcated.

ISRE Binding Across the promoter regions

The identified promoter regions were analyzed with respect to the enrichment of the interferon-stimulated response element (ISRE) and TFs individually. The binding site's information of TFs, IRF family (IRF1-9), and STAT1/2 of the human genome (hg19) were extracted from the latest version of the JASPAR database (Fornes et al., 2020). The individual TFs binding regions were generated by bedtools merge (at least 1 base pair overlap), whereas the general ISRE was generated using the bedtools merge in all the IRF family of TFs and STAT1/STAT2. The TF binding occurrences in the different categories of promoters were computed by counting intersects (bedtool intersect) of individual TFs binding regions and ISRE across corresponding promoter regions.

Positional distribution of IFNa-specific TFs binding sites

Overlapping binding sites of STAT1, STAT2, and IRF9 in IFNa-stimulated K562 cells were defined using the 'intersect' parameter from BEDTools v2.28.0 (Quinlan, 2014), based on ChIP-seq data. For each induced gene, we identified the closest merged peak. The location of the closest merged peak was categorized into three groups: i) same promoter, located within ± 1 kb from the TSS of the induced gene; ii) intergenic, located at >1 kb from the TSS of any RefSeq gene and; iii) other promoter, located ± 1 kb from the TSS of any other RefSeq gene. Statistical significance of the positional distribution was assessed using the OLOGRAM tool (Ferre et al., 2019) from the Pygtftk package (Lopez et al., 2019).

Proximity of IFNa-induced genes in the human genome

To find the distribution of induced genes in the genome, the same number of IFNa-stimulated genes was extracted randomly from the human genome (hg19). The randomly selected genes were further clustered based on distance (≤ 100 kb) between the TSS of two induced genes. Further, the distance density distributions of observed and randomly selected genes were compared using Kolmogorov–Smirnov (KS) test.

Identification of IFNa-dependent clusters

The TSS coordinates of IFNa-induced genes or associated with induced Epromoters in K562 cells (as defined in Table S2) were extracted with reference to the RefSeq annotations. The distance between each pair of TSSs was computed using a Custom Python script. A cluster of IFNa-induced loci was defined if at least two TSS of induced genes or Epromoters was found to be closer than 100 kb. The clusters were then classified in function of the type of genes they contained: “induced gene only”, “induced gene & Epromoter”, “induced Epromoter only”, or whether they contained a constitutive Epromoter (Table S1). To assess whether the clustered genes were located within the same TAD, we used the coordinates of TAD borders identified in K562 cells from highly-resolutive Hi-C data (Rao et al., 2014). Overlapping domains were merged into one single TAD domain like previously described in (Paulsen et al., 2017). The detailed clusters are summarized in Table S5.

Identification of LPS dependent clusters in primary macrophages

LPS-induced genes ($\log_2(\text{FC}) > 1$; P val. < 0.01) and binding sites of interferon associated TFs (IRF1, IRF8, STAT1) were extracted from the Mancino et al. study (Mancino et al., 2015). The dataset comprises gene expression and TF binding sites (mm9) in mouse primary macrophages before and after LPS stimulation for 4h. The TSS coordinates of LPS-induced genes were extracted with reference to the RefSeq annotations. LPS-induced clusters were identified based on distance (≤ 100 Kb) as described for the IFNa-response clusters. The promoters binding with TF(s) were considered only if the TF binding site(s) overlapped at least 1 bp with the promoter region (± 1 Kb from the TSS). We then determined the number of promoters per cluster that bind a given TF. The detailed description of the LPS-induces clusters is described in Table S6.

3D chromatin organization of the *IL15RA/IL2RA* locus

CTCF ChIP-seq and Hi-C data in THP-1+PMA cells from the data of (Phanstiel et al., 2017) were visualized using the New WashU Epigenome Browser (epigenomegateway.wustl.edu) (Li et al., 2019a). TADs called from Hi-C experiments (HindIII) THP-1 cells were taken from (Lin et al., 2018). DNA interactions centered on the *IL15RA* promoter with the associated CHiCAGO score were taken from Promoter Capture Hi-C were taken from (Javierre et al., 2016) and visualized using the New WashU Epigenome Browser.

Statistical analysis

GraphPad 7.0 was used for the statistical analysis of luciferase and gene expression assays. Two-tailed unpaired t-test was used to two groups with equal size and small n. *P* value < 0.05 was considered statistically significant. Mean and standard deviation are shown as error bars.

Acknowledgments

We thank the Transcriptomics and Genomics Marseille-Luminy (TGML) platform for sequencing the CapSTARR-seq and ChIP-seq samples and the Marseille-Luminy cell biology platform for the management of cell culture. TGML is a member of the France Genomique consortium (ANR-10-INBS-0009). Work in the laboratory of S. Spicuglia was supported by recurrent funding from Institut National de la Santé et de la Recherche Médicale and Aix-Marseille University and by specific grants from A*MIDEX (ANR-11-IDEX-0001-02), INCA (PLBIO018-031 INCA_12619), Ligue contre le Cancer (*Equipe Labelisée 2018*), ANR (ANR-18-CE12-0019 and ANR-17-CE12-0035) and Bettencourt Schueller Foundation (*Prix coup d'élan pour la recherche française*). This work was also supported by ECOS/ANUIES/SEP/CONACYT grant M17S02/291235. Work in A. Medina-Rivera's lab was supported by the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica – Universidad Nacional Autónoma de México (PAPIIT-UNAM) grant [IA201119]. David Santiago-Algarra was supported by a PhD fellowship from the Conacyt (Mexico).

Author contributions

SS conceived and supervised the project and secured funding. DS, CS, and S.S. conceptualized and designed the experiments. DS, CS, and LTMD performed the experiments. HS performed the majority of the bioinformatic analyses, with the help of DS and CS. NS and GC contributed to the processing of NGS data. JC, AMR, and LR contributed to motif analyses. DS, CS, and SS wrote the manuscript. All authors contributed to reading, discussion, and commenting on the manuscript.

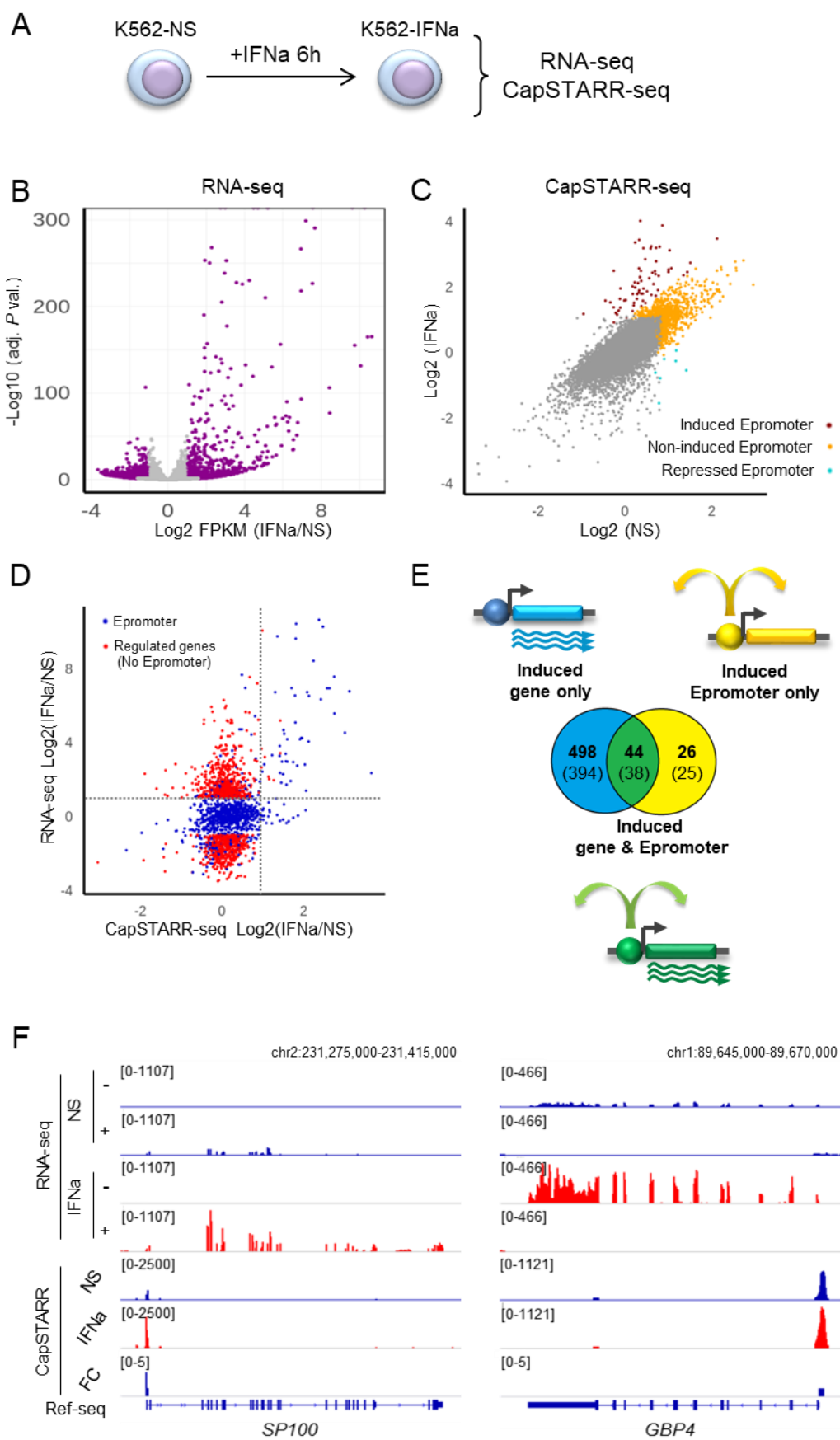


Figure 1. Comparison of RNA-seq and CapSTARR-seq in non-stimulated and IFNa-stimulated K562 cells.

(A) Experimental approach to assess gene expression and enhancer activity of human promoters in non-stimulated (K562-NS) and IFNa-stimulated (K562+IFNa) K562 cells.

(B) Volcano plot of RNA-seq data in K562. The data plotted is the fold change K562+IFNa over K562-NS (log₂ scale) and the adjusted *P* value (-log₁₀ scale). Genes significantly regulated (Adj. *P* val. < 0.001) are highlighted in purple.

(C) Scatter plot showing the CapSTARR-seq signal in K562-NS and K562+IFNa. The data plotted is the CapSTARR-seq signal over the input (log₂ scale). IFNa-induced Epromoters (brown), non-induced Epromoters (yellow) and repressed Epromoters (cyan) are highlighted.

(D) Scatter-plot of RNA-seq data and CapSTARR-seq signal. The data plotted is the RNA-seq and the CapSTARR-seq fold change of K562+IFNa over K562-NS (log₂ scale). Differentially expressed genes that are not associated with Epromoters (red) and Epromoters active in unstimulated or IFNa-stimulated K562 cells (blue), are shown. Dot lines indicated a fold change of 2.

(E) Venn diagram showing the overlap between significantly induced genes and induced Epromoters. The number of promoters (bold) and the associated genes (under brackets) are shown.

(F) Examples of IFNa-induced genes with induced and non-induced Epromoters. The genomic tracks show the RNA-seq and CapSTARR-seq signal in non-stimulated (blue) and IFNa-stimulated (red) K562 cells.

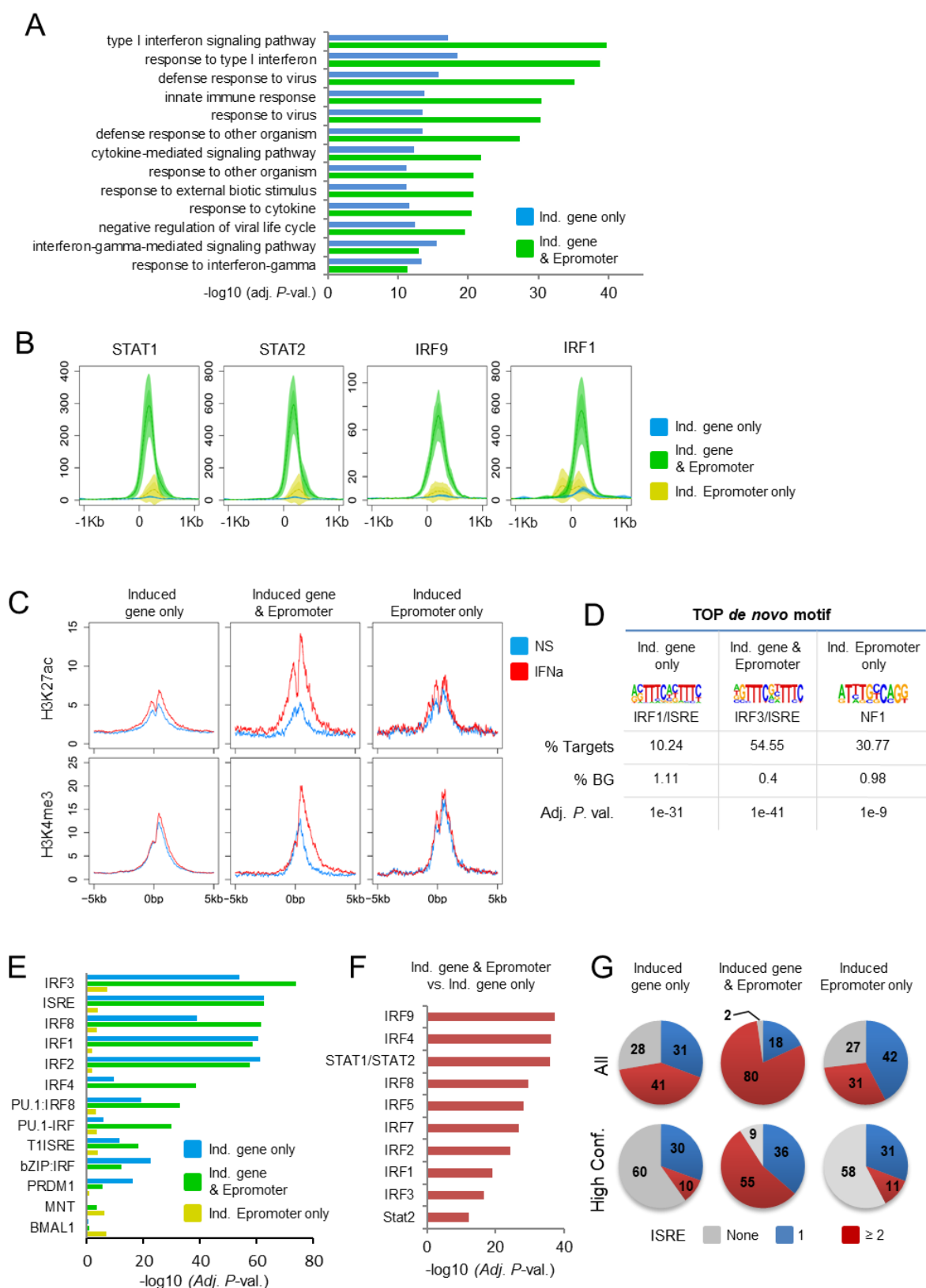


Figure 2. Genomic and epigenomic characteristics of IFNa-induced genes and Epromoters.

(A) Top10 Gene Ontology enrichment for the biological process of genes associated with “induced gene only” (blue) and “induced Epromoter & gene expression” (green). No enrichment was found in the “induced Epromoter only” category. Data plotted is the adjusted P value (-log10 scale) of the enrichment.

(B) Average profiles of ChIP-seq signals for the TFs STAT1, STAT2, IRF9, and IRF1 in K562 cells stimulated with IFNa for 6h. The groups shown were defined in figure 1E. The solid line represents the mean of the signal while the coloured area represents the 95% confidence interval.

(C) Average profiles of ChIP-seq signals for the histone modifications H3K4me3 and H3K27ac in K562 cells non-stimulated and stimulated with IFNa for 6h.

(D) Top de novo motifs found in “induced gene only”, “induced gene & Epromoter” and “induced Epromoter only” using the HOMER tool (Heinz et al., 2010). The enrichment adjusted P value as well as the percentage found in the targets and background (BG) sets are shown.

(E) Motif enrichment analysis in promoter regions from the groups defined in Fig. 1E using the HOMER tool. Only the top 10 motifs for each promoter set is shown. Data shown is the adjusted P value (-log10).

(F) Comparison of the relative number of binding sites per promoter between the “induced Epromoter and induced gene” versus the “induced genes only” sets. Data shown is the adjusted P value (-log10) of the comparison.

(G) Pie-charts showing the percentage of promoters from the indicated data sets that contains either none, one or two or more ISRE binding sites (merged sites for IRF1-9 and STA1-2 motifs) based on the Jaspar 2020 annotation (Fornes et al., 2020) by default (All ISRE sites) or with a P value < 1e-4 (High confidence).

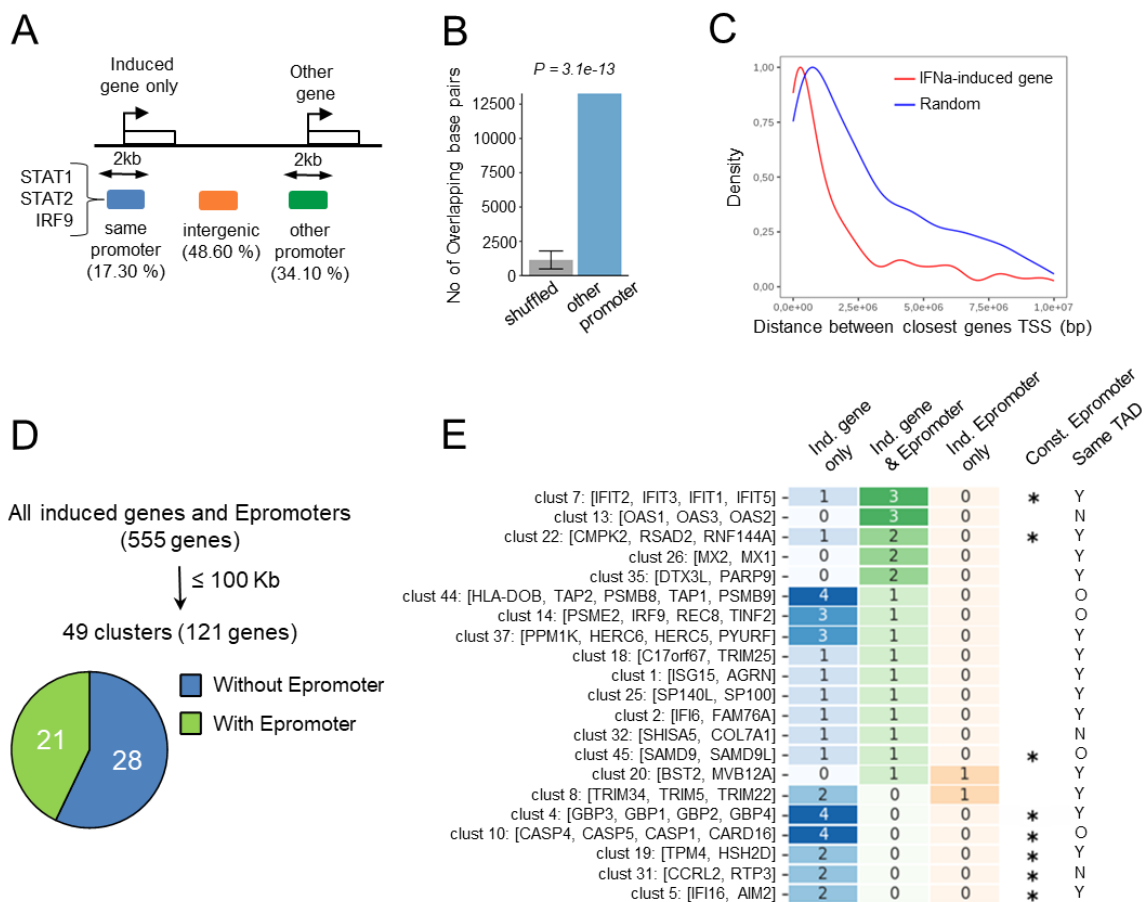


Figure 3. Clusterization of IFNa-induced Epromoters and genes.

(A) Schematic diagram showing the percentage of the closest TF binding with respect to the TSS of “induced gene only” set. TF binding peaks in the same promoter (blue; ≤ 1 Kb from the TSS of “induced gene only”; other promoter (green; ≤ 1 Kb from the TSS of other genes), or intergenic region (orange; >1 Kb from any TSS of a coding gene).

(B) Assessment of the significance of TF binding in other gene promoters (from Figure 3a, green) by the OLOGRAM tool (Ferre et al., 2019). The observed intersection is shown in blue and the random intersections of shuffled regions are shown in grey. Error bars represent the standard deviation of the shuffled distribution. Statistical significance was calculated against a negative binomial model.

(C) The density distribution of the distant between pairs of closest genes using the list of IFNa-induced genes (red) or a set of same number of randomly selected genes (blue) from the human genome.

(D) Summarized diagram of the clusterization pipeline of all induced genes and Epromoter-associated genes from which the distance between the TSS was less than 100 Kb were considered to belong to the same cluster. The pie-chart shows the number of clusters that contains (green) or not (blue) at least one Epromoter. The detailed list of clusters is provided in Table S5.

(E) Gene clusters harbouring at least one Epromoter. Data shown contain the cluster number and the genes within the cluster. The heatmap shows the number of genes in each category. The induced gene (blue), induced gene & Epromoter (green), and induced Epromoter only (orange) groups are illustrated. The number of genes for each category is indicated and their value represented by color intensity. Presence of constitutive Epromoters in the clusters is shown by a star (*). Whether the clustered genes were found within the same TAD (Y), different TADs (N) or outside any TAD (O) is indicated.

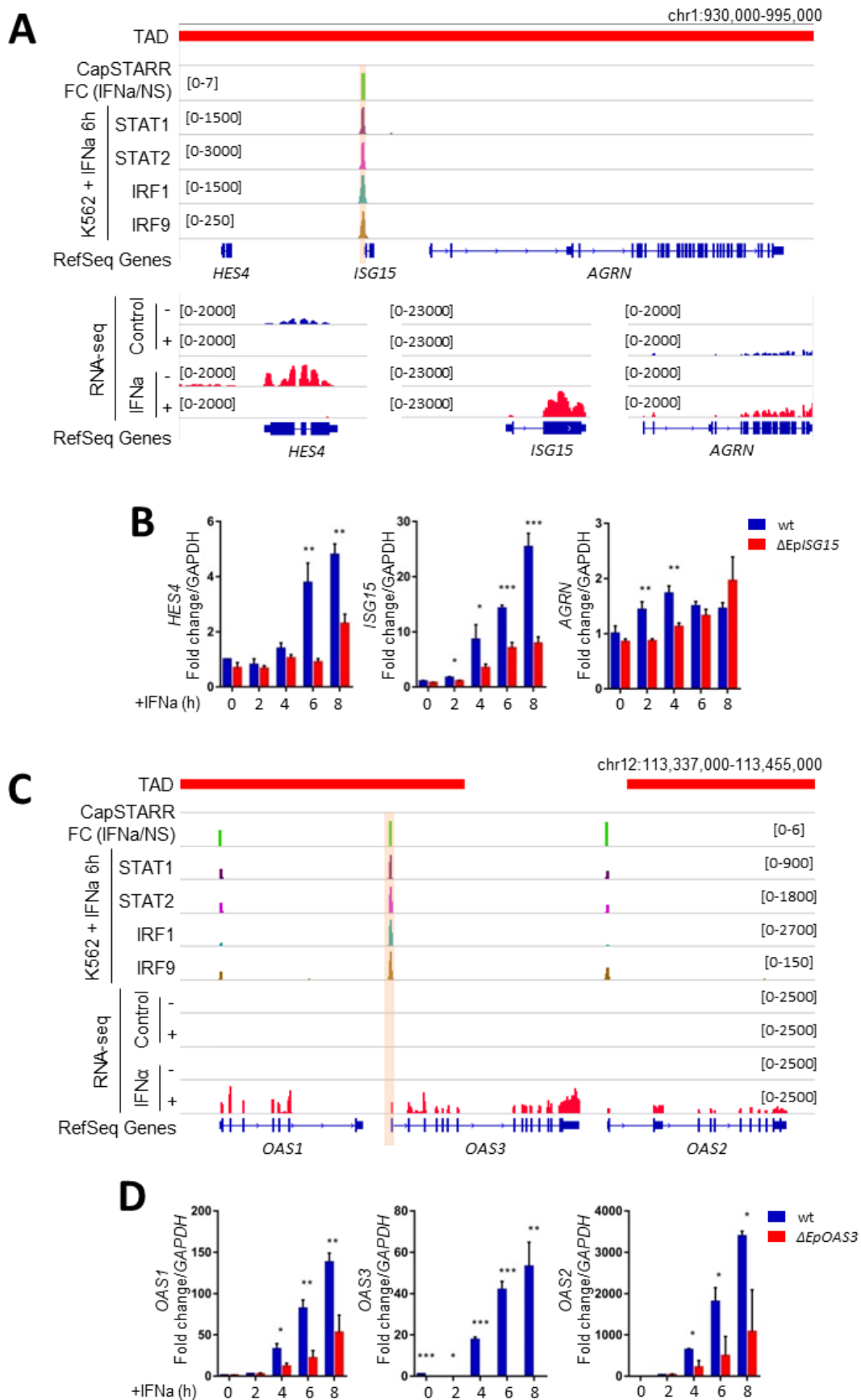


Figure 4. Genomic visualization and kinetic analysis of IFNa-induced genes clustered with induced Epromoters.

(A and C) Genomic tracks centred on the *HES4/ISG15/AGRN* (A) and *OAS1/OAS2/OAS3* (C) loci. Top panels show the TAD, CapSTARR-seq signal fold change (IFNa over non-stimulated) and the ChIP-seq signal for the indicated TFs after IFNa induction. The induced Epromoter of the *ISG15* and *OAS3* genes are highlighted. Bottom panels show the RNA-seq signals of the induced genes.

(B and D) qPCR analysis of gene expression of the indicated genes in K562 wild-type and $\Delta pISG15$ (B) or $\Delta pOAS3$ (D) mutants in non-stimulated conditions or after different times of IFNa stimulation. Values represent the relative expression levels as compared to the unstimulated conditions and normalised by the *GAPDH* housekeeping gene. Error bars indicated s.d. of three independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-sided Student's *t* test.

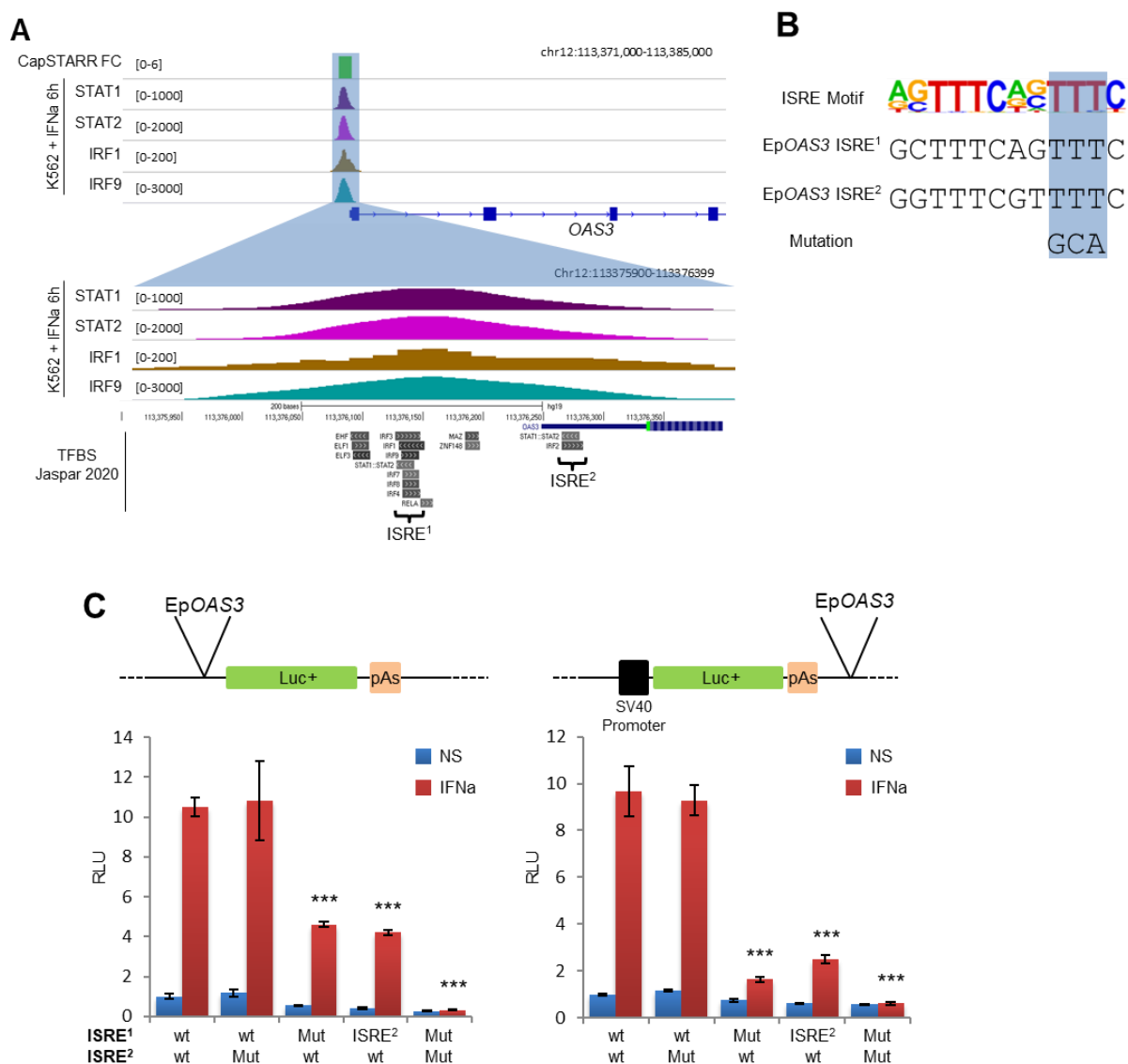


Figure 5. Luciferase assay of promoter and enhancer activity in wild-type and *OAS3* promoter mutants.

(A) Genome browser tracks showing the fold change of CapSTARR-seq activity and the ChIP-seq signal of the indicated TFs surrounding the *OAS3* Epromoter. Lower panel shows the location of significant TF binding sites (Jaspar 2020) and the relative position of the two ISRE motifs in comparison with the ChIP-seq peaks.

(B) Consensus sequence of the ISRE motif as well as the genomic sequence of the two identified ISREs in EpOAS3, as well as the mutation made for the luciferase test.

(C) Luciferase assays to quantify the promoter (left) and the enhancer (right) activity of the wild-type *OAS3* Epromoter or with the indicated mutations in unstimulated or IFNa-stimulated K562 cells. Statistical significance between the mutants *versus* the wild type in stimulation conditions is shown. No significant difference was found in non-stimulated conditions. Error bars represent the s.d. of three independent replicates. *** $P < 0.001$, two-sided Student's *t* test.

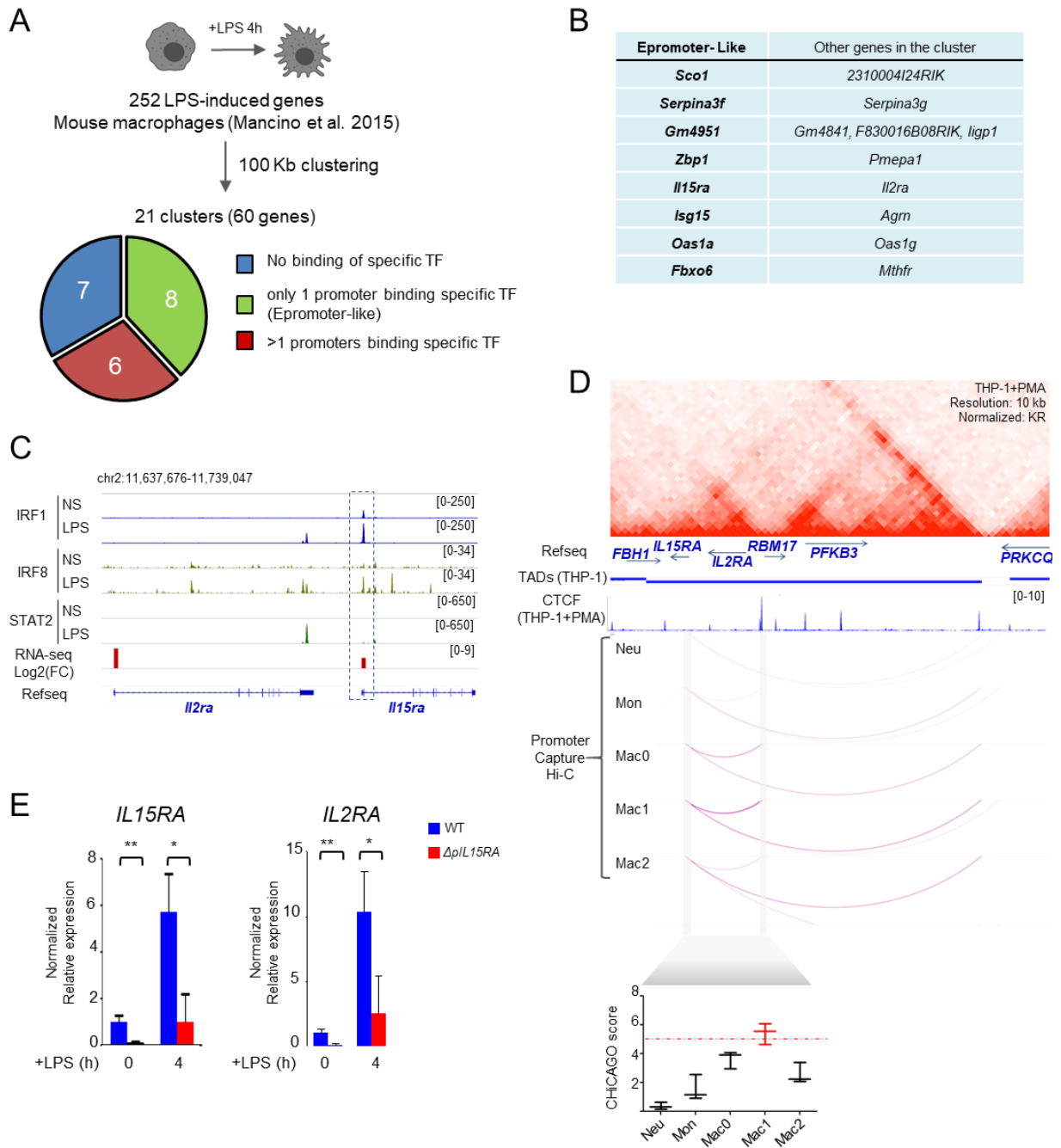


Figure 6. Epromoter-like clusters identification and validation of enhancer activity in LPS stimulated macrophages

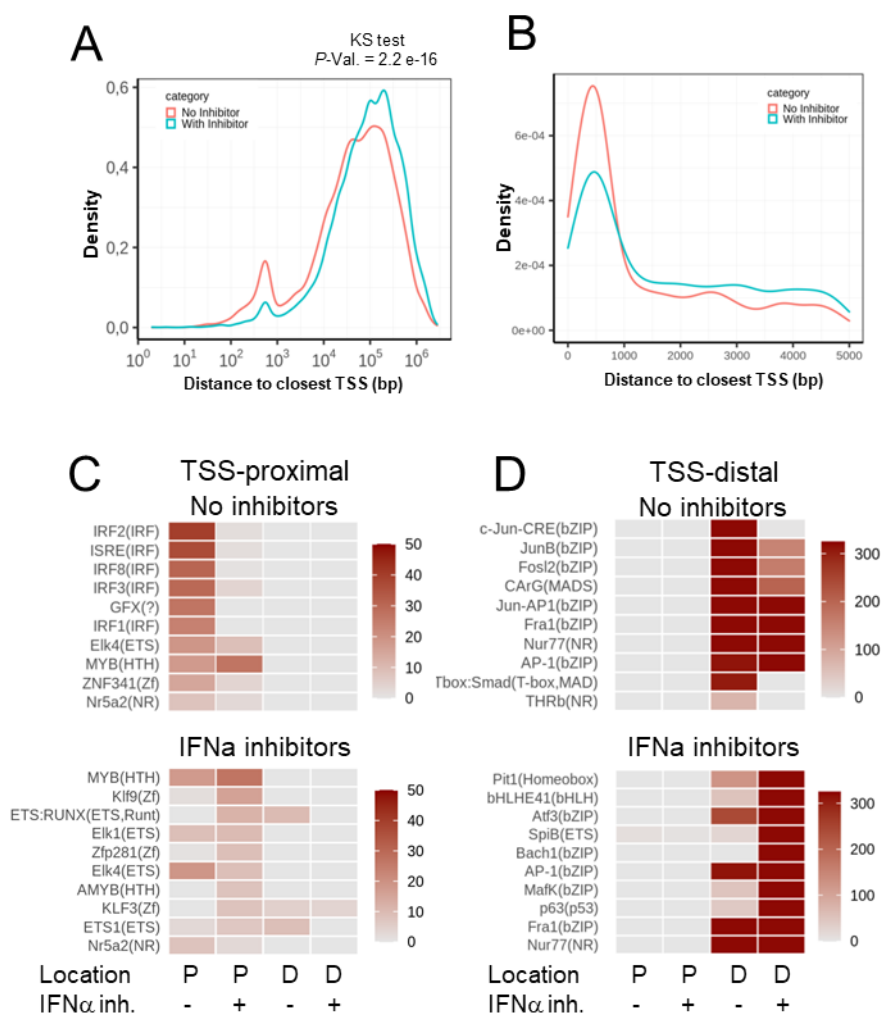
(A) Schematic description of the clusterization of the LPS induced genes in the mouse macrophages and pie-chart with the proportion of promoters that is bound by none, 1 or more than one of the TFs IRF1, IRF8 and STAT2, using data from Mancino et al., (Mancino et al., 2015). LPS induced genes in the mouse macrophages were clustered together if at least two of their promoters were located within less than 100 kb.

(B) Table showing the list of the 8 LPS-induced clusters with putative Epromoter elements as defined in (A).

(C) Genomic tracks centered on the mouse *Il15ra/Il2ra* cluster and showing the IRF1, IRF8 and STAT2 ChIP-seq signal in unstimulated (NS) and after 4 hours of LPS stimulation (LPS) of mouse macrophages, as well as, the log₂ fold change of the RNA-seq.

(D) Top panel: genomic tracks showing Hi-C triangular matrix and ChIP-seq CTCF signal in *in vitro* differentiated human THP-1 macrophages (Phanstiel et al., 2017). Topological associated domains (TADs) from THP-1 cells (Hong et al., 2020). Middle panel: DNA interactions detected by CHiCAGO at the *IL15RA* from Promoter Capture Hi-C from Javierre et al., (Javierre et al., 2016) in human primary neutrophils (Neu), monocytes (Mon), *in vitro* differentiated macrophages unstimulated (Mac0), stimulated with LPS (Mac1) or with IL-13 (Mac2). Bottom panel: scatter plot showing the mean of the CHiCAGO score of the three interactions between *IL15RA* to the three bins surrounding the CTCF site downstream *IL2RA*. The significant threshold as defined in (Phanstiel et al., 2017) is indicated by a red line.

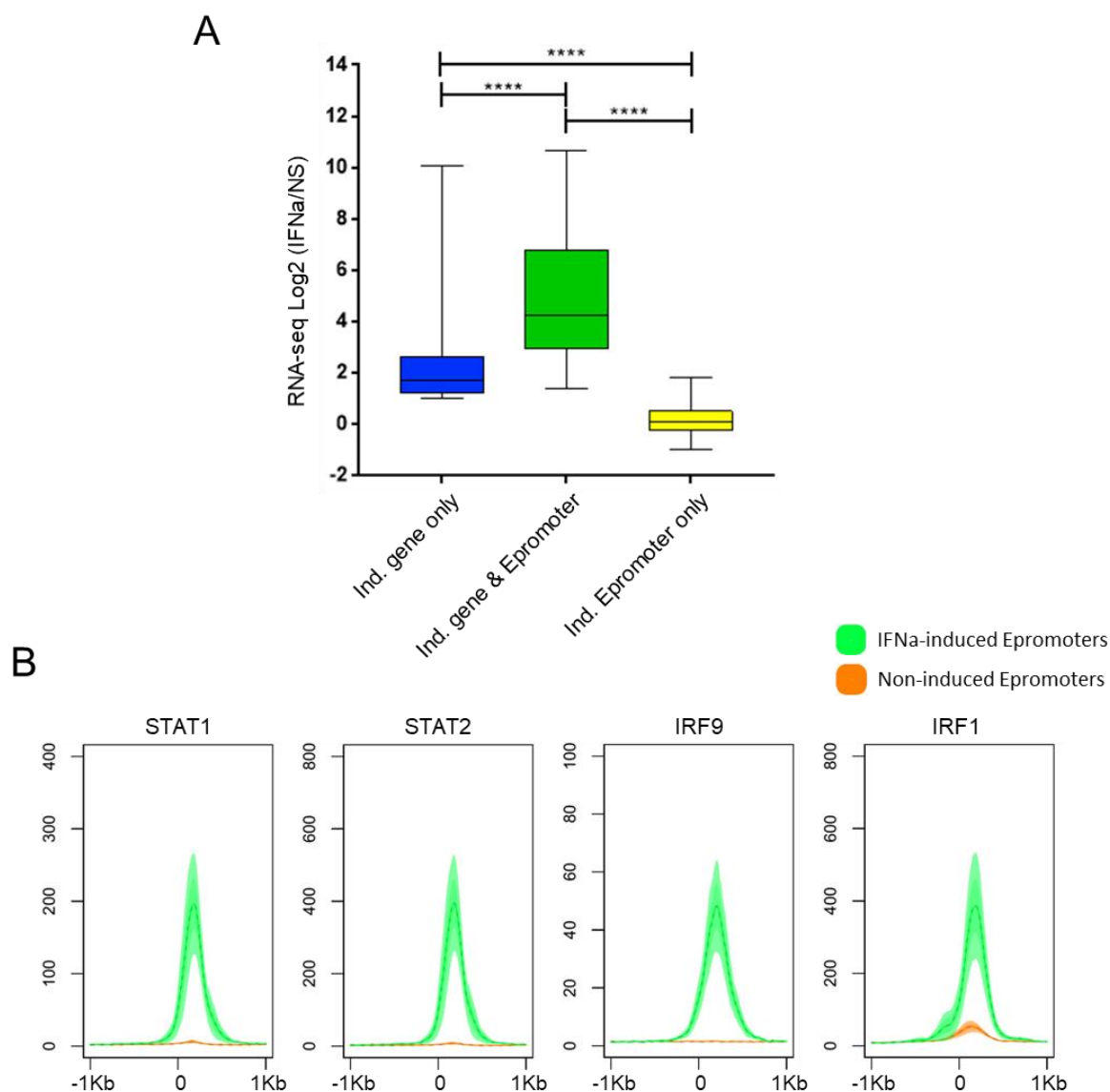
(E) qPCR analysis of gene expression of *IL15RA* and *IL2RA* genes in wild-type and $\Delta pIL15RA$ mutants of *in vitro* differentiated THP-1 macrophages in unstimulated or LPS stimulated conditions. Values represent the relative expression levels as compared to the unstimulated conditions and normalised by the *GAPDH* housekeeping genes. Error bars indicated s.d. of three independent experiments. * $P < 0.05$, ** $P < 0.01$, two-sided Student's *t* test.



Supplemental Figure S1. Distribution and transcription factor motif enrichment of active enhancers in HeLa cells.

(A-B) Distribution of active enhancers in HeLa-S3 cells in function of the distance to the closest TSS with (blue) or without IFN response inhibitor (red), based on whole-genome STARR-seq data (Muerdter et al. 2018). Data plotted is the density of all active enhancers and the distance to the closest TSS (A; log10 scale) and the density of active enhancers located at least than 5 kb to the closest TSS (B). Kolmogorov–Smirnov test was used to compare the distribution of active enhancers in the presence or absence of interferon inhibitors.

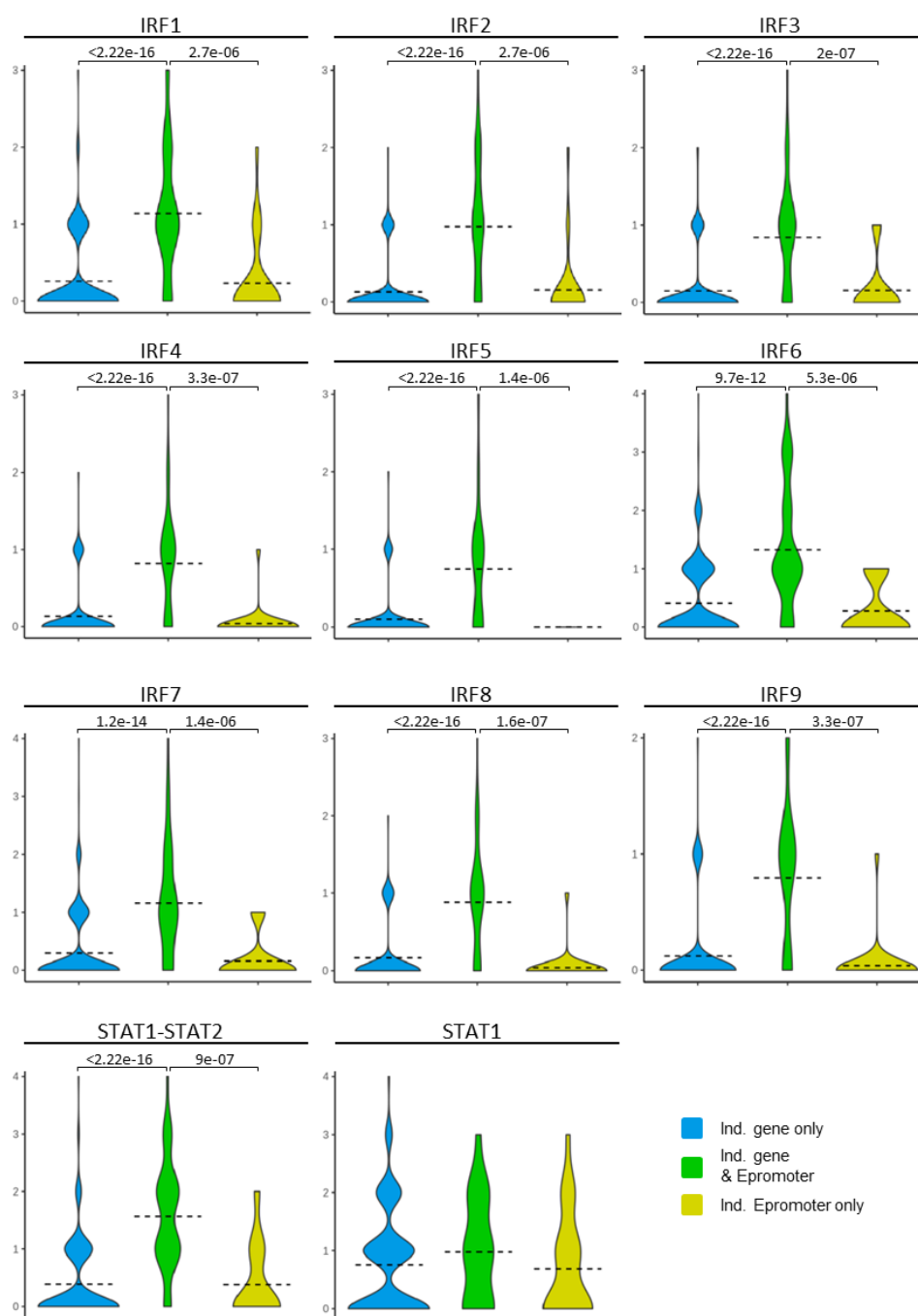
(C-D) Motif enrichment analysis of TSS-proximal (C; ≤ 1 kb from the TSS) and distal (D; > 1 kb from any TSS) enhancers in HeLa-S3 cells in the presence (bottom) or absence (top) of interferon inhibitors, using the HOMER tool. The Top 10 motifs for each condition was retrieved.



Supplemental Figure S2. Interferon response in K562 cells.

(A) IFNa-induced gene expression. Boxplot of RNA-seq signal of the different IFNa-induced sets. Data plotted is the log₂ of the FPKM ratio between K562+IFNa and K562-NS. Kruskal-Wallis test; ****, *P*-value < 0.0001

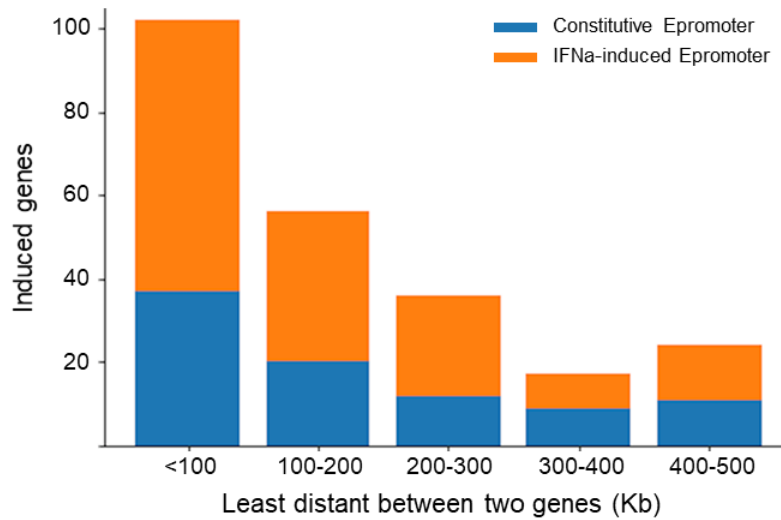
(B) Average profiles of CHIP-seq signals for the transcription factors STAT1, STAT2, IRF9, and IRF1 in K562 cells stimulated or not with IFNa for 6h, for the constitutive and IFNa-induced Epromoters as defined in Table S1. The solid line represents the mean of the signal while the coloured area represents the 95% CI.



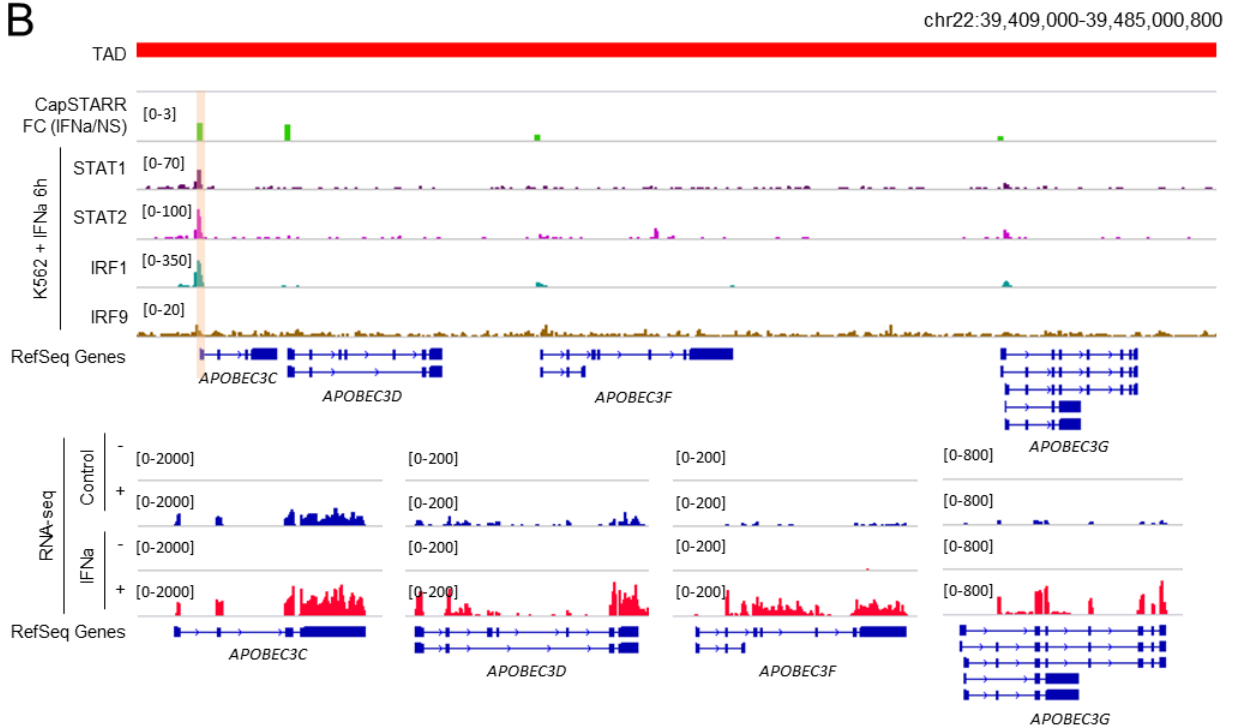
Supplemental Figure S3. Violin plot of the distribution of the binding sites of IRF1-9, and STAT1/STAT2 transcription factors per promoter.

Data shown is the number of the TFs binding sites in the groups defined in Table S2. Dashed lined, mean. Only significant *P* values in the Kruskal-wallis test are shown.

A

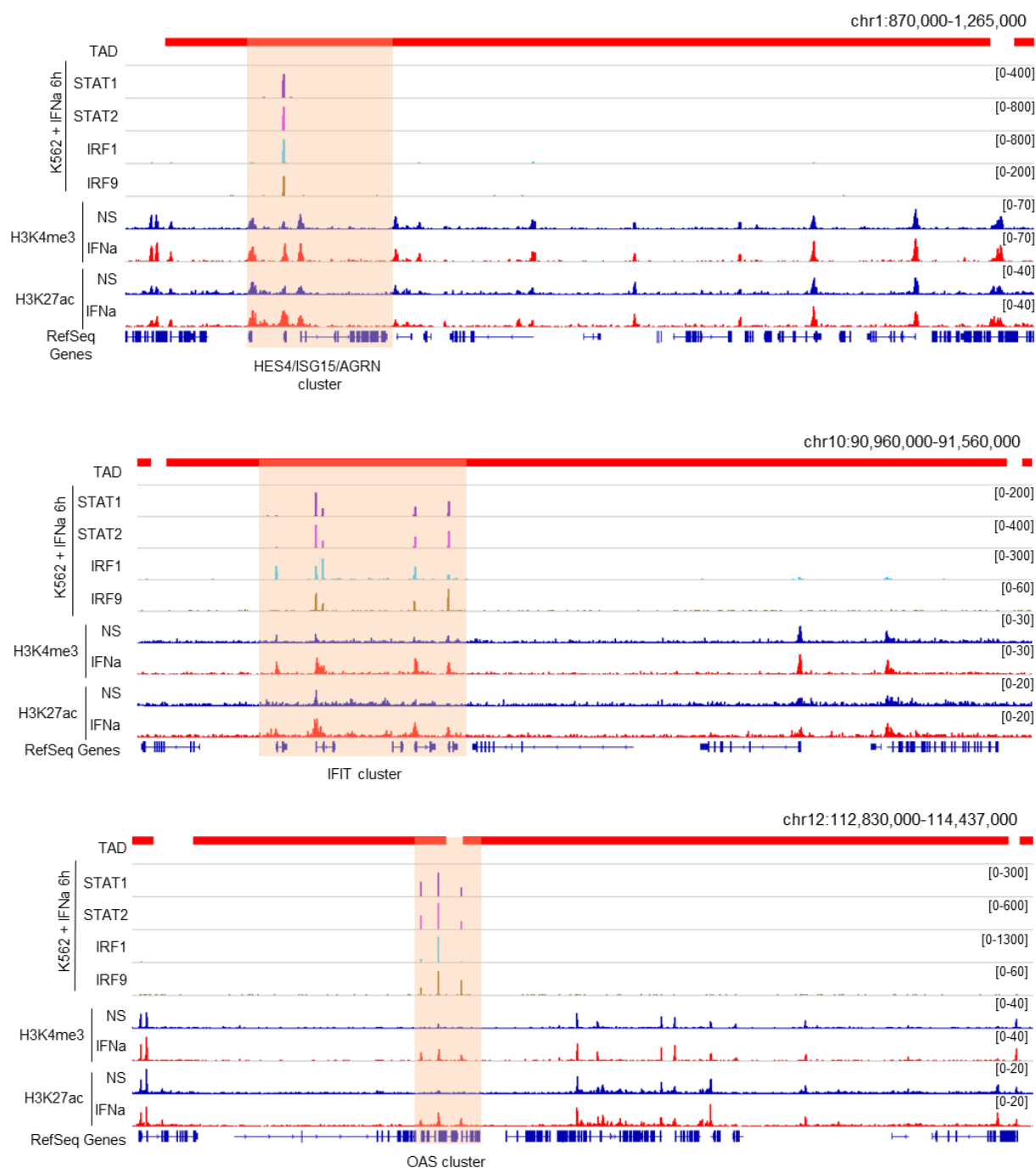


B



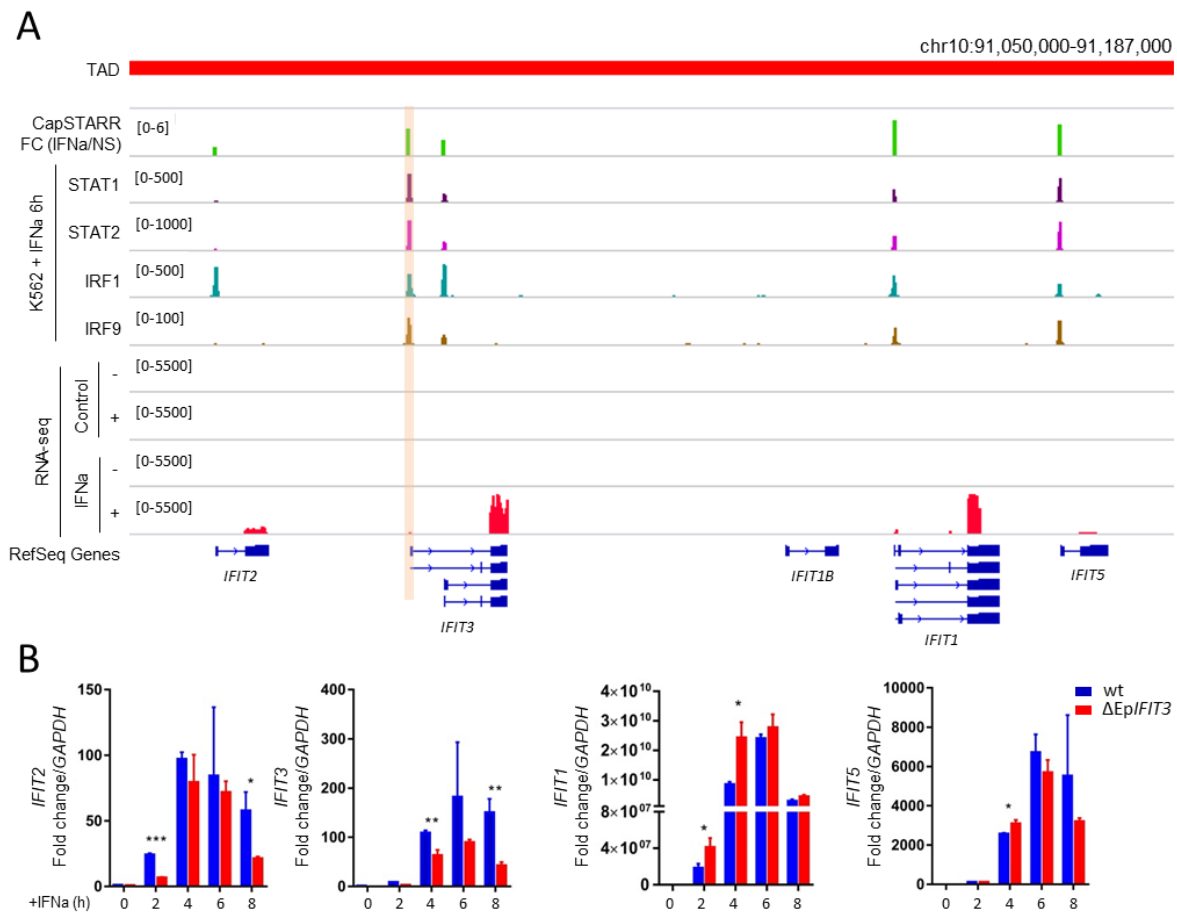
Supplemental Figure S4. Genomic visualization of the IFNa-induced cluster *APOBEC3*

Top panels show the TAD, CapSTARR signal fold change (IFNa over non-stimulated) and the ChIP-seq signal for the indicated transcription factors after IFNa induction. The induced Epromoter of the *APOBEC3C* is highlighted. Bottom panels show the RNA-seq signals of the induced genes.



Supplemental Figure S5. Genomic visualization of TADs harboring IFN α -induced clusters.

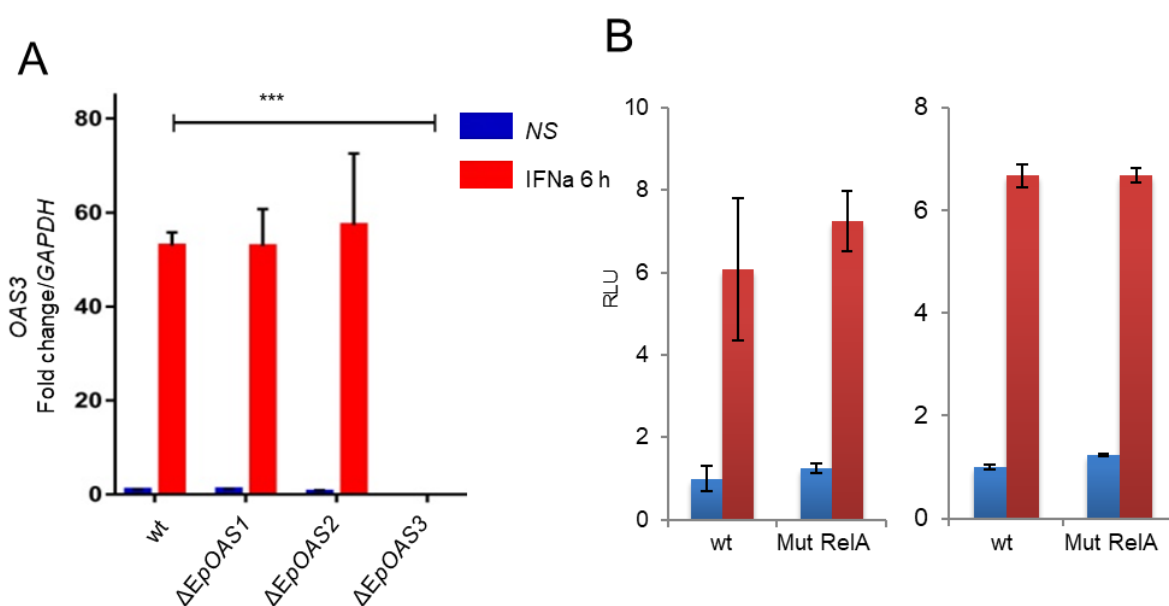
Genomic tracks showing the TAD, the ChIP-seq signal for the indicated transcription factors after IFN α induction and the ChIP-seq signal for the histone modifications in H3K4me3 and H3K27ac in non-stimulated and IFN α -stimulated K562 cells. The cluster of HES4/ISG15/AGRN (upper panel), IFIT genes (middle panel) and OAS genes (lower panel) are highlighted.



Supplemental Figure S6. Genomic visualization and kinetic analysis of the IFNa induced clusters IFIT

(A) Genomic tracks centered on the *IFIT* loci, showing the TAD, CapSTARR signal fold change (IFNa over non-stimulated), the ChIP-seq signal for the indicated transcription factors after IFNa induction and the RNA-seq signal. The deleted Epromoter of the *IFIT3* gene is highlighted. Bottom panels show the RNA-seq signals of the induced genes.

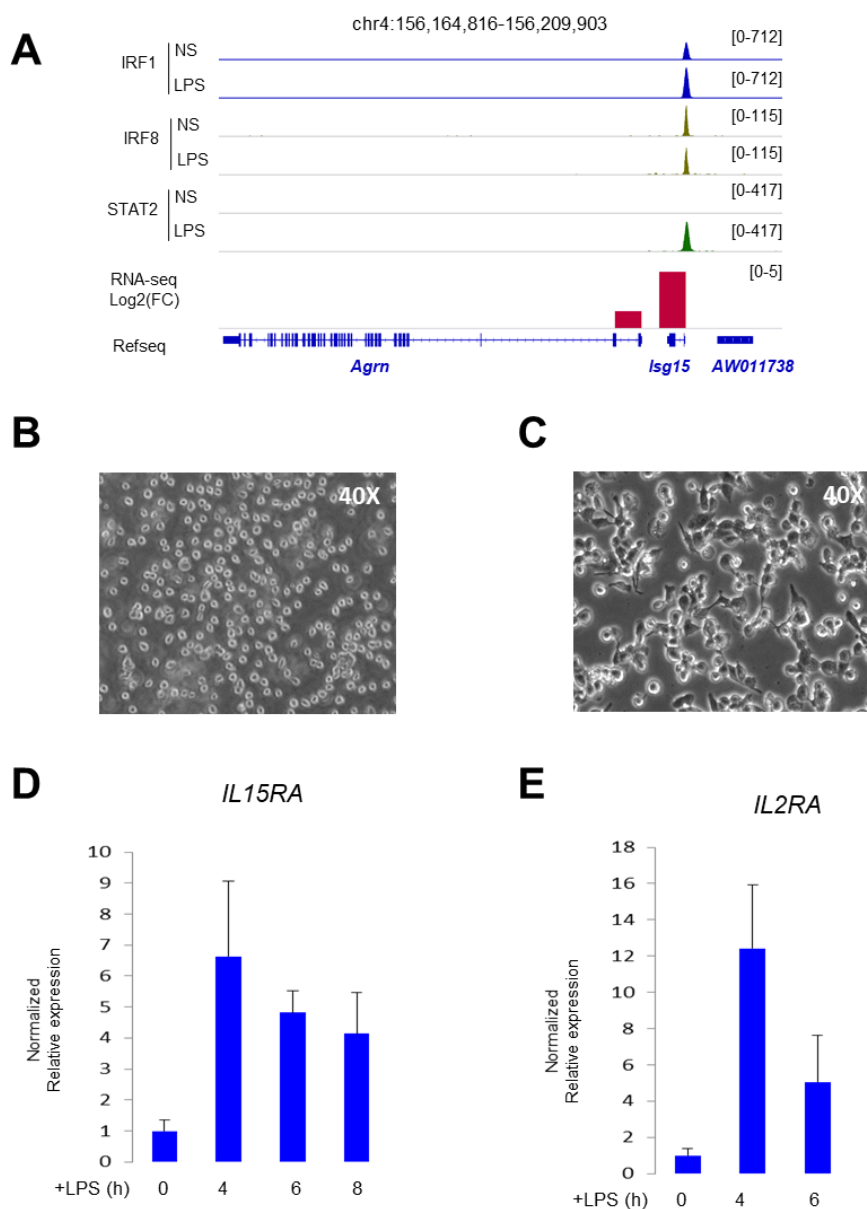
(B) qPCR analysis of gene expression of the indicated genes in K562 wild-type and Δ Ep*IFIT3* mutants in non-stimulated conditions or after different times of IFNa stimulation. Values represent the relative expression levels as compared to the unstimulated conditions and normalized by the *GAPDH* housekeeping gene. Error bars indicated s.d. of three independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-sided Student's t test.



Supplemental Figure S7. Gene expression of *OAS3* in the different mutants of the *OAS* cluster

(A) qPCR analysis of gene expression of *OAS3* in K562 wild-type versus mutant Δ Ep*OAS1*, Δ Ep*OAS2*, and Δ Ep*OAS3*, before and after 6 hours of IFN α stimulation. Error bars indicated s.d. of three independent experiments. ***, $P < 0.001$, two-sided Student's t test.

(B) Luciferase assays to quantify the promoter (left) and the enhancer (right) activity of the wild-type *OAS3* Epromoter or with the RelA mutation in unstimulated or IFN α -stimulated K562 cells. No significant difference was found in between the mutants versus the wild type in stimulated and non-stimulated conditions. Error bars represent the s.d. of three independent replicates. two-sided Student's t test.



Supplemental Figure S8. Macrophage stimulation by LPS.

(A) Genomic tracks at the *Isg15* cluster showing the ChIP-seq signal of the transcription factors IRF1, IRF8 and STAT2, before (NS) and after 4 hours of LPS stimulation (LPS), and the log₂ fold change of the RNA-seq before and after the stimulation in the mouse macrophages (data taken from Mancino et al., 2015). *Isg15* and *Agrn* are significantly induced by LPS, but only the *Isg15* promoter bind the specific transcription factors.

(B-C) Microscopy images showing changes in morphology of the *in vitro* differentiated THP-1 macrophages. (B) THP-1 monocytes cell lines in suspension and (C) *in vitro* differentiated THP-1 macrophages upon 48 hours of treatment with PMA at 10 ng/mL.

(D-E) qPCR analysis of gene expression of (D) *IL15RA* and (E) *IL2RA* in the *in vitro* differentiated THP-1 macrophages after the stimulation with LPS at 0, 4, 6 and 8 hours. The induction of both genes reach the highest level after 4 hours of stimulation.

V. Complementary results

Functional study of the APOBEC3 cluster of IFNa-induced genes

As mentioned in the manuscript, we identify that the apolipoprotein B editing complex family (APOBEC3) locus has three IFNa-induced genes (APOBEC3D, APOBEC3F, and APOBEC3G) and the non-induced Epromoter of APOBEC3C. The APOBEC3 genes are cytidine deaminases important in the response to infection by retroviruses (Stavrou and Ross, 2015). We hypothesized that even if the Epromoter activity of APOBEC3C was not induced by the IFNa stimulation, it could still regulate the expression of the co-expressed genes in the same locus since we did not find any other Epromoter or IFNa-induced gene within the same TAD (Fig. extra 1A). Like in previous examples, the Epromoter APOBEC3C recruit the transcription factors STAT1, STAT2, and IRF1, whereas the IFNa-induced genes APOBEC3D, APOBEC3F, and APOBEC3G display low binding of this factors (Fig. S4). Interestingly, the entire TAD does not recruit IRF9. To test this hypothesis, we deleted the Epromoter of APOBEC3C and analysed the gene expression of the associated genes in different times of IFNa stimulation. Surprisingly, the deletion of the Epromoter of APOBEC3C has no negative effect in the expression of the other genes but an increase of the expression of APOBEC3G and APOBEC3H at six hours post-stimulation, compared with the wild-type (Fig. extra 1B). It is possible that in the absence of the APOBEC3C Epromoter the transcription factors are redistributed to the other promoters of the cluster. Another possibility is the existence of distal regulators within the same TAD that can recruit the key transcription factors.

CTCF binding sites affect the enhancer activity of OAS3 promoter

A key question regarding the Epromoter activity is whether the enhancer activity of Epromoters is affected by the looping and whether this looping is preformed or formed by IFNa-stimulation. In parallel to the study of the 3D organization of the locus (ongoing experiments), we decided to study the effect of deleting the CTCF sites surrounding the OAS locus and assess whether it perturb the activation of the OAS genes. By visual inspection of the OAS cluster, we identified two CTCF binding peaks flanking the cluster in K562 cells (Fig. extra 2A). The corresponding CTCF binding sites were positioned in the opposite direction as proposed for looping formation. This suggests that this CTCF might form a local loop exclusive for the OAS genes. To test this hypothesis, we mutated the two CTCF binding sites flanking the OAS cluster in the same cellular clones and evaluate the expression of the OAS1, OAS2, and OAS3 genes (Fig. extra 2A). We decided to perform this experiment in the semi-haploid cell line HAP1 to make sure that the effect of the mutated allele will not be masked by not mutated alleles (K562 contains 3 copies of chromosome 12). Surprisingly, the mutation of both CTCF binding sites did not affect the expression of any of the OAS genes. One possible explanation is that the mutated CTCF sites do not contribute to the formation of the looping or that CTCF-mediated looping is not required for the Epromoter function.

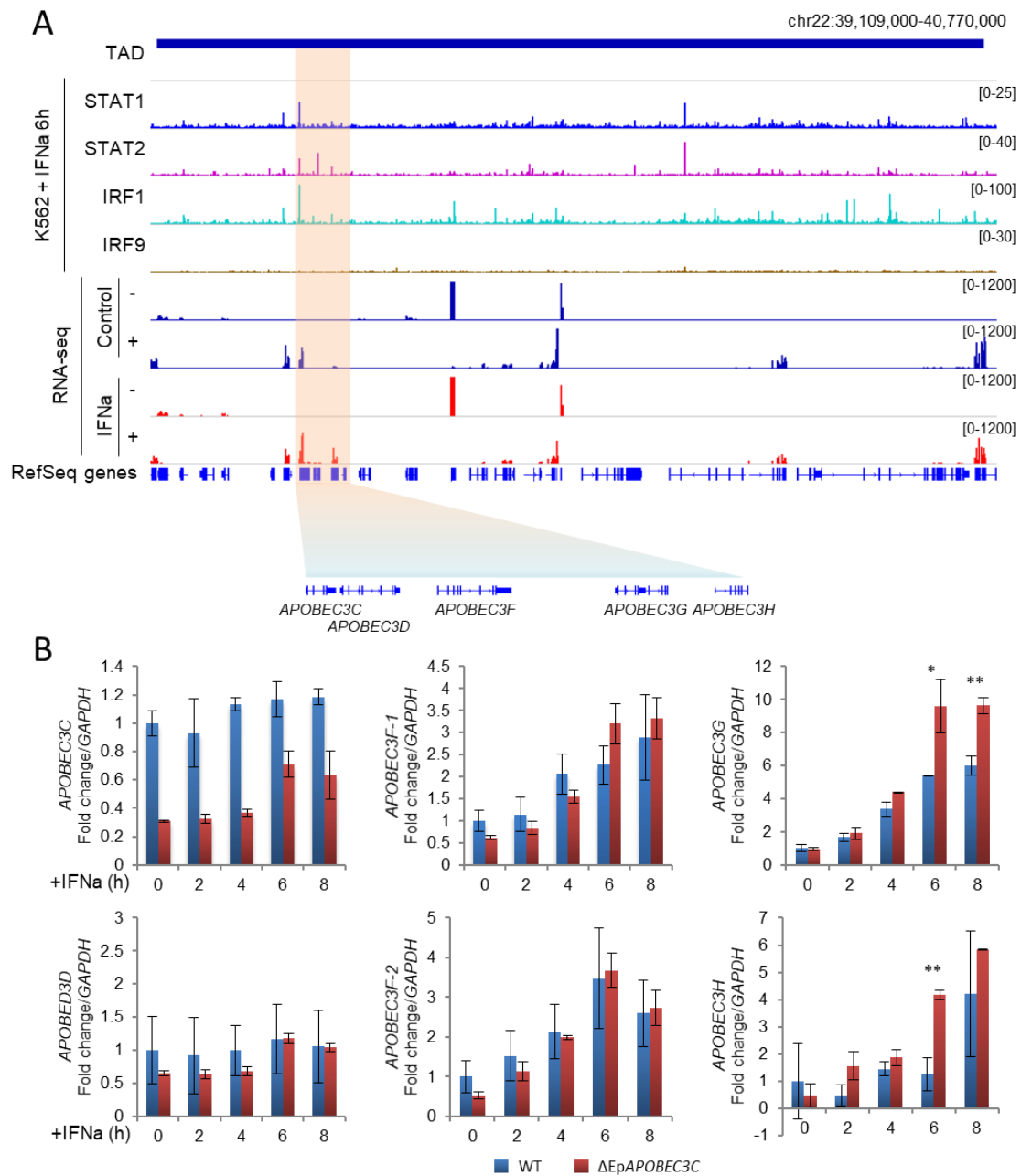


Figure extra 1. Genomic visualization and kinetic analysis of *APOBEC3* locus.

(A) Genomic tracks in the TAD that contains the *APOBEC3* locus. Top panels show the TAD, the ChIP-seq signal for the indicated transcription factors after IFNa induction and the RNA-seq in non-stimulated and IFNa-stimulated K562 cells. The locus of *APOBEC3* genes are highlighted. Bottom panels show the gene distribution of the *APOBEC3* genes.

(B) qPCR analysis of gene expression of the indicated genes in K562 wild-type and Δ p*APOBEC3C* mutants in non-stimulated conditions or after different times of IFNa stimulation. Values represent the relative expression levels as compared to the unstimulated conditions and normalised by the *GAPDH* housekeeping gene. Error bars indicated s.d. of three independent experiments. * $P < 0.05$, ** $P < 0.01$, two-sided Student's *t* test.

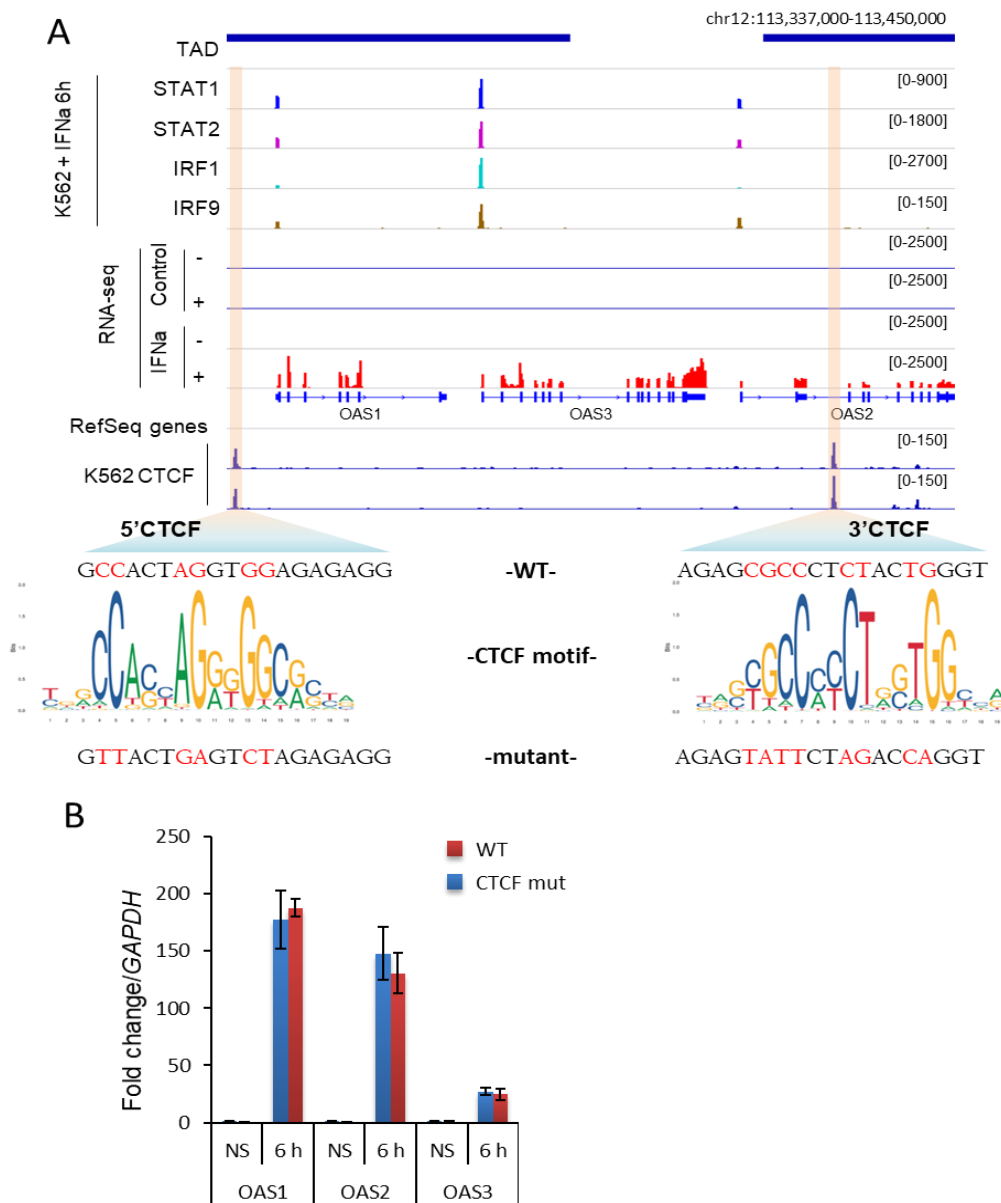


Figure extra 2. Genomic visualization and CTCF mutant analysis in *OAS3* locus.

(A) Genomic tracks in the TAD that contains the *OAS3* locus. Top panels show the TAD, the ChIP-seq signal for the indicated transcription factors after IFNa induction, the RNA-seq in non-stimulated and IFNa-stimulated in K562 cells and the ChIP-seq signal for the CTCF in K562. The CTCF peaks flanking the *OAS3* locus are highlighted. Bottom panels show the wild-type sequence, the CTCF motif and the mutant sequence.

(B) qPCR analysis of gene expression of the indicated genes in Hap1 wild-type and CTCF mutants in non-stimulated conditions or six hours after the IFNa stimulation. Values represent the relative expression levels as compared to the unstimulated conditions and normalised by the *GAPDH* housekeeping gene. Error bars indicated s.d. of three independent experiments. Two-sided Student's *t* test.

Chapter 5. General discussion and perspectives

I. Epromoters activity in the type I interferon response

Following the discovery of the Epromoters as promoter elements able to regulate the expression of proximal and distal genes, we raised the hypothesis that type of regulatory elements might be involved in the rapid activation of gene clusters in response to external stimuli. We choose to use as a model the type I interferon response for two main reasons: 1) previous studies suggested a significant association of Epromoters with interferon-response genes and 2) the type I interferon response is a ubiquitous mechanism of antiviral defence which is highly conserved in most cell types. Thanks to the use of the CapSTARR-seq and the RNA-seq we were able to assess the gene expression and the enhancer activity of the coding-gene promoters during the type I interferon response in K562 cells and found a subset of IFN α -response genes associated with IFN α -induced Epromoters. The IFN α -induced Epromoters preferentially bound to the key transcription factors involved in the type I interferon response pathway (STAT1, STAT2, IRF1, and IRF9). After validating that this IFN α -induced Epromoters regulates the expression of IFN α -induced genes *in vivo*, we concluded that this induced Epromoter serves as a hub to efficiently recruit the transcription factor and efficiently induce the expression of the neighbour genes within the same cluster.

In figure 5.1 we illustrate the general model of action of Epromoter after the stimulation with IFN α .

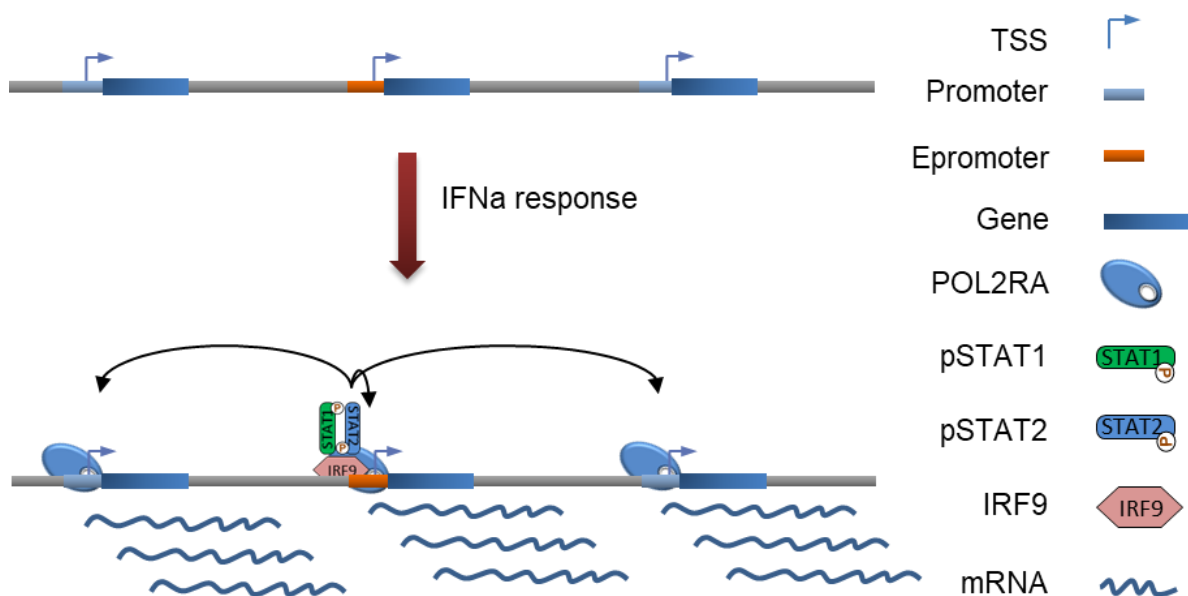


Figure 5.1 General model of Epromoter activity during IFN α response.

II. Epromoters role in the formation of 3D interactions

Nowadays is known that the long-range interactions involving promoters and enhancers that lie within the same TAD often show coordinated activity (Zhan et al., 2017). Our original approach consisted in identify the convergently orientated boundary elements associated with the CCCTC-binding factor (CTCF) flanking the OAS cluster and mutate the two binding sites. By doing so, the expression within the OAS cluster after the IFN α stimulation was not affected. This seems to correlate with some reports suggesting that not all CTCF-bound sites act as boundaries and, importantly, not all TADs are flanked by convergent CTCF sites (Harrold et al., 2020).

Nonetheless, our results show that the Epromoters recruit most of the transcription factor than the other coregulated gene in the same cluster, but we do not know the role that Epromoters play in the formation of the long-range interactions in a co-regulated cluster of genes. Are Epromoters necessary for the stabilization of the transcription factors and the looping formation, or they are just necessary for the stabilization of the transcription factors? (Fig 5.2).

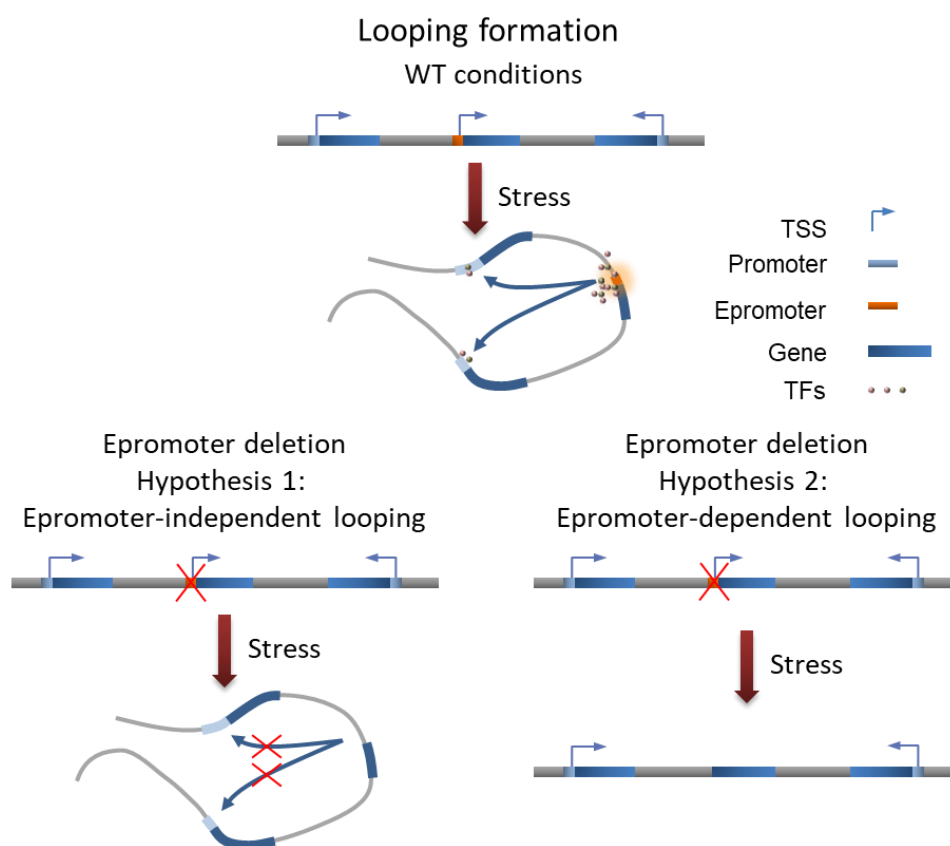


Figure 5.2 Role of the Epromoters in the long-range interactions of co-induced genes

To test answer these questions our team, in collaboration with Eric soler's team (IGMM, Montpellier, France), we are conducting several 3C-based techniques:

- To performed 4C with the viewpoint in the OAS3 Epromoter in non-stimulated and IFN α -stimulated conditions. I choose the primers and the restriction enzymes and performed the 4C experiments in K562 wt and EpOAS3 deleted cells. As of today, we are in the

sequencing and analysis step. This should indicate if the interactions between the OAS cluster exist in normal conditions or after the stress signaling.

- The team is also performing a local capture Hi-C within the OAS cluster and surrounding regions in *wild-type* and $\Delta EpOAS3$ K562 cells non-stimulated and IFN α -stimulated. One postdoc of the lab, Charbel Souaid, oversees the design and the analysis. The results should also be ready shortly. This result might indicate general changes in the 3D-conformation in the OAS cluster before and after stimulation.
- In collaboration with the team of Erick Soler (IGMM, Montpellier, France) we selected several Epromoters to perform promoter capture 4C in K562 cells non-stimulated and IFN α -stimulated. This result should give us an interaction map of several Epromoters with their target genes.

III. Genetic variants in the Epromoters

It has long been known that malfunction of regulatory elements by point mutations can lead to disease. Therefore, the identification of single-nucleotide polymorphisms (SNPs) at regulatory elements can have implications for the treatment of disease. The combination of genome-wide association study (GWAS) and chromatin signatures of enhancers has been used to identify single nucleotide polymorphisms (SNPs) at enhancers associated with conditions such as cancer, diabetes, arthritis, and Alzheimer's disease (Kikuchi et al., 2019a; Lim and Kim, 2019; Michailidou et al., 2017; Sun et al., 2018). Moreover, in prostate cancer, an SNP was found in the promoter of the short isoform of the lncRNA PCAT19 that changed his natural binding site with YY1 and NKX3.1, resulting in a new binding with HOXA2. This SNP disrupted the normal activity of the promoter reducing its promoter activity. Strikingly the SNP stronger enhancer activity that activates the long transcript of PCAT19 and CEACAM2 gene transcription and is associated with aggressive prostate cancer (fig. 5.3) (Bulger, 2018; Gao et al., 2018; Hua et al., 2018).

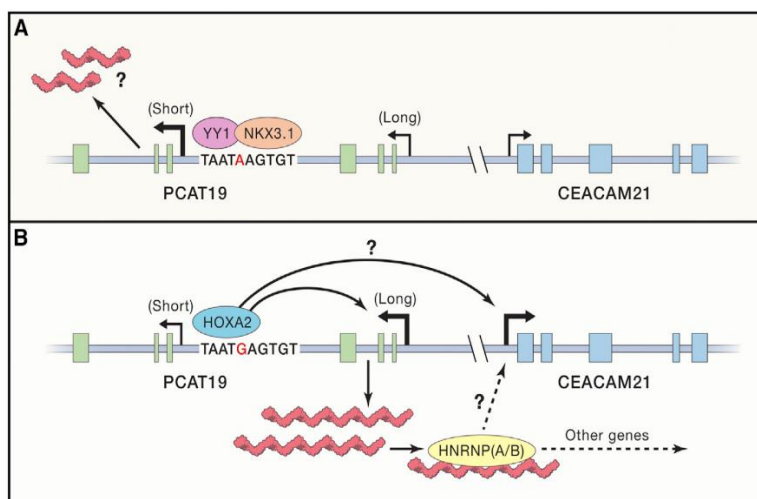


Figure 5.3 Effect of genetic variants within the Epromoters and their relationship with disease. (A) Wildtype activity. (B) Variant activity. Adapted from (Bulger, 2018).

This suggests that a single nucleotide mutation in Epromoters could affect its enhancer or promoter activities by recruiting a different transcription factor and directly or indirectly lead to a disease (Fig. 5.4). In principle, the effect of a mutation or regulatory variant within an Epromoter can have a different outcome on the relative enhancer or promoter active, either by

increasing or reducing both activities simultaneously or by differentially impacting those activities. Therefore, it is expected that genetic variation at Epromoters might have a large and complex effect on phenotypic consequences.

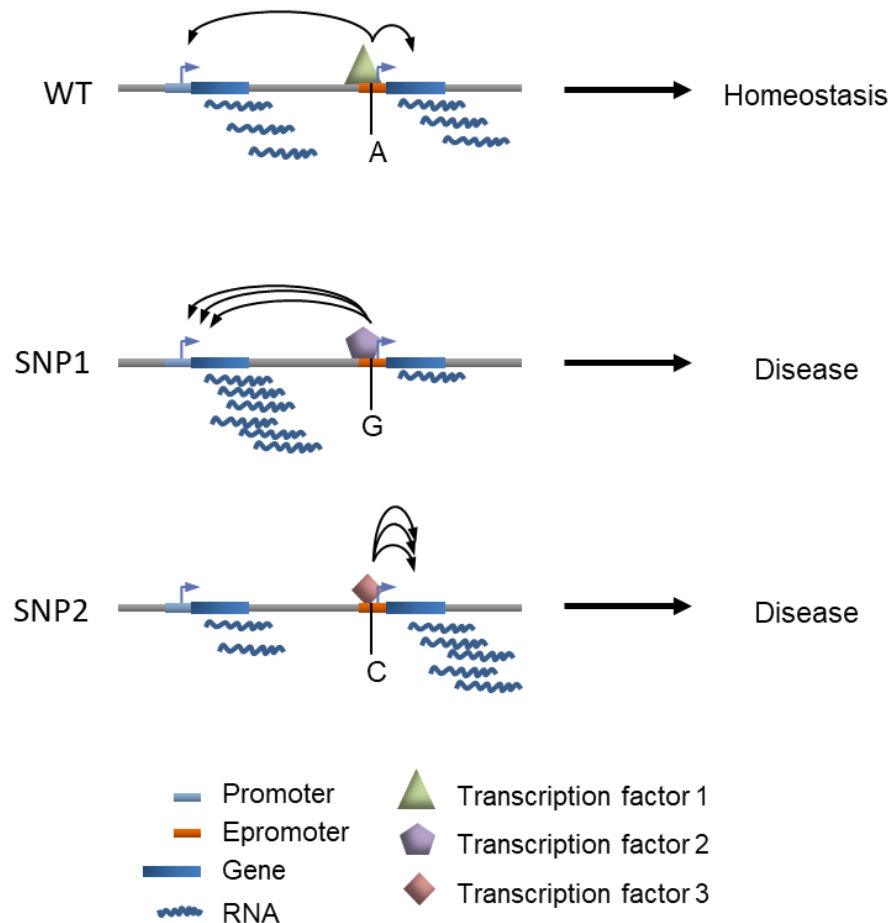


Figure 5.4 Effect of genetic variants within the Epromoters and their relationship with disease

To start addressing this issue, in collaboration with Alejandra Medina's team (LIIGH, Queretaro, Mexico) we searched for SNPs in the Epromoters identified in HeLa (Dao et al., 2017), K562, and K562+IFN α , and search for their association in expression quantitative trait loci (eQTL) and with inflammatory disease GWAS. This analysis allows us to identify disease-associated mutations within Epromoters that regulates distal genes. The effect of these mutations in the expression of distal genes can be further analysed and tested using CRISPR-Cas9-mediated single-nucleotide editing (Fig. 5.5).

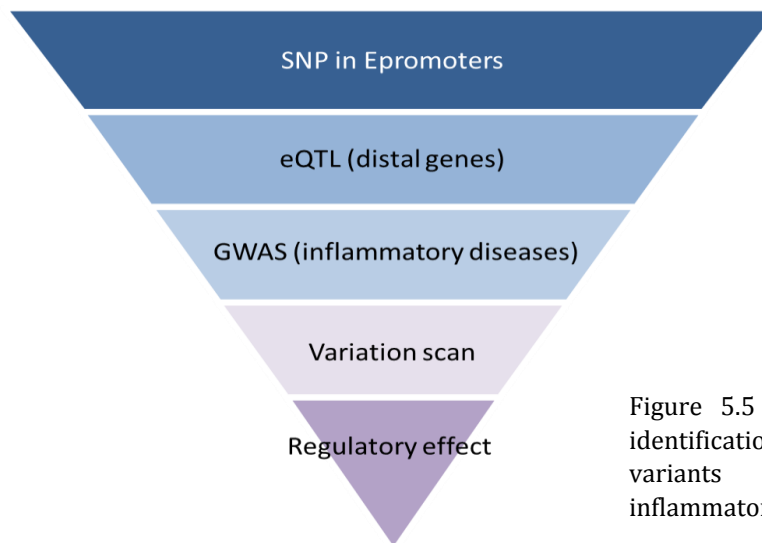


Figure 5.5 Flowchart for the identification of genetic variants associated with inflammatory diseases.

Thanks to this strategy, we were able to identify the SNP rs3771180 in the Epromoter of *IL1RL1*. This Epromoter was found as active in the HeLa cell line in normal conditions. This SNP was correlated with reduced expression of *IL1RL1* and *IL18R1*, and with increased expression of *IL18RAP*. Using a variation scan from RSAT we were able to predict that this SNP affects the binding of TFs. Furthermore, the effect of this SNP in its regulatory activity was validated by SuRE (van Arensbergen et al., 2019) (Fig 5.6A).

This SNP in *IL1RL* was associated with asthma in GWAS. *IL1RL1*, also called ST2, is the natural receptor for IL-33, and its gene codes for a transmembrane receptor, that transduce the IL-33 signal, and for a soluble receptor, that serves as a “decoy” receptor for the IL-33. The soluble receptor is used as a cardiovascular biomarker that predicts both rehospitalization and mortality (Tian et al., 2020). The association of ST2/IL-33 has been implicated in several immune and inflammatory diseases, while some SNPs of *IL1RL1* are strongly associated with asthma (Ho et al., 2013; Savenije et al., 2011). *IL18R1* and *IL18RAP* form the alpha and beta chain of the IL-18 receptor. Interestingly, five of the six members of the IL-1 receptor family share the same cluster (Fig 5.6B), but only *IL1RL1*, *IL18R1*, and *IL18RAP* have been associated with asthma (Reijmerink et al., 2008).

We predict that the involvement of SNP rs3771180 in asthma might be mediated by its impact on the expression of the three genes in the cluster. To test this hypothesis, we will need a cellular model where all three genes can be stimulated in physiologically relevant conditions. Then, the SNP can be introduced using CRISPR/Cas9 mediated homologous recombination.

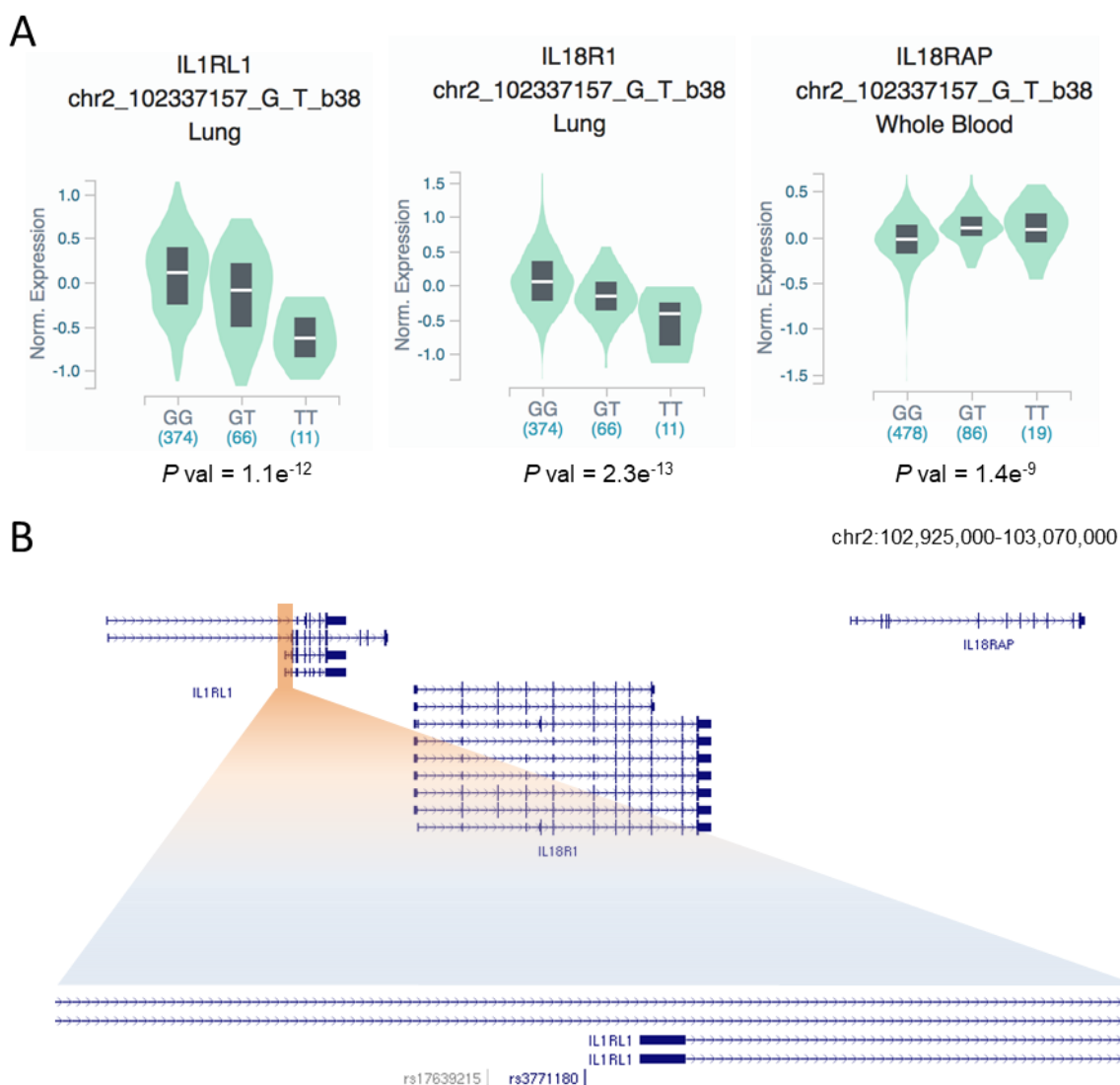


Figure 5.6 Genetic variant in the Epromoter of IL1RL1 associated with inflammatory diseases. (A) eQTL analysis of the SNP rs3771180 in the expression of IL1RL1, IL18R1, and IL18RAP. (B) Visual inspection of the locus affected by the SNP rs3771180 and its position in the Epromoter (alternative promoter) of IL1RL1.

IV. Epromoters in other cellular stress models

Provocative model... simple clusters....

Although our work was focused on the interferon response, the Epromoters could likely be more generally involved in the rapid response of genes to cellular stress. The specific regulatory network will then depend on the cell type and cellular context. This leads to the hypothesis that early stress response genes might be organized in clusters of co-regulated genes with some of their promoters having enhancer properties. Based on our results, we concluded that in the IFN α response the Epromoters recruits preferentially the key IFN α response transcription factors and can regulate the expression of the genes within the same cluster. Following this concept, we reasoned that the identification of Epromoters in other cell types or cellular context might not

require performing a high-throughput reporter assay to assess the enhancer activity of promoters. We hypothesize that the Epromoter-like function can be predicted by identifying the promoters that preferentially recruit the transcription factors in any cluster of induced genes, in response to extracellular signaling. To test this hypothesis, we did a literature search for stress signaling that included gene expression data (RNA-seq, PRO-seq, etc.) and ChIP-seq data for the key transcription factors in non-stimulated and stimulated conditions. We used those data sets to create a pipeline that allows us to isolate the cluster of stress-induced genes, the number of promoters per cluster that bound to the key transcription factors, and compute an Epromoter-likeness score (Fig. 5.7). **The bioinformatic pipeline was made by a post-doc in the lab, Himanshu Singh, while I contribute to defining the parameters, selecting the datasets, and interpreting the results (a manuscript describing these results is under preparation).**

In the heat shock response in the K562 cell line, we identify sixteen induced gene clusters that contained one Epromoter-like each [data from (Vihervaara et al., 2017)]. Visual inspection of the clusters confirmed that the Epromoter-like preferentially recruited the heat shock factor 1 (HSF1) and HSF2, while the other induced genes did not. A Master student in the lab, Juliette Malfait, was able to validate the enhancer activity of the predicted Heat-Shock dependent Epromoters using a luciferase assay (unpublished results).

Another example of Epromoter-like activity was found in the response to interferon-gamma (IFN γ) in mouse macrophages [data from (Piccolo et al., 2017)]. In this cluster of co-induced genes, we observed that the Epromoter-like region of *Cd274* (also called *Pdl1*) recruits all the transcription factors and we predicted it can regulate the expression of the genes *Jak2* and *Pdcd1lg2* (*Pdl2*). This cluster is conserved between mouse and human and was validated in the THP-1 cell line by a post-doc in the lab, Charbel Souaid. Human PDL1 is an immune inhibitory transmembrane receptor-ligand important for preventing autoimmunity by maintaining homeostasis of the immune response. This interaction provides an immune escape for tumor cells through cytotoxic T-cell inactivation and inducing a Th2 immune response. High expression levels of this gene in cancer cells are related to bad prognostic, including pancreatic, lung, and breast cancers (Birnbaum et al., 2016; Mittendorf et al., 2014; Yu et al., 2016). PDL2, like PDL1, also plays a role in the negative regulation of the immune response (Latchman et al., 2001). JAK2 is a tyrosine kinase expressed virtually in all tissues and is involved in the JAK-STAT signaling pathway of multiple cytokines, including type II interferon, IL-6, etc. (Latchman et al., 2001). Both loss and abnormal increase of JAK2 activity is associated with diverse diseases (Perner et al., 2019). To resume, the correct regulation of this gene cluster has serious clinical implications. Preliminary results by Charbel Souaid, using CRISPR-Cas9 promoter deletion, suggest that the human Epromoter-like of PDL1 is indeed an IFN γ -induced Epromoter and regulates the expression of *JAK2* and *PDL2* (Fig. 5.8).

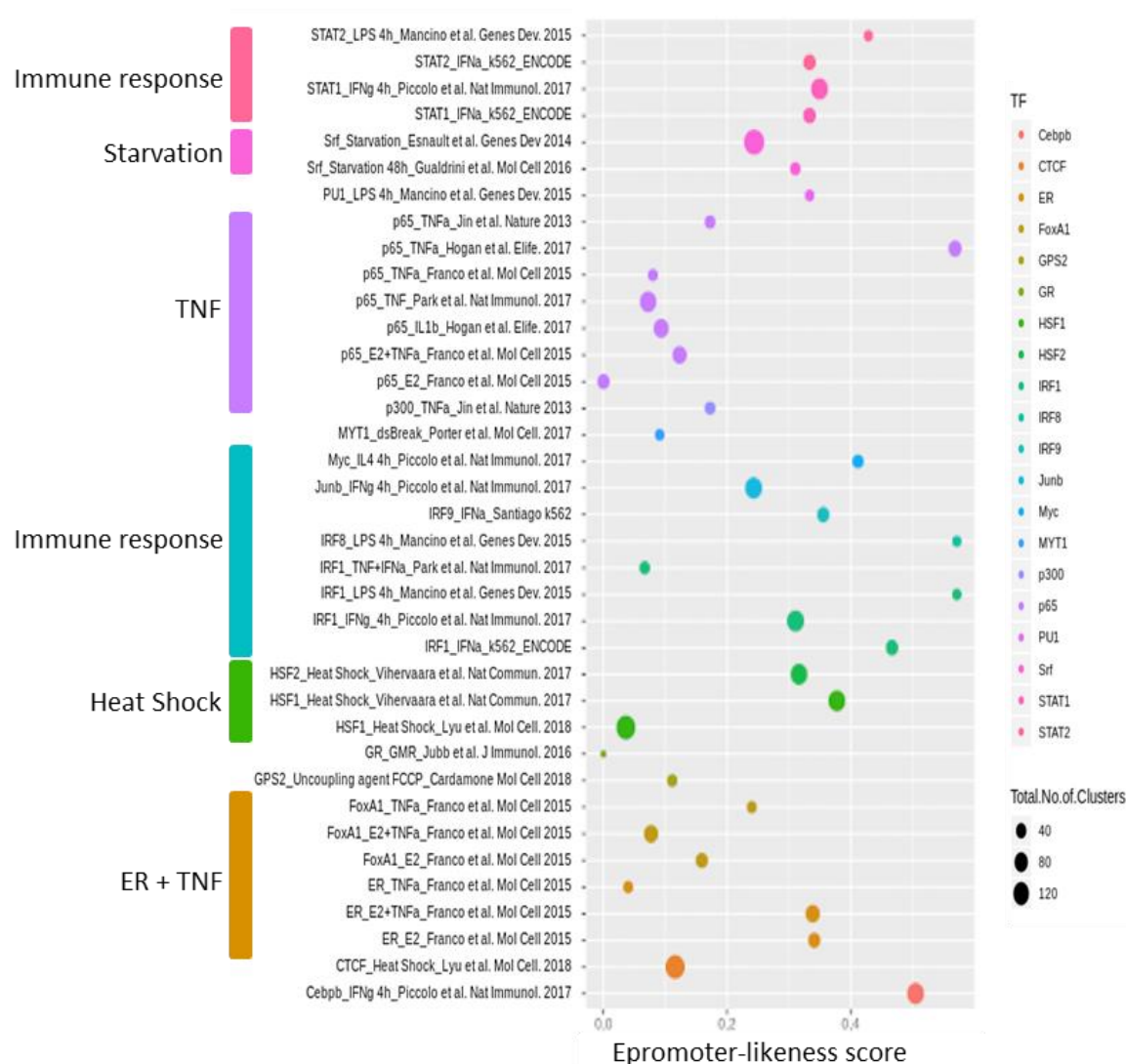


Figure 5.7 Identification of Epromoters-like active in different experimental

Finally, this pipeline was also used to identify the LPS response cluster IL15RA/IL2RA, for which we predicted that the IL15RA promoter might function as an Epromoter. Validation of this Epromoter was included in the article presented in the Results section. Overall, we developed a pipeline allowing the identification of gene-clusters potentially regulated by Epromoters in different cell types and stimulatory conditions. Preliminary validation of the predicted Epromoters supports the validity of the bioinformatic strategy and suggests that Epromoters may function in different cellular contexts. Continuing exploring the results generated by this pipeline will provide a comprehensive view of the different pathways in which Epromoters might be functionally relevant.

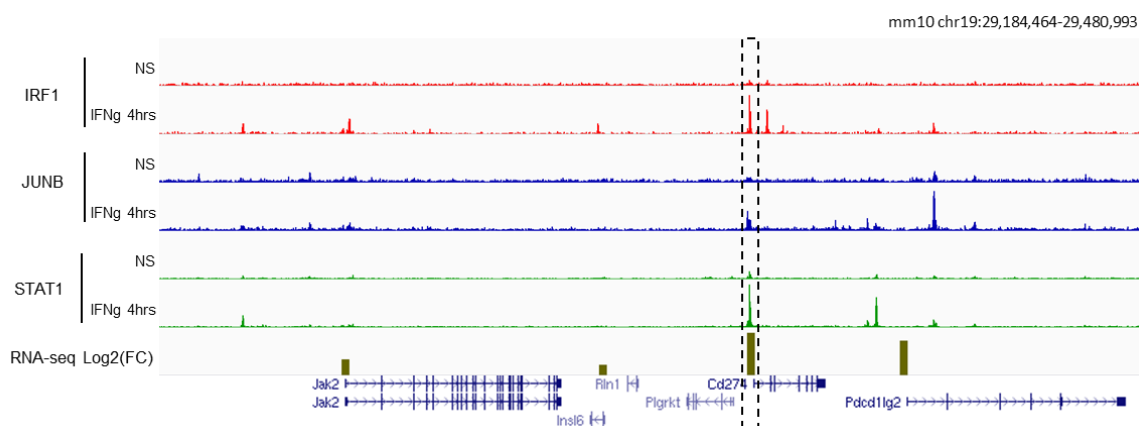


Figure 5.8 Genomic visualization of the cluster *Jak2/Pdl1(Cd274)/Pdl2(Pdcd1lg2)* in mouse macrophages stimulated with IFNg. Genomic tracks for transcription factors IRF1, JUNB, and STAT1, and the RNA-seq log2 are shown.

V. Molecular bases of promoter versus enhancer activity of Epromoters

Features defining the enhancer versus promoter activity of regulatory elements are a fundamental question in the gene regulation field and a focus of extensive research. In particular, it will be important to define the molecular basis that discriminates between the enhancer and promoter activity of Epromoters. Previous work has suggested that the intrinsic and external factors controlling the promoter and enhancer activity of a given Epromoter are not necessarily the same. Until now, the relation between the potential of each regulatory element as a local promoter and its ability to enhance transcription at the distal promoter remains poorly understood. Do the enhancer and the promoter activities positively correlate with the Epromoter activity, in which the enhancer activity rises as the promoter activity does; or is it a negative correlation, where the enhancer activity diminishes when the promoter rises and vice versa? (Fig 5.9)

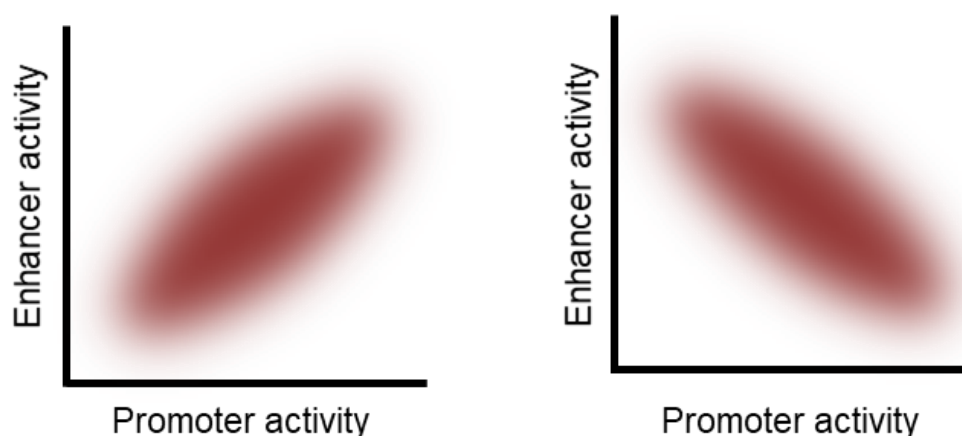


Figure 5.9 Positive and negative correlation between the enhancer and promoter activity in Epromoters

Another important question is whether enhancer and promoter functions in the Epromoters are coupled or even interdependent (fig 5.10). To start addressing this question, we mutate the two binding sites of the ISG3 factor in the *OAS3* Epromoter. These results suggested that the enhancer activity of the Epromoter is dependent on the presence of a binding site with the consensus motif, while the promoter activity does not require a high-quality binding site.

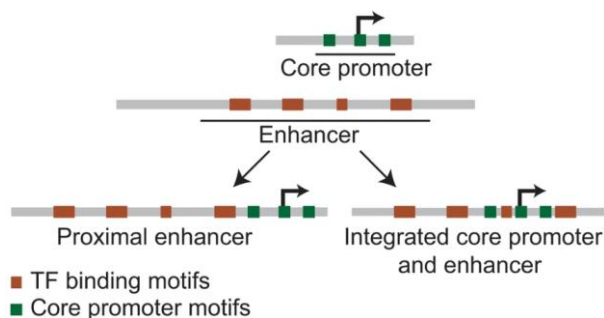


Figure 5.10 Combinatory of the core promoter elements and enhancer transcription factors in Epromoters. Modified from (Catarino et al., 2017)

As mentioned before genetic variants located in the Epromoters might also have an unanticipated role in the coordinated activation of co-regulated gene clusters. However, establishing the relationship between the causal regulatory variant and its target gene(s) is a major challenge. The discovery of disease-associated mutations at Epromoters affecting distal gene expression will represent a breakthrough in the paradigm of genetic studies.

To test the relationship between the enhancer and promoter activity in the Epromoters and the effect of natural variants (SNPs) our team is developing a new reporter assay that may test both activities in a single assay. The basic principle of this assay requires a construct that includes two different reporter genes cassettes (GFP and RFP): GFP gene expression will be ruled by a minimal promoter (TATA, ORI or SCP1), while the RFP expression will be controlled by the test regions (reference and alternative alleles generated by massive oligo synthesis). If the tested variant affects the enhancer activity the RFP/GFP ratio will be high, whereas if the promoter activity is affected the RFP/GFP will be low. This can be analyzed in different cell lines or stimulatory conditions by targeted Next-Generation Sequencing of FACS' cell sorted populations. The general structure of the dual reporter vector and the possible outcomes are exemplified in figure 5.11. The design and experimental data analysis of this dual reporter vector is part of the project of Juliette Malfait, now a Ph.D. student in our team.

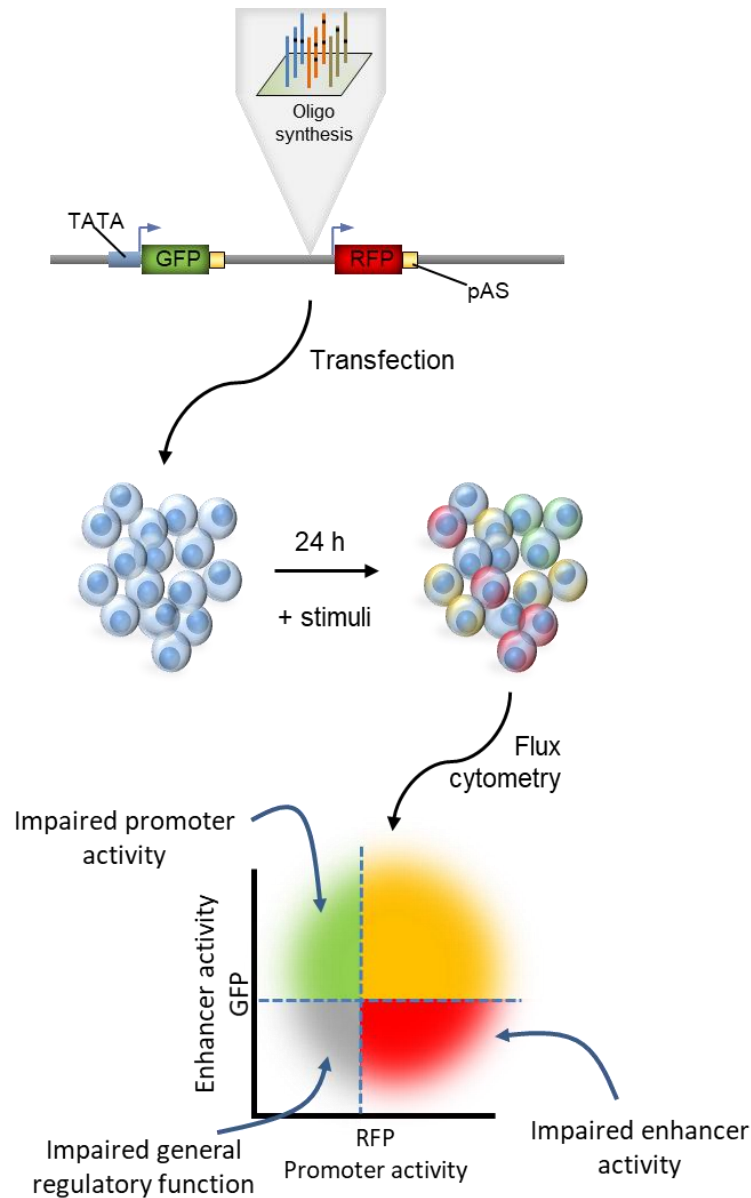


Figure 5.11 Scheme of a dual reporter assay for enhancer and promoter function in Epromoters

VI. Concluding remarks

The results obtained during my Ph.D. provided further evidence supporting the biological roles of Epromoters by highlighting their key functional involvement in the regulation of interferon-induced clusters. My work also provided the basis for developing a bioinformatic pipeline leading to the identification of Epromoters in different cellular and stimulatory contexts.

The implementation of high-throughput reporter assays has led to the unbiased discovery of a new type of regulatory element (i.e. Epromoters). In the future it will be important to further dig into the relevance of Epromoters by studying their function in physiological contexts and assess the impact of overlapping genetic variants, thus providing a valuable resource to understand the basic mechanisms of Epromoter function, but also the discovery of

their potential involvement in diseases. We forecast that Epromoters might have an unanticipated key role in the coordinated activation of co-regulated gene clusters. Since the advent of GWAS, there has been a significant number of SNPs identified as being associated with a phenotype or pathology. Strikingly most of these SNPs are in the non-coding genome and potentially cis-regulatory regions. However, establishing the relationship between the causal regulatory variant and its target gene(s) is a major challenge. The discovery of disease-associated mutations at Epromoters affecting distal gene expression represents a breakthrough in the paradigm of genetic studies. The studies of Epromoter variants might provide new rationales for the interpretation of genetic variation within promoters, as the “disease” gene might not necessarily be the closest one.

More generally, the observations of broad similarities between enhancers and gene promoters call for a revision of established distinctions between them. While they do not differ categorically in their ability to encode enhancer or promoter activities, their function might be genomic context-dependent and they could be considered in the same class of regulatory elements. A unifying model of the regulatory element has been proposed trying to understand the complexity of transcription regulation (Andersson et al., 2015a; Core et al., 2014). Follow this hypothesis, regulatory elements often interact in close physical proximity in RNAPII foci and the activity of each regulatory element can be influenced by the others. Depending on the context, the elements can act as either promoters or enhancers.

Bibliography

- A, P., and Weber, S.C. (2019). Evidence for and against Liquid-Liquid Phase Separation in the Nucleus. *Noncoding RNA* 5.
- Adey, A., Morrison, H.G., Asan, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., *et al.* (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 11, R119.
- Adiconis, X., Haber, A.L., Simmons, S.K., Moonshine, A.L., Ji, Z., Busby, M.A., Shi, X., Jacques, J., Lancaster, M.A., Pan, J.Q., *et al.* (2018). Author Correction: Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat Methods* 15, 1126.
- Alerasool, N., Segal, D., Lee, H., and Taipale, M. (2020). An efficient KRAB domain for CRISPRi applications in human cells. *Nat Methods*.
- Alleva, D.G., Kaser, S.B., Monroy, M.A., Fenton, M.J., and Beller, D.I. (1997). IL-15 functions as a potent autocrine regulator of macrophage proinflammatory cytokine production: evidence for differential receptor subunit utilization associated with stimulation or inhibition. *J Immunol* 159, 2941-2951.
- Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* 51, 786-794.
- Andersson, R., Chen, Y., Core, L., Lis, J.T., Sandelin, A., and Jensen, T.H. (2015a). Human Gene Promoters Are Intrinsically Bidirectional. *Molecular cell* 60, 346-347.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461.
- Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature reviews Genetics* 21, 71-87.
- Andersson, R., Sandelin, A., and Danko, C.G. (2015b). A unified architecture of transcriptional regulatory elements. *Trends in genetics : TIG* 31, 426-433.
- Andreakos, E., and Tsiodras, S. (2020). COVID-19: lambda interferon against viral load and hyperinflammation. *EMBO molecular medicine* 12, e12465.
- Andrey, G., Schopflin, R., Jerkovic, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M., *et al.* (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome research* 27, 223-233.
- Arenas-Mena, C. (2017). The origins of developmental gene regulation. *Evol Dev* 19, 96-107.
- Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074-1077.
- Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729-740.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299-308.
- Banninger, G., and Reich, N.C. (2004). STAT2 nuclear trafficking. *The Journal of biological chemistry* 279, 39199-39206.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research* 21, 381-395.
- Barakat, T.S., Halbritter, F., Zhang, M., Rendeiro, A.F., Perenthaler, E., Bock, C., and Chambers, I. (2018). Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell stem cell* 23, 276-288 e278.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.

- Barrat, F.J., Crow, M.K., and Ivashkiv, L.B. (2019). Interferon target-gene expression and epigenomic signatures in health and disease. *Nature immunology* 20, 1574-1583.
- Barrat, F.J., Elkou, K.B., and Fitzgerald, K.A. (2016). Importance of Nucleic Acid Recognition in Inflammation and Autoimmunity. *Annu Rev Med* 67, 323-336.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
- Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153-155.
- Baxter, J.S., Leavy, O.C., Dryden, N.H., Maguire, S., Johnson, N., Fedele, V., Simigdala, N., Martin, L.A., Andrews, S., Wingett, S.W., *et al.* (2018). Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat Commun* 9, 1028.
- Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremens, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome research* 27, 1139-1152.
- Bhattacharya, S., Eckner, R., Grossman, S., Oldread, E., Arany, Z., D'Andrea, A., and Livingston, D.M. (1996). Cooperation of Stat2 and p300/CBP in signalling induced by interferon-alpha. *Nature* 383, 344-347.
- Bienias, M., Bruck, N., Griep, C., Wolf, C., Kretschmer, S., Kind, B., Tungler, V., Berner, R., and Lee-Kirsch, M.A. (2018). Therapeutic Approaches to Type I Interferonopathies. *Curr Rheumatol Rep* 20, 32.
- Bird, A. (1992). The essentials of DNA methylation. *Cell* 70, 5-8.
- Birnbaum, D.J., Finetti, P., Lopresti, A., Gilabert, M., Poizat, F., Turrini, O., Raoul, J.L., Delperro, J.R., Moutardier, V., Birnbaum, D., *et al.* (2016). Prognostic value of PDL1 expression in pancreatic cancer. *Oncotarget* 7, 71198-71210.
- Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799-816.
- Blat, Y., and Kleckner, N. (1999). Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* 98, 249-259.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 132, 311-322.
- Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985). Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 41, 41-48.
- Brandt, F., Carlson, L.A., Hartl, F.U., Baumeister, W., and Grunewald, K. (2010). The three-dimensional organization of polyribosomes in intact human cells. *Molecular cell* 39, 560-569.
- Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS biology* 6, 2928-2939.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213-1218.
- Bulger, M. (2018). Changing the Guard at a Prostate Cancer SNP. *Cell* 174, 501-502.
- Carelli, F.N., Liechti, A., Halbert, J., Warnefors, M., and Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. *Nat Commun* 9, 4066.
- Carrozza, M.J., Li, B., Florens, L., Sukanuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.

- Castaneda-Delgado, J.E., Bastian-Hernandez, Y., Macias-Segura, N., Santiago-Algarra, D., Castillo-Ortiz, J.D., Aleman-Navarro, A.L., Martinez-Tejada, P., Enciso-Moreno, L., Garcia-De Lira, Y., Olguin-Calderon, D., *et al.* (2017). Type I Interferon Gene Response Is Increased in Early and Established Rheumatoid Arthritis and Correlates with Autoantibody Production. *Front Immunol* *8*, 285.
- Catarino, R.R., Neumayr, C., and Stark, A. (2017). Promoting transcription over long distances. *Nature genetics* *49*, 972-973.
- Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* *147*, 358-369.
- Chatterjee, S., and Ahituv, N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annual review of genomics and human genetics*.
- Chaudhri, V.K., Dienger-Stambaugh, K., Wu, Z., Shrestha, M., and Singh, H. (2020). Charting the cis-regulome of activated B cells by coupling structural and functional genomics. *Nature immunology* *21*, 210-220.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* *12*, 326-328.
- Chen, H.M., Tanaka, N., Mitani, Y., Oda, E., Nozawa, H., Chen, J.Z., Yanai, H., Negishi, H., Choi, M.K., Iwasaki, T., *et al.* (2009). Critical role for constitutive type I interferon signaling in the prevention of cellular transformation. *Cancer Sci* *100*, 449-456.
- Chen, K., Liu, J., and Cao, X. (2017). Regulation of type I interferon signaling in immunity and inflammation: A comprehensive review. *Journal of autoimmunity* *83*, 1-11.
- Chen, S.S., Ruteshouser, E.C., Maity, S.N., and de Crombrughe, B. (1997). Cell-specific in vivo DNA-protein interactions at the proximal promoters of the pro alpha 1(I) and the pro alpha2(I) collagen genes. *Nucleic acids research* *25*, 3261-3268.
- Chu, V.T., Graf, R., Wirtz, T., Weber, T., Favret, J., Li, X., Petsch, K., Tran, N.T., Sieweke, M.H., Berek, C., *et al.* (2016). Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line. *Proc Natl Acad Sci U S A* *113*, 12514-12519.
- Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57-74.
- Cook, P.R. (2010). A model for all genomes: the role of transcription factories. *J Mol Biol* *395*, 1-10.
- Cook, P.R., and Marenduzzo, D. (2018). Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic acids research* *46*, 9895-9906.
- Cooles, F.A.H., Anderson, A.E., Lendrem, D.W., Norris, J., Pratt, A.G., Hilkens, C.M.U., and Isaacs, J.D. (2018). The interferon gene signature is increased in patients with early treatment-naive rheumatoid arthritis and predicts a poorer response to initial therapy. *J Allergy Clin Immunol* *141*, 445-448 e444.
- Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., *et al.* (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* *14*, 959-962.
- Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics* *46*, 1311-1320.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* *322*, 1845-1848.
- Corrales, M., Rosado, A., Cortini, R., van Arensbergen, J., van Steensel, B., and Filion, G.J. (2017). Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome research* *27*, 1153-1161.

- Corthay, A., Skovseth, D.K., Lundin, K.U., Rosjo, E., Omholt, H., Hofgaard, P.O., Haraldsen, G., and Bogen, B. (2005). Primary antitumor immune response mediated by CD4+ T cells. *Immunity* 22, 371-383.
- Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. (2006a). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* 3, 503-509.
- Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., *et al.* (2006b). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome research* 16, 123-131.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA*, DOI: 10.1073.
- da Rocha, S.T., and Gendrel, A.V. (2019). The influence of DNA methylation on monoallelic expression. *Essays in biochemistry* 63, 663-676.
- Dao, L.T.M., Galindo-Albarran, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., *et al.* (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature genetics* 49, 1073-1081.
- Dao, L.T.M., and Spicuglia, S. (2018). Transcriptional regulation by promoters with enhancer function. *Transcription*.
- Dao, L.T.M., Vanhille, L., Griffon, A., Fernandez, N., and Spicuglia, S. (2015). CapStarr-seq protocol. Protocol Exchange.
- Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H. (2008). The functional consequences of alternative promoter use in mammalian genomes. *Trends in genetics : TIG* 24, 167-177.
- De Gobbi, M., Viprakit, V., Hughes, J.R., Fisher, C., Buckle, V.J., Ayyub, H., Gibbons, R.J., Vernimmen, D., Yoshinaga, Y., de Jong, P., *et al.* (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* 312, 1215-1217.
- de Villiers, J., Olson, L., Tyndall, C., and Schaffner, W. (1982). Transcriptional 'enhancers' from SV40 and polyoma virus show a cell type preference. *Nucleic acids research* 10, 7965-7976.
- de Villiers, J., and Schaffner, W. (1981). A small segment of polyoma virus DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 base pairs. *Nucleic acids research* 9, 6251-6264.
- de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Versteegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Molecular cell* 60, 676-684.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306-1311.
- Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* 19, 2418-2423.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* 22, 1775-1789.
- Diamond, M.S., and Farzan, M. (2013). The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol* 13, 46-57.
- Dianov, G.L., and Hubscher, U. (2013). Mammalian base excision repair: the forgotten archangel. *Nucleic acids research* 41, 3483-3490.
- Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., *et al.* (2017a). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods*.

- Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., *et al.* (2017b). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* *14*, 629-635.
- Ding, L., Bailey, M.H., Porta-Pardo, E., Thorsson, V., Colaprico, A., Bertrand, D., Gibbs, D.L., Weerasinghe, A., Huang, K.L., Tokheim, C., *et al.* (2018). Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell* *173*, 305-320 e310.
- DiStefano, J.K. (2018). The Emerging Role of Long Noncoding RNAs in Human Disease. *Methods Mol Biol* *1706*, 91-110.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* *485*, 376-380.
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* *489*, 101-108.
- Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. *Nat Commun* *11*, 1061.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., *et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome research* *16*, 1299-1309.
- Dryden, N.H., Broome, L.R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I., *et al.* (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome research* *24*, 1854-1868.
- Du, Q., Luu, P.L., Stirzaker, C., and Clark, S.J. (2015). Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* *7*, 1051-1073.
- Duncan, C.J., Mohamad, S.M., Young, D.F., Skelton, A.J., Leahy, T.R., Munday, D.C., Butler, K.M., Morfopoulou, S., Brown, J.R., Hubank, M., *et al.* (2015). Human IFNAR2 deficiency: Lessons for antiviral immunity. *Science translational medicine* *7*, 307ra154.
- Duttke, S.H., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., Heinz, S., Kadonaga, J.T., and Ohler, U. (2015). Human promoters are intrinsically directional. *Molecular cell* *57*, 674-684.
- Ea, V., Baudement, M.O., Lesne, A., and Forne, T. (2015). Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes* *6*, 734-750.
- Edmondson, D.G., Davie, J.K., Zhou, J., Mirnikjoo, B., Tatchell, K., and Dent, S.Y. (2002). Site-specific loss of acetylation upon phosphorylation of histone H3. *The Journal of biological chemistry* *277*, 29496-29502.
- El-Mounadi, K., Morales-Floriano, M.L., and Garcia-Ruiz, H. (2020). Principles, Applications, and Biosafety of Plant Genome Editing Using CRISPR-Cas9. *Frontiers in plant science* *11*, 56.
- Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*.
- Erdel, F., and Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* *114*, 2262-2270.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., *et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* *473*, 43-49.
- Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T.S., and Kellis, M. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol* *34*, 1180-1190.
- Espana, A.P., Santiago-Algarra, D., Pradel, L., and Spicuglia, S. (2017). [High-throughput approaches to study cis-regulating elements]. *Biol Aujourdhui* *211*, 271-280.

- Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M., and Burckstummer, T. (2014). Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome research* 24, 2059-2065.
- Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606-620.
- Fernando, M.B., Ahfeldt, T., and Brennand, K.J. (2020). Modeling the complex genetic architectures of brain disease. *Nature genetics* 52, 363-369.
- Ferraiuolo, M.A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X.Q., Nadler, M., Blanchette, M., and Dostie, J. (2010). The three-dimensional architecture of Hox cluster silencing. *Nucleic acids research* 38, 7472-7484.
- Ferre, Q., Charbonnier, G., Sadouni, N., Lopez, F., Kermezli, Y., Spicuglia, S., Capponi, C., Ghattas, B., and Puthier, D. (2019). OLOGRAM: Determining significance of total overlap length between genomic regions sets. *Bioinformatics*.
- Feuerborn, A., and Cook, P.R. (2015). Why the activity of a gene depends on its neighbors. *Trends in genetics : TIG* 31, 483-490.
- Field, A., and Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annual review of biochemistry* 89, 213-234.
- Flaishon, L., Hershkovich, R., Lantner, F., Lider, O., Alon, R., Levo, Y., Flavell, R.A., and Shachar, I. (2000). Autocrine secretion of interferon gamma negatively regulates homing of immature B cells. *J Exp Med* 192, 1381-1388.
- Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranasic, D., *et al.* (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research* 48, D87-D92.
- Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., *et al.* (2014). A promoter-level mammalian expression atlas. *Nature* 507, 462-470.
- Fraser, J., Rousseau, M., Shenker, S., Ferraiuolo, M.A., Hayashizaki, Y., Blanchette, M., and Dostie, J. (2009). Chromatin conformation signatures of cellular differentiation. *Genome Biol* 10, R37.
- Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L.A. (2017). Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harbor symposia on quantitative biology* 82, 45-55.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., *et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58-64.
- Furuichi, Y. (2015). Discovery of m(7)G-cap in eukaryotic mRNAs. *Proceedings of the Japan Academy Series B, Physical and biological sciences* 91, 394-409.
- Fuster-Garcia, C., Garcia-Bohorquez, B., Rodriguez-Munoz, A., Millan, J.M., and Garcia-Garcia, G. (2020). Application of CRISPR Tools for Variant Interpretation and Disease Modeling in Inherited Retinal Dystrophies. *Genes* 11.
- Gabory, A., Ripoche, M.A., Le Digarcher, A., Watrin, F., Ziyat, A., Forne, T., Jammes, H., Ainscough, J.F., Surani, M.A., Journot, L., *et al.* (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* 136, 3413-3421.
- Gao, P., Xia, J.H., Sipeky, C., Dong, X.M., Zhang, Q., Yang, Y., Zhang, P., Cruz, S.P., Zhang, K., Zhu, J., *et al.* (2018). Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* 174, 576-589 e518.
- Gariglio, P., Buss, J., and Green, M.H. (1974). Sarkosyl activation of RNA polymerase activity in mitotic mouse cells. *FEBS letters* 44, 330-333.
- Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., *et al.* (2010). A map of open chromatin in human pancreatic islets. *Nature genetics* 42, 255-259.

- Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., *et al.* (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 159, 647-661.
- Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442-451.
- Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33, 717-728.
- Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* 17, 877-885.
- Gisselbrecht, S.S., Palagi, A., Kurland, J.V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in *Drosophila* Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Molecular cell* 77, 324-337 e328.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *AnnuRevBiochem* 74, 481-514.
- Gomez-Velazquez, M., Badia-Careaga, C., Lechuga-Vieco, A.V., Nieto-Arellano, R., Tena, J.J., Rollan, I., Alvarez, A., Torroja, C., Caceres, E.F., Roy, A.R., *et al.* (2017). CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart. *PLoS genetics* 13, e1006985.
- Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284-288.
- Goodbourn, S., Zinn, K., and Maniatis, T. (1985). Human beta-interferon gene expression is regulated by an inducible enhancer element. *Cell* 41, 509-520.
- Gotthardt, D., and Sexl, V. (2016). STATs in NK-Cells: The Good, the Bad, and the Ugly. *Front Immunol* 7, 694.
- Guiducci, C., Ghirelli, C., Marloie-Provost, M.A., Matray, T., Coffman, R.L., Liu, Y.J., Barrat, F.J., and Soumelis, V. (2008). PI3K is critical for the nuclear translocation of IRF-7 and type I IFN production by human plasmacytoid dendritic cells in response to TLR activation. *J Exp Med* 205, 315-322.
- Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature reviews Molecular cell biology* 19, 621-637.
- Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., *et al.* (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* 369, 718-724.
- Halfon, M.S. (2020). Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends in genetics : TIG* 36, 149-151.
- Harrold, C.L., Gosden, M.E., Hanssen, L.L.P., Stolper, R.J., Downes, D.J., Telenius, J.M., Biggs, D., Preece, C., Alghadban, S., Sharpe, J.A., *et al.* (2020). A functional overlap between actively transcribed genes and chromatin boundary elements. 2020.2007.2001.182089.
- Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Molecular cell* 10, 1467-1477.
- Heidari, N., Phanstiel, D.H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q., and Snyder, M.P. (2014). Genome-wide map of regulatory interactions in the human genome. *Genome research* 24, 1905-1917.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 38, 576-589.

- Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* 32, 26-41.
- Henson, D.M., Chou, C., Sakurai, N., and Egawa, T. (2014). A silencer-proximal intronic region is required for sustained CD4 expression in postselection thymocytes. *J Immunol* 192, 4620-4627.
- Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., *et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 6, 283-289.
- Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochemical and biophysical research communications* 52, 504-510.
- Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33, 510-517.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., *et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454-1458.
- Ho, J.E., Chen, W.Y., Chen, M.H., Larson, M.G., McCabe, E.L., Cheng, S., Ghorbani, A., Coglianese, E., Emilsson, V., Johnson, A.D., *et al.* (2013). Common genetic variation at the IL1RL1 locus regulates IL-33/ST2 signaling. *The Journal of clinical investigation* 123, 4208-4218.
- Hodawadekar, S.C., and Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* 26, 5528-5540.
- Hoffman, M.M., Buske, O.J., Wang, J., Weng, Z., Bilmes, J.A., and Noble, W.S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* 9, 473-476.
- Hoffmann, H.H., Schneider, W.M., and Rice, C.M. (2015). Interferons and viruses: an evolutionary arms race of molecular interactions. *Trends in immunology* 36, 124-138.
- Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* 5, e1000566.
- Horii, T., Morita, S., Kimura, M., Kobayashi, R., Tamura, D., Takahashi, R.U., Kimura, H., Suetake, I., Ohata, H., Okamoto, K., *et al.* (2013). Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *PeerJ* 1, e230.
- Hornung, V., Hartmann, R., Ablasser, A., and Hopfner, K.P. (2014). OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat Rev Immunol* 14, 521-528.
- Horz, W., and Altenburger, W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic acids research* 9, 2643-2658.
- Hsieh, T.H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162, 108-119.
- Hsu, Y.A., Huang, C.C., Kung, Y.J., Lin, H.J., Chang, C.Y., Lee, K.R., and Wan, L. (2016). The anti-proliferative effects of type I IFN involve STAT6-mediated regulation of SP1 and BCL6. *Cancer Lett* 375, 303-312.
- Hua, J.T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., Lu, J., Zhou, S., Wang, M., Li, H., *et al.* (2018). Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. *Cell* 174, 564-575 e518.
- Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S.H., Xu, J., and Yuan, G.C. (2018). Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun* 9, 943.
- Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature genetics* 46, 205-212.

- Hull, E.E., Montgomery, M.R., and Leyva, K.J. (2016). HDAC Inhibitors as Epigenetic Regulators of the Immune System: Impacts on Cancer Therapy and Inflammatory Diseases. *BioMed research international* 2016, 8797206.
- Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merckenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome research* 25, 504-513.
- Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., and Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome research* 27, 38-52.
- Inoue, F., Kreimer, A., Ashuach, T., Ahituv, N., and Yosef, N. (2019). Identification and Massively Parallel Characterization of Regulatory Elements Driving Neural Induction. *Cell stem cell* 25, 713-727 e710.
- Inukai, S., Kock, K.H., and Bulyk, M.L. (2017). Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* 43, 110-119.
- Isaacs, A., and Lindenmann, J. (1957). Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci* 147, 258-267.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology* 169, 5429-5433.
- Ivashkiv, L.B., and Donlin, L.T. (2014). Regulation of type I interferon responses. *Nat Rev Immunol* 14, 36-49.
- Jager, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H.E., Heindl, A., Whiffin, N., Carnicer, M.J., Broome, L., Dryden, N., *et al.* (2015). Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun* 6, 6178.
- Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* 43, 1565-1575.
- Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., *et al.* (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369-1384 e1319.
- Jefferies, C.A. (2019). Regulating IRFs in IFN Driven Disease. *Front Immunol* 10, 325.
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews Genetics* 10, 161-172.
- Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature genetics* 41, 941-945.
- Jin, P., Han, T.H., Ren, J., Saunders, S., Wang, E., Marincola, F.M., and Stroncek, D.F. (2010). Molecular signatures of maturing dendritic cells: implications for testing the quality of dendritic cell therapies. *Journal of translational medicine* 8, 4.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497-1502.
- Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., *et al.* (2017). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548, 343-346.
- Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., *et al.* (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature genetics* 51, 1442-1449.
- Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental biology* 339, 225-229.

- Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. *Wiley interdisciplinary reviews Developmental biology* 1, 40-51.
- Kalita, C.A., Brown, C.D., Freiman, A., Isherwood, J., Wen, X., Pique-Regi, R., and Luca, F. (2018). High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. *Genome research* 28, 1701-1708.
- Kang, H., Minder, P., Park, M.A., Mesquitta, W.T., Torbett, B.E., and Slukvin, II (2015). CCR5 Disruption in Induced Pluripotent Stem Cells Using CRISPR/Cas9 Provides Selective Resistance of Immune Cells to CCR5-tropic HIV-1 Virus. *Molecular therapy Nucleic acids* 4, e268.
- Kaniskan, H.U., Martini, M.L., and Jin, J. (2018). Inhibitors of Protein Methyltransferases and Demethylases. *Chemical reviews* 118, 989-1068.
- Karimian, A., Azizian, K., Parsian, H., Rafieian, S., Shafiei-Irannejad, V., Kheyrollah, M., Yousefi, M., Majidinia, M., and Yousefi, B. (2019). CRISPR/Cas9 technology as a potent molecular tool for gene therapy. *J Cell Physiol* 234, 12267-12277.
- Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* 107, 2926-2931.
- Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., Gylfe, A.E., Ristolainen, H., Hanninen, U.A., Cajuso, T., *et al.* (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nature genetics* 47, 818-821.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001). Functional annotation of a full-length mouse cDNA collection. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. *Nature* 409, 685-690.
- Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637-650.
- Kawasaki, T., and Kawai, T. (2014). Toll-like receptor signaling pathways. *Front Immunol* 5, 461.
- Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome research* 22, 2497-2506.
- Keshet, I., Lieman-Hurwitz, J., and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell* 44, 535-543.
- Kikuchi, M., Hara, N., Hasegawa, M., Miyashita, A., Kuwano, R., Ikeuchi, T., and Nakaya, A. (2019a). Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med Genomics* 12, 128.
- Kikuchi, M., Kizaki, K., Shigeno, S., Toji, N., Ishiguro-Oonuma, T., Koshi, K., Takahashi, T., and Hashizume, K. (2019b). Newly identified interferon tau-responsive Hes family BHLH transcription factor 4 and cytidine/uridine monophosphate kinase 2 genes in peripheral blood granulocytes during early pregnancy in cows. *Domestic animal endocrinology* 68, 64-72.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.
- Kim, Y.H., Marhon, S.A., Zhang, Y., Steger, D.J., Won, K.J., and Lazar, M.A. (2018). Rev-erbalpha dynamically modulates chromatin looping to control circadian gene transcription. *Science* 359, 1274-1277.
- Klein, J.C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., and Shendure, J. (2020). A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat Methods*.
- Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature reviews Genetics* 20, 207-220.
- Koch, F., Jourquin, F., Ferrier, P., and Andrau, J.C. (2008). Genome-wide RNA polymerase II: not genes only! *Trends Biochem Sci* 33, 265-273.
- Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393-407.

- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schubeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular cell* 67, 411-422 e414.
- Kreimer, A., Yan, Z., Ahituv, N., and Yosef, N. (2019). Meta-analysis of massively parallel reporter assays enables prediction of regulatory function across cell types. *Human mutation* 40, 1299-1313.
- Kruesi, W.S., Core, L.J., Waters, C.T., Lis, J.T., and Meyer, B.J. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* 2, e00808.
- Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. *Advances in genetics* 70, 27-56.
- Kurata, M., Yamamoto, K., Moriarity, B.S., Kitagawa, M., and Largaespada, D.A. (2018). CRISPR/Cas9 library screening for drug target discovery. *Journal of human genetics* 63, 179-186.
- Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950-953.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* 12, 34-44.
- Laimins, L., Holmgren-Konig, M., and Khoury, G. (1986). Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. *Proc Natl Acad Sci U S A* 83, 3151-3155.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170-182.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Latchman, Y., Wood, C.R., Chernova, T., Chaudhary, D., Borde, M., Chernova, I., Iwai, Y., Long, A.J., Brown, J.A., Nunes, R., *et al.* (2001). PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nature immunology* 2, 261-268.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021-1036.
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in genetics : TIG* 32, 42-56.
- Lazear, H.M., Schoggins, J.W., and Diamond, M.S. (2019). Shared and Distinct Functions of Type I and Type III Interferons. *Immunity* 50, 907-923.
- Lea, A.J., Vockley, C.M., Johnston, R.A., Del Carpio, C.A., Barreiro, L.B., Reddy, T.E., and Tung, J. (2018). Genome-wide quantification of the effects of DNA methylation on human gene regulation. *Elife* 7.
- Lee-Kirsch, M.A. (2017). The Type I Interferonopathies. *Annu Rev Med* 68, 297-315.
- Lee, A.J., and Ashkar, A.A. (2018). The Dual Nature of Type I and Type II Interferons. *Front Immunol* 9, 2061.
- Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annual review of genetics* 34, 77-137.
- Lewis, B.A., Sims, R.J., 3rd, Lane, W.S., and Reinberg, D. (2005). Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Molecular cell* 18, 471-481.
- Li, D., Hsu, S., Purushotham, D., Sears, R.L., and Wang, T. (2019a). WashU Epigenome Browser update 2019. *Nucleic acids research* 47, W158-W165.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926.

- Li, L., Suzuki, T., Mori, N., and Greengard, P. (1993). Identification of a functional silencer element involved in neuron-specific expression of the synapsin I gene. *Proc Natl Acad Sci U S A* *90*, 1460-1464.
- Li, W., Wong, W.H., and Jiang, R. (2019b). DeepTACT: predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic acids research* *47*, e60.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289-293.
- Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* *18*, 1606-1617.
- Lim, J., and Kim, K. (2019). Genetic variants differentially associated with rheumatoid arthritis and systemic lupus erythematosus reveal the disease-specific biology. *Sci Rep* *9*, 2739.
- Lin, D., Hong, P., Zhang, S., Xu, W., Jamal, M., Yan, K., Lei, Y., Li, L., Ruan, Y., Fu, Z.F., *et al.* (2018). Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nature genetics* *50*, 754-763.
- Lin, J.D., Yang, S.F., Wang, Y.H., Fang, W.F., Tang, K.T., and Cheng, C.W. (2019). Associations of gene polymorphisms in interferon-alpha signature-related genes with autoimmune thyroid diseases. *Clin Endocrinol (Oxf)* *91*, 860-868.
- Lin, S.R., Yang, H.C., Kuo, Y.T., Liu, C.J., Yang, T.Y., Sung, K.C., Lin, Y.Y., Wang, H.Y., Wang, C.C., Shen, Y.C., *et al.* (2014). The CRISPR/Cas9 System Facilitates Clearance of the Intrahepatic HBV Templates In Vivo. *Molecular therapy Nucleic acids* *3*, e186.
- Liu, B., Saber, A., and Haisma, H.J. (2019). CRISPR/Cas9: a powerful tool for identification of new targets for cancer treatment. *Drug discovery today* *24*, 955-970.
- Liu, S., Liu, Y., Zhang, Q., Wu, J., Liang, J., Yu, S., Wei, G.H., White, K.P., and Wang, X. (2017a). Systematic identification of regulatory variants associated with cancer risk. *Genome Biol* *18*, 194.
- Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., *et al.* (2017b). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* *355*.
- Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell* *167*, 233-247 e217.
- Liu, Y., Irie, T., Yada, T., and Suzuki, Y. (2017c). A new computational method to predict transcriptional activity of a DNA sequence from diverse datasets of massively parallel reporter assays. *Nucleic acids research* *45*, e124.
- Lo, W.S., Trievel, R.C., Rojas, J.R., Duggan, L., Hsu, J.Y., Allis, C.D., Marmorstein, R., and Berger, S.L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Molecular cell* *5*, 917-926.
- Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* *126*, 403-413.
- Lomvardas, S., and Thanos, D. (2001). Nucleosome sliding via TBP DNA binding in vivo. *Cell* *106*, 685-696.
- Lopez, F., Charbonnier, G., Kermezli, Y., Belhocine, M., Ferre, Q., Zweig, N., Aribi, M., Gonzalez, A., Spicuglia, S., and Puthier, D. (2019). Explore, edit and leverage genomic annotations using Python GTF toolkit. *Bioinformatics*.
- Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* *153*, 320-334.
- Lu, C.C., Chen, M.Y., Lee, W.S., and Chang, Y.L. (2020). Potential therapeutic agents against COVID-19: What we know so far. *Journal of the Chinese Medical Association : JCMSA* *83*, 534-536.
- Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* *161*, 1012-1025.

- Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., *et al.* (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* 12, 71-78.
- Madan, N., Ghazi, A.R., Kong, X., Chen, E.S., Shaw, C.A., and Edelstein, L.C. (2019). Functionalization of CD36 cardiovascular disease and expression associated variants by interdisciplinary high throughput analysis. *PLoS genetics* 15, e1008287.
- Madsen, J.G.S., Madsen, M.S., Rauch, A., Traynor, S., Van Hauwaert, E.L., Haakonsson, A.K., Javierre, B.M., Hyldahl, M., Fraser, P., and Mandrup, S. (2020). Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nature genetics*.
- Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature protocols* 11, 1455-1476.
- Majic, P., and Payne, J.L. (2020). Enhancers Facilitate the Birth of De Novo Genes and Gene Integration into Regulatory Networks. *Molecular biology and evolution* 37, 1165-1178.
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* 1, 7.
- Malek, T.R. (2008). The biology of interleukin-2. *Annual review of immunology* 26, 453-479.
- Mancino, A., Termanini, A., Barozzi, I., Ghisletti, S., Ostuni, R., Prosperini, E., Ozato, K., and Natoli, G. (2015). A dual cis-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev* 29, 394-408.
- Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends in plant science* 24, 1102-1125.
- Maricque, B.B., Dougherty, J.D., and Cohen, B.A. (2017). A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic acids research* 45, e16.
- Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., *et al.* (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun* 6, 10069.
- Mathieu, E.L., Belhocine, M., Dao, L.T., Puthier, D., and Spicuglia, S. (2014). [Functions of lncRNA in development and diseases]. *Medecine sciences : M/S* 30, 790-796.
- Matsushita, H., Hosoi, A., Ueha, S., Abe, J., Fujieda, N., Tomura, M., Maekawa, R., Matsushima, K., Ohara, O., and Kakimi, K. (2015). Cytotoxic T lymphocytes block tumor growth both by lytic activity and IFN γ -dependent cell-cycle arrest. *Cancer Immunol Res* 3, 26-36.
- Mattioli, K., Volders, P.J., Gerhardinger, C., Lee, J.C., Maass, P.G., Mele, M., and Rinn, J.L. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome research* 29, 344-355.
- Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541-554.
- McCarty, N.S., Graham, A.E., Studena, L., and Ledesma-Amaro, R. (2020). Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat Commun* 11, 1281.
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A., and O'Garra, A. (2015). Type I interferons in infectious disease. *Nat Rev Immunol* 15, 87-103.
- Medina-Rivera, A., Santiago-Algarra, D., Puthier, D., and Spicuglia, S. (2018). Widespread Enhancer Activity from Core Promoters. *Trends Biochem Sci*.
- Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *Elife* 8.
- Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* 12, R113.

- Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A., *et al.* (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* *551*, 92-94.
- Michalska, A., Blaszczyk, K., Wesoly, J., and Bluysen, H.A.R. (2018). A Positive Feedback Amplifier Circuit That Regulates Interferon (IFN)-Stimulated Gene Expression and Controls Type I and Type II IFN Responses. *Front Immunol* *9*, 1135.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., *et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics* *47*, 598-606.
- Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* *32*, 42-57.
- Miranda, T.B., and Jones, P.A. (2007). DNA methylation: the nuts and bolts of repression. *J Cell Physiol* *213*, 384-390.
- Mitchelmore, J., Grinberg, N.F., Wallace, C., and Spivakov, M. (2020). Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters. *Nucleic acids research* *48*, 2866-2879.
- Mittendorf, E.A., Philips, A.V., Meric-Bernstam, F., Qiao, N., Wu, Y., Harrington, S., Su, X., Wang, Y., Gonzalez-Angulo, A.M., Akcakanat, A., *et al.* (2014). PD-L1 expression in triple-negative breast cancer. *Cancer Immunol Res* *2*, 361-370.
- Miyoshi, T., Hiratsuka, K., Saiz, E.G., and Morizane, R. (2020). Kidney organoids in translational medicine: Disease modeling and regenerative medicine. *Developmental dynamics : an official publication of the American Association of Anatomists* *249*, 34-45.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* *60*, 174-182.
- Mojica, F.J., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular microbiology* *9*, 613-621.
- Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proc Natl Acad Sci U S A* *110*, 20693-20698.
- Montgomery, K.T., Tardiff, J., Reid, L.M., and Krauter, K.S. (1990). Negative and positive cis-acting elements control the expression of murine alpha 1-protease inhibitor genes. *Molecular and cellular biology* *10*, 2625-2637.
- Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome research* *27*, 246-258.
- Morera, L., Lubbert, M., and Jung, M. (2016). Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin Epigenetics* *8*, 57.
- Mortier, E., Advincula, R., Kim, L., Chmura, S., Barrera, J., Reizis, B., Malynn, B.A., and Ma, A. (2009). Macrophage- and dendritic-cell-derived interleukin-15 receptor alpha supports homeostasis of distinct CD8+ T cell subsets. *Immunity* *31*, 811-822.
- Muerdter, F., Boryn, L.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., *et al.* (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* *15*, 141-149.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* *13*, 919-922.

- Nafee, T.M., Farrell, W.E., Carroll, W.D., Fryer, A.A., and Ismail, K.M. (2008). Epigenetic control of fetal gene expression. *BJOG : an international journal of obstetrics and gynaecology* *115*, 158-168.
- Nakato, R., and Sakata, T. (2020). Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* (San Diego, Calif).
- Natarajan, A., Yardimci, G.G., Sheffield, N.C., Crawford, G.E., and Ohler, U. (2012). Predicting cell-type-specific gene expression from regions of open chromatin. *Genome research* *22*, 1711-1722.
- Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. *Immunity* *33*, 12-24.
- Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annual review of genetics* *46*, 1-19.
- Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., *et al.* (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* *489*, 83-90.
- Netea, M.G., Balkwill, F., Chonchol, M., Cominelli, F., Donath, M.Y., Giamarellos-Bourboulis, E.J., Golenbock, D., Gresnigt, M.S., Heneka, M.T., Hoffman, H.M., *et al.* (2017). A guiding map for inflammation. *Nature immunology* *18*, 826-831.
- Neumayr, C., Pagani, M., Stark, A., and Arnold, C.D. (2019). STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Curr Protoc Mol Biol* *128*, e105.
- Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *The EMBO journal* *31*, 522-533.
- Nguyen, T.A., Jones, R.D., Snavely, A., Pfenning, A., Kirchner, R., Hemberg, M., and Gray, J.M. (2016a). High-throughput functional comparison of promoter and enhancer activities. *Genome research*.
- Nguyen, T.A., Jones, R.D., Snavely, A.R., Pfenning, A.R., Kirchner, R., Hemberg, M., and Gray, J.M. (2016b). High-throughput functional comparison of promoter and enhancer activities. *Genome research* *26*, 1023-1033.
- Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* *161*, 526-540.
- Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* *169*, 930-944 e922.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* *485*, 381-385.
- Novo, C.L., Javierre, B.M., Cairns, J., Segonds-Pichon, A., Wingett, S.W., Freire-Pritchett, P., Furlan-Magaril, M., Schoenfelder, S., Fraser, P., and Rugg-Gunn, P.J. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell reports* *22*, 2615-2627.
- Oh, S., Oh, C., and Yoo, K.H. (2017). Functional roles of CTCF in breast cancer. *BMB reports* *50*, 445-453.
- Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature reviews Genetics* *12*, 283-293.
- Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. *Cell* *133*, 1145-1148.
- Orlando, G., Law, P.J., Cornish, A.J., Dobbins, S.E., Chubb, D., Broderick, P., Litchfield, K., Hariri, F., Pastinen, T., Osborne, C.S., *et al.* (2018). Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nature genetics* *50*, 1375-1380.

- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev* *10*, 2657-2683.
- Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell* *152*, 157-171.
- Pancaldi, V., Carrillo-de-Santa-Pau, E., Javierre, B.M., Juan, D., Fraser, P., Spivakov, M., Valencia, A., and Rico, D. (2016). Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol* *17*, 152.
- Parsi, K.M., Hennessy, E., Kearns, N., and Maehr, R. (2017). Using an Inducible CRISPR-dCas9-KRAB Effector System to Dissect Transcriptional Regulation in Human Embryonic Stem Cells. *Methods Mol Biol* *1507*, 221-233.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., *et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* *473*, 394-397.
- Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M., *et al.* (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* *30*, 265-270.
- Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., and Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* *27*, 1173-1175.
- Paulsen, J., Sekelja, M., Oldenburg, A.R., Barateau, A., Briand, N., Delbarre, E., Shah, A., Sorensen, A.L., Vigouroux, C., Buendia, B., *et al.* (2017). Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol* *18*, 21.
- Pegram, H.J., Andrews, D.M., Smyth, M.J., Darcy, P.K., and Kershaw, M.H. (2011). Activating and inhibitory receptors of natural killer cells. *Immunol Cell Biol* *89*, 216-224.
- Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO journal* *30*, 4198-4210.
- Peng, T., Zhai, Y., Atlasi, Y., Ter Huurne, M., Marks, H., Stunnenberg, H.G., and Megchelenbrink, W. (2020). STARR-seq identifies active, chromatin-masked, and dormant enhancers in pluripotent mouse embryonic stem cells. *Genome Biol* *21*, 243.
- Perner, F., Perner, C., Ernst, T., and Heide, F.H. (2019). Roles of JAK2 in Aging, Inflammation, Hematopoiesis and Malignant Transformation. *Cells* *8*.
- Perng, Y.C., and Lenschow, D.J. (2018). ISG15 in antiviral immunity and beyond. *Nature reviews Microbiology* *16*, 423-439.
- Petrykowska, H.M., Vockley, C.M., and Elnitski, L. (2008). Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. *Genome research* *18*, 1238-1246.
- Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., Polo, J.M., Ford, E., and Lister, R. (2018). A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome research* *28*, 1193-1206.
- Phanstiel, D.H., Van Bortle, K., Spacek, D., Hess, G.T., Shamim, M.S., Machol, I., Love, M.I., Aiden, E.L., Bassik, M.C., and Snyder, M.P. (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Molecular cell* *67*, 1037-1048 e1036.
- Piccolo, V., Curina, A., Genua, M., Ghisletti, S., Simonatto, M., Sabo, A., Amati, B., Ostuni, R., and Natoli, G. (2017). Opposing macrophage polarization programs show extensive epigenomic and transcriptional cross-talk. *Nature immunology* *18*, 530-540.
- Platanias, L.C. (2005). Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol* *5*, 375-386.
- Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., *et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* *515*, 402-405.

- Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic acids research* 39, 7179-7193.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851-1854.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183.
- Qu, J., Ouyang, Z., Wu, W., Li, G., Wang, J., Lu, Q., and Li, Z. (2020). Functions and Clinical Significance of Super-Enhancers in Bone-Related Diseases. *Front Cell Dev Biol* 8, 534.
- Queen, C., and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. *Cell* 33, 741-748.
- Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47, 11 12 11-34.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279-283.
- Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J., Gifford, D.K., and Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. *Nat Biotechnol* 34, 167-174.
- Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380-1389.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. *Nature protocols* 8, 2281-2308.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.
- Reijmerink, N.E., Postma, D.S., Bruinenberg, M., Nolte, I.M., Meyers, D.A., Bleecker, E.R., and Koppelman, G.H. (2008). Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. *J Allergy Clin Immunol* 122, 651-654 e658.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306-2309.
- Rennie, S., Dalby, M., Lloret-Llinares, M., Bakoulis, S., Dalager Vaagenso, C., Heick Jensen, T., and Andersson, R. (2018). Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities. *Nucleic acids research*.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., *et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *NatMethods* 4, 651-657.
- Rodero, M.P., and Crow, Y.J. (2016). Type I interferon-mediated monogenic autoinflammation: The type I interferonopathies, a conceptual overview. *J Exp Med* 213, 2527-2538.
- Rossetto, D., Avvakumov, N., and Cote, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* 7, 1098-1108.

- Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013). Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic acids research* 41, D1040-1046.
- Saber, A., Liu, B., Ebrahimi, P., and Haisma, H.J. (2020). CRISPR/Cas9 for overcoming drug resistance in solid tumors. *Daru : journal of Faculty of Pharmacy, Tehran University of Medical Sciences* 28, 295-304.
- Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., *et al.* (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3, 511-518.
- Saccani, S., and Natoli, G. (2002). Dynamic changes in histone H3 Lys 9 methylation occurring at tightly regulated inducible inflammatory genes. *Genes Dev* 16, 2219-2224.
- Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature reviews Molecular cell biology* 16, 129-143.
- Sakuma, T., Masaki, K., Abe-Chayama, H., Mochida, K., Yamamoto, T., and Chayama, K. (2016). Highly multiplexed CRISPR-Cas9-nuclease and Cas9-nickase vectors for inactivation of hepatitis B virus. *Genes to cells : devoted to molecular & cellular mechanisms* 21, 1253-1262.
- Salje, H., Tran Kiem, C., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., Andronico, A., Hoze, N., Richet, J., Dubost, C.L., *et al.* (2020). Estimating the burden of SARS-CoV-2 in France. *Science* 369, 208-211.
- Sams, D.S., Nardone, S., Getselter, D., Raz, D., Tal, M., Rayi, P.R., Kaphzan, H., Hakim, O., and Elliott, E. (2016). Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. *Cell reports* 17, 2418-2430.
- Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., *et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112, E6456-6465.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature reviews Genetics* 8, 424-436.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11, 783-784.
- Santiago-Argarra, D., Dao, L.T.M., Pradel, L., Espana, A., and Spicuglia, S. (2017). Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res* 6, 939.
- Saprunauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research* 39, 9275-9282.
- Sati, S., and Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* 126, 33-44.
- Savenije, O.E., Kerkhof, M., Reijmerink, N.E., Brunekreef, B., de Jongste, J.C., Smit, H.A., Wijga, A.H., Postma, D.S., and Koppelman, G.H. (2011). Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. *J Allergy Clin Immunol* 127, 750-756 e751-755.
- Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* 396, 311-327.
- Schmidt, S.F., Larsen, B.D., Loft, A., Nielsen, R., Madsen, J.G., and Mandrup, S. (2015). Acute TNF-induced repression of cell identity genes is mediated by NFkappaB-directed redistribution of cofactors from super-enhancers. *Genome research* 25, 1281-1294.
- Schneider, W.M., Chevillotte, M.D., and Rice, C.M. (2014). Interferon-stimulated genes: a complex web of host defenses. *Annual review of immunology* 32, 513-545.
- Schoenborn, J.R., and Wilson, C.B. (2007). Regulation of interferon-gamma during innate and adaptive immune responses. *Adv Immunol* 96, 41-101.

- Schoenfelder, S., Javierre, B.M., Furlan-Magaril, M., Wingett, S.W., and Fraser, P. (2018). Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *Journal of visualized experiments : JoVE*.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132, 887-898.
- Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular cell* 58, 1101-1112.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* 322, 1849-1851.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458-472.
- Shaw, A.E., Hughes, J., Gu, Q., Behdenna, A., Singer, J.B., Dennis, T., Orton, R.J., Varela, M., Gifford, R.J., Wilson, S.J., *et al.* (2017). Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS biology* 15, e2004086.
- Shigaki, D., Adato, O., Adhikari, A.N., Dong, S., Hawkins-Hooker, A., Inoue, F., Juven-Gershon, T., Kenlay, H., Martin, B., Patra, A., *et al.* (2019). Integration of multiple epigenomic marks improves prediction of variant impact in saturation mutagenesis reporter assay. *Human mutation* 40, 1280-1291.
- Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., *et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* 100, 15776-15781.
- Siersbaek, R., Madsen, J.G.S., Javierre, B.M., Nielsen, R., Bagge, E.K., Cairns, J., Wingett, S.W., Traynor, S., Spivakov, M., Fraser, P., *et al.* (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Molecular cell* 66, 420-435 e425.
- Siersbaek, R., Rabiee, A., Nielsen, R., Sidoli, S., Traynor, S., Loft, A., Poulsen, L.C., Rogowska-Wrzęsinska, A., Jensen, O.N., and Mandrup, S. (2014). Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell reports* 7, 1443-1455.
- Siggers, T., and Gordan, R. (2014). Protein-DNA binding: complexities and multi-protein codes. *Nucleic acids research* 42, 2099-2111.
- Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature protocols* 7, 256-267.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics*.
- Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6.
- Smale, S.T. (2009). Nuclear run-on assay. *Cold Spring Harbor protocols* 2009, pdb prot5329.
- Smale, S.T., and Natoli, G. (2014). Transcriptional control of inflammatory responses. *Cold Spring Harbor perspectives in biology* 6, a016261.
- Smith, A.D., Sumazin, P., Xuan, Z., and Zhang, M.Q. (2006). DNA motifs in human and mouse proximal promoters predict tissue-specific expression. *Proc Natl Acad Sci U S A* 103, 6275-6280.
- Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., and Meissner, A. (2017). Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* 549, 543-547.
- Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Graf, S., Huss, M., Keefe, D., *et al.* (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome research* 21, 1757-1767.
- Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature reviews Molecular cell biology* 19, 262-274.

- Sparber, P., Filatova, A., Khantemirova, M., and Skoblov, M. (2019). The role of long non-coding RNAs in the pathogenesis of hereditary diseases. *BMC Med Genomics* 12, 42.
- Stavrou, S., and Ross, S.R. (2015). APOBEC3 Proteins in Viral Immunity. *J Immunol* 195, 4565-4570.
- Steen, H.C., and Gamero, A.M. (2013). STAT2 phosphorylation and signaling. *JAKSTAT* 2, e25790.
- Stormo, G.D. (2013). Modeling the specificity of protein-DNA interactions. *Quant Biol* 1, 115-130.
- Sun, W., Yao, S., Tang, J., Liu, S., Chen, J., Deng, D., and Zeng, C. (2018). Integrative analysis of super enhancer SNPs for type 2 diabetes. *PloS one* 13, e0192105.
- Suzuki, Y., and Sugano, S. (2003). Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol Biol* 221, 73-91.
- Swygert, S.G., and Peterson, C.L. (2014). Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochimica et biophysica acta* 1839, 728-736.
- Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Science advances* 5, eaaw1668.
- Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., *et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS genetics* 7, e1002154.
- Takahashi, H., Lassmann, T., Murata, M., and Carninci, P. (2012). 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nature protocols* 7, 542-561.
- Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178-192.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernet, B., *et al.* (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75-82.
- Tian, P., Wang, J., Shen, X., Rey, J.F., Yuan, Q., and Yan, Y. (2017). Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synth Syst Biotechnol* 2, 219-225.
- Tian, X., Guo, Y., Wang, X., Pei, L., Wang, X., Wu, J., Sun, S., Li, Y., Ning, M., Buonanno, F.S., *et al.* (2020). Serum soluble ST2 is a potential long-term prognostic biomarker for transient ischaemic attack and ischaemic stroke. *Eur J Neurol*.
- Tovey, M.G., Streuli, M., Gresser, I., Gugenheim, J., Blanchard, B., Guymarho, J., Vignaux, F., and Gigou, M. (1987). Interferon messenger RNA is produced constitutively in the organs of normal individuals. *Proc Natl Acad Sci U S A* 84, 5038-5042.
- Tucci, V., Isles, A.R., Kelsey, G., and Ferguson-Smith, A.C. (2019). Genomic Imprinting and Physiological Processes in Mammals. *Cell* 176, 952-965.
- Tyagi, S., Kumar, R., Kumar, V., Won, S.Y., and Shukla, P. (2021). Engineering disease resistant plants through CRISPR-Cas9 technology. *GM crops & food* 12, 125-144.
- Valitutti, S., Carbone, A., Castellino, F., Maggiano, N., Ricci, R., Larocca, L.M., and Musiani, P. (1989). The expression of functional IL-2 receptor on activated macrophages depends on the stimulus applied. *Immunology* 67, 44-50.
- van Arensbergen, J., FitzPatrick, V.D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H.J., and van Steensel, B. (2017). Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol* 35, 145-153.
- van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Vosa, U., Franke, L., *et al.* (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nature genetics* 51, 1160-1169.
- van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., *et al.* (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* 9, 969-972.
- van Ouwkerk, A.F., Bosada, F.M., Liu, J., Zhang, J., van Duijvenboden, K., Chaffin, M., Tucker, N.R., Pijnappels, D., Ellinor, P.T., Barnett, P., *et al.* (2020). Identification of Functional Variant Enhancers Associated With Atrial Fibrillation. *Circulation research* 127, 229-243.

- Vanhille, L., Griffon, A., Maqbool, M.A., Zacarias-Cabeza, J., Dao, L.T., Fernandez, N., Ballester, B., Andrau, J.C., and Spicuglia, S. (2015a). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6, 6905.
- Vanhille, L., Griffon, A., Maqbool, M.A., Zacarias-Cabeza, J., Dao, L.T.M., Fernandez, N., Ballester, B., Andrau, J.C., and Spicuglia, S. (2015b). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6, 6905.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nature reviews Genetics* 10, 252-263.
- Venter, G.J., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., and Holt, R.A. (2001). The sequence of the human genome. *Science* 291, 1304-1351.
- Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. *Critical reviews in biochemistry and molecular biology* 44, 117-141.
- Vermersch, E., Jouve, C., and Hulot, J.S. (2020). CRISPR/Cas9 gene-editing strategies in cardiovascular cells. *Cardiovascular research* 116, 894-907.
- Vernimmen, D., and Bickmore, W.A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends in genetics : TIG* 31, 696-708.
- Vihervaara, A., Mahat, D.B., Guertin, M.J., Chu, T., Danko, C.G., Lis, J.T., and Sistonen, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat Commun* 8, 255.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., *et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854-858.
- Volpi, S., Picco, P., Caorsi, R., Candotti, F., and Gattorno, M. (2016). Type I interferonopathies in pediatric rheumatology. *Pediatr Rheumatol Online J* 14, 35.
- Wakeling, E.L. (2011). Silver-Russell syndrome. *Archives of disease in childhood* 96, 1156-1161.
- Wammers, M., Schupp, A.K., Bode, J.G., Ehling, C., Wolf, S., Deenen, R., Kohrer, K., Haussinger, D., and Graf, D. (2018). Reprogramming of pro-inflammatory human macrophages to an anti-inflammatory phenotype by bile acids. *Sci Rep* 8, 255.
- Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., *et al.* (2018a). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362.
- Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., *et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120-124.
- Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular cell* 19, 631-642.
- Wang, W., Ye, C., Liu, J., Zhang, D., Kimata, J.T., and Zhou, P. (2014). CCR5 gene disruption via lentiviral vectors expressing Cas9 and single guided RNA renders cells resistant to HIV-1 infection. *PloS one* 9, e115987.
- Wang, X., He, L., Goggin, S.M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., and Kellis, M. (2018b). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun* 9, 5380.
- Weingarten-Gabbay, S., Nir, R., Lubliner, S., Sharon, E., Kalma, Y., Weinberger, A., and Segal, E. (2019). Systematic interrogation of human promoters. *Genome research* 29, 171-183.
- West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes Dev* 16, 271-288.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307-319.

- Wijchers, P.J., Krijger, P.H.L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., *et al.* (2016). Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Molecular cell* *61*, 461-473.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., *et al.* (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* *32*, 670-676.
- Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? *Cell* *155*, 990-996.
- Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature reviews Genetics* *18*, 517-534.
- Wurtele, H., and Chartrand, P. (2006). Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* *14*, 477-495.
- Xie, C., Zhang, Y.E., Chen, J.Y., Liu, C.J., Zhou, W.Z., Li, Y., Zhang, M., Zhang, R., Wei, L., and Li, C.Y. (2012). Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. *PLoS genetics* *8*, e1002942.
- Yamagami, M., Otsuka, M., Kishikawa, T., Sekiba, K., Seimiya, T., Tanaka, E., Suzuki, T., Ishibashi, R., Ohno, M., and Koike, K. (2018). ISGF3 with reduced phosphorylation is associated with constitutive expression of interferon-induced genes in aging cells. *NPJ Aging Mech Dis* *4*, 11.
- Yamamoto, T., and Saitoh, N. (2019). Non-coding RNAs and chromatin domains. *Curr Opin Cell Biol* *58*, 26-33.
- Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S., Sugano, S., Nakai, K., and Suzuki, Y. (2011). Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. *Genome research* *21*, 775-789.
- Yang, J., Rajan, S.S., Friedrich, M.J., Lan, G., Zou, X., Ponstingl, H., Garyfallos, D.A., Liu, P., Bradley, A., and Metzakopian, E. (2019). Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming. *Stem Cell Reports* *12*, 757-771.
- Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Molecular cell* *47*, 648-655.
- Yu, H., Boyle, T.A., Zhou, C., Rimm, D.L., and Hirsch, F.R. (2016). PD-L1 Expression in Lung Cancer. *J Thorac Oncol* *11*, 964-975.
- Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* *518*, 556-559.
- Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research* *21*, 1273-1283.
- Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Bluthgen, N., Stadler, M., Tian, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome research* *27*, 479-490.
- Zhang, P., Xia, J.H., Zhu, J., Gao, P., Tian, Y.J., Du, M., Guo, Y.C., Suleman, S., Zhang, Q., Kohli, M., *et al.* (2018). High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. *Nat Commun* *9*, 2022.
- Zhao, G.N., Jiang, D.S., and Li, H. (2015). Interferon regulatory factors: at the crossroads of immunity, metabolism, and disease. *Biochimica et biophysica acta* *1852*, 365-378.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* *322*, 750-756.
- Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh, S.K., Singh, U., *et al.* (2006). Circular chromosome conformation capture (4C) uncovers

- extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature genetics* *38*, 1341-1347.
- Zhen, S., Hua, L., Takahashi, Y., Narita, S., Liu, Y.H., and Li, Y. (2014). In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by CRISPR/Cas9. *Biochemical and biophysical research communications* *450*, 1422-1426.
- Zheng, C.X., Wang, S.M., Bai, Y.H., Luo, T.T., Wang, J.Q., Dai, C.Q., Guo, B.L., Luo, S.C., Wang, D.H., Yang, Y.L., *et al.* (2018). Lentiviral Vectors and Adeno-Associated Virus Vectors: Useful Tools for Gene Transfer in Pain Research. *Anat Rec (Hoboken)* *301*, 825-836.
- Zhou, J.H., Wang, Y.N., Chang, Q.Y., Ma, P., Hu, Y., and Cao, X. (2018a). Type III Interferons in Viral Infection and Antiviral Immunity. *Cell Physiol Biochem* *51*, 173-185.
- Zhou, Q., Chen, V., Shannon, C.P., Wei, X.S., Xiang, X., Wang, X., Wang, Z.H., Tebbutt, S.J., Kollmann, T.R., and Fish, E.N. (2020). Interferon-alpha2b Treatment for COVID-19. *Front Immunol* *11*, 1061.
- Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews Genetics* *12*, 7-18.
- Zhou, W., Dinh, H.Q., Ramjan, Z., Weisenberger, D.J., Nicolet, C.M., Shen, H., Laird, P.W., and Berman, B.P. (2018b). DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics* *50*, 591-602.
- Zhou, X., Cain, C.E., Myrthil, M., Lewellen, N., Michelini, K., Davenport, E.R., Stephens, M., Pritchard, J.K., and Gilad, Y. (2014). Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biol* *15*, 547.
- Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nature reviews Genetics* *17*, 551-565.
- Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* *138*, 1122-1136.

ANNEX 1

Review

Widespread Enhancer Activity from Core Promoters

Alejandra Medina-Rivera,¹ David Santiago-Algarra,^{2,3} Denis Puthier,^{2,3} and Salvatore Spicuglia^{2,3,*}

Gene expression in higher eukaryotes is precisely regulated in time and space through the interplay between promoters and gene-distal regulatory regions, known as enhancers. The original definition of enhancers implies the ability to activate gene expression remotely, while promoters entail the capability to locally induce gene expression. Despite the conventional distinction between them, promoters and enhancers share many genomic and epigenomic features. One intriguing finding in the gene regulation field comes from the observation that many core promoter regions display enhancer activity. Recent high-throughput reporter assays along with clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-related approaches have indicated that this phenomenon is common and might have a strong impact on our global understanding of genome organisation and gene expression regulation.

Highlights

Promoters and enhancers share architectural and functional properties.

When tested on episomal reporters, many promoters display enhancer activity.

In vivo experiments demonstrated that enhancer like promoters function as *bona fide* enhancers.

Genetic variants lying in enhancer-like promoters might impact on physiological traits or diseases by altering the expression of distal genes.

Similarities between Enhancers and Promoters

The regulation of gene transcription in higher eukaryotes is accomplished through the involvement of **transcription start site** (TSS)-proximal (promoters) and -distal (enhancers) regulatory elements [1,2]. The classical distinction between **enhancers** (see [Glossary](#)) and **promoters** generally relies on their location with respect to the 5' end of genes and the enrichment of specific histone modifications. From a functional point of view, an enhancer implies the property of activating a distal promoter, independently of location and orientation with respect to the target genes. In contrast, promoters must be able to initiate transcription locally and induce efficient transcription elongation toward the direction of the gene. However, this basic dichotomy of *cis*-regulatory elements has been challenged by broad similarities between genetic and epigenetic properties of promoters and enhancers and has been the topic of several recent reviews [3–6] ([Table 1](#)).

Like promoters, active enhancers are bound by RNA polymerase II (RNAPII) and general transcription factors (GTFs), and transcribe noncoding **enhancer RNAs** (eRNAs) [7–12]. Promoters and enhancers are demarcated by divergent transcription initiation and a well-positioned array of surrounding nucleosomes [7,10,13]. While enhancers are generally depleted of CpG islands, they recruit master regulators like CpG-poor promoters [7] and are enriched in **core promoter** elements [10]. Histone modifications have been commonly used to discriminate between enhancers and promoters [14–16]. For instance, enhancers are enriched in monomethylation of histone H3 Lys4 (H3K4me1) and acetylation of histone H3 Lys27 (H3K27ac). In contrast, gene promoters typically exhibit trimethylated H3K4 (H3K4me3). As a consequence, the presence of H3K27ac accompanied by high levels of H3K4me1 and low H3K4me3 have been used as a proxy for active enhancers [17]. However, recent work has demonstrated that the presence of H3K4me3 is fully compatible with **enhancer activity** [10,11,18–20]; the level of H3K4me3 being actually positively correlated with the enhancer

¹Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Mexico

²Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France

³Equipe Labélisée, Ligue Contre le Cancer, Paris, France

*Correspondence: salvatore.spicuglia@inserm.fr (S. Spicuglia).

Table 1. Features Associated with Active Promoters and Enhancers

Features (active elements)	Promoter	Enhancer
Intrinsic property	Induce transcription of a heterologous reporter gene	Activate a distal (heterologous) promoter
Transcription initiation	Unidirectional or divergent	Mainly divergent
Ratio between sense and antisense transcripts	Biased toward sense transcription	Equilibrated
Transcription elongation	Produce long polyadenylated transcripts	Some enhancers can produce low levels of polyadenylated transcripts
Histone modifications	H3K27ac (H3K4me1 < H3K4me3)	H3K27ac (H3K4me1 > H3K4me3)
RNAPII and GTF	Present	Present
GpG islands	Majority	Very rare

strength and eRNA level [7,10,12,21]. Thus, the current view postulates that similar regulatory mechanisms are at play at enhancers and promoters, but differences in H3K4 methylation patterns simply reflect differences in transcription levels between the two types of elements.

Besides the shared architectural characteristics between promoters and enhancers, some promoter elements have been shown to function as enhancers in ectopic enhancer reporter assays and to form long-range contacts with other promoters [4,22]. However, whether this fraction of promoters could function as distal-acting enhancers *in vivo* has remained unclear. More recently, high-throughput functional screens and *in vivo* genetic experiments have highlighted the commonality and physiological functions of these enhancer-like promoters, also referred as **ePromoters** (see below). In the present review, we describe the different evidence for the existence of enhancer-like promoters and discuss whether they might define a new type of regulatory elements, the implications for the understanding of complex gene regulation in normal development and disease, as well as, for the topological organisation of the genome.

Initial Evidence of Enhancer Activity from Promoters

Initial characterisation of enhancer elements from the early 1980s consisted in isolating DNA sequences able to stimulate transcription of a heterologous promoter using episomal reporter assays [23,24]. For instance, the first identified enhancer by Schaffner and collaborators in 1981 corresponded to the promoter of a simian virus 40 (SV40) early gene [25]. They showed that a 72-repeat sequence motif was sufficient to increase expression of ectopic β -globin gene by 200-fold and to function over long distances in an orientation-independent fashion relative to the β -globin gene.

It is worth noting that many of the early characterised enhancers are located close to, or overlapping with, the promoter region of inducible genes, such as metallothioneins, histones of early cleavage stages, viral immediate-early genes (from some papovaviruses, cytomegaloviruses, and retroviruses), heat-shock genes, and the antiviral interferon genes [24] (Table 2). A characteristic example is the *IFN β* enhancer, which is one of the most well-studied enhancers [26]. Although located immediately upstream of the *IFN β* gene, it can also function as a classical enhancer element conferring virus-infection-dependent activation of heterologous promoters, even when it is placed kilobases away from the targeted promoter [27,28]. Enhancer activity of the *IFN β* promoter depends on loop formation mediated by critical sequence-specific

Glossary

Core promoter: short sequence of ~50 bp that serves as a binding platform for the transcriptional machinery consisting of RNAPII and its associated GTFs.

Enhancer RNA (eRNA): noncoding RNAs produced by the enhancers. They are generally non-polyadenylated, low in abundance, unspliced, and retained within the nucleus.

Enhancer/promoter activity: this makes reference to any functional experiment that assesses the propensity of a given regulatory element to act as an enhancer or promoter.

Enhancers: regulatory elements that activates transcription over large distances and independently of orientation. These *cis*-regulatory elements are generally located distally with respect to the 5' end of genes.

ePromoter: promoter element that displays enhancer activity in a functional experimental setting.

Expression quantitative trait locus (eQTL): genetic polymorphisms whose alleles are associated with gene expression variability.

Genome-wide association study (GWAS): A GWAS is intended to detect genomic variants that are found to be associated with a trait or disease.

Promoters: regulatory element capable of inducing gene expression. These *cis*-regulatory elements are generally located in close proximity to the 5' end of genes.

Transcription factories: describe the discrete sites where transcription occurs in the nucleus. The factories contain RNA polymerase (under active or inactive status) and the necessary transcription factors (activators and repressors) for transcription.

Transcription start site (TSS): defines the nucleotide position of any transcription initiation event. However, it generally refers to the position of the main 5' end of an mRNA.

Table 2. Individual Examples of Enhancer Activity from Promoter Elements

Gene	Origin	Size (bp)	Distance from TSS	Refs
<i>Early gene</i>	SV40	196	~200	[25,95]
<i>Early gene</i>	Cytomegalovirus	406	–524 to –118	[96]
<i>Hsp70</i>	<i>Xenopus</i>	160	–260 to –100	[97]
<i>Fos</i>	Human	340	–404 to –64	[98]
<i>hMT-IIA</i>	Human	327	–366 to –39	[99]
<i>Mmt-IA</i>	Mouse	114 155	–187 to –73 –194 to –39	[99]
<i>H2A</i>	Urchin	28	–139 to –111	[100]
<i>IFNβ</i>	Human	40	–77 to –37	[28]

transcription factors bound to the regulatory sequences [29]. A more recent study reported that a promoter located upstream of the adeno-associated virus type 2 (AAV2) genome also display liver-specific enhancer activity; a finding that might explain the pathogenic association between AAV2 integration events and human hepatocellular carcinoma through insertional dysregulation of cancer driver genes via enhancer-mediated effects [30].

A common characteristic of most of the aforementioned promoters is that they are associated with inducible genes that have to quickly respond to environmental stress, which might take more time or be less efficient with a remote enhancer [24]. These early studies already highlighted that enhancers and promoters are very similar entities with some gene promoters having the intrinsic properties to work as enhancers and raised the possibility that enhancer-like promoters could regulate distal genes in their natural context.

Promoter–Promoter Interactions Suggest Distal Regulation by Gene Promoters

Mammalian genomes are intricately and dynamically organised into higher-order conformation inside the micron-sized nuclear space [31]. Such 3D organisation of the genome is thought to have a role in the mechanisms of transcription regulation and coordination by mediating dynamic looping between distantly located *cis*-regulatory elements while enabling fine-tuning of gene expression. The development of different molecular methods for capturing the spatial

Box 1. Genome Topology and 3C-based Approaches

Interacting genomic regions can be identified by 3C and its derivative methods, which involve crosslinking distal interacting DNA pieces, proximity ligation, and sequencing to map the interactions ([32] and references therein). Variations of 3C can focus on interactions for a small number of genomic bait regions (4C), interactions within specific genomic domains (5C), or analyse the whole set of chromosomal interactions within a cell population (Hi-C). Since the HiC technique requires high sequencing coverage, alternative methods have been developed to allow exploration of the contacts of a subset of genomic regions, with higher resolution at the same cost. Chromatin interaction analysis by ChIA-PET [34] or HiChIP [40] considers only those interactions that are mediated by a protein of interest by pulling down by ChIP only the interacting fragments that include this protein. Other capture approaches have been developed that enable selective enrichment for genome-wide interactions involving, on one end, specific regions of interest; these included capture Hi-C (CHI-C) [37,38] and HiCap [74]. In these later approaches, promoter elements or DNase hypersensitive sites are generally captured using sequence-specific beads, thus providing a comprehensive view of genomic regions interacting with *cis*-regulatory elements. A major finding of these studies is that the genome contains regions that are defined by high levels of chromatin interactions occurring within a domain, interspersed with genomic regions with fewer interactions. These regions are generally referred to as topologically associating domains (TADs), and studies have shown that their borders are conserved across mammalian cell types and even across mammalian species [32].

organisation of the genome (Box 1), such as chromosome conformation capture (3C) and related techniques has provided an unprecedented view of the 3D organisation of the genome as well as the spatial resolution of interacting regions [31,32].

Besides the expected interactions between distal enhancers and promoters of target genes, several observations have led to the notion that promoters participate in long-range regulation of distal genes through promoter–promoter (P-P) interactions. Different 3C-based methods such as 3C carbon copy (5C) [33], chromatin interaction analysis with paired-end-tag sequencing (ChIA-PET) [34–36], promoter capture Hi-C (CHi-C) [37–39], or HiChIP [40] have revealed extensive P-P interactions. In fact, based on promoter capture Hi-C approaches, P-P interactions represent ~30% of all promoter-centred interactions [41], suggesting that this particular type of multigene regulatory networks is common in mammalian cells.

In general, promoters contact other promoters with similar expression levels [34,36,38], indicating that 3D contacts between promoters are nonrandom. Therefore, promoter interaction networks may facilitate the coordinated expression control of associated genes and allow for regulatory crosstalk between them. Within this hypothesis it is plausible that a fraction of these P-P interactions represent a more specific regulatory circuitry, whereby a given promoter might regulate the activity of distal neighbour genes. Epigenetic analyses of P-P interactions identified by RNAPII based ChIA-PET experiments revealed a strong bias toward higher H3K4me1/me3 ratio [34], thus suggesting potential enhancer-like activity for a fraction of interacting promoters. In this study, two promoters involved in P-P interactions were shown to function as enhancers of the other associated promoter by luciferase reporter assays.

High-throughput Reporter Assays Highlight Frequent Enhancer Activity from Promoter Elements

In recent years, various powerful techniques that incorporate high-throughput sequencing into reporter assays have enabled quantitative and straightforward measurements of enhancer activity of thousands of regulatory elements [42] (Box 2). In particular, two approaches have been widely used in recent years: massively parallel reporter assay (MPRA) and self-transcribing active regulatory region sequencing (STARR-seq). One interest of high-throughput enhancer assays is the possibility to explore enhancer function without preconceived notions, thus potentially leading to new unforeseen findings. Indeed one intriguing and recurrent observation of several episomal assays is that many core promoter regions display enhancer activity [22,42–50].

Using STARR-seq, Zabidi *et al.* screened the whole fly genome with the use of different core promoters from either ubiquitously expressed housekeeping genes or developmentally regulated and cell-type-specific genes [44]. They found that promoter–proximal enhancers mainly regulate promoters of housekeeping genes, while promoters of developmental genes required distally located enhancers. Several independent studies in mammals also reported widespread enhancer activity from TSS-proximal regions. Ernst *et al.* assessed the enhancer activity of a large selection of DNase I hypersensitivity sites (DHSs) across several human cell lines and found that a significant subset of active enhancers overlap the TSS of genes [51]. Nguyen *et al.* performed a functional comparison of a subset of promoters and enhancers in mouse neurons using an integrative MPRA approach [45]. Gene promoters and distal regulatory regions generated similar enhancer activity. By performing STARR-seq on enriched targets, we found that TSS-proximal and distal DHSs were similarly enriched for active enhancers [46]. Further systematic assessment of all human core promoters of coding genes demonstrated that 2–3% of promoters displayed enhancer activity in a given cell line [46], and these promoters were

Box 2. High-throughput Reporter Assays

Episomal reporter assays have been widely used to characterise putative regulatory regions. Several high-throughput strategies have been developed, enabling the simultaneous analysis of hundreds of thousands of reporter plasmids at once (Figure 1). In particular, MPRA and STARR-seq have been widely used in recent years. The MPRA method consists of the generation of a library of reporter constructs based on microarray synthesis of DNA sequences (generally, tested sequences are cloned upstream of a basal promoter) and unique sequence tags or barcodes [placed in the 3' untranslated region (UTR) of the reporter gene]. To increase the sensitivity and reproducibility, several barcodes could be added to any given sequence. The reporter library is then transfected into cell lines of interest and RNA sequencing of the barcodes is performed, thus providing a quantitative readout of the regulatory activity of the tested regions. STARR-seq is a massively parallel reporter assay [93] aimed to identify and quantify transcriptional enhancers directly based on their activity across whole genomes. A bulk of DNA fragments from arbitrary sources is cloned downstream of a core promoter and into the 3' UTR of a GFP reporter gene. Once in cellular context, active enhancers will activate the promoter and transcribe themselves resulting in reporter transcripts among cellular RNAs. Thus, each reporter transcript contains the reporter gene and the barcode of itself. These reporter transcripts can be isolated separately by targeted PCR and eventually detected by deep sequencing. The main advantage over the classical MPRA is that the tested sequence itself is used as a barcode, substantially simplifying the whole procedure of quantifying the enhancer activity. Capture-based approaches can be used to enrich for particular region of interest. For recent reviews on these methods, see [2,42].

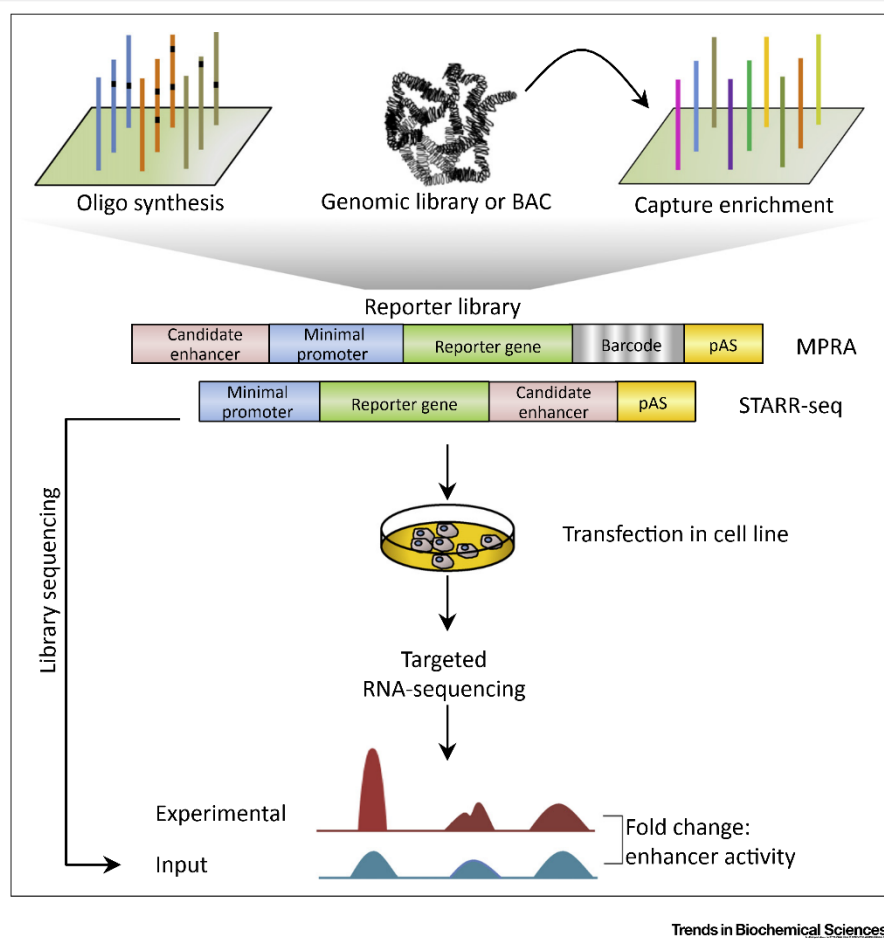


Figure 1. High-throughput Reporter Strategies to Assess Enhancer Activity.

denoted ePromoters. Consistent with these results, two recent whole genome STARR-seq studies performed in human cancer cell lines, LNCaP and HeLa, found that between 650 and 1000 of functionally identified enhancers overlapped a TSS [47,48], representing 1% and 6% of all active enhancers detected in the respective cell lines.

High-throughput reporter assays have several intrinsic caveats that might over- or underestimate the actual number of promoters with enhancer-like activity [2,42]. These caveats include, the size of the tested fragments, the heterologous promoters used in the assays, and the fact that candidate enhancers are studied outside their endogenous chromatin context, which is likely required for their *in vivo* function.

Another potential concern is that the enhancer activity in the reporter assays actually reflects intrinsic properties of the promoter (e.g., acting as a hotspot for the recruitment of transcription factors), which does not necessarily imply enhancer activity *in vivo*. Certainly, an equally valid argument is that episomal reporter assays allow us to study enhancer function without bias, independently of any perturbing chromatin or genomic context. In any case, it would be interesting to systematically assess enhancer activity from gene promoters using chromatinised episomal or virus-based high-throughput reporter assays [45,52–54].

In vivo Assessment of Distal Gene Regulation by Promoter Elements

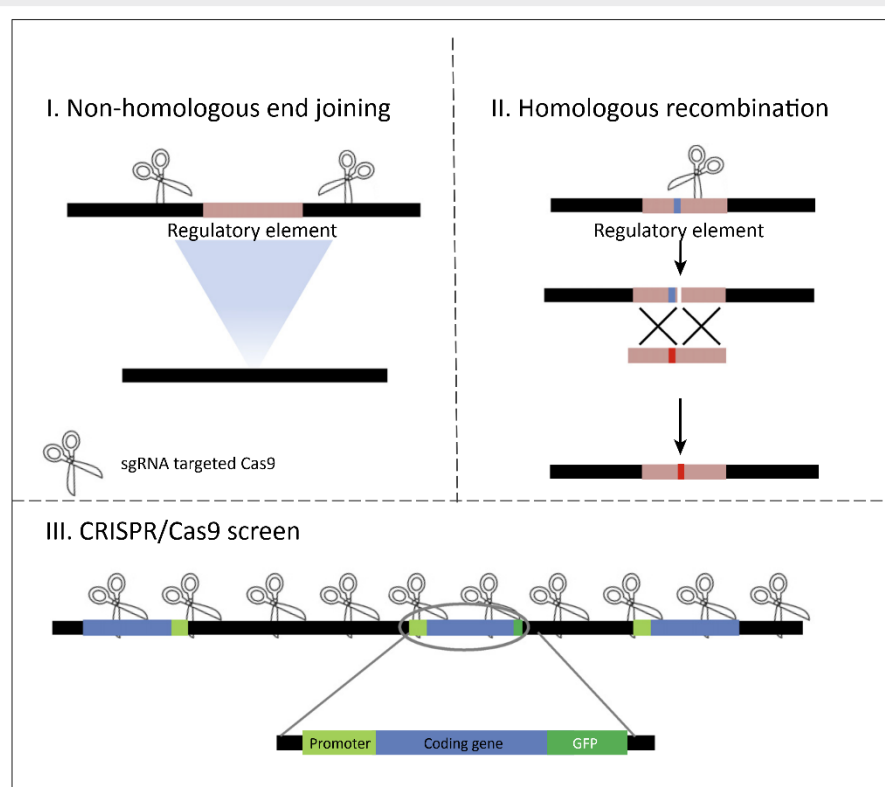
As mentioned above, the fact that some promoters might display enhancer capacity, when tested in episomal reporter assays, does not necessarily imply that they could influence other promoters *in vivo*. Therefore, a critical issue is whether gene promoters are able to function as *bona fide* enhancers by regulating distal gene expression in their endogenous context. A pioneer study showed that one enhancer of the α -globin locus located within the intron of the *Nbl1* gene harbours intrinsic **promoter activity** and induces the expression of a noncoding isoform [55]; however, the physiological function of this noncoding transcript remains elusive.

The advent of CRISPR/Cas9 genome editing methods allows us to systematically study the role of *cis*-regulatory elements in their endogenous context [56,57] (Box 3). Several independent studies using CRISPR genome editing demonstrated that some promoters function as enhancers in their endogenous context (Figure 1A and Table 3). Using a CRISPR/Cas9-based promoter deletion strategy, we showed that selected promoters of coding genes with enhancer activity identified in a human STARR-seq reporter assay (i.e., ePromoters), are indeed involved in *cis*-regulation of distal gene expression in their natural context, therefore functioning as *bona fide* enhancers [46]. These ePromoters were shown to physically interact with the promoters of the regulated genes, in some cases involving several target genes, implying that in these P-P interactions, one promoter acts as an active regulatory element of the others. Inversion of one of the model promoters still retained significant enhancer activity, suggesting that, like classical distal enhancers, enhancer-like promoters might display orientation independent enhancer activity.

Engreitz and colleagues performed systematic genomic editing of promoters of long noncoding RNAs (lncRNAs) co-regulated with neighbour coding genes. Out of 12 deleted lncRNA promoters, five resulted in significant reduction in the expression of the associated neighbour gene [58]. Further genetic manipulation of the loci by inserting a polyadenylation site downstream the promoter of the lncRNA, thus blocking transcription without affecting the integrity of the promoter, demonstrated that regulation of the target genes does not require the specific lncRNA transcripts themselves, but instead involves enhancer-like activity of the lncRNA promoters [58]. Another study found similar results for the promoter of a lncRNA located downstream of the *Cdkn1b* gene [59]. Nevertheless, as for the α -globin locus mentioned

Box 3. CRISPR/Cas9-based Approach to Study *cis*-Regulatory Elements

Since its discovery, CRISPR/Cas9 technology has been widely used for genome editing. This method permits us to target genome DNA using a small RNA fragment (sgRNA). The Cas9 enzyme recognises the sgRNA/DNA complex and cuts the DNA, triggering the DNA repair system of the cell. This strategy can help to study the *cis*-regulatory elements in their natural context (Figure I): (i) deletion of a *cis*-regulatory element by nonhomologous end-joining repair using two sgRNA flanking the regulatory region of interest [46,58]; (ii) CRISPR-mediated mutagenesis permits us to create single base mutations by the homologous recombination repair system using a sgRNA targeting the *cis*-regulatory element [94] and a donor template containing the mutation; and (iii) genomic tile-deletion screening using multiple pairs of sgRNA to identify *cis*-regulatory elements of any gene fused with a reporter marker, such as the GFP [60].

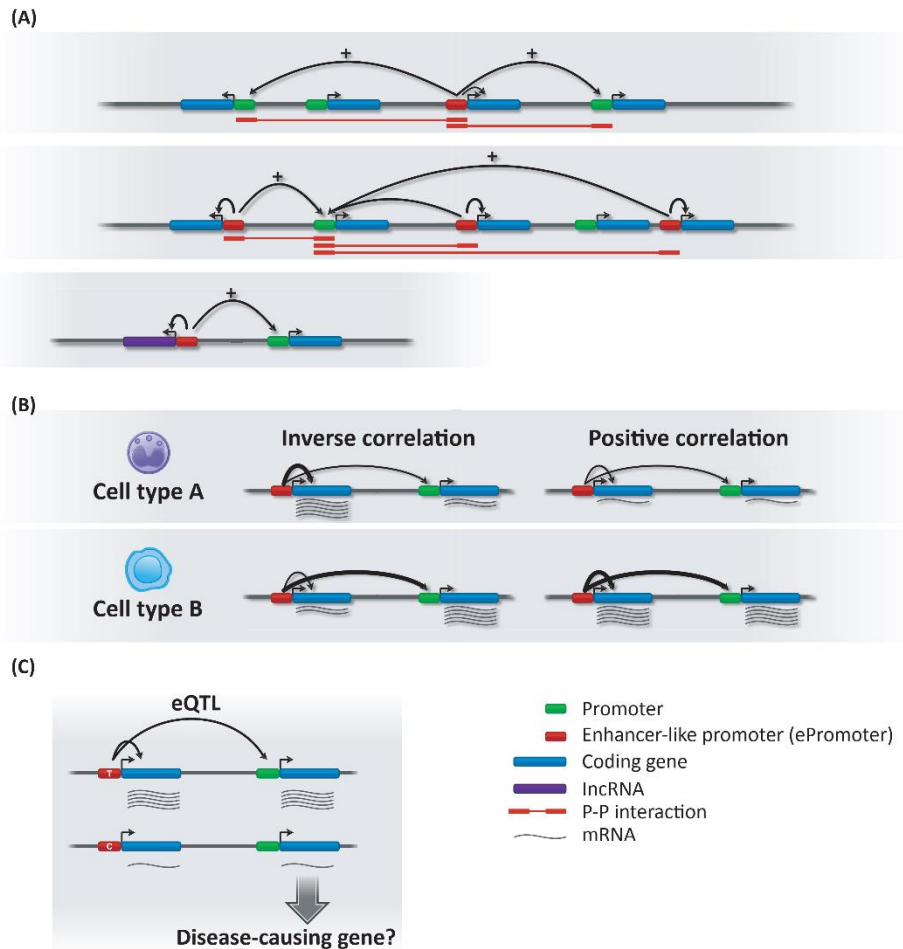


Trends in Biochemical Sciences

Figure I. CRISPR/Cas9-based Approaches to Study *cis*-Regulatory Elements.

above, it is difficult to ascertain whether the tested regulatory element is a functional promoter of the lncRNA or rather a distal enhancer associated with a long eRNA.

The CRISPR/Cas9 approach has been implemented to assess enhancer function within large genomic regions surrounding a given gene of interest [42,56]. In these studies, a reporter gene introduced at the place of the target gene is used to monitor gene expression. Then, a tiling single guide RNA (sgRNA) library covering the surrounding genomic regions is screened to identify deleted regions with potential enhancer elements. Two independent studies performing such screens of *cis*-regulatory elements also found that the expression of some genes is controlled, at least partially, by distal gene promoters [60,61] (Table 3). In particular, interrogation of a 2 Mb genomic region surrounding the *POU5F1* locus, using a high-throughput tiling-deletion strategy in human embryonic stem cells identified 45 sequences regulating *POU5F1* expression in *cis* [60]. Of these, 17 sequences corresponded to promoters of functionally



Trends in Biochemical Sciences

Figure 1. Role of Enhancer-like Promoters in Gene Regulation. (A) ePromoter (red) interacts with one or more distal promoters (green) and activates the expression of neighbour genes (top). A given gene might be regulated by several ePromoters located in the neighbourhood (middle). Promoters of lncRNAs (purple) can also have enhancer-like activity and positively regulate the expression of a nearby gene (bottom). (B) The enhancer and promoter activities of ePromoters could be dissociated (inverse correlation); in this case the same regulatory element displays enhancer activity in one cell type and promoter activity in another cell type. In contrast, the enhancer and promoter activities could be linked (positive correlation); in this case the ePromoter exhibits both enhancer and promoter activities in the same cell type. The later model might result in the coordinated regulation of neighbour genes upon stress or cell-type specific signalling. (C) Genetic variants (e.g. eQTL or genome-wide association study single nucleotide polymorphism) lying within an ePromoter might influence the expression of neighbour genes. It is plausible that the physiological impact (trait or disease) of the variant could rely on the deregulation of a distal gene. Abbreviations: ePromoter, enhancer-like promoter; eQTL, expression quantitative trait locus; lncRNA, long noncoding RNA; P-P, promoter-promoter.

unrelated genes. Fourteen of 17 *POU5F1*-regulating promoters had significant level of chromatin interactions with the *POU5F1* promoter, confirming that enhancer-like activity of promoters require long-range chromatin interactions.

An alternative strategy to assess enhancer activity in the endogenous context is to use a nuclease-deactivated Cas9 (dCas9) fused to an activator or repressor domain to precisely modify gene expression from promoters or distal regulatory elements [57]. By using this

Table 3. List of Experimentally Validated Promoters with Enhancer-like Activity in Their Natural Context

Strategy	Validation	Cell type or line	Gene associated with the enhancer-like promoter (target gene)	Refs
Characterisation of DHS associated with α -globin locus	Knock-out mice	Mouse erythrocytes	<i>Npr13</i> (α -globin)	[55]
Co-regulated genes	CRISPR deletion and pAS insertion	mESC	<u><i>Bend4</i></u> ^a ; <i>Slc30a9</i> (<i>Bend4</i>) <i>Snhg171</i> (<i>Snhg11</i>) <i>Linc14051</i> (<i>Snhg11</i>) <u><i>Linc1405</i></u> (<i>Eomes</i>) <i>Gpr19</i> (<i>Cdkn1b</i>)	[58]
Transgenic reporter	Reporter assay	Mouse cardiomyocytes	<i>Nppb</i> (<i>Nppa</i>)	[88]
CRISPR screening (MERA)	None	mESC	<i>Lrrc2</i> (<i>Tdgf1</i>)	[61]
Co-regulated genes	CRISPR deletion pAS Insertion	G1E	<i>Lockd</i> (<i>Cdkn1b</i>)	[59]
Reporter assay (CapSTARR-seq)	CRISPR deletion	HeLa K562	<i>FAF2</i> (<i>RNF44</i>) <i>TAGLN2</i> (<i>PIGM</i> ; <i>PEA15</i>) <i>CSDE1</i> (<i>BCAS2</i> ; <i>SIKE1</i>) <i>BAZ2 B</i> (<i>MARCH7</i>) <i>YPEL4</i> (<i>UBE2L6</i>) <i>METTL21A</i> (<i>CCNYL1</i>)	[46]
CRISPR screening (CREST-seq)	CRISPR deletion	hESC	17 promoters (<i>POU5F1</i>)	[60]
HiChIP	CRISPRa	Jurkat	<i>CLEC2D</i> ; <i>CLEC2 B</i> (<i>CD69</i>)	[40]

^aPromoters of lncRNAs are underlined.

approach, another study assessed the functional relevance of two heterologous promoters interacting with the promoter of the T cell inducible gene *CD69* and demonstrated that these distal promoters indeed regulate the expression of *CD69* after T cell activation [40].

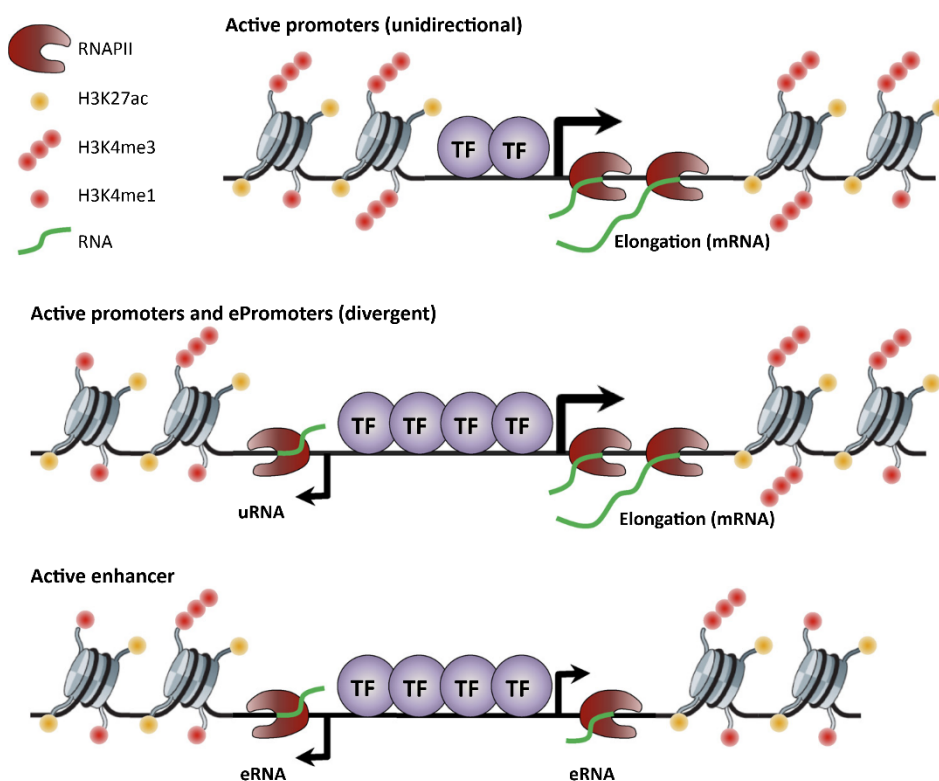
Features of Enhancer-like Promoters

It is clear that not all gene promoters display enhancer activity. For instances, in the Engreitz *et al.* study only a subset of tested promoters had significant enhancer activity [58]. Similarly, in the Dao *et al.* study, while the ePromoter of the *FAF2* gene is required for the expression of *RNF44* gene, deletion of the *RNF44* promoter did not have any impact on *FAF2* expression [46]. Therefore, what defines enhancer-like promoters and what are the underlying characteristics that entail their enhancer function? First of all, enhancer-like promoters appear to be preferentially associated with housekeeping and stress response genes, including interferon response genes [44,46,48,49]. Consistently, a study in *Drosophila* using random insertion of reporter constructs found that expression of the reporter gene depends on chromosomal contacts with endogenous promoters of housekeeping genes [62], suggesting that promoters of housekeeping genes might influence the expression of neighbour loci.

In comparison to classical promoters and distal enhancers, the enhancer-like promoters (ePromoters) display distinct genomic and epigenomic features. They differ in motif content, transcription factor binding and histone modifications [45,46,48]. Indeed, enhancer-like promoters bind higher levels of p300, a cofactor usually associated with active enhancers [17] and display increased ratio of H3K27ac over H3K4me3 [46], this ratio correlating with enhancer activity in different cell lines. Consistent with housekeeping and stress response functions, the enhancer-like promoters are preferentially bound by general inducible transcription factors

such as AP1, STAT, and the ATF/CREB family of transcription factors [45,46,48]. High-throughput reporter assays using synthetic sequences with tandem repeats of DNA motifs assessed the intrinsic properties of transcription factor binding sites to display promoter or enhancer activities [45]. The study found that distinct DNA motifs were required for either type of activity. For example, the presence of the AP1 motif resulted in significant enhancer activity, but little promoter activity, while motifs for EGR, CREB, and RFX families of transcription factors generated preferential promoter activity. Thus, it is plausible that within the same regulatory sequence different motifs might provide specific enhancer or promoter functions. Another striking feature of enhancer-like promoters is that they harbour a higher density of distinct motifs and bound transcription factors, key properties shared with distal enhancers [63].

The advent of high-throughput sequencing has allowed us to map transcription initiation with an unprecedented sensitivity and resolution [5]. This has revealed that *cis*-regulatory elements are commonly associated with transcriptional initiation sites flanking the regulatory sequences (Figure 2). Promoters can be associated with either unidirectional or bidirectional transcription,



Trends in Biochemical Sciences

Figure 2. Chromatin Structure of Active Regulatory Elements. Unidirectional promoters (top) have a main TSS (arrow) and are associated with high levels of H3K4me3 and H3K27ac. Bidirectional promoters (middle) have two unbalanced TSSs defining a larger promoter region than unidirectional promoters and allow the recruitment of a higher number of transcription factors. They are also associated with H3K4me3 and H3K27ac, but the upstream region is also enriched in H3K4me1. The ePromoters belong to this category. Active enhancers (bottom) have two balanced TSSs, produced eRNAs in both direction and are enriched for all three histone marks. Abbreviations: ePromoter, enhancer-like promoter; eRNA, enhancer RNA; H3K4me1, monomethylation of histone H3 Lys4; H3K4me3, trimethylated H3K4; H3K27ac, acetylation of histone H3 Lys27; RNAPII, RNA polymerase II; TF, transcription factor; TSS, transcription start site; uRNA, upstream RNA.

in the latter the signal intensity being biased toward the sense of the gene. Enhancers produce RNAs (eRNAs) *in vivo* [8,9,11] with an initiation and chromatin architecture similar to that of promoters [7,10,12,64]. In particular, enhancers have been shown to generally produce bidirectional unstable transcripts with no particular orientation bias. While the functional relevance of eRNAs is not fully understood, it is clear that their relative abundance is positively correlated with enhancer activity [7,12,64].

In macrophages, promoters highly induced during the immune challenge are characterised by the presence of divergent transcription initiation in which the sense and antisense TSSs are separated by large distances [65]. This in turn correlates with enlarged nucleosome-depleted regions and enhancer-like features such as higher transcription factor occupancy, binding of p300 and high level of H3K4me1 (Figure 2, middle panel). Thus, the size of the nucleosome-depleted region in bidirectional promoters appears to contribute toward enhancer-like properties. Reminiscent of these findings, promoters with enhancer activity are predominantly associated with bidirectional transcription [46]. Similarly, testing gene promoters for enhancer activity in *Drosophila* embryos revealed that when bidirectionally transcribed, promoters could function as enhancer *in vivo*, while unidirectional promoters generally cannot [64]. Overall, these results point toward an unifying model whereby there is a continuum of *cis*-regulatory activity with some elements acting strictly as either enhancer or promoter, while others function predominantly as an enhancer with weak promoter activity or vice versa, yet others can have both strong promoter and enhancer activities [4–6,10,64] (Figure 2). This spectrum of activities might be highly correlated with the directionality of transcription, which likely reflects the underlying sequence properties. In this context, bidirectional transcription at enhancer-like promoters might provide enlarged nucleosome depleted regions serving as hubs for transcription factor binding and establishment of highly active chromatin to further regulate or enhance proximal and distal gene expression (Figure 2, middle panel). This would be particularly relevant in the case of rapid and coordinated regulation of gene expression in response to environmental or intrinsic cellular stimuli.

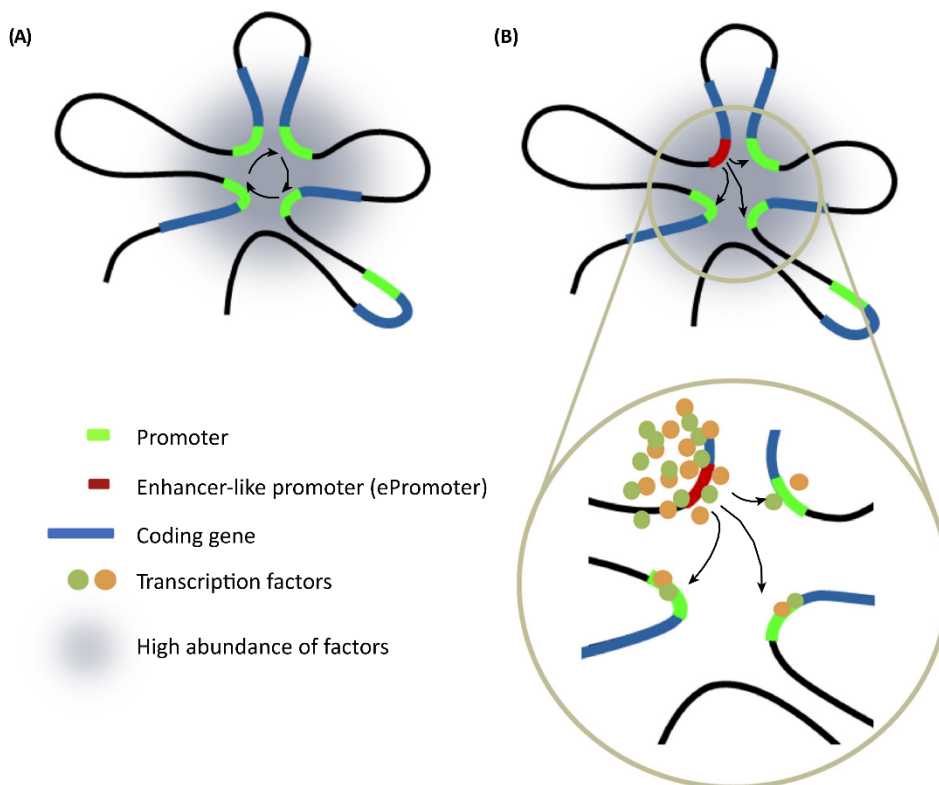
Another outstanding question is whether promoter and enhancer activities of enhancer-like promoters are correlated (Figure 1B). Nguyen *et al.* compared the enhancer and promoter activities of defined promoter elements using distinct reporter assays. They observed a clear positive correlation between enhancer and promoter activity [45]. Similarly, a recent study developed a transgenic assay in *Drosophila* embryos with dual vectors that simultaneously assesses the ability of elements to function as an enhancer and a promoter *in vivo* [64]. Some of the tested promoters harboured concomitant promoter and enhancer activity. Comparison of enhancer activity of STARR-seq defined ePromoters with the expression level of the associated gene (as a proxy of the promoter activity) did not show a strict correlation [46]. However, some of the ePromoters displayed high levels of both promoter and enhancer activity, whereas for others ePromoters, both activities were anticorrelated. Consistently, integrative analysis of epigenomes across human tissues revealed that a given genomic region could have epigenetic features of enhancer or promoter in different tissues, suggesting that the type of regulatory activity (i.e., enhancer or promoter) might be tissue specific [66]. Therefore, it is plausible that depending on the locus, enhancer-like promoters might either coordinate the mRNA expression of clusters of genes (for instance, upon stress response signalling) or display context-dependent enhancer or promoter activities (Figure 1B).

As it could be expected, enhancer-like promoters interact with the promoters of regulated genes [40,46,60]. The frequency of P-P interactions is higher when the interaction involves at least one enhancer-like promoter [46]. This suggests that one of the properties defining

enhancer-like promoter might be to favour P-P interactions, likely by recruiting key transcription factors such as ZNF143 or YY1, which are two factors involved in looping [67,68] and enriched at enhancer-like promoters [46]. However, in a given cell type, the number of promoters involved in P-P interactions surpass the number of enhancer-like promoters that can be found in the same cells [46]. It is therefore likely that not all P-P interactions require an enhancer-like promoter. Alternatively, it is possible that not all enhancer-like promoters are detected by the enhancer reporter assays. Finally, whether enhancer-like promoters represent a hub of interactions with multiple genes need to be explored in the future.

Promoter-centred Transcription Factories

The expression of interacting genes within multigene complexes is generally well correlated, suggesting that 3D gene organisation contributes to coordination of gene expression programmes. Evidence from *in situ* fluorescence studies in the last decade suggests that transcription is not evenly distributed and is instead concentrated within large discrete foci in mammalian nuclei, raising the possibility that genes are organised into **transcription factories** containing RNAPII and other components for transcription [69] (Figure 3A). In the current model of transcription factories, regulatory regions of neighbour genes are clustered together and contribute to the expression of each other by increasing the local concentration of regulatory



Trends in Biochemical Sciences

Figure 3. Model of ePromoters and Gene Regulation. (A) Chromatin interactions place promoters in close physical proximity (transcription factories), facilitating the recruitment of transcription factors and RNAPII necessary for the transcription of their associated genes. (B) The presence of an ePromoter inside the transcription factory could favour the recruitment of high levels of transcription factors and RNAPII. Abbreviations: ePromoter, enhancer-like promoter; RNAPII, RNA polymerase II.

factors and RNAPs that might form non-membrane-bound compartments with transcription activating and repressing microenvironments [70]. Such clustering has been reported for NF- κ B-regulated genes in response to TNF- α stimulation [71]. Experimental removal of a gene from the NF- κ B-dependent multigene complex was shown to directly affect the transcription of its interacting genes, suggesting that co-association of co-regulated genes might contribute to a hierarchy of gene expression control [72]. Building up on the transcription factory model, Hinisz and collaborators recently proposed a phase separation model for transcriptional control, whereby clusters of enhancers and promoters mediate multimolecular assemblies of protein–nucleic acid complexes providing a general regulatory mechanism to compartmentalise membraneless nuclear compartments [73]. However, the precise contribution of enhancer-like promoters within these transcription factories is currently unknown.

As mentioned above, the widespread occurrence of P-P interactions suggests that promoter-centred chromatin structure contributes to the 3D organisation of the genome and has provided a structural framework for the postulated transcription factories [34]. Indeed, the P-P interactions appear to define a subset of co-regulated promoters sharing genomic and structural regulatory properties, which may be critical for stabilising the local 3D interactions and the activity of transcription factories. For instance, compared to the interactions between enhancer and promoters, the P-P interactions form a higher-order chromatin structure involving many loci, have highly coordinated expression, and are more resistant to external changes [34,37,38,74–76]. In these promoter-centred transcription factories, promoter-interacting multigene clusters might represent topological units of transcriptional coordination where co-regulated genes might come to close vicinity by P-P interactions, resulting in an optimal stoichiometry of chromatin factors required for modulation of gene expression (Figure 3A). The interacting regions can be established or maintained by chromatin bridging proteins such as cohesins and CTCF, which are enriched at the interacting promoters [35,41,77].

Given the overall contribution of enhancer-like promoters to the regulation of neighbour genes [40,46,58,60], as well as the intrinsic features described in the previous section (frequently involved in P-P interactions; high density of transcription factor binding, etc.), it is tempting to speculate that this type of promoters might play a key role within the transcription factories (Figure 3B). In this model, the enhancer-like promoters could either facilitate the assembly or maintenance of the transcription factories by tightening the P-P interactions or bring specific transcriptional regulators required for the regulation of the neighbour genes. In any case, it will be essential to investigate the specific contribution of enhancer-like promoters to the functioning of transcription factories.

Genetic Variation within Promoters Influence Distal Gene Regulation

One of the major endeavours in genomic research in the past decade was the advent of **genome-wide association studies** (GWASs) in order to identify genetic variants associated with candidate genes for human diseases. Most of these variants are located in noncoding regions [78,79], hence are more likely to be modifying gene expression regulatory mechanisms [2,80]. It is possible that genetic variants outside coding regions play a regulatory role, but the target genes of these variants are difficult to identify, in particular when the location of the hit is far away from the neighbouring genes. Regardless of this, most GWAS studies establish plausible causality mechanisms by selecting the closest gene to the associated variant, especially when the variant lies within an intronic region, or in the vicinity of a TSS. However, this assumption has been proven to be biased in several examples [81,82]. In a similar way, it might be envisioned that GWAS variants lying within enhancer-like promoters might regulate the expression of distal disease-causing genes.

While GWAS-reported genetic variants are not easily connected to effects on gene function, genetic polymorphisms can be associated with gene expression variability, these variants are known as **expression quantitative trait loci** (eQTLs). eQTLs with the higher probability to be causal of gene expression variation, tend to be located in open chromatin regions, such as promoters and enhancers [83], supporting the hypothesis of a possible effect through changes in gene expression regulatory mechanisms. Using the set of ePromoters defined in Dao *et al.* [46], we observed that it is more likely to find an eQTL associated with the expression of a distal gene within an ePromoter as compared to other promoters. Given the functional characteristics of eQTLs it is possible to use the reported effect (β value) of the eQTL as a proxy of the effect a variant could have on its putative target genes. eQTLs lying within ePromoters tend to have stronger effects on distal gene expression than those in other promoters. Moreover, eQTLs potentially affecting transcription factor binding within ePromoters were biased toward having a positive effect on distal gene expression. Specifically, allelic replacement using CRISPR/Cas9 homologous recombination (Box 3) of the reference eQTL allele of two of these ePromoters recapitulated the regulatory function of the eQTL variant in the regulation of distal gene expression.

Several examples from the literature might point toward the relation between disease-associated variants and disrupted regulatory mechanisms. The type 2 diabetes associated variant rs11603334 lies within the *ARAP1* promoter and affects PAX6/PAX4 binding in human pancreatic islets [84]. The *ARAP1* promoter displayed enhancer activity in STARR-seq assays [46], and the rs11603334 variant is reported in the Genotype-Tissue Expression (GTEx; <http://www.gtexportal.org>) database as an eQTL affecting both *ARAP1* and *PDE2* genes; the latter was already suspected by Kulzer *et al.* to be of possible relevance for type 2 diabetes. The *NPPB-NPPA* cluster is associated with several cardiovascular diseases and multiple GWAS variants have been reported within the *NPPB* promoter [85–87]. Functional analysis of double-reporter transgenic mice revealed that the *Nppb* promoter is required for heart hypertrophy-induced *Nppa* expression [88], raising the possibility that the causal mechanism of *NPPB*-promoter variants might be due to dysregulation of both *NPPB* and *NPPA* mRNAs. Mumbach *et al.* [40] integrated 3D genome wide interaction maps in primary human cells to identify regulatory connectomes linking intergenic mutations to target genes. One of the identified interactions mapped to the rs56375023 and rs17293632 variants associated with Crohn's disease and lying within a *SMAD3* alternative promoter. Interestingly, this *SMAD3* promoter interacts with another, more upstream, *SMAD3* promoter as well as the *AAGAB* promoter, while functional association was supported by eQTL data.

Besides genetic variants, other types of genomic alterations such as enhancer hijacking by chromosomal translocation, genomic rearrangement or insulator disruption, are common molecular mechanisms resulting in disease-related gene deregulation, including overexpression of oncogenes [89,90]. It is likely expected that enhancer-like promoters could impact on disease through related mechanisms. Integrating information about enhancer-like promoters (e.g., using high-throughput reporter assays) along with 3D interaction data, eQTLs, and disease-associated variants (e.g., GWASs) might lead to the discovery of disease-associated regulation by distal promoters (Figure 1C).

Another way distal promoter regulation might have pathological relevance is by indirect perturbation of genome topology. For instances, Cornelia de Lange syndrome (CdLS) is a complex multisystem developmental disorder caused by mutations in cohesin subunits and regulators [91]. Some of the genes deregulated in CdLS are not directly associated with cohesin subunits but are positioned within reach of cohesin-occupied regions through P-P

interactions [92], suggesting that wide gene expression deregulation rely on enhancer-like function of cohesin-bound promoters.

Concluding Remarks

Overall, the reviewed results reveal the commonality and widespread use of promoters as distal enhancers. Furthermore, these findings extend and support the increasing evidences pointing toward a unified model of transcriptional regulation, highlighting broad similarities between enhancers and promoters [3,4,6,10]. Although several of these regulatory elements have been validated *in vivo*, more systematic studies using CRISPR/Cas9-based technology will be needed to assess the actual proportion of promoters functioning as *bona fide* enhancers. For instances, recent developments combining CRISPR/Cas9 screening and single-cell RNA-seq [44], thus enabling high-throughput interrogation of enhancers at single cell resolution and directly linking enhancer function with its target genes might help to provide a more comprehensive view of enhancer-like promoters function in living cells. Whether this phenomenon uncovers nonspecific contribution of promoters to gene regulation (e.g., keeping open chromatin structure or a defined 3D topology) or rather a specific enhancer-like activity (defining new types of regulatory elements; i.e., ePromoters), will require further investigations.

These findings also open up the intriguing possibility that developmental traits or disease-associated variants lying within a subset of promoters might directly impact on distal gene expression. While there is already work to be done on the understanding of the molecular mechanisms that govern the enhancer-like activity from promoters in cell type or response specific regulatory systems (see Outstanding Questions), the ePromoters concept stresses the fact that the identification of regulatory variant target genes in the context of disease is not a straightforward task, and the door should remain open for new association studies and more complex regulatory networks than previously foreseen.

Acknowledgements

We thank Mauricio Guzmán Araiza for help in figure design. Work in S.S.'s laboratory was supported by recurrent funding from the Inserm and Aix-Marseille University and by ARC (PJA 20151203149), Plan Cancer (P036496) and 'Equipe Labellisée Ligue Contre le Cancer' grants. A.M.-R.'s laboratory is supported by CONACYT (269449) and 'Programa de Apoyo a Proyectos de investigación e innovación tecnológica' – Universidad Nacional Autónoma de México (PAPIIT-UNAM) (IA206517) grants. This review was written with the support of the SEP-CONACYT-ANUIES-ECOS-NORD project M17S02. D.S.A. was a fellow of CONACYT-Mexico.

References

- Vermimmen, D. and Bickmore, W.A. (2015) The hierarchy of transcriptional activation: from enhancer to promoter. *Trends Genet.* 31, 696–708
- Chatterjee, S. and Ahituv, N. (2017) Gene regulatory elements, major drivers of human disease. *Annu. Rev. Genomics Hum. Genet.* 18, 45–63
- Kim, T.K. and Shiekhattar, R. (2015) Architectural and functional commonalities between enhancers and promoters. *Cell* 162, 948–959
- Andersson, R. (2015) Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays* 37, 314–323
- Andersson, R. *et al.* (2015) A unified architecture of transcriptional regulatory elements. *Trends Genet.* 31, 426–433
- Tippen, N.D. *et al.* (2018) Enhancer transcription: what, where, when, and why? *Genes. Dev.* 32, 1–3
- Andersson, R. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461
- De Santa, F. *et al.* (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8, e1000384
- Kim, T.K. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187
- Core, L.J. *et al.* (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* 46, 1311–1320
- Koch, F. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18, 956–963
- Henriques, T. *et al.* (2018) Widespread transcriptional pausing and elongation control at enhancers. *Genes. Dev.* 32, 26–41
- He, H.H. *et al.* (2010) Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* 42, 343–347
- Heintzman, N.D. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112

Outstanding Questions

What are the specific components within the promoter region driving promoter versus enhancer activity?

Are promoter and enhancer activities correlated across different tissues?

Do ePromoter–promoter interactions rely on similar mechanisms as previously shown for enhancer–promoter interactions?

Are enhancer-like promoters a hub of P-P interactions?

Are enhancer-like promoters involved in particular biological processes?

Is the enhancer activity of promoters dependent on the genomic context?

Is the regulation by enhancer-like promoters a specific process or rather an unspecific contribution to gene expression within transcription factories?

Is enhancer activity from promoters evolutionary conserved? Could enhancer-like promoters be associated with evolutionarily new genes originated from distal enhancer elements?

What are the contributions of enhancer-like activity of promoters to disease?

15. Creighton, M.P. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936
16. Rada-Iglesias, A. *et al.* (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283
17. Heintzman, N.D. and Ren, B. (2009) Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.* 19, 541–549
18. Wang, Z. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903
19. Ernst, J. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49
20. Pekowska, A. *et al.* (2011) H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* 30, 4198–4210
21. Vanhille, L. *et al.* (2015) High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.* 6, 6905
22. Catarino, R.R. *et al.* (2017) Promoting transcription over long distances. *Nat. Genet.* 49, 972–973
23. Marriott, S.J. and Brady, J.N. (1989) Enhancer function in viral and cellular gene regulation. *Biochim. Biophys. Acta* 989, 97–110
24. Schaffner, W. (2015) Enhancers, enhancers – from their discovery to today’s universe of transcription enhancers. *Biol. Chem.* 396, 311–327
25. Banerji, J. *et al.* (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308
26. Thanos, D. and Maniatis, T. (1995) Virus induction of human IFN-gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100
27. Fan, C.M. and Maniatis, T. (1989) Two different virus-inducible elements are required for human beta-interferon gene regulation. *EMBO J.* 8, 101–110
28. Goodbourn, S. *et al.* (1985) Human beta-interferon gene expression is regulated by an inducible enhancer element. *Cell* 41, 509–520
29. Nolis, I.K. *et al.* (2009) Transcription factors mediate long-range enhancer-promoter interactions. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20222–20227
30. Logan, G.J. *et al.* (2017) Identification of liver-specific enhancer-promoter activity in the 3′ untranslated region of the wild-type AAV2 genome. *Nat. Genet.* 49, 1267–1273
31. Mishra, A. and Hawkins, R.D. (2017) Three-dimensional genome architecture and emerging technologies: looping in disease. *Genome Med.* 9, 87
32. Sati, S. and Cavalli, G. (2017) Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* 126, 33–44
33. Sanyal, A. *et al.* (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109–113
34. Li, G. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98
35. Handoko, L. *et al.* (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638
36. Kieffer-Kwon, K.R. *et al.* (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507–1520
37. Mifsud, B. *et al.* (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606
38. Schoenfelder, S. *et al.* (2015) The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25, 582–597
39. Javierre, B.M. *et al.* (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167, 1369–1384 e1319
40. Mumbach, M.R. *et al.* (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612
41. Pancaldi, V. *et al.* (2016) Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol.* 17, 152
42. Santiago-Algarra, D. *et al.* (2017) Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res.* 6, 939
43. Arnold, C.D. *et al.* (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–1077
44. Zabidi, M.A. *et al.* (2015) Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559
45. Nguyen, T.A. *et al.* (2016) High-throughput functional comparison of promoter and enhancer activities. *Genome Res.* 26, 1023–1033
46. Dao, L.T.M. *et al.* (2017) Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* 49, 1073–1081
47. Liu, Y. *et al.* (2017) Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol.* 18, 219
48. Muerdter, F. *et al.* (2018) Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* 15, 141–149
49. Barakat, T.S. *et al.* (2017) Functional dissection of the enhancer repertoire in human embryonic stem cells. *bioRxiv* <http://dx.doi.org/10.1101/146696>
50. Wang, X. *et al.* (2017) High-resolution genome-wide functional dissection of transcriptional regulatory regions in human. *bioRxiv* <http://dx.doi.org/10.1101/193136>
51. Ernst, J. *et al.* (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat. Biotechnol.* 34, 1180–1190
52. Shen, S.Q. *et al.* (2016) Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* 26, 238–255
53. Murtha, M. *et al.* (2014) FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat. Methods* 11, 559–565
54. Inoue, F. *et al.* (2017) A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* 27, 38–52
55. Kowalczyk, M.S. *et al.* (2012) Intragenic enhancers act as alternative promoters. *Mol. Cell* 45, 447–458
56. Montalbano, A. *et al.* (2017) High-throughput approaches to pinpoint function within the noncoding genome. *Mol. Cell* 68, 44–59
57. Lo, A. and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Res.* 6, <http://dx.doi.org/10.12688/f1000research.11113.1> eCollection2017
58. Engreitz, J.M. *et al.* (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455
59. Paralkar, V.R. *et al.* (2016) Unlinking an lncRNA from its associated cis element. *Mol. Cell* 62, 104–110
60. Diao, Y. *et al.* (2017) A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* 14, 629–635
61. Rajagopal, N. *et al.* (2016) High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* 34, 167–174
62. Corrales, M. *et al.* (2017) Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome Res.* 27, 1153–1161



63. Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* 13, 469–483
64. Mikhaylichenko, O. *et al.* (2018) The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* 32, 42–57
65. Scruggs, B.S. *et al.* (2015) Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* 58, 1101–1112
66. Leung, D. *et al.* (2015) Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 518, 350–354
67. Whalen, S. *et al.* (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* 48, 488–496
68. Weintraub, A.S. *et al.* (2017) YY1 is a structural regulator of enhancer-promoter loops. *Cell* 171, 1573–1588 e1528
69. Rieder, D. *et al.* (2012) Transcription factories. *Front. Genet.* 3, 221
70. Feuerborn, A. and Cook, P.R. (2015) Why the activity of a gene depends on its neighbors. *Trends Genet.* 31, 483–490
71. Papanonis, A. *et al.* (2010) Active RNA polymerases: mobile or immobile molecular machines? *PLoS Biol.* 8, e1000419
72. Fanucchi, S. *et al.* (2013) Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606–620
73. Hnisz, D. *et al.* (2017) A phase separation model for transcriptional control. *Cell* 169, 13–23
74. Sahlen, P. *et al.* (2015) Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 16, 156
75. Zhu, Y. *et al.* (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.* 7, 10812
76. Barbieri, M. *et al.* (2017) Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 24, 515–524
77. Merkschlager, M. (2010) Cohesin: a global player in chromosome biology with local ties to gene regulation. *Curr. Opin. Genet. Dev.* 20, 555–561
78. MacArthur, J. *et al.* (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res.* 45, D896–D901
79. Maurano, M.T. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195
80. Deplancke, B. *et al.* (2016) The genetics of transcription factor DNA binding variation. *Cell* 166, 538–554
81. Claussnitzer, M. *et al.* (2015) FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373, 895–907
82. Gupta, R.M. *et al.* (2017) A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* 170, 522–533 e515
83. Battle, A. *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature* 550, 204–213
84. Kulzer, J.R. *et al.* (2014) A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am. J. Hum. Genet.* 94, 186–197
85. Del Greco, M.F. *et al.* (2011) Genome-wide association analysis and fine mapping of NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB gene cluster. *Hum. Mol. Genet.* 20, 1660–1671
86. Ellis, K.L. *et al.* (2011) Association of genetic variation in the natriuretic peptide system with cardiovascular outcomes. *J. Mol. Cell. Cardiol.* 50, 695–701
87. Fox, A.A. *et al.* (2009) Natriuretic peptide system gene variants are associated with ventricular dysfunction after coronary artery bypass grafting. *Anesthesiology* 110, 738–747
88. Sergeeva, I.A. *et al.* (2016) Identification of a regulatory domain controlling the Nppa-Nppb gene cluster during heart development and stress. *Development* 143, 2135–2146
89. Bradner, J.E. *et al.* (2017) Transcriptional addiction in cancer. *Cell* 168, 629–643
90. Smith, E. and Shilatifard, A. (2014) Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* 21, 210–219
91. Boyle, M.I. *et al.* (2015) Cornelia de Lange syndrome. *Clin. Genet.* 88, 1–12
92. Boudaoud, I. *et al.* (2017) Connected gene communities underlie transcriptional changes in Cornelia de Lange syndrome. *Genetics* 207, 139–151
93. Muerdter, F. *et al.* (2015) STARR-seq – Principles and applications. *Genomics* 106, 145–150
94. Canver, M.C. *et al.* (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192–197
95. Benoist, C. and Chambon, P. (1981) *In vivo* sequence requirements of the SV40 early promoter region. *Nature* 290, 304–310
96. Boshart, M. *et al.* (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41, 521–530
97. Bienz, M. and Pelham, H.R. (1986) Heat shock regulatory elements function as an inducible enhancer in the *Xenopus* hsp70 gene and when linked to a heterologous promoter. *Cell* 45, 753–760
98. Deschamps, J. *et al.* (1985) Identification of a transcriptional enhancer element upstream from the proto-oncogene fos. *Science* 230, 1174–1177
99. Serfling, E. *et al.* (1985) Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO J.* 4, 3851–3859
100. Grosschedl, R. and Birnstiel, M.L. (1982) Delimitation of far upstream sequences required for maximal *in vitro* transcription of an H2A histone gene. *Proc. Natl. Acad. Sci. U. S. A.* 79, 297–301

ANNEX 2



REVIEW

Recent advances in high-throughput approaches to dissect enhancer function [version 1; peer review: 3 approved]

David Santiago-Algarra, Lan T.M. Dao, Lydie Pradel, Alexandre España , Salvatore Spicuglia 

Aix-Marseille University, TAGC, Marseille, France

v1 **First published:** 19 Jun 2017, 6(F1000 Faculty Rev):939 (<https://doi.org/10.12688/f1000research.11581.1>)
Latest published: 19 Jun 2017, 6(F1000 Faculty Rev):939 (<https://doi.org/10.12688/f1000research.11581.1>)

Abstract

The regulation of gene transcription in higher eukaryotes is accomplished through the involvement of transcription start site (TSS)-proximal (promoters) and -distal (enhancers) regulatory elements. It is now well acknowledged that enhancer elements play an essential role during development and cell differentiation, while genetic alterations in these elements are a major cause of human disease. Many strategies have been developed to identify and characterize enhancers. Here, we discuss recent advances in high-throughput approaches to assess enhancer activity, from the well-established massively parallel reporter assays to the recent clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based technologies. We highlight how these approaches contribute toward a better understanding of enhancer function, eventually leading to the discovery of new types of regulatory sequences, and how the alteration of enhancers can affect transcriptional regulation.

Keywords

gene transcription, enhancer function, MPRA, STARR-seq, CRISPR

Open Peer Review

Reviewer Status 

	Invited Reviewers		
	1	2	3
version 1 published 19 Jun 2017			

F1000 Faculty Reviews are written by members of the prestigious F1000 Faculty. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

- 1 **Alvaro Rada-Iglesias**, University of Cologne, Cologne, Germany
- 2 **Douglas Vernimmen**, University of Edinburgh, Edinburgh, UK
- 3 **Bart Deplancke**, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Any comments on the article can be found at the end of the article.

Corresponding author: Salvatore Spicuglia (salvatore.spicuglia@inserm.fr)

Competing interests: The authors declare that they have no competing interests.

Grant information: Work in Salvatore Spicuglia's laboratory was supported by recurrent funding from the Inserm and Aix-Marseille University and by the ARC (PJA 20151203149) and the Plan Cancer (P036496) grants. David Santiago-Algarra and Lan T.M. Dao were supported by the CONACYT-Mexico and Vietnam International Education Development, respectively.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2017 Santiago-Algarra D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Santiago-Algarra D, Dao LTM, Pradel L *et al.* **Recent advances in high-throughput approaches to dissect enhancer function [version 1; peer review: 3 approved]** F1000Research 2017, 6(F1000 Faculty Rev):939 (<https://doi.org/10.12688/f1000research.11581.1>)

First published: 19 Jun 2017, 6(F1000 Faculty Rev):939 (<https://doi.org/10.12688/f1000research.11581.1>)

Introduction

Gene expression is precisely regulated by a combination of promoters and gene-distal regulatory regions, known as enhancers^{1,2}. With the increasing awareness of the important role of enhancers in normal development as well as in disease, there is strong scientific interest in identifying and characterizing these elements. This is a challenging task because an enhancer does not have to be located directly adjacent to the gene it regulates. Putative enhancers can be identified across entire genomes based on open chromatin regions (e.g. based on DNase I-seq or ATAC-seq) or chromatin signatures (H3K4me1, H3K27ac), which map the potentially active enhancers³. Although useful, these approaches do not provide direct proof of enhancer function, nor do they allow insights into the discrete sequences required for enhancer activity. Therefore, it is crucial to test whether genomic regions actually function as *bona fide* enhancers in living cells or tissues.

In recent years, various powerful techniques that incorporate high-throughput sequencing into reporter assays have enabled quantitative and straightforward measurements of enhancer activity

of thousands of regulatory elements. More recently, the advent of clustered regularly interspaced short palindromic repeats (CRISPR)-related approaches allows massively assessing the relevance of enhancer function in the endogenous context. This review summarizes the assays developed for functional genome-wide testing of enhancer activity and their limitations as well as the main findings that have been gathered using these techniques.

Principle of high-throughput reporter assays

Episomal reporter assays have been widely used to characterize putative regulatory regions. Several high-throughput strategies have been developed, permitting the simultaneous analysis of hundreds of thousands of reporter plasmids at once. These have been the focus of several comprehensive reviews (e.g. 4–6). These methods can be either qualitative (usually based on cell sorting) or quantitative (based on RNA-seq) and designed to test enhancer or promoter activity. Here, we will focus on recent quantitative methods aiming to characterize enhancers. In particular, two approaches have been widely used in recent years (Figure 1; Table 1): massively parallel reporter assay (MPRA) and self-transcribing active regulatory region sequencing (STARR-seq).

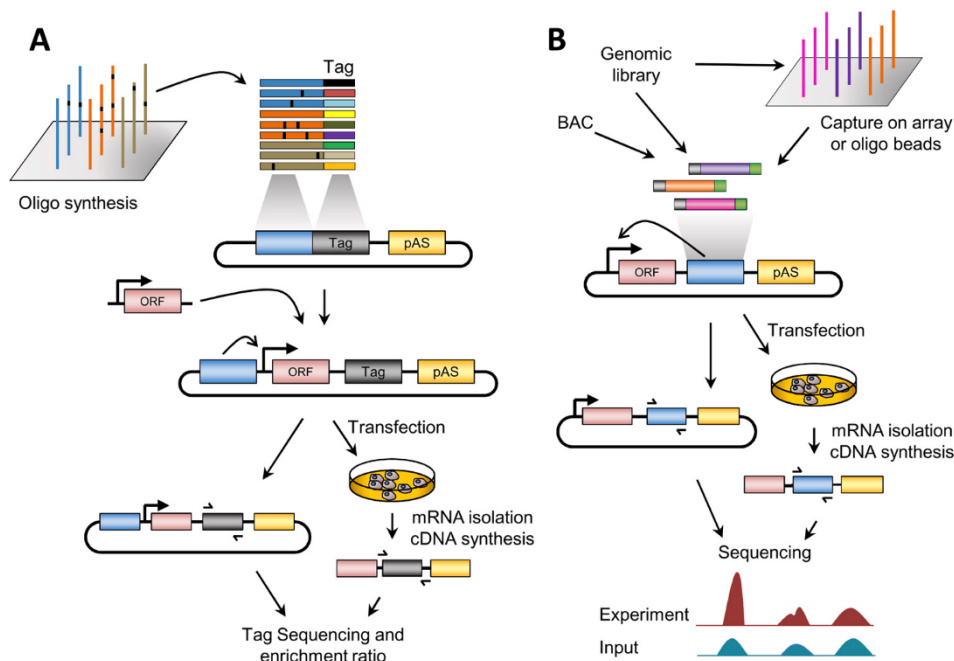


Figure 1. Principle of high-throughput assays for enhancer activity. (A) Overview of massively parallel reporter assay (MPRA). The test sequences (wild-type, variants, etc.) are generally synthesized *in silico* by massive oligonucleotide synthesis with unique barcode tags and cloned into the plasmid backbone. Tags can be synthesized along with the test sequences or added after synthesis by polymerase chain reaction (PCR) amplification. A basal promoter and a reporter open reading frame (ORF) are inserted between the tested element and tag sequences. The reporter library is then transfected into cultured cells. Subsequently, mRNA is isolated and cDNA synthesized. The tags are sequenced before (plasmid library pool, for normalization) and after the transfection. The difference in the enrichment of each barcode is proportional to the enhancer activity of the test sequence. In the case of post-synthesis addition of barcodes, an additional sequencing step is required at the first cloning step. (B) Overview of self-transcribing active regulatory region sequencing (STARR-Seq). A genomic or bacterial artificial chromosome (BAC) library is cloned in the reporter plasmid, downstream of the ORF and upstream of the polyadenylation site (pAS). Alternatively, the regions of interest might be enriched by a capture approach. The reporter library is transfected into cultured cells. Subsequently, mRNA is isolated and cDNA synthesized. The cloned regions are sequenced from the plasmid library pool (input) and the cDNA. Differences in the enrichment with respect to the input are proportional to the enhancer activity. In both panels, the effect of the enhancer on the basal promoter is indicated by an arrow.

Table 1. Examples of high-throughput functional assays of mammalian enhancers.

Original technique	Specific Name	DNA origin	Application	Specific features	No. of regions ¹	Size (bp)	Cell lines	Promoter	Species	Ref.
MPRA		Synthetic	Characterize putative enhancers and their active region	Centered on TF sites	2,104	145	K562, HepG2	SV40	Human	12
MPRA	CRE-seq	Synthetic	Characterize genomic regions with predicted function as enhancers	Mutation of TF binding sites	2,100	130	K562	Hsp68	Human	9
FIREWACH		Genomic	Identification of enhancers from restriction enzyme accessible regions	Isolation of GFP-positive regions	84,240	154	ESC	Minimal FpG5	Mouse	25
STARR-seq	CapSTARR-seq	Genomic	Identification of enhancers from DHS	Capture of DHS	7,542	330–430	P5424, 3T3	SCP1	Mouse	22
STARR-seq		Genomic	Characterize genetic variants in regulatory elements from 95 subjects	Capture of DHS	104	402	HepG2	SCP1	Human	23
MPRA	CRE-seq	Genomic	Identification of regulatory elements	Capture of DHS	4,000	464	Retina	Minimal Rho	Mouse	24
MPRA	Sharp-MPRA	Synthetic	Characterize nucleotides as activators or repressors in regulatory elements	Tiled oligonucleotide synthesis (30 or 5 bp resolution)	15,720	145	HepG2, K562	Minimal TATA, SV40	Human	13
MPRA		Synthetic	Identification of regulatory variants associated with eQTLs	Centered on eQTL variants	3,642	150	Lymphoblastoid, HepG2	Minimal TATA	Human	14
MPRA		Synthetic	Identification of regulatory variants associated with red blood cell traits	Three sliding windows with respect to the GWAS variants	2,756	145	K562, K562+GATA1	Minimal TATA	Human	15
STARR-seq	CapSTARR-seq	Genomic	Identification of promoter regions with enhancer activity	Capture of -200 to +50 bp with respect to the TSS of coding genes	20,719	250	K562, HeLa	SCP1	Human	33
MPRA	LentMPRA	Synthetic	Identification of enhancers by chromosomal integration	Centered on several ChIP-seq peak signals	2,236	171	HepG2	pGL4.23 promoter	Human	26
CRISPR-Cas9		Synthetic	Identify endogenous enhancers bound by p53 and ER α TFs	Target p53 binding site in putative enhancers	685	N/A	BJ-RASg12v	N/A	Human	44

¹Number of targeted DNA sequences, not necessarily the number of unique fragments that are tested

Bp, base pair; capSTARR-seq, capture-based self-transcribing active regulatory region sequencing; CHIP-seq, chromatin immunoprecipitation sequencing; CRE, cis-regulatory elements; CRISPR, clustered regularly interspaced short palindromic repeats; DHS, DNase I hypersensitive sites; eQTL, expression quantitative trait loci; ER α , estrogen receptor alpha; ESC, embryonic stem cell; FIREWACH, functional identification of regulatory elements within accessible chromatin; GFP, green fluorescent protein; GWAS, genome-wide association study; Hsp68, heat shock promoter 68; lentMPRA, lentiviral massively parallel reporter assay; MPRA, massively parallel reporter assay; N/A, not applicable; Sharp, systematic high-resolution activation and repression profiling with reporter-tiling; STARR-seq, self-transcribing active regulatory region sequencing; SCP1, super core promoter 1; SV40, simian virus 40; TF, transcription factor; TSS, transcription start site.

The MPRA method consists of the generation of a library of reporter constructs based on microarray synthesis of DNA sequences (generally, tested sequences are cloned upstream of a basal promoter) and unique sequence tags or barcodes (placed in the 3' UTR of the reporter gene). To increase the sensitivity and reproducibility, several barcodes could be added to any given sequence. The reporter library is then transfected into cell lines of interest and RNA sequencing of the barcodes is performed, thus providing a quantitative readout of the regulatory activity of the tested regions (Figure 1A).

MPRAs have been used to investigate a number of biological questions. Initially, MPRA was designed to dissect the functional components of previously identified enhancers at single-nucleotide resolution^{7,8}. Subsequently, a similar approach (also named CRE-seq) was used to functionally test ~2,000 genomic segments predicted by ENCODE to be enhancers, weak enhancers, or repressed elements⁹ as well as test synthetic enhancers to model grammatical rules of regulatory sequences^{10,11}. MPRA can be used to systematically assess the relevance of predicted regulatory motifs within enhancers. Kheradpour *et al.* tested ~2,000 predicted enhancers along with engineered enhancer variants containing targeted motif disruptions for key transcription factors (TFs)¹². In a follow-up study, Kellis' lab developed a high-resolution MPRA approach (also named Sharpr-MPRA) that allowed genome-scale mapping of activating and repressive nucleotides in regulatory regions¹³. Here, by synthesizing dense tiling of overlapping MPRA constructs, they managed to infer the regulatory effects of functional regulatory nucleotides with either activating or repressive properties^{11,13}. Finally, MPRA can be used to test the impact of single nucleotide polymorphisms (SNPs) in order to identify functional regulatory variants linked to human traits or diseases. Two recent studies from the Broad Institute provide proof-of-concept for such approaches. Tewhey *et al.* used an improved version of the MPRA to analyze thousands of human expression quantitative trait loci (eQTL) to identify alleles that impact gene expression in lymphoblastoid cell lines¹⁴. Ulirsch *et al.* used MPRA to test 2,756 variants linked to 75 genome-wide association studies (GWAS) loci involved in red blood cell traits¹⁵. In both cases, CRISPR-mediated genetic engineering confirmed the relevance of the MPRA findings. Interestingly, some of the identified regulatory variants did not lie within known motifs, suggesting that they can influence DNA structure instead of binding of TFs¹⁵.

A new innovative method named STARR-seq was introduced by Alexander Stark and colleagues¹⁶. STARR-seq is an MPRA (reviewed in 17) aimed to identify and quantify transcriptional enhancers directly based on their activity in entire genomes (Figure 1B). In brief, a bulk of DNA fragments from arbitrary sources is cloned downstream into the 3' UTR of a GFP reporter gene. Once in cellular context, active enhancers will activate the upstream promoter and transcribe themselves, resulting in reporter transcripts among cellular RNAs. Thus, each reporter transcript contains the reporter gene and the "barcode" of itself. These reporter transcripts can be isolated separately by targeted PCR and eventually detected by high-throughput sequencing. In this way, the activity of millions of putative enhancers can be measured

simultaneously without being affected by the location of the candidate sequences and their orientation. The main advantage over the classical MPRA is that the tested sequence itself is used as a "barcode", substantially simplifying the whole procedure to quantify enhancer activity. Stark's lab used the STARR-seq approaches to ask several basic mechanistic questions of enhancer biology in *Drosophila*, including (i) identification and characterization of cell-type-specific^{16,18} and hormone-responsive enhancers¹⁹, (ii) the impact of *cis*-regulatory sequence variation on enhancer activity and evolution²⁰, and (iii) dissecting the basis of enhancer core-promoter specificity²¹.

STARR-seq has been applied to human cells by utilizing selected bacterial artificial chromosomes (BACs)¹⁶; however, with the complexity and size of mammalian genomes, this technique is not easily implemented, making the formulation of representative libraries a challenge and a very high sequencing depth a necessity. To avoid this issue, we developed a capture-based approach (named CapSTARR-seq) to assess a subset of mouse DNase I hypersensitive sites (DHSs) found in developing thymocytes²². Here, the regions of interest are captured using custom-designed microarrays and cloned into the STARR-seq vector, thus providing cost-effective and accurate quantification of enhancer activity in mammals. Similar approaches have been published by other labs, including capture of natural genomic variants²³ and test of DHSs from the central nervous system using a capture approach with oligo-baits²⁴. Alternatively, it could be possible to directly clone open chromatin regions, as described in the functional identification of regulatory elements within accessible chromatin (FIREWACH) method²⁵.

Potential caveats of high-throughput reporter assays

The DNA sources used for testing are a potential issue of high-throughput reporter assays. Most MPRA approaches have used massive oligonucleotide synthesis (Figure 1A), which allows the precise definition of tested regions as well as custom modifications of underlying sequences. However, there are currently two limitations to this approach. On the one hand, the size of the tested fragment is limited to ~200 bp (including the adaptors), which might prevent testing the full regulatory regions. On the other hand, there is a limitation in the number of oligonucleotides that can be synthesized (currently <100,000). These constraints are expected to be overcome in the near future with the improvement of oligonucleotide synthesis technologies.

In the STARR-seq approach, the DNA fragments are cloned within the transcribed region (Figure 1B), which is very convenient because their sequences provide direct information about enhancer activity. However, it also introduces a source of potential artifacts, as some sequences might influence transcript stability instead of enhancing transcription. This potential bias could be avoided by comparing the results of tested regions on both orientations, allowing one to filter out the effects of strand-specific transcript-stabilizing effects.

A general concern about the episomal reporter assays is that they may not accurately reflect the function of enhancer elements in their endogenous context. To partially circumvent this caveat,

chromatinized adeno-associated^{11,24} and lentiviral MPRA^{25,26} have also been performed. These methods capacitate reporter assays within cells or tissues that are difficult to transfect. Certainly, an equally valid argument is that episomal reporter assays allow the unbiased study of enhancer function independently of any “perturbing” chromatin or genomic context. Interestingly, a recent study performed a systematic comparison of chromosomal versus episomal enhancer activity using integrative and non-integrative versions of a lentiviral-based reporter assay²⁶. Although the chromosomally based reporter assay was more predictable by epigenomic and sequence-based models, both reporter assays remained relatively well correlated. Another alternative approach is the introduction of reporter genes throughout the genome using transposition systems (e.g. 27–29). Although these approaches do not directly assess enhancer activity, they allow one to infer the regulatory context of endogenous loci.

Some relevant findings

One interest of high-throughput enhancer assays is the possibility to explore enhancer function without preconceived notions, thus potentially leading to new unforeseen findings. A common observation of several studies is that many predicted enhancer regions did not show reporter activity^{9,12,13,22,26}. For example, only 26% of predicted enhancers based on chromatin signatures in K562 cells displayed enhancer activity in the reporter assays performed in the same line⁹, suggesting that, in addition to histone modifications, additional sequence specificity, such as TF-binding sites, are essential determinants of *cis*-regulatory activity. Indeed, the concentration of TF-binding sites or motifs is highly predictive of strong enhancer activity^{13,22,30}. Alternatively, this could also indicate that not all of the required sequences are present in the tested regions or that endogenous promoter contexts are essential to the enhancer activity. Finally, it is also plausible that some open chromatin regions, while contributing to transcriptional regulation, have enhancer-independent functions³¹ or lack classical enhancer functions³².

Perhaps the most intriguing finding of functional enhancer assays comes from the observation that many core promoter regions display enhancer activity^{11,16,21,33}. The original definition of enhancers implies the ability to activate gene expression at a distance, while promoters entail the capability to induce local gene expression. However, this basic dichotomy of *cis*-regulatory elements has been challenged by broad similarities between promoters and enhancers, such as DNA sequence features, chromatin marks, Pol II recruitment, and bidirectional transcription³⁴. For instance, H3K4me3, a histone modification generally found at promoter regions, has been also associated with active enhancers^{35–37}. Assessment of enhancer activity by Cap-STARR-seq showed that strong transcription start site (TSS)-distal enhancers are indeed associated with H3K4me3 enrichment at the endogenous loci²². Several studies have also suggested that some promoters might play enhancer functions³⁴. The extent of this type of promoter and whether it actually functions to regulate the expression of distal genes have remained elusive. Now, several independent studies based on massive reporter assays reported widespread enhancer activity from TSS-proximal regions. By

applying STARR-seq, Zabidi *et al.* screened the whole fly genome with the use of different core promoters obtained from either ubiquitously expressed housekeeping genes or developmentally regulated and cell-type-specific genes²¹. They found that promoters of housekeeping genes were mainly regulated by promoter-proximal enhancers, while promoters of developmental and cell-type-specific genes required distally located enhancers. Ernst *et al.* found that active enhancers were enriched in DNase I sites overlapping TSS in human cell lines¹³. Nguyen *et al.* performed a functional comparison of a subset of promoters and enhancers in mouse neurons using an integrative MPRA approach¹¹. Interestingly, gene promoters generated similar enhancer activity as compared to distal regulatory regions. In a recent study, we found that 2–3% of all human core promoters displayed enhancer activity in a given cell line³³. Compared to classical promoters and distal enhancers, these TSS-overlapping enhancers displayed distinct genomic and epigenomic features and were associated with housekeeping and stress response genes. CRISPR genomic deletions demonstrated that several core promoters with enhancer activity in the reporter assay are indeed involved in *cis*-regulation of distal gene expression in their natural context, therefore functioning as *bona fide* enhancers. Furthermore, human genetic variation within this type of promoter was associated with a strong effect on distal gene expression. Concomitantly, another study, using comprehensive genetic manipulation of promoter regions, reported frequent distal *cis*-regulation by loci associated with promoters of lncRNAs and, to a lesser extent, coding genes³². Finally, two recent studies performing screens of *cis*-regulatory elements by CRISPR/Cas9-based approaches (see below) have found that the expression of some genes is controlled by distal gene promoters^{38,39}. Overall, these findings open up the intriguing possibility that developmental traits or disease-associated variants lying within a subset of promoters might directly impact on distal gene expression.

CRISPR-based approaches to assess enhancer function

As discussed above, a potential limitation of reporter assays is that candidate enhancers are studied outside their endogenous genomic context, which is likely required for their *in vivo* function. The advent of CRISPR-based technologies now allows the circumvention of this caveat. Several studies have performed systematic dissection of individual enhancers using either TALEN⁴⁰ or CRISPR-mediated mutagenesis⁴¹. In these studies, a tiling single guide RNA (sgRNA), or TALEN, library covering selected enhancers was designed to perform *in situ* saturating mutagenesis screens, pinpointing sequences with either positive or negative impact on enhancer function. The CRISPR approach was subsequently extended to assess enhancer function within large genomic regions surrounding key loci^{38,39,42,43} or to screen for enhancer elements involved in specific gene regulation pathways⁴⁴ (Table 1).

CRISPR-mediated mutagenesis is limited by the fact that a high density of sgRNAs is required to saturate all possible regulatory elements and specific mutations are difficult to implement genome-wide. In some cases, there might also be a bias with respect to

the regions that can be targeted by the designed sgRNAs or limitations owing to transfection efficiency in particular cell types. Alternatively, nuclease-deactivated Cas9 (dCas9) can be fused to activator or repressor domains to precisely modify gene expression from promoters and distal regulatory elements⁴⁵. Based on this property, repressor and activator domains fused with dCas9 combined with a pool of sgRNAs have been used for comprehensive CRISPR-inactivation (CRISPRi) and CRISPR-activation screens targeting DHSs of a gene of interest⁴⁶ or an entire locus⁴⁷. In the former study, a reporter gene introduced at the place of the target genes was used to monitor enhancer activity. In the latter study, the screening criteria were based on the growth advantage or disadvantage provided by the change in expression of the enhancer-associated gene, thus providing proof-of-concept for screening of functional regulatory regions genome-wide. A current limitation of these approaches is that the screening strategy might be based on phenotypic features (such as cell growth fitness, developmental markers, etc.) instead of directly assessing the expression levels of regulated genes. To overcome this limitation, a new powerful method combined CRISPRi and single-cell RNA-seq⁴⁸, enabling high-throughput interrogation of enhancers at single-cell resolution and directly linking enhancer function, and their combinations, with its target gene(s). Although these approaches have been used so far to scan restricted genomic areas, they will likely be implemented in true genome-wide screens of regulatory elements in the coming future.

Concluding remarks

The implementation of high-throughput reporter assays and CRISPR-based screens allows the experimental validation of enhancer activity in different cell types and cellular contexts (Table 1). These assays are now robust and sensitive enough to be widely used as part of the toolkit for researchers interested in gene regulation. These approaches also led to unpredicted

discoveries, such as the role of core promoters as enhancer-like regulators. One main limitation of these approaches remains the fact that they do not provide direct information towards uncovering enhancer target genes. Therefore, the combination of enhancer assays with recently developed 3C-related methodologies, such as 4C-seq, Hi-C or Capture Hi-C⁴⁹, should greatly facilitate the assignment of discovered enhancers to their putative target genes. Finally, with the expected decrease in the cost of sequencing and oligo synthesis, it will be possible to systematically test the impact of regulatory variants in different diseases and developmental contexts.

Abbreviations

CRISPR, clustered regularly interspaced short palindromic repeats; DHS, DNase I hypersensitive site; MPRA, massively parallel reporter assay; sgRNA, single guide RNA; STARR, self-transcribing active regulatory region; TF, transcription factor; TSS, transcription start site.

Competing interests

The authors declare that they have no competing interests.

Grant information

Work in Salvatore Spicuglia's laboratory was supported by recurrent funding from the Inserm and Aix-Marseille University and by the ARC (PJA 20151203149) and the *Plan Cancer* (P036496) grants. David Santiago-Algarra and Lan T.M. Dao were supported by the CONACYT-Mexico and Vietnam International Education Development, respectively.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References



- Kim TK, Shiekhattar R: **Architectural and Functional Commonalities between Enhancers and Promoters.** *Cell.* 2015; **162**(5): 948–59.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vernimmen D, Bickmore WA: **The Hierarchy of Transcriptional Activation: From Enhancer to Promoter.** *Trends Genet.* 2015; **31**(12): 696–708.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chatterjee S, Ahituv N: **Gene Regulatory Elements, Major Drivers of Human Disease.** *Annu Rev Genomics Hum Genet.* 2017; **18**.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dailey L: **High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics.** *Genomics.* 2015; **106**(3): 151–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Inoue F, Ahituv N: **Decoding enhancers using massively parallel reporter assays.** *Genomics.* 2015; **106**(3): 159–64.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- White MA: **Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences.** *Genomics.* 2015; **106**(3): 165–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Patwardhan RP, Hiatt JB, Witten DM, *et al.*: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nat Biotechnol.* 2012; **30**(3): 265–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Melnikov A, Murugan A, Zhang X, *et al.*: **Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay.** *Nat Biotechnol.* 2012; **30**(3): 271–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kwasnieski JC, Fiore C, Chaudhari HG, *et al.*: **High-throughput functional testing of ENCODE segmentation predictions.** *Genome Res.* 2014; **24**(10): 1595–602.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith RP, Taher L, Patwardhan RP, *et al.*: **Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model.** *Nat Genet.* 2013; **45**(9): 1021–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nguyen TA, Jones RD, Snively AR, *et al.*: **High-throughput functional comparison of promoter and enhancer activities.** *Genome Res.* 2016; **26**(8): 1023–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kheradpour P, Ernst J, Melnikov A, *et al.*: **Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay.** *Genome Res.* 2013; **23**(5): 800–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ernst J, Melnikov A, Zhang X, *et al.*: **Genome-scale high-resolution mapping of**

- activating and repressive nucleotides in regulatory regions. *Nat Biotechnol.* 2016; 34(11): 1180–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. **F** Tewhey R, Kotliar D, Park DS, *et al.*: **Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay.** *Cell.* 2016; 165(6): 1519–29.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 15. **F** Ulirsch JC, Nandakumar SK, Wang L, *et al.*: **Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits.** *Cell.* 2016; 165(6): 1530–45.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 16. **F** Arnold CD, Gerlach D, Stelzer C, *et al.*: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science.* 2013; 339(6123): 1074–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 17. Muerdtter F, Bonyá LM, Arnold CD: **STARR-seq - principles and applications.** *Genomics.* 2015; 106(3): 145–50.
[PubMed Abstract](#) | [Publisher Full Text](#)
 18. Yáñez-Cuna JO, Arnold CD, Stampfel G, *et al.*: **Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features.** *Genome Res.* 2014; 24(7): 1147–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Shlyueva D, Stelzer C, Gerlach D, *et al.*: **Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin.** *Mol Cell.* 2014; 54(1): 180–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
 20. Arnold CD, Gerlach D, Spies D, *et al.*: **Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution.** *Nat Genet.* 2014; 46(7): 685–92.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 21. Zabihi MA, Arnold CD, Scherzhuber K, *et al.*: **Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation.** *Nature.* 2015; 518(7540): 556–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
 22. Vanhille L, Griffon A, Maqbool MA, *et al.*: **High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq.** *Nat Commun.* 2015; 6: 6905.
[PubMed Abstract](#) | [Publisher Full Text](#)
 23. Vockley CM, Guo C, Majoros WH, *et al.*: **Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort.** *Genome Res.* 2015; 25(8): 1206–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 24. Shen SQ, Myers CA, Hughes AE, *et al.*: **Massively parallel cis-regulatory analysis in the mammalian central nervous system.** *Genome Res.* 2016; 26(2): 238–55.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 25. Murtha M, Tokcaer-Keskin Z, Tang Z, *et al.*: **FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells.** *Nat Methods.* 2014; 11(5): 559–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Inoue F, Kircher M, Martin B, *et al.*: **A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity.** *Genome Res.* 2017; 27(1): 38–52.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 27. **F** Ruf S, Symmons O, Uslu VV, *et al.*: **Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor.** *Nat Genet.* 2011; 43(4): 379–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 28. **F** Akhtar W, de Jong J, Pindyurin AV, *et al.*: **Chromatin position effects assayed by thousands of reporters integrated in parallel.** *Cell.* 2013; 154(4): 194–27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 29. **F** Symmons O, Uslu VV, Tsujimura T, *et al.*: **Functional and topological characteristics of mammalian regulatory domains.** *Genome Res.* 2014; 24(3): 390–400.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 30. Dogan N, Wu W, Morrissey CS, *et al.*: **Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility.** *Epigenetics Chromatin.* 2015; 8: 16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. Bevington SL, Cauchy P, Cockerill PN: **Chromatin priming elements establish immunological memory in T cells without activating transcription: T cell memory is maintained by DNA elements which stably prime inducible genes without activating steady state transcription.** *Bioessays.* 2017; 39(2).
[PubMed Abstract](#) | [Publisher Full Text](#)
 32. **F** Engreitz JM, Haines JE, Perez EM, *et al.*: **Local regulation of gene expression by lncRNA promoters, transcription and splicing.** *Nature.* 2016; 539(7629): 452–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 33. Dao LTM, Galindo-Albarrán AO, Castro-Mondragon JA, *et al.*: **Genome-wide characterization of mammalian promoters with distal enhancer functions.** *Nat Genet.* 2017.
[PubMed Abstract](#) | [Publisher Full Text](#)
 34. Andersson R: **Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model.** *Bioessays.* 2015; 37(3): 314–23.
[PubMed Abstract](#) | [Publisher Full Text](#)
 35. Pekowska A, Benoukraf T, Zacarias-Cabeza J, *et al.*: **H3K4 tri-methylation provides an epigenetic signature of active enhancers.** *EMBO J.* 2011; 30(20): 4198–210.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Koch F, Fenouil R, Gut M, *et al.*: **Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters.** *Nat Struct Mol Biol.* 2017; 18(8): 956–63.
[PubMed Abstract](#) | [Publisher Full Text](#)
 37. Core LJ, Martins AL, Danko CG, *et al.*: **Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers.** *Nat Genet.* 2014; 46(12): 1311–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 38. Diao Y, Fang R, Li B, *et al.*: **A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells.** *Nat Methods.* 2017; 14(6): 629–35.
[PubMed Abstract](#) | [Publisher Full Text](#)
 39. Rajagopal N, Srinivasan S, Kooshesh K, *et al.*: **High-throughput mapping of regulatory DNA.** *Nat Biotechnol.* 2016; 34(2): 167–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. Vierstra J, Reik A, Chang KH, *et al.*: **Functional footprinting of regulatory DNA.** *Nat Methods.* 2015; 12(10): 927–30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Canver MC, Smith EC, Sher F, *et al.*: **BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis.** *Nature.* 2015; 527(7577): 192–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 42. Sanjana NE, Wright J, Zheng K, *et al.*: **High-resolution interrogation of functional elements in the noncoding genome.** *Science.* 2016; 353(6307): 1545–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 43. Diao Y, Li B, Meng Z, *et al.*: **A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening.** *Genome Res.* 2016; 26(3): 397–405.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Korkmaz G, Lopes R, Ugalde AP, *et al.*: **Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9.** *Nat Biotechnol.* 2016; 34(2): 192–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
 45. Hsu PD, Lander ES, Zhang F: **Development and applications of CRISPR-Cas9 for genome engineering.** *Cell.* 2014; 157(6): 1262–78.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Klann TS, Black JB, Chellappan M, *et al.*: **CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome.** *Nat Biotechnol.* 2017; 35(6): 561–568.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 47. **F** Fulco CP, Munschauer M, Anyoha R, *et al.*: **Systematic mapping of functional enhancer-promoter connections with CRISPR interference.** *Science.* 2016; 354(6313): 769–73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 48. Xie S, Duan J, Li B, *et al.*: **Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells.** *Mol Cell.* 2017; 66(2): 285–299.e5.
[PubMed Abstract](#) | [Publisher Full Text](#)
 49. Sati S, Cavalli G: **Chromosome conformation capture technologies and their impact in understanding genome function.** *Chromosoma.* 2017; 126(1): 33–44.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: 

Editorial Note on the Review Process

F1000 Faculty Reviews are written by members of the prestigious F1000 Faculty. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

The reviewers who approved this article are:

Version 1

- Bart Deplancke**
Laboratory of Systems Biology and Genetics, Institute of Bioengineering, Swiss Federal Institute of Technology in Lausanne (EPFL), Lausanne, Switzerland
Competing Interests: No competing interests were disclosed.
- Douglas Vernimmen**
The Roslin Institute, University of Edinburgh, Edinburgh, UK
Competing Interests: No competing interests were disclosed.
- Alvaro Rada-Iglesias**
Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany
Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research