



**HAL**  
open science

## High-throughput study of silencer elements in T cells

Saadat Hussain

► **To cite this version:**

Saadat Hussain. High-throughput study of silencer elements in T cells. Life Sciences [q-bio]. Aix-Marseille Université, 2019. English. NNT : 2019AIXM0215 . tel-04426960

**HAL Id: tel-04426960**

**<https://amu.hal.science/tel-04426960>**

Submitted on 30 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 UNIVERSITY

FACULTY OF SCIENCE

DOCTORAL SCHOOL OF LIFE SCIENCES AND HEALTH

THEORIES AND APPROACHES OF GENOMIC COMPLEXITY

DOCTORAL THESIS

GENOMICS AND BIOINFORMATICS

---

High-throughput study of silencer elements in T cells

---

BY

**SAADAT HUSSAIN**

JURY MEMBERS:

**Dr. Eric PINAUD**

**President and Reporter**

**Dr. Alvaro RADA-IGLESIAS**

**Reporter**

**Dr. Simona SACCANI**

**Examiner**

**Dr. Salvatore SPICUGLIA**

**Supervisor**

# Acknowledgements

First and foremost I acknowledge jury members **Eric PINAUD**, **Alvaro RADA-IGLESIAS** and **Simona SACCANI**, who dedicated their time for critical reading, proof and assessing my dissertation. I am grateful for their valuable input and candid advice that actually helps enlarging my vision of science.

I would like to offer my sincerest gratitude to my supervisor, **Dr. Salvatore SPICUGLIA**, who has supported me throughout my thesis with his patience and knowledge. I attribute the level of my PhD degree to his encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor.

This project would also have been impossible without the contribution of **Nori SADOUNI**. He helps me a lot with bioinformatics parts. Most of the bioinformatics analyses were performed by him. Without him this project would never have done.

I offer my enduring gratitude to the lab mate **Lan T. M. DAO** for her indispensable help in conducting my experiments. My PhD is credited to her support and devoted guidance, always providing coherent answers to my questions.

I owe particular thanks to **Magali TORRES** who always help me to do my work and her valuable guidelines. I am honored to meet a person like her.

Thanks to our collaborators **Mohamed BELHOCINE**, **Eve-lyne MATHIEU**, **Ariel GALINDO**, **Guillaume CHARBONNIER**, **Alexander ESPANA**, **Iris MANOSALVA**, **David ABAD**, **Himanshu Narayan SINGH**, **Lydie PRADEL**, **Denis PUTHIER**, **Quentin FERRÉ**, **David SANTIAGO**, **Wiam SAADI**, **Yasmina KERMEZLI** who have been providing intellectual materials and moral support throughout.

Thanks to all members of the **TGML** for performing RNA-Seq. Many thanks to the staffs of **TAGC** for creating an environment where I can focus on doing great science, working with you have been extremely fun, stimulating and educational. I am very happy to spend the most beautiful of my life with you. Thanks so much for all.

I would like to acknowledge funding from **Higher Education Commission of Pakistan**. Without their support I would never had a chance to carry out my PhD in France. For the last 4 years, I have had the privilege of being a part of the best graduate program at Aix-Marseille University. This fundamental education provides the solid foundation for my future careers.

Last but not least, I would like to express my special thanks to my entire **dear friends** everywhere for their supports, kindness and amity. All of you have been an indispensable part of my life. My life is happier and meaning when I meet you.

Above all, thank to **my parents** who gave me a life to live with so much of love and joy. **My wife, brother and sisters**, who stood with me all this time. There are not enough words to describe just how important they are for me. Thanks so much to **Soul of My Father's**, hard to say how much I missed you, he was waiting for my graduation but unfortunately he left us few days back. This dissertation is a gift to all my family members specially my father.

# Abstract

The control of gene expression is fundamental to mammalian cell life. Although much of this control occurs at transcriptional and post transcriptional level. Gene expression is controlled by cis regulatory elements along with numbers of transcription factors combinatorial in mammals. The transcription can only be initiated by the assembly of RNAPII machinery around transcription start site of a gene which is also known as core promoter. Other cis regulatory elements like enhancers, silencers and insulators along with several transcription factors are also involved in gene regulation.

There has been increasing number of studies relating enhancer and promoters but silencer has not been studied as much. Previous findings have suggested that some gene is regulated by silencers in T cell differentiation. The stagnation of silencer studies raises several questions whether silencers represent a major genomic strategy of gene regulation. There is lack of high throughput screening technique to identify silencers.

My project has carried out aiming to answer these above questions. Firstly, I have repurposed the CapSTARR seq technique to identify activity silencer elements; previously this technique has developed in the lab, which used as an approach to exploiting a high-throughput enhancer activity. Performing CapStarr-seq in mouse p5424 cell line, I found a substantial proportion of regions displaying silencer activity. They display silencer activity in dual luciferase assay. The REST motif found to be enriched in several silencer elements. Moreover, by using comprehensive CRISPR/Cas9 genomic deletion approach, I demonstrated that silencer is generally involved in the regulation of nearby genes. Taken together, our results identify a new high throughput reporter assay to identify silencer element.

My thesis is structured into 7 chapters. In chapter 1, I summarize the current understanding about gene regulation in mammals and the important factors contributing to transcriptional regulation. In chapter 2, I focus particularly on the regulatory role of silencer elements on gene expression. Here, the classical view of silencer and the classical reporter assays to identify silencer elements are discussed. There is an overview about silencer found in the hematopoietic system specifically during T cell differentiation. In chapter 3, I discussed respectively about the High throughput techniques that can be used to study regulatory elements. Chapter 4 is focused on the T cell differentiation. The results are organized in chapter 5. This chapter tells about how we identify silencer elements using CapSTARR-Seq technique, and functional validation of silencer elements. In Chapter 6, I give general discussion about results. Finally, in chapter 7, perspectives are mentioned. My work continues to identify silencer elements in different cell lines.

# Résumé

Le contrôle de l'expression des gènes est fondamental pour la vie des cellules des mammifères. La majeure partie de ce contrôle se produisant au niveau de la transcription puis la transcription. Chez les mammifères, l'expression des gènes est contrôlée par des éléments cis-régulateurs ainsi que des combinaisons de facteurs de transcription. La transcription ne peut être initiée que par l'assemblage de la machinerie RNAPII autour du site d'initiation de la transcription du gène également appelé région promotrice. D'autres éléments cis-régulateurs tels que les enhancers, les silencers et les insulateurs associés à plusieurs facteurs de transcription sont également impliqués dans la régulation des gènes.

Alors que les enhancers et les promoteurs sont largement étudiés par la communauté scientifique, les mécanismes d'actions des silencers restent à l'heure actuelle méconnus. Des recherches antérieures ont suggéré que certains gènes sont réprimés par des silencers dans la différenciation des lymphocytes T. Le manque de connaissance sur les silencers soulève plusieurs questions sur l'importance au sein de la régulation génomique. À l'heure actuelle, peu de techniques permettent l'identification et l'étude des éléments silencers.

Mon projet a été réalisé dans le but de répondre à ces questions. Premièrement, j'ai réutilisé la technique du CapSTARR-seq afin d'identifier des éléments silencers; Cette technique avait été mise au point dans le laboratoire. Cette technique était utilisée pour l'étude des éléments enhancers. En réalisant le CapStarr-seq dans la lignée cellulaire p5424 de souris, j'ai pu identifier une proportion non négligeable de régions montrant une activité répressive. Des résultats similaires ont été démontrés via un reporter de luciférase. Le motif REST s'est révélé enrichi de plusieurs régions silencers. De plus, en utilisant l'approche de la suppression génomique CRISPR / Cas9, j'ai démontré que les silencers sont généralement impliqués dans la régulation des gènes proches. L'ensemble de nos résultats nous ont permis d'identifier une nouvelle technique de séquence à haut débit pour l'identification d'éléments silencers.

Ma thèse est structurée en 7 chapitres. Au chapitre 1, je résume les connaissances actuelles sur la régulation des gènes chez les mammifères et les facteurs importants contribuant à la régulation de la transcription. Dans le chapitre 2, je me concentre particulièrement sur le rôle régulateur des éléments silencers dans l'expression des gènes. Dans ce chapitre, nous discutons du mode d'action « classique » des silencers ainsi que les techniques existant afin de les étudier. On y trouve également un résumé sur les silencers trouvés dans le système hématopoïétique, lors de la différenciation des cellules T. Au chapitre 3, j'aborde les différentes techniques de séquençage à haut débit pouvant être utilisées pour étudier les éléments de régulation génomique. Le chapitre 4 est consacré à la différenciation des cellules T. Les résultats sont présentés au chapitre 5. Ce chapitre explique comment nous identifions les régions silencers avec la technique du CapSTARR-Seq et sur la validation fonctionnelle des silencers. Au chapitre 6, discute d'une façon globale les différents résultats obtenus. Les perspectives sont exposées dans le chapitre 7. Mon travail d'identifications des régions silencers dans différentes lignées cellulaires continue.

# Contents

	Page
<b>Introduction</b>	
<b>Chapter 1: Transcriptional regulation in mammals</b>	
I. The importance of transcriptional regulation	8
II. Transcription regulatory factors	9
1. Chromatin structure, histone modifications, DNA methylation	9
2. Transcription factors	15
3. Cis-regulatory elements	17
3.1. Promoters	18
3.2. Enhancers	19
3.3. Silencers	23
3.4. Insulators	23
4. Long non-coding RNAs	25
5. Long-range chromatin interactions	28
<b>Chapter 2 : Silencer elements</b>	
1. Historical Overview	32
2. Models of silencer function	33
3. Repressors	37
3.1 REST	37
4. Silencers in hematopoietic system and specifically during T cell differentiation	38
4.1 CD4 silencer	39
4.2 Runx1 and Runx3 at the CD4 silencer	40
4.3 ThPOK silencer promotes CD4 commitment	41
4.4 Key role for ThPOK silencer	42
4.5 Opposing effects of Runx and ThPOK on CD4/CD8 on lineage specification	42
4.6 L2a silencer CD8 gene	43
4.7 Silencer and anti-silencers regulate <i>Rag1</i> and <i>Rag2</i> genes	44
<b>Chapter 3 : High-throughput reporter assays</b>	
1. Overview	46
2. Classical reporter assays	46
3. High-throughput reporter assays	47

3.1. Massively Parallel Reporter Assays (MPRAs)	47
3.2. Self-Transcribing Active Regulatory Region Sequencing (STARR-seq)	48
3.3. ChIP-STARR-seq	51
3.4. Survey of regulatory elements (SuRE)	52
3.5. High-Definition Reporter Assay (HiDRA)	53
4. Other High throughput reporter assays	56
5. Medium and High-throughput approaches to identify silencers	56
<b>Chapter 4 : T cell differentiation</b>	<b>58</b>
<b>Chapter 5 : Results</b>	
1 Objective	62
2. Main results	62
3. Contributions	63
Manuscript: <i>High-throughput identification of silencer elements</i>	64
<b>Chapter 6 : Discussion</b>	<b>102</b>
<b>Chapter 7 : Perspectives</b>	
1. Identifying functional silencers	107
2. Exploring the function of Sdhaf1_silencer	107
3. Developing an improved pipeline to identify silencers	109
4. Future experiments	109
5. Remaining questions	110
<b>Bibliography</b>	<b>113</b>
<b>Annex: CRISPR/cas9 screening table</b>	<b>130</b>

# **Introduction**



# Chapter 1

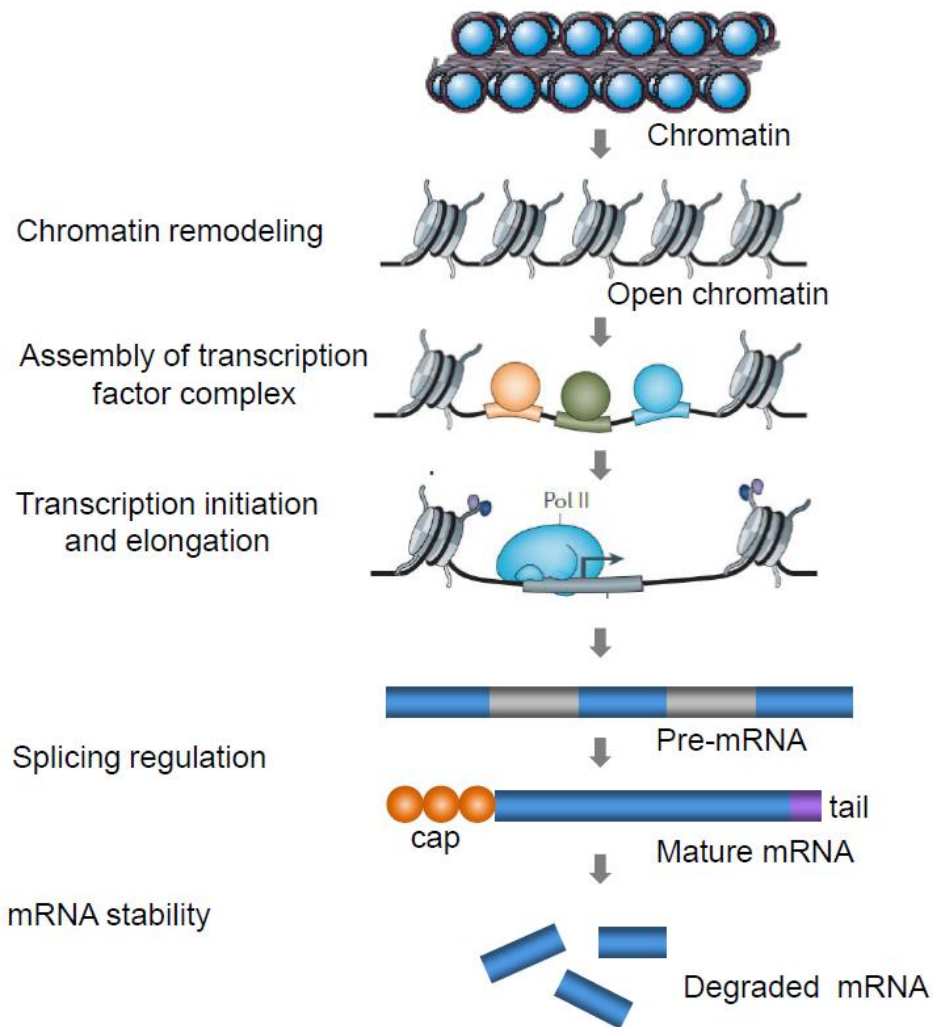
## Transcriptional regulation in mammals

### I. The importance of transcriptional regulation

Regulated gene expression is crucial to ensure normal cellular and organismal function. Studies on transcriptional regulation are perhaps the most investigated field in the understanding of the complexity of mammalian gene expression. Today, we know that the unique molecule of DNA contains the genetic and vital data within its bases as constitutional structures of nucleotides. Both eukaryotic and prokaryotic chromosomes are made up of DNA molecules. In addition to DNA, the role and importance of RNA is also important. It is thought that its informational functions were subsequently devolved to the more stable and easily replicable DNA, and its catalytic functions to the more chemically versatile polypeptides (Kohler et al., 2017).

However, each individual is constructed by a hundred specialized cell types and many organs with that shared genome. In the attempts to understand what makes this variety, researchers found that the basis of diversity throughout differentiation and development is by differential gene expression. According to the center dogma of biology, the genetic information hidden in the format of the gene can be expressed through two vital processes of transcription (production of coding molecules of mRNA) and translation (in which nucleotides are replaced by amino acids). RNA molecules are not only temporal and transient transcripts that are obtained from the genes, but also they act as pivotal modulators that mediate pre-transcriptional and posttranscriptional steps and have a direct affect on gene regulation and gene expression processes (Bunch, 2018).

Transcription in mammalian cells is regulated at multiple stages and several protein factors and cofactors are involved at each stage. In general, a transcriptionally active gene is controlled by a stretch of DNA sequence mostly situated at the upstream of the transcription start site (−500 bp to −1000 bp) defined as the promoter, which is a docking site of several proteins termed as transcription factors (TF). Mammalian cells synthesize around 3000 transcription factors (Baatar et al., 2002), and each one harbors a specific DNA sequence binding motif. In short, several steps can be described in the pathway from DNA to RNA to protein, including the modulations of chromatin states, transcription initiation and elongation, mRNA processing, transport, translation and stability (Fig. 1.1). This chapter aims to give an overview of the factors contributing to transcriptional regulation.



**Figure 1.1 | Different levels of transcriptional regulation.** Chromatin structure decides whether gene is available for transcription. The specific binding of transcription factors (TFs) and other protein complexes at promoters controls the transcription initiation. Transcription elongation is regulated by RNA polymerase II (RNAPII). The primary transcripts are then going through splicing regulation. The stability of mRNA within cytoplasm is defined by degradation frequency.

## II. Transcription regulatory factors

### 1. Chromatin structure, histone modifications, DNA methylation

#### Chromatin structure

In eukaryotes, the modification of chromatin structure provides unique transcription regulatory mechanisms. Chromatin is a dynamic structure that not only helps to package the entire eukaryotic genome into the confines of the nucleus but also regulates the accessibility of DNA for transcription, recombination, DNA repair and replication. Although the structure

of nucleosomes appears rigid at the cytological level, these repeating subunits of chromatin are very dynamic (Luger et al., 2012).

Furthermore, Nucleosome consists of a protein core of basic histone proteins, around which the DNA is wound 1.65 times. The core is an octamer that comprises two copies of each of the canonical histones: H3, H4, H2A and H2B. The four proteins interact in an ordered manner during nucleosome assembly, giving rise to the modular nature of the nucleosome. There are variants for the core histones H3, H2A and H2B, and for the linker histone H1. These variants may differ from the core histones either by the alteration of a few amino acids or by the addition of larger domains. The regulation of transcription is a result of the combined effects of structural properties and the interaction of transcription factors (TFs) to the regulatory elements. The state of chromatin is a key player that can limit or facilitate the accessibility of TFs and RNA polymerase to reach the regulatory sequences. Chromatin accessibility is influenced by several factors, including the interaction of chromatin with DNA-binding proteins (activators and repressors), DNA methylation, histone variants, histone modifications and the presence of non-coding RNAs. The combination of these factors determines the different states of chromatin that are open, closed or poised. In order to initiate gene transcription, gene's regulatory sequences must be exposed, thus, the chromatin must be unwrapped.

### **Histone posttranslational modification**

Posttranslational modification of histone proteins is one of the very well-studied epigenetic modifications. As described above, mammalian chromosomes are compacted into the nucleus by forming the primary and several higher order structures by the building block of nucleosomes. Posttranslational modifications affect nucleosome stability by altering the chemical interactions within nucleosomes (Williams et al., 2008) or with neighboring nucleosomes, resulting in open chromatin or closed chromatin. Posttranslational modifications may also prevent proteins from binding to chromatin (Venkatesh et al., 2012). Depending on their effect on transcription, most marks are classified as activating or repressing (Smolle and Workman, 2013). As each nucleosome comprises a histone octamer surrounded by 146 base pairs of DNA, so the amino terminal (N-terminal) part of histone protein protrudes out of the histone-DNA assembly. The N-terminal modification of the core histone proteins is very common and those modifications are acetylation, phosphorylation, methylation, sumoylation and ubiquitination.

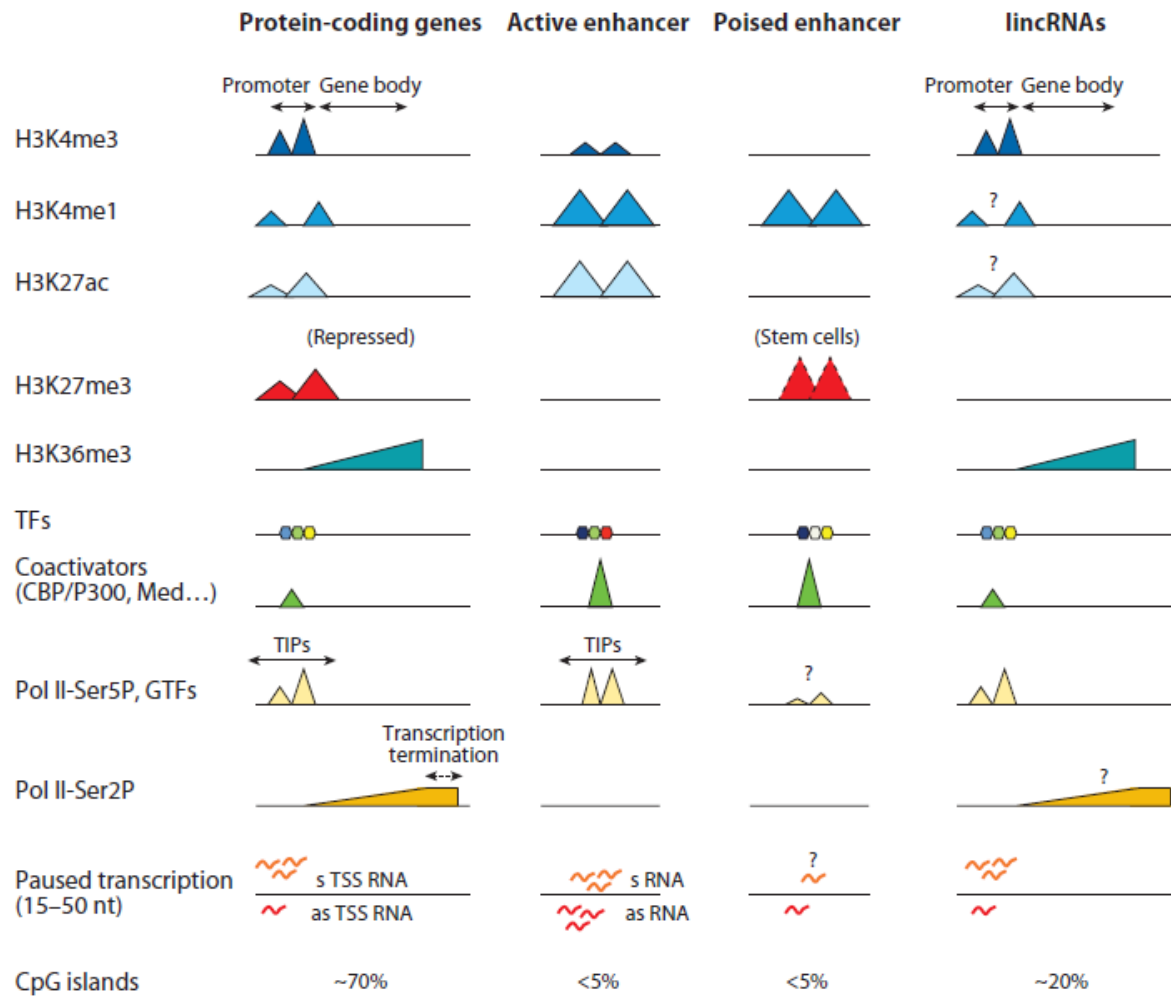
All these modifications are unique in a sense because each one of them introduces a specific change in the secondary and higher order structure of the chromatin, which in turn contributes to gene expression. Lysine acetylation correlates with chromatin accessibility and transcriptional activation. The acetylation of lysine 27 (H3K27ac) has shown to mark at active promoters and distal regulatory elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011). Tri-methylation of histone H3 lysine 4 (H3K4me3) and H3 lysine 36 (H3K36me3) are both associated with transcribed chromatin, however, H3K4me3 marks mainly promoter regions whereas H3K36me3 is found along gene body of transcribed genes (Barski et al., 2007; Mikkelsen et al., 2007). In contrast to these active marks, tri-methylation of H3 lysine 9

(H3K9me3), H3 lysine 27 (H3K27me3) and H4 lysine 20 (H4K20me3) are generally related to gene repression (Mikkelsen et al., 2007; Zhu et al., 2012). Posttranslational histone modifications associated with their location and functions are shown in (Table 1.1).

**Table 1.1: Post-translational histone modifications associated with their location and functions.** Figure partly adapted from (Lawrence et al., 2016).

<b>Modification</b>	<b>Position</b>	<b>Associated effects</b>
H3K4me1	<i>Cis</i> -regulatory elements	activation
H3K4me2	<i>Cis</i> -regulatory elements	Permissive euchromatin
H3K4me3	<i>Cis</i> -regulatory elements	Transcriptional elongation; active euchromatin
H3K9me3	Heterochromatin	Transcriptional repression; imprinting; DNA methylation
H3K9ac	<i>Cis</i> -regulatory elements	Histone deposition; transcriptional activation
H3K27ac	<i>Cis</i> -regulatory elements	Transcriptional activation
H3K27me3	<i>Cis</i> -regulatory elements	Transcriptional silencing; X-inactivation; bivalent genes/gene poising
H3K36me3	Gene body	Transcriptional elongation
H3K79me3	Gene body	activation

Most histone posttranslational modifications are reversible, as the cell contains separate enzymes that add and remove these marks (known as ‘writers’ and ‘erasers’, respectively). This separation not only regulates the genomic localization of these marks, which ensures the removal of aberrantly added marks, but also controls how long a Posttranslational modification remains at a particular genomic location. For example, acetylation of the variant H2A.Z by promoter bound acetyltransferases and prevents its removal from promoter regions by the inositol requiring protein 80 (INO80) remodeler complex (Papamichos-Chronakis et al., 2011). The Addition of Posttranslational modifications on histones may be the consequence of the activity of histone-modifying enzymes at a specific location rather than the cause of their genomic localization. The H3.3 variant-containing nucleosome is enriched for active chromatin marks (such as H4 acetylation, and H3 lysine 4 (H3K4) and H3K36 methylation) owing to its localization to regions undergoing active transcription (McKittrick et al., 2004). Epigenetic and transcriptional features associated with protein-coding genes, enhancers, and long intergenic noncoding RNAs (lincRNAs) are mentioned in (Fig 1.2).



**Figure 1.2 | Epigenetic and transcriptional features associated with protein-coding genes, enhancers, and long intergenic noncoding RNAs (lincRNAs).** Based on the literature, the main features of protein-coding genes (both promoters and gene bodies), active and poised enhancers, and lincRNAs are indicated on each line. In general, H3K4me1 nucleosomes are more spread around the boundaries of promoters and cover larger areas than does H3K4me3 on both promoters and enhancers. The amount of H3K4me1 relative to H3K4me3 is also higher on enhancers, irrespective of their poised/active status. The concept of poised enhancers carrying H3K27me3, and eventually shifting to H3K27ac in the active state, was developed mainly in embryonic stem cells and remains to be verified in other systems. Initiating Pol II (Ser5P), H3K27ac and bidirectional paused transcription in combination with H3K4me1 high/H3K4me3 low are the best indicators of active enhancers in various systems, whereas the occurrence of H3K4me3 followed by H3K36me3 regions corresponds to lincRNAs and protein-coding genes. It remains to be established if transcription elongation, as determined by Pol II-Ser2P, is similar at lincRNAs and protein-coding genes. Combinatorial assembly of transcription factors (TFs) and coactivators is also likely distinct at the various regulatory elements presented here. Transcription initiation platforms (TIPs) represent large arrays of initiating Pol II transcription (larger than 400 bp and up to 10 kb) that were characterized at both active and tissue-specific promoters and

enhancers. Finally, CpG islands are mainly associated with promoters. Although they are also found at lincRNA promoters, they are not found at enhancers. Abbreviations: GTFs, general transcription factors; Med, mediator complex; TFs, transcription factors; TSS, transcription start site. Figure adapted from (Natoli and Andrau, 2012)

Histone acetylation occurs at the lysine residue by the enzyme histone acetyl transferases (HATs) and it is associated with transcription activation. However, the histone deacetylases (HDACs) remove the acetyl group and thereby suppresses the transcription. Dynamic regulation of acetylation and deacetylation of chromosomes have been shown to play a very important role in the regulation of gene expression and propagation of disease.

Like histone acetylation, the phosphorylation of histones is highly dynamic and is controlled by kinases (RPS6KA5, RPS6KA4 and BAZ1B) and diphosphatases (serine/threonine protein phosphatase, PPPs) that add and remove the modification, respectively. The phosphatase group readers are BRCT domain-containing proteins (BRCTs).

Ubiquitination has also been shown to be involved in transcriptional activation *in vivo*. H2BK120 ubiquitination has been associated with the transcribed regions of highly expressed genes (Minsky et al., 2008). The role of this modification seems to function through multiple mechanisms to enhance transcription. For example, H2BK120ub increases the methylation state of H3 associated with sites of active transcription (Dover et al., 2002). This modification is also associated with increases in transcriptional elongation through a mechanism involving H2A/H2B displacement (Pavri et al., 2006). In addition, H2B ubiquitination has been suggested to be involved in chromatin condensation during DNA repair (Moyal et al., 2011).

There are two main ways of how histone modifications can affect transcriptional regulation. Firstly, histone modifications may directly change the chromatin structure or its dynamics. For example, acetylation of a lysine neutralizes its positive charge and reduces the affinity of positive charge on histone to the negative charge on DNA, therefore loosen the chromatin (Choi and Howe, 2009). Secondly, histone modifications can act indirectly as signals to be recognized by “readers” who translate these modifications into transcriptional outcome (Strahl and Allis, 2000).

### **DNA Methylation**

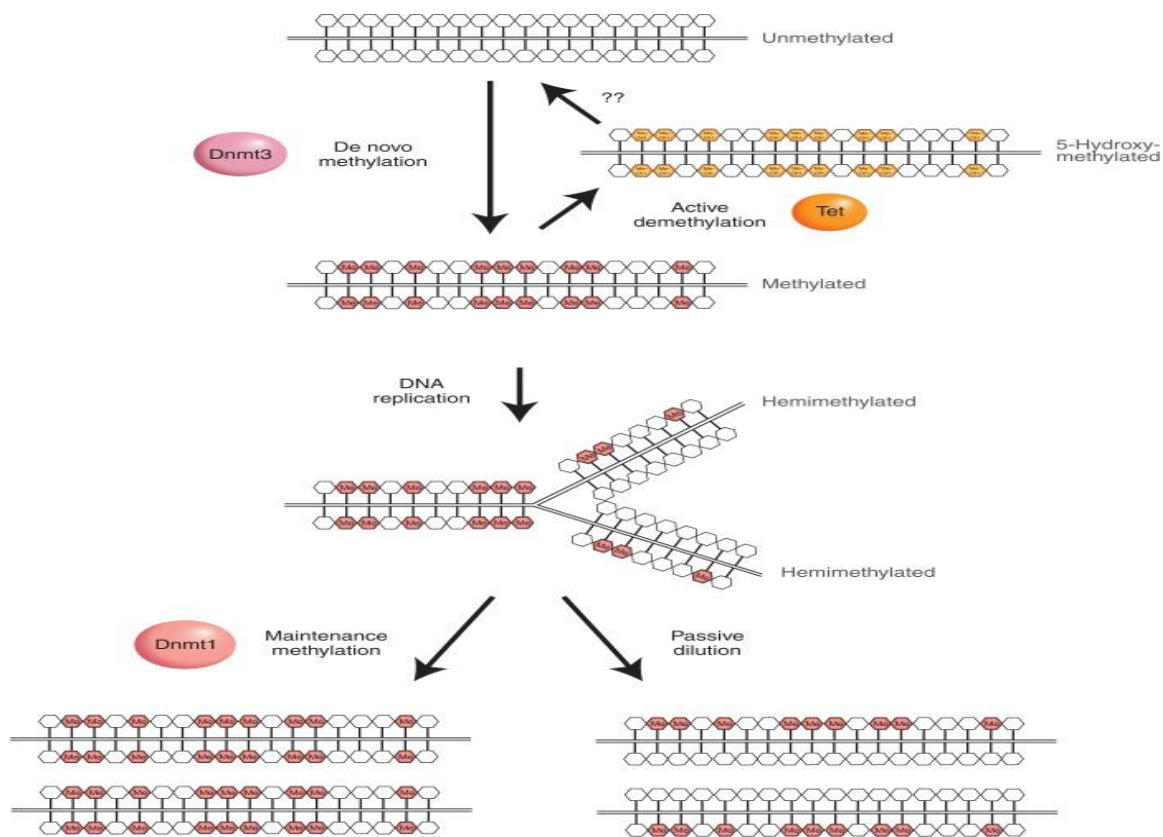
The DNA of vertebrate animals can be covalently modified by methylation of the cytosine base in the dinucleotide sequence 5'CpG3'. In mammals, 5-methylcytosine (5mC) is the major form of DNA modification, and it has important roles in development and disease. CpG is an abbreviation for cytosine and guanine separated by a phosphate, which links the two nucleotides together in DNA. In mammals, DNA methylation patterns are established during embryonic development by *de novo* methylating enzymes called DNA methyltransferase 3a (Dnmt3a) and DNA methyltransferase 3b (Dnmt3b) (Ooi and Bestor, 2008). They are maintained by a Dnmt1-mediated copying mechanism when cells divide. The heritability of DNA methylation pattern provides an epigenetic marking of the genome that is stable through multiple cell divisions and therefore constitutes a form of cellular memory (Fig. 1.3). For this

reason, historically, DNA methylation has represented the archetypal mechanism of epigenetic inheritance.

Molecular and genetic studies in mammals have shown that DNA cytosine methylation (abbreviated to 5mC, for 5-methyl cytosine) is associated with gene silencing (Wu and Zhang, 2011). About 60–80% of the CpG sites in the mammalian genome are modified by 5mC (Tahiliani et al., 2009). It also plays an important role in developmental processes such as X-chromosome inactivation and genomic imprinting. The methyl moiety of methyl cytosine resides in the major groove of the DNA helix where many DNA-binding proteins make contact with DNA. The methylation, therefore, likely exerts its effect by attracting or repelling various DNA-binding proteins. A family of proteins, known as methyl-CpG binding domain proteins (or MBDs), are attracted to and bind DNA-containing methylated CpG dinucleotides and have been shown to recruit repressor complexes to methylated promoter regions, thereby contributing to transcriptional silencing (Wu and Zhang, 2011). Conversely, regions of CpG methylation are known to prevent protein binding of certain transcription factors, thereby preventing transcription.

Certain regions of the genome contain clusters of CpG sequences, termed CpG islands, and are mostly found directly upstream of gene promoters. In general, CpG islands are DNA methylation-free. Certain transcription factors have been discovered to bind to nonmethylated CpG-containing DNA sequences via a CXXC binding domain motif and contribute to creating a transcriptionally competent chromatin configuration, preventing DNA methylation from occurring at these regions (Blackledge et al., 2010).

Although DNA methylation patterns can be transmitted from cell to cell, they are not permanent. In fact, changes in DNA methylation patterns can occur throughout the life of an individual. Some changes can be a physiological response to environmental changes, whereas others might be associated with a pathological process such as oncogenic transformation or cellular aging. DNA methylation marks can be removed by either an active demethylation mechanism involving a family of DNA hydroxylases called Tet proteins or a passive demethylation process by inhibition of the maintenance methyltransferase, Dnmt1, during cell divisions (Fig. 1.3). DNA methylation patterns fit into an epigenetic framework directly, but also indirectly through their intimate link to other epigenetic mechanisms such as histone lysine methylation and acetylation.



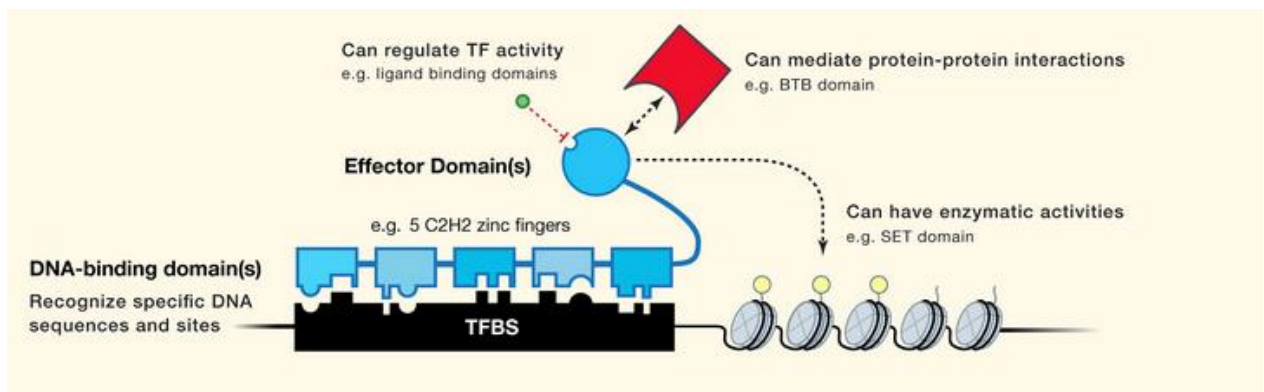
**Figure 1.3 | De novo methylation and maintenance methylation of DNA.** A stretch of genomic DNA is shown as a line with self-complementary CpG pairs marked as vertical strokes. Unmethylated DNA (*top*) becomes methylated “de novo” by Dnmt3a and Dnmt3b to give symmetrical methylation at certain CpG pairs. On semiconservative DNA replication, a progeny DNA strand is base-paired with one of the methylated parental strands (the other replication product is not shown). Symmetry is restored by the maintenance DNA methyltransferase, Dnmt1, which completes half-methylated sites, but does not methylate unmodified CpG (Li and Zhang, 2014).

## 2. Transcription factors

Historically, the term transcription factor has been applied to describe any protein involved in transcription and/or capable of altering gene-expression levels. In the current vernacular, however, the term is reserved for proteins capable of (1) binding DNA in a sequence-specific manner and (2) regulating transcription (Fig. 1.4)(Fulton et al., 2009). Transcription factors (TFs) directly interpret the genome, performing the first step in decoding the DNA sequence. Many function as “master regulators” and “selector genes”, exerting control over processes that specify cell types and developmental patterning (Lee and Young, 2013) and controlling specific pathways such as immune responses (Singh et al., 2014). Mutations in TFs and TF-binding sites underlie many human diseases. Their protein sequences, regulatory regions, and physiological roles are often deeply conserved among metazoans (Carroll, 2008), suggesting that global gene regulatory “networks” may be similarly conserved. And yet, there is high turnover in individual regulatory sequences (Weirauch and Hughes, 2010), and over longer timescales, TFs duplicate and



diverge. The same TF can regulate different genes in different cell types (e.g., ESR1 in breast and endometrial cell lines (Gertz et al., 2012), indicating that regulatory networks are dynamic even within the same organism. TFs can have 1,000-fold or greater preference for specific binding sequences relative to other sequences (Gertz et al., 2012). Because TFs can act by occluding the DNA-binding site of other proteins (e.g., the classic lambda, lac, and trp repressors (Ptashne, 2011)), the ability to bind specific DNA sequences alone is often taken as an indicator of ability to regulate transcription. Schematic of a prototypical TF is shown in (Fig 1.4)



**Figure 1.4 | Schematic of a prototypical TF** (Lambert et al., 2018).

During transcription, transcription initiation by RNA Polymerase II (RNAP<sub>II</sub>) minimally requires the basal initiation factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Thomas and Chiang, 2006). These so-called general transcription factors (GTFs) mediate promoter recognition and unwinding, and together with RNAP<sub>II</sub> and promoter DNA comprise the Pre-Initiation Complex (PIC) (Sikorski and Buratowski, 2009). The mediator complex works as a co-activator that bridges regulatory factors and the basal machinery to allow high levels of activator-dependent transcription (Thomas and Chiang, 2006). Mediator is not required for basal transcription in purified systems but can stimulate transcription in these systems even in the absence of activators (Kornberg, 2005). Mediator may also directly stabilize PIC assembly intermediates (Baek et al., 2006). Two transcription repressors, Mot1/BTAF1 and NC2, act through direct interactions with TBP. Mot1/BTAF1 is a Snf2 family ATPase that removes TBP from promoters. It also behaves genetically as a repressor. NC2 is a heterodimer that blocks TFIIA and TFIIB from associating with the TBP-TATA complex. Its genetic properties are also consistent with transcription repression (Auble, 2009). To explain the apparent paradox of inhibitory complexes binding to active promoters, it has been proposed that these repressors (particularly Mot1) displace TBP from cryptic TATA sequences or other inappropriate genomic locations in order to make it available to weaker promoters (Dasgupta et al., 2002; Spedale et al., 2012). Several recent reports support this model. NC2 alters the conformation of the TBP/DNA complex, allowing it to slide away from the TATA boxes (Schluesche et al., 2007). TBP mutants with a decreased ability to form PICs suppress the

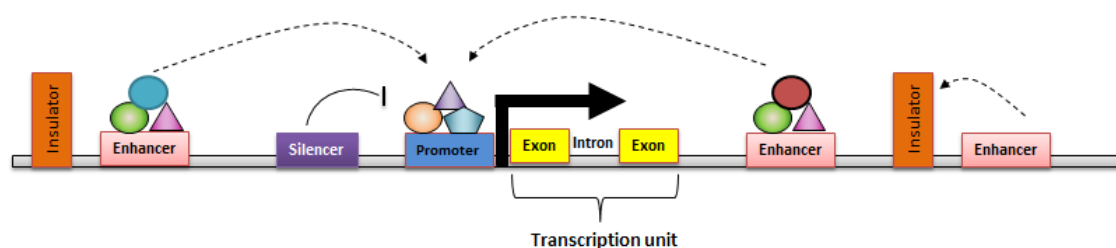
gene expression defects seen in a *mot1* mutant (Sprouse et al., 2009). FRAP experiments show that the rapid exchange of TBP associated with chromatin is dependent upon Mot1 (Sprouse et al., 2008).

The most crucial property of TFs is that they contain DNA binding domains (DBDs) which attach to specific sequences of DNA when conformed. Factors that lack DBDs are not classified as transcription factors. In addition, TFs possess activation or repression domains that function either to up- or down-regulate the expression of adjacent genes by binding to their regulatory elements. TFs recognize DNA sequences by a short DNA binding motif (6-12 bp) which is called transcription factor binding site (TFBS). TFBS is a consensus sequence in which certain positions are relatively constrained (4 to 6 bp) and others are more variable.

TF DNA-binding specificities are frequently summarized as “motifs” models representing the set of related, short sequences preferred by a given TF, which can be used to scan longer sequences (e.g., promoters) to identify potential binding sites. There are various mechanisms by which TFs regulate gene expression. For example, TFs can recruit and stabilize the binding of RNAPII, or catalyze the acetylation or deacetylation of histone proteins or recruit coactivator or corepressor proteins through protein-protein interactions to the transcription factor DNA complex.

### 3. Cis-regulatory elements

Cis-regulatory elements (CREs) are regions of non-coding DNA which regulate the transcription of neighboring genes. CREs is found in the vicinity of the genes that they regulate. CREs typically regulates gene transcription by binding to transcription factors. A single transcription factor may bind to many CREs, and hence control the expression of many genes (pleiotropy). CREs consists of enhancer, promoter, silencer and insulators. These elements act in cooperation with one another to govern a coordinated expression pattern of a gene. They are summarized in (Fig 1.5).

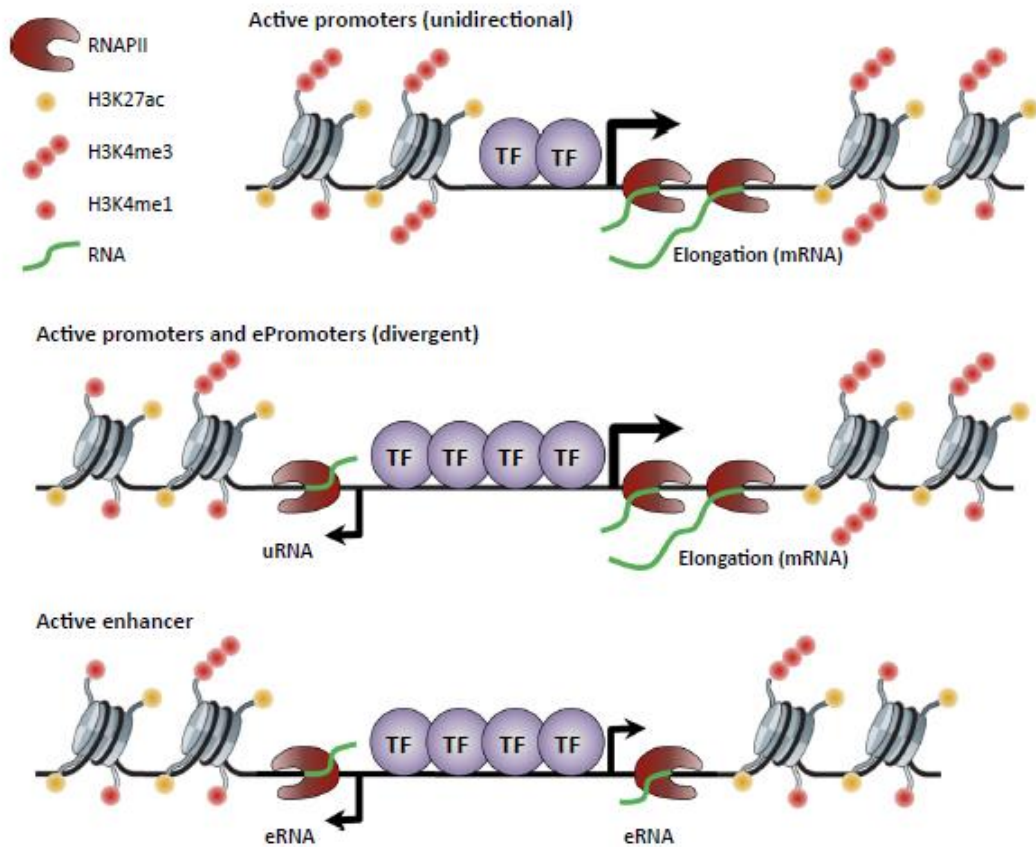


**Figure 1.5 | Landscape of *cis*-acting DNA regulatory elements.** Depending on the context, the promoter may be activated by the enhancers or kept in a repressed state by a silencer. In the other case, enhancers or silencers cannot influence activity due to the shielding effect of insulator.

### 3.1. Promoters

Transcription of a gene in eukaryotes is a highly complex process that requires precise coordination in the assembly of trans-acting factors through the recognition of various types of regulatory DNA sequences. The promoter and the enhancer represent DNA regulatory regions responsible for ensuring proper spatiotemporal expression patterns of eukaryotic genes. The promoter generally refers to a DNA region that allows accurate transcription initiation of a gene (Smale and Kadonaga, 2003). The core promoter is a minimal stretch of DNA sequences (e.g., the TATA box, initiator, and downstream core promoter element) surrounding the transcription start site that directly interacts with the components of the basal transcription machinery, including RNA polymerase II (RNAPII). Although the DNA sequences or motifs comprising the core promoter region for individual genes can be structurally and functionally diverse, its universal role is thought to drive accurate transcription initiation (Smale and Kadonaga, 2003). Transcription factors that bind ~100–200 bp upstream of the core promoter can increase the rate of transcription by facilitating the recruitment or assembly of the basal transcription machinery onto the core promoter or by mediating the recruitment of specific distal regulatory DNA sequences to the core promoter (Akbari et al., 2008).

The promoter of an active gene is commonly located in a nucleosome free region (NFR) flanked by two well-positioned nucleosomes (+1 and -1 nucleosomes). This nucleosome-free region makes the promoter more accessible and facilitates transcription machinery assembly. These promoter-associated nucleosomes display specific histone variants and histone modifications. Particularly, in mammals, the two adjacent nucleosomes are enriched for histone variants H3.3/H2A.Z (Jin et al., 2009). Another important feature of nucleosome flanking promoters is that they are marked by specific histone modifications, which have been shown to correlate with promoter activity. While the H3K4me3 mark is associated with active promoters, H3K27me3 is associated with repressed promoters (Barski et al., 2007; Lenhard et al., 2012). The other well studied mark is H3K27ac which is also associated with active promoters. In mammals, the vast majority of promoters overlap with CpG islands (Lenhard et al., 2012). The advent of high-throughput sequencing has allowed us to map transcription initiation with an unprecedented sensitivity and resolution. This has revealed that cis-regulatory elements are commonly associated with transcriptional initiation sites flanking the regulatory sequences (Fig. 1.6).



**Figure 1.6 | Chromatin Structure at 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. Abbreviations: ePromoter, enhancer-like promoter; RNAPII, RNA polymerase II; TF, transcription factor; TSS, transcription start site; uRNA, upstream RNA; eRNA, enhancer RNA. Figure adapted from (Medina-Rivera et al., 2018).

### 3.2. Enhancers

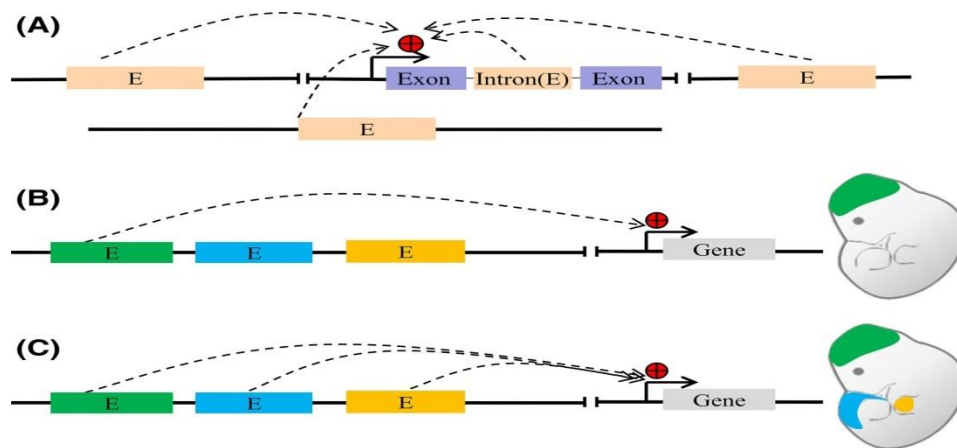
Enhancers are regulatory elements that activate transcription over large distances and independently of orientation. These cis-regulatory elements are generally located distally with respect to the 5' end of genes (Medina-Rivera et al., 2018). They can positively regulate spatiotemporal gene expression during development through either cis- or trans interaction manner (Fig. 1.7) (Ong and Corces, 2011). In 1981, enhancer was first described as a 72-bp repeated sequence in simian virus 40 (SV40) genome, which could increase the ectopic expression of a reporter gene by ~200-fold (Banerji et al., 1981; Benoist and Chambon, 1981). In 1983, enhancer was discovered within a mouse immunoglobulin heavy chain gene in mammals (Schaffner, 2015). Subsequently, different enhancers in various cells and tissues have been reported (Shlyueva et al., 2014a).

Enhancer activity is usually linked with certain properties of chromatin. Active enhancers are typically bound with transcription factors (TFs) (Spitz and Furlong, 2012). The flanking of enhancers are commonly marked by histone modifications such as histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac) (Calo and Wysocka, 2013). Active enhancers are marked by both H3K4me1 and H3K27ac, with depletion of histone H3 lysine 4 trimethylation (H3K4me3) (Calo and Wysocka, 2013); inactive, poised enhancers are marked only with H3K4me1 (Rada-Iglesias et al., 2011). In addition, enhancers are typically depleted of nucleosomes and sensitive to DNase I digestion (Thurman et al., 2012). Distal enhancers are brought into close proximity with their target promoters through chromatin looping (Shlyueva et al., 2014a), which is facilitated by mediators and cofactors (Spitz and Furlong, 2012). Moreover, active enhancer can recruit RNAPII and produce RNAs that contributes to its function and gene regulation (Natoli and Andrau, 2012).

H3K4me1 and H3K27ac are commonly used hallmarks to identify putative enhancers genome-wide (Buecker and Wysocka, 2012). H3K4me1 and H3K27ac are conferred by the mixed lineage leukemia (MLL) family of methyltransferase (MLL2/3/4) and the CREB-binding protein (CBP)/P300 acetyltransferases, respectively (Buecker and Wysocka, 2012). Knocking out H3K4 methyltransferases MLL3 and MLL4 resulted in a global loss of H3K4me1 binding, and subduction of H3K27ac, mediators and RNAPII bindings as well (Hu et al., 2013a). It has been found that H3K4me1 can facilitate recruitment of the cohesion complex to chromatin, which provides a potential mechanism for MLL3/4 to promote chromatin interactions between enhancers and promoters (Yan et al., 2018). In addition, a recent study has suggested that H3K4me1 might play a fine-tuning role in enhancer activity by facilitating the binding of the BAF complex and possibly other chromatin regulators (Yan et al., 2018). Meanwhile, active enhancers in both flies and mice are not necessarily marked by H3K27ac, but H3K27ac has been supposed to affect enhancer activity through destabilizing nucleosomes or recruiting H3K27ac-binding proteins (Catarino and Stark, 2018). All these evidences imply that H3K4me1 and H3K27ac themselves are not required for enhancer activity. However, it is demonstrated that the presence of H3K4me3 is fully compatible with enhancer activity (Core et al., 2014; Ernst et al., 2011; Koch et al., 2011; Pekowska et al., 2011; Wang et al., 2008b); the level of H3K4me3 is actually positively correlated with the enhancer strength (Andersson et al., 2014; Core et al., 2014; Henriques et al., 2018; Vanhille et al., 2015b). 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 (Andersson, 2015; Medina-Rivera et al., 2018) (Fig 1.6).

Super-enhancers have emerged as clusters of enhancers that are densely occupied by the master regulators and mediators, which is speculated to act as switches to determine cell identity and fate (Hnisz et al., 2013). This notion was first described as genomic regions with high levels of five master transcription factors (Oct4, Sox2, Nanog, Klf4, and Esrrb) and the Mediators in mESCs (Whyte et al., 2013). Subsequent studies have extended the concept of super-enhancers as genomic regions densely occupied by high levels of

H3K4me1, H3K27ac, p300 or master transcription factors in multiple cell types and tissues (Hnisz et al., 2013).



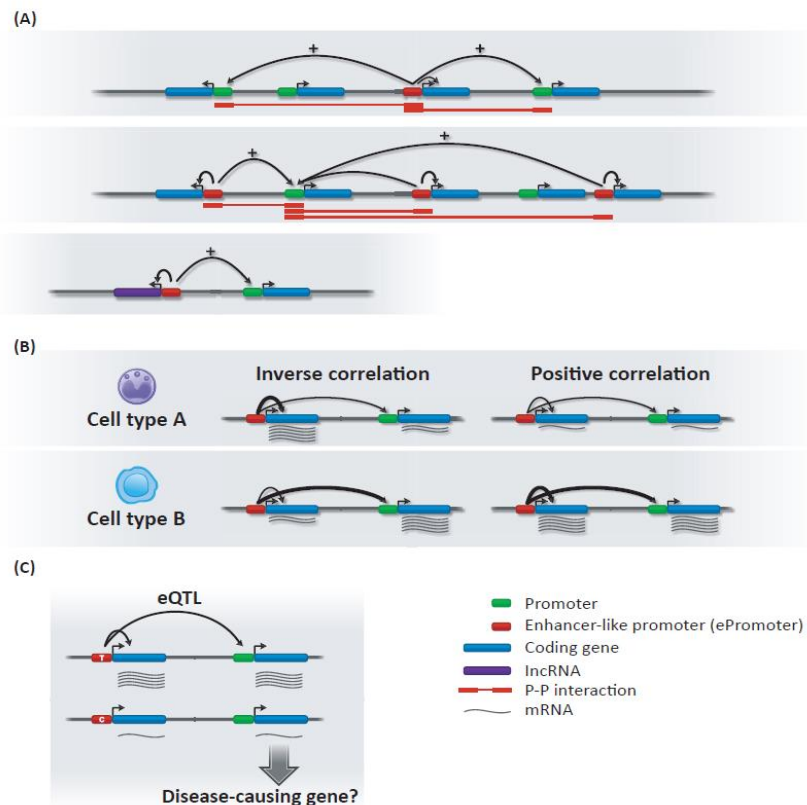
**Figure 1.7** | (A) Enhancers are *cis*-regulatory elements that can increase expression of target genes in *cis* and *trans*-acting manner; (B and C) Enhancer regulate spatiotemporal gene expression (Peng and Zhang, 2018).

Enhancer-promoter interactions can be commonly found to determine spatiotemporal gene expression patterns in eukaryotes (Novo et al., 2018). This has been well presented by studies of the globin locus control region (LCR) and its target gene (Snetkova and Skok, 2018). During erythroid development, LCR activates distinct globin genes in a stage specific manner through the formation of DNA looping. LCR- $\beta$ -globin interactions are established dependent on gene-specific transcription factors, including the hematopoietic-specific factors GATA1 and FOG1, KLF, and the widely expressed factor LDB1 (Song et al., 2007). The depletion of LDB1 has been previously reported to disrupt long-range LCR loop formation, and thus affect gene transcription (Song et al., 2007). There are other examples of specific gene regulation involving enhancer-promoter looping. The *Satb1* gene is silent when its promoter does not contact with enhancers in the brain, whereas it is highly expressed when enhancer-promoter looping has been de novo formed in the thymus (van de Werken et al., 2012).

The protein yin and yang (YY1) has been recognized as a structural mediator of DNA looping (Weintraub et al., 2017). YY1 could globally mediate enhancer-promoter interactions by binding to DNA and facilitate the formation of chromatin loops, probably through its dimerization (Weintraub et al., 2017). In addition, YY1 has been further indicated to positively regulate transcription by targeting promoters and enhancers through the BAF complexes in embryonic stem cells (Weintraub et al., 2017).

Despite the conventional distinction between them, promoters and enhancers share many genomic and epigenomic features (Andersson, 2015; Medina-Rivera et al., 2018). 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 approaches have indicated that this phenomenon is common and might have a strong impact on our global

understanding of genome organization and gene expression regulation (Dao et al., 2017; Diao et al., 2017; Medina-Rivera et al., 2018). Several independent studies using CRISPR genome editing demonstrated that some promoters function as enhancers in their endogenous context (Fig 1.8).



**Figure 1.8 | Role of Enhancer-like Promoters in Gene Regulation.** (A) ePromoter (red) interacts with one or more distal promoters (green) and activates the expression of neighbor genes (top). A given gene might be regulated by several ePromoters located in the neighborhood (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 results in the coordinated regulation of neighbor genes upon stress or cell-type specific signaling. (C) Genetic variants (e. g. eQTL or genome-wide association study single nucleotide polymorphism) lying with in an ePromoter might influence the expression of neighbor 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. Figure adapted from (Medina-Rivera et al., 2018).

### 3.3. Silencers

An understanding of transcriptional repression is also essential for a complete understanding of promoter structure and the regulation of gene expression. Transcriptional repression in eukaryotes is achieved through 'silencers', which are sequence-specific elements that induce a negative effect on the transcription of its particular gene (Ogbourne and Antalis, 1998). There are two main types of silencers, namely 'silencer elements' and 'negative regulatory elements' (NREs). Silencer elements are classical, position-independent elements that direct an active repression mechanism, and NREs are position dependent elements that direct a passive repression mechanism. Silencers form an intrinsic part of many eukaryotic loci. The knowledge of their interactive role with promoters and enhancers, as well as other transcriptional elements, is essential for our understanding of gene regulation in eukaryotes. In Chapter 2, I provide a detailed description of silencer functions and mechanisms of action.

### 3.4. Insulators

Insulators (also known as boundary elements) function to block genes from being affected by the transcriptional activity of neighboring genes. They thus limit the action of transcriptional regulatory elements to defined domains and partition the genome into discrete realms of expression. Insulators have two main properties: (a) they can block enhancer-promoter communication (i.e., enhancer-blocking activity), and (b) they can prevent the spread of repressive chromatin (i.e., heterochromatin-barrier activity). For at least some insulators, these two activities can be separable (Recillas-Targa et al., 2002). Typically, insulators are ~0.5–3 kb in length, and function in a position dependent or orientation independent manner. In vertebrates, the most well characterized insulator element is the chicken  $\beta$ -globin insulator, 5'HS4 (Felsenfeld et al., 2004); a homologous element resides in the human  $\beta$ -globin gene locus (Li et al., 1999). Insulator elements have also emerged as a recurrent feature of a number of imprinted loci in the human genome; the most well-characterized example is the imprinting control region (ICR) located upstream of the H19 gene that modulates allele-specific transcription of H19 and another gene, Igf2 (Ling et al., 2006). The number of insulator elements in the human genome is not known. It is now thought that genuine insulator elements may be less common than initially envisaged, and found only in regions with a high density of coding or regulatory information (Fourel et al., 2004). Although a number of transacting factors that mediate insulator activity have been identified in *Drosophila*, the only known protein to mediate such activity in vertebrates is CTCF (CCCTC-binding factor). CTCF has been implicated to play a role in many different loci, including chicken globin 5'HS4 (Parelho et al., 2008); and the mammalian H19/Igf2ICR (Bell and Felsenfeld, 2000). The activity of CTCF can be regulated by a number of means, including DNA methylation, post-translational modification, and interaction with cofactors (West and Fraser, 2005).

Moreover, the predominant models of how insulators work are the looping model and the decoy model (Herold et al., 2012). The looping model explains that two or more insulator sites physically interact with each other, thereby establishing loops that alter the 3D conformation of the chromatin fiber in a manner that affects the interaction between enhancers and promoters. In the decoy model, insulators interact directly with an



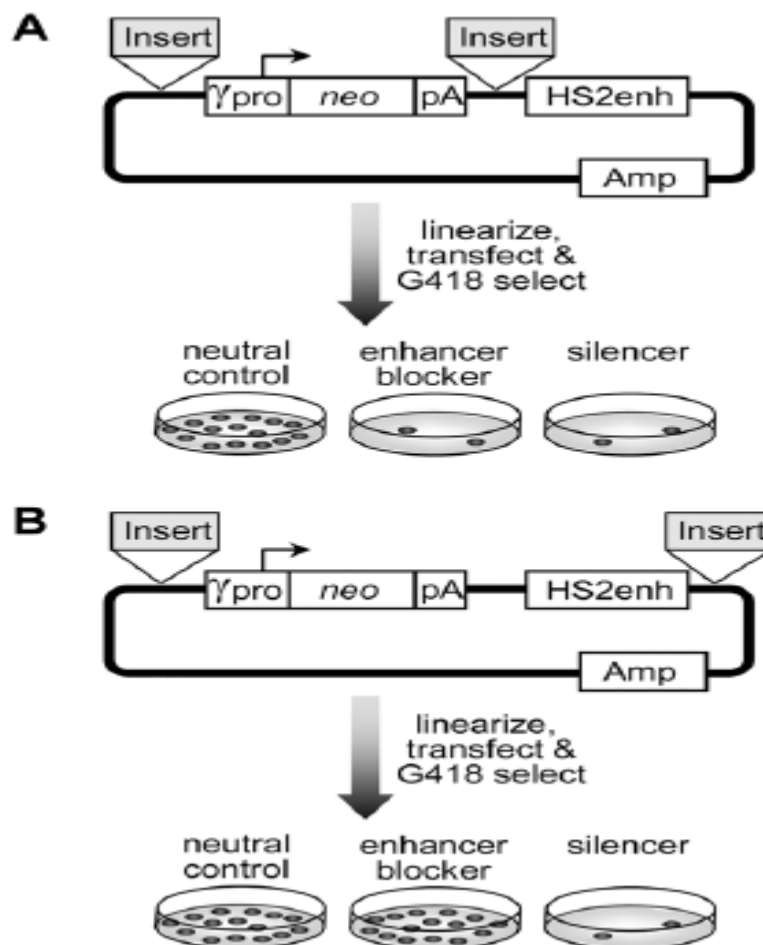
enhancer or promoter, thus, interfering with their communication. The insulating function is thought to depend on the binding of the CTCF factor to these regions.

Notably, experiments and mathematical models suggested that enhancer-promoter interactions could be sufficiently reproduced by a topological domain model wherein the topological domains were assumed to be formed by chromatin loop forming protein binding to insulator sequences (Gohl et al., 2011). In particular, the promoter could associate with the enhancer in the same chromatin domain more frequently than with that in the neighboring chromatin domain even if the genomic distances from both enhancers to the promoter were identical (Mukhopadhyay et al., 2011). To function as chromatin domain boundaries, it is necessary for the insulator sequences to associate with other insulator sequences via their binding proteins, such as the CTCF-cohesion complex (Merkenschlager and Nora, 2016). Alternatively, some nucleosomeexcluding sequences such as (A)<sub>n</sub>or (CCGNN)<sub>n</sub> repeat sequences and the core 182 bp region of *Ars* insulator sequences (*ArsInsC*) found in sea urchin genome have been shown to independently exhibit insulator activities (Bi et al., 2004). For these sequences, the specific binding of proteins such as CTCF to form chromatin loops was not identified. Moreover, the insulator activities of CTCF binding sequences are highly dependent on their intra-genomic direction (Rao et al., 2014), whereas those of nucleosome excluding sequences are robust (Bi et al., 2004). This indicates the possibility that such nucleosomeexcluding sequences may exhibit insulator activities without the formation of chromatin loops; herein these are therefore termed as nucleosome excluding non-looping insulator sequences (NENLIS). However, the physical mechanism of the NENLIS-mediated insulator activities remained unknown.

Furthermore, the *Ars* insulator sequence was demonstrated as showing insulator activities not only in sea urchin cells but also in cells of various organisms including humans and plants (Yajima et al., 2007). This suggested that NENLIS might be distributed in various chromosomal regions and exhibit insulator activities across different organisms. However, despite the exhaustive epigenome and 3D chromosome structure data obtained via Hi-C experiments from various research projects, little progress has been made regarding the genome-wide analysis of such sequences (Bernstein et al., 2010).

An important issue is that there can be a potential overlap between Insulators and Silencers, especially when tested in reporter assays, as both types of elements are predicted to negatively influence gene expression. To overcome this problem, it is essential to consider the relative position of the tested elements with respect to the enhancers and promoters. For examples, Qi et al., used a reporter assay strategy to distinguish between enhancer-blocking and silencer activity (Fig.1.9). The reporter plasmid used for this assay contains an expression cassette for the bacterial drug resistance gene *neo* transcribed by an erythroid promoter from the  $\gamma$ -globin gene *HBG1* and terminated with the SV40 polyadenylation signal. An erythroid HS2 enhancer from the  $\beta$ -globin locus control region is located 3' of this cassette. Candidate elements are inserted 5' of the expression cassette and 3' between the expression cassette and the enhancer. After being linearized, the plasmid is transfected into erythroid K562 cells,

which are then analyzed for the frequency of colony formation under selection with the neomycin drug analog G418. Both enhancer-blocking insulator and silencers are expected to reduce the rate of G418 resistant colony formation (Fig.1.9A). The reporter plasmid is the same as for the enhancer-blocking assay, except that candidate elements are inserted 5' of the neo expression cassette and 3' of the HS2 enhancer in a manner that brackets both elements and allows the enhancer to interact with the promoter. As above, the plasmids are linearized and transfected into K562 cells, which are then analyzed for the frequency of drug resistant colony formation. In this case, only silencers are expected to reduce the rate of G418-resistant colony formation (Fig.1.9B) (Qi et al, 2015).



**Figure 1.9 | Schema for drug-resistant colony assay.** (A) Enhancer-blocking and silencer assay. (B) Silencer-only assay (Qi et al, 2015).

#### 4. Long non-coding RNAs

lncRNAs are long RNA transcripts that do not encode proteins and longer than 200 nt. Many lncRNAs are very much like mRNAs: they are transcribed by RNA polymerase II (Pol II) from genomic loci with similar chromatin states to mRNAs (Guttman et al., 2009); they are often 5'-capped, spliced and polyadenylated; in most instances, they lack any biochemical distinction from mRNAs besides the absence of a translated ORF. But

there are also general trends that discriminate lncRNAs from mRNAs: lncRNAs tend to be shorter than mRNAs, have fewer but longer exons, be expressed at relatively low levels and exhibit poorer primary sequence conservation (Derrien et al., 2012). Of course, exceptions to all of these trends abound. As a class, lncRNAs run the gamut from mRNA-like to truly exotic, such as chemically circular RNAs (Salzman et al., 2012), lncRNAs spanning 100 kb and abundant lncRNAs with a restricted subnuclear localization (Hutchinson et al., 2007).

Transcriptome wide studies showed that lncRNAs in general exhibit more specific expression profiles than mRNAs (Derrien et al., 2012); that is, they are expressed in a cell type, tissue, developmental stage or disease state specific manner (Yang et al., 2014). This trend is true even after correcting for the markedly lower expression levels of lncRNAs (Derrien et al., 2012). Furthermore, lncRNA expression patterns are often correlated with mRNA expression patterns both in *cis* (neighboring genomic environment) and in *trans* (distant sites of action), suggesting that certain lncRNAs may be co-regulated in expression networks (Guttman et al., 2009).

From the perspective of the chromatin state, it appears that lncRNAs follow the same rules as protein-coding genes (Derrien et al., 2012). That is to say, expressed lncRNA promoters are enriched for active histone modifications (for example, H3K4me3, H3K9ac and H3K27ac) similar to their protein-coding counterparts, and histone modification patterns can be used to identify lncRNAs (Derrien et al., 2012). However, recent work has revealed that certain transcription factors and chromatin remodeling enzymes globally regulate lncRNA expression.

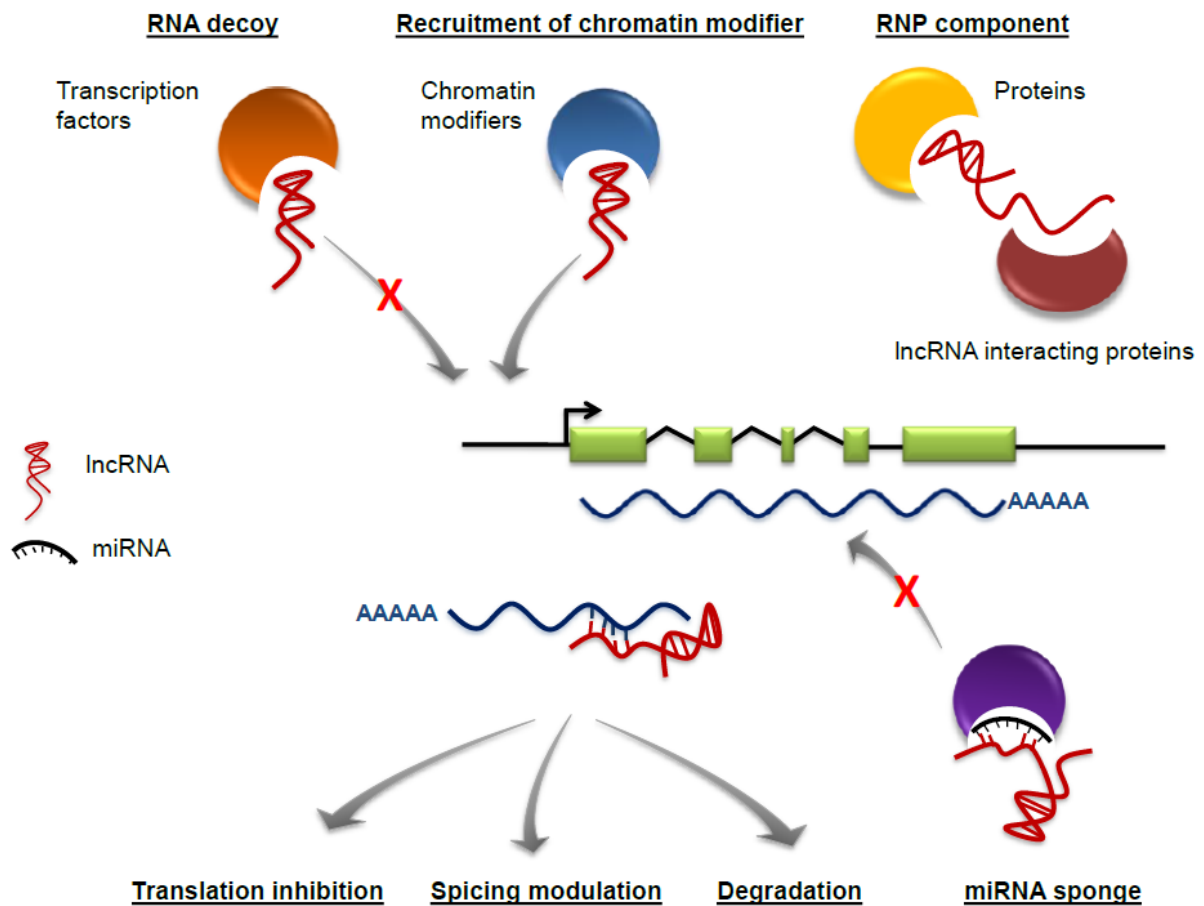
In addition, from a genetic RNA interference screen in yeast, four mechanistically distinct chromatin remodeling complexes (Swr1, Isw2, Rsc, and Ino80) were identified as global repressors of ncRNA transcription, particularly of lncRNAs that overlap protein-coding genes (Alcid and Tsukiyama, 2014). Disruption of these complexes led to the derepression of antisense lncRNAs and resulted in a marked decrease in levels of their overlapping mRNAs. Therefore, in yeast, these chromatin remodelers may repress such lncRNAs in order to activate the sense-strand mRNAs in *cis*. Other studies using different reporter-based screens in yeast also identified chromatin remodeling and nucleosome assembly factors as key regulators of intragenic cryptic unstable transcripts (CUTs) and divergent lncRNAs (Cheung et al., 2008).

To reach their mature forms, nascent RNA transcripts undergo extensive co-transcriptional and post-transcriptional processing events, such as 5'-capping, splicing, polyadenylation and chemical base modification. During this impressionable period in the life of RNA, some lncRNAs experience alternative forms of processing that distinguish them from other transcripts (Wilusz, 2016).

Many lncRNA species have very well defined subcellular localizations, including XIST (on the inactive X), Gomafu (also known as MIAT; subnuclear domains) (Tsuiji et al., 2011), BORG (restricted to the nucleus) (Costantini et al., 2007) and GAS5 (exported to

the cytoplasm) (Liu et al., 2018). More generally, compared to mRNAs, lncRNAs are more enriched in the nucleus relative to the cytoplasm (Derrien et al., 2012).

Studies on the function of lncRNA expression showed that they are differentially expressed during differentiation, development or in response to stimuli (Dinger et al., 2008). Many lncRNAs have found to be involved in a wide variety of cellular processes, cell differentiation and implicated in many diseases, reviewed in (Batista and Chang, 2013). lncRNAs might mediate epigenetic modifications of DNA by acting as modular scaffolds for recruiting chromatin remodeling complexes to specific loci (Wang et al., 2011). Many lncRNAs seem to bind to specific combinations of regulatory proteins, potentially acting as scaffold elements within ribonucleoprotein complexes (Batista and Chang, 2013). Homology between the lncRNA and its target promoter region can elicit transcriptional silencing. lncRNA transcripts can have critical functions on expression and repression of nearby genes. For instance, the transcription of ncRNA from the upstream promoter region involves both a direct interaction with transcription factor binding and promoter-specific interference (Patil et al., 2014). Some enhancer-associated lncRNAs that are expressed during embryonic stem cell differentiation to form cardiomyocytes (Di Salvo et al., 2015). Enhancer lncRNA-HIT (Hotchon) regulates the expression of nearby genes (H3K27ac and 5'HoxA) by recruiting chromatin-modifying proteins (CMPs) (p100 and CBP) to increase the accessibility of the chromosomal region to gene-regulatory factors (Carlson et al., 2015). lncRNAs can compete with miRNA for their binding to mRNA, thus, act as miRNA sponge (Cesana et al., 2011) (Fig 1.10 ).

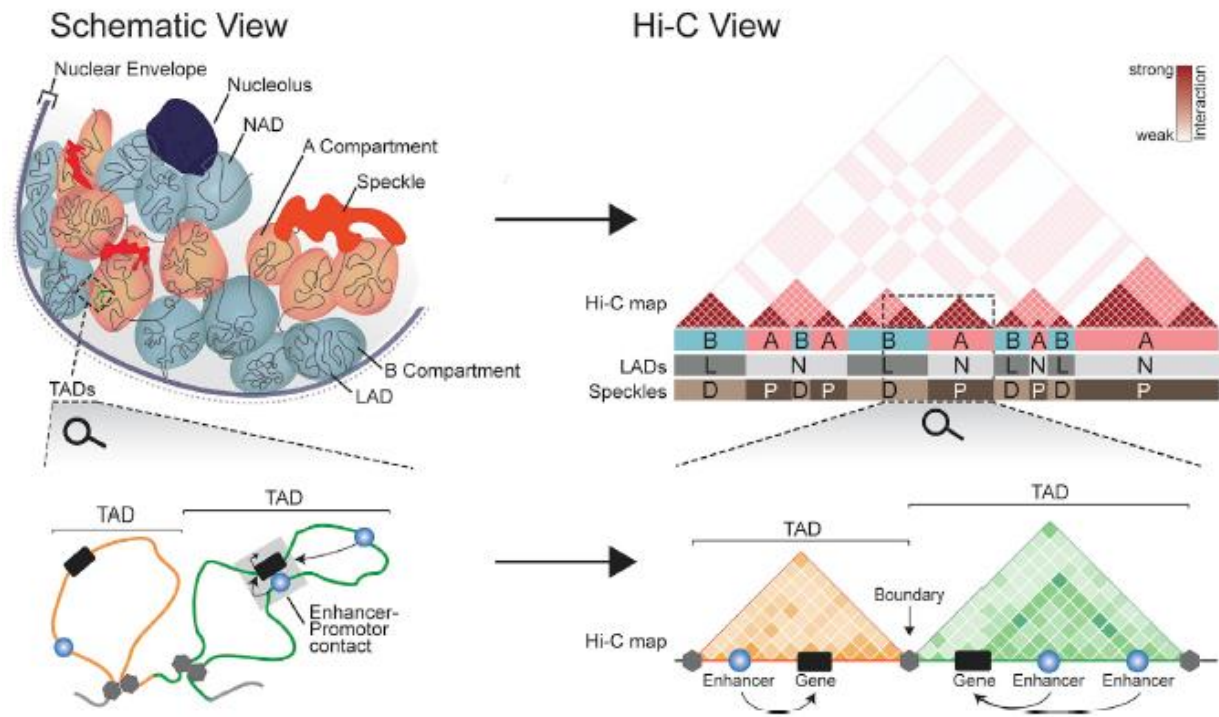


**Figure 1.10| Regulation of gene expression by lncRNAs.** Figure adapted from (Mathieu et al., 2014).

## 5. Long-range chromatin interactions

In living cells, chromosomes are well-organized in three dimensions inside the nucleus forming separated chromosome territories (CTs) (Cremer and Cremer, 2001). In each territory, the interchromosomal interaction of particular chromosomes and long-range interactions between genomic regions is often occurred. The position of CTs is thought to correlate with transcriptional activity. Transcriptionally inactive regions are located at nuclear periphery (nuclear lamina) (Padeken and Heun, 2014), while regions with similar transcription activity are co-localized in the nuclear space called transcription factories where they are likely sharing transcription machinery (Edelman and Fraser, 2012; Papantonis and Cook, 2013). At increasing resolution, each chromosome is comprised of many distinct chromatin domains which referred to as topological associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012). A TAD can expand a few hundred kilobases to several megabases region of high local contact frequency and separated from other TADs by sharp boundaries. At the level of genes, transcription is regulated by *cis*-regulatory elements such as enhancers and promoters. Remarkably, enhancers can communicate their defined activities across large genomic distances by physically contacting distal promoters via chromatin

folding. To achieve this, regulatory landscapes are highly organized in 3D nuclear space at a number of scales, each of which differently influences enhancer-promoter communication (Fig 1.11).



**Figure 1.11 | The Hierarchical 3D Organization of the Genome Schematic (left) and Hi-C (right) views of genome organization.** Upper panel: at higher-order scales, chromatin with a transcriptionally active or repressive signature separates into A (red) and B (blue) compartments, respectively. B compartments frequently overlap with nucleolar-associated domains (NADs) and LADs (L) but are distant from speckles (D). Compartments are coincident with non-LADs (N) and are speckle-proximal (P). Lower panel: at smaller scales, enhancers transmit regulatory information to genes by physical proximity within, but not between, TADs. Separated by insulating boundaries, TADs preferentially internally self-associate to create discrete functional and structural blocks. Figure adapted from (Robson et al., 2019).

To fully understand genome function, studying the linear genome map as well as the spatial map chromosome organization is extremely critical. There are increasing evidences that looping of chromosomes is important for transcriptional regulation and gene activation mechanisms by distant regulatory elements (Lomvardas et al., 2006; Tolhuis et al., 2002). It has been demonstrated that transcriptionally active genes contact enhancer-like elements, whereas transcriptionally inactive genes interact with elements marked by repressive features that may act as long-range silencers (Mifsud et al., 2015).

In order to better understand the physical organization of chromosomes in the native cellular state, the chromosome conformation capture (3C) and its derivative techniques have been

developed as valuable tools for uncovering functional elements in the whole genome. The most advantage of 3C is converting the physical chromosomal interactions into specific DNA ligation products bearing information of interacted genomic sequences that can be detected by PCR. Only over the past decade, a series of related techniques have been developed from 3C with an increase in throughput and resolution, the later the fancier name than the last. Variations of the 3C-based techniques include 4C, 5C, Hi-C, Capture-C, and ChIA-PET which are capturing the interactions in different scales and address different biological questions. (See Fig. 1.12 for the methodology summary of 3C based techniques).

Data from 3C-derived technologies have allowed researchers to answer the questions about genome organization that were previously beyond reach. A rich landscape of interactions between specific genomic loci is readily detectable. The first demonstration of long range interaction at  $\beta$  globin gene clusters using 3C technique reveals that promoter of active gene interacts with an upstream LCR (Tolhuis et al., 2002). Subsequently, a number of studies have shown similar interactions between *cis*-regulatory elements (Handoko et al., 2011; Li et al., 2012; Sanyal et al., 2012). By using different 3C-based technology approaches, these studies consistently demonstrated that reproducible interactions are common in mammalian genomes and interacting loci are often enriched between *cis*-regulatory elements. It is noted that the vast majority of these interactions are within a TAD, consistent with the role of TAD boundary in constraining 3D interactions (Dixon et al., 2012; Nora et al., 2012).

Altogether, 3C and 3C variants generate comprehensive and genome-wide interaction maps for studying higher-order chromatin structure. As the evolution of chromosome conformation capture is continuing, the C methods become more refined and their use expands, they will be a valuable tool in the understanding of how chromatin structure, protein interactions and DNA sequence all work together to control gene expression for years to come.





## Chapter 2

### Silencer elements

Eukaryotic transcription is controlled by the combined activities of gene promoters and distal cis-regulatory elements, such as enhancers and silencers. Whereas the biological functioning of promoters and enhancers is relatively well understood, the understanding of the function and importance of silencer elements have remained elusive.

Silencers were initially defined as sequence elements which are capable of repressing promoter activity in an orientation and position independent fashion, in the context of a native or a heterologous promoter (Brand et al., 1985). More recently it has become clear that many different types of silencers exist that are capable of affecting many aspects of gene regulation, such as activity of positive acting TFs, chromatin structure, intron splicing, 3' upstream untranslated signal recognition, as well as GTF assembly, to ultimately down-regulate gene expression (Ogbourne and Antalis, 1998).

Regulatory elements capable of silencing gene expression have also been described in several species (Maston et al., 2006). However, the number of silencers identified in the mammalian genomes has been very limited, resulting in a lack of information regarding silencer-specific chromatin profiles and mechanisms of action. They include silencer elements derived from humans, mouse, rats, chickens, and yeast. A list of representative silencers identified in mouse and human is provided in (Table 2.1).

A survey of the published literature on transcriptional repression sheds the light on the different subclasses of silencer elements. Initially silencers were sub classified as classical, position-independent element that directs an active repression mechanism (generally by interfering with GTF assembly) should be referred to as a 'silencer element', and that a non-classical, position-dependent element directs a passive repression mechanism (generally by interfering with upstream elements) be referred to as an 'NRE', and that the associated TF be known as a 'repressor' (Ogbourne and Antalis, 1998).

#### 1. Historical Overview

In 1985, the yeast mating type loci revealed that distal silencer elements could control gene expression from afar (Brand et al., 1985). The role of distal silencer elements in mammals was demonstrated shortly thereafter through a silencer element located several kilobases upstream of the rat insulin gene (Laimins et al., 1986). A decade later, a series of key experiments identified a silencer in the intron of the mouse and human CD4 genes, and revealed the important role that silencers can play in lineage specificity and cell fate determination, as this silencer represses CD4 expression in CD8<sup>+</sup> T cells (Donda et al., 1996; Taniuchi and Littman, 2004). Later, several studies identified genomic sequences with silencer properties that are the opposite of enhancers across many species

(Ogbourne and Antalis, 1998). Although many silencers were identified in the late 80's and 90's (Table 2.1), the silencer field has been somehow hampered in the last years by the lack of high-throughput functional assays, as reflected by a PubMed survey between the literature related to silencers and other cis-regulatory elements (Fig 2.4).

**Table 2.1: Examples of silencer elements validated by reporter assays in mouse and human.** Adapted from (Ogbourne and Antalis, 1998; Qi et al., 2015).

Target gene	Size(bp)	References
Human plasminogen activator inhibitor type-2	302	(Antalis et al., 1996)
Human sperm histone H2b-1	28	(Barberis et al., 1987)
Human apolipoprotein A-II	100	(Bossu et al., 1996)
Human papilloma late mRNA	79?	(Dietrich-Goetz et al., 1997)
Human YD promoter	17	(Dirks et al., 1996)
T-cell receptor Vb2.2	39	(Dombret et al., 1996)
Human CD4	15	(Donda et al., 1996)
Human neuronal $\alpha$ 1-chimaerin	30	(Dong and Lim, 1996)
Human insulin	21	(Goodman et al., 1996)
Human collagen type 4	21	(Haniel et al., 1995)
Human <i>c-Fes</i>	13	(He et al., 1996)
Human thyrotropin $\beta$	352	(Kim et al., 1996)
Human platelet-derived growth factor A-chain	30	(Liu et al., 1996)
Human Pi Class glutathione S-transferase	7	(Moffat et al., 1996)
Human T-cell activation gene 3	19	(Oh et al., 1997)
Human hypoxanthine phosphorylase transferase	59	(Rincon-Limas et al., 1995)
Human synapsin I	36	(Schoch et al., 1996)
Human <i>c-myc</i>	9	(Takimoto et al., 1989)
Human platelet-derived-growth-factor A-chain	25	(Wang et al., 1994)
Human interleukin-2	9	(Williams et al., 1991)
Human interleukin-8	7	(Wu et al., 1997)
Human interferon- $\gamma$	25	(Ye et al., 1996)
CCND1	953	(French et al., 2013)
MECP2F3	985	(Liu and Francke, 2006)
TSHB	353	(Kim et al., 1996)
PDGFA	31	(Liu et al., 1996)
PPARD	500	(Yadav et al., 2018)
Mouse bone morphogenetic protein 4	1230	(Feng et al., 1995)
Mouse mammary-tumour-virus long terminal repeat	13	(Giffin et al., 1994)
Mouse thyroid HR $\beta$ 1	14	(Nagasawa et al., 1997)
Mouse thyroid HR $\beta$ 1	8	(Nagasawa et al., 1997)
Mouse major inducible <i>Hsp70</i>	1044	(Shimokawa and Fujimoto, 1996)

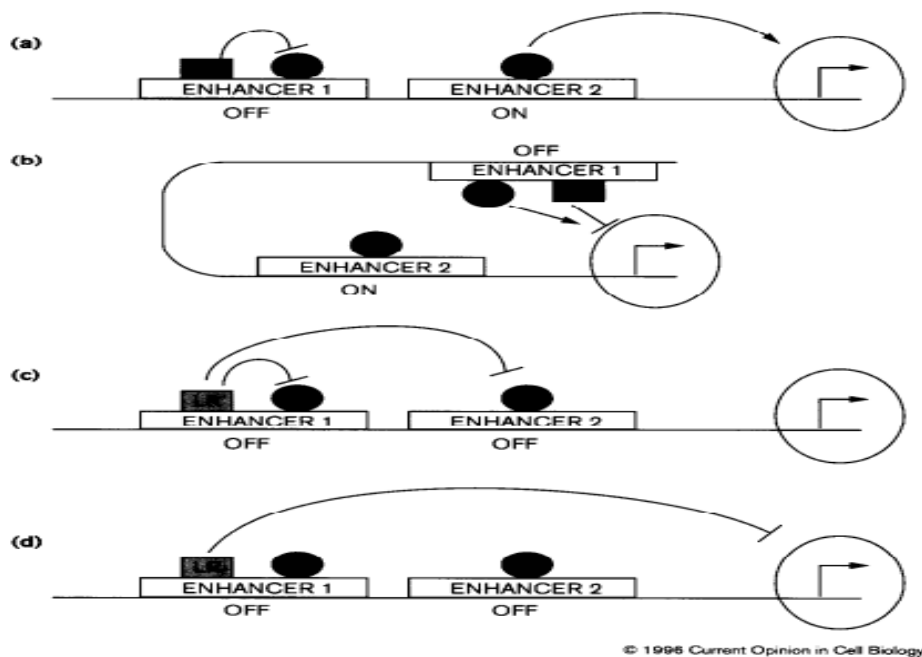
## 2. Models of silencer function

Typically, silencers function independently of orientation and distance from the promoter, although some position-dependent silencers have been encountered. They can be situated as part of a proximal promoter, as part of a distal enhancer, or as an independent distal regulatory module; in this regard, silencers can be located far from their target gene, in its intron, or in its 3'-untranslated region. Finally, silencers may

cooperate in binding to DNA (Harris et al., 2005), and they can act synergistically (Sertil et al., 2003).

To study silencer elements a number of models have been proposed for repressor function. In some cases, repressors appear to function by blocking the binding of a nearby activator (Harris et al., 2005). Alternatively, a repressor may prevent activators and/or GTFs from accessing a promoter by establishing a repressive chromatin structure through the recruitment of histone-modifying activities or chromatin-stabilizing factors (Weintraub et al., 2017).

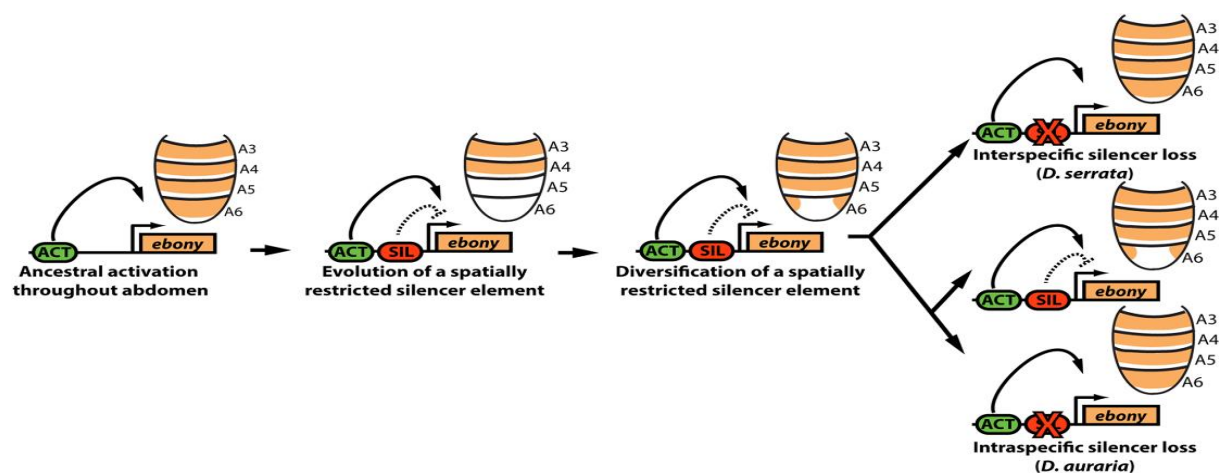
Studies on transcriptional repression in the *Drosophila* embryo suggest that there may be two basic forms of repression, namely long-range and short-range repression (summarized in Fig. 2.1). Short-range repressors work over distances of less than 100-150 base pairs (bp) to inhibit upstream activators in a local fashion. Long-range repressors function over distances of at least 500 bp to silence the transcription complex or to inhibit upstream activator(s) that are bound to promoter-proximal regions. In principle, long-range repressors can function in a dominant fashion to block multiple enhancers in a complex modular promoter reviewed in (Gray and Levine, 1996).



**Figure 2.1| Quenching versus direct repression models for short-range and long-range repression.** (a) Short-range quenching. A short-range repressor (SRI) bound to enhancer 1 interferes with the function of a nearby activator (A). Activators bound to enhancer 2 are out of the range of the SR and can therefore contact the transcription complex (represented by an arrow within an oval). (b) Short-range direct repression. When enhancer 1 loops into the transcription complex, the repressor interferes with the neighboring activator by competing with common components of the transcription complex. As the interaction of the repressor with the transcription complex is weak and transient, activators bound to enhancer 2 are

unaffected. (c) Long-range quenching. This differs from short-range quenching in that the long-range repressor (LR) can interfere with the function of activators that are bound kilobase distances away from, in addition to nearby, the LR. This interference could involve the recruitment of one or more corepressors, or perhaps changes in chromatin structure. (d) Long-range direct repression. Here, the LR functions as a dominant repressor by stably interacting with one or more components of the transcription complex (Gray and Levine, 1996).

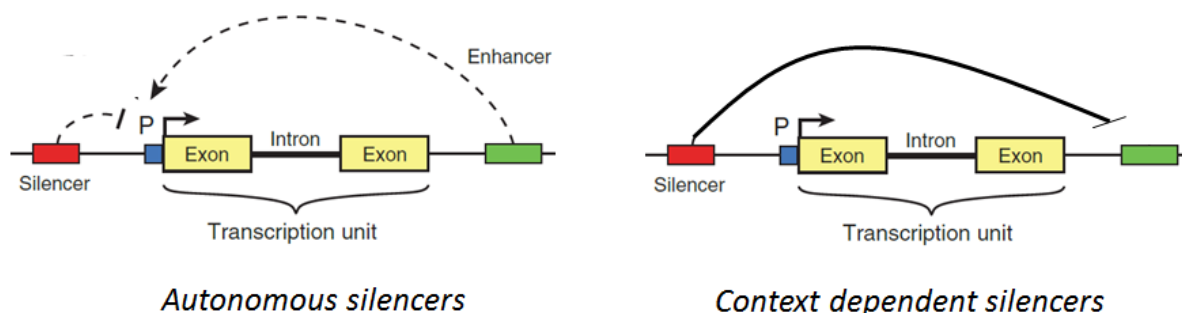
Johnson et al., shown how a transcriptional silencer that participates in the sexually dimorphic patterning of gene expression has experienced repeated inactivation events that increased its target gene's expression (Fig. 2.2). Despite highly varied male pigmentation phenotypes in *D. auraria*, two genes that often correlate with pigment pattern (*yellow* and *tan*) were expressed similarly among light and dark strains. In contrast, *ebony* expression uniquely correlated with pigmentation, it was showed that the correlation between expression and pigmentation phenotypes were due to *cis*-regulatory mutations within *ebony*. Mapping the *ebony* regulatory region of *D. auraria* established the conservation of a silencer element with male-specific activity that carves out sexually dimorphic expression from a ubiquitously activating enhancer. Within this conserved silencer, functional changes were localized responsible for allelic differences in expression. Moreover, it is showed that the secondary loss of male-specific pigmentation in *D. serrata* occurred through the parallel inactivation of this upstream silencer element of *ebony* (Fig. 2.2). These findings highlight the under-appreciated role of silencers in the evolution of gene expression and morphology (Johnson et al., 2015).



**Figure 2.2| Model for the parallel loss of the *ebony* upstream silencer element.**

Within the ancestral *ebony* gene, a silencer element evolved that adopted a repressive role in in the male posterior abdomen. Within the *montium* subgroup, this silencer's activity was modified through changes in *cis* that altered the silencer's spatial domain of activity. Within *D. serrata*, this silencer was inactivated, leading to increased *ebony* expression. In contrast, this silencer was maintained in the species *D. auraria*, in which intraspecific variation exists for an allele that inactivated this element (Johnson et al., 2015).

Limited information suggests there are two general categories of silencers (Fig 2.3): elements that only function in a specific context to repress promoter activity (termed negative regulatory elements (Ogbourne and Antalis, 1998); and elements that function in an autonomous, context-independent manner. Negative regulatory elements function by recruiting proteins that disrupt or inactivate the formation of functional Pol II transcription complexes at otherwise accessible promoters. This is accomplished by recruiting repressor proteins, blocking the nearby binding of activator proteins, or competing directly with activator proteins for the same binding site (Maston et al., 2006). Autonomous silencers function by establishing a repressive chromatin state that can be stably inherited (Dean, 2011). This is typically accomplished by recruiting proteins capable of modifying DNA (e.g. methyltransferases) or histone tails (e.g. histone deacetylases) in a manner that supports the formation of heterochromatin, or proteins that help stabilize and propagate heterochromatin (e.g. polycomb group proteins, heterochromatin protein 1). This in turn prevents activators and transcription factors from accessing gene promoters. Other types of silencers might interfere with additional cis-regulatory elements. For example, the CD4 silencer has been shown to interfere with the enhancer-promoter interactions at this locus (Taniuchi et al., 2002b).

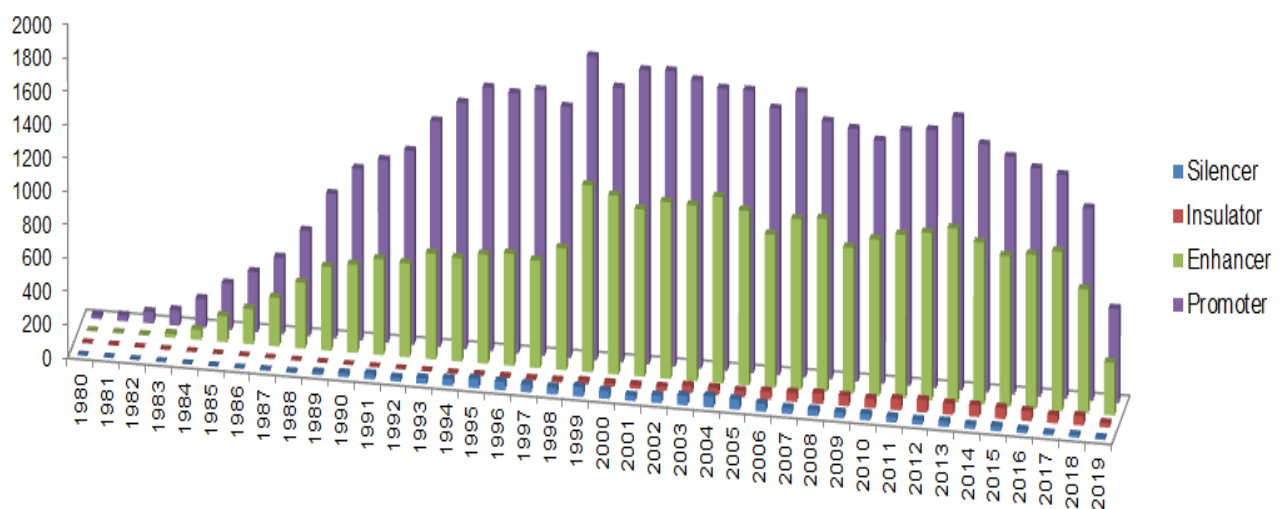


**Figure 2.3 | Two possible mechanisms of silencer action**

The availability of a very limited number of autonomous mammalian silencer elements described in the scientific literature has made it difficult to identify common mechanisms of action, as well as common chromatin profiles. Advances toward identifying and characterizing autonomous silencer elements have also been hampered by the lack of independently-validated functional assays. A question even remains as to whether silencer activity determined by transient transfection of reporter constructs can accurately predict whether a particular sequence is capable of providing silencer activity in the setting of intact chromatin.

Later, using drug-resistant colony assays and the GFP expression assay the T39 human silencer element has been identified. The T39 silencer exhibits a distinct chromatin profile, including high occupancy by CTCF and DHS formation in multiple tissues (Liu et al., 2015; Qi et al., 2015). This silencer is located in the first intron of the ARHGAP6

gene. It has been suggested that silencers located in introns could suppress transcription by blocking transcriptional elongation, by preventing recognition of intronic splice sites or by abrogating basal transcriptional apparatus assembly (Ogbourne and Antalis, 1998). Publicly available CTCF interactome data indicate that the T39 element physically interacts with other CTCF sites located at the terminus of the ARHGAP6 gene and within the neighboring gene MSL3, but not with the promoters of these genes. It is of interest to speculate that the T39 silencer may modulate the expression of MSL3, a gene contributing to the X-inactivation complex of *Drosophila* and humans (Kadlec et al., 2011).



**Figure 2.4 | Number of PubMed citations per year associated with the different cis-regulatory elements (July 2019).**

### 3. Repressors

The role of repressors and repressor complexes are obviously paramount in this area. There are several groups of repressor complexes and how they regulate gene expression, as well as several examples of TFs which can behave as activators or repressors. One of them i.e. REST is detailed below.

#### 3.1 REST

RE-1 silencing transcription factor (REST), which is also known as neuron-restrictive silencer factor (NRSF), is a transcription repressor that binds to the 21-bp RE1 sites in the regulatory regions of its target genes (Ooi and Wood, 2007). REST is known to have a central role in regulating neurogenesis, neural differentiation, and preservation of the unique neural phenotype (Song et al., 2015). In addition to neuronal development, NRSF/REST may have other roles in cardiac development (Kuwahara et al., 2003), pancreatic islet development

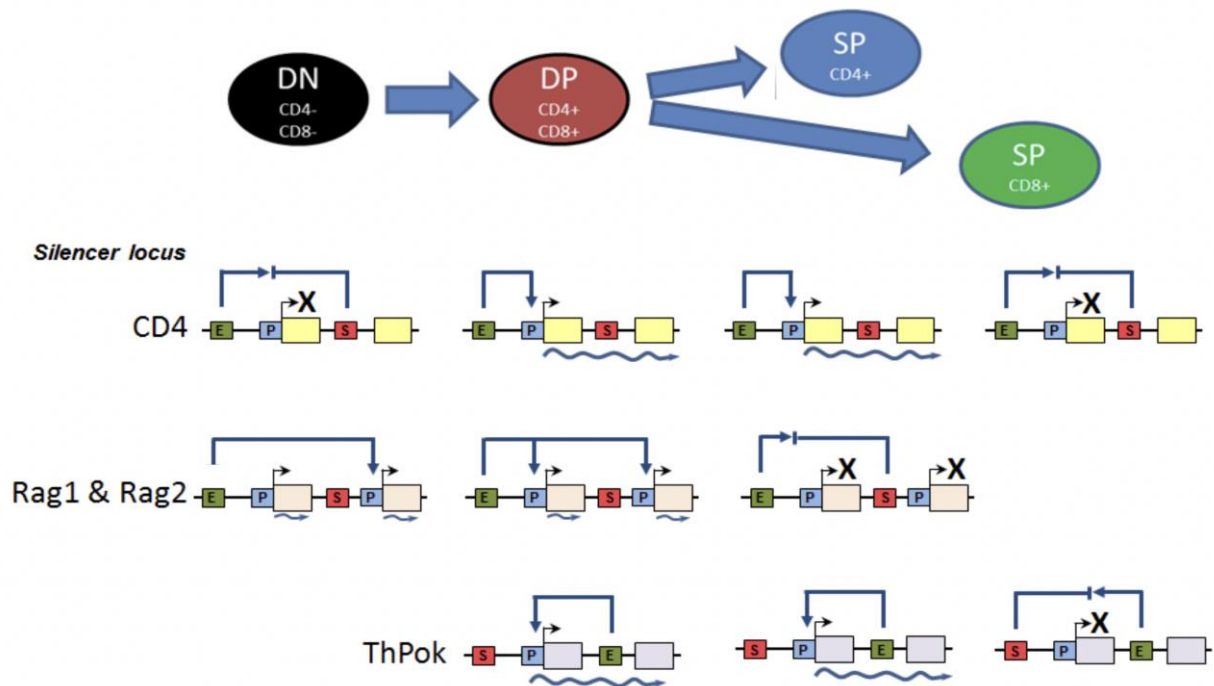
(Abderrahmani et al., 2001), and perhaps B- or T-cell lineages (Scholl et al., 1996). Downregulation of REST during neural differentiation is necessary for the correct development of certain classes of neurons (Terry and Beltran, 2014). REST is known to repress its target genes by interacting with subunits of several transcription regulatory complexes, including CoREST and mSin3 corepressor complexes, the SWItch/Sucrose Non-Fermentable (SWI/SNF) complex, and polycomb repressive complex 1 (PRC1) and PRC2 (Dietrich et al., 2012; Ooi and Wood, 2007). Another isoform, REST4, is thought to act in a dominant negative fashion (Hersh and Shimojo, 2003) These REST-interacting proteins were independently identified using yeast two-hybrid screening or co-immunoprecipitation under different experimental conditions.

In pluripotent stem cells and neural progenitors, REST actively represses a large array of coding and noncoding neuron-specific genes important to synaptic plasticity and structural remodeling including miR-132 (Ballas and Mandel, 2005; Ooi and Wood, 2007). A subset of microRNAs including miR-9, miR-124a, and miR-132 contains RE-1 sites within their promoter regions in mouse and humans and are functionally validated targets of REST in mammalian cell lines (Conaco et al., 2006). Hwang *et al* demonstrate that ischemic insults trigger activation of REST and REST-dependent silencing of miR-132 in selectively-vulnerable CA1 neurons and that REST-dependent repression of miR-132 is critical to ischemia-induced neuronal death in a clinically-relevant model of ischemic stroke *in vivo* (Hwang et al., 2014).

In another repression role of REST was shown by Zhang *et al*, 2012 repress *CART* expression. They found that NRSF utilizes a dual NRSE pattern to efficiently repress *CART* expression. They proved that both the promoter NRSE (pNRSE) and the intron NRSE (iNRSE) are *bona fide* regulatory elements of *CART* that play equivalent roles in *CART* regulation. NRSF recruits co-repressor complexes to the NRSE sites to ensure repression efficiency. The complexes include HDAC1, mSin3, and CoREST, which is the first time shown to bind to the *CART* gene. They noted that similar but not identical co-repressors were recruited to the pNRSE (pNRSE) and iNRSE (iNRSE), *i.e.* (a) NRSF and CoREST displayed stronger affinity to the iNRSE than to the pNRSE (about 8-folds); (b) mSin3A was specifically recruited to pNRSE but not to iNRSE (Zhang et al., 2012).

#### **4. Silencers in the hematopoietic system and specifically during T cell differentiation**

It is noticeable that among the few silencers that have been described in the literature, many are associated with the regulation of T cell specific genes (Table 2.3). Several of these silencers have been shown to play an important role in cell lineage restriction and their study has allow a better understanding of the mechanisms underlying silencer function during development and cell differentiation (Fig 2.5).



**Figure 2.5 | Best characterized silencers involved in the regulation of key T cell differentiation genes**

#### 4.1 CD4 silencer

During the development of T lymphocytes, differential regulation of expression of the CD4 and CD8 glycoproteins is coupled to the choice of one of two pathways of differentiation. Thymocytes that express both of these coreceptors commit to either the helper lineage, shutting off CD8, or the cytotoxic lineage, shutting off CD4. In earlier work, a T cell specific enhancer was identified and characterized located at a DNAase I hypersensitive site 13 kb upstream from the transcriptional initiation site of the mouse CD4 (mCD4) gene (Sawada et al., 1994). It was shown that this enhancer and the homologous human element are required for the correct expression of the CD4 gene during T cell differentiation in transgenic mice (Killeen and Littman, 1993). Several lines of evidence suggested that the CD4 gene has additional elements required for its negative regulation in CD8 SP lineage T cells. In transient transfection assays, combinations of the CD4 enhancer and promoter were functional in both CD4 SP and CD8 SP T cell lines (Sawada et al., 1994). In transgenic mice, the CD4 enhancer/promoter linked to a human CD4(hCD4) cDNA/genomic chimeric gene, lacking introns 1-4, directed expression in both CD4 and CD8 SP cells, in contrast, a similar construct containing the hCD4 gene, with intact introns 1, 2, and 4, was correctly regulated, with expression detected in mature CD4 SP but not in CD8 SP T cells (Killeen and Littman, 1993).

To localize the negative regulatory element in the CD4 gene, a transgenic system using the hCD2 cDNA as a reporter was developed (Sawada et al., 1994). Analysis of hCD2 expression in T cell subsets in the transgenic animals has permitted to identify a short intronic sequence that coincides with a DNAase I hypersensitive site and that promotes repression of the hCD2

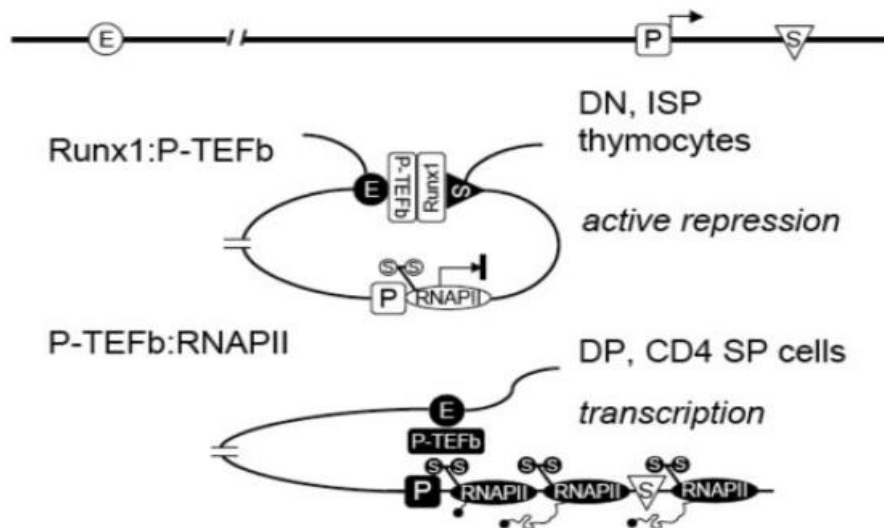


transgene in CD8 SP cells as well as in immature thymocytes that express neither CD4 nor CD8 (double negative or DN cells). This negative regulatory region functions in an orientation and position independent manner and combination with heterologous cis-acting regulatory elements, indicating that it is a developmentally regulated transcriptional silencer (Sawada et al., 1994).

#### **4.2 Runx1 and Runx3 at the CD4 silencer**

Binding of Runx1 and Runx3 to the Cd4 silencer element located in the first intron of the Cd4 gene serves as an instructive example of the ability of Runx proteins to mediate repression by two different mechanisms at different stages of thymocyte development (Taniuchi et al., 2004). The CD4 silencer confers lineage-specific expression of Cd4 in thymocytes by repressing the CD4 promoter and enhancers twice during T-cell development in the thymus. In DN cells this repression is transient and reversible (as thymocytes have to upregulate CD4 in the subsequent stage of development to DP cells) and depends on the continued presence of the silencer; on the other hand, during DP to SP transition, the silencer mediates heritably stable (irreversible) silencing in CD8SP cells (as these cells must not express CD4 at subsequent stages of differentiation in the periphery), and silencing persists even if the silencer element is removed in mature CD8 SP cells. Runx1 and Runx3 bind to the CD4 silencer in DN and CD8SP cells, respectively, and loss of Runx-binding sites in the CD4 silencer causes Cd4 derepression in both DN and CD8SP cells (Taniuchi et al., 2002b).

A mechanism for how Runx1 transiently represses CD4 was proposed (Jiang and Peterlin, 2008) (Fig 2.6). In DN cells, Runx1 bound at the CD4 silencer interacts with the positive elongation factor, P-TEFb at the CD4 enhancer. This interaction is thought to prevent polymerase II preloaded at the CD4 promoter from engaging with P-TEFb and initiating productive transcription (Jiang and Peterlin, 2008). On the other hand, in mature CD8+ T cells, the silent CD4 promoter is associated with both of the repressive histone modifications, H3K9me3 and H3K27me3 (Yu et al., 2008). The presence of these modifications at the CD4 locus is likely to depend on Runx3, given its crucial role in CD4 silencing in CD8SP cells. Thus, changes of chromatin structure at the level of histone modifications are likely to be a part of the mechanism by which Runx3 mediates long-term gene repression. However, whether the epigenetic modifications of the CD4 locus are mediated by direct recruitment of Polycomb proteins and/or HP-1 by Runx3 is currently unknown.



**Figure 2.6 | A model for active repression by Runx1.** In DN and ISP thymocytes, the interaction between Runx1 and P-TEFb brings the enhancer (E) into the close proximity of the silencer (S) and prevents P-TEFb from activating RNAPII, which is arrested at the promoter (P), thus actively repressing transcription elongation. In DP and CD4 SP thymocytes, the lack of Runx1 binding to the CD4 silencer frees P-TEFb to interact with RNAPII and to activate transcription elongation. Only RNAPII on the CD4 promoter is depicted in the model. Figure adapted from (Jiang and Peterlin, 2008).

### 4.3 ThPOK silencer promotes CD4 commitment

Identification by Kappes *et al.*, in the late 1990s, of a spontaneous mutant mouse line (helper deficient, or HD) that lacked CD4<sup>+</sup> helper cells due to a single recessive mutation was the first step in identifying ThPOK as the master regulator of CD4 lineage choice (Dave et al., 1997). Subsequent backcrossing of HD mice to an MHC class I-deficient background revealed that lack of CD4 T cells in HD mice was not due to a block in the development of MHC class II-restricted thymocytes but rather to their redirection to the CD8 lineage. This finding led to the conclusion that lineage choice and positive selection were two mechanistically distinct processes. Finally, in 2005 the mutation was mapped to a point mutation in the gene encoding the transcription factor ThPOK (T helper-inducing POZ Krueppel factor), which resulted in a single amino acid substitution (R to G) in its DNA binding domain (He et al., 2005).

These and later studies from another group established that ThPOK functions as a “master regulator” of lineage commitment, whose presence or absence is necessary and sufficient to drive the development of immature thymocytes to the CD4 or CD8 lineages, respectively (Kappes et al., 2005). The structure and expression pattern of the ThPOK gene appears to have been conserved since the divergence of bony fish from other vertebrate lineages.

#### **4.4 Key role for ThPOK silencer**

Given that ThPOK functions as the master regulator of CD4 lineage commitment and is regulated primarily at the level of transcription, it became critical to elucidate the molecular mechanism of its transcriptional regulation. A 20kb region surrounding the ThPOK locus and bounded by CTCF insulators is sufficient to reconstitute normal lineage commitment when introduced as a transgene into HD mice. Further investigation of this region identified six DNase hypersensitivity (DHS) sites, including two promoters and four other putative cis regulatory elements (He et al., 2008a), which collectively seem necessary and sufficient for the stage-specific regulation of ThPOK expression during thymopoiesis. Of the six DHS sites, a 400 bp silencer element at DHS site A (Sil ThPOK) gained prominence, as its deletion from a ThPOK GFP reporter transgene abolished lineage-specific regulation, leading to promiscuous expression in both CD4 and CD8 T cells. Furthermore, knockout mice lacking the Sil ThPOK exhibit complete derepression of ThPOK in class I-restricted thymocytes, resulting in a severe reduction in the number of CD8 T cells due to redirection of class I-restricted thymocytes to the CD4 lineage. Interestingly, conditional deletion of the Sil ThPOK in mature CD8 T cells does not derepress ThPOK expression, indicating that CD8 lineage commitment entails permanent epigenetic silencing of the locus (Sun et al., 2005). Collectively these results indicate that Sil ThPOK function is controlled in part by stage- and lineage-specific chromatin modifications and by sequential recruitment of DNA binding proteins. The precise transcription factors that control Sil ThPOK function remain incompletely characterized. One important factor is Runx3, which has previously been shown to be essential for CD8 differentiation. The Sil ThPOK region has two putative Runx binding sites, and deletion of these motifs impairs silencer activity (He et al., 2008a). Interestingly, chromatin immunoprecipitation (ChIP) assays revealed that Runx complexes are bound to Sil ThPOK at all stages of thymic T cell development, indicating that Runx binding is not sufficient to account for stage specific regulation of the Sil ThPOK. Hence, the molecular basis by which differential TCR signals specifically modulate the activity of the Sil ThPOK has yet to be established.

#### **4.5 Opposing effects of Runx and ThPOK on CD4/CD8 on lineage specification**

Several lines of evidence support that ThPOK functions as a master transcriptional regulator of CD4 lineage specification and as an antagonist of Runx3. ThPOK has been implicated in the regulation of multiple genes involved in CD4 lineage specification, including CD4 and ThPOK, at least in some cases via direct binding to the target locus. Direct binding of ThPOK to CD4 and ThPOK silencers antagonizes their function and initiates positive-feedback loops that support CD4 surface expression and increased transcription of ThPOK itself. Induced expression of ThPOK in DN and CD4<sup>+</sup>CD8<sup>lo</sup> thymocytes mediates derepression of CD4 expression, and in peripheral CD8 T cells causes activation of several CD4 lineage-restricted genes including GATA-3 (Wildt et al., 2007). Further, ectopic ThPOK expression leads to severe functional defects by peripheral CD8 T cells, due to down modulation of many CD8 specific genes including CD8, Perforin, Granzyme B, and Eomes (Kundu et al., 2005).

ThPOK has also been shown to bind to its own silencer and antagonize its function, indicative of a positive feedback regulatory loop. Diminished ThPOK expression in mature CD4 T cells, on the other hand, results in downregulation of CD4 lineage-specific genes including CD4, upregulation of cytotoxic lineage genes such as Runx3 and Eomes, acquisition of high IFN- $\gamma$  production capacity and transdifferentiation into CD8 T cells upon transfer into T cell deficient mice. In vitro, cotransfection assay showed that ThPOK antagonizes Runx-mediated repression of a reporter gene controlled by the CD4 silencer (Wildt et al., 2007). Furthermore, ThPOK represses Runx3 distal promoter activity during the differentiation of class II-restricted cells; although it is not clear whether the effect is direct (Cruz-Guilloty et al., 2009). Interestingly, the antagonistic effect of ThPOK on Runx-mediated silencing can be effectively ablated by the histone deacetylase inhibitor TsA, suggesting that antagonism depends on transcriptional repression by ThPOK of an unknown factor that acts in conjunction with Runx, a repressor-of repressor model (Rui et al., 2012). In summary, current data indicate that ThPOK and Runx3 antagonize each other's functions by supporting CD4 and CD8 lineage-specific gene expression programs, respectively. Regulatory loops initiated by ThPOK explain in part how the initial lineage specification signals that induce limited expression of ThPOK at the CD4+CD8lo stage are amplified to drive full CD4 commitment.

#### **4.6 L2a silencer CD8 gene**

The mechanism underlying CD4 and CD8 T cell lineage commitment has been the subject of intense investigation over many years. Identification of cis-acting regulatory elements and factors controlling CD4 and CD8 gene transcription has been critical to the understanding of CD4/CD8 lineage differentiation. DHS measurements and transgenic reporter assays of the ~80 kb region spanning the murine genomic CD8 $\alpha$  and CD8 $\beta$  locus identified four clusters (CI–CIV) of DHS. Subsequently, four enhancers (E8I–E8IV), which overlap DHS between CD8 $\alpha$  and CD8 $\beta$ , were shown to be involved in CD8 gene expression in CD8 $\alpha\beta$ + T cells (Kioussis and Ellmeier, 2002). Each enhancer is CD8 lineage-specific and is active at a defined T cell developmental stage.

Targeted disruption of E8I (CIII-1,2) in vivo had no effect on CD8 $\alpha$  and CD8 $\beta$  expression in thymus derived T cells, but CD8 expression in IELs was eliminated, suggesting that other cis-acting elements could compensate for the loss of E8I in thymus-derived T cells (Ellmeier et al., 2002). Combined deletion of E8I and E8II resulted in variegated expression of CD8SP and DP thymocytes as well as in reduced CD8 expression on mature CD8+ T cells (Ellmeier et al., 2002). Further studies demonstrated that E8I and E8II deletion led to impaired chromatin remodeling during T cell development (Bilic et al., 2006). Combined targeted deletion of E8II and E8III did not significantly alter expression levels of CD8 $\alpha$  and CD8 $\beta$  in thymocytes or T cells. DH cluster II was inactive in transgenic reporter analysis, but as observed in E8I–E8II double deletion mice, cluster II deletion led to impaired CD8SP frequencies with a small % of variegated expression observed in peripheral CD8+ T cells (Garefalaki et al., 2002).

The L2a element is located ~4.5 kb upstream of the CD8a gene where it resides within the first DHS of cluster II (CII-1), a short distance upstream of two GATA-3 binding sites (Landry et al., 1993). L2a was defined by in vitro nuclear matrix binding assays as a nuclear matrix attachment region (MAR) (Banan et al., 1997). MARs frequently reside in close proximity to promoters and enhancers and have been shown to be important transcriptional regulators during chromatin remodeling (Hart et al., 1997). In stable transfection studies, inclusion within constructs of a 900 bp fragment that spanned L2a significantly reduced the frequency of CD8<sup>+</sup> T cells, suggesting that L2a may negatively regulate CD8 transcription (Lee et al., 2004). This interpretation is consistent with the observations that deletion of the DHS cluster II/CII-1,2 region (which spans L2a) results in variegated CD8 expression (Garefalaki et al., 2002), and that inclusion of a fragment spanning DHS cluster II enables E81 to activate reporter gene expression in DP thymocytes (Erman et al., 2002).

#### 4.7 Silencer and anti-silencers regulate *Rag1* and *Rag2* genes

There are two developmental windows of *Rag1* and *Rag2* (hereafter, *Rag*) gene expression during T and B lymphocyte development (Kuo and Schlissel, 2009). In developing thymocytes, the *Rag* genes are first expressed at the CD4<sup>-</sup>CD8<sup>-</sup> double-negative (DN) stage to promote recombination of the *Tcrb*, *Tcrg*, and *Tcrd* genes. Productive *Tcrb* recombination causes *Rag* gene downregulation, cellular proliferation, and differentiation to the CD4<sup>+</sup>CD8<sup>+</sup> double-positive (DP) stage. *Rag* genes are then reexpressed in DP thymocytes to promote the recombination of *Tcra* genes. After productive *Tcra* gene assembly and positive selection of TCR-expressing DP thymocytes, *Rag* genes are silenced during differentiation to the CD4<sup>+</sup>CD8<sup>-</sup> or CD4<sup>-</sup>CD8<sup>+</sup> single-positive (SP) stage.

The intergenic region between *Rag1* and *Rag2* contains a silencer that strongly represses expression of *Rag2* in DP T cells and attenuates expression in DN T cells (Yannoutsos et al., 2004). The interval between 71 kb and 86 kb 5' of *Rag2* contains a *cis* element named anti-silencer (ASE) required to suppress the activity of the intergenic silencer in DP thymocytes (Yannoutsos et al., 2004). The intergenic silencer suppresses both *Rag1* and *Rag2* expression in the DP stage of thymocyte development and the ASE has no apparent effect on gene expression in the absence of this silencer (Yannoutsos et al., 2004). *Rag1* and *Rag2* gene expression in CD4<sup>+</sup>CD8<sup>+</sup> double-positive (DP) thymocytes depends on the activity of a distant anti-silencer element (ASE) that counteracts the activity of an intergenic silencer (Yannoutsos et al., 2004) (Hao et al., 2015).

Similar to the *Cd4* silencer, the sequence of *Rag* silencer has little conservation between mouse and human, however the location of Runx binding sites are conserved in both species (SCHLUTER and MARCHALONIS, 2003). Thus, *Rag* silencing resembles *Cd4* silencing in the way that Runx is essential for negative regulation. However, the *Cd4* silencer has a dominant effect that prevents gene expression, whereas the positively acting ASE is dominant over the silencer in the *Rag* locus (Yannoutsos et al., 2004).

*Rag* gene expression in DP thymocytes depends on the assembly of a multi-component chromatin complex, or hub, containing the ASE and both *Rag* gene promoters. The chromatin organizer SATB1 plays an important role in the assembly of this complex and in *Rag* gene transcription, functioning at least in part by stimulating RNA pol II loading to the *Rag* promoters (Hao et al., 2015).

**Table 2.2: Validated silencers in T cells**

Target gene	Species	Methods	Associated Repressor	Mechanism	References
<b>Il2</b>	Human	Luciferase and CAT reporter assay	Zinc finger containing protein	The NRE-A element inhibit Il2 activation upon T cell stimulation	(Williams et al., 1991)
<b>Vb2.2 segment from the TCRb locus</b>	Human	CAT assay, Gel shift assay, Electrophoretic mobility shift assay, Methylation interference analysis		The downstream part of the Vb2.2 gene sequence exerts a dominant silencing activity in the Jurkat and MOLT4 cell lines, but not in the immature HSB2 cell line.	(Dombret et al., 1996)
<b>TCA3/CCL1</b>	Mouse	CAT reporter assay, EMSA, DNase footprinting		The NRE element control the expression of TCA3 in activated mast cells	(Oh et al., 1997)
<b>CD4</b>	Mouse	Transgenic mice; Knock-out mice	Runx 1	Differential looping mediate active repression of the CD4 gene during thymocyte development. Deletion of the CD4 silencer resulted in stochastic CD4 expression in immature and CD8+ thymocytes	Adlam and Siu (2003); (Collins et al., 2011; Jiang and Peterlin, 2008; Leung et al., 2001; Sawada et al., 1994; Xing et al., 2018)
<b>Rag1 and Rag2</b>	Mouse	Transgenic mice Dual-Luciferase Reporter Assay,	Runx 1	Runx-dependent intergenic silencer suppress Rag1 and Rag2 expression in developing T cells. An anti-silencer element (ASE) physically interacts with the distant Rag1 and Rag2 promoters in DP thymocytes, overcoming the silencer effect	(Hao et al., 2015; Yannoutsos et al., 2004)
<b>TAL1</b>	Human	Electrophoretic Mobility Shift Assay, UV Cross-linking, in vivo Footprinting LMPCRs, EMSAs	PU.1	Functions of the human tal-1 39-UTR is to inhibit the expression of transcripts encoding the TAL-1 proteins in cells from erythroid and megakaryocytic lineages.	(Courtes et al., 2000; Le Clech et al., 2006)
<b>CD8</b>	Mouse	Transgenic mice, EMSA, DNase I footprinting	SATB1 and CDP	The L2A element is CD8 silencer that interacts with MAR-binding proteins SATB1 and CDP	(Yao et al., 2010)
<b>ThPOK</b>	Mouse	GFP reporter ,Flow Cytometry, DNase I Hypersensitivity Analysis. Knock-out mice	Runx1 Tle/Groucho	A distal regulatory element (DRE) is necessary for suppression of ThPOK expression in class I restricted +thymocytes and sufficient for its induction in class II-restricted thymocytes. Silencer-mediated alterations of chromatin structures in cytotoxic-lineage thymocytes establish a repressive state that is epigenetically inherited in peripheral CD8+ T cells even after removal of the silencer	(Collins et al., 2011; He et al., 2008a; Tanaka et al., 2013; Xing et al., 2018)

# Chapter 3

## High-throughput reporter assays

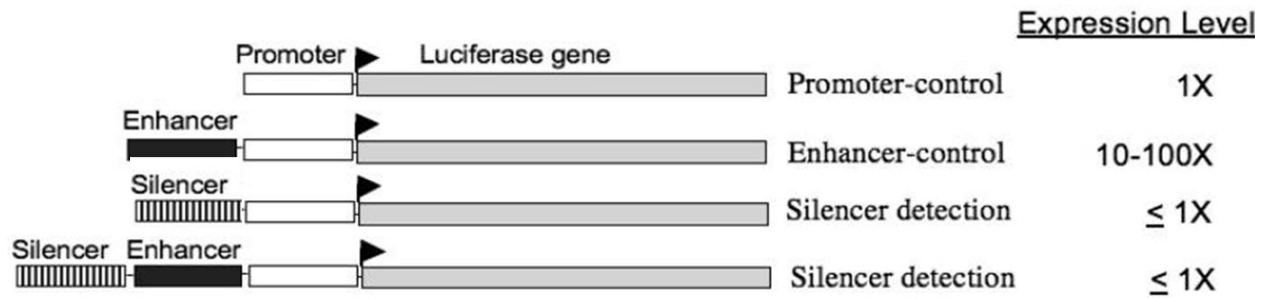
### 1. Overview

With the increasing awareness of the important role of *cis*-regulatory elements 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 some of the *cis*-regulatory elements do not have to be located directly adjacent to the gene it regulates. With recent technological advances, these can be identified across entire genomes by many techniques such as DNase-seq, which detect the DNase I accessible regions; ATAC-seq, detects transposase accessible sites or ChIP-seq for sites enriched in histone modifications or transcription factors. However, it is also crucial to test whether these genomic regions are actually functioning in living cells or tissues. Traditional approaches for testing *cis*-regulatory elements, specially enhancer activity, rely on functional assays of individually transfected reporter plasmids harboring putative regulatory regions for a gene of interest or *in vivo* testing by generating transgenic animals, reviewed in (Dailey, 2015). In recent years, various powerful techniques that incorporate high-throughput sequencing into reporter assays enabled quantitative and straightforward measures of enhancer activity for thousands of regulatory elements. This chapter will summarize the recent powerful assays for functional testing of *cis*-regulatory element activity.

### 2. Classical reporter assays

The classical reporter assays remain one of the most frequently used methods to validate Enhancer and silencer activity. In these assays, the tested candidate region is placed upstream or downstream from a minimal promoter and a reporter gene (GFP, LacZ, luciferase or others). The reporter construct is then introduced into living cells/embryos of interest. If the candidate sequence acts as an enhancer, it will activate the promoter and lead to the transcription of the reporter gene. If the candidate sequence acts as a silencer, it will deactivate the promoter and lead to the repression of the reporter gene. The levels of reporter gene product (mRNA or protein) can be detected by LacZ staining, fluorescence or *in situ* hybridization or quantified using bioluminescence such as in the luciferase assays. Traditionally drug-resistant colony assay was also used to detect silencer and insulator activity. The expression level of reporter products reflects the strength of the element used (Fig 3.1).

The traditional reporter assay serves as a simple, fast and efficient method to test the enhancer or silencer activity and is still used. However, they are considered as low throughput method and time consuming because every single candidate has to be cloned into reporter construct and tested one-by-one.



**Figure 3.1 | Traditional reporter assays for enhancer or silencer discovery.** The tested DNA fragment is cloned into a construct which contains a reporter gene encode for a gene product which is easily detected and quantified transcriptional activation after transfect/inject into living cells. Expression levels are high for enhancer driven expression (10–100 $\times$ ) and low for repressive activity (1 $\times$ ). Figure adapted from (Petrykowska et al., 2008).

### 3. High-throughput reporter assays

High-throughput strategies have been developed to study mammalian enhancers (Santiago-Algarra et al., 2017), permitting the simultaneous analysis of hundreds of thousands of reporter plasmids at once (Table 3.1). These have been the focus of several comprehensive reviews (Dailey, 2015; Inoue and Ahituv, 2015; White, 2015). 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. A general concern about the episomal reporter assays is that they may not accurately reflect the function of regulatory elements in their endogenous context. These limitations include lack of proper chromatin context, lack of interaction with other regulatory elements, no target gene involved in the assay and no TAD structures. However, it provides valuable information about the intrinsic properties of the sequence itself, which is essential for the initial assessment of any cis-regulatory element. Here, we will focus on recent quantitative methods aiming to characterize enhancers.

#### 3.1. Massively Parallel Reporter Assays (MPRAs)

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 (Fig 3.2).

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-



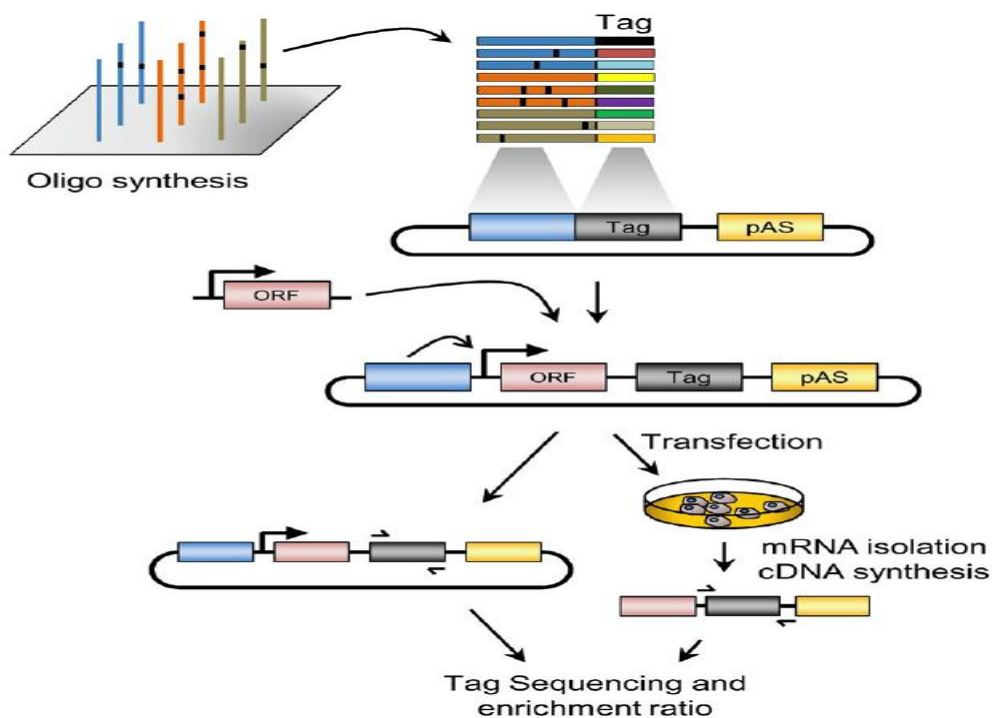
nucleotide resolution (Melnikov et al., 2012; Patwardhan et al., 2009). Subsequently, a similar approach (also named CRE-seq) was used to functionally test ~2,000 genomic segments predicted by ENCODE to be enhancers, weak enhancers, or repressed elements (Kwasnieski et al., 2014), as well as test synthetic enhancers to model grammatical rules of regulatory sequences (Nguyen et al., 2016; Smith et al., 2013). 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) (Kheradpour et al., 2013). 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 regions (Ernst et al., 2016). 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 (Ernst et al., 2016; Nguyen et al., 2016). 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 lines (Tewhey et al., 2016). Ulirsch *et al.* used MPRA to test 2,756 variants linked to 75 genome wide association studies (GWAS) loci involved in red blood cell traits (Ulirsch et al., 2016). In both cases, CRISPR-mediated genetic engineering confirmed the relevance of the MPRA findings (Ulirsch et al., 2016).

Most MPRA approaches have used massive oligonucleotide synthesis (Fig 3.2), 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.

### **3.2. Self-Transcribing Active Regulatory Region Sequencing (STARR-seq)**

An innovative method named STARR-seq was introduced by Alexander Stark and colleagues (Arnold et al., 2013)(reviewed (Muerdter et al., 2015)). STARR-seq aimed to identify and quantify transcriptional enhancers directly based on their activity in entire genomes (Fig 3.3). In brief, a bulk of DNA fragments from arbitrary sources is cloned downstream into the 3' UTR of a GFP reporter gene. Once in a 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 (Arnold et al., 2013; Yanez-Cuna et al., 2014), and hormone-responsive enhancers (Shlyueva et al., 2014b), (ii) the impact of *cis*-regulatory sequence variation on enhancer activity and evolution (Arnold et al., 2014), and (iii) dissecting the basis of enhancer core-promoter specificity (Zabidi et al., 2015).

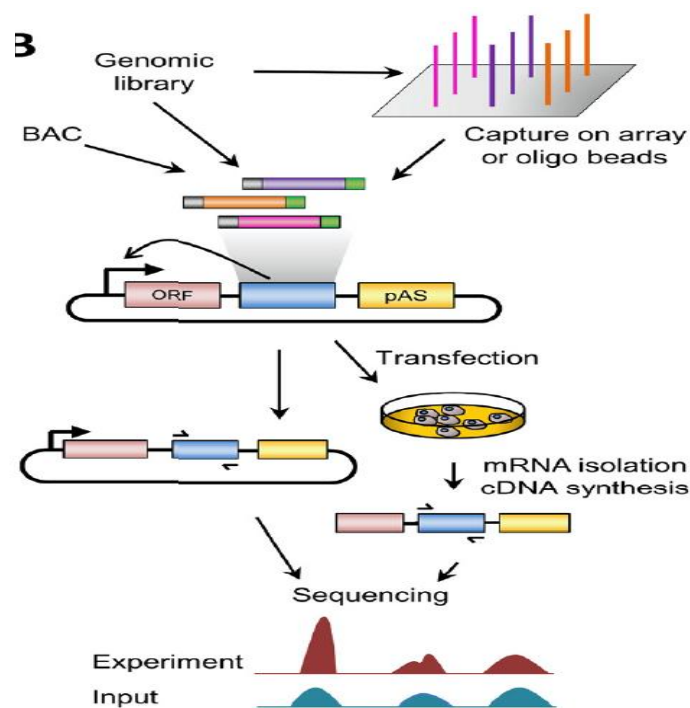


**Figure 3.2 | Overview of massively parallel reporter assay (MPRA).** The test sequences (wild-type, variants, etc.) are generally synthesized *in silico* by massive oligonucleotide synthesis with unique barcode tags and cloned into the plasmid backbone. Tags can be synthesized along with the test sequences 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. Figure adapted from (Santiago-Algarra et al., 2017).

STARR-seq has been applied to human cells by utilizing selected bacterial artificial chromosomes (BACs)(Arnold et al., 2013); 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, a capture-based approach (named CapSTARR-seq) to assess a subset of mouse DNase I hypersensitive sites (DHSs) found in developing thymocytes was developed (Vanhille et al., 2015b). 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 variants (Vockley et al., 2015), and test of DHSs from the central nervous system using a capture approach with oligo-baits (Shen et al., 2016). Alternatively, it could be possible to directly clone open chromatin regions, as described in the functional identification of regulatory elements within the accessible chromatin (FIREWACH) method (Murtha et al., 2014).

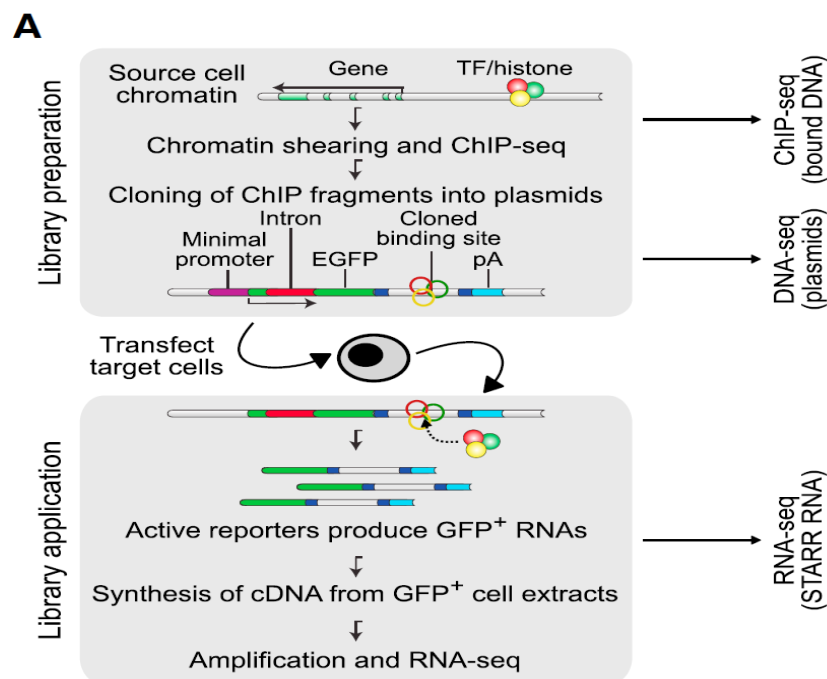
In the STARR-seq approach, the DNA fragments are cloned within the transcribed region, which is very convenient because their sequences provide direct information about enhancer activity. However, it also introduces a source of potential artifacts, as some sequences might influence transcript stability 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.



**Figure 3.3 | 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. Figure adapted from (Santiago-Algarra et al., 2017).

### 3.3. ChIP-STARR-seq

Application of STARR-seq to explore mammalian genomes is hindered by genome size which means that enhancer sequences would be infrequently sampled. This issue can be alleviated by combining ChIP with STARR-seq (Vockley et al., 2015). Using a similar approach (refer to as “ChIP-STARR-seq”), generates a resource of genome-wide activity maps of functional enhancers in ESCs. This identified highly active enhancers with major changes in activity patterns between primed and naive ESCs. Moreover, some transposable element (TE) families are enriched at highly active enhancers. It also identified the functional components within super enhancers (SEs) and uncovers a previously unidentified set of enhancers, including some associated with housekeeping functions. This resource encompasses an extensive collection of functional enhancer sequences in ESCs, providing a knowledge base for systematic analysis of the transcriptional circuitry underlying ESC maintenance and differentiation (Fig. 3.4). By using ChIP-STARR-seq, Barakat *et al.*, assessed the ability of sequences bound by OCT4, NANOG, or marked by H3K4me1 and H3K27ac to function as enhancers (Barakat et al., 2017).



**Figure 3.4 | ChIP-STARR-Seq in Human Embryonic Stem Cells.** Outline of the ChIP-STARR-seq approach combining antibodies against TFs or histone modifications (colored balls) with the STARR-seq plasmid. Figure adapted from (Barakat et al., 2017).

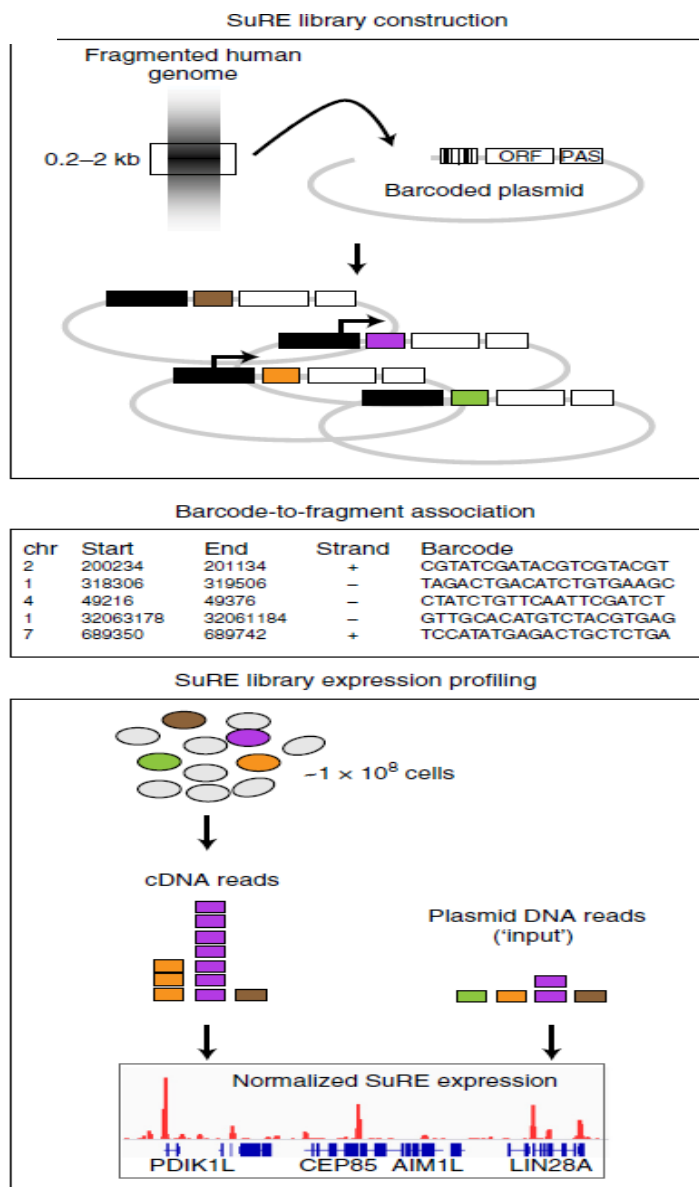
**Table 3.1: Examples of high-throughput functional assays of mammalian enhancers.**  
Table adapted from (Santiago-Algarra et al., 2017)

Original technique	Specific Name	DNA origin	Application	Specific features	No. of regions	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	(Kheradpour et al., 2013)
MPRA	CRE-seq	Synthetic	Characterize genomic regions with predicted function as enhancers	Mutation of TF binding sites	2,100	130	K562	Hsp68	Human	(Kwasnieski et al., 2014)
FIREWACH		Genomic	Identification of enhancers from restriction enzyme accessible regions	Isolation of GFP-positive regions	84,240	154	ESC	Minimal FpG5	Mouse	(Murtha et al., 2014)
STARR-seq	CapSTARR-seq	Genomic	Identification of enhancers from DHS	Capture of DHS	7,542	330–430	P5424, 3T3	SCP1	Mouse	(Vanhille et al., 2015)
STARR-seq		Genomic	Characterize genetic variants in regulatory elements from 95 subjects	Capture of DHS	104	402	HepG2	SCP1	Human	(Vockley et al., 2015)
MPRA	CRE-seq	Genomic	Identification of regulatory elements	Capture of DHS	4,000	464	Retina	Minimal Rho	Mouse	(Smith et al., 2013)
MPRA	Sharpr-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	(Ernst et al., 2016)
MPRA		Synthetic	Identification of regulatory variants associated with eQTLs	Centered on eQTL variants	3,642	150	Lymphoblastoid, HepG2	Minimal TATA	Human	(Tewhey et al., 2016)
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	(Ulirsch et al., 2016)
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	(Dao et al., 2017)
MPRA	LentiMPRA	Synthetic	Identification of enhancers by chromosomal integration	Centered on several ChIP-seq peak signals	2,236	171	HepG2	pGL4.23 promoter	Human	(Inoue and Ahituv, 2015)
CRISPR-Cas9		Synthetic	Identify endogenous enhancers bound by p53 and ER $\alpha$ TFs	Target p53 binding site in putative enhancers	685	N/A	BJ-RASg12v	N/A	Human	(Korkmaz et al., 2016)

### 3.4. Survey of regulatory elements (SuRE)

SuRE provides a genome wide map of promoter activity. Instead of short synthetic promoter sequences, SuRE queries random genomic fragments ranging in size from 0.2 to 2 kb, which is long enough to include most elements that constitute fully functional promoters. Moreover, with SuRE, it is possible to achieve a throughput of  $>10^8$  fragments, which is sufficient to redundantly scan the entire human genome at an average base coverage of  $\sim 55$ -fold.

SuRE data can be interpreted as maps of promoter ‘autonomy’, that is, the degree to which sequences across the genome can act as promoters in the absence of other regulatory elements. Additionally, because each promoter is represented by many partially overlapping random fragments, it is possible to delineate the regions that contribute to its activity. The SuRE maps provide unique opportunities to gain new insights into the biology of human promoters and enhancers. Schematic representation of the SuRE experimental strategy is shown in (Fig. 3.5). Functional screens based on reporter assays have previously been of insufficient throughput to test the vast space of SNPs for possible effects on regulatory element activity. In a subsequent work, the same team used the SuRE approach to systematically assess the regulatory effect of 5.9 million human SNPs, including 57% of the known common SNPs (van Arensbergen et al., 2019).



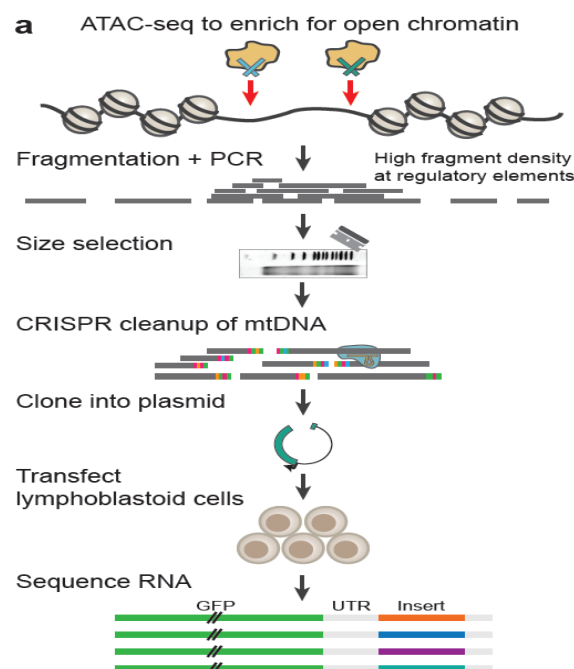
**Figure 3.5 | SuRE provides a genome wide map of autonomous promoter activity.**

Schematic representation of the SuRE experimental strategy. ORF, open reading frame; PAS, polyadenylation signal. Colors indicate different barcodes. Figure adapted from (van Arensbergen et al., 2017)

### 3.5. High-Definition Reporter Assay (HiDRA)

HiDRA, is a novel high-resolution global screen for transcriptional regulatory activity in accessible regions, building on several key ideas from previous technologies to overcome their limitations and combine their advantages, enabling high-efficiency, high-throughput, and high-resolution inference of regulatory activity. HiDRA overcomes these limitations by combining components of Sharpr-MPRA (Ernst et al., 2016) and STARR-Seq (Arnold et al., 2013) with genome-wide selection of accessible regions from ATAC-Seq (Buenrostro et al., 2013). The accessible DNA regions from ATAC-Seq were extracted (Buenrostro et al., 2013), size-select for constructs 150-500nt long, and insert them downstream of episomal reporter's genes to test their activity and exploit their overlapping nature for high-resolution inferences. This approach overcomes the construct-length and region count limitations of synthesis-based technologies, and the ATAC-seq selection of open chromatin regions concentrates the signal on likely regulatory regions and enables high-resolution inferences.

HiDRA was applied to infer genome-wide regulatory activity across ~7 million DNA fragments preferentially selected from accessible chromatin in the GM12878 lymphoblastoid cell line, resulting in ~95,000 active fragments clustering in ~65,000 regions showing significant regulatory function. These are enriched for endogenous active histone marks (including H3K9ac, H3K27ac), regulatory sequence motifs, and regions bound by immune regulators. The ATAC-based selection approach resulted in highly-overlapping fragments, with up to 370 fragments per region, enabling to pinpoint driver regulatory nucleotides. So, ~13,000 high-resolution driver elements were discovered, which are enriched for regulatory motifs and evolutionarily conserved nucleotides, and help predict causal genetic variants underlying disease from genome-wide association studies. Overall, HiDRA provides a general, scalable, high-throughput, and high-resolution approach for experimental dissection of regulatory regions and driver nucleotides in the context of human biology and disease. The overview of HiDRA library preparation is shown in (Fig. 3.6).



**Figure 3.6 | Overview of HiDRA library preparation.** The Tn5 transposase preferentially fragments genomic DNA at regions of open chromatin. Fragments are then size-selected on an agarose gel and mtDNA contamination is removed by selective CRISPR-Cas9 degradation. The fragment library is amplified by PCR and cloned into enhancer reporter vector. Figure adapted from (Wang et al., 2018).

**Table 3.2: Recent studies using high-throughput reporter assays**

Tech.	DNA origin	Application	Specific features	No. of regions	Size (bp)	Cell lines	Promoter	Species	Ref.
Sharpr-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, pSV40	Human	(Ernst et al., 2016)
SuRE	Genomic	Mapping autonomous promoter activity genome wide	Assays more than 10 <sup>8</sup> DNA fragments for their ability to derive autonomous transcription	>10 <sup>8</sup>	200-2000	K562	Sv40	Human	(van Arensbergen et al., 2017)
STAP-seq	Genomic	Assessing the sequence-intrinsic enhancer responsiveness of millions of candidate sequences at single base-pair resolution	Finding TSS-proximal or TSS-overlapping enhancer.		100-250	S2	Minimal promoter, pGL3 Sv40	D. melanogaster	(Arnold et al., 2017)
ChIP-STARR-seq	Genomic	Generate a resource of genome-wide activity maps of functional enhancers in ESCs		350000	~221	H9 ESCs cells	SCP1	Human	(Barakat et al., 2017)
MPRA	Synthetic	Systematic dissection of genomic features determining transcription factor binding and enhancer function	Systematically evaluate the factors that govern PPAR $\gamma$ binding and regulation in adipocytes	32115	145-150	3T3-L1	Minimal promoter	Mouse	(Grossman et al., 2017)
CERES (CRISPR/Cas9)	Synthetic	Targeting DHSs surrounding genes of interest to identify endogenous regulatory elements	Loss- and gain- of-function epigenome editing	433, 281		K562, A431, HEK293T	PGK, HBE1, HER2	Human	(Klann et al., 2017)
WHG-STARR-seq	Genomic	Obtain an assessment of enhancer activity	Whole human genome STARR-seq	94,527	350-650	LNCaP	SCP1, miniCMV	Human	(Liu et al., 2017)
methyl-STARR-seq	Genomic	Assay the causal relationship between initial differences in DNA methylation and subsequent regulatory activity in high-throughput, within the cellular environment	Identify thousands of MD regulatory elements in the human genome		300-700	K562	CpG-free EF1 promoter	Human	(Lea et al., 2018)
SNPs-seq	Genomic	Simultaneously screen hundreds to thousands of SNPs for their allele-dependent protein-binding differences	Examine SNP dependent transcriptional regulation at candidate SNP regions		161	LNCap,22Rv1, VCaP	minimal promoter, RGS17, pGL4	Human	(Zhang et al., 2018)
BiT-STARR-seq	Synthetic	characterization of genetic effects on DNA-protein binding and gene transcription	Different categories of regulatory variants including eQTLs, CentiSNPs, ASB SNPs, variants associated with complex traits in GWAS	50,609	230	Lcls	Plasmid pGL4	Human	(Kalita et al., 2018)
MPRA	Synthetic	assess the enhancer capacity across lincRNA and mRNA loci	Enhancer capacity of lincRNAs tested on their loci at 50-bp resolution. Full gene body of each locus and 1000 bp upstream.	22,000	100-500	C2C12	Minimal promoter	Mouse	(Groff et al., 2018)
HIDRA	Genomic	High-resolution genome-wide functional dissection of transcriptional regulatory regions in human	Assay the regulatory activity of millions of open chromatin-derived fragments located genome-wide	65,000	169-477	GM12878		Human	(Wang et al., 2018)
STARR-seq	Genomic	Whole genome STARR-Seq	ORI is an optimal core-promoter for STARR-seq and luciferase assays	9,613	1000-1500	HeLa-S3,	ORI	Human	(Muerdter et al., 2018)
STARR-seq	Synthetic	Functionally screen thousands of sequences for regulatory activity in parallel	Conventional episomal	2,236	192-678	HepG2, HEK293T	Minimal promoter, pGL4.23,ORI, HSS	Human	(Klein et al., 2019)



#### **4. Other High throughput reporter assays**

Several other high-throughput assays have also been developed for example for performing genome-scale tests of methylation dependent regulatory activity; mSTARR-seq (Lea et al., 2018) was developed. To characterize variants associated with complex traits in noncoding regions that contribute to phenotypes by disrupting regulatory sequences, Biallelic Targeted STARR-seq (BiT-STARR-seq) assay was developed (Kalita et al., 2018). To generate the first whole-genome-scale enhancer mapping and activity quantification in human cell lines whole human genome STARR-seq (WHG-STARR-seq) (Liu et al., 2017), was developed. Functional characterization of disease causing variants at risk loci has been a significant challenge. A high-throughput single-nucleotide polymorphisms sequencing (SNPs-seq) (Zhang et al., 2018), technology to simultaneously screen hundreds to thousands of SNPs for their allele-dependent protein-binding differences have been also developed. Weingarten *et al.*, devised a high-throughput assay to quantify the activity of approximately 15,000 fully designed sequences that were integrated and expressed from a fixed location within the human genome (Weingarten-Gabbay et al., 2019). This method is used to investigate thousands of native promoters and preinitiation complex (PIC) binding regions followed by in-depth characterization of the sequence motifs underlying promoter activity, including core promoter elements and TF binding sites.

#### **5. Medium and High-throughput approaches to identify silencers**

Here are some of the medium and high throughput techniques used to identify silencer elements (Table 3.3). Petrykowska *et al.*, examined 47 sequences from a 1.8-Mb region encompassing the *CFTR* gene of human chromosome 7 for silencer and enhancer blocker (EB) activities using luciferase reporter assay. The majority of functional elements displayed directional and promoter-specific activities. A limited number of sequences acted in a dual manner, as both silencers and EBs. (Petrykowska et al., 2008). While using an enhancer-blocking assay, Qi et al. identified T39 human autonomous silencer in ARHGAP6 gene and validated by Drug-resistant colony assay and GFP reporter assay (Qi et al, 2015). Later, Capture Hi-C (CHi-C), an adapted genome conformation assay was used, to examine the long-range interactions of almost 22,000 promoters in 2 human blood cell types. Over 1.6 million shared and cell type-restricted interactions spanning hundreds of kilobases between promoters and distal loci were identified. Transcriptionally active genes contact enhancer-like elements, whereas transcriptionally inactive genes interact with previously uncharacterized elements marked by repressive features that may act as long-range silencers (Mifsud et al., 2015). Then Sharpr-MPRA technique was used to test 4.6 million nucleotides spanning 15,000 putative regulatory regions tiled at 5-nucleotide resolution in two human cell types. Known cell-type-specific regulatory motifs and evolutionarily conserved nucleotides, and distinguished known activating and repressive motifs were recovered. In the same study Ernst et al., demonstrated that endogenous chromatin state and DNA accessibility are both predictive of regulatory function in reporter assays, identified retroviral elements with activating roles, and uncovered 'attenuator' motifs with repressive roles within active chromatin regions (Ernst et al., 2016). Similarly, a simple but powerful computational

approach to identify putative silencers genome-wide was used (Jayavelu et al 2018). Jayavelu et al., used a series of consortia data to predict silencers in over 100 human and mouse cell or tissue types. Motif enrichment analyses on putative silencers determined that motifs belonging to known transcriptional repressors are enriched, as well as overlapping known transcription repressor binding sites. Next, silencer activity was validated using massively parallel reporter assays (MPRAs) (Jayavelu et al 2018). Recently, Huang et al. 2019 utilized H3K27me3-DNase I hypersensitive site (DHS) peaks with tissue specificity negatively correlated with the expression of nearby genes across 25 diverse cell lines. These regions are predicted to be silencers since they are physically linked using (Hi-C loops) or associated using (expression quantitative trait loci (eQTL) results) with a decrease in gene expression much more frequently than general H3K27me3-DHSs. Also, these regions are enriched for the binding sites of repressors (such as CTCF, MECOM, SMAD4, and SNAI3) and depleted of the binding sites of transcriptional activators. Using sequence signatures of these regions, a computational model was constructed they identified that silencers commonly interact with enhancers to affect the transcriptional dynamics of tissue-specific genes and to facilitate fine-tuning of transcription in the human genome (Huang et al., 2019).

**Table 3.3: Identification of silencers using medium and high-throughput approaches**

Experimental approach	Validation	Target gene	Reference
<b>Tested 47 sequences from a 1.8-Mb region surrounding the human CFTR locus by Luciferase Reporter Assay</b>	None	CFTR	(Petrykowska et al., 2008)
<b>Testing of CTCF binding sites using an enhancer-blocking assay</b>	Drug-resistant colony assay, GFP reporter assay	ARHGAP6, (T39 silencer)	(Liu et al., 2015; Qi et al., 2015)
<b>Genomic regions interacting with inactive genes by Capture Hi-C</b>	Luciferase Reporter Assay	BCL6	(Mifsud et al., 2015)
<b>CRISPRi screening of the genomic region surrounding the MYC locus</b>	None	MYC	(Fulco et al., 2016)
<b>Massively parallel reporter assays (MPRAs) using the SV40 promoter</b>	None		(Ernst et al., 2016)
<b>simple subtractive analysis (SSA) approach and ENCODE consortium to get silencer from open chromatin data either from DNase-seq or ATAC-seq, ChIP-seq for any cell type or organism</b>	MPRA	Not described	(Jayavelu et al, 2018; <a href="http://dx.doi.org/10.1101/252304">http://dx.doi.org/10.1101/252304</a> )
<b>H3K27me3-DHS peaks with tissue specificity negatively correlated with the expression of nearby genes</b>	Luciferase Reporter Assay	Not described	(Huang et al., 2019)
<b>CapSTARR-seq using the PGK promoter</b>	Luciferase assay, CRISPR/Cas9	HCST, Nfkbid (Sdhaf1 silencer)	This study (see results section)

## Chapter 4

### T cell differentiation

The thymus is the organ specialized to make T cells. T cells originate from hematopoietic stem and precursor cells in the bone marrow or fetal liver, which migrate to the thymus and acquire T-cell identity. Relatively small numbers of T-cell progenitors migrate into the thymus per day, but they respond to the new environment by undergoing multiple rounds of proliferation while initiating the T-cell differentiation program (Petrie and Zuniga-Pflucker, 2007; Rothenberg, 2014). They then undergo T-cell lineage commitment, begin T-cell receptor (TCR) rearrangements, and thus generate  $\alpha\beta$ TCR- or  $\gamma\delta$ TCR-expressing T cells. The  $\alpha\beta$  T cells further diverge into different sublineages, as CD4 T cells, CD8 T cells, natural killer T (NKT) cells and regulatory T (Treg) cells, ultimately to act as key players of cell-mediated immunity. T-cell differentiation is subjected to several developmental checkpoints that are controlled by either pre-TCR or TCR complexes and can be followed by the expression of surface markers, such as CD4 and CD8. In particular, (pre)-TCR downstream signaling triggers complex intracellular pathways resulting in wide changes in the transcriptional and epigenetic program of genes associated with cell survival, proliferation and differentiation (Carpenter and Bosselut, 2010). Therefore, T cell differentiation provides an excellent model for the study of gene regulatory programs

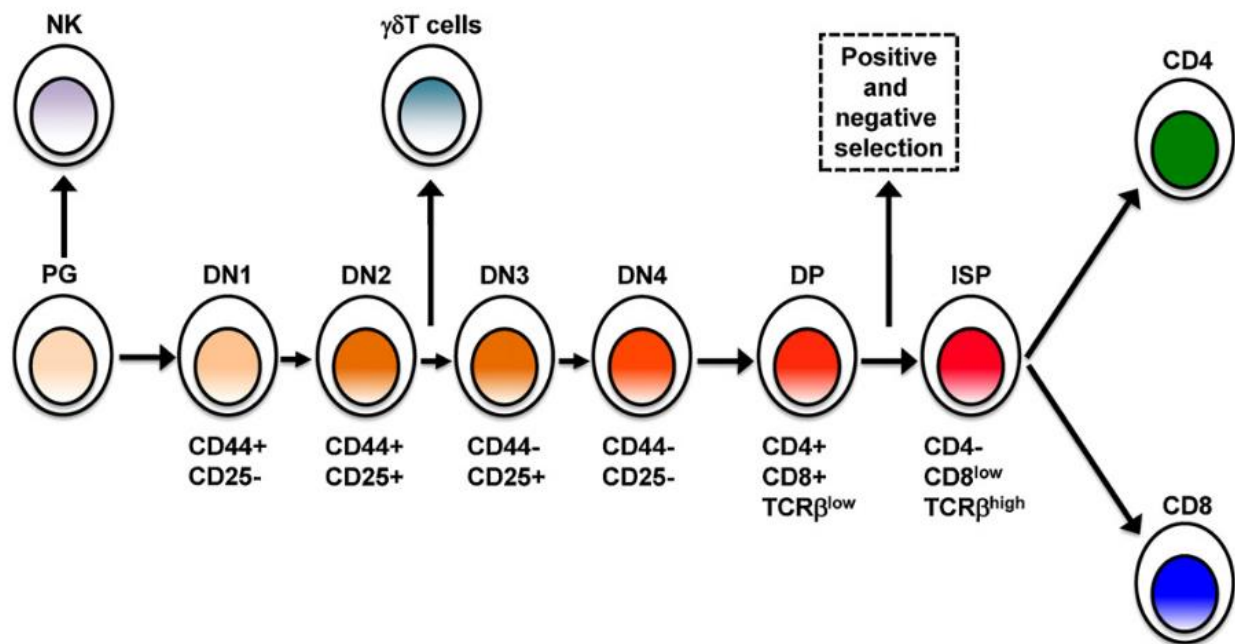
Thymocytes are divided into multiple phenotypically distinct stages that are defined by the expression of CD4, CD8, and other markers (Hayday and Pennington, 2007; Rothenberg, 2014; Rothenberg et al., 2008) (Fig 4.1). T-cell development is initiated from the subpopulation that lacks the expression of both CD4 and CD8, thus called double-negative (DN) cells, which then become CD4<sup>+</sup> CD8<sup>+</sup> double-positive (DP) and subsequently differentiate into mature CD4 or CD8 single-positive (SP) cells. The earliest T-cell precursors in the thymus, called early thymic progenitor (ETP) or Kit-high double-negative 1 (KIT<sup>++</sup> DN1; CD44<sup>+</sup> CD25<sup>-</sup>), still harbor the potential to gain access to non-T-alternative fates. These cells start expressing T-cell markers in the next stage, DN2a (KIT<sup>++</sup> CD44<sup>+</sup> CD25<sup>+</sup>), but commitment to the T-cell lineage occurs only at the following stage, DN2b (Kit<sup>+</sup>CD44<sup>+</sup> CD25<sup>+</sup>). Then in the DN3a (KIT<sup>-</sup> CD44<sup>-</sup> CD25<sup>+</sup>) stage, TCR $\beta$  gene rearrangement begins. This process enables some cells to express either a pre-TCR (TCR $\beta$  with invariant pre-TCR $\alpha$ ) or a  $\gamma\delta$ TCR. Pre-TCR mediated signal transduction triggers the transition of DN3a cells through DN3b into DN4 (Kit<sup>-</sup> CD44<sup>-</sup> CD25<sup>-</sup>), followed by progression to the DP stage. DP thymocytes undergo TCR $\alpha$  gene rearrangement and begin to express fully assembled  $\alpha\beta$ TCR. Then, they are subjected to a selection process, which is known as positive selection, to identify cells that express TCR with potentially useful ligand specificities. Positively selected thymocytes are allowed to differentiate into either CD4 helper T cells or CD8 cytotoxic T cells, known as CD4/CD8-lineage choice.

## Positive and negative selection of $\alpha\beta$ lineage cells

The differentiation of  $\alpha\beta$  lineage cells, emerge from  $\beta$ -selection as DP thymocytes and are the precursors of conventional CD4 and CD8 cells. Three key events mark the developmental progression of these cells: (i) positive selection, the rescue from programmed cell death of DP thymocytes whose TCR $\alpha\beta$  productively interacts with self MHC peptide complexes (MHCp) expressed by the thymic epithelium (or with other MHC or MHC-like molecules (Starr et al., 2003) (ii) negative selection, the elimination of self-reactive cells, and (iii) acquisition of functional competence, notably marked by the termination of either CD4 or CD8 expression ('lineage differentiation') and its matching to MHC specificity.

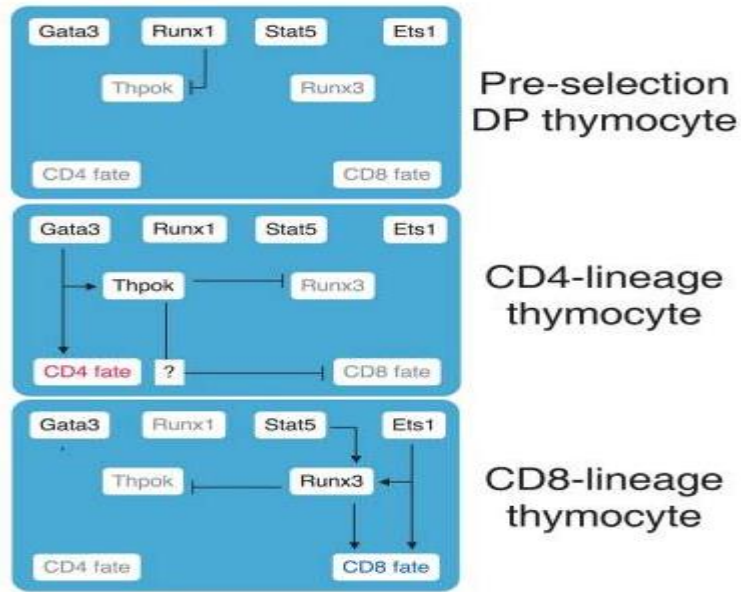
The need for positive selection is a direct consequence of the random nature of TCR rearrangement and of the high diversity of MHC alleles. As a result, most DP thymocytes in a given individual fail to productively interact with MHCp and die by 'neglect' within a few days, even though evolutionary pressure has resulted in a 'germline-encoded' MHC reactivity of TCR variable regions, thereby increasing the yield of positive selection (Feng et al., 2007). CD4 and CD8 molecules provide an additional guard against the selection of non-MHC reactive cells, as they sequester the tyrosine kinase Lck required to initiate TCR signaling, thereby restricting its activity to TCRs engaged by MHCp, which, unlike non MHC-ligands, co-engage CD4 or CD8 (Van Laethem et al., 2007).

In addition to being rescued from cell death (positive selection *per se*) TCR signaled thymocytes undergo functional differentiation into mature T cells. One key aspect of this process is the differentiation into the CD4 or CD8 lineage. This includes the termination of expression of either coreceptor, and the initiation of gene expression programs characteristic of helper (CD4) or cytotoxic (CD8) cells, two events that have long been recognized as being mechanistically coupled (Corbella et al., 1994). Because a functional immune system requires that this lineage 'decision' be matched to MHC specificity, so that MHC II-restricted thymocytes become helper CD4 cells, and MHC I-restricted thymocytes cytotoxic CD8 cells, its mechanisms have attracted much attention over the last two decades. The last few years have seen significant progress in the elucidation of the transcriptional circuits that promote CD4 or CD8 differentiation.



**Figure 4.1 | Schematic representation of T cell development in the thymus.** Progenitor (PG) cells from the bone marrow colonize the thymus where they commit to natural killer cells (NK) or T cell lineage. Here they undergo progressive differentiation from double-negative (DN) to double-positive (DP) to single-positive CD4 or CD8 thymocytes. The immature DN thymocytes can be further divided into DN1 to DN4 based on their expression of CD44 and CD25. The gd lineage diverge from the ab at the DN2/DN3 stage. Progression from DP stage to SP can occur through an immature CD8+ single positive (ISP) cell intermediate. Figure adapted from (D'Acquisto and Crompton, 2011).

Two transcription factors, *Thpok* and *Runx3*, specifically expressed in CD4 and CD8 differentiating thymocytes, respectively, are important for this process (Taniuchi et al., 2002a). *Thpok* is required for CD4 commitment and acts at least in part by repressing expression of CD8 lineage genes, including *Runx3* (Taniuchi et al., 2002a). *Runx3* is important for the silencing of *Cd4* in CD8 cells (Woolf et al., 2003), and the complete disruption of *Runx* activity (that is of *Runx3* and the partly redundant factor *Runx1*) prevents CD8 cell development (Wang et al., 2008a). While this effect is due in part to unrestrained *Thpok* expression in *Runx*-deficient thymocytes (Setoguchi et al., 2008), cells lacking both *Runx* and *Thpok* activities fail to become CD8, indicating a specific role of *Runx* proteins in CD8-lineage differentiation (Egawa and Littman, 2008). These findings have led to the proposal that a dual negative regulatory loop involving *Thpok* and *Runx3*, which mutually prevent expression of each other, results in lineage commitment (Fig 4.2).



**Figure 4.2 | CD4-CD8 lineage differentiation:** Components of the transcriptional circuitry that promotes CD4-CD8 differentiation are schematically depicted and interconnected at three stages of T cell development. In preselection cells, Runx1-nucleated activities repress *Thpok* expression. In CD4-differentiating cells, Runx1-mediated *Thpok* expression is relieved, although Runx1 is still expressed in CD4 cells in which it binds the *Thpok* gene. Gata3 promotes both *Thpok* expression and additional developmental events required for CD4 cell differentiation. *Thpok* prevents *Runx3* up-regulation and CD8 differentiation. In CD8-differentiating cells, *Thpok* repression is maintained, presumably through Runx3. Ets1 promotes *Runx3* expression, and binds the *Runx3* locus, whereas Stat5 has been reported to relay IL-7 signaling to *Runx3* (Park et al., 2010). Grey lettering indicates factors not expressed at a particular stage. Other factors (including Tox) are omitted for clarity. Arrows or block signs do not imply direct effects. Figure adapted from (Carpenter and Bosselut, 2010).

## **Results**

# Chapter 5

## Results

### 1 Objective

Like other cis-regulatory elements, silencers activity can be assessed by reporter assays where a DNA fragment is tested for its ability to reduce the activity of a promoter element associated with a reporter gene. While several high-throughput reporter assays have been developed to systematically assess enhancer function, no massive reporter strategy for silencers has been published. Our team has previously set-up a high-throughput approach, named CapSTARR-seq (Vanhille et al., 2015), to quantitatively assess enhancer activity in mammals coupling capture of regions of interest to massive paralleled Starr-seq reporter assay. Using a mouse leukemic cell line it was demonstrated that this technique provides accurate quantification of enhancer activity from hundreds of thousands of DNA fragments in parallel. Therefore, we repurposed the CapStarr-seq technique in order to quantify the silencer activity.

Identification and characterization of silencer elements along with their actual biological functions have become the central focus of my study. There are some outstanding questions that I will address in this chapter.

- a) Could the High throughput screening identify silencer elements?
- b) Do silencers represent a major genomic strategy of gene regulation?
- c) What are the main genomic and epigenomic features of silencers?

### 2. Main results

Using the captured DHS genomic library in CapSTARR-seq technique with four different promoters we identified a substantial number of putative silencers in p5424 cell line. Performing luciferase reporter assay of twenty-eight silencer candidates, twenty-two showed silencer activity. The motif analysis studies showed enrichment of REST motif in pPGK and pSCP1 based silencers, which is validated through luciferase reporter assay after site specific mutagenesis. Our lab performed RNA-seq experiments with control and PMA/Ionomycin treated P5424 cells. Analyses of this dataset reveal the regulation of key T cell factors suggesting PMA/Ionomycin mimic T cell differentiation and beta-selection. Using



CRISPR/Cas9 technique, we try to delete four silencer elements, but only *Dpp9* and *Sdhaf1* silencers got homozygous clones (see Annex). However, *Dpp9* homozygous clones were lost upon culture. The *Sdhaf1* silencer is associated with two genes *Hcst* and *Nfkbid*, which are highly regulated during T cell differentiation and involved in T cell activation and function. Deletion resulted in up-regulation of *Hcst* in unstimulated condition and *Nfkbid* in stimulated condition. These results show CapSTARR-seq is a better high through-put approach to screen silencer elements and silencer are a major component of gene regulation with important transcription factors enrichment.

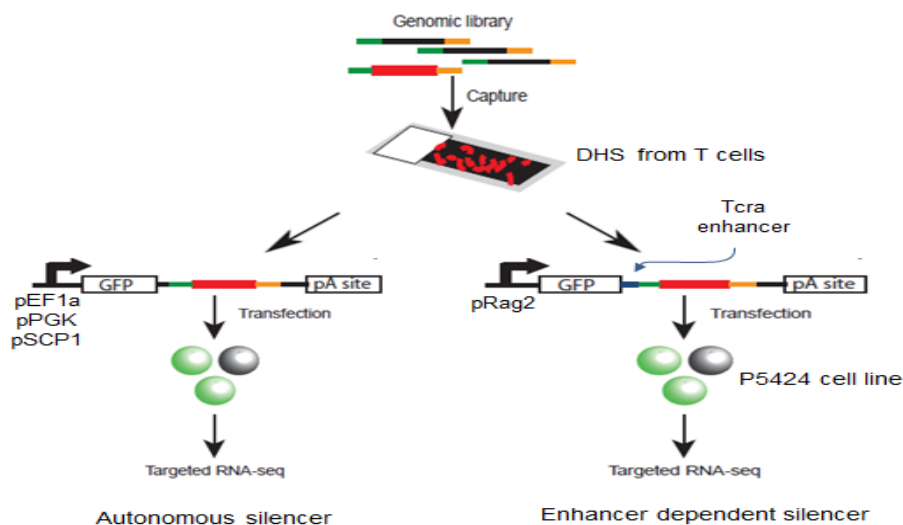
### 3. Contributions

*Experimental contribution:* To carry out this project, my supervisor and I routinely discuss conceptual and experimental designs and I performed almost all experimental work,

which includes:

- Performing CapSTARR-seq with four distinct promoter-based reporter vectors in mouse p5424 cell line (Fig 5.1).
- Cloning, Luciferase assay
- Design and perform CRISPR/Cas9-mediated knock-out strategies and screening.
- RT-qPCR analysis, gene expression analysis.

Manuscript contribution: Contribute to write the manuscript and editing figures



**Figure 5.1| CapSTARR-seq experimental scheme.**

## Manuscript: High-throughput identification of silencer elements

**Authors:** Saadat Hussain<sup>1</sup>, Nori Sadouni<sup>1</sup>, Dominic van Essen<sup>3</sup>, Pascal Lopez, Lan T.M. Dao<sup>1,2,4</sup>, Guillaume Charbonnier<sup>1</sup>, Magali Torres<sup>1,2</sup>, Simona Sacconi<sup>3</sup>, Salvatore Spicuglia<sup>1,2\*</sup>

<sup>1</sup>Aix-Marseille University, Inserm, TAGC, UMR1090, Marseille, France.

<sup>2</sup>Equipe Labélisée Ligue Contre le Cancer, Marseille, France.

<sup>3</sup>Institute for Research on Cancer and Ageing, IRCAN 06107 Nice

<sup>4</sup>Present address: Vinmec Research Institute of Stem cell and Gene technology (VRISG), Vietnam.

\***Correspondence:** [salvatore.spicuglia@inserm.fr](mailto:salvatore.spicuglia@inserm.fr)

### Introduction

The precise regulation of gene expression during normal development and cell differentiation underpins the action of cis-regulatory elements with either activation or repression functions (Chatterjee and Ahituv, 2017; Maston et al., 2006; Ogbourne and Antalis, 1998). Gene activation by the combined activities of promoters and distal enhancers has been extensively studied in normal and pathological contexts. In sharp contrast, gene repression by cis-acting **silencers**, defined as genetic elements that negatively regulate gene transcription in a position independent fashion, is less well understood. Silencers were first described more than 3 decades ago in yeast and vertebrates (Brand et al., 1985; Kadesch et al., 1986; Steiner et al., 1987). Since then, several silencers have been discovered to control the expression of key developmental and immunological model genes, and some progress has been made to characterize various features of a few of these individual silencers (e.g. (Ogbourne and Antalis, 1998; Setoguchi et al., 2008; Taniuchi et al., 2002a). Nevertheless, despite the widely-held belief that silencers likely represent important and critical general regulators of gene expression, this view is still largely conjectural, and their genome-wide distribution, mechanisms of action and involvement in oncogenic transformation are largely unknown.

A breakthrough in the analysis of distal cis-regulatory elements was made several years ago by the identification of chromatin ‘signatures’ which are predictive of enhancer activity genome wide (Natoli and Andrau, 2012). Similarly, the discovery of CTCF protein as an

essential factor of Insulator function leads to the comprehensive mapping and characterization of these regulatory elements (Ong and Corces, 2014). These findings have enabled a multitude of functional studies of enhancers and insulators, and led to an explosion of discoveries, including their roles in the regulation of many disease-related genes and their involvement in the development of diverse pathologies and cancer (Herz et al., 2014; Robson et al., 2019; Smith and Shilatifard, 2014). So far, no chromatin ‘signature’ or mark has been identified which could predict potential silencers. As a result, we have neither a global map of silencers nor any rough estimate of their abundance in the genome or of the extent of gene regulation by silencers.

Noticeable, among the few silencers that have been described in the literature, many are associated with the regulation of T cell specific genes. These included silencer associated with the expression of TCA3/CCL1 (Oh et al., 1997), Il2 (Williams et al., 1991), CD4 (Sawada et al., 1994), TCRb (Dombret et al., 1996), ThPOK (He et al., 2008a), Rag1-Rag2 (Yannoutsos et al., 2004) and CD8 (Yao et al., 2010) loci. Several of these silencers have been shown to play an important role in cell lineage restriction: for instance, the CD4 and CD8 silencers repress the expression of the associated genes in CD8<sup>+</sup> and CD4<sup>+</sup> T cells, respectively. T cell differentiation thus provides an excellent model for the implementation of a high-throughput strategy to identify silencers.

In line with their operational definitions, assays for enhancers and silencers measure their ability to enhance or silence gene expression in cis, when driven by an independent promoter (Chatterjee and Ahituv, 2017; Petrykowska et al., 2008; Qi et al., 2015). Episomal reporter assays have been widely used to characterize putative regulatory regions; however, their implementation has been traditionally expensive, laborious, low-throughput, and low resolution. Several recent high-throughput reporter assays for enhancer function enable testing of thousands of distinct DNA sequences simultaneously, by cloning variable DNA fragments into common reporter constructs, and using high-throughput sequencing to quantify fragment activity (Santiago-Algarra et al., 2017). In particular, the Self Transcribing Active Regulatory Region Sequencing (STARR-seq) approach allows direct genome-wide investigation of enhancer activity using DNA fragments directly collected from genomic DNA (Arnold et al., 2013). We recently developed CapStarr-seq (Vanhille et al., 2015b), a high-throughput reporter strategy coupling capture of regions of interest to massive paralleled

Starr-seq reporter assay, providing a cost-effective method to assess cis-regulatory functions in mammals.

Here we repurposed the CapStarr-seq approach to identify active silencers from isolated genomic regions. We systematically tested DNase hypersensitive sites (DHS) from developing T cells, using four distinct promoter-based reporter vectors. We compared the set of silencers identified with the different vectors and evaluated their association with genomic and epigenomic features. The robustness of the approaches was extensively assessed by independent episomal end integrative reporter assays. Finally, CRISPR–Cas9 genomic manipulation demonstrated the involvement of one silencer in the concomitant repression of two key T cell genes. Overall, we provide a general, scalable, and high-throughput approach for the high-resolution experimental dissection of silencer elements in the context of human biology and disease.

## Results

### *Experimental design to identify silencers*

To test the silencer activity we repurposed the CapStarr-seq technique, a novel approach coupling capture of defined regions to the previously developed Starr-seq technique (Arnold et al., 2013) (Figure 1a; see also Methods section). We replaced the basic pSCP1 promoter present in the enhancer-STARR-seq vector by two ubiquitous promoters from the human *Efl1 $\alpha$*  and *Pgk* genes (Qin et al., 2010) and a tissue-specific promoter enhancer pairs pRag2-E $\alpha$  (Wei et al., 2005) (**Figure 1A**). Analyses of GFP expression by FACS validated all the promoter vectors (**Supplemental figure 1A**). Although little is known about the general biochemical properties of silencers, it seems reasonable to assume that the majority of silencers may be occupied by sequence-specific transcription factors and/or lie within nucleosome-depleted genomic regions, and consequently may overlap with DNaseI Hypersensitive Sites (DHS). Indeed, several known silencers have been identified as laying within DHS sites (e.g. (Donda et al., 1996; He et al., 2008b; Qi et al., 2015)). Therefore, to isolate silencer elements genome-wide, we designed a captured library (Agilent) containing 28055 DHSs from developing thymocytes (DP), plus 437 randomly selected non-DHS regions (as negative controls).

In brief, DNA fragments of ~400 bp were captured on a custom-designed microarray covering all the DHS and cloned by homologous recombination into the four different Starr-seq vectors. The CapStarr-seq libraries were transfected in triplicate into the mouse T-cell line P5424 and sequenced by targeted RNA-seq. The P5424 cell line is originated from early developing T cells and resembles DP thymocytes at phenotypic and transcriptomic levels and have been previously used in CapStarr-seq experiments (Mombaerts et al., 1995; Vanhille et al., 2015b). As controls, we sequenced the cloned DHS libraries obtained before transfection (hereafter, input). Cis-regulatory activity was assessed by computing the log<sub>2</sub> fold change between the targeted RNA-seq and the input signals for each library (hereafter referred to as CapStarr-seq signal) after normalization and filtering (See Methods section; **Supplemental Table 1**). A good correlation was obtained between the replicates of the same library, with Spearman's correlation coefficient ( $\rho$ ) ranging between 0.45 and 0.82 (**Figure 1B**). The pPGK vector provided the highest reproducibility; however, correlation between different libraries was relatively poor, with pPGK and pEF1 $\alpha$  libraries providing the best cross-correlation. For subsequent analyses, the signals from the replicates of the same library were merged.

To assess the meaningfulness of the CapSTARR-seq results obtained with the different libraries, we associated the signals with the expression of the surrounding genes using available RNA-seq data from P5424 cells (Saadi et al., 2019) (**Figure 1C**). CapStarr-seq signal obtained with the pSCP1 library was positively correlated with gene expression, suggesting that the pSCP1 vector is able to detect both enhancer and silencer activity. The opposite was observed with the T cell-specific pRag2-E $\alpha$  vector, suggesting that the presence of the E $\alpha$  enhancer strongly interfere with additional cis-regulatory sequences. The pPGK and pEF1 $\alpha$  vectors provided an intermediate situation whereby lower CapStarr-seq signals correlated with reduced gene expression, but there was no correlation between high CapStarr-seq signals and gene expression. Consistent results were obtained when subdividing the DHS regions in function of the log<sub>2</sub>(FC) of the CapStarr-seq signal (**Figure 1D**). Therefore, the two strong ubiquitous promoters (pPGK and pEF1 $\alpha$ ) appeared to be suitable to assess for silencer, but not for enhancer, activity.

Based on the previous results, we defined the putative silencers as either the DHS regions with the lowest CapStarr-seq signal (bottom 500) or with a log<sub>2</sub>(FC) lower than or equal to -1. The later definition resulted in a set of 136, 564, 333, 1372 putative silencers for the

pEF1 $\alpha$ , pPGK, pRag2-E $\alpha$  and pSCP1 libraries, respectively. Comparison of the putative silencers obtained with each library is provided in Figure 1E. The highest overlap was obtained between the pPGK and pEF1 $\alpha$  libraries (**Figure 1G**), consistent with the overall correlation between the two data sets (**Figure 1B**) and the similar association with gene expression (**Figure 1C**). Few regions displayed consistent silencer activity in all the libraries (e.g. **Figure 1F**).

To assess whether there were biases on the genomic location of silencers, we computed the specific enrichment of putative silencers obtained with each library with respect to the whole set of DHS (**Figure 1H, Supplemental Figure 1B**). While pSCP1 and pEF1 $\alpha$  based silencers were not biased with respect to the DHS, the pPGK and pRag2-E $\alpha$  based silencers were significantly depleted on genic regions and enriched on promoter and intergenic regions. Importantly, none of the silencer sets were enriched for terminator sequences, which could represent a potential bias of the approach by artificially interfering with the quantification of the Starr-seq vector-derived transcripts.

### *Chromatin features associated with silencers*

To initially explore whether CapStarr-Seq results reflect the epigenetic status of endogenous silencers, we analyzed several epigenomic features from the P5424 cell line and primary DP thymocytes. pSCP1-based silencers were associated with lower levels of active histone marks (H3K4me1, H3K4me2, H3K4me3 and H3K27ac) (**Figure 2; Supplemental figure 2A**) and RNA polymerase II (Pol II) (**Supplemental figure 2B-C**), as compared with putative enhancers and DHS with intermediated CapStarr-seq signals. On the contrary, the repressive mark H3K27me3, but not H3K9me3, was highly enriched at putative silencers (**Figure 2; Supplemental figure 2A**), along with a higher density of nucleosome occupancy and similar levels of DNaseI accessibility (**Supplemental figure 2B-C**). The pEF1 $\alpha$  and pPGK-based silencers displayed more variable but consistent profiles, while pRag2-E $\alpha$ -based silencers were generally not correlated with the chromatin features (**Figure 2; Supplemental figure 2**), consistent with the lack of association with gene expression observed previously (**Figure 1 C-D**). Overall, our results suggested that silencer elements are found in a relatively open chromatin configuration associated with Polycomb-mediated silencing.

### ***Validation of CapStarr-seq identified silencers***

To independently evaluate the accuracy of CapStarr-seq to identify silencers, we have selected at least five individual candidates and five overlapped silencer candidates from each promoter strategy used. In total, we tested 39 DHS in a classical luciferase reporter assay in the P5424 cell line (**Figure 3A**). 78.6% (22 out of 28) putative silencers and 36.3% (4 out of 11) control regions, displayed significant silencer activity in the luciferase assay. Overall, CapStarr-seq identified silencers displayed a statistically higher silencer activity in the luciferase assay as compared with the control DHS set (**Figure 3B**). Importantly, several CapStarr-seq-defined silencers (*Cfdp1*, *Jkamp*, *Dirc2*, *Rbfox3*, *Dpp9*, and *Sdhaf1*) displayed a strong silencer effect resulting in luciferase expression close to background levels, while this was not observed for any of the control regions. Moreover, reporter assays of CapStarr-seq-defined silencers confirmed that the silencer activity was independent of their orientation (**Figure 3C**). To more quantitatively assess the silencer activity we cloned one of the validated silencers (*Dpp9*) into a GFP-containing reporter vector. As shown in Figure 3D, the *Dpp9* silencer almost completely repressed the GFP expression. As an additional validation, 16 CapStarr-seq-defined silencers were assessed in a retroviral reporter assay. Of these, 56% (9 out of 16) putative silencers displayed significant silencer activity in the Beko T cell line (**Supplemental Table 2**). Thus, consistency between the two independent reporter assays indicates that high-throughput assessment of silencer activity by CapStarr-seq is highly accurate.

### ***Transcription factors associated with silencers***

To assess whether putative silencers were enriched for transcription factor binding sites (TFBS) we performed motif enrichment analyses using Homer tool (Heinz et al., 2010)(**Figure 4A**). All four sets of defined silencers were enriched in TFBS commonly found at promoters such as Sp1, KLF and NFY binding sites, reflecting the enrichment in promoter elements at DHS. Strikingly, the pPGK and pSCP1 based silencers were enriched in binding sites for the RE1-Silencing Transcription factor (REST), which is a transcriptional repressor mainly involved in the repression of neural genes in non-neuronal cells (Chong et al., 1995a; Coulson, 2005; Schoenherr and Anderson, 1995). On the other hand, the pRag2-E $\alpha$  based silencers were specifically enriched in binding sites for the ETS family of lymphoid-specific factors, consistent with the tissue-restricted activity of this vector.

To explore the biological significance of REST motif-containing silencers, we mutated the REST binding site at two REST-containing silencers and tested the silencer activity by luciferase assay. In both cases, mutation of the REST site resulted in derepression of luciferase expression (**Figure 4B-C**), suggesting that the silencer activity of these elements is mediated by the REST transcription factor. Note, however, that the *Rbfox3* silencer contains two REST sites, but only the most upstream (REST#1) display strong silencer activity. This may be due to the fact that the second REST site partially overlapped with other TFBS and therefore the mutation might also affect the overall activity of the regulatory element.

To further explore the role of REST in the activity of identified silencers, we tested by the retroviral reporter assay a set of 36 CapStarr-seq defined silencers containing or not a REST binding site (**Supplemental Table 2**). As shown in figure 4D, REST-containing silencers systematically displayed silencer activity in both a T cell derived cell line (Beko) and mouse fibroblasts (NIH-3T3). However, putative silencers without REST site preferentially displayed silencer activity in the Beko cell line. These results strongly suggest that REST-containing silencers might work as ubiquitous silencers, while the other REST-deprived silencers identified by CapStarr-seq might display more tissue-restricted silencer activity.

We next explored whether silencer activity was linked to the occupancy of lymphoid TFs using ChIP-seq data for six TFs in DP thymocytes (**Supplemental Figure 4**). All TFs were enriched at active enhancers found in the pSCP1-based CapStarr-seq experiment (Top 500), but not consistent association was found between these TFs and the silencers identified with the different CapStarr-seq settings. In addition, CapStarr-seq-defined silencers were not enriched in CTCF binding (**Supplemental Figure 2B-C**), indicating that CTCF does not contribute significantly to the silencer activity detected by CapStarr-seq.

### ***Transposable elements associated with silencers***

Large portions of mammalian genomes are derived from TEs which are linked to TF binding sites (Barakat et al., 2017; Glinsky, 2015; Kunarso et al., 2010). TE elements have been associated with both enhancer (Barakat et al., 2017; Wang et al., 2018) and repressive activities (Adoue et al., 2019), but whether these elements are linked to silencer activity genome-wide have not been determined. To assess CapStarr-seq silencers for the occurrence of TE sequences, we used the RepeatMasker annotation (Kent et al., 2002). The number of TE-derived sequences in silencer regions was compared to the number detected in all DHS



regions (**Supplemental Figure 4B-C**). pSCP1-based silencers were mainly enriched in LTR while pEF1 $\alpha$ -based silencers were enriched in small RNA and simple repeats. These results indicate that certain families of TEs are overrepresented at silencers. However, different TEs are enriched depending on the library used to identify the silencers.

### **The *Sdhaf1*\_silencer regulates two genes involved in T cell function**

To start exploring the role of silencers in normal T cell development, we searched for CapStarr-seq-based silencers associated with relevant genes. We reasoned that lymphoid genes regulated by silencers might have a high expression variance across T cell populations. We isolated the top 5% of highly variable genes based on available RNA-seq from different stages of thymic and peripheral T cell differentiation (Hu et al., 2013b)(**Figure 5A**). We then retrieved the variable genes located in a window of 200kb around the CapStarr-seq-defined silencer (**Figure 5B**). We obtained 309 genes associated with 544 silencers. Of these, we identified the *Sdhaf1*\_silencer (named accordingly to the neighbour gene) associated with *Hcst* (Hematopoietic cell signal transducer, also known as DAP10) and *Nfkbid* (Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, delta), two genes involved with T cell maturation and activation. The *Hcst* gene encodes for a transmembrane signaling adaptor which forms part of the immune recognition receptor complex (Diefenbach et al., 2002; Wu et al., 1999). This receptor complex has a role in cell survival and proliferation by activation of NK and T cell responses. The *Nfkbid* gene encodes for a member of the atypical inhibitors of NF- $\kappa$ B TF and is particularly involved in the regulation of T cell activation and development of regulatory T cells (Schuster et al., 2013; Schuster et al., 2012). Strikingly, the expression of both genes is anticorrelated through T cell differentiation. In particular, *Hcst* is induced during T helper (Th) maturation, while *Nfkbid* is repressed (**Figure 5C**).

Importantly, both genes are located within the TAD containing the *Sdhaf1*\_silencer (**Figure 6A; Supplemental Figure 5**), suggesting that they could be directly regulated by this silencer. Luciferase reporter assay demonstrated a strong silencer activity for the *Sdhaf1*\_silencer (**Figure 6B**). To explore the *in vivo* function of the *Sdhaf1* silencer we deleted this element in the P5424 cell line using the CRISPR/Cas9 technology (**Figure 6C**). We analyzed the expression of all genes contained in the same TAD as the *Sdhaf1* silencer (**Figure 6D**; only genes with detectable expression are shown). Only the *Hcst* gene appeared to be significantly up-regulated in two independent *Sdhaf1*\_silencer-deleted clones. We previously showed that PMA/ionomycin stimulation of P5424 cells partially mimic early T cell differentiation (Saadi et al., 2019). Interestingly, we observed that *Nfkbid* was specifically

upregulated by the PMA/ionomycin stimulation (**Figure 6A**), consistent with its induction between the DN and DP stages (**Figure 5**). Analyses of expression in wild-type and mutated P5424 cells stimulated by PMA/ionomycin revealed a role of *Sdhaf1\_silencer* in repressing the *Nfkbid* gene (**Figure 6E**). Therefore, the *Sdhaf1\_silencer* appears to regulate both *Hcst* and *Nfkbid* genes in different stimulatory context. However, the effect on the gene expression has appears to be mild (~2 fold) in comparison with the expression changes that the target genes display during T cell differentiation. This, might be due to the presence of additional repressive elements or that the *Sdhaf1\_silencer* not being responsible for complete repression of these genes but rather modulate their transcription levels.

We next looked for genomics and epigenomics features associated with the *Sdhaf1\_silencer* (**Supplemental Figure 6A**) in DP thymocytes. The *Sdhaf1\_silencer* was not associated with active histone marks or lymphoid TFs, but was surrounded by H3K27me3 enriched regions. Interestingly, the *Sdhaf1\_silencer* was found to be associated with CTCF in different tissues, including thymus (**Supplemental Figure 6A-B**). To look for other potential regulators with explored TFBS sites overlapping the *Sdhaf1\_silencer* (**Supplemental Figure 7**). In addition to CTCF, we found two potentially relevant factors. GF11 (Growth Factor Independent 1 Transcriptional Repressor) is a nuclear zinc finger protein that functions as a transcriptional repressor and is essential for haematopoiesis and involved in Th2 differentiation pathway and T-cell receptor signalling (Anguita et al., 2017; Chiang and Ayyanathan, 2013; Ebihara and Taniuchi, 2019; Moroy and Khandanpour, 2011). ZNF263 is a C2H2 ZNF that contains 9 finger domains and a KRAB repression domain (Frietze et al., 2010), and is expressed in T cells. Strikingly, the *Sdhaf1\_silencer* contains a tandem repeat of 13 ZNF263, which is reminiscent of a previous study showing that high density of identical motifs of ZEB1 in tandem repeats can make them suitable platforms for recruitment of transcriptional repressors (Balestrieri et al., 2018).

## Methods

### Cell culture

Mouse P5424 T cells (Mombaerts et al., 1995) (kindly provided by Dr. Eugene Oltz, Washington, USA) were cultured in RPMI medium (Life technologies) supplemented with heat-inactivated 10% FBS (PAA) at 37 °C, 5% CO<sub>2</sub>. The P5424 Cells were passaged every 2-3 d and frequently tested for mycoplasma contamination.

### **Stimulation of P5424 cells**

P5424 cells were grown at density of  $3 \times 10^5$  cells/ml. Cells were treated by DMSO or PMA at 10 ng/ml (P1585, Sigma) and ionomycin at 0.5  $\mu$ g/ml (I3909, Sigma) for 4 hours in triplicates as previously described (Saadi et al., 2019).

### **Cloning of the STARR-Seq vectors**

Construction and capture of the genomic library is described previously (Vanhille et al., 2015a). The STARR-Seq mammalian screening vector (Arnold et al., 2013) has been kindly provided by Alexander Stark (Vienna, Austria). The synthetic SCP1 promoter present in the enhancer-STARR-seq vector was replaced by the ubiquitous promoters of the human EF1 $\alpha$  and PGK genes as defined in (Qin et al., 2010) and a tissue-specific promoter/enhancer pair (Rag2-E $\alpha$ ) as defined in (Wei et al., 2005) using In-Fusion homologous recombination. In the Rag2-E $\alpha$  construct the TCR $\alpha$  enhancer was cloned downstream the GFP cassette.

### **CapStarr-seq library cloning**

The principle of CapStarr-seq was explained earlier (Vanhille et al., 2015a). In-Fusion, a cloning technology developed by Clontech was used for insertion of promoters in STARR-Seq vector. The detailed step-by-step protocol is accessible on Protocol Exchange (Dao et al., 2015). The four screening vectors were linearized with AgeI-HF and Sall-HF (New England Biolabs) by 6 h digestion, followed by agarose gel electrophoresis, extracted by QIAquick gel extraction (Qiagen) and cleaned-up with Qiagen Minelute PCR purification Kit (Qiagen). After, 500 ng of amplified captured DNA was recombined by (Clontech In-Fusion HD) with 500 ng of linearized screening vectors (Qiagen) in a total of ten 10- $\mu$ l reactions (each having 50 ng of captured DNA and 50 ng of screening vector). All the recombination reactions were pooled together and purified with Agencourt AMPureXP DNA beads and then eluted in 29  $\mu$ l. Thirteen aliquots (20  $\mu$ l each) of MegaX DH10B Electrocompetent Bacteria of Life technologies were transformed with 2  $\mu$ l of DNA each, according to manufacturer's recommendation. After 1 h recovery at 37°C, the transformations were pooled together and transferred to 2 l of LB overnight for each specific promoter vector. An aliquot of each transformation was plated on LB AMP medium to estimate the number of cloned fragments. Usually, a total of 6-8 million colonies are achieved by library. Finally, plasmid libraries were extracted using Qiagen Plasmid Plus Maxi Kit (Qiagen).

### **CapStarr-seq library transfection**

Using Neon Transfection System (Life technologies) the library transfection of (5ug/1x10<sup>6</sup> cells) was performed. For each experiment, total of 50x10<sup>6</sup> cells were transfected. The P5424 cells were transfected with 1600v-20ms-1 pulse. After transfection, cells were transferred to complete growth medium and incubated for 24 h before isolation of RNA, 3 replicates (except pSCP1).

### **CapStarr-seq RNA and plasmid isolation from transfected cells**

RNA extraction was done using RNeasy miniprep kit (Qiagen) with the on-column DNaseI treatment. The PolyA RNA fraction was isolated by mMACS mRNA isolation kit (Milteny) following the manufacturer's recommendations. PolyA RNA treatment was done with Ambion turboDNase (Ambion) and purified with RNeasy Minelute kit (Qiagen). Finally, using Qubit RNA HS Kit (Life technologies) mRNA was quantified. Later, using the QIAGEN plasmid plus mini kit (Qiagen), the Plasmid DNA was extracted from 5x10<sup>6</sup> cells.

### **CapStarr-seq reverse transcription and sequencing library preparation**

cDNA first strand synthesis was performed with superscript III (Life technologies) using a reporter- RNA specific primer (5'-CAAACATCAATGTATCTTATCATG-3') and polyA RNA 0.2 to 0.3 ug of per reaction for a total of 10 reactions. After the reverse transcription, RNaseH 1 ul was added and incubated at 37°C for 1 h). The cDNA was then purified with QIAquick PCR purification kit and using Qubit ssDNA Kit (Life technologies) concentration was determined. The cDNA was amplified using the KAPA Hifi Hot Start Ready Mix in a 2-step nested PCR. In first PCR (98°C, 2min; followed by 15 cycles of 98°C for 20 s, 65°C for 20 s, 72°C for 30 s), cDNA of 5 ng per reaction was amplified using two reporter-specific primers (fw: 5'-GGGCCAGCTGTTGGGGTG\*T\*C\*C\*A\*C-3' and rw: 5'-CTTATCATGTCTGCTCGA\*A\*G\*C-3'), one of which spans splice junction of synthetic intron, in total of 10 reactions. Purification of PCR products were done on gel using QIAquick gel extraction Kit (Qiagen) followed by a clean-up with QIAquick mini elute PCR purification kit (Qiagen), to remove any residual contamination of plasmid or cDNA. Generating Ion Torrent libraries, purified PCR product was used as a template for the second PCR (5 ng/PCR, for a total of 10 PCR reactions; 98°C, 2min; followed by 10 cycles of 98°C for 20 s, 65°C for 20 s, 72°C for 30 s) with KAPA Hifi Hot Start Ready Mix and Ion Torrent library amplification primer mix (Life technologies, T\_PCR\_A: 5'-CCA TCT CAT CCC TGC GTG TC-3' and P1amp: 5'-CCA CTACGC CTC CGC TTT CCT CTC TAT G-3').

Generating the INPUT control, 10 reactions with 5 ng of reporter constructs (library) per reaction were amplified using same conditions as mentioned above, at the exception of forward primer in first PCR (fw: 5'-GGGCCAGCTGTTGGGGTG\*A\*G\*T\*A\*C-3'). To assess potential biased in library composition caused by electroporation, 10 reactions with 5 ng per reaction of the reporter constructs isolated from transfected cells were amplified as explained above. For sequencing the libraries, the sequencing indexes are added to the libraries by one simple PCR reaction using Multiplex oligos for Illumina (NEBNext). So, the non-transfected (plasmid input) and transfected libraries were single-end sequenced on the Illumina NextSeq 500 platform, and reads were mapped using standard procedures.

### **CapStarr-seq data processing**

Row CapStarr-seq data processing was described previously (Dao et al., 2015; Vanhille et al., 2015a). To quantify the coverage of each captured region, we intersected the elongated mapped reads (that is, fragments) with the DHS using BedTools (v2.28.0) ((Quinlan and Hall, 2010)) and calculated FPKM values (Fragment Per Kilobase per Million mapped reads). All conditions and replicates were processed separately. To filter out low represented regions, all regions with a FPKM lower than 10 in the input sample were excluded, thereby removing 13,687 regions for pEF1 $\alpha$ , 14,349 for pPGK, 15,283 for pRag2-E $\alpha$  and 12,096 for pSCP1. A total of 14,805, 14,143, 13,209, 16,396 regions were then retained for pEF1 $\alpha$ , pPGK, pRag2-E $\alpha$  and pSCP1, respectively, for further analyses. We pooled all replicates and computed the FPKM mean values for all subsequent analyses. The enrichment of CapStarr-seq samples over input was computed by calculating the log<sub>2</sub> ratio of FPKM and centered on 0 for all conditions (denoted CapStarr-seq signal; Supplementary figure 1B; Supplementary Table 1 CapStarr-seq data). To assess reproducibility, we generated a correlogram using ggplot2 and corrplot package (<https://cran.r-project.org/web/packages/corrplot>) with a Spearman non-parametric test. The log<sub>2</sub> (FC) was used to determine the silencer activity of each DHS. A search for between conditions was performed. Venn diagram comparing the shared silencers was made using the online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn>). All the regions were annotated with the two nearby genes using GREAT. To visualize the CapStarr-seq signal per individual cloned fragments we generated a bed file with a color code proportional to the activity.

### **Definition of silencers**

Putative silencers were defined as DHS regions displaying a CapStarr-seq signal ( $\text{Log}_2(\text{FC})$ ) lower than or equal to -1 or belonging to the bottom 500 signal in any of the CapStarr-seq experiments. In order to compare putative silencers, other groups were created based on the  $\text{log}_2(\text{FC})$ . We clustered other DHS which have  $\text{log}_2(\text{FC})$  included between  $]-1;0]$  and DHS with  $\text{log}_2(\text{FC}) > 0$ . Based on the rank, we selected the inactive regions (500 regions around the  $\text{log}_2(\text{FC})$  median) and the top 500 regions with the highest  $\text{log}_2(\text{FC})$ .

### **Epigenomic analyses**

ChIP-seq datasets from P5424 cells and MNase-seq, DNaseI-seq and ChIP-seq datasets from mouse DP thymocytes were downloaded from Gene Expression Omnibus (GEO) database as detailed in (Supplementary Table 3). ChIP-seq data were processed in the same way as in (Vanhille et al., 2015a). For average profiles, Wiggle files were converted to BigWiggle files using wig to BigWig (Kent et al., 2010). Average profiles were generated with deepTools (Ramirez et al., 2014) using MNase-seq, DNaseI-seq and ChIP-seq signal from bigwig files around DHS center ( $\pm 5$  kb for histone modifications,  $\pm 1$  kb for TFs and chromatin features).

### **RNA-seq data**

Public RNA-seq data from P5424 cells (Saadi et al., 2019) and mouse T cell differentiation (Hu et al., 2013b) were downloaded from GEO database (accession numbers GSE120655 and GSE48138, respectively). The raw RNA-seq data was processed as described in (Saadi et al., 2019).

### **Motif research analysis.**

The HOMER (Heinz et al., 2010) software was used to perform research motif analysis with the option “findMotifsGenome.pl input.file.bed mm9 output.file.bed  $-\text{len } 6,8,10,12,15$   $-\text{size given}$   $-\text{bg dhs}$ ”. We choose as background all the DHS sites in order to isolate the silencer’s specific binding sites. Then we used the known motifs output to create a heatmap using Matplotlib and Seaborn libraries (<https://ieeexplore.ieee.org/document/4160265>)

### **Genomic enrichment of candidates regions**

Genomic distribution was analyzed using a new homemade program : OLOGRAM, included in pygtftk toolkit (Lopez et al., 2019). OLOGRAM, performs overlap statistics between sets

of genomic regions described in BEDs or GTF. It uses Monte Carlo simulation, taking into account both the distributions of region and inter-region lengths, to fit a negative binomial model of the total overlap length. Exclusion of user-defined genomic areas during the shuffling is supported in order to compare with a specific set of sequence. We decided to exclude all the regions which are not DHS to compare our candidates with the DHS distribution. Mm9 GTF files from Ensembl was provided to assess the genomic distribution of our candidates. We plot with ggplot2 (<https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.147>) the  $\log_{10}(\text{p-value})$  with a factor  $\pm 1$  depending on whether it is enriched or depleted to get the enrichment score.

### **Repetitive element analysis**

Repetitive elements were retrieved from the UCSC repeat masker files for mm9. A hypergeometric test was computed from the overlap between silencers candidates and each class of repetitive elements. We then plotted the  $-\log_{10}(\text{p-value})$  using ggplot2 with a factor  $\pm 1$  depending on whether there was enrichment or depletion.

### **Association of silencers with variable genes**

RefSeq genes in a window of 200kb around the all silencers candidates ( $\log_2(\text{FC}) \leq -1$  or bottom 500) were retrieved using bedtools. In the different conditions were annotated, using mm9 file from UCSC. To retrieve the top 5% most variable genes, the variance through T cell differentiation was computed for each gene using merged replicates of the RNA-seq dataset from (Hu et al., 2013b).

### **Luciferase reporter assays**

Silencer candidates or control regions were amplified from mouse genomic DNA or synthesized *in vitro* and cloned downstream of the luciferase gene at the BamHI site in the pGL3-Promoter vector (Promega) (**Supplementary Table 4**) and verified by Sanger sequencing. A total of  $1 \times 10^6$  P5424 cells were cotransfected with 1  $\mu\text{g}$  of the tested construct and 200 ng of Renilla vector using the Neon Transfection System (Thermo Fisher Scientific). Electroporation conditions for P5424 cells were maintained at 1600v-20ms-1 pulse. After Twenty-four hours of transfection, luciferase activity was measured using the Dual-Luciferase Reporter Assay kit (Promega) on a TriStar LB-941 Reader. For all measurements, firefly luciferase values were first normalized to Renilla luciferase values (controlling for transfection efficiency and cell number). Data are represented as the fold

decrease in relative luciferase signal over the pGL3-Promoter vector (pSV40). All experiments were performed in triplicates.

### **Site-Specific Mutagenesis**

Mutagenesis was performed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs). For the *Rbfox3* and *Dpp9* silencers, the REST motifs were mutated. The mutagenesis primers were designed using NEBaseChanger tool and are shown in Supplementary Table 5.

### **CRISPR-Cas9 genome editing**

For the CRISPR–Cas9 experiments of targeted silencer regions, two gRNAs were designed for each end of targeted region using CRISPRdirect tool (Naito et al., 2015). The designed gRNAs were cloned into a gRNA cloning vector (Addgene, 41824) as described previously (Mali et al., 2013). One million cells were transfected with 1 µg of the each gRNA and 1 µg of hCas9 vector (Addgene, 41815) using Neon Transfection System (Thermo Fisher Scientific). After three days of transfection, the transfected cells were plated in 96-well plates at limiting dilution (0.5 cells per 100 µl per well) for the clonal expansion. Individual cell clones were screened for homologous allele deletion after 10-14 d, by direct PCR using Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific) according to manufacturer's protocol. Allowing the detection of knockout or wild-type alleles, forward and the reverse primers were designed bracketing the targeted regions. The clones were considered to have undergone homologous allele deletion if they had no wild-type band and at least one deletion band of expected size. All the gRNAs and primers are listed in Supplementary Table 5.

### **Gene expression analyses**

Total RNA from P5424 from wt and deleted clones in unstimulated or PMA/Ionomycin treated conditions was extracted using Rneasy mini kit (Qiagen). Three micrograms of RNA was treated with DNase I (Ambion) and was quickly reverse transcribed into cDNA using Superscript VILO Master Mix (Thermo Fisher Scientific). The Real-time PCR was performed using Power SYBR Master Mix (Thermo Fisher Scientific) on a QuantStudio 6 Flex (life technology) instrument. Primer sequences are listed in Supplementary Table 5. Expression of the gene was normalized to that of Rlp32. Relative expression was calculated by the  $\Delta C_T$  method, and all the data shown are reported as fold change over the control. For each of the cell clone, from the three independent RNA/cDNA preparations, the Student's t test was performed (unpaired, two-tailed, 95% confidence interval). Data are represented with s.d. For



the conventional RT-PCR, one-twentieth of the synthesized cDNA was used as template for one reaction.

## Legends

### Figure 1. CapStarr-seq for silencer assessment.

(A) Schematic of the CapSTARR-seq strategy to assess the silencer activity in the P5424 mouse T cell line.

(B) Correlogram of CapStarr-seq signal between all conditions and replicates. The Spearman's correlation coefficient is displayed. Samples are clustered using Hierarchical Clustering.

(C) Smoothed data using Generalized Additive Model (GAM) to visualize the correlation between CapStarr-seq signal for each DHS and expression of the two closest genes in the P5424 cell line. In X-axis regions are sorted ascending by CapStarr-seq signal ( $\text{Log}_2(\text{FC})$ ). RNA-seq data of associated genes are represented in Y-axis using TPM value.

(D) Violin plot comparing the expression in the P5424 cell line of genes associated with DHS subgroups based on the indicated categories of CapStarr-seq signal ( $\text{log}_2(\text{FC})$ ). Statistical analysis was performed using Wilcoxon test, P-values are displayed.

(E) Venn diagrams of silencer candidates identified with the indicated promoter-based CapStarr-seq strategy. The threshold used was based either on the DHS with the bottom lower CapStarr-seq signal (left panel) or a  $\text{log}_2(\text{FC})$  lower or equal than -1.

(F) Example of a putative silencer identified with the 4 promoter-based CapStarr-seq strategies. The signal for each promoter-based CapStarr-seq and the corresponding Input are showed.

(G) Bar plot showing the percentage of shared silencer candidates (bottom 500) between each pair of CapStarr-seq conditions.

(H) Genomic distribution of silencer candidates (bottom 500) compare to the whole set of DHS. Bar plots represent the  $-\log_{10}(\text{p-value})$  of negative binomial test. If the silencers are depleted in a particular feature compared to the DHS region, we multiplied by a factor 1.

### Figure 2. Epigenetic profiles of silencers

(A-B) Average profiles of H3K27ac, H3K4me3 in the mouse P5424 cell line (A) and H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9me3 in the mouse DP thymocytes (B) for captured regions identified as putative silencers (bottom 500, red), inactive regions (Middle 500, black) and putative enhancers (Top 500, green).

### Figure 3. Validation of the CapStarr-seq approach.

(A) Luciferase silencer assays in P5424 cells of DHSs defined as putative silencers by CapStarr-seq (green bars) or with CapStarr-seq signal ( $\log_2(\text{FC})$ ) close to zero (black plots). The promoter-based CapStarr-seq where the silencer was identified is indicated in the bottom panel. Data represent the normalized fold change over the vector control. Error bars show s.d. from three independent transfections ( $***P < 0.001$ ,  $**P < 0.01$ ,  $*P < 0.1$ ; two-sided Student's  $t$  test). *Rbfox3* and *Dpp9* silencers are enriched in REST motif indicated by an arrow.

(B) Comparison of luciferase activity between silencer candidates and control regions. The two-sided Student's  $t$  test is shown (C) Assessment of orientation dependent silencer activity for a subset of identified silencers.

(D) FACS analysis for GFP expression using *Dpp9*\_silencer element in a GFP reporter construct.

### Figure 4. TFBS associated with silencers.

(A) Heatmap showing the enrichment ( $-\log_{10}(\text{p-value})$ ) of known motifs obtained with HOMER software for each set of silencers (bottom 500).

(B, C) The UCSC genomic track of Mouse NCBI37/mm9 around the *Rbfox3* (A, top panel) and *Dpp9* (B, top panel) DHS sites (black) indicating the CapStarr-seq signal obtained for each captured region (the color scale is indicated) and the Jasp motifs. The REST motifs are indicated. The effect of REST mutation was assessed by Luciferase assay (A-B, bottom panels). Data represent the normalized fold change over the wild type vector. Error bars show s.d. from three independent transfections ( $***P < 0.001$ ,  $**P < 0.01$ ,  $*P < 0.1$ ; two-sided Student's  $t$  test). D: Experimental validation of silencer activity using a retroviral system. A total of 36 and 16 regions were tested in NIH-3T3 and Beko cell lines, respectively. Silencer effect was determined by significant decrease of a reporter marker assessed by FACS.

### Figure 5. Identification of *Sdhaf1* silencer regulating *Hcst* and *Nfkbid* during T-cell differentiation

(A) Variance of genes ranked by the expression variance during mouse T-cell differentiation using RNA-seq data from (Hu et al., 2013b). The top 5% variable genes are highlighted in blue. *Nfkbid* and *Hcst* genes associated with the *Sdhaf1*\_silencer are indicated.

(B) All genes associated with silencers candidates (880) in a window of +/- 100kB were retrieved. The overlap between silencer's associated genes and top 5% variable genes is

shown as well as the number of corresponding silencers. The genes associated with the *Sdhaf1*\_silencer are indicated. (C) Log<sub>2</sub> of the relative expression of *Nfkbid* and *Hcst* genes during T-cell differentiation. Level of expression were normalized regarding to the DN\_01 sample.

**Figure 6. *Sdhaf1*\_silencer functions as a *bona fide* silencer.**

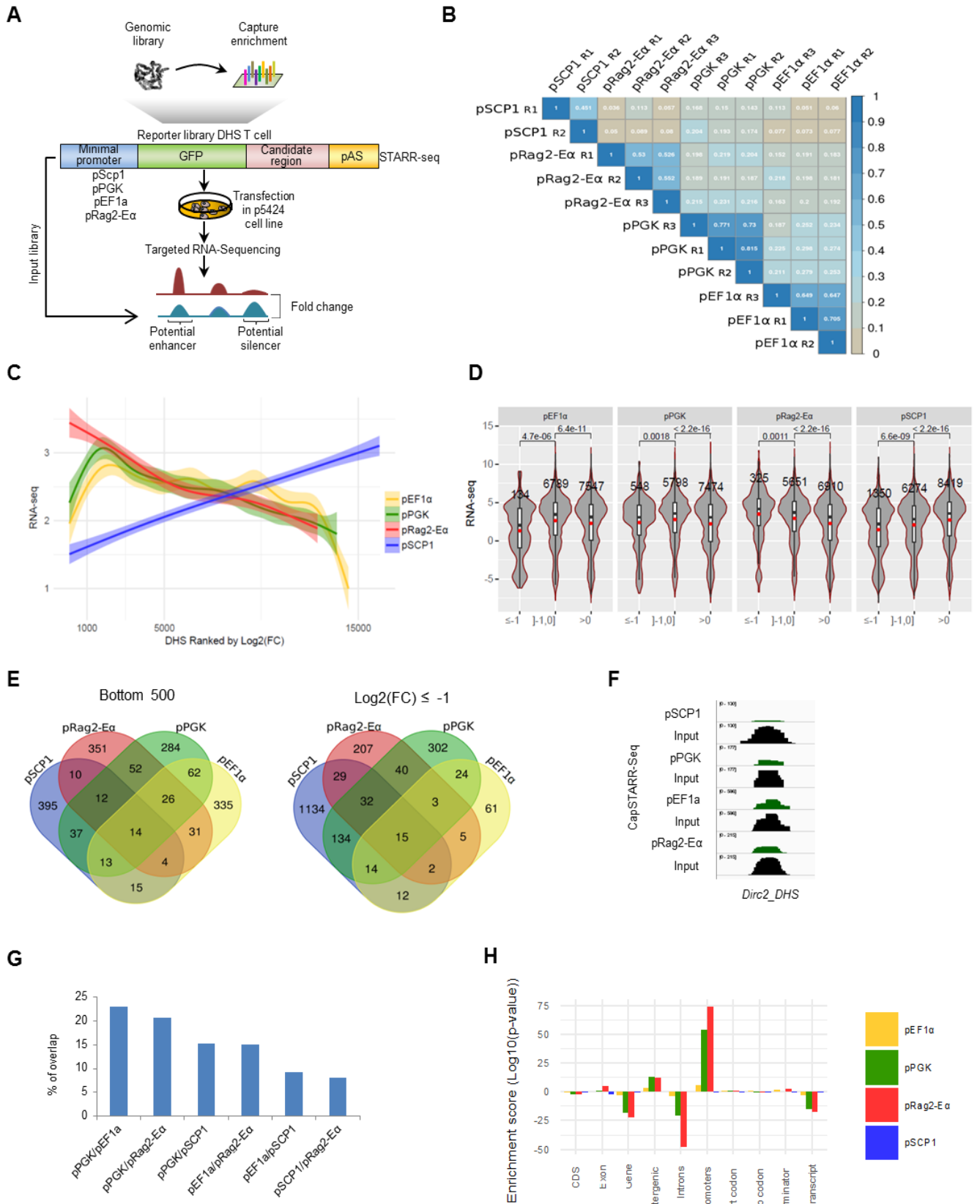
(A) Genomic tracks for Hi-C in mouse ES cells and RNA-seq and ChIP-seq from P5424 cells around the *Sdhaf1* locus.

(B) Luciferase reporter assays testing the silencer activity of the *Sdhaf1*\_silencer in both orientations.

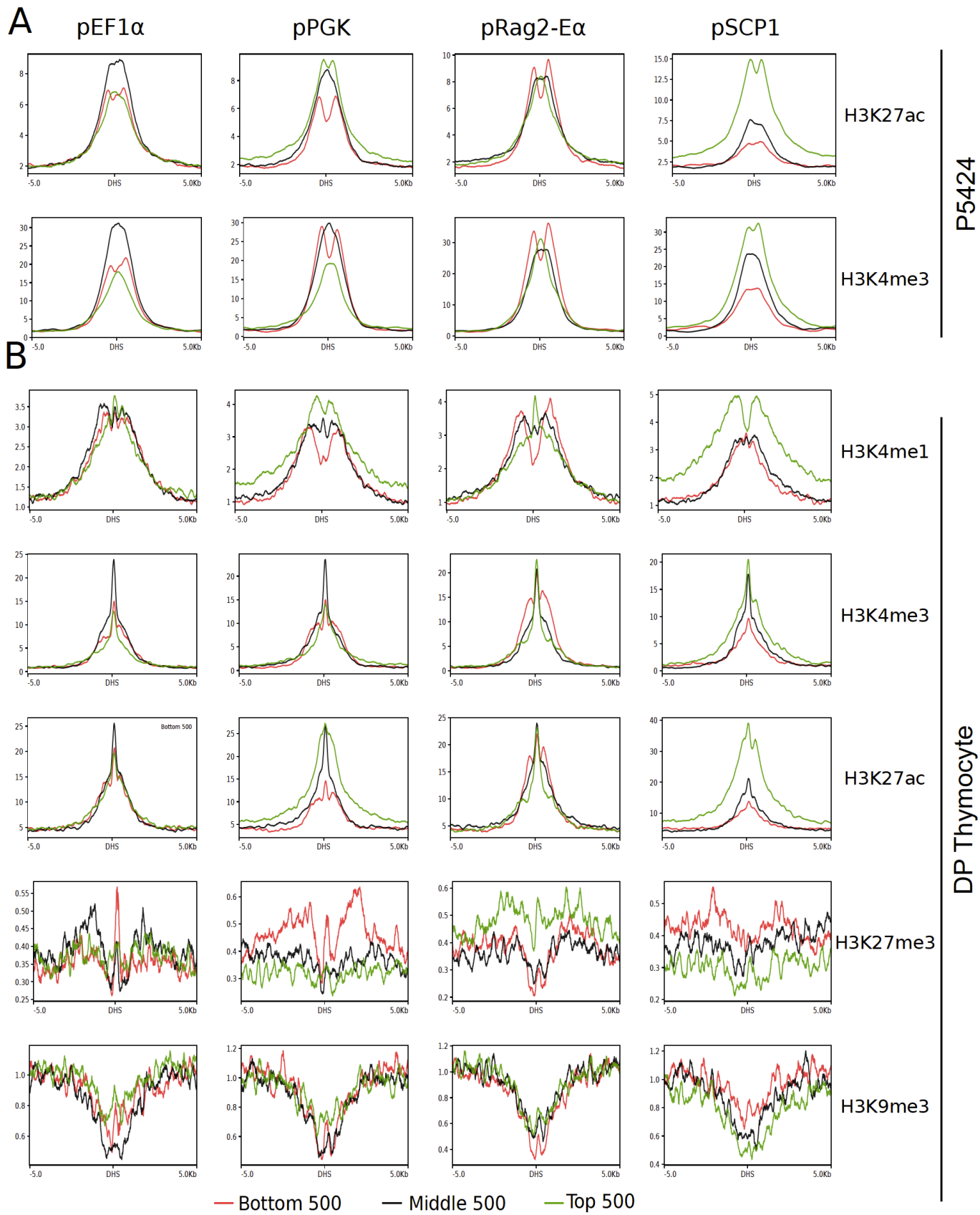
(C) *Sdhaf1*\_silencer knockout by CRISPR–Cas9 system. The homozygous and heterozygous genomic deletion of the *Sdhaf1* silencer is detected by PCR using primers P1 and P2. Details on the gRNA sequences, PCR primers and expected PCR fragments are provided in (Supplementary Table 5).

(D) qPCR analysis of gene expression of genes around the *Sdhaf1* locus in wild-type, Δsil S26 and Δsil S70 *p5424* clones.

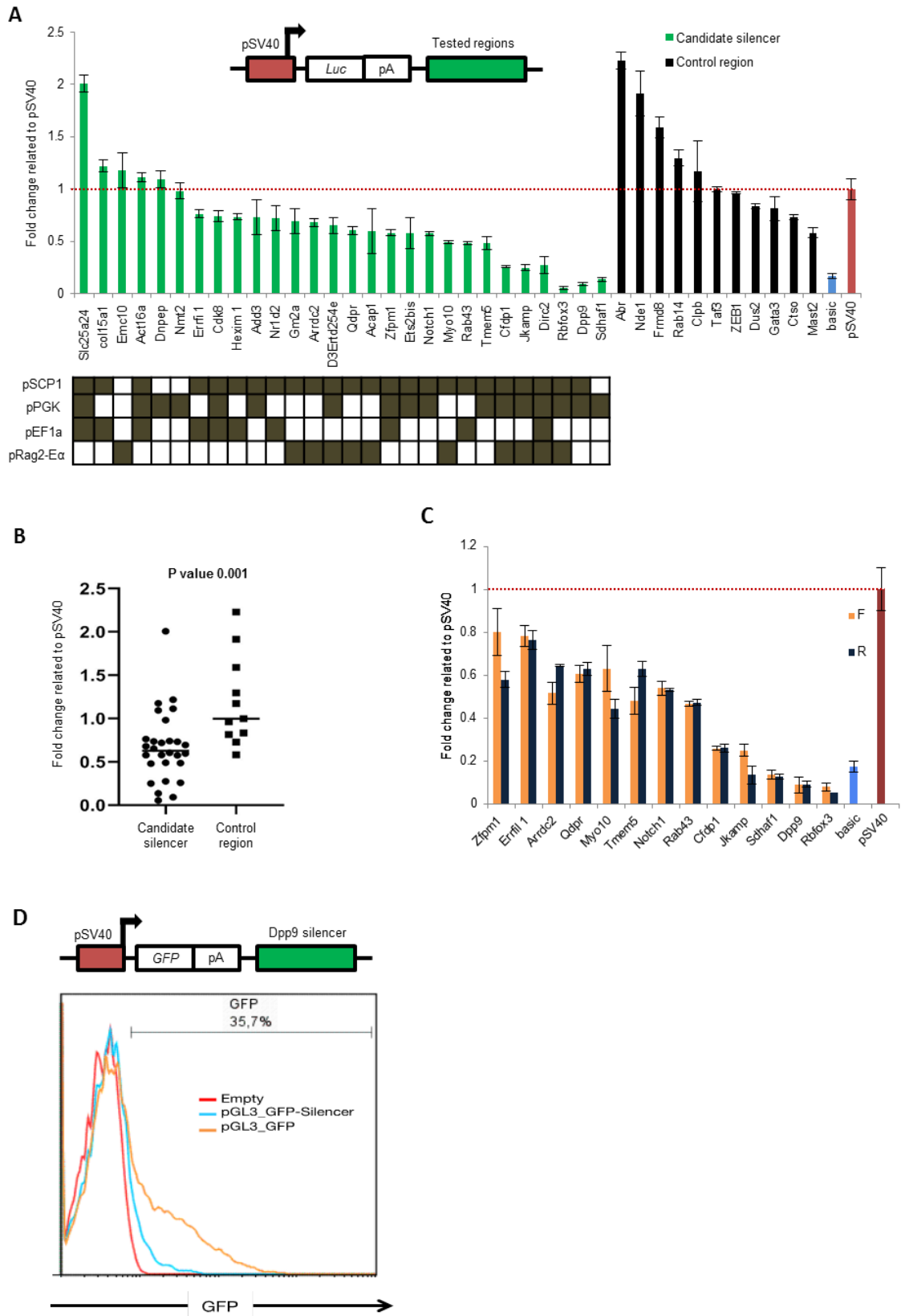
(E) qPCR analysis of gene expression of *Nfkbid* in wild-type, Δsil S26 and Δsil S70 *p5424* clones stimulated or not with PMA and Ionomycin. Error bars, s.d.: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.1$ , two-sided Student's *t* test.



**Figure 1**



**Figure 2**



**Figure 3**

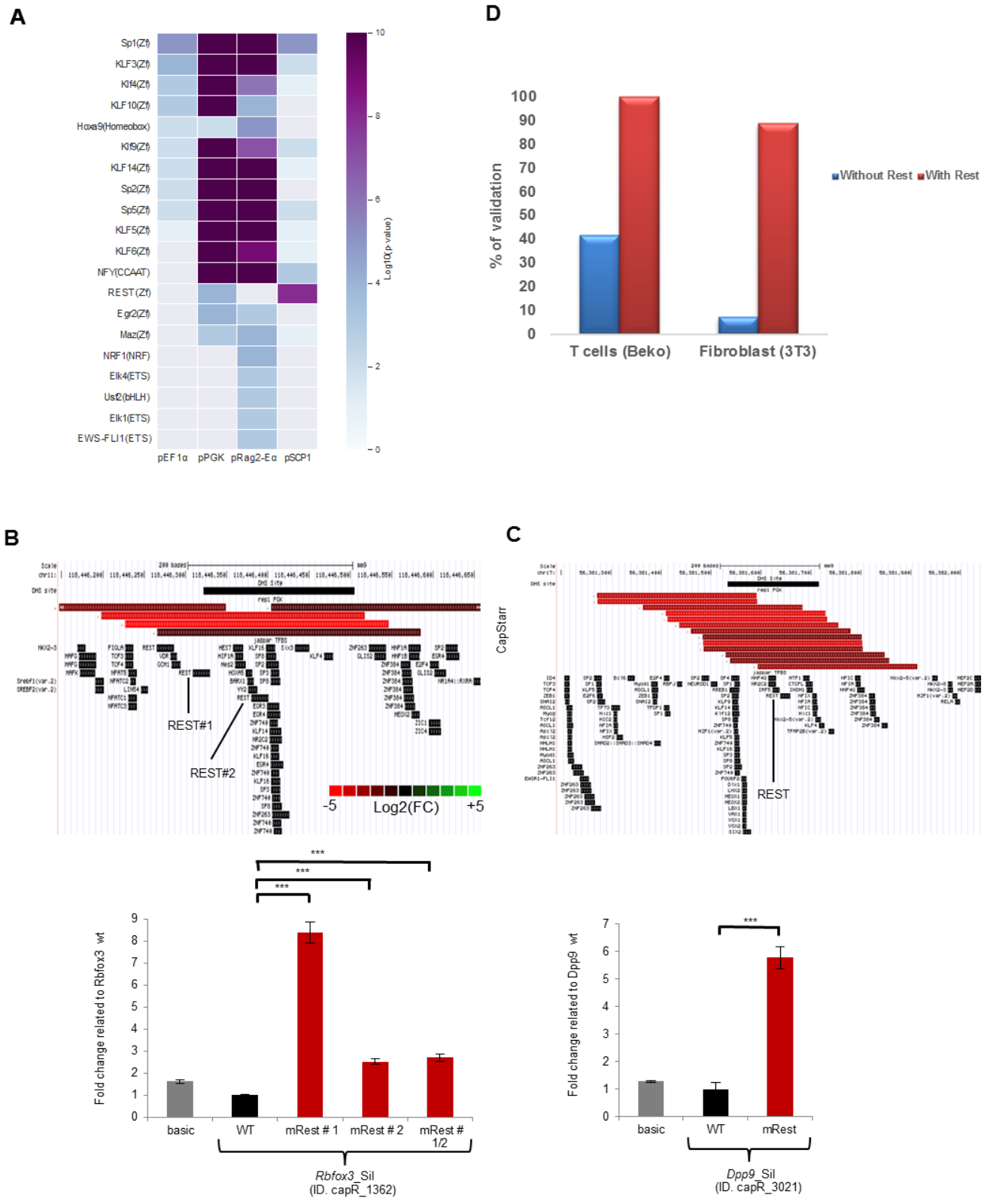
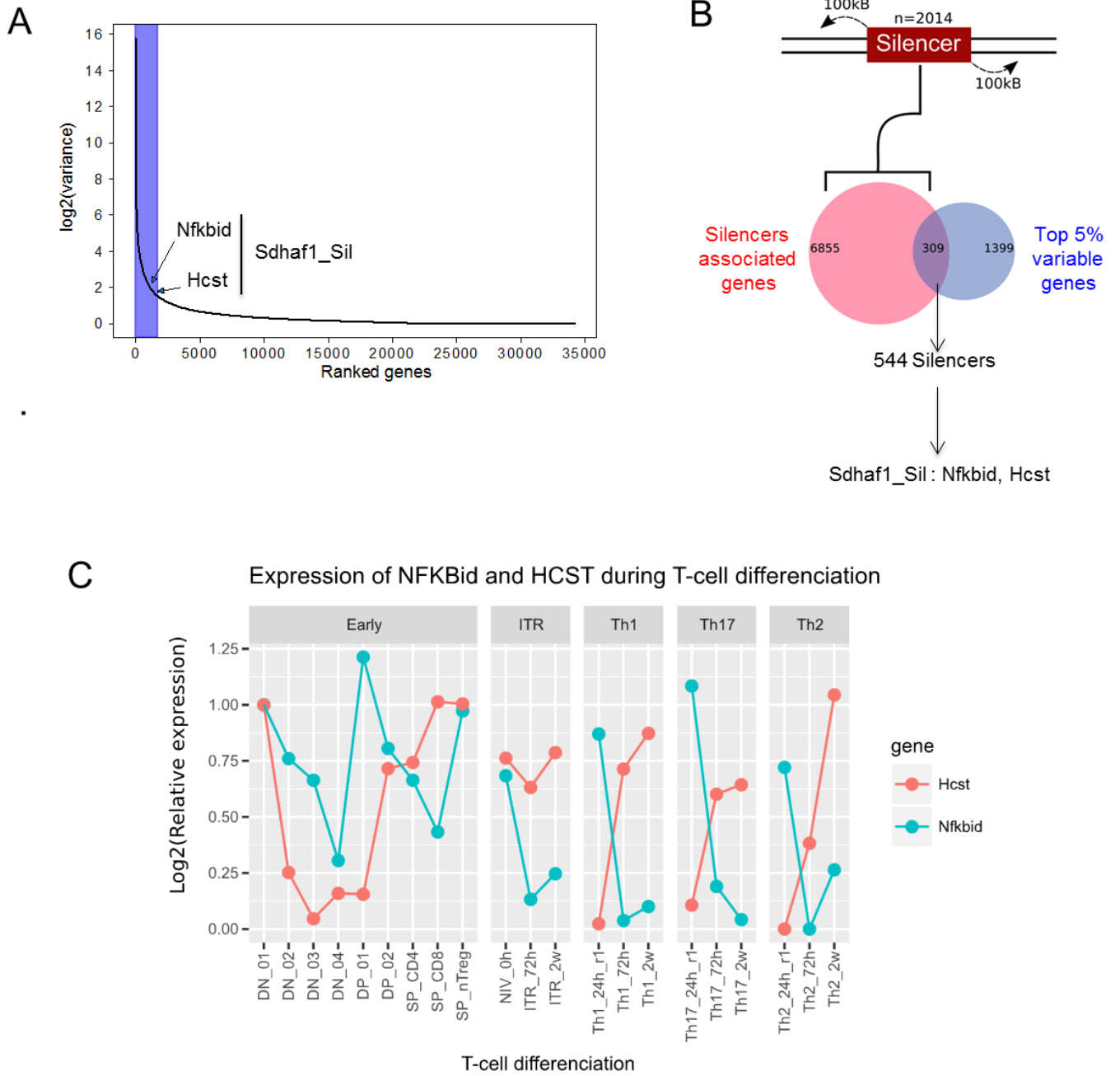
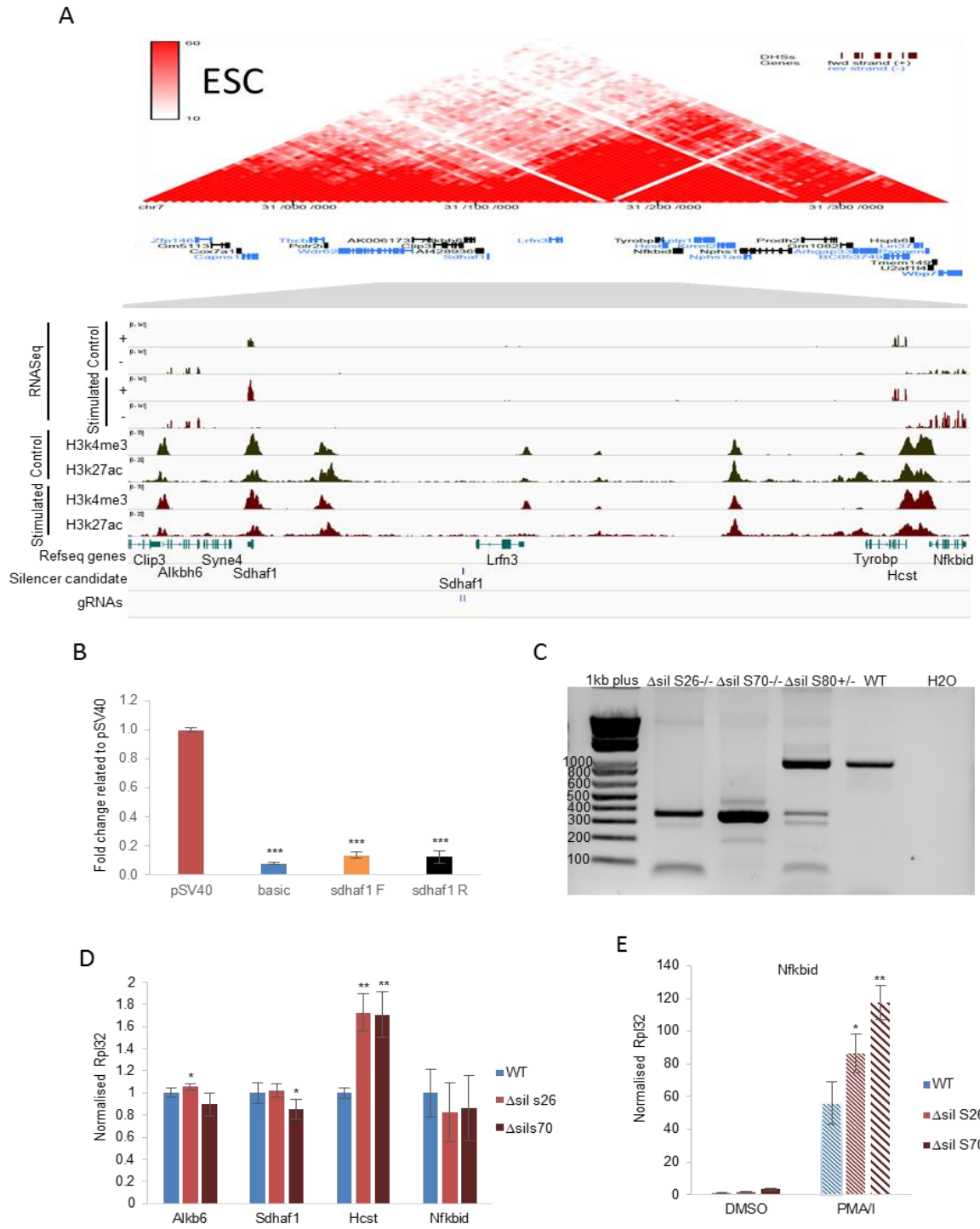


Figure 4

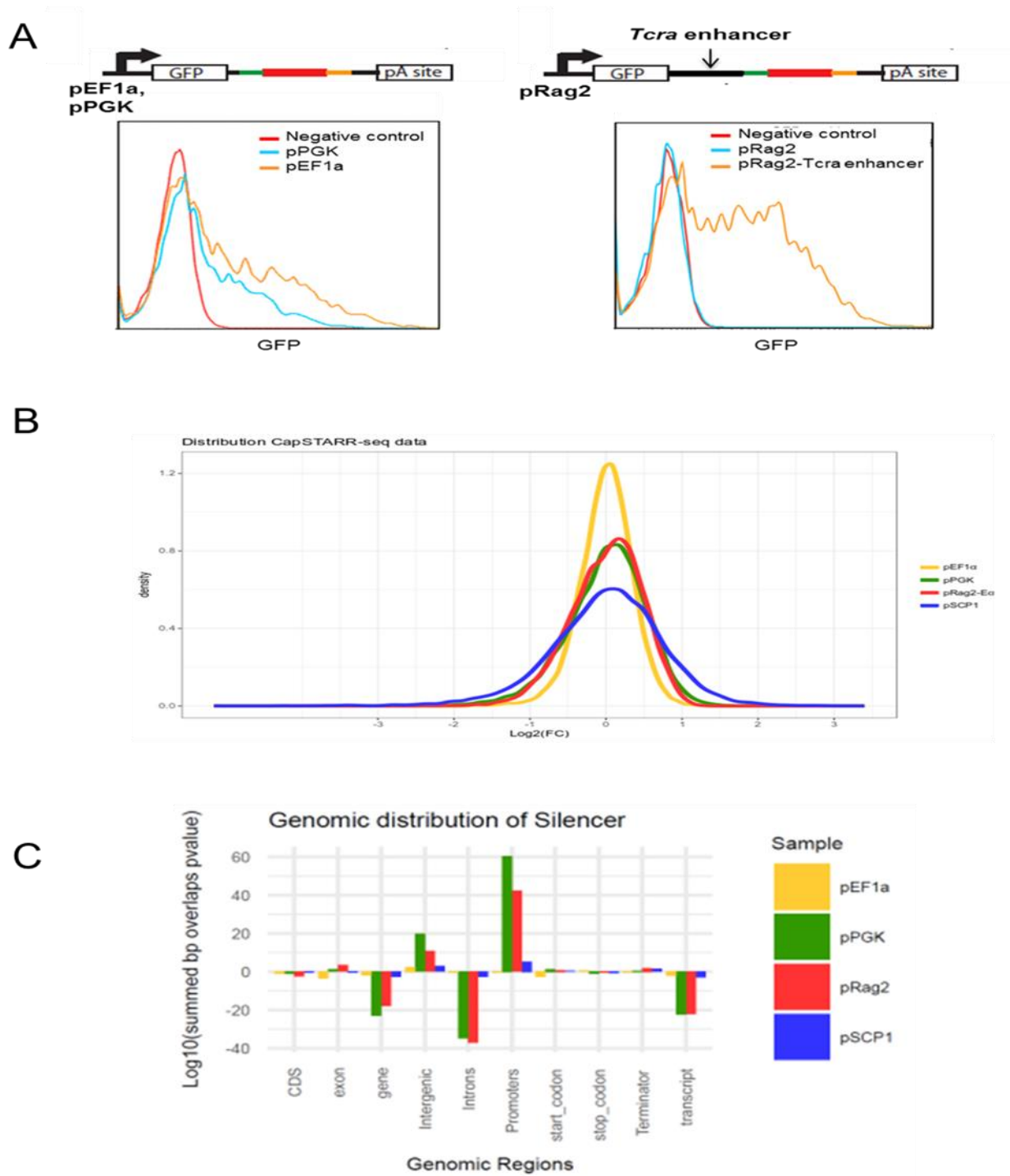




**Figure 5**

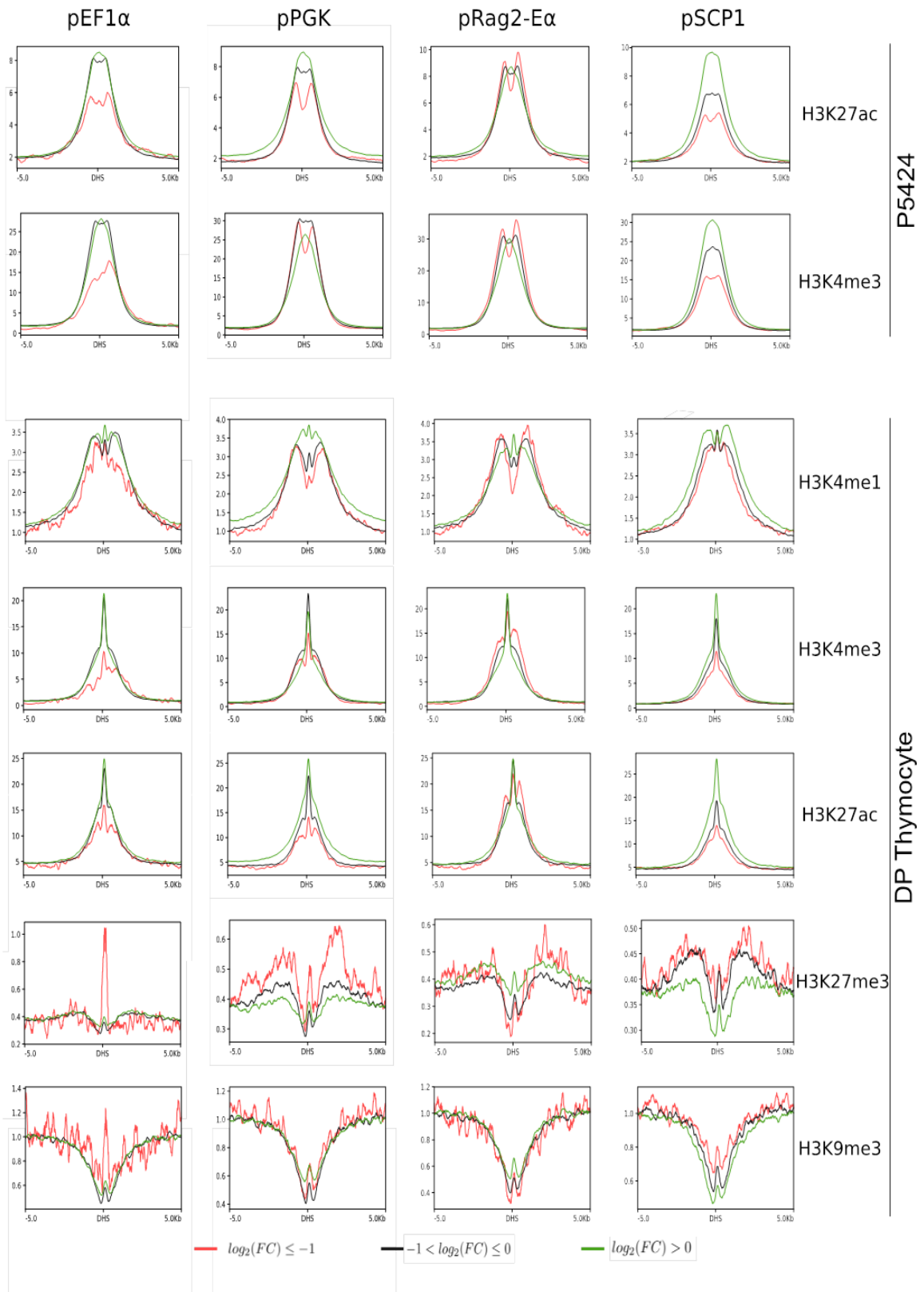


**Figure 6**

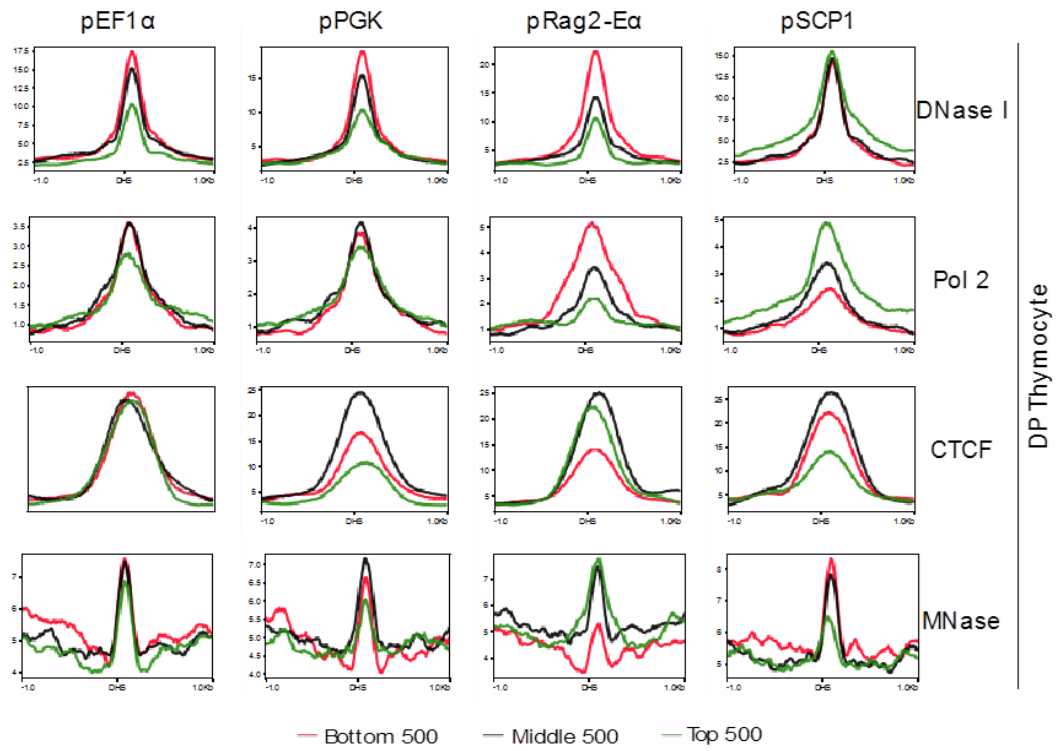


**Supplementary figure 1: (A)** Validation of the CapStarr-seq vectors by FACS analysis of GFP expression **(B)** Distribution of normalized CapStarr-seq data. **(C)** Genomic enrichment of captured region with a  $\log_2(\text{FC}) \leq -1$  (putative silencers).

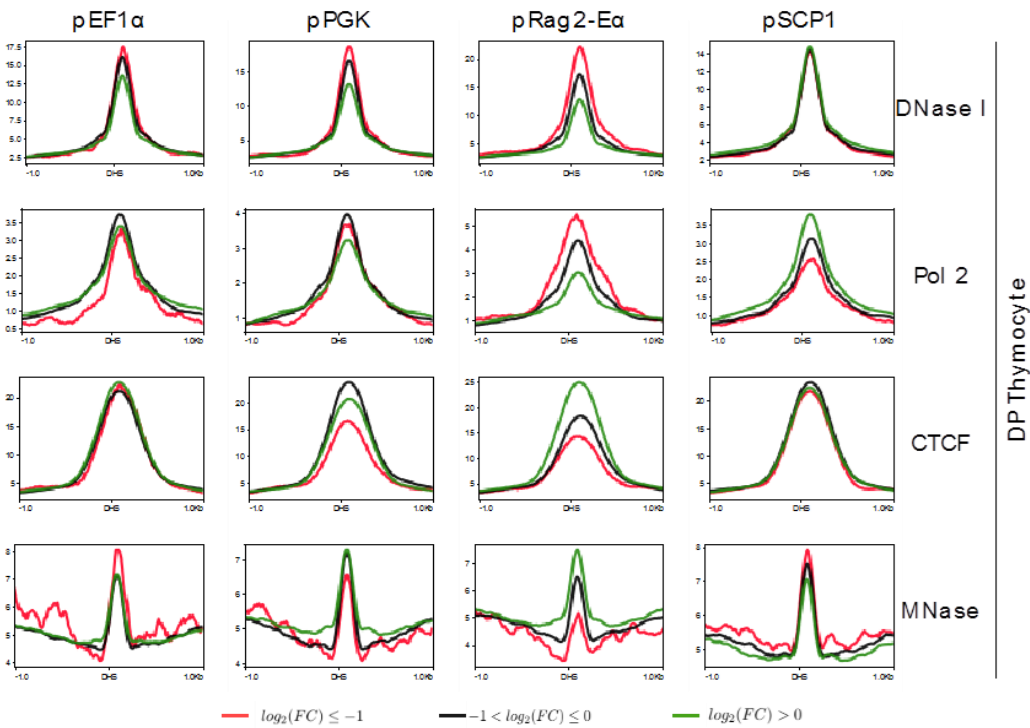
A



**B**



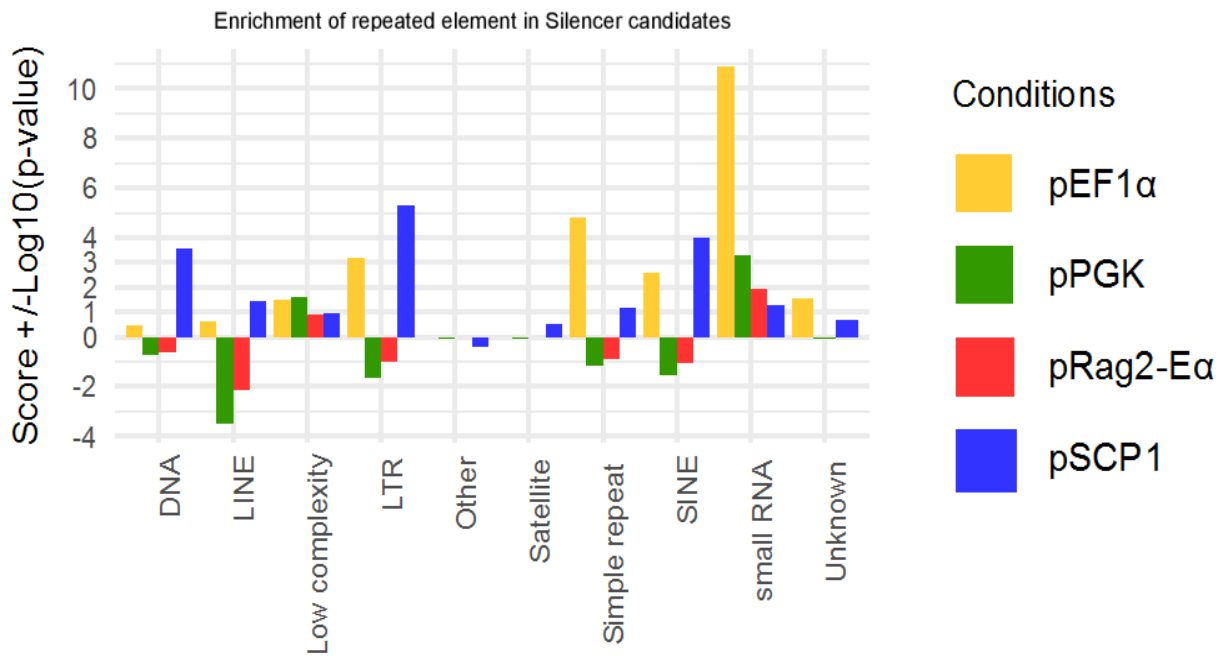
**C**



**Supplementary figure 2. :** (A) Average ChIP-seq profiles of different histone marks based on CapStarr-seq signal  $\log_2(FC)$ . (B-C) Average profiles of ChIP-seq and chromatin features based on CapStarr-seq signal rank (B) or  $\log_2(FC)$  (C).

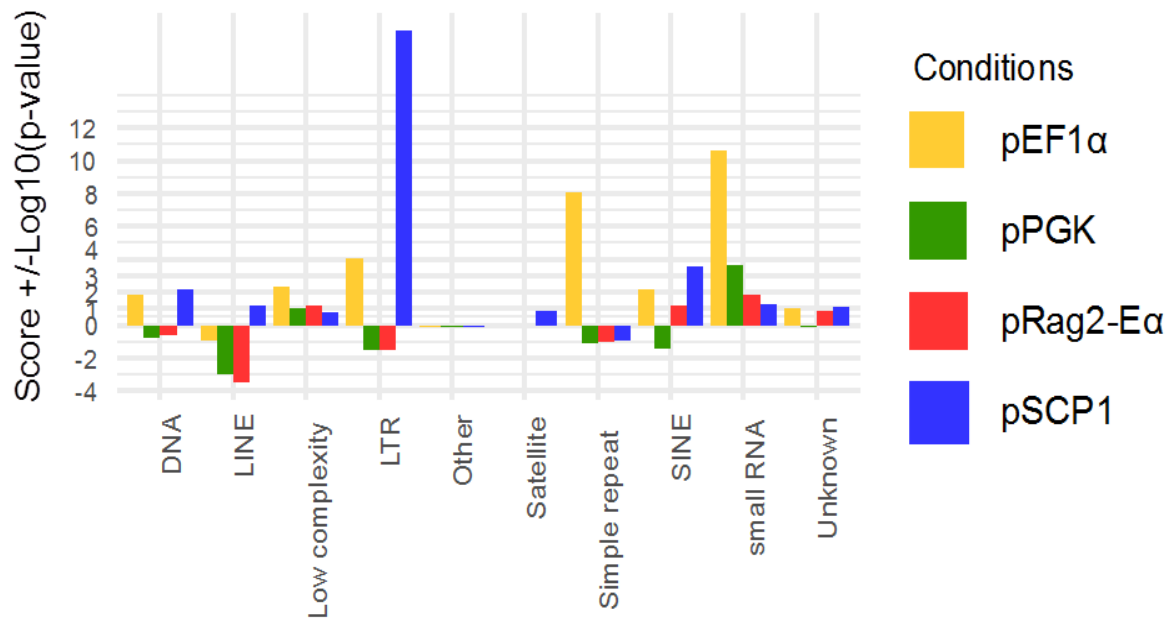
**A**

$\log_2(\text{FC}) \leq -1$



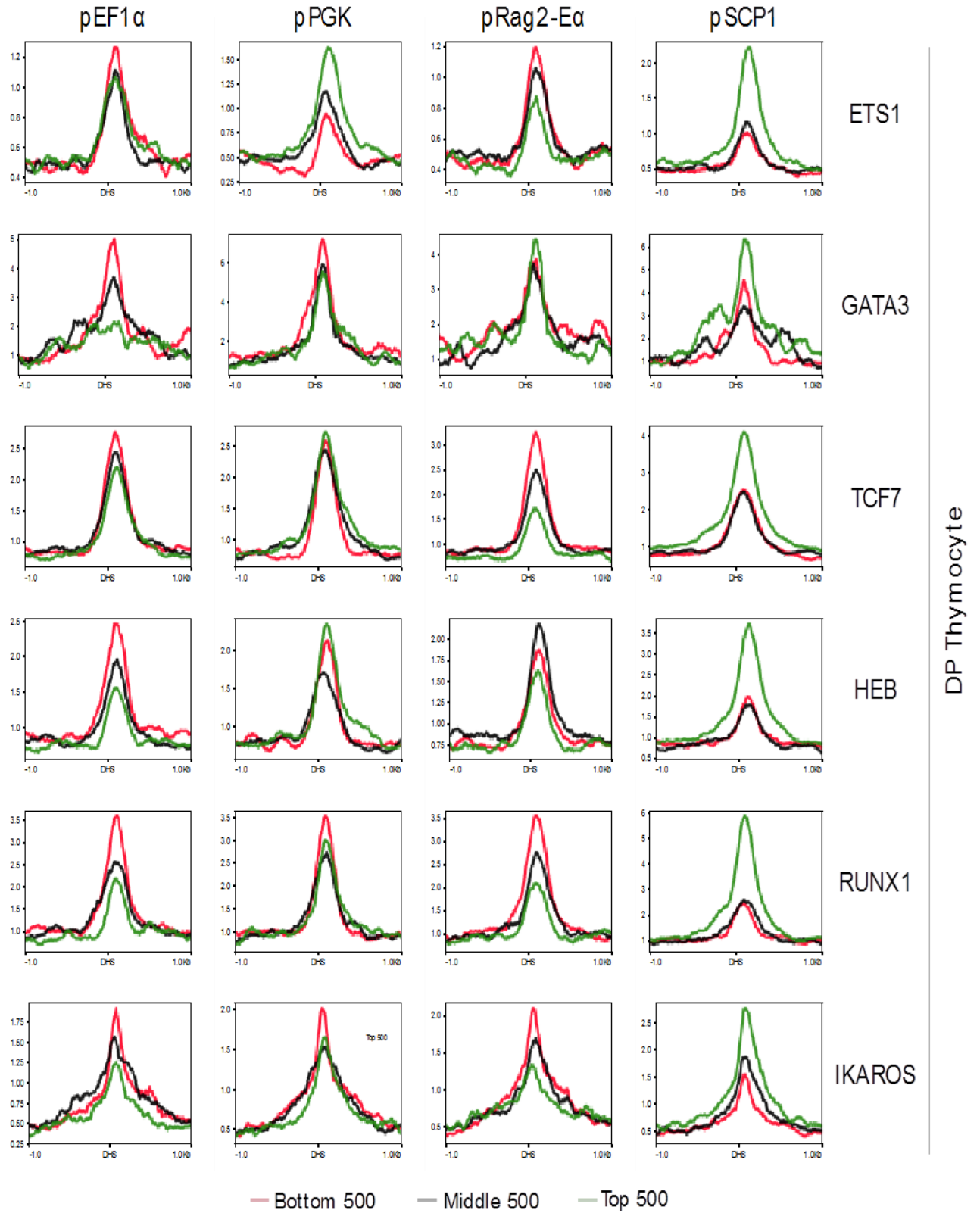
**B**

Bottom 500



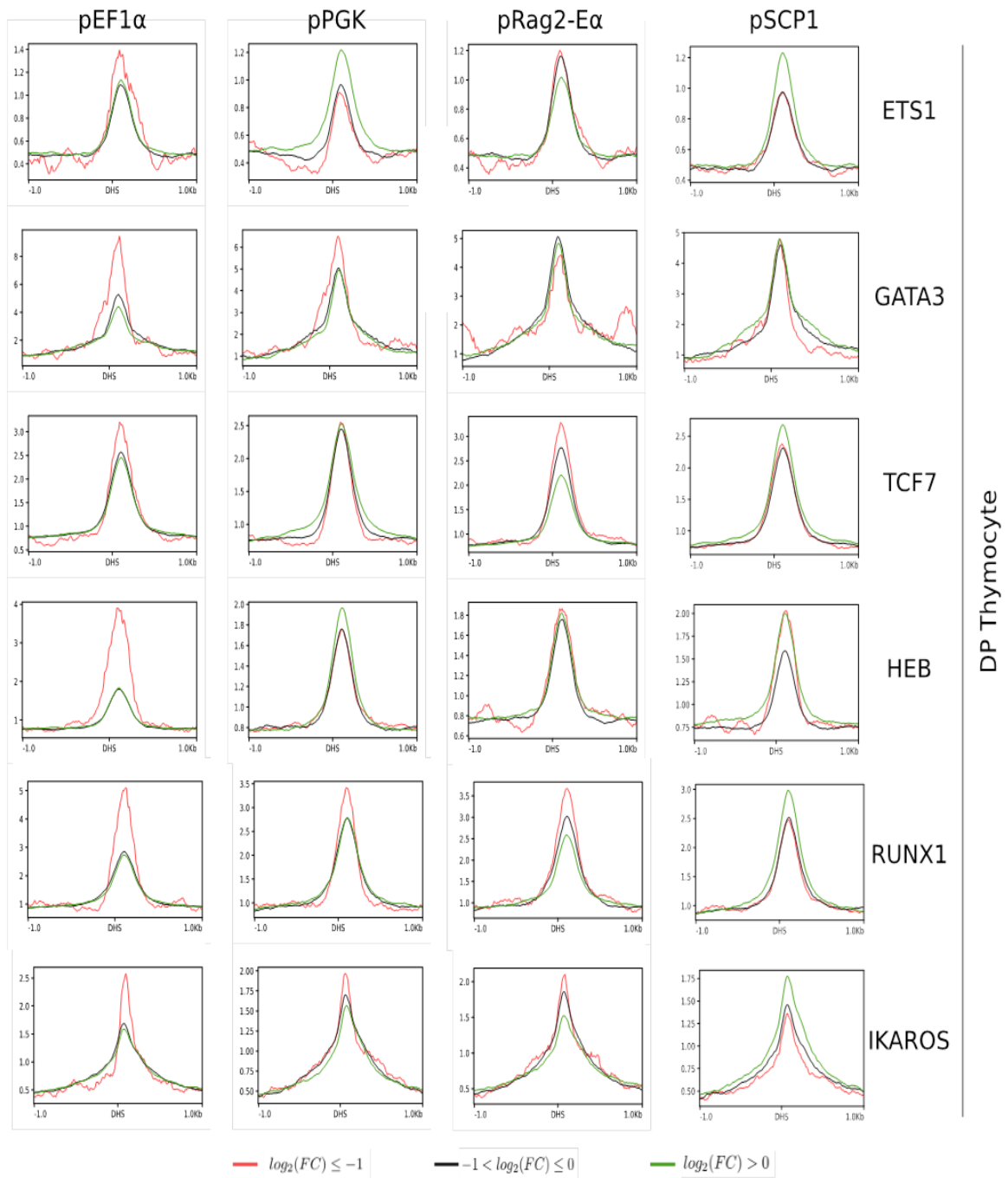


C

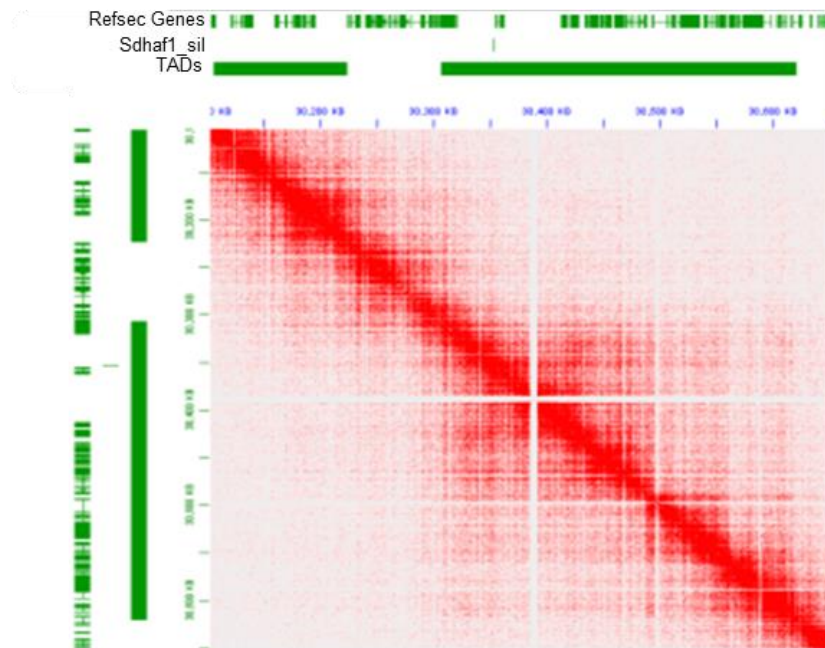
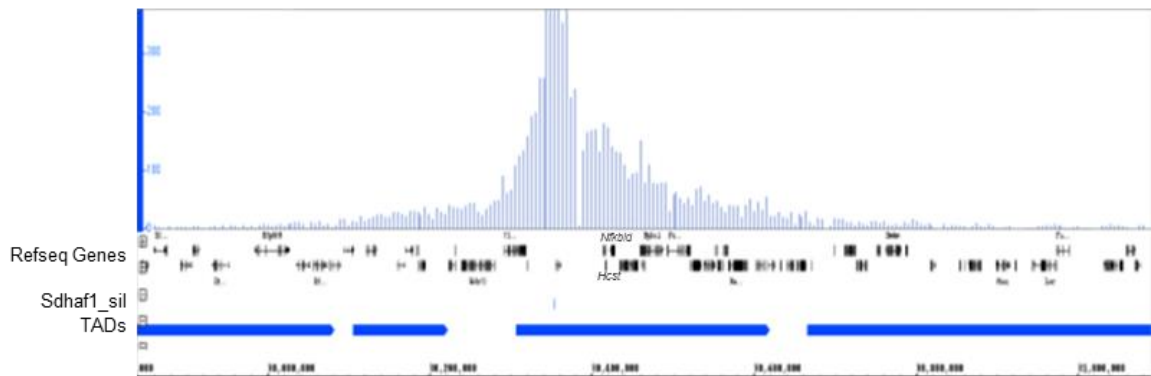




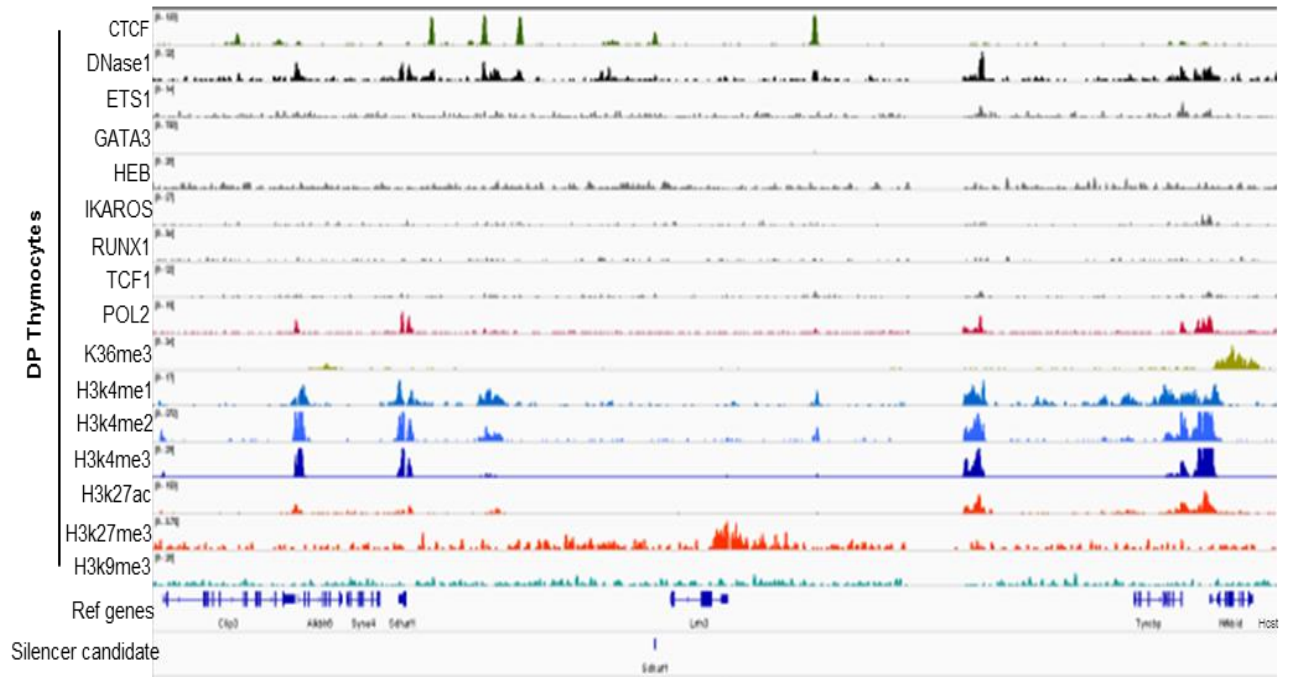
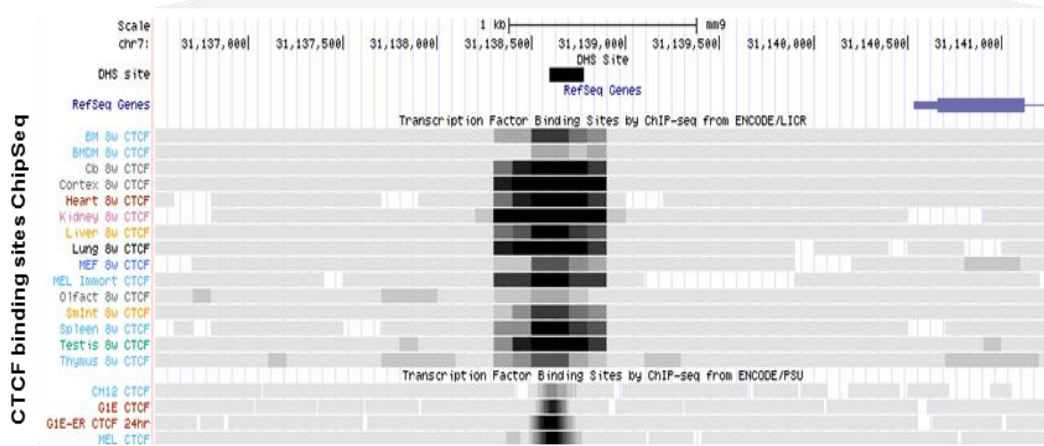
D



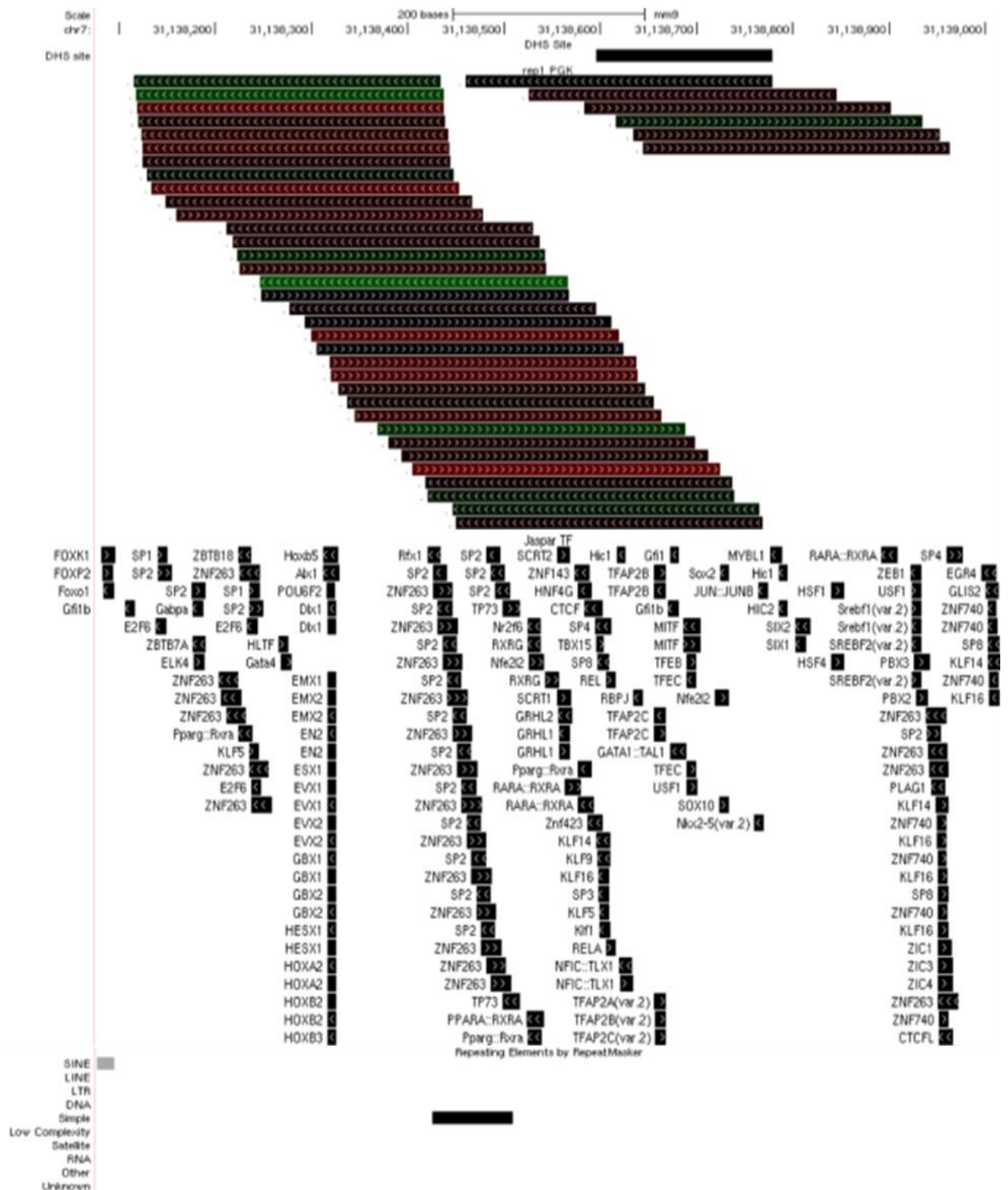
**Supplementary figure 4. (A-B).** Enrichment of repetitive elements at putative silencers based on CapStarr-seq signal rank (A) or  $\log_2(FC)$  (B). (C-D) Average ChIP-seq profiles of lymphoid transcription factors in DP thymocytes based on CapStarr-seq signal rank (C) or  $\log_2(FC)$  (D).

**A****B**

**Supplementary figure 5.** (A) Hi-C matrix at 1 kb resolution in ESCs using juicebox tool. (B) Virtual 4C at 1 kb resolution, surrounding the silencer regions. TADs coordinates are from Bonive *et al.* data.

**A****B**

**Supplementary figure 6: (A)** Genomic tracks for Histone marks and TFs from DP thymocytes around the *Sdhaf1* locus. All the tracks scale is fixed using the TCR $\alpha$  enhancer. **(B)** Tracks from UCSC genome browser (NCBI37/mm9) of CTCF ChIP-Seq from ENCODE/LICR and PSU.



**Supplementary figure 7.** UCSC genome browser (NCBI37/mm9) displaying the the pPGK CapStarr-seq signal, the Jaspar TFBS and the repetitive elements tracks around *the Sdhaf1*\_silencer.

**Supplementary Table 2: CapSTARR-Seq silencers with or without REST binding sites validated by Retroviral reporter assay in fibroblast and T cell .**

Fibroblast cells (3t3)	Tested	Validated	percentage
Total	36	10	27.7
Without Rest	27	2	7.4
With Rest	9	8	88.9

T cells (Beko)	Tested	Validated	percentage
Total	16	9	56.2
Without Rest	12	5	41.6
With Rest	4	4	100

**Supplementary Table 3: Chipseq data**

Sample	Cell type	GEO ID	Reference
GATA3	DP Thymocytes	GSE20898	Wei, et al. (2011). <i>Immunity</i> 35(2): 299-311
Ikaros	DP Thymocytes	GSE32311	Zhang, et al. (2012). <i>Nat Immunol</i> 13(1): 86-94
CTCF	DP Thymocytes	GSE27214	Ebert, et al. (2011). <i>Immunity</i> 34(2): 175-187
RUNX1	DP Thymocytes	GSE44578	Lepoivre et al. (2013). <i>BMC genomics</i> 14: 914
ETS1	DP Thymocytes	GSE29362	Koch, et al. (2011). <i>N. S. M. B.</i> 18(8): 956-963
H3K4me1			
H3K4me3			
Polymerase II			
H3K27me3	DP Thymocytes	GSE38577	Fenouil, et al. (2012). <i>Genome Res</i> 22(12): 2399-2408
MNase			
H3K27ac	DP Thymocytes	GSE63732	Lepoivre et al. (2013). <i>BMC genomics</i> 14: 914
HEB	DP Thymocytes	GSE63732	Vanhille, L et al. (2015). <i>Nature communications</i> , 6, 6905.
TCF1			
DNase			
H3K9me3			
H3K4me3	P5424	GSE120655	Saadi, W. et al. <i>Leukemia &amp; lymphoma</i> , 60(8), 2002-2014.
H3K27ac			

### Supplementary Table 4: All tested silencers and coordinates of Luciferase experiments

Name corres. to nearest genes	Condition	Captured coordinates	Coordinates cloned	Tested dual orientation	Control region
Slc25a24	pPGK, pEF1a, pSCP1	chr3:109126760-109126943	chr3:109126487-109127016		
Col15a1	pEF1a, pSCP1	chr4:47332572-47332788	chr4:47332325-47333037		
Emc10	pRag2- Eα	chr7:51751911-51752094	chr7:51751549-51752373		
Actl6a	pPGK, pEF1a, pSCP1	chr3:32607416-32607599	chr3:32607203-32607749		
Dnpep	pPGK	chr1:75314205-75314388	chr1:75313899-75314437		
Nmt2	pPGK	chr2:3201204-3201387	chr2:3201023-3201576		
Errfi1	pEF1a, pSCP1	chr4:150183355-150183538	chr4:150183112-150183779	Yes	
Cdk8	pPGK, pEF1a, pSCP1	chr5:147072539-147072842	chr5:147072368-147073047		
Hexim1	pEF1a, pSCP1	chr11:102977482-102977591	chr11:102977319-102978009		
Add3	pPGK, pSCP1	chr19:53216896-53217079	chr19:53216801-53217392		
Nr1d2	pEF1a, pSCP1	chr14:19071668-19071851	chr14:19071510-19072076		
Gm2a	pRag2- Eα, pSCP1	chr11:54910998-54911181	chr11:54910870-54911645		
Arrdc2	pRag2- Eα, pSCP1	Chr8:73363529-73363712	Chr8:73363329-73363956	Yes	
D3Erttd254e	pPGK, pEF1a, pSCP1	chr3:36049917-36050026	chr3:36049670-36050132		
Qdpr	pRag2-Eα, pSCP1	chr5:45841471-45841654	chr5:45841156-45841896	Yes	
Acap1	pRag2-Eα, pSCP1	chr11:69709004-69709187	chr11:69708824-69709410		
Zfp1	pPGK, pEF1a, pSCP1	chr8:124805512-124805695	chr8:124805225-124805869	Yes	
Ets2bis	pPGK, pSCP1	chr16:96037720-96037903	chr16:96037479-96037989		
Notch1	pPGK, pSCP1	chr2:26359460-26359643	chr2:26359168-26359723	Yes	
Myo 10	pRag2-Eα, pSCP1	chr15:25552136-25552319	chr15:25551850-25552510	Yes	
Rab43	pEF1a, pSCP1	chr6:87761975-87762158	chr6:87761650-87762305	Yes	
Tmem5	pPGK, pSCP1	chr10:121534331-121534514	chr10:121534224-121534717	Yes	
Cfdp1	pPGK, pEF1a, pSCP1	chr8:114345202-114345407	chr8:114345003-114345643	Yes	
Jkamp	pPGK, pEF1a, pSCP1	chr12:73186582-73186765	chr12:73186314-73186796	Yes	
Dirc2	pPGK, pRag2-Eα, pEF1a pSCP1	chr16:35770112-35770295	chr16:35769967-35770422		
Rbfox3	pPGK, pRag2-Eα, pSCP1	chr11:118446322-118446505	chr11:118446148-118446659	Yes	
Dpp9	pPGK, pSCP1	chr17:56381533-56381716	chr17:56381335-56381900	Yes	
Sdhaf1	pPGK	chr7:31138596-31138779	chr7:31138157-31138782	Yes	
Abr		chr11:76307697-76307880	chr11:76307509-76308239		Yes
Nde1		chr16:14162950-14163133	chr16:14162696-14163563		Yes
Frmd8		chr19:5803071-5803254	chr19:5802822-5803654		Yes
Rab14		chr2:35070226-35070409	chr2:35070052-35070766		Yes
Clpb		chr7:108828346-108828529	chr7:108828010-108828682		Yes
Taf3		chr2:9515681-9515868	chr2:9515498-9515991		Yes
ZEB1		chr18:5591575-5591816	chr18:5591070-5591573		Yes
Dus2		chr8:108541315-108541498	chr8:108546558-108547042		Yes
GATA3		chr2:9486202-9486517	chr2:9486049-9486550		Yes
Ctso		chr3:81736747-81736930	chr3:81736409-81737152		Yes
Mast2		chr4:116121226-116121409	chr4:116120898-116121544		Yes

**Supplementary Table 5: Guide RNA and primer sequences for genome editing**

Guide RNA sequences	Sequence (5'-3')	Coordinates	Expected deleted region (bp)
Sdhaf1_G1	TATCATATGCCGTTCCATGT	chr7:31138272-31138291	696
Sdhaf1_G2	GAGAAGCTACTTGCATGGGG	chr7:31138989-31139008	

**Primer sequences for detection of genome editing**

Name	Sequence (5'-3')	Coordinates	Expected band size (bp)	
			With editing	Without editing
Sdhaf1_del_F	GAATGGAATCCTCCCTCCTC	chr7:31138131-31138150	338	1034
Sdhaf1_del_R	GCCTCATGTCTTTCCTGAGC	chr7:31139145-31139164		



## Chapter 6

### Discussion

A definitive assay for the silencer function of a genetic element is to isolate it and measure its ability to repress promoter activity in different cell types (Petrykowska et al., 2008; Qi et al., 2015). To screen for silencers in a manner optimizing throughput and functional information, we adapted a high throughput reporter assay based on the CapStarr-seq approach developed previously (Vanhille et al., 2015). Here we identified DNA fragments capable of negatively regulating their own transcription in a minimal, episomal context in transfected cells, allowing for genome-wide screening of putative silencers from several tens of thousands of genomic regions. As an initial assessment of our approach, we analyzed a set of 28055 DHS from primary developing mouse thymocytes in a T cell line previously used in CapStarr-seq (Vanhille et al., 2015). We tested 4 reporter vectors containing 4 different promoters. This suggests that the presence of the E $\alpha$  enhancer might interfere with the silencer activity. We extensively assessed the accuracy of CapStarr-seq to quantify silencer activity and demonstrated its robustness to identify *bona fide* silencers. Overall, the basic pScp1 vector performed well in the identification of both enhancers and silencers, while the ubiquitous promoters-harboring vectors pEF1 $\alpha$  and pPGK allowed only identifying silencers. In contrast, the tissue specific vector pRag2-E $\alpha$  failed to identify neither silencer nor enhancers. We need to finish the bioinformatic analyses (specially the new pipeline) to be able to fully conclude that which promoter system is the most reliable. Initial results suggest that SCP1 and PGK promoter systems might be the most reliable. pSCP1 library signals were positively correlated with gene expression of available RNA-seq data from P5424 cells and the highest number of silencer candidates were also identified using SCP1 and PGK promoter as compare to others. In particular, the silencer candidates with the highest luciferase activity (*Cfdp1*, *Jkamp*, *Dirc2*, *Rbfox3*, *Dpp9*, *Sdhaf1*) were identified by the pSCP1 and pPGK-based vectors. Furthermore, pSCP1 and pPGK based silencers candidates were enriched in REST motif.

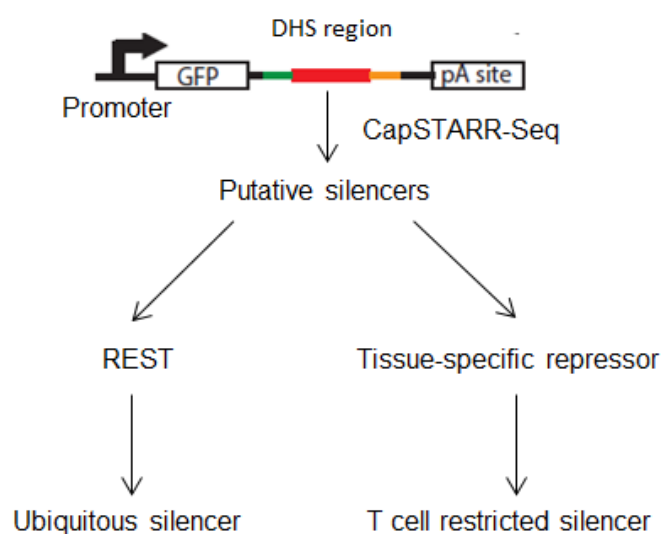
Several previous works have identified silencers using indirect high-throughput approaches. Fulco et al. performed CRISPRi (KRAB-dCas9) to tile gRNAs across a large region around GATA1 and MYC in K562 cells and measured the effects using a proliferation assay (Fulco et al., 2016). They identified two repressive elements associated with Myc locus. Subsequent

work by the same team. A study using Capture Hi-C (ChI-C) technology suggested that transcriptionally inactive genes interact with previously uncharacterized elements marked by repressive features that may act as long-range silencers (Mifsud et al., 2015). They further showed that a genomic region located 1.2 Mb from the *BCL6* promoter and associated with Polycomb complex displayed silencer activity when tested in a reporter assay. The Stamatoyannopoulos team developed a genomic approach for the identification of compact sequence elements that function as insulators (Liu et al., 2015). They found that some of these regions have silencer activity instead of enhancer-blocking insulation. The same team further reported the characterization and functional validation of a constitutive autonomous silencer element, named T39 (Qi et al., 2015). A recent study identified a genetic variant associated with vascular diseases (Gupta et al., 2017). Interestingly, the deletion of the genomic region surrounding the genetic variant resulted in the overexpression of the *EDN1* gene located 600 kb from the variant, suggesting this region might work as a silencer element. However, none of these studies directly focused on the study of silencer, nor provided global epigenomic or genomic features associated with these elements.

A recent study performed a systematic high-resolution activation and repression profiling with reporter tiling using Massively Parallel Reporter Assay (MPRA), and provided a genome-scale high-resolution mapping of activating and repressive nucleotides (Ernst et al., 2016). Interestingly, one of the major features associated with repression was found to be the REST motif. However, in this study the authors did not identify silencer elements per se neither provide in vivo validation of their findings. In a recent study, putative silencers were identified by isolating H3K27me3-DHS peaks with tissue specificity negatively correlated with the expression of nearby genes across 25 diverse cell lines (Huang et al., 2019). However, experiments luciferase reporter assay did not demonstrate increased silencer activity over randomly selected DHS regions. Another, unpublished study (Jayavelu *et al.*; <http://dx.doi.org/10.1101/252304>), have also attempted to identify silencer across several human and mouse cell types. The authors developed a computational pipeline to infer putative silencers. Subsequently, they performed Starr-seq with a subset of predicted silencers, but the overall signal obtained with the putative silencers was not significantly different to the signal obtained with randomly selected regions.

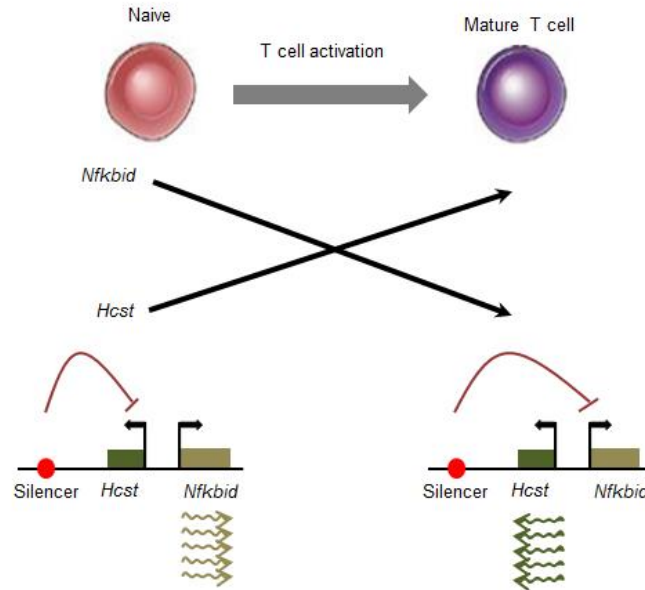
Our results suggested the existence of two main types of silencer elements. One type of silencers contains the binding site for the REST repressor and appears to have ubiquitous

silencer activity. This is in agreement with the wide spread role of REST. REST is expressly involved in the repression of neural genes in non-neuronal cells (Chong et al., 1995b; Coulson, 2005; Schoenherr and Anderson, 1995). However REST function is not restricted to repression of neuronal genes and might be involved in the regulation of distinct developmental pathways (Kuwahara et al., 2003; Martin et al., 2003). Moreover, the REST-containing putative silencers displayed a very low rate of false positive in the retroviral validation assay, indicating that the presence of REST binding site provide universal silencer activity (Figure 6.1). The other type of silencers does not contain a REST binding site and appears to have more tissue-restricted silencer activity. Some known silencers have been shown to recruit tissue specific transcription factors with repressive activities (Collins et al., 2009), but the overall set of proteins which collaborate to impart silencer function are essentially unknown. In particular, members of the Runx family have been shown to mediated silencer in T cell specific genes, such as CD4 and ThPOK (Jiang and Peterlin, 2008; Jiang et al., 2005; Setoguchi et al., 2008; Taniuchi et al., 2002a; Telfer et al., 2004; Wildt et al., 2007). However, we did not observed significant enrichment for Runx motifs in our set of putative silencers. Although, we cannot exclude that some of the identified silencers might harbour RUNX-dependent silencer activity, it is plausible that RUNX-mediated silencing is not detected with the reporter assay. For instances, the RUNX-dependent CD4 silencer work by quenching the CD4 enhancers and therefore interfering with the enhancer-promoter looping (Jiang and Peterlin, 2008), such “indirect” mechanism might be difficult to reproduce in the reporter assay.



**Figure 6.1 | A scheme showing REST-dependent and tissue-specific silencers.**

Although reporter assays, either episomal or integrative, are capable of demonstrating the potential of a test region to positively- or negatively-regulate transcription, they do not alone prove that the endogenous copy is active, or even required, in any particular cellular setting. Similarly, there are some limitations using reporter assays already discussed in Chapter 3. This is why we performed genetic engineering using CRISPR/cas9 technology to know the functional role of the element *in vivo* and an integrative reporter assays using a retroviral system. Another alternative approach is to develop a site-specific integrative reporter based on a recently described strategy discussed in perspective (Weingarten-Gabbay et al., 2017; Weingarten-Gabbay et al., 2019). While this has been successfully undertaken for many enhancers (Bauer et al., 2013; Canver et al., 2015; Diao et al., 2016; Huang et al., 2016; Zhou et al., 2014), it has only seldom been performed for a silencer (e.g. (Leung et al., 2001; Tanaka et al., 2013)). This is however a critical issue because silencers might be highly stage specific or function only under given cellular context, and in view of the many-to-one relationship between distal regulatory elements and genes, some degree of redundancy is also expected. Therefore, to increase the chances of isolating *in vivo* functional silencers, we crossed our set of putative silencers with the most highly variable genes during T cell differentiation and maturations. We identified the *Sdhaf1*\_silencer which was associated with two genes *Hcst* and *Nfkbid*, both highly regulated during T cell differentiation and involved in T cell activation and function. Deletion of the *Sdhaf1*\_silencer resulted in up-regulation of both genes, either in unstimulated (*Hscst*) or stimulated (*Nfkbid*) conditions. Given the fact that the two genes are anti-correlated during T cell maturation and that they might play antagonist roles, it is tempting to speculate that the *Sdhaf1*\_silencer might control the extend of T cell maturation and/or activation by regulating the relative balance of *Hcst* versus *Nfkbid* genes (**Figure 6.2**).



**Figure 6.2 | Coordinated regulation of *Hcst* and *Nfkbid* gene expression by the *Sdhaf1\_silencer* during T cell maturation.**

Examination of TF binding at the *Sdhaf1\_silencer* revealed the presence of a CTCF binding site which appears to be occupied in several cell types. Interestingly, in a previous study, the T39 region bound by CTCF has been shown to function as a strong silencer, but is devoid of insulator activity, thus suggesting that in some cases CTCF might be specifically required for silencer activity (Liu et al., 2015; Qi et al., 2015). The precise mechanisms of repression mediated by *Sdhaf1\_silencer* will need to be investigated in the near future.

Overall, we provided experimental evidence that at least one of the CapStarr-seq identified silencers function in the endogenous context as a *bona fide* silencer element and might play a key physiological role in the regulation of T cell function.

Our study represents an initial step toward the understanding of molecular basis driving silencer activity, including epigenetic features and binding of transcription factors. Together our findings illustrate how the CapStarr-seq technique will further help to the functional assessment of mammalian silencers that are active in different cellular systems. This approach will be also useful for thorough characterization of silencer subsets involved in specific pathways or induced by specific stimuli.

# Chapter 7

## Perspectives

### *1. Identifying functional silencers*

An important milestone of our study will be to identify several *bona fide* silencers that regulate endogenous genes. Combining our set of silencers with highly variable genes throughout T cell differentiation has led to the identification of a silencer (*Sdhaf1\_silencer*) which might have functional *in vivo* relevance. Attempts to generate other silencer deletions by CRISPR/Cas9 have currently failed.

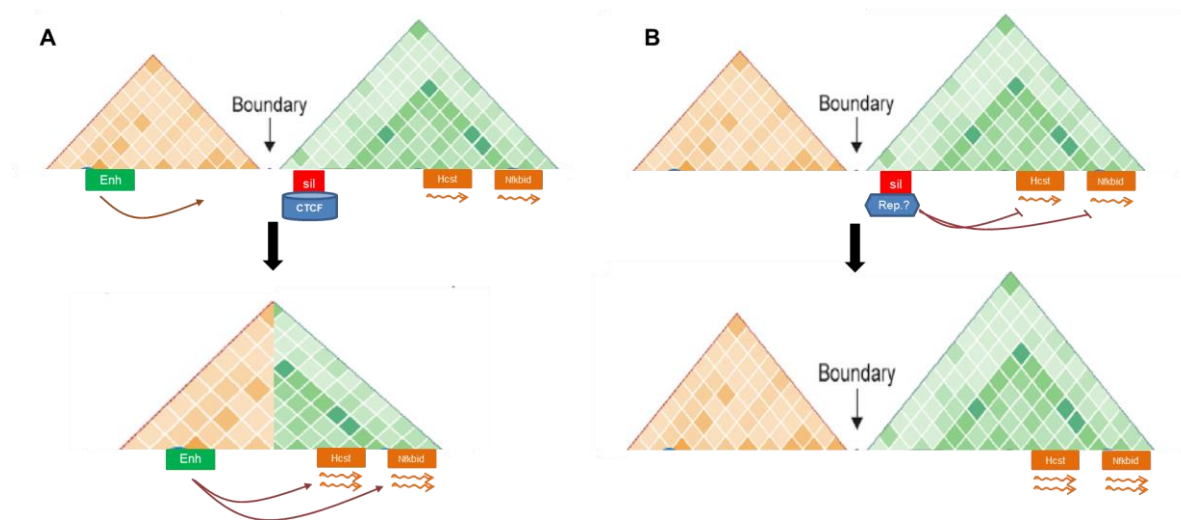
Huang et al. have developed a pipeline whereby silencers are predicted by negative correlation between H3K27me3-(DHS) peaks and the expression of nearby genes across several tissues (Huang et al., 2019). On the other hand, silencers might be identified by searching for genomic elements marked by repressive features and involved in cell type-restricted interactions as determined by Capture Hi-C (Mifsud et al., 2015). We propose that these approaches could be improved by integrating the correlation between putative silencers identified by CapStarr-seq with epigenetic profiles, gene expression and genome-wide 3D interactions. To this end, Capture HiC will be performed with our selection of putative silencers in collaboration with the team of Tom Sexton (IGBMC, Strasbourg). Another alternative strategy is to search for silencers associated with repressor-dependant genes. This approach has been undertaken by our collaborators from IRCAN (Simona Sacconi). They have crossed our list of putative silencers with a list of REST-dependant genes and experimentally validated a REST-dependant silencer in its endogenous context.

### *2. Exploring the function of Sdhaf1\_silencer*

Our experimental results have provided solid evidences that *Sdhaf1\_silencer* regulates the expression of two key genes involved in T cell activation, *Hcst* and *Nfkbid*. However, several questions remain regarding the mechanism of action of this silencer and its physiological function.

First, is the *Sdhaf1\_silencer* directly contacting the *HCST* and *Nfkbid* promoters or if the silencer function mediated by an indirect mechanism? For example, by interfering with

another *cis*-regulatory element or the TAD border. The presence of CTCF binding within the silencer as well as its proximity to a TAD boundary suggests that the effect of *Sdhaf1\_Sil* deletion on gene expression might be due to an enhancer adoption mechanism due to the disruption of the TAD boundary (Fig 7.1A). Our preliminary observations using available Hi-C data suggest that *Sdhaf1\_silencer* might interact in vivo with the two regulated genes. To investigate this possibility we should explore the 3D topology of the locus in the presence or absence of the *Sdhaf1\_silencer*. Current collaboration with the team of Tom Sexton (IGBMC, Strasbourg) aims to perform 4C experiments in P5424 wild-type and *Sdhaf1* knock-out cells.



**Figure 7.1|** Model (A) showing the enhancer adoption mechanism due to deletion of an insulator after performing Hi-C experiment in p5424 wild type and *Sdhaf1\_silencer* knockout cells. Model (B) showing the effect of *Sdhaf1\_silencer* on gene expression before and after site directed mutagenesis of potential repressors.

Second, what are the factors required for the silencer activity? We have observed that CTCF binds to *Sdhaf1\_silencer*, which could mediate silencer activity as previously shown (Qi et al., 2015). We have observed a reproducible and consistent silencer effect using CapStarr-seq and luciferase reporter assays, which suggest that the *Sdhaf1\_sil* harbor intrinsic silencer properties which are not dependent on the 3D organization of the locus (Fig 7.1B). However, we have also identified other potential repressors such as ZNF263 and GFI1 which could also be involved in the silencer function of *Sdhaf1\_silencer*. To better explore the sites required for the silencer activity a directed mutagenesis should be performed to interrogate the effects of disrupting each of the aforementioned sites.

Third, what is the function of *Sdhaf1\_silencer* in vivo? Given the important roles play by *Hcst* and *Nfkbid* in T cell function it is plausible that modulating their expression by the *Sdhaf1\_silencer* might be essential to proper T cell activation. To address this issue, the effect of *Sdhaf1\_silencer* deletion might be studied in a more physiological context by generating knock-out mice models. Finally, it is tempting to speculate that genetic alteration of *Sdhaf1\_silencer* (if conserved in humans) might be associated with immune diseases. It will be worth to assess whether any GWAS associated SNP overlap with the human counterpart of the *Sdhaf1\_silencer*.

### ***3. Developing an improved pipeline to identify silencers***

Identifying putative silencers based on the CapStarr-seq signal is not a trivial issue, because the selection might be influenced by several confounding parameters, including the setting of the signal threshold, the representation of the region in the input library, consistency between replicates and strand orientation biases due to the presence of RNA-stabilization or destabilization sequences. In the present study, we employed two empirical threshold based either on the ranked signal (bottom 500) or the fold change ( $<-1$ ). Although somehow arbitraries, these threshold have allow to already gain new insides on the epigenomics and genomics features of silencers and to perform a series of experimental validations.

To account for the aforementioned issues, we aim to develop a new analytical pipeline for a robust identification of silencers. The pipeline will associate a strong statistical test based on the comparison of individual DHS fragments with the random captured fragments and account for consistency between strand orientation and experimental replicates. The pipeline is currently being developed in collaboration with the lab of Simona Sacconi (IRCAN, Nice).

### ***4. Future experiments***

Once we have fully validated our strategy, it will be worth to perform a similar approach to identify silencer in different human cell lines. Instead of capturing DHS sites, it will be more convenient to directly clone the ATAC fragments into the Starr-seq vector, as recently published (Wang et al., 2018). This will provide a comprehensive dataset of silencers, which will allow to better assess the genomic and epigenomics features associated with silencers by taking advantage of the wealth of genomic data available in human cell lines (e.g. ENCODE). This will also allow to identify silencers associated with specific diseases by



crossing the list of silencers with GWAS SNPs or studying the alteration of silencers in cancer samples.

A potential issue with our study is that silencers might not be generally present within DHSs site, or that at least some silencers are not within open chromatin regions. Therefore to test for candidate regions that might not be within DHSs, and to be able to estimate the abundance of silencers across the whole genome, we will use pool(s) of BACs covering large genomic intervals. Our aim is to cover up to 1% of the mouse genome (30 Mb, or approximately a collection of  $\approx 384$  BACs [contained in a single library plate]). The BACs will be fragmented by sonication and cloned directly into the Starr-seq vectors. We have already performed such experiments in collaboration with the team of Simona Sacconi (IRCAN, Nice), but we need to wait for the development of the new analytical pipeline to properly process the data. If our initial study led to the conclusion that a substantial number of silencers might be located outside DHS, we might envision the possibility of performing Starr-seq with the whole genome.

Another general concern about the episomal reporter assays is that they may not accurately reflect the function of silencer elements in their endogenous context. 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. In any case, it will be of interest to compare episomal and integrative reporter assays, using for instances retroviral vectors. Another alternative approach is to develop a site-specific integrative reporter based on a recently described strategy (Weingarten-Gabbay et al., 2017; Weingarten-Gabbay et al., 2019). Briefly, a genomic library such as the one generated for the CapStarrseq, could be integrated into the “safe harbor” AAVS1 site by inducing a double strand break using specific Zinc Finger Nucleases (ZFNs) followed by genomic integration of a reporter cassette by homologous recombination.

### ***5. Remaining questions***

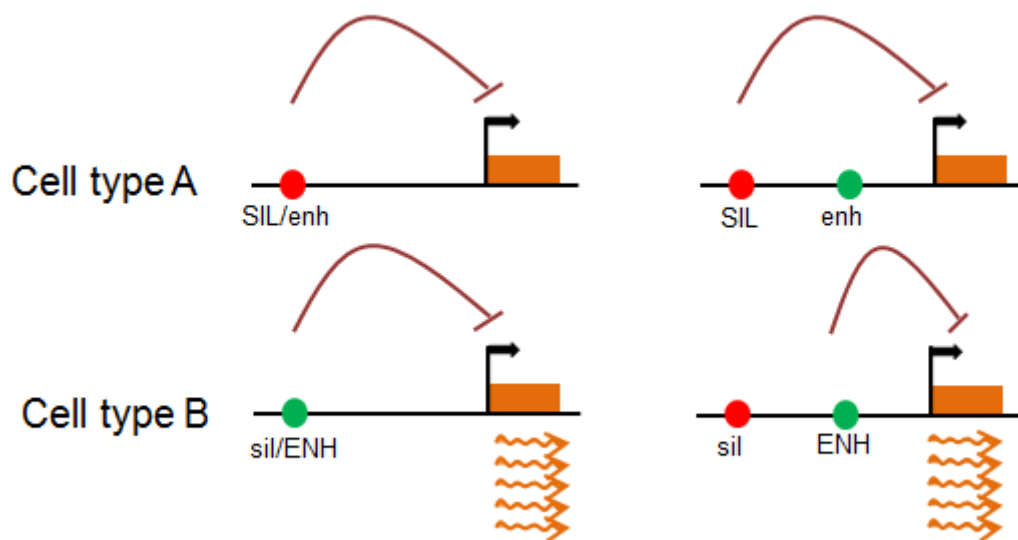
No general picture has yet emerged of a common mechanism of silencer function. Our study will open new avenues for further interrogation of several aspects of silencer biology.

*How do silencers interact with surrounding genomic regions to regulate distant genes?* it is not known if silencers form loops to directly contact their target promoters (analogously to

enhancers; (Bergeron et al., 2015; Dean, 2011; Tiwari et al., 2008), if they act by repressing the activities of enhancers, either by directly repressive effects, or by inhibiting enhancer-promoter contact (Jiang and Peterlin, 2008), or if they nucleate larger-scale domains of repressive chromatin by recruitment of chromatin modifiers such as Polycomb complexes (Dietrich et al., 2012; Li and Arnosti, 2011; Tiwari et al., 2008). Indeed, it remains possible that each of these mechanisms may be utilized by distinct silencer classes.

*-What are the molecular factors and pathways that are functionally utilized by silencers?* Some silencers recruit transcription factors with repressive activities, such as REST/NRSF and members of the Runx family (Chong et al., 1995b; Collins et al., 2009), but the overall set of proteins which collaborate to impart silencer function is essentially unknown.

*-Are silencers fully independent regulatory elements?* Whether silencers function as autonomous regulatory elements independent of enhancers or whether the same regulatory elements can display silencer or enhancer activity depending on the cellular context will need to be investigated (Figure 7.2). Recent studies aimed at dissecting functional subsequences within selected enhancers have revealed that some enhancers can be bound by a spectrum of activating and repressing transcription factors, depending on the cellular context (Canver et al., 2015; Ernst et al., 2016; Kheradpour et al., 2013). Thus, it even remains possible that some enhancers and silencers may comprise an overlapping set of genetic elements, which exhibit divergent behaviours under different biological conditions (Li et al., 2014).



**Figure 7.2 | Models of silencer versus enhancer function.** Capital letters indicate active regulatory element.

*-Could promoter overlapping silencers regulate distal genes?* Our study has identified many silencers close (or overlapping to) to gene promoters. Our laboratory has recently identified a subset of core promoters (named Epromoters) that also display enhancer function to regulate distal genes (preliminary results from the lab (Dao et al., 2017)). Thus, similar to the Epromoters, some promoter overlapping silencers might concomitantly regulate distal genes. Preliminary results from the group of Pascal Rihet at the TAGC suggest that this might be the case.

*-Is deregulation of silencers a mechanism of oncogenic activation?* Functional disabling of regulatory elements required for expression of key cancer-related genes, by direct mutation or impaired function of regulatory factors, is emerging as a major oncogenic mechanism (Bradner et al., 2017; Flavahan et al., 2016; Flavahan et al., 2017; Hnisz et al., 2016). Ectopic expression of master oncogenes might similarly arise from disruption of silencer elements. For instances, a silencer element located downstream of *TALI* proto-oncogene has been suggested to play a role in T cell leukemia (Le Clech et al., 2006). Although perturbation of silencer elements has been proposed as a potential mechanism for proto-oncogene activation (Teitell and Pandolfi, 2009), to our knowledge no published study have formally demonstrated such mechanism.

# Bibliography

- Abderrahmani, A., Steinmann, M., Plaisance, V., Niederhauser, G., Haefliger, J.A., Mooser, V., Bonny, C., Nicod, P., and Waeber, G. (2001). The transcriptional repressor REST determines the cell-specific expression of the human MAPK8IP1 gene encoding IB1 (JIP-1). *Molecular and cellular biology* 21, 7256-7267.
- Adlam, M., and Siu, G. (2003). Hierarchical interactions control CD4 gene expression during thymocyte development. *Immunity* 18, 173-184.
- Adoue, V., Binet, B., Malbec, A., Fourquet, J., Romagnoli, P., van Meerwijk, J.P.M., Amigorena, S., and Joffre, O.P. (2019). The Histone Methyltransferase SETDB1 Controls T Helper Cell Lineage Integrity by Repressing Endogenous Retroviruses. *Immunity* 50, 629-644.e628.
- Akbari, O.S., Bae, E., Johnsen, H., Villaluz, A., Wong, D., and Drewell, R.A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo. *Development* 135, 123-131.
- Alcid, E.A., and Tsukiyama, T. (2014). ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. *Genes Dev* 28, 2348-2360.
- Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays : news and reviews in molecular, cellular and developmental biology* 37, 314-323.
- 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.
- Anguita, E., Candel, F.J., Chaparro, A., and Roldan-Etcheverry, J.J. (2017). Transcription Factor GF11B in Health and Disease. *Frontiers in oncology* 7, 54.
- Antalis, T.M., Costelloe, E., Muddiman, J., Ogbourne, S., and Donnan, K. (1996). Regulation of the plasminogen activator inhibitor type-2 gene in monocytes: localization of an upstream transcriptional silencer. *Blood* 88, 3686-3697.
- Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nature genetics* 46, 685-692.
- 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.
- Auble, D.T. (2009). The dynamic personality of TATA-binding protein. *Trends Biochem Sci* 34, 49-52.
- Baatar, D., Jones, M.K., Pai, R., Kawanaka, H., Szabo, I.L., Moon, W.S., Kitano, S., and Tarnawski, A.S. (2002). Selective cyclooxygenase-2 blocker delays healing of esophageal ulcers in rats and inhibits ulceration-triggered c-Met/hepatocyte growth factor receptor induction and extracellular signal-regulated kinase 2 activation. *The American journal of pathology* 160, 963-972.
- Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *The Journal of biological chemistry* 281, 15172-15181.
- Balestrieri, C., Alfarano, G., Milan, M., Tosi, V., Prosperini, E., Nicoli, P., Palamidessi, A., Scita, G., Diaferia, G.R., and Natoli, G. (2018). Co-optation of Tandem DNA Repeats for the Maintenance of Mesenchymal Identity. *Cell* 173, 1150-1164.e1114.
- Ballas, N., and Mandel, G. (2005). The many faces of REST oversee epigenetic programming of neuronal genes. *Current opinion in neurobiology* 15, 500-506.
- Banan, M., Rojas, I.C., Lee, W.H., King, H.L., Harriss, J.V., Kobayashi, R., Webb, C.F., and Gottlieb, P.D. (1997). Interaction of the nuclear matrix-associated region (MAR)-binding proteins, SATB1 and CDP/Cux, with a MAR element (L2a) in an upstream regulatory region of the mouse CD8a gene. *The Journal of biological chemistry* 272, 18440-18452.

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.

Barakat, T.S., Halbritter, F., Zhang, M., Rendeiro, A.F., Bock, C., and Chambers, I. (2017). Functional dissection of the enhancer repertoire in human embryonic stem cells. *bioRxiv*.

Barberis, A., Superti-Furga, G., and Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* 50, 347-359.

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.

Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152, 1298-1307.

Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., *et al.* (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 342, 253-257.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482-485.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. *Nature* 290, 304-310.

Bergeron, K.F., Cardinal, T., Toure, A.M., Beland, M., Raiwet, D.L., Silversides, D.W., and Pilon, N. (2015). Male-biased aganglionic megacolon in the TashT mouse line due to perturbation of silencer elements in a large gene desert of chromosome 10. *PLoS genetics* 11, e1005093.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., *et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28, 1045-1048.

Bi, X., Yu, Q., Sandmeier, J.J., and Zou, Y. (2004). Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. *Molecular and cellular biology* 24, 2118-2131.

Bilic, I., Kisters, K., Unger, B., Sekimata, M., Hertweck, A., Maschek, R., Wilson, C.B., and Ellmeier, W. (2006). Negative regulation of CD8 expression via CD8 enhancer-mediated recruitment of the zinc finger protein MAZR. *Nat Immunol*.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Molecular cell* 38, 179-190.

Bossu, J.P., Chartier, F.L., Fruchart, J.C., Auwerx, J., Staels, B., and Laine, B. (1996). Two regulatory elements of similar structure and placed in tandem account for the repressive activity of the first intron of the human apolipoprotein A-II gene. *The Biochemical journal* 318 ( Pt 2), 547-553.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. *Cell* 168, 629-643.

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.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends in genetics : TIG* 28, 276-284.

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.

Bunch, H. (2018). Gene regulation of mammalian long non-coding RNA. *Molecular genetics and genomics : MGG* 293, 1-15.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Molecular cell* 49, 825-837.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., *et al.* (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192-197.

Carlson, H.L., Quinn, J.J., Yang, Y.W., Thornburg, C.K., Chang, H.Y., and Stadler, H.S. (2015). LncRNA-HIT Functions as an Epigenetic Regulator of Chondrogenesis through Its Recruitment of p100/CBP Complexes. *PLoS genetics* 11, e1005680.

Carpenter, A.C., and Bosselut, R. (2010). Decision checkpoints in the thymus. *Nature immunology* 11, 666-673.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25-36.

Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev* 32, 202-223.

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*.

Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS biology* 6, e277.

Chiang, C., and Ayyanathan, K. (2013). Snail/Gfi-1 (SNAG) family zinc finger proteins in transcription regulation, chromatin dynamics, cell signaling, development, and disease. *Cytokine & growth factor reviews* 24, 123-131.

Choi, J.K., and Howe, L.J. (2009). Histone acetylation: truth of consequences? *Biochemistry and cell biology = Biochimie et biologie cellulaire* 87, 139-150.

Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuler, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995a). REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 80, 949-957.

Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuler, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995b). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 80, 949-957.

Collins, A., Hewitt, S.L., Chaumeil, J., Sellars, M., Micsinai, M., Allinne, J., Parisi, F., Nora, E.P., Bolland, D.J., Corcoran, A.E., *et al.* (2011). RUNX transcription factor-mediated association of Cd4 and Cd8 enables coordinate gene regulation. *Immunity* 34, 303-314.

Collins, A., Littman, D.R., and Taniuchi, I. (2009). RUNX proteins in transcription factor networks that regulate T-cell lineage choice. *Nature reviews Immunology* 9, 106-115.

Conaco, C., Otto, S., Han, J.J., and Mandel, G. (2006). Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc Natl Acad Sci U S A* 103, 2422-2427.

Corbella, P., Moskophidis, D., Spanopoulou, E., Mamalaki, C., Tolaini, M., Itano, A., Lans, D., Baltimore, D., Robey, E., and Kioussis, D. (1994). Functional commitment to helper T cell lineage precedes positive selection and is independent of T cell receptor MHC specificity. *Immunity* 1, 269-276.

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.

Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P.A., and Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* 6, 712-722.

Coulson, J.M. (2005). Transcriptional regulation: cancer, neurons and the REST. *Current biology : CB* 15, R665-668.

Courtes, C., Lecointe, N., Le Cam, L., Baudoin, F., Sardet, C., and Mathieu-Mahul, D. (2000). Erythroid-specific inhibition of the tal-1 intragenic promoter is due to binding of a repressor to a novel silencer. *The Journal of biological chemistry* 275, 949-958.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews Genetics* 2, 292-301.

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.

Cruz-Guilloty, F., Pipkin, M.E., Djuretic, I.M., Levanon, D., Lotem, J., Lichtenheld, M.G., Groner, Y., and Rao, A. (2009). Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. *J Exp Med* 206, 51-59.

D'Acquisto, F., and Crompton, T. (2011). CD3+CD4-CD8- (double negative) T cells: saviours or villains of the immune response? *Biochemical pharmacology* 82, 333-340.

Dailey, L. (2015). High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics. *Genomics* 106, 151-158.

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., Vanhille, L., Griffon, A., Fernandez, N., and Spicuglia, S. (2015). CapStarr-seq protocol. Protocol Exchange.

Dasgupta, A., Darst, R.P., Martin, K.J., Afshari, C.A., and Auble, D.T. (2002). Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. *Proc Natl Acad Sci U S A* 99, 2666-2671.

Dave, V.P., Cao, Z.S., Browne, C., Alarcon, B., Fernandez-Miguel, G., Lafaille, J., De la Hera, A., Tonegawa, S., and Kappes, D.J. (1997). CD3 $\delta$  deficiency arrests development of the  $\alpha$  but not the  $\beta$  T cell lineage. *The EMBO journal* 16, 1360-1370.

Dean, A. (2011). In the loop: long range chromatin interactions and gene regulation. *Briefings in functional genomics* 10, 3-10.

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.

Di Salvo, T.G., Guo, Y., Su, Y.R., Clark, T., Brittain, E., Absi, T., Maltais, S., and Hemnes, A. (2015). Right ventricular long noncoding RNA expression in human heart failure. *Pulm Circ* 5, 135-161.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., *et al.* (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* 14, 629-635.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A.Y., Dixon, J., Maliskova, L., Guan, K.L., Shen, Y., and Ren, B. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome research* 26, 397-405.

Diefenbach, A., Tomasello, E., Lucas, M., Jamieson, A.M., Hsia, J.K., Vivier, E., and Raulet, D.H. (2002). Selective associations with signaling proteins determine stimulatory versus costimulatory activity of NKG2D. *Nature immunology* 3, 1142-1149.

Dietrich-Goetz, W., Kennedy, I.M., Levins, B., Stanley, M.A., and Clements, J.B. (1997). A cellular 65-kDa protein recognizes the negative regulatory element of human papillomavirus late mRNA. *Proc Natl Acad Sci U S A* 94, 163-168.

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Sodersten, E., and Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS genetics* 8, e1002494.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C., *et al.* (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome research* 18, 1433-1445.

Dirks, R.P., Klok, E.J., van Genesen, S.T., Schoenmakers, J.G., and Lubsen, N.H. (1996). The sequence of regulatory events controlling the expression of the gamma D-crystallin gene during fibroblast growth factor-mediated rat lens fiber cell differentiation. *Developmental biology* 173, 14-25.

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.

Dombret, H., Font, M.P., and Sigaux, F. (1996). A dominant transcriptional silencer located 5' to the human T-cell receptor V-2.2 gene segment which is activated in cell lines of thymic phenotype. *Nucleic Acids Res* 24, 2782-2789.

Donda, A., Schulz, M., Burki, K., De Libero, G., and Uematsu, Y. (1996). Identification and characterization of a human CD4 silencer. *Eur J Immunol* 26, 493-500.

Dong, J.M., and Lim, L. (1996). The human neuronal alpha 1-chimaerin gene contains a position-dependent negative regulatory element in the first exon. *Neurochemical research* 21, 1023-1030.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *The Journal of biological chemistry* 277, 28368-28371.

Ebihara, T., and Taniuchi, I. (2019). Transcription Factors in the Development and Function of Group 2 Innate Lymphoid Cells. *International journal of molecular sciences* 20.

Edelman, L.B., and Fraser, P. (2012). Transcription factories: genetic programming in three dimensions. *Curr Opin Genet Dev* 22, 110-114.

Egawa, T., and Littman, D.R. (2008). ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. *Nature immunology* 9, 1131-1139.

Ellmeier, W., Sunshine, M.J., Maschek, R., and Littman, D.R. (2002). Combined deletion of CD8 locus cis -regulatory elements affects initiation but not maintenance of CD8 expression. *Immunity* 16, 623-634.

Erman, B., Feigenbaum, L., Coligan, J.E., and Singer, A. (2002). Early TCR $\zeta$  expression generates TCR $\zeta$  g complexes that signal the DN-to-DP transition and impair development. *Nat Immunol* 3, 564-569.

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.

Felsenfeld, G., Burgess-Beusse, B., Farrell, C., Gaszner, M., Ghirlando, R., Huang, S., Jin, C., Litt, M., Magdinier, F., Mutskov, V., *et al.* (2004). Chromatin boundaries and chromatin domains. *Cold Spring Harbor symposia on quantitative biology* 69, 245-250.

Feng, D., Bond, C.J., Ely, L.K., Maynard, J., and Garcia, K.C. (2007). Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nature immunology* 8, 975-983.

Feng, J.Q., Chen, D., Cooney, A.J., Tsai, M.J., Harris, M.A., Tsai, S.Y., Feng, M., Mundy, G.R., and Harris, S.E. (1995). The mouse bone morphogenetic protein-4 gene. Analysis of promoter utilization in fetal rat calvarial osteoblasts and regulation by COUP-TFI orphan receptor. *The Journal of biological chemistry* 270, 28364-28373.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110-114.

Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science* 357.

Fourel, G., Magdinier, F., and Gilson, E. (2004). Insulator dynamics and the setting of chromatin domains. *BioEssays : news and reviews in molecular, cellular and developmental biology* 26, 523-532.

French, J.D., Ghousaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M., *et al.* (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet* 92, 489-503.



Frietze, S., Lan, X., Jin, V.X., and Farnham, P.J. (2010). Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *The Journal of biological chemistry* 285, 1393-1403.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769-773.

Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol* 10, R29.

Garefalaki, A., Coles, M., Hirschberg, S., Mavria, G., Norton, T., Hostert, A., and Kioussis, D. (2002). Variegated expression of CD8  $\alpha$  resulting from in situ deletion of regulatory sequences. *Immunity* 16, 635-647.

Gertz, J., Reddy, T.E., Varley, K.E., Garabedian, M.J., and Myers, R.M. (2012). Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome research* 22, 2153-2162.

Giffin, W., Torrance, H., Saffran, H., MacLeod, H.L., and Hache, R.J. (1994). Repression of mouse mammary tumor virus transcription by a transcription factor complex. Binding of individual components to separated DNA strands. *The Journal of biological chemistry* 269, 1449-1459.

Glinsky, G.V. (2015). Transposable Elements and DNA Methylation Create in Embryonic Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers and Noncoding RNAs. *Genome biology and evolution* 7, 1432-1454.

Gohl, D., Aoki, T., Blanton, J., Shanower, G., Kappes, G., and Schedl, P. (2011). Mechanism of chromosomal boundary action: roadblock, sink, or loop? *Genetics* 187, 731-748.

Goodman, P.A., Medina-Martinez, O., and Fernandez-Mejia, C. (1996). Identification of the human insulin negative regulatory element as a negative glucocorticoid response element. *Molecular and cellular endocrinology* 120, 139-146.

Gray, S., and Levine, M. (1996). Transcriptional repression in development. *Curr Opin Cell Biol* 8, 358-364.

Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., *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.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223-227.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., *et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics* 43, 630-638.

Haniel, A., Welge-Lussen, U., Kuhn, K., and Poschl, E. (1995). Identification and characterization of a novel transcriptional silencer in the human collagen type IV gene COL4A2. *The Journal of biological chemistry* 270, 11209-11215.

Hao, B., Naik, A.K., Watanabe, A., Tanaka, H., Chen, L., Richards, H.W., Kondo, M., Taniuchi, I., Kohwi, Y., Kohwi-Shigematsu, T., *et al.* (2015). An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J Exp Med* 212, 809-824.

Harris, M.B., Mostecky, J., and Rothman, P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *The Journal of biological chemistry* 280, 13114-13121.

Hart, C.M., Zhao, K., and Laemmli, U.K. (1997). The scs' boundary element: Characterization of boundary element-associated factors. *Molecular and cellular biology* 17, 999-1009.

Hayday, A.C., and Pennington, D.J. (2007). Key factors in the organized chaos of early T cell development. *Nature immunology* 8, 137-144.

He, X., He, X., Dave, V.P., Zhang, Y., Hua, X., Nicolas, E., Xu, W., Roe, B.A., and Kappes, D.J. (2005). The zinc finger transcription factor Th-POK regulates CD4 versus CD8 T-cell lineage commitment. *Nature* 433, 826-833.

He, X., Park, K., Wang, H., He, X., Zhang, Y., Hua, X., Li, Y., and Kappes, D.J. (2008a). CD4-CD8 lineage commitment is regulated by a silencer element at the ThPOK transcription-factor locus. *Immunity* 28, 346-358.

He, X., Park, K., Wang, H., Zhang, Y., Hua, X., Li, Y., and Kappes, D.J. (2008b). CD4-CD8 lineage commitment is regulated by a silencer element at the ThPOK transcription-factor locus. *Immunity* 28, 346-358.

He, Y., Borellini, F., Koch, W.H., Huang, K.X., and Glazer, R.I. (1996). Transcriptional regulation of c-Fes in myeloid leukemia cells. *Biochimica et biophysica acta* 1306, 179-186.

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.

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development* 139, 1045-1057.

Hersh, L.B., and Shimojo, M. (2003). Regulation of cholinergic gene expression by the neuron restrictive silencer factor/repressor element-1 silencing transcription factor. *Life sciences* 72, 2021-2028.

Herz, H.M., Hu, D., and Shilatifard, A. (2014). Enhancer malfunction in cancer. *Molecular cell* 53, 859-866.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934-947.

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.

Hu, D., Gao, X., Morgan, M.A., Herz, H.M., Smith, E.R., and Shilatifard, A. (2013a). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology* 33, 4745-4754.

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T.M., Muljo, S.A., Zhu, J., and Zhao, K. (2013b). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nature immunology* 14, 1190-1198.

Huang, D., Petrykowska, H.M., Miller, B.F., Elnitski, L., and Ovcharenko, I. (2019). Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome research* 29, 657-667.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.C., *et al.* (2016). Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Developmental cell* 36, 9-23.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC genomics* 8, 39.

Hwang, J.-Y., Kaneko, N., Noh, K.-M., Pontarelli, F., and Zukin, R.S. (2014). The gene silencing transcription factor REST represses miR-132 expression in hippocampal neurons destined to die. *Journal of molecular biology* 426, 3454-3466.

Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics* 106, 159-164.

Jiang, H., and Peterlin, B.M. (2008). Differential chromatin looping regulates CD4 expression in immature thymocytes. *Molecular and cellular biology* 28, 907-912.

Jiang, H., Zhang, F., Kurosu, T., and Peterlin, B.M. (2005). Runx1 Binds Positive Transcription Elongation Factor b and Represses Transcriptional Elongation by RNA Polymerase II: Possible Mechanism of CD4 Silencing. *Molecular and cellular biology* 25, 10675-10683.

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.

Johnson, W.C., Ordway, A.J., Watada, M., Pruitt, J.N., Williams, T.M., and Rebeiz, M. (2015). Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between *Drosophila* Species. *PLoS genetics* *11*.

Kadesch, T., Zervos, P., and Ruezinsky, D. (1986). Functional analysis of the murine IgH enhancer: evidence for negative control of cell-type specificity. *Nucleic acids research* *14*, 8209-8221.

Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S., and Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nature structural & molecular biology* *18*, 142-149.

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.

Kappes, D.J., He, X., and He, X. (2005). CD4-CD8 lineage commitment: an inside view. *Nat Immunol* *6*, 761-766.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome research* *12*, 996-1006.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* *26*, 2204-2207.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome research* *23*, 800-811.

Killeen, N., and Littman, D.R. (1993). Helper T-Cell Development in the Absence of CD4p56lck Association. *Nature* *364*, 729-732.

Kim, M.K., Lesoon-Wood, L.A., Weintraub, B.D., and Chung, J.H. (1996). A soluble transcription factor, Oct-1, is also found in the insoluble nuclear matrix and possesses silencing activity in its alanine-rich domain. *Molecular and cellular biology* *16*, 4366-4377.

Kioussis, D., and Ellmeier, W. (2002). Chromatin and CD4, CD8A and CD8B gene expression during thymic differentiation. *Nat Rev Immunol* *2*, 909-919.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., *et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature structural & molecular biology* *18*, 956-963.

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science* *356*, 194-197.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* *30*, 235-239.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature genetics* *42*, 631-634.

Kundu, M., Compton, S., Garrett-Beal, L., Stacy, T., Starost, M.F., Eckhaus, M., Speck, N.A., and Liu, P.P. (2005). Runx1 deficiency predisposes mice to T-lymphoblastic lymphoma. *Blood* *106*, 3621-3624.

Kuwahara, K., Saito, Y., Takano, M., Arai, Y., Yasuno, S., Nakagawa, Y., Takahashi, N., Adachi, Y., Takemura, G., Horie, M., *et al.* (2003). NRSF regulates the fetal cardiac gene program and maintains normal cardiac structure and function. *The EMBO journal* *22*, 6310-6321.

Kwasnieski, J.C., Fiore, C., Chaudhari, H.G., and Cohen, B.A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome research* *24*, 1595-1602.

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.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* **172**, 650-665.

Landry, D.B., Engel, J.D., and Sen, R. (1993). Functional GATA-3 binding sites within murine CD8 alpha upstream regulatory sequences. *J Exp Med* **178**, 941-949.

Le Clech, M., Chalhoub, E., Dohet, C., Roure, V., Fichelson, S., Moreau-Gachelin, F., and Mathieu, D. (2006). PU.1/Spi-1 binds to the human TAL-1 silencer to mediate its activity. *J Mol Biol* **355**, 9-19.

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, M.S., Hanspers, K., Barker, C.S., Korn, A.P., and McCune, J.M. (2004). Gene expression profiles during human CD4 + T cell differentiation. *Int Immunol* **16**, 1109-1124.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature reviews Genetics* **13**, 233-245.

Leung, R.K., Thomson, K., Gallimore, A., Jones, E., Van den Broek, M., Sierro, S., Alsheikhly, A.R., McMichael, A., and Rahemtulla, A. (2001). Deletion of the CD4 silencer element supports a stochastic mechanism of thymocyte lineage commitment. *Nature immunology* **2**, 1167-1173.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harbor perspectives in biology* **6**, a019133.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., *et al.* (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98.

Li, L.M., and Arnosti, D.N. (2011). Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. *Current biology : CB* **21**, 406-412.

Li, Q., Lewandowski, J.P., Powell, M.B., Norrie, J.L., Cho, S.H., and Vokes, S.A. (2014). A Gli silencer is required for robust repression of gremlin in the vertebrate limb bud. *Development* **141**, 1906-1914.

Li, Q., Zhang, M., Duan, Z., and Stamatoyannopoulos, G. (1999). Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics* **61**, 183-193.

Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M., and Hoffman, A.R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269-272.

Liu, B., Maul, R.S., and Kaetzel, D.M., Jr. (1996). Repression of platelet-derived growth factor A-chain gene transcription by an upstream silencer element. Participation by sequence-specific single-stranded DNA-binding proteins. *The Journal of biological chemistry* **271**, 26281-26290.

Liu, H., Li, H., Jin, L., Li, G., Hu, S., Ning, C., Guo, J., Shuai, S., Li, X., and Li, M. (2018). Long Noncoding RNA GAS5 Suppresses 3T3-L1 Cells Adipogenesis Through miR-21a-5p/PTEN Signal Pathway. *DNA and cell biology* **37**, 767-777.

Liu, J., and Francke, U. (2006). Identification of cis-regulatory elements for MECP2 expression. *Hum Mol Genet* **15**, 1769-1782.

Liu, M., Maurano, M.T., Wang, H., Qi, H., Song, C.Z., Navas, P.A., Emery, D.W., Stamatoyannopoulos, J.A., and Stamatoyannopoulos, G. (2015). Genomic discovery of potent chromatin insulators for human gene therapy. *Nat Biotechnol* **33**, 198-203.

Liu, Y., Yu, S., Dhiman, V.K., Brunetti, T., Eckart, H., and White, K.P. (2017). Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol* **18**, 219.

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.

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*.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews Molecular cell biology* **13**, 436-447.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* *339*, 823-826.

Martin, D., Tawadros, T., Meylan, L., Abderrahmani, A., Condorelli, D.F., Waeber, G., and Haefliger, J.A. (2003). Critical role of the transcriptional repressor neuron-restrictive silencer factor in the specific control of connexin36 in insulin-producing cell lines. *The Journal of biological chemistry* *278*, 53082-53089.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* *7*, 29-59.

Mathieu, E.L., Belhocine, M., Dao, L.T., Puthier, D., and Spicuglia, S. (2014). [Functions of lncRNA in development and diseases]. *Med Sci (Paris)* *30*, 790-796.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci USA* *101*, 1525-1530.

Medina-Rivera, A., Santiago-Algarra, D., Puthier, D., and Spicuglia, S. (2018). Widespread Enhancer Activity from Core Promoters. *Trends Biochem Sci*.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B., *et al.* (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* *30*, 271-277.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual review of genomics and human genetics* *17*, 17-43.

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.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* *448*, 553-560.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *NatCell Biol* *10*, 483-488.

Moffat, G.J., McLaren, A.W., and Wolf, C.R. (1996). Functional characterization of the transcription silencer element located within the human Pi class glutathione S-transferase promoter. *The Journal of biological chemistry* *271*, 20740-20747.

Mombaerts, P., Terhorst, C., Jacks, T., Tonegawa, S., and Sancho, J. (1995). Characterization of immature thymocyte lines derived from T-cell receptor or recombination activating gene 1 and p53 double mutant mice. *Proc Natl Acad Sci U S A* *92*, 7420-7424.

Moroy, T., and Khandanpour, C. (2011). Growth factor independence 1 (Gfi1) as a regulator of lymphocyte development and activation. *Seminars in immunology* *23*, 368-378.

Moyal, L., Lerenthal, Y., Gana-Weisz, M., Mass, G., So, S., Wang, S.Y., Eppink, B., Chung, Y.M., Shalev, G., Shema, E., *et al.* (2011). Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Molecular cell* *41*, 529-542.

Muerdter, F., Boryn, L.M., and Arnold, C.D. (2015). STARR-seq - Principles and applications. *Genomics*.

Mukhopadhyay, S., Schedl, P., Studitsky, V.M., and Sengupta, A.M. (2011). Theoretical analysis of the role of chromatin interactions in long-range action of enhancers and insulators. *Proc Natl Acad Sci U S A* *108*, 19919-19924.

Murtha, M., Tokcaer-Keskin, Z., Tang, Z., Strino, F., Chen, X., Wang, Y., Xi, X., Basilico, C., Brown, S., Bonneau, R., *et al.* (2014). FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat Methods* *11*, 559-565.

Nagasawa, T., Takeda, T., Minemura, K., and DeGroot, L.J. (1997). Oct-1, silencer sequence, and GC box regulate thyroid hormone receptor beta1 promoter. *Molecular and cellular endocrinology* *130*, 153-165.

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* *31*, 1120-1123.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* 46, 1-19.

Nguyen, T.A., Jones, R.D., Snavely, A.R., Pfenning, A.R., Kirchner, R., Hemberg, M., and Gray, J.M. (2016). High-throughput functional comparison of promoter and enhancer activities. *Genome research* 26, 1023-1033.

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.

Ogbourne, S., and Antalis, T.M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *The Biochemical journal* 331 ( Pt 1), 1-14.

Oh, C.K., Neurath, M., Cho, J.J., Semere, T., and Metcalfe, D.D. (1997). Two different negative regulatory elements control the transcription of T-cell activation gene 3 in activated mast cells. *The Biochemical journal* 323 ( Pt 2), 511-519.

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.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature reviews Genetics* 15, 234-246.

Ooi, L., and Wood, I.C. (2007). Chromatin crosstalk in development and disease: lessons from REST. *Nature reviews Genetics* 8, 544-554.

Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. *Cell* 133, 1145-1148.

Padeken, J., and Heun, P. (2014). Nucleolus and nuclear periphery: velcro for heterochromatin. *Curr Opin Cell Biol* 28, 54-60.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* 144, 200-213.

Papantonis, A., and Cook, P.R. (2013). Transcription factories: genome organization and gene regulation. *Chemical reviews* 113, 8683-8705.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., *et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422-433.

Park, J.H., Adoro, S., Guinter, T., Erman, B., Alag, A.S., Catalfamo, M., Kimura, M.Y., Cui, Y., Lucas, P.J., Gress, R.E., *et al.* (2010). Signaling by intrathymic cytokines, not T cell antigen receptors, specifies CD8 lineage choice and promotes the differentiation of cytotoxic-lineage T cells. *Nature immunology* 11, 257-264.

Patil, V.S., Zhou, R., and Rana, T.M. (2014). Gene regulation by noncoding RNAs. *Critical reviews in biochemistry and molecular biology* 49, 16-32.

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.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* 125, 703-717.

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, Y., and Zhang, Y. (2018). Enhancer and super-enhancer: Positive regulators in gene transcription. *Animal models and experimental medicine* 1, 169-179.

Petrie, H.T., and Zuniga-Pflucker, J.C. (2007). Zoned out: functional mapping of stromal signaling microenvironments in the thymus. *Annual review of immunology* 25, 649-679.

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.

Ptashne, M. (2011). Principles of a switch. *Nature chemical biology* 7, 484-487.

Qi, H., Liu, M., Emery, D.W., and Stamatoyannopoulos, G. (2015). Functional validation of a constitutive autonomous silencer element. *PloS one* 10, e0124588.

Qin, J.Y., Zhang, L., Clift, K.L., Hular, I., Xiang, A.P., Ren, B.Z., and Lahn, B.T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PloS one* 5, e10611.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

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.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* 42, W187-191.

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.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* 99, 6883-6888.

Rincon-Limas, D.E., Amaya-Manzanares, F., Nino-Rosales, M.L., Yu, Y., Yang, T.P., and Patel, P.I. (1995). Ubiquitous and neuronal DNA-binding proteins interact with a negative regulatory element of the human hypoxanthine phosphoribosyltransferase gene. *Molecular and cellular biology* 15, 6561-6571.

Robson, M.I., Ringel, A.R., and Mundlos, S. (2019). Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Molecular cell* 74, 1110-1122.

Rothenberg, E.V. (2014). Transcriptional control of early T and B cell developmental choices. *Annual review of immunology* 32, 283-321.

Rothenberg, E.V., Moore, J.E., and Yui, M.A. (2008). Launching the T-cell-lineage developmental programme. *NatRevImmunol* 8, 9-21.

Rui, J., Liu, H., Zhu, X., Cui, Y., and Liu, X. (2012). Epigenetic silencing of CD8 genes by ThPOK-mediated deacetylation during CD4 T cell differentiation. *J Immunol* 189, 1380-1390.

Saadi, W., Kermezli, Y., Dao, L.T.M., Mathieu, E., Santiago-Algarra, D., Manosalva, I., Torres, M., Belhocine, M., Pradel, L., Loriod, B., *et al.* (2019). A critical regulator of Bcl2 revealed by systematic transcript discovery of lncRNAs associated with T-cell differentiation. *Scientific Reports* 9, 4707.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PloS one* 7, e30733.

Santiago-Algarra, 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.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109-113.

Sawada, S., Scarborough, J.D., Killeen, N., and Littman, D.R. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* 77, 917-929.

Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* 396, 311-327.

Schluesche, P., Stelzer, G., Piaia, E., Lamb, D.C., and Meisterernst, M. (2007). NC2 mobilizes TBP on core promoter TATA boxes. *Nature structural & molecular biology* 14, 1196-1201.

SCHLUTER, S.F., and MARCHALONIS, J.J. (2003). Cloning of shark RAG2 and characterization of the RAG1/RAG2 gene locus. *The FASEB Journal* 17, 470-472.

Schoch, S., Cibelli, G., and Thiel, G. (1996). Neuron-specific gene expression of synapsin I. Major role of a negative regulatory mechanism. *The Journal of biological chemistry* *271*, 3317-3323.

Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* *267*, 1360-1363.

Scholl, T., Stevens, M.B., Mahanta, S., and Strominger, J.L. (1996). A zinc finger protein that represses transcription of the human MHC class II gene, DPA. *J Immunol* *156*, 1448-1457.

Schuster, M., Annemann, M., Plaza-Sirvent, C., and Schmitz, I. (2013). Atypical I $\kappa$ B proteins - nuclear modulators of NF- $\kappa$ B signaling. *Cell communication and signaling : CCS* *11*, 23.

Schuster, M., Glaben, R., Plaza-Sirvent, C., Schreiber, L., Annemann, M., Floess, S., Kuhl, A.A., Clayton, L.K., Sparwasser, T., Schulze-Osthoff, K., *et al.* (2012). I $\kappa$ B(NS) protein mediates regulatory T cell development via induction of the Foxp3 transcription factor. *Immunity* *37*, 998-1008.

Sertil, O., Kapoor, R., Cohen, B.D., Abramova, N., and Lowry, C.V. (2003). Synergistic repression of anaerobic genes by Mot3 and Rox1 in *Saccharomyces cerevisiae*. *Nucleic acids research* *31*, 5831-5837.

Setoguchi, R., Tachibana, M., Naoe, Y., Muroi, S., Akiyama, K., Tezuka, C., Okuda, T., and Taniuchi, I. (2008). Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. *Science* *319*, 822-825.

Shen, S.Q., Myers, C.A., Hughes, A.E., Byrne, L.C., Flannery, J.G., and Corbo, J.C. (2016). Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome research* *26*, 238-255.

Shimokawa, T., and Fujimoto, H. (1996). Identification of a transcriptional silencer in the protein-coding region of the mouse major inducible Hsp70 gene. *Biochemical and biophysical research communications* *221*, 843-848.

Shlyueva, D., Stampfel, G., and Stark, A. (2014a). Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews Genetics* *15*, 272-286.

Shlyueva, D., Stelzer, C., Gerlach, D., Yanez-Cuna, J.O., Rath, M., Boryn, L.M., Arnold, C.D., and Stark, A. (2014b). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Molecular cell* *54*, 180-192.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* *21*, 344-351.

Singh, H., Khan, A.A., and Dinner, A.R. (2014). Gene regulatory networks in the immune system. *Trends in immunology* *35*, 211-218.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annual review of biochemistry* *72*, 449-479.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature structural & molecular biology* *21*, 210-219.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature genetics* *45*, 1021-1028.

Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. *Biochimica et biophysica acta* *1829*, 84-97.

Snetkova, V., and Skok, J.A. (2018). Enhancer talk. *Epigenomics* *10*, 483-498.

Song, S.H., Hou, C., and Dean, A. (2007). A Positive Role for NLI/Ldb1 in Long-Range  $\gamma$ -Globin Locus Control Region Function. *Molecular cell* *28*, 810-822.

Song, Z., Zhao, D., Zhao, H., and Yang, L. (2015). NRSF: an angel or a devil in neurogenesis and neurological diseases. *Journal of molecular neuroscience : MN* *56*, 131-144.

Spedale, G., Meddens, C.A., Koster, M.J., Ko, C.W., van Hooff, S.R., Holstege, F.C., Timmers, H.T., and Pijnappel, W.W. (2012). Tight cooperation between Mot1p and NC2beta in regulating genome-wide transcription, repression of transcription following heat shock induction and genetic interaction with SAGA. *Nucleic acids research* *40*, 996-1008.



Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. *Nature reviews Genetics* *13*, 613-626.

Sprouse, R.O., Karpova, T.S., Mueller, F., Dasgupta, A., McNally, J.G., and Auble, D.T. (2008). Regulation of TATA-binding protein dynamics in living yeast cells. *Proc Natl Acad Sci U S A* *105*, 13304-13308.

Sprouse, R.O., Wells, M.N., and Auble, D.T. (2009). TATA-binding protein variants that bypass the requirement for Mot1 in vivo. *The Journal of biological chemistry* *284*, 4525-4535.

Starr, T.K., Jameson, S.C., and Hogquist, K.A. (2003). Positive and negative selection of T cells. *Annual review of immunology* *21*, 139-176.

Steiner, C., Muller, M., Baniahmad, A., and Renkawitz, R. (1987). Lysozyme gene activity in chicken macrophages is controlled by positive and negative regulatory elements. *Nucleic acids research* *15*, 4163-4178.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.

Sun, G., Liu, X., Mercado, P., Jenkinson, S.R., Kyriiotou, M., Feigenbaum, L., Galera, P., and Bosselut, R. (2005). The zinc finger protein cKrox directs CD4 lineage differentiation during intrathymic T cell positive selection. *Nat Immunol* *6*, 373-381.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* *324*, 930-935.

Takimoto, M., Quinn, J.P., Farina, A.R., Staudt, L.M., and Levens, D. (1989). fos/jun and octamer-binding protein interact with a common site in a negative element of the human c-myc gene. *The Journal of biological chemistry* *264*, 8992-8999.

Tanaka, H., Naito, T., Muroi, S., Seo, W., Chihara, R., Miyamoto, C., Kominami, R., and Taniuchi, I. (2013). Epigenetic Thpok silencing limits the time window to choose CD4(+) helper-lineage fate in the thymus. *The EMBO journal* *32*, 1183-1194.

Taniuchi, I., Ellmeier, W., and Littman, D.R. (2004). The CD4/CD8 lineage choice: new insights into epigenetic regulation during T cell development. *Adv Immunol* *83*, 55-89.

Taniuchi, I., and Littman, D.R. (2004). Epigenetic gene silencing by Runx proteins. *Oncogene* *23*, 4341-4345.

Taniuchi, I., Osato, M., Egawa, T., Sunshine, M.J., Bae, S.C., Komori, T., Ito, Y., and Littman, D.R. (2002a). Differential requirements for RUNX proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* *111*, 621-633.

Taniuchi, I., Sunshine, M.J., Festenstein, R., and Littman, D.R. (2002b). Evidence for distinct CD4 silencer functions at different stages of thymocyte differentiation. *Molecular cell* *10*, 1083-1096.

Teitell, M.A., and Pandolfi, P.P. (2009). Molecular genetics of acute lymphoblastic leukemia. *Annual review of pathology* *4*, 175-198.

Telfer, J.C., Hedblom, E.E., Anderson, M.K., Laurent, M.N., and Rothenberg, E.V. (2004). Localization of the domains in runx transcription factors required for the repression of CD4 in thymocytes. *J Immunol* *172*, 4359-4370.

Terry, S., and Beltran, H. (2014). The many faces of neuroendocrine differentiation in prostate cancer progression. *Frontiers in oncology* *4*, 60.

Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., *et al.* (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell* *165*, 1519-1529.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology* *41*, 105-178.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., *et al.* (2012). The accessible chromatin landscape of the human genome. *Nature* *489*, 75-82.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS biology* *6*, 2911-2927.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active  $\gamma$ -globin locus. *Molecular cell* 10, 1453-1465.

Tsuiji, H., Yoshimoto, R., Hasegawa, Y., Furuno, M., Yoshida, M., and Nakagawa, S. (2011). Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1. *Genes to cells : devoted to molecular & cellular mechanisms* 16, 479-490.

Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S., *et al.* (2016). Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* 165, 1530-1545.

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 Laethem, F., Sarafova, S.D., Park, J.H., Tai, X., Pobezinsky, L., Guintier, T.I., Adoro, S., Adams, A., Sharrow, S.O., Feigenbaum, L., *et al.* (2007). Deletion of CD4 and CD8 coreceptors permits generation of alphabetaT cells that recognize antigens independently of the MHC. *Immunity* 27, 735-750.

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.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* 489, 452-455.

Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Jr., and Reddy, T.E. (2015). Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome research* 25, 1206-1214.

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, X., He, L., Goggin, S.M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., and Kellis, M. (2018). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun* 9, 5380.

Wang, X., Xiao, G., Zhang, Y., Wen, X., Gao, X., Okada, S., and Liu, X. (2008a). Regulation of Tcr $\beta$  recombination ordering by c-Fos-dependent RAG deposition. *Nature immunology* 9, 794-801.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., *et al.* (2008b). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* 40, 897-903.

Wang, Z.Y., Masaharu, N., Qiu, Q.Q., Takimoto, Y., and Deuel, T.F. (1994). An S1 nuclease-sensitive region in the first intron of human platelet-derived growth factor A-chain gene contains a negatively acting cell type-specific regulatory element. *Nucleic acids research* 22, 457-464.

Wei, X.C., Dohkan, J., Kishi, H., Wu, C.X., Kondo, S., and Muraguchi, A. (2005). Characterization of the proximal enhancer element and transcriptional regulatory factors for murine recombination activating gene-2. *Eur J Immunol* 35, 612-621.

Weingarten-Gabbay, S., Nir, R., Lubliner, S., Sharon, E., Kalma, Y., Weinberger, A., and Segal, E. (2017). Deciphering Transcriptional Regulation of Human Core Promoters. *bioRxiv*, 174904.

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.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., *et al.* (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* *171*, 1573-1588 e1528.

Weirauch, M.T., and Hughes, T.R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends in genetics : TIG* *26*, 66-74.

West, A.G., and Fraser, P. (2005). Remote control of gene transcription. *HumMolGenet* *14 Spec No 1*, R101-R111.

White, M.A. (2015). Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* *106*, 165-170.

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.

Wildt, K.F., Sun, G., Grueter, B., Fischer, M., Zamisch, M., Ehlers, M., and Bosselut, R. (2007). The transcription factor Zbtb7b promotes CD4 expression by antagonizing Runx-mediated activation of the CD4 silencer. *J Immunol* *179*, 4405-4414.

Williams, S.K., Truong, D., and Tyler, J.K. (2008). Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A* *105*, 9000-9005.

Williams, T.M., Moolten, D., Burlein, J., Romano, J., Bhaerman, R., Godillot, A., Mellon, M., Rauscher, F.J., 3rd, and Kant, J.A. (1991). Identification of a zinc finger protein that inhibits IL-2 gene expression. *Science* *254*, 1791-1794.

Wilusz, J.E. (2016). Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochimica et biophysica acta* *1859*, 128-138.

Woolf, E., Xiao, C., Fainaru, O., Lotem, J., Rosen, D., Negreanu, V., Bernstein, Y., Goldenberg, D., Brenner, O., Berke, G., *et al.* (2003). Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. *Proc Nat Acad Sci USA* *100*, 7731-7736.

Wu, G.D., Lai, E.J., Huang, N., and Wen, X. (1997). Oct-1 and CCAAT/enhancer-binding protein (C/EBP) bind to overlapping elements within the interleukin-8 promoter. The role of Oct-1 as a transcriptional repressor. *The Journal of biological chemistry* *272*, 2396-2403.

Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* *25*, 2436-2452.

Wu, J., Song, Y., Bakker, A.B., Bauer, S., Spies, T., Lanier, L.L., and Phillips, J.H. (1999). An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* *285*, 730-732.

Xing, S., Shao, P., Li, F., Zhao, X., Seo, W., Wheat, J.C., Ramasamy, S., Wang, J., Li, X., Peng, W., *et al.* (2018). The corepressors are differentially partitioned to instruct CD8(+) T cell lineage choice and identity. *J Exp Med* *215*, 2211-2226.

Yadav, D.K., Shrestha, S., Dadhwal, G., and Chandak, G.R. (2018). Identification and characterization of cis-regulatory elements 'insulator and repressor' in PPAR $\delta$  gene. *Epigenomics* *10*, 613-627.

Yajima, M., Kiyomoto, M., and Akasaka, K. (2007). An insulator protects transgenes from long-term silencing in sea urchin larva. *Development genes and evolution* *217*, 331-336.

Yan, J., Chen, S.A., Local, A., Liu, T., Qiu, Y., Dorighi, K.M., Preissl, S., Rivera, C.M., Wang, C., Ye, Z., *et al.* (2018). Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell research* *28*, 387.

Yanez-Cuna, J.O., Arnold, C.D., Stampfel, G., Boryn, L.M., Gerlach, D., Rath, M., and Stark, A. (2014). Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome research* *24*, 1147-1156.

Yang, L., Froberg, J.E., and Lee, J.T. (2014). Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci* *39*, 35-43.

Yannoutsos, N., Barreto, V., Misulovin, Z., Gazumyan, A., Yu, W., Rajewsky, N., Peixoto, B.R., Eisenreich, T., and Nussenzweig, M.C. (2004). A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nature immunology* 5, 443-450.

Yao, X., Nie, H., Rojas, I.C., Harriss, J.V., Maika, S.D., Gottlieb, P.D., Rathbun, G., and Tucker, P.W. (2010). The L2a element is a mouse CD8 silencer that interacts with MAR-binding proteins SATB1 and CDP. *Mol Immunol* 48, 153-163.

Ye, J., Cippitelli, M., Dorman, L., Ortaldo, J.R., and Young, H.A. (1996). The nuclear factor YY1 suppresses the human gamma interferon promoter through two mechanisms: inhibition of AP1 binding and activation of a silencer element. *Molecular and cellular biology* 16, 4744-4753.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202-206.

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.

Zhang, J., Wang, S., Yuan, L., Yang, Y., Zhang, B., Liu, Q., Chen, L., Yue, W., Li, Y., and Pei, X. (2012). Neuron-restrictive silencer factor (NRSF) represses cocaine- and amphetamine-regulated transcript (CART) transcription and antagonizes cAMP-response element-binding protein signaling through a dual NRSE mechanism. *The Journal of biological chemistry* 287, 42574-42587.

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.

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes & development* 28, 2699-2711.

Zhu, Y., van Essen, D., and Sacconi, S. (2012). Cell-type-specific control of enhancer activity by H3K9 trimethylation. *Molecular cell* 46, 408-423.

## **Annex: CRISPR/cas9 Screening**

**CRISPR/cas9 screening table: Numbers of tested clones**

<b>Silencer candidates</b>	<b>Tested clones</b>	<b>Heterozygous</b>	<b>Homozygous</b>
Sdhaf1	135	6 (4.44%)	2
Dpp9	141	7 (4.96%)	2*
Midn	102	4 (3.98%)	0
Kif27	138	0	0

**\*Clones lost.**