BIOINFORMATIC ANALYSIS OF EPIGENETIC REGULATORY MECHANISMS IN DEVELOPMENT AND DISEASE

Mar González Ramírez

TESI DOCTORAL UPF / 2020

Directors de la tesi:

Dr. Luciano Di Croce

Dr. Enrique Blanco

EPIGENETIC EVENTS IN CANCER GROUP, GENE REGULATION, STEM CELLS AND CANCER PROGRAM, CENTRE FOR GENOMIC REGULATION (CRG)

UNIVERSITAT POMPEU FABRA, DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

ABSTRACT (English)

Appropriate regulation of gene expression is necessary for correct development and homeostasis of organisms. Epigenetic mechanisms represent an additional layer of information, besides the genetic sequence, crucial for the correct functioning of each cell. Histone modifications, which modulate and are associated to transcriptional activation or repression, are a major epigenetic feature. Thanks to predictive modelling, we have studied which histone modifications relate better to enhancer or promoter function in mouse embryonic stem cells, during differentiation and in animal development. We have found that different histone modifications relate better to enhancers or promoters, respectively. We have studied the role of poised enhancers during differentiation and development. We have seen that poised enhancer activation is not exclusive of the neural lineage, but a general mechanism implicated in differentiation of every cell type. We have characterized the epigenetic landscape of Cushing's syndrome. We have found persistent epigenetic and transcriptional alterations after long-term remission of the disease, related to a deep alteration of the circadian rhythm. These findings promise to be relevant for future therapeutic advances.

RESUM (Català)

Una regulació apropiada de l'expressió gènica és necessària per a un correcte desenvolupament i homeòstasi dels organismes. Els mecanismes epigenètics representen una informació addicional, a més de la seqüència genètica, crucial per al correcte funcionament de cada cèl·lula. Les modificacions d'histones, que modulen i s'associen a activació o repressió transcripcionals, són una característica epigenètica important. Gràcies al modelatge predictiu, hem estudiat quines modificacions d'histones es relacionen millor amb la funció dels *enhancers* o promotors en cèl·lules mare embrionàries de ratolí, durant la diferenciació i en el desenvolupament animal. Hem trobat que modificacions d'histones diferents es relacionen millor amb *enhancers* o promotors, respectivament. Hem estudiat el rol dels *poised enhancers* durant la diferenciació i el desenvolupament. Hem vist que l'activació dels *poised enhancers* no és exclusiva del llinatge neural, sinó un mecanisme implicat en la diferenciació de tot tipus cel·lular. Hem caracteritzat el paisatge epigenètic de la síndrome de Cushing. Hem trobat alteracions epigenètiques i transcripcionals després d'una remissió de la malaltia a llarg termini, relacionades amb una profunda alteració del ritme circadiari. Aquestes troballes prometen ser rellevants per a futurs avenços terapèutics.

TABLE OF CONTENTS

INTRODUCTION

If the present arrangements of society will not admit of woman's free development, then society must be remodelled.

Elizabeth Blackwell (1821-1910)

1. Chromatin and gene regulation

1.1 The basic structure of chromatin

Chromatin is the structure in which the genome is packed inside of the nucleus of a eukaryotic cell. It is a complex made of proteins and DNA. Nucleosomes, the basic repetitive units of the chromatin, consist of around 200 bp of DNA and five types of histone proteins [1-4]. The nucleosome core is formed by 147 bp of DNA wrapped around a histone octamer, which is constituted by two copies of the four core histones: H2A, H2B, H3 and H4. The linker histone H1 binds the linker DNA that connects the nucleosome cores.

According to its level of compaction, chromatin can be divided into euchromatin and heterochromatin. While euchromatin is open and highly transcribed, heterochromatin has a high level of compaction and gene expression inside is limited [5]. Chromatin is indeed a dynamic structure involved in several processes such as gene regulation [6], DNA repair [7] and DNA replication [8].

1.2 The concept of Epigenetics

The term epigenetics was originally defined in 1942 as the study of the causal mechanisms that produce the phenotype from the genotype in a developmental context [9]. Nowadays, it is widely accepted that this term refers to the inheritance of different chromatin states without affecting the DNA sequence [10]. Therefore, epigenetics corresponds to an additional layer of information, besides the genetic sequence, important for the correct functioning of each cell. Examples of epigenetic mechanisms are DNA methylation [11], histone post-translational modifications [12] and non-coding RNAs (ncRNAs) [13]. DNA methylation, the addition of methyl groups to the DNA molecule, is generally associated with gene silencing and usually occurs at CpG islands. ncRNAs usually modify chromatin and target gene expression through pathways of RNA interference, and can also act as recruiters of chromatin modifying enzymes.

1.2.1 Histone post-translational modifications

Histones can be chemically modified at the N-terminal tail by specific enzymes. Several types of post-translational modifications have been identified so far, including methylation, acetylation, phosphorylation and ubiquitylation [12]. These modifications –also called histone marks– affect

inter-nucleosomal interactions that ultimately alter the chromatin structure, and also facilitate or prevent the binding of several factors and protein complexes [12]. Certain modifications have been associated to an active transcriptional state of the chromatin, such as trimethylation of histone H3 at lysine 4 (H3K4me3) [14-16], acetylation of histone H3 at lysine 27 (H3K27ac) [17] and trimethylation of histone H3 at lysine 36 (H3K36me3) [18, 19]. Other histone marks have been associated to gene repression instead: trimethylation of histone H3 at lysine 27 (H3K27me3) [20] and trimethylation of histone H3 at lysine 9 (H3K9me3) [21]. Moreover, during the last decade, several studies have shown that gene expression and histone modification signals are quantitatively related [22-32].

1.2.2 ChIP-seq and chromatin segmentation

Chromatin immunoprecipitation (ChIP) is the principal tool to assess protein–DNA binding *in vivo* [33, 34]. In here, antibodies recognizing specific proteins –or even chemical modifications of them, such as histone modifications– are employed to enrich for DNA fragments that are precisely bound to these proteins. ChIP followed by massive sequencing (ChIP-seq) allows the genome–wide identification of these sequences [35-38]. Thus, ChIP-seq is the main technique to profile DNA-binding proteins, histone modifications, etc., and it

therefore becomes a fundamental tool to understand gene regulation and epigenetic mechanisms [39, 40].

The computational analysis of ChIP-seq experiments is divided into two different steps: mapping and peak calling. The mapping tool is in charge of processing the raw data to identify the exact position of each read in a chromosome of the genome. Specific software has been developed to perform this step, such as BOWTIE [41], BWA [42, 43] or GEM [44]. Peak calling, instead, consists in the detection of genomic regions presenting a statistically significant high enrichment of reads in a certain experiment (named peaks) that are not observed in a second sample of control. A high variety of peak callers are available, such as MACS [45], SICER [46] and HOMER [47]. Next, differential peak analysis between ChIP-seq experiments on two different conditions can be used to identify specific peaks of any of them. DiffBind [48] is an R package widely used to perform differential analysis of ChIP-seq peaks. Once the location of the ChIP-seq peaks of one or more experiments is uncovered, further bioinformatic analysis is necessary to be performed: genome distribution of peaks, putative target genes, functional analysis, motif enrichment, occupancy meta-plots, heat maps, etc.

Finally, chromatin segmentation methods are useful to infer novel knowledge from the combination of multiple ChIP-seq experiments. Briefly, chromatin segmentation takes as input

multiple sources of ChIP-seq data and generate a division of the genome into segments that are assigned to a specific chromatin state. The signature of a chromatin state is defined by the particular configuration of ChIP-seq features that are present in this class of segment. The most widely used software for chromatin segmentation is ChromHMM [49]. This software is based on a multivariate Hidden Markov Model (HMM) that explicitly models the presence or absence of each chromatin feature. In ChromHMM, state emissions are defined by probability to find a specific feature in the regions which belong to that particular state.

1.3 3D organization of the chromatin

Another important aspect of chromatin is its 3D distribution inside of the nucleus, which is thought to influence gene regulation by bringing together distant regions of the genome [50]. Remarkably, physical interactions tend to occur between regions with same chromatin features [51, 52].

The genome can be classified into two compartments according to the pattern of physical interactions [53]. The A compartment, which is associated to transcriptional activation of genes and resembles to euchromatin; and the B compartment, which is associated to gene silencing and resembles to heterochromatin. Moreover, the genome can

further be divided into topologically associating domains (TADs), which are megabase-sized domains that preferentially interact within themselves and are generally maintained across cell types and species [54]. TAD boundaries serve as insulators to prevent inter-TAD interactions [55, 56].

1.4 Regulatory regions

A correct regulation of gene expression is necessary to establish the appropriate developmental programs and homeostasis of organisms. Thus, the genome contains thousands of functional non-coding DNA sequences that regulate gene transcription, such as promoters, enhancers, silencers and insulators [57]. Promoters are elements that initiate transcription and are located in the surroundings of a transcription start site (TSS) of a gene. Enhancers are DNA sequences that substantially amplify the expression of genes and can be located in distal areas up to approximately 1 Mb from their target gene. Silencers act similarly to enhancers but instead of amplifying gene expression, they diminish it. Finally, insulators are boundary elements that prevent enhancers and silencers from modulating gene expression from genes in the other side of the boundary. Moreover, regulatory regions are susceptible of undergoing epigenetic changes such as histone modifications that will make them more or less accessible to

transcription factors (TFs), which will decisively influence gene expression [58].

1.4.1 Enhancers

The term “enhancer” first appeared in 1981 [59], although by that time it was already known that regions of open chromatin far from promoters existed [60]. Originally, the term referred to non-coding regions that could enhance transcription in a reporter assay at any orientation in many positions, even downstream of the TSS [59, 61]. However, nowadays multiple genomic characteristics have been identified to be common to most enhancers and serve as markers to identify enhancer candidates *in silico* [57, 62]. Some of these features are (Figure I1): (i) open regions free of nucleosomes [63]; (ii) monomethylation of histone H3 at lysine 4 (H3K4me1) [64]; (iii) H3K27ac, which distinguishes active from inactive enhancers [65]; (iv) a high density of TF binding and presence of TF motifs [66]; and (v) enhancer transcription [67]. Therefore, the results of next generation sequencing techniques such as DHS-seq, ATAC-seq, ChIP-seq, GRO-seq, etc. that capture these common features can be used to identify putative enhancers genome-wide. Next, the enhancer activity of the most interesting candidates can be validated by reporter assays or CRISPR-Cas9 genome engineering [57, 62].

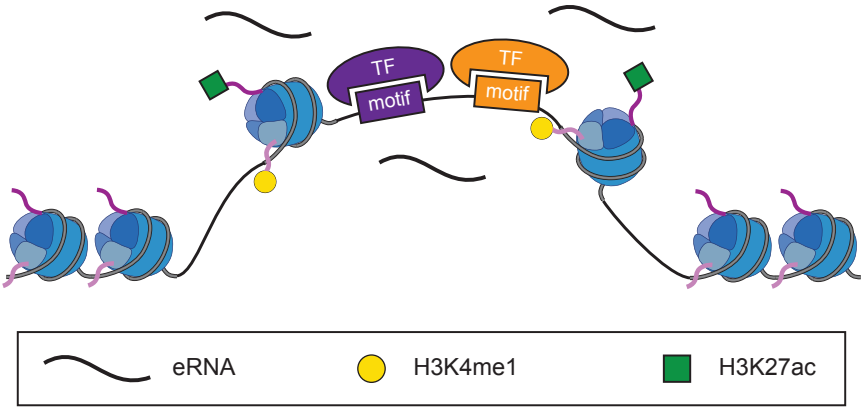

Figure I1: Schematic representation of an enhancer. Enhancers are regions of open chromatin, free of nucleosomes, covered by H3K4me1 and H3K27ac, contain TF motifs, are bound by TFs and have eRNA expression.

The most accepted model for enhancer-gene communication is through 3D proximity between the enhancer and the promoter of the gene [68-70]. Therefore, genome conformation techniques capturing 3D proximity between genomic regions (such as ChIA-PET and 3C technologies) can be used to assign putative target genes to an enhancer. However, when 3D proximity data is not available, other strategies can be used to assign putative target genes, such as the closest gene or those genes closer to a certain distance in the linear genome. Several bioinformatic tools combine different types of NGS data to predict enhancers and their target genes by correlating epigenetic information and gene expression across cell types, tissues, etc. [71-73].

The contacts between enhancers and promoters are restricted inside of the same TAD [74]. Indeed, when rearranging TADs,

a promoter-enhancer rewiring that may cause pathogenic phenotypes is produced [75, 76]. Importantly, enhancer-promoter communication does not seem a one to one relationship. Rather than that, hubs of several enhancers and promoters might influence each other and serve as traps for complexes needed for gene expression [69]. Moreover, certain promoters are controlled by multiple enhancers, each of them driving gene expression in a different place and/or a different time [68]. However, in most cases there are also multiple redundant enhancers, which confer resistance to genetic variation so that the expression of the gene is ensured [70]. The presence of redundant enhancers may allow the accumulation of mutations, which is important for evolution [77].

Enhancers permanently contribute to regulate gene expression programs in specific cell types at particular stages of development [68-70]. Interestingly, enhancer-promoter contacts are established in different ways during development. While some are pre-established before gene activation, the majority of contacts appears concomitantly with the activation of the gene [68, 70]. The specific activation of enhancers during development is given by their DNA sequence, not only through the presence or absence of a TF binding site, but also through the order, orientation and spacing of the TF motifs inside of the enhancer sequence [78-80].

1.5 Gene expression prediction with epigenetic features

There is an intense effort in the field of epigenetics to understand the relationship between gene expression and chromatin features. In this regard, several works have constructed predictive models of gene expression using as predictive variables the ChIP-seq levels of certain epigenetic features in multiple cellular contexts (Figure I2). Their main goal is to identify which epigenetic features in particular correlate better with gene expression. These studies generally measure ChIP-seq signal in the genomic region that comprises the gene body and the promoter, and model it with gene expression using different predictive methods (Table I1). Most of these studies use RNA-seq to measure gene expression.

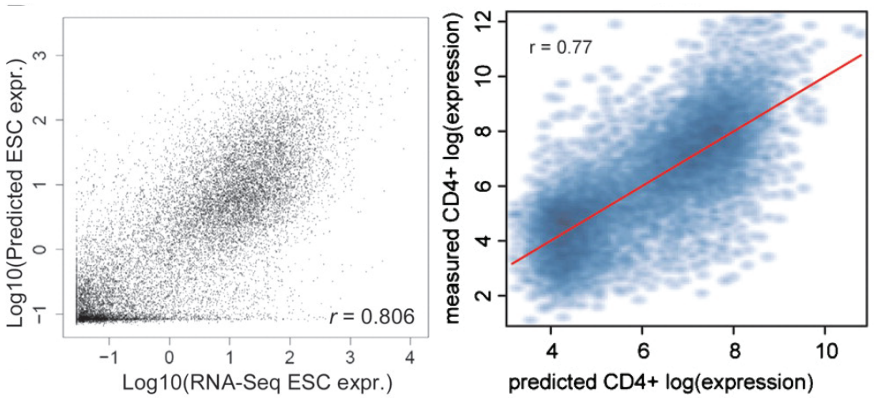

Figure I2: Performance of predictive models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Example of the first predictive model of gene expression using ChIP-seq of TF in mESCs, adapted from [81] (left). Example of the first predictive model of gene expression using ChIP-seq of histone modifications in CD4+ cells, adapted from [22] (right).

Table I1: List of gene expression prediction publications

Publication	Features	Predictive model
Ouyang et al. [81]	TFs ChIP-seq	PC-regression
Karlic et al. [22]	Histone modifications ChIP-seq	Linear regression
Cheng et al. [23]	Histone modifications and TFs ChIP-seq	Support vector machine (SVM) and support vector regression (SVR)
Cheng et al. [24]	Histone modifications and TFs ChIP-seq	SVR
Wang et al. [25]	Histone modifications ChIP-seq	Linear regression
Tippmann et al. [26]	Histone modifications and TF ChIP-seq	Linear regression
Dong et al. [27]	Histone modifications ChIP-seq, open chromatin DNase-seq, CpG islands	Random forests classification and linear regression
Zhou et al. [28]	Histone modification and TF ChIP-seq	Bayesian variable selection regression
Budden et al. [29]	Histone modifications and TFs ChIP-seq, open chromatin DNase-seq and predicted TF binding	Linear regression and SVR
Budden et al. [30]	Histone modifications ChIP-seq	Linear regression with least-squares fitting and regularised least squares regression
Singh et al. [31]	Histone modification ChIP-seq	Convolution neural network (CNN)
Read et al. [32]	Nucleosome positioning MNase-seq, distance inferred by Hi-C (to telomeres, centromeres and center), histone modifications ChIP-seq, GC content and TF motifs	Logistic regression with elastic net regularization, tree model with gradient boosting, multi-layer perceptron model

Type of data used to predict gene expression and the predictive method are reported for each publication.

Recently, predictive models combining enhancer and promoter information have shown good performance (Table I2). These studies suggest that chromatin features such as chromatin accessibility and TF binding at enhancers might have also a quantitative relationship with gene expression. Moreover, other epigenetic features such as histone modifications, could be introduced in the modelling.

Table I2: List of gene expression prediction publications introducing enhancer information

Publication	Features	Predictive model
Duren et al. [82]	DNase-seq and predicted TF binding	Linear regression
Schmidt et al. [83]	DNase-seq and predicted TF binding	Linear regression

Type of data used to predict gene expression and the predictive method are reported for each publication.

2. Mouse embryonic stem cells and bivalency

2.1 Mouse embryonic stem cells and development

Mouse embryonic stem cells (mESCs) were first isolated in 1981 [84, 85]. They have two main characteristics: pluripotency and self-renewal (Figure I3). Pluripotency refers to the capacity of differentiating into any type of cell in the developing and adult organism, whereas self-renewal refers to the ability to proliferate indefinitely while maintaining the undifferentiated state. The mESC gene expression program is controlled by TFs, cofactors, chromatin regulators and ncRNAs which altogether ensure pluripotency and self-renewal [86]. mESCs are derived from the inner cell mass (ICM) of the blastocyst embryo (Figure I3).

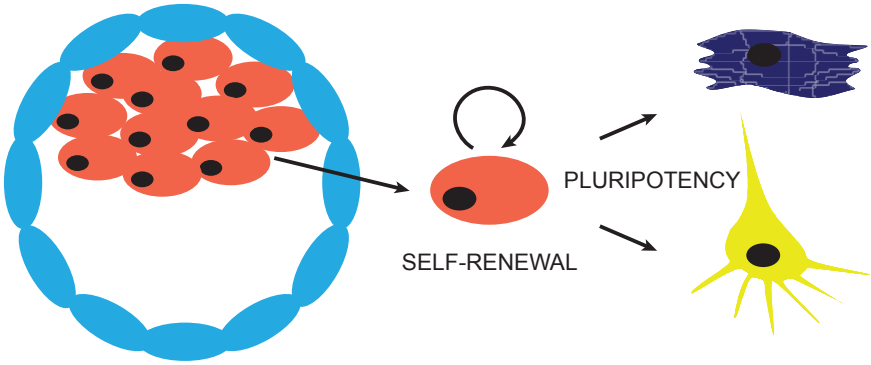

Figure I3: Origin of mESC and properties. mESCs are obtained from the blastocyst embryo (left), and their main characteristics are self-renewal and pluripotency (right).

Embryo development starts with the fertilization of the oocyte by a sperm cell, which forms the zygote. Then, the zygote (Figure I4) starts several rounds of division with subsequent cell compaction, and reaches the morula stage (4-16 cells). At E3.0 stage (3 days after fertilization, around 16 cells) cavitation starts. Cavitation consists in the formation of the blastocel, which is a cavity filled with fluid. After that (E3.5 stage), the embryo is in blastocyst stage (32 cells). Two different structures can be found in the blastocyst: the ICM and the outer layer. The cells in the outer layer will give rise to the trophoectoderm (TE). The division between ICM and TE is the first cell fate decision. The second cell fate decision occurs right after, when ICM cells become either epiblast (EPI) cells or primitive endoderm (PrE) cells. TE and PrE cells will give rise to the extra-embryonic tissues, whereas EPI cells will develop into all the cell types in the future embryo. Indeed, mESCs are derived from the EPI cells [87]. The embryo is implanted in the uterus around E4.5 stage, and gastrulation starts at E6.5 stage. During gastrulation, cells start to differentiate into the cells of the three germ layers: endoderm, mesoderm and ectoderm. These cells will finally give rise to all cell types of the organism.

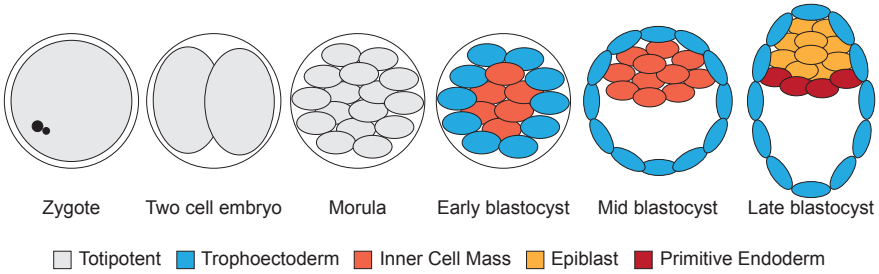

Figure I4: Early mouse development. Initial stages of mouse development, from zygote to blastocyst.

2.2 Bivalent domains

Bivalent domains are defined by the presence of H3K4me3 and H3K27me3 at the same region, and were first described in mESCs in 2006 [88, 89]. The “bivalency” term was coined for this colocalization of two opposing marks, an active one (H3K4me3) and a repressive one (H3K27me3) [88]. Moreover, in the initial studies, bivalent domains were proposed to silence developmental genes in mESCs and prepare them for future activation during differentiation [88, 89]. Indeed, it was shown that the bivalent state is resolved upon mESC differentiation when genes losing H3K27me3 were activated and genes losing H3K4me3 remained repressed [88] (Figure I5). In 2007 these observations were confirmed genome-wide with the application of the technology of ChIP-seq [38]. That same year, bivalent domains were identified also in human embryonic stem cells (hESCs) genome-wide [90, 91]. Importantly, in

2012, colocalization of H3K4me3 and H3K27me3 in the same nucleosome was confirmed [92].

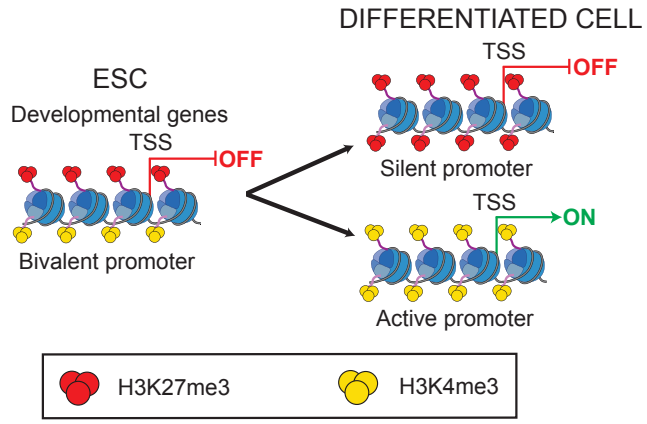

Figure I5: Resolution of bivalency in mESC differentiation. After differentiation, bivalent promoters of developmental genes resolve their bivalent state toward a silent one where H3K27me3 is conserved but H3K4me3 is lost, or an active one in which H3K4me3 is still present but H3K27me3 is lost. When the bivalent promoter becomes active, the gene is transcribed.

Further characterization of bivalent domains revealed that they are rich in CpG islands [93]. Moreover, they are usually found in the surrounding region of a TSS (bivalent promoters), and their genes are generally larger and have larger introns than active genes [94]. Bivalent promoters are occupied by paused RNA polymerase II, whose function is also poising genes for future activation [95]. Interestingly, although bivalent domains are associated to gene repression, they are found in regions of open chromatin that associate with the active compartment [51, 96]. Bivalent promoters mainly present two types of 3D contacts: intragene interactions, preferentially towards the

transcription termination site (TTS); and intra-TAD interactions, mainly between TSS of different genes [97]. Therefore, it has been proposed that bivalency provides an open chromatin architecture that allows the proper modulation of developmental gene expression programs [97].

Despite bivalency has been mostly studied in embryonic stem cells (ESCs), it has also been identified in other cellular contexts. Indeed, bivalent domains have been described also in neural precursors (NPCs) and mouse embryonic fibroblasts [38]. There are also bivalent genes in primary human $CD4^+$ central memory T cells [36, 98, 99]. Moreover, during differentiation from mESCs into NPCs, a pattern of gain and loss of bivalent domains was observed [100]. Bivalent domains where also found in mouse adult tissues such as brain, kidney, liver and lung [101], and in purified neural cells from mice [102]. Finally, bivalent domains have also been identified in adult stem cells such as hematopoietic stem cells, where their progression was monitored during differentiation into erythrocyte precursors [103]. Therefore, bivalency is a mechanism used by several types of cells to tightly regulate lineage-specific genes expression [94].

The protein complexes in charge of depositing H3K4me3 and H3K27me3 are Trithorax group (TrxG) and Polycomb group (PcG), respectively [104]. TrxG is composed by the SWI/SNF complex and the COMPASS family. The MLL1/MLL2

COMPASS-like complex is responsible of depositing H3K4me3 at bivalent domains, especially MLL2 [105]. Interestingly, loss of MLL2 in mESCs does not affect distribution of A/B compartments and TAD borders; nonetheless, it leads to increased PcG occupancy, redistribution of long-range interactions, and failure to differentiate [97].

2.3 Polycomb group proteins

Polycomb (*Pc*), the first PcG gene, was discovered in Drosophila melanogaster in 1947 [106]. In 1978, it was shown the importance of *Pc* in repressing *Homeotic* (*Hox*) genes [107]. Later, other genes with similar roles as *Pc* were found, which led to the definition of the PcG proteins [104]. Finally, many more PcG ortholog genes were found in mammals, most probably as a result of duplication events during metazoan evolution [108].

PcG proteins are associated in two main complexes: Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2). At the same time, PRC1 in mammals is subdivided into canonical PRC1 (cPRC1) and non-canonical PRC1 (ncPRC1) [109]. The core components of PRC1 are one of the RING1A/B and one of the Polycomb group ring-finger domain proteins (PCGF1-PCGF6) [104]. RING1A/B have E3

ubiquitin ligase activity which mediates the ubiquitylation of histone H2A on lysine 119 (H2AK119ub), a mark that is important for gene silencing [110]. On the one hand, cPRC1 contain PCGF2/4, a chromobox protein (CBX2/4/6/7/8) and a Polyhomeotic (Ph) homologous protein (PHC1-PHC3) [104]. On the other hand, ncPRC1 contain the zinc-finger domain and YY1-binding protein (RYBP) or its paralog YAF2, and PCGF1/3/5/6 [104]. The different set of subunits specific to each complex modulates DNA-binding affinities and regulatory functions [104].

The core components of PRC2 are one enhancer of zeste (EZH1/2), embryonic ectoderm development (EED), suppressor of zeste (SUZ12) and one retinoblastoma-binding protein (RBBP4/7) [111]. Ezh1/2 contains a SET domain and is the catalytic subunit of the complex in charge of depositing H3K27me3 [112, 113]. At the same time, PRC2 can be classified into two variants: PRC2.1 and PRC2.2, according to the presence or absence of certain accessory proteins. The emergence of these accessory proteins is correlated with the increase in cell and tissue complexity of metazoans, which suggests that regulation of PRC2 by different accessory subunits is important for cell fate and cell identity [104]. On the one hand, PRC2.1 is composed by one of the Polycomb-like proteins (PCL1/2/3, also named PHF1, MTF2 and PHF19, respectively), Elongin BC and Polycomb repressive complex 2-associated protein (EPOP) or PRC2-associated LCOR

isoform 1 (PALI1/2) [111]. On the other hand, PRC2.2 contains Jumonji and AT-rich interaction domain 2 (JARID2) and adipocyte enhancer-binding protein 2 (AEBP2) [111].

Loss of both, PRC1 and PRC2 subunits in mESC showed an important role of PcG in maintaining a pluripotent state and also for proper differentiation [114-119].

2.4 Bivalency, PcG and chromatin architecture

As for MLL2, loss of PcG subunits in mESC does not lead to changes in A/B compartment distribution or TAD borders, and instead, it also causes changes in interactions between bivalent regions. For example, deletion of several PRC1 subunits reduced the number of intrachromosomal contacts with consequent derepression of target genes [51, 120-122].

When bivalent genes activate during mESC differentiation, there is a displacement of PcG and a consequent reduction of interactions between PcG-bound sites [120, 122, 123]. On the contrary, mESC-specific genes and bivalent genes that become further repressed gain PcG occupancy and even increase their contacts with PcG-bound regions [122]. Therefore, PcG occupancy is associated to the establishment of a repressive chromatin architecture [94].

Despite most interactions between enhancers and their target gens appear concomitantly with gene activation and are disrupted when genes are repressed, some pre-set chromatin interactions prior transcriptional activation occur at developmental genes [52, 123]. Indeed, loops facilitated by PRC2 between bivalent promoters and poised enhancers (PE) are pre-established in mESC and maintained throughout differentiation [124].

2.5 Poised enhancers

PEs were originally described in mESC in 2010, and were defined as enhancers that are primed for future activation during development [65]. In this study, the authors distinguished active enhancers from poised ones by presence or absence of H3K27ac in H3K4me1 sites, respectively. In 2011, the colocalization of H3K4me1 and H3K27me3 at enhancers was demonstrated in hESC and mESC, as well as their role in poising for future gene activation [125, 126]. Since then, the term PE usually refers to enhancers marked by H3K4me1 and H3K27me3, in absence of H3K27ac. In consequence, a third type of enhancers marked only by H3K4me1, called intermediate enhancers, arose [126]. In the literature, intermediate enhancers have also been called primed enhancers [124]. In this PhD thesis, we will refer to enhancers marked by H3K4me1 in absence of H3K27ac and

H3K27me3 as intermediate, enhancers marked by H3K4me1 and H3K27me3 in absence of H3K27ac as poised, and finally, enhancers marked by H3K4me1 and H3K27ac in absence of H3K27me3 as active (Figure I6).

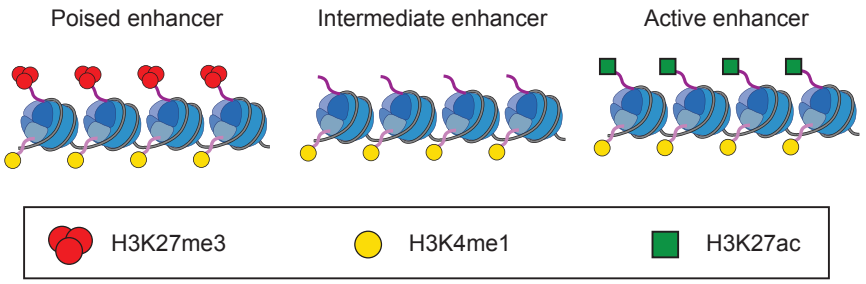

Figure I6: Types of enhancers in mESC. There are three types of enhancers in mESC that can be distinguished by their combination of histone modifications: poised (presence of H3K4me1 and H3K27me3 and absence of H3K27ac), intermediate (presence of H3K4me1 and absence of H3K27me3 and H3K27ac) and active (presence of H3K4me1 and H3K27ac and absence of H3K27me3)

Several PcG subunits have been found in PEs [124-126], as well as other factors associated to gene activation. For example, chromatin remodelers and histone modifiers involved in active transcription such as p300 [124-126], CHD7 [126], and BRG1 [125] have been found in PEs.

Similar to bivalent promoters, PEs also resolve their poised state during differentiation towards neural lineage (Figure I7): either they become active by losing H3K27me3 and gaining H3K27ac, or they remain inactive by retaining H3K27me3 and diminishing p300 [124-127]. Interestingly, recent evidence

suggests that this mechanism of PE activation is not exclusive of the neural lineage [128].

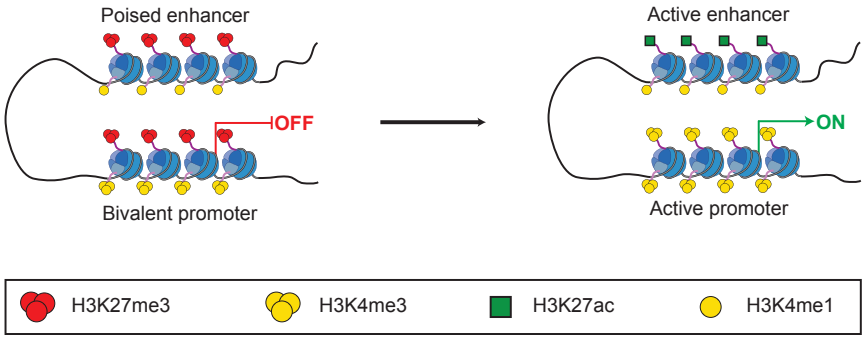

Figure I7: Coordinated activation of a bivalent promoter and its poised enhancer in differentiation. During differentiation from mESC, there is a coordinated activation of the bivalent promoters and their poised enhancers.

PEs have also been identified in other cellular contexts different from ESC. Indeed, PEs becoming active during differentiation towards effector/memory T cells were identified in mouse naïve $CD8^+$ T cells [129]. Moreover, a group of PEs was found in neural crest cells [127]. However, true colocalization of H3K27me3 and H3K4me1 in the same chromatin fragment has only been confirmed in ESC [94]. Finally, their function in cell types different from ESC is still controversial as they could play either a repressive or a poised role [94, 130, 131].

3. Epigenetics and disease

3.1 Epigenetic alterations and therapies in disease

Changes in DNA methylation, histone modifications and ncRNA are common in disease [10]. Indeed, several mutations that affect chromatin factors such as chromatin remodelers, histone modifiers and even histone variants have been identified in diseases such as cancer [132]. Therefore, chromatin proteins and modifications can be targeted for therapy [133]. Indeed, histone deacetylase (HDAC) inhibitors have already been approved by the Food and Drug Administration (FDA) to treat diseases such as breast cancer, lymphomas and multiple myeloma [134]. Moreover, other HDAC inhibitors, histone methyltransferase (HMT) inhibitors, etc. are in clinical trials [134].

3.2 Cushing's syndrome

Endogenous Cushing's syndrome (CS) is a rare disease that produces hypercortisolism and is caused by a pituitary or an adrenal tumour [135]. Hypercortisolism, which refers to overexposure to glucocorticoids (GCs), is the cause of several adverse cardiometabolic effects such as truncal obesity,

insulin resistance, altered glucose homeostasis, dyslipidaemia and increased cardiovascular morbidity and mortality [136, 137]. Moreover, around 1-2% of the population receives long-term GC treatment against a broad spectrum of inflammatory and autoimmune diseases, with similar side effects as hypercortisolism caused by CS [138, 139]. Strikingly, even after long-term CS remission, persistent cardiometabolic effects have been described [140-144].

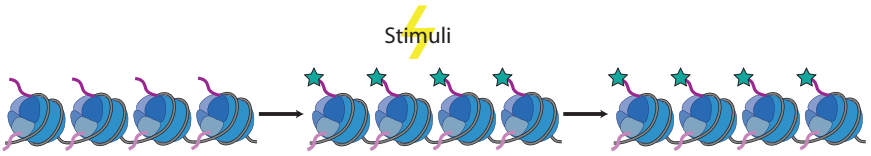

Figure I8: Epigenetic memory. Even in absence of the stimuli that originated it, the epigenetic change (depicted with stars) is maintained. In Cushing's syndrome, the stimuli would correspond to high levels of glucocorticoids.

Chromatin carries epigenetic information that can propagate active and repressed states during cell division, even in absence of the original signal that established them [10] (Figure I8). Therefore, epigenetic changes could explain why there is not a complete clinical remission after the resolution of hypercortisolism. Indeed, it has been shown that chronic hypercortisolism reshapes the epigenetic landscape by inducing lasting changes in DNA methylation in whole blood human samples [145, 146] or by regulating histone modifying enzymes in a rat model of CS [147, 148].

OBJECTIVES

If I had not been discriminated against or had not suffered persecution, I would never have received the Nobel Prize.

Rita Levi-Montalcini (1909-2012)

1. Deciphering which histone modifications correlate better with enhancer function

Modelling gene expression from enhancer epigenetic information might help to understand how the contribution to gene expression differs between promoters and enhancers and, more broadly, how enhancers function. However, predictive models using only information at enhancers have not been obtained yet. Thus, we set out to explore the quantitative relationship between histone modifications and gene expression, focusing on enhancer regions. In particular, our main goal is to decipher which histone modifications correlate better with enhancer function. To do so, we ask the following questions:

- Are histone modifications at enhancers predictive of gene expression?
- Which histone modifications are more predictive in the enhancer models?
- Are the same histone modifications also important for the promoter predictive models?
- Is an enhancer predictive model learned in a specific cell type useful to predict gene expression in another one?

2. Understanding the biological role of poised enhancers during differentiation and development

Much effort has been invested into understanding the role of bivalent promoters throughout differentiation from mESCs. However, little attention has been brought to poised enhancers. To overcome this limitation, we aim to explore the biological role of poised enhancers during differentiation and development. Thus, we focus on the following issues:

- Characterization in terms of epigenetic landscape, chromatin accessibility, conservation and overlap with CpG islands, of poised enhancers in mESCs
- Tracking the fate of poised enhancers that become active during differentiation along different lineages, both *in vitro* and *in vivo*
- Studying the implication of poised enhancer resolution in differentiation

3. Identification of a persistent epigenetic fingerprint of Cushing's syndrome

To address this issue, we established a collaboration with researchers and medical doctors from IDIBAPS and Hospital Clínic in Barcelona. We merged their expertise in the physiology of Cushing's syndrome (CS) and a mouse model of the disease, with our expertise in epigenetics, gene regulation and Next Generation Sequencing (NGS) data analysis. All together we aim to identify an epigenetic fingerprint in CS that could explain the persistent effects after remission of the disease. To address this issue, we focus on the following objectives:

- Construction of a pipeline to obtain statistically significant consensus peaks across several replicates
- Integration and comparison of expression and epigenomic data retrieved in different species, and during active CS and after remission
- Identification of the epigenetic fingerprint that could explain the persistent effects after remission

RESULTS

Nothing was given to minorities or women. It took some fighting to get that equal opportunity and we're still fighting today.

Annie Easley (1933-2011)

CHAPTER 1

The results shown in this chapter correspond to objective number 1 of this thesis.

González-Ramírez M., Ballaré C., Mugianesi F., Beringer M., Santanach A., Blanco E. and Di Croce L. Differential contribution to gene expression prediction of histone modifications at enhancers or promoters. In revision at Nucleic Acids Res.

1.1 Computational design

We developed a novel computational approach based on the combination of chromatin segmentation and linear regression to infer gene expression using ChIP-seq data from histone modifications at enhancers and promoters (Figure R1.1). To construct proper predictive models, the full spectrum of active and repressed regions is needed. Therefore, we took advantage of mESCs for which active and repressed regulatory regions can be identified. mESCs contain active enhancers (AEs) and poised (repressed) enhancers (PEs), which respectively coordinate with active promoters (APs) and bivalent (repressed) promoters (BPs) to regulate gene expression [94].

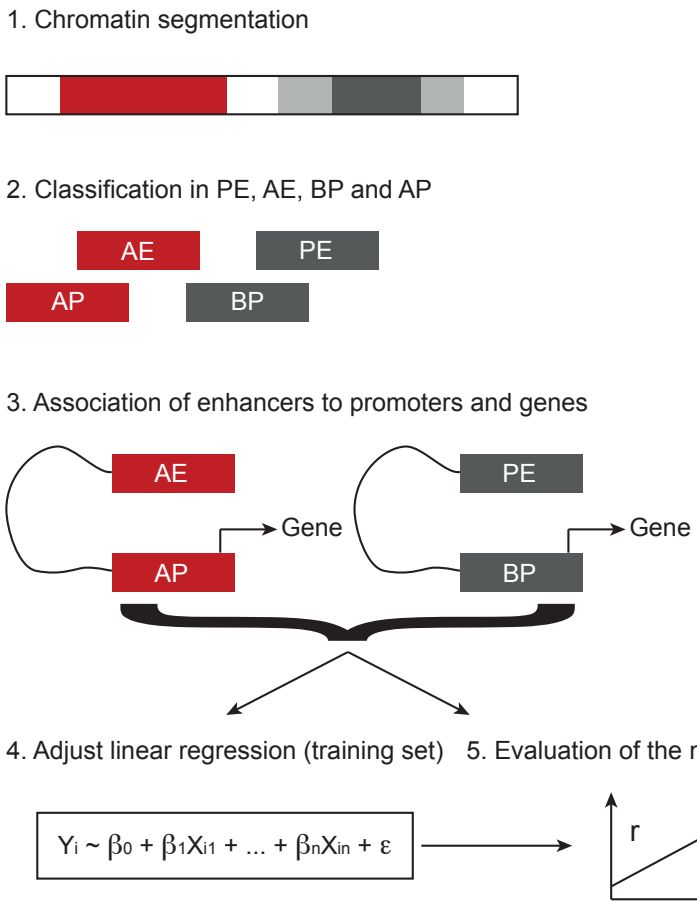

Figure R1.1: Schematic diagram of the modelling pipeline to predict gene expression in mESC. Chromatin segmentation was used to identify active (red) and repressed (dark grey) regions. Next, these regions were classified into PE, AE, BP and AP. Enhancers were associated to their target promoters and genes using Hi-C interactions. Finally, the set of enhancers was divided into training and test sets. The training set was used to learn the predictive model and the test set to evaluate it. Y_i is the $\log_2$ expression of gene i plus a pseudocount of 0.1. β_0 to β_n are the coefficients to be inferred. X_{i1} to X_{in} are the log2 of the ChIP-seq signal strength plus a pseudocount of 0.1. ε is the error. r is the Pearson's correlation coefficient between the predicted expression and the measured one.

1.2 Construction of a chromatin segmentation map

First of all, ChIP-seq experiments of H3K4me3, H3K27me3, H3K27ac, and H3K4me1 in mESC were performed by Cecilia Ballaré (our lab) to identify the different types of regulatory regions. This set of histone modifications has been previously used to distinguish between AEs, PEs, APs, and BPs in mESC [124, 126]. We next generated a 9-state chromatin segmentation model of mESCs with the ChromHMM software [49] (Figure R1.2). As expected, active states mark transcriptionally active regions, while repressed states denote transcriptionally repressed regions (Figure R1.2B; see also our previously published RNA-seq data [118]).

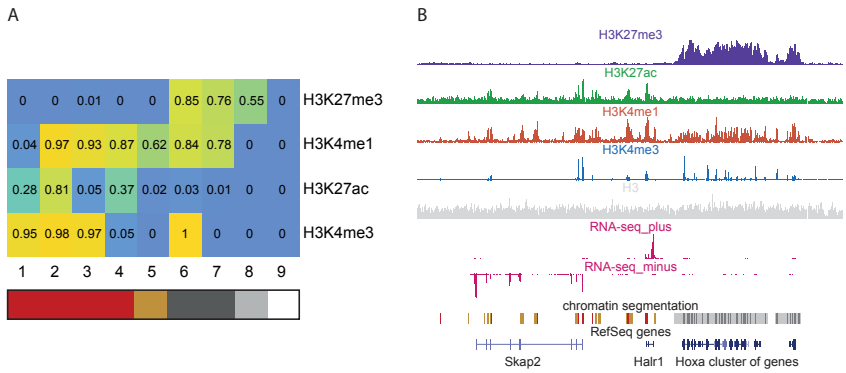

Figure R1.2: Chromatin segmentation model for mESC. (A) State definition of the chromatin segmentation model in mESCs. The values represent the probability (from 0 to 1) of finding each histone modification (vertical) in genomic segments of states 1 to 9 (horizontal). Red: states with histone modifications associated to activation (active, 1–4); dark yellow, H3K4me1-only state (Intermediate, 5); grey, states in which H3K27me3 was present (repressed, 6–8); dark grey, poised states, in which H3K27me3 colocalized with H3K4me3 and/or H3K4me1 (states 6 and 7); light grey, H3K27me3-only regions (state 8); and white, unmarked state (9). (B) Example of a genomic region containing two expressed genes (*Skap2* and *Halr1*), which are covered by active states (in red), and a cluster of repressed genes (HoxA), which are covered by repressed states (in grey). Active chromatin segments integrate the signal of H3K27ac, H3K4me3, and H3K4me1 and lack H3K27me3. Repressed chromatin segments integrate the signal of H3K27me3, H3K4me3, and H3K4me1 and lack H3K27ac. Expression of *Skap2* and *Halr1*, and silencing of HoxA genes, were confirmed by the RNA-seq profiles [118].

To understand the resulting map of states, we designed a Python script that calculates the matrix of transition enrichments between all states in the model. The transition value between two different states, x and y, is defined as the number of times that a segment of state y is found after a segment of state x, as measured from left to right in the linear genome. The enrichment score is defined as the ratio between the observed number of transitions and the expected number of transitions by chance.

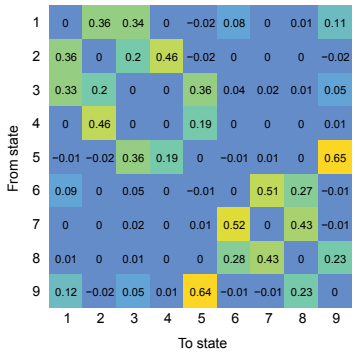

Figure R1.3: State transitions. Enrichment of state transitions (number of observed transitions divided by the number of expected transitions by chance) from the segments of one state (vertical) towards the segments of another state (horizontal) in the linear chromatin.

Two groups of states (active, 1–4, and repressed, 6–8) clearly emerged from the global picture (Figure R1.3), as well as state 5 (most likely representing intermediate enhancers). The high enrichment of transitions between states belonging to the same category suggests that they might mark the same functional regulatory regions, rather than be caused by different functional regions separated by the unmarked state. Indeed, visual inspection confirmed that states 1 to 4 marked the same active regulatory regions, revealing that differences in the state definition are due to differences in the shape of the ChIP-seq peaks (Figure R1.4A). We also observed that the repressed states 6 and 7 decorated poised or bivalent regulatory regions (e.g., marked with H3K27me3, in combination with H3K4me3 and/or H3K4me1), whereas state 8 was generated by the tail-end of broad peaks of H3K27me3 (Figure R1.4B). Similarly, state 5 was associated with the tail-

end of broad peaks of H3K4me1 in active regions, but it also associated with single peaks of H3K4me1 near active regions, which might correspond to intermediate enhancers (Figure R1.4A).

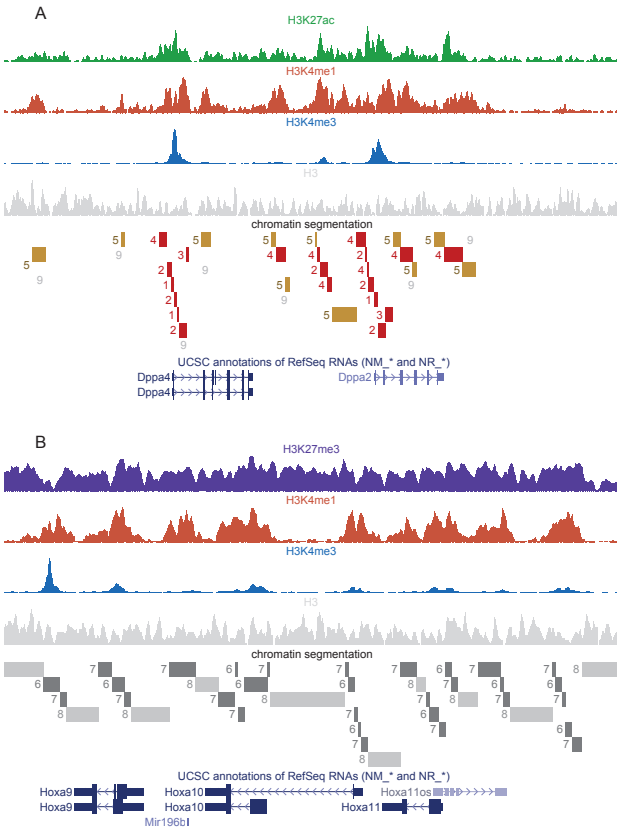

Figure R1.4: Functional regions are covered by more than one class of state. (A) Segments of active states 1–4 cover the same functional regions delimited by peaks of H3K4me3, H3K27ac and H3K4me1. Differences in the definition of active states are due to the shape of the peaks over the same functional elements. Segments of state 5 denote broad peaks of H3K4me1 in active regions, or single peaks of H3K4me1. (B) Segments of repressed states 6 and 7 denote the sharp peaks of H3K4me3 and H3K4me1 found inside broad regions covered by H3K27me3. State 8 corresponds to the fraction of H3K27me3 peaks that does not overlap with the other two marks. Differences in the definition of repressed states 6 and 7 are due to the shape of the peaks over the same functional elements.

Based on these results, we decided to merge the contiguous segments of states 1 to 4 in a single unit to constitute the list of potential active regulatory regions, and the segments of states 6 and 7 in a single unit to constitute the list of potential poised or bivalent regulatory regions.

1.3 Identification of poised enhancers, active enhancers, bivalent promoters and active promoters

We reasoned that functional regions should have a minimum length of 600 bp and thus discarded shorter regions, which we considered as artefacts. We also discarded those cases in which an active region and a poised region were contiguous, as this was ambiguous. Promoters were defined as those regions that overlapped by at least 1 bp to a region $\pm$ 500 bp around a TSS according to RefSeq [149]. Enhancers were defined as regions that were not classified as promoters and overlapped by at least 1 bp with a peak of the enhancer mark p300 [124]. As H3K4me3 can be present in enhancers [36, 129, 150, 151], we did not discard enhancers containing H3K4me3, although this histone modification has been traditionally only associated with promoters. In total, we found 9,421 APs, 3,344 BPs, 16,904 AEs, and 2,699 PEs in mESCs.

1.4 Association of enhancers to promoters and target genes

Next, we matched our sets of promoters and enhancers to their potential target genes. Using RNA-seq data [118], we confirmed that genes associated with APs are expressed in mESC, while genes associated with BPs are silenced (Figure R1.5A). Gene ontology (GO) term enrichment analysis performed with Enrichr [152] confirmed that genes with APs are involved in housekeeping roles, while genes with BPs are mostly related to development and differentiation (Figure R1.5B), as is expected in mESCs.

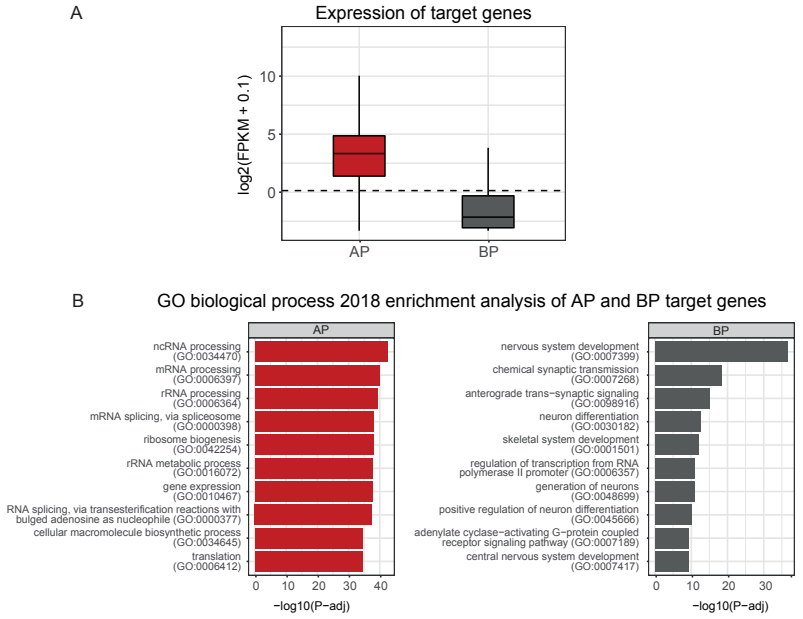

Figure R1.5: Target genes of APs and BPs. (A) Expression of genes associated to active promoters (AP; 10,786 genes) or bivalent promoters (BP; 3,459 genes). The dotted line represents 1 FPKM. (B) Top GO biological process (2018 categories) for each list of genes in A.

We also used available high-throughput chromosome conformation capture (3C) data, of Hi-C [123], to link AEs and PEs with target genes. Francesca Mugianesi (our lab) analyzed Hi-C data and provided the list of significant interactions (total of 43,892,155 significant Hi-C interactions). As interacting enhancers and promoters have been shown to match their chromatin state [52], we developed a Python script that associates an enhancer with the target gene of a promoter when both enhancer and promoter are in the same category (e.g., both active, or both repressed) and each one overlaps with one of the two sides of the same Hi-C significant interaction. We confirmed that PEs were significantly enriched in interactions with BPs over APs (p < 2.2e-16, Exact Binomial Test), whereas AEs were significantly enriched in interactions with APs over BPs (p < 2.2e-16, Exact Binomial Test). In total, we found 10,786 genes associated to APs, 3,459 genes to BPs, 10,206 genes to AEs, and 2,526 genes to PEs. Likewise, 15,841 AEs and 2,466 PEs were associated to at least one gene, and 8,931 APs and 2,443 BPs, to at least one enhancer.

1.5 Models using all the interactions to associate enhancers to promoters (Hi-C–all)

After identifying the set of enhancers and promoters in mESCs, we assessed several gene expression predictive models. We first performed additional ChIP-seq experiments for other

histone modifications, in order to have additional variables to predict gene expression. Specifically, we performed ChIP-seq experiments for trimethylation of histone H3 at lysine 36 (H3K36me3), ubiquitination of histone H2B (H2Bub), monomethylation of histone H3 at lysine 27 (H3K27me1), dimethylation of histone H3 at lysine 27 (H3K27me2), trimethylation of histone H4 at lysine 20 (H4K20me3), and dimethylation of histone H3 at lysine 79 (H3K79me2). This new set of experiments was performed by Malte Beringer and Alexandra Santanach (former members of our lab). The input of our predictive models consisted of the ChIP-seq data of these six histone marks as well as the four histone marks previously used to define promoters and enhancers (H3K4me3, H3K4me1, H3K27ac and H3K27me3), together with our previously-published RNA-seq expression data [118]. A total of 11,387 mouse protein–coding genes previously associated to a promoter (active or bivalent) and at least one enhancer (active or poised) entered the modelling. We divided the set of genes into two subsets: training and test. The Pearson's correlation coefficient (*r*) between the measured expression in the test subset and the predicted one was used to assess the performance of a predictive model.

We first generated a predictive model for the promoters identified using our approach (named Hi-C–all promoter model) to confirm that histone modifications at these elements are predictive of gene expression, as has been previously

shown in several cell types from different model organisms, including mESCs [22-32]. As a control, we repeated the predictive model learning in a training subset in which expression values were randomized. We obtained an *r*-value of 0.81 for the promoter model, and an *r*-value of –0.07 for the random promoter model (Figure R1.6A). The low performance of the random promoter model strongly indicated that the high predictive power of the Hi-C–all promoter model was not due to random structures in the data. Importantly, the performance of our promoter model was comparable to previously described predictive models [22-32]. Coefficients and *p*-values of the predictors for all the predictive models generated in this study can be found in the Appendix of this thesis.

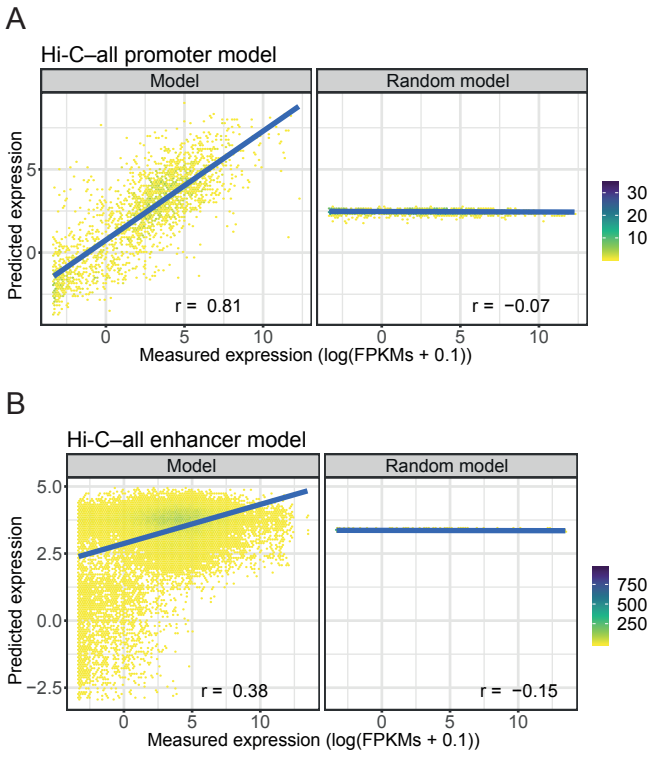

Figure R1.6: Performance of the Hi-C–all models in mESC. Predicted expression of the test subset of genes calculated by the models versus their measured expression by RNA-seq. Model performances are represented by the Pearson's correlation (*r*) between predicted and measured expression values. (A) Left, the model trained on the promoter regions associated to at least one enhancer using all significant interactions of Hi-C (Hi-C–all promoter model). Right, the performance of the same model after randomizing the expression of the training subset of genes. (B) Left, the model trained on the enhancer regions associated to at least one promoter using all the significant interactions of Hi-C (Hi-C–all enhancer model). Right, the performance of the same model after randomizing the expression of the training subset of genes.

Next, we trained a second predictive model of gene expression (the Hi-C–all enhancer model) using the levels of histone modifications at the previously-defined enhancers as predictors. We obtained a performance in the test subset of *r* = 0.38 (Figure R1.6B). Although the *r*-value of this model is

relatively modest, this is the first time to our knowledge that enhancers have been shown to be predictors of gene expression through their histone modification levels. Critically, when the expression data for learning the model were randomized, the performance was poor (r = –0.15, Figure R1.6B).

1.6 Models using the best interactions to associate enhancers to promoters (Hi-C–top)

We hypothesized that the modest performance of the Hi-C–all enhancer model could be due to some enhancer–promoter associations that were simply regions in close 3D proximity but not functionally linked. To enrich the set of interactions for functional promoter–enhancer loops, we applied a more restrictive threshold on the Hi-C significant interactions (FDR = 0 and ln(p-value) ≤ –100). We obtained a total of 5,555,844 interactions (8% of the total interactions). We then recalculated the enhancer-promoter-gene associations and obtained 1,846 PEs associated to 1,382 BPs and to 1,434 target genes, and 11,777 AEs associated to 7,211 APs and to 8,254 target genes. We selected the protein-coding genes included in the new associations (a total of 8,639 genes) to recalculate the predictive models (hereon in termed Hi-C–top models) and random models for promoters (Figure R1.7A) and enhancers (Figure R1.7B). While the Hi-C–top promoter model performed

similarly to the previous model (r = 0.79 vs. r = 0.81, respectively), the outcome of the Hi-C–top enhancer model was significantly improved (r = 0.49 vs. r = 0.38). This result further confirmed that enhancers, as well as promoters, possess a quantitative relationship with gene expression.

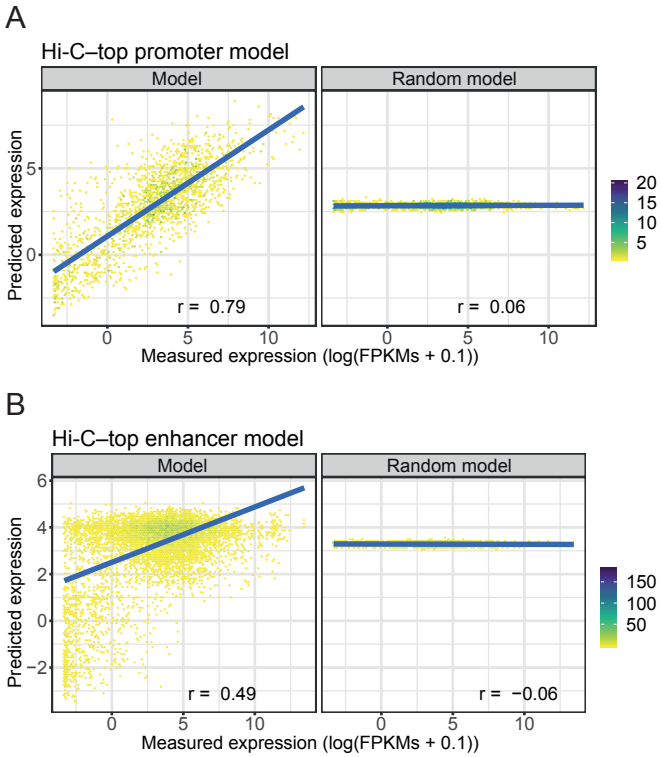

Figure R1.7: Performance of the Hi-C–top models in mESC. Predicted expression of the test subset of genes calculated by the models versus their measured expression by RNA-seq. Model performances are represented by the Pearson's correlation (r) between predicted and measured expression values. (A) Left, the model trained on the promoter regions associated to at least one enhancer using the top significant interactions of Hi-C (Hi-C–top promoter model). Right, the performance of the same model after randomizing the expression of the training subset of genes. (B) Left, the model trained on the enhancer regions associated to at least one promoter using the top significant interactions of Hi-C (Hi-C–top enhancer model). Right, the performance of the same model after randomizing the expression of the training subset of genes.

Finally, from the Hi-C–top interactions we selected those involving a distal enhancer (> 5 Kb from a TSS, a total of 14,861 AEs and 2,287 PEs) to confirm that the predictive capacity was not exclusive of proximal enhancers. We generated a new distal enhancer model (Hi-C–top_distal; 7,925 protein-coding genes) that properly predicted gene expression with a $r = 0.41$ (Figure R1.8).

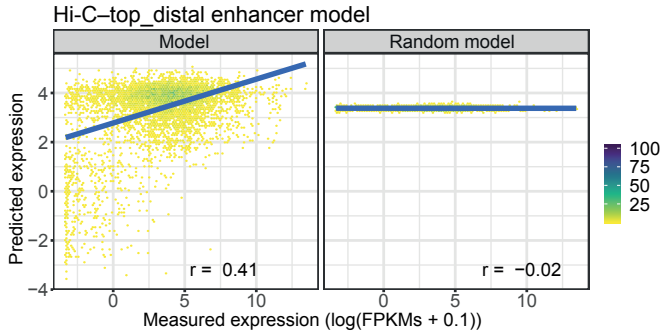

Figure R1.8: Performance of the Hi-C–top_distal enhancer model in mESC. Predicted expression of the test subset of genes calculated by the models versus their measured expression by RNA-seq. Model performances are represented by the Pearson's correlation (*r*) between predicted and measured expression values. Left, the model trained on the distal enhancer regions associated to at least one promoter using the top significant interactions of Hi-C (Hi-C–top enhancer model). Right, the performance of the same model after randomizing the expression of the training subset of genes.

1.7 Models using 1 Mb distance to associate enhancers to promoters

We now know that an enhancer preferentially interacts with promoters located in the same TAD rather than those located

in neighboring domains [74]. Moreover, TADs have an average size of around 1 Mb. Therefore, the assignment of genes to enhancers located in a distance lower than 1 Mb seems appropriate. Thus, we developed a Python script that associates enhancers to promoters of the same chromatin state that are closer than 1 Mb. Indeed, when evaluating a new mESC predictive model with this new set of enhancer–gene associations (1 Mb model; 11,986 protein-coding genes), we achieved a performance of r = 0.34 (Figure R1.9). Nonetheless, that performance is lower than the mESC models based on Hi-C data that we have built previously (r = 0.38 and r = 0.49 for Hi-C–all and Hi-C–top enhancer models, respectively). This suggests that matching genes to regulatory elements by 1 Mb distance leads to some false–positive associations, yet maintaining its predictive capacity.

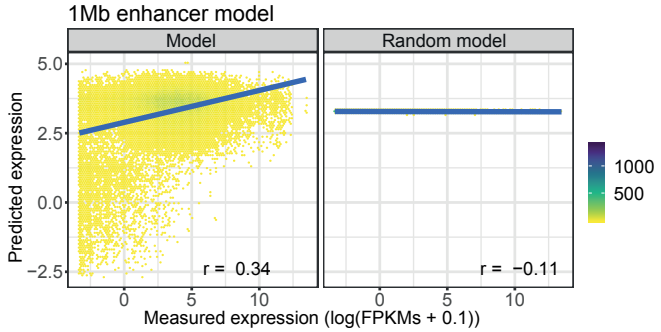

Figure R1.9: Performance of the 1 Mb enhancer model in mESC. Predicted expression of the test subset of genes calculated by the models versus their measured expression by RNA-seq. Model performances are represented by the Pearson's correlation (r) between predicted and measured expression values. Left, the model trained on the distal enhancer regions associated to at least one promoter using 1 Mb distance (1 Mb enhancer model). Right, the performance of the same model after randomizing the expression of the training subset of genes.

1.8 H3K27me3 is the most informative mark for predicting gene expression in mESC

Enhancers and promoters exhibit similar histone modification patterns. We therefore wondered whether the histone modifications mostly contributing to the prediction of expression are the same ones as well, or there are differences in the contributions between the enhancer and the promoter predictive models. To address this issue, we assessed variable importance in the Hi-C–top model for promoters and enhancers. Variable importance refers to the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistic for each model parameter. Notably, H3K27me3—a histone modification associated with transcriptional gene repression—was the prevalent mark in both classes of regulatory elements (Figure R1.10). In contrast to promoters, in which H3K27me3 has a relatively similar importance as other marks (e.g., H2Bub, H3K4me3, and H3K36me3, Figure R1.10A), H3K27me3 in enhancers represented up to 55% of the total importance (Figure R1.10B). Interestingly, H3K27ac—considered the canonical marker of enhancer activation—had little importance in the enhancer model. Therefore, even though promoters and enhancers contribute to predict gene expression mostly through H3K27me3, this contribution seems to be uniquely driven by H3K27me3 in the enhancer predictive model while it

is shared by other histone marks in the promoter predictive model.

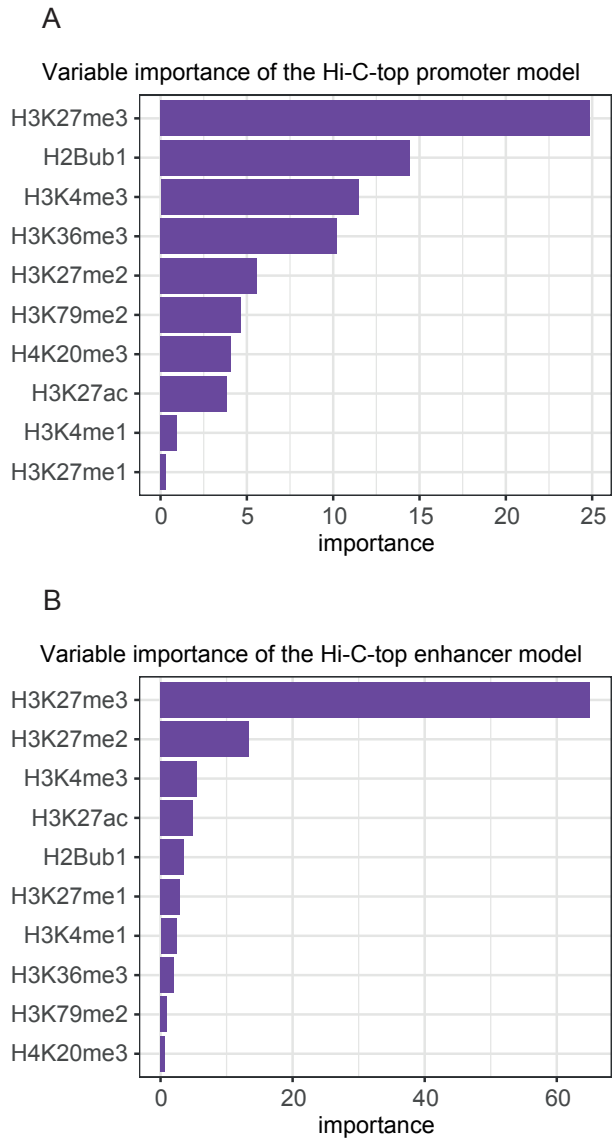

Figure R1.10: Variable importance of each histone modification in the Hi-C–top predictive models. (A) Importance of each histone modification used to train the Hi-C–top promoter predictive model. (B) Importance of each histone modification used to train the Hi-C–top enhancer predictive model. In both A and B, importance is defined as the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistic for each model parameter.

1.9 LOESS normalization of ChIP-seq and RNA-seq data from heterogeneous sources

We next aimed to determine whether enhancers are predictive of gene expression in other cellular contexts besides mESCs, and whether a predictive model learned in one cell type could predict gene expression in another. As true colocalization of H3K27me3 with H3K4me1 or p300 in the same DNA fragment has been only studied in mESCs, it is still not clear whether PEs exist in other developmental scenarios [94]. To address this issue, we took advantage of the capacity of PEs and BPs to switch into an active state for certain cell types, in a lineage-specific manner during differentiation from mESCs, while remaining inactive in others [94]. We hypothesized that differentiation data could be used to obtain predictive models exclusively from PEs and BPs. We focused on several time points for two cell differentiation mouse models (Figure R1.11): i) cardiac lineage: mesoderm (MES), cardio precursors (CPs), and cardiomyocytes (CMs) [153]; and ii) neural lineage: neural precursors (NPCs) and cortical neurons (CNs) [123].

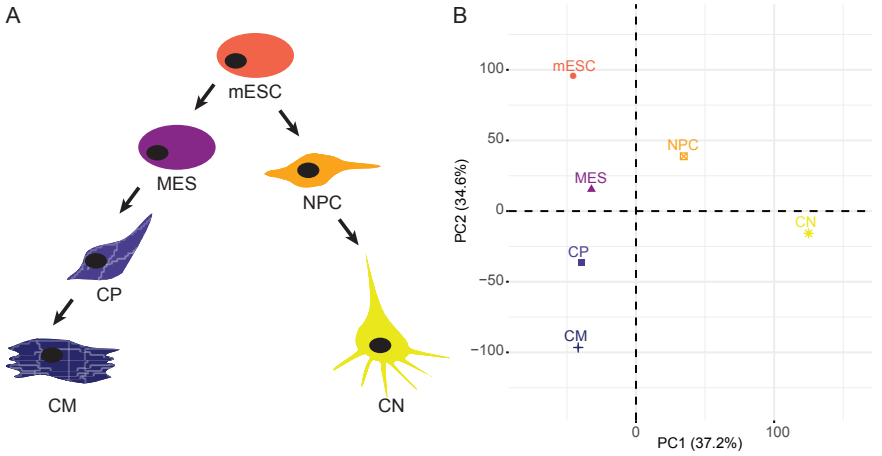

Figure R1.1: Differentiation from mESC towards cardiac lineage and neural lineage. (A) Schematic representation of the differentiation stages from mESC towards: i) cardiac lineage: mesoderm (MES), cardio precursors (CPs), and cardiomyocytes (CMs); and ii) neural lineage: neural precursors (NPCs) and cortical neurons (CNs). (B) Principal component analysis (PCA) on the expression data of the differentiation stages in A. The PCA plot was generated by our SeqCode webserver (http://ldicrocelab.crg.eu/02_DataDist/PCAplotter/index.php).

We downloaded RNA-seq and ChIP-seq data of five histone modifications (H3K27me3, H3K4me3, H3K27ac, H3K4me1, and H3K36me3) that were available in the literature for both differentiation models. To remove potential biases (e.g., due to the source of data generation or to a batch effect), we normalized the sequencing samples of the same feature at all the available time points. For this, we designed a new normalization approach based on a local regression (LOESS) that was originally proposed for the pairwise normalization of expression microarrays [154] but generalized for multiple arrays [155]. LOESS normalization is based on a MA methodology, where M is the $\log_2$ ratio of the intensities of the samples, and A is the $\log_2$ of the average intensity. It assumes

that the intensities of the two samples should be equal, therefore M = 0. Finally, corrections based on a LOESS are applied to obtain a MA plot in where the regression line approximates M = 0.

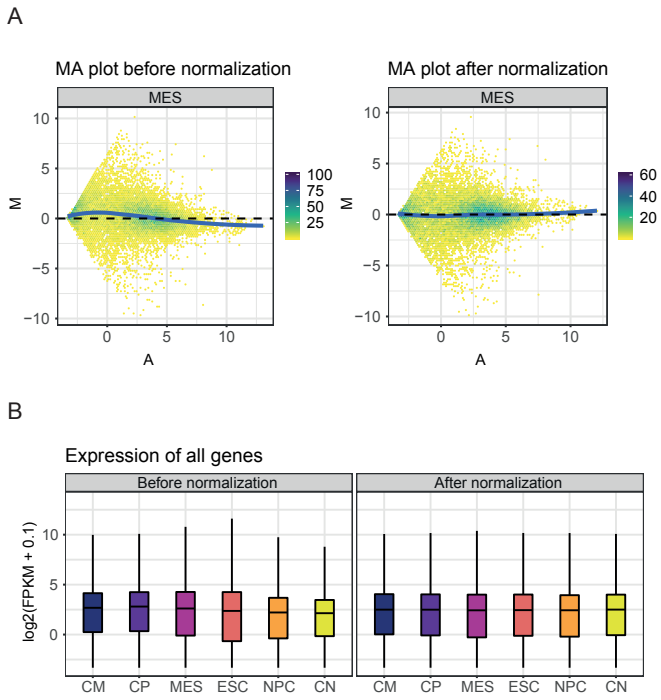

Figure R1.12: Performance of LOESS normalization in RNA-seq data. (A) MA plot before (left) and after (right) normalization of expression data at MES against mESCs. M represents the $\log_2$ ratio of the intensities of the two samples and A is the $\log_2$ of the average intensity. Intensity is determined in FPKMs. After normalization, the regression line tends to M = 0. (B) Boxplot of expression of 15,065 protein-coding genes before and after LOESS normalization. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors.

We first applied this normalization method over the expression of a set of 20,706 protein-coding genes in the mouse genome, at each differentiation time point, and mESCs (an example of

MA plots of MES against mESCs before and after LOESS normalization is shown in Figure R1.12A). As LOESS normalization assumes that expression is equal in all samples, a general balance in global expression distribution of all time points is expected (R1.12B).

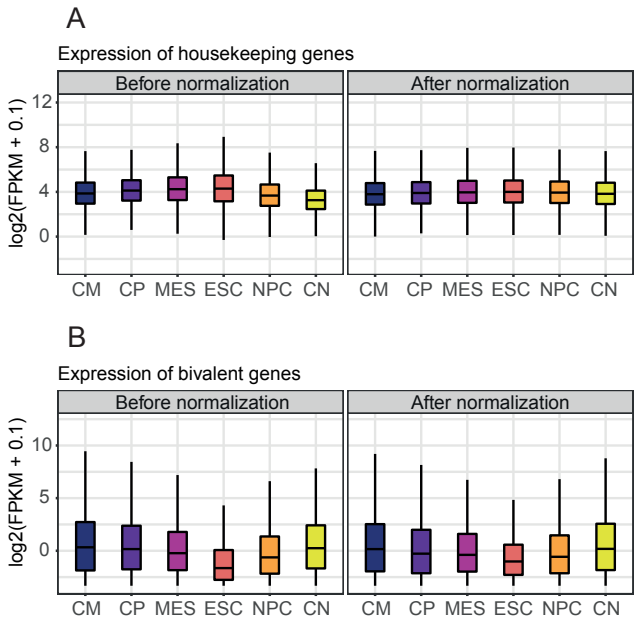

Figure R1.13: RNA-seq data before and after LOESS normalization. (A) Raw and normalized expression of 3,277 housekeeping genes along cardiac and neural differentiation from mESCs. (B) Raw and normalized expression of 3,459 bivalent genes along the same time points as A.

To further confirm the normalization efficiency, we tested its performance on two different subsets of genes: housekeeping genes and bivalent genes. We hypothesized that housekeeping genes would show a balanced distribution of expression after normalization, while bivalent genes would

increase their expression globally during differentiation (as some of them are activated). For this, we extracted a list of mouse housekeeping genes across 14 mouse tissues from the literature [156] to check their expression. We also evaluated the normalization on our list of bivalent genes (e.g., those associated to BPs). Indeed, after LOESS normalization, the expression of housekeeping genes was correctly balanced (Figure R1.13A), whereas the expression of bivalent genes maintained the characteristic pattern of increased expression across time (Figure R1.13B).

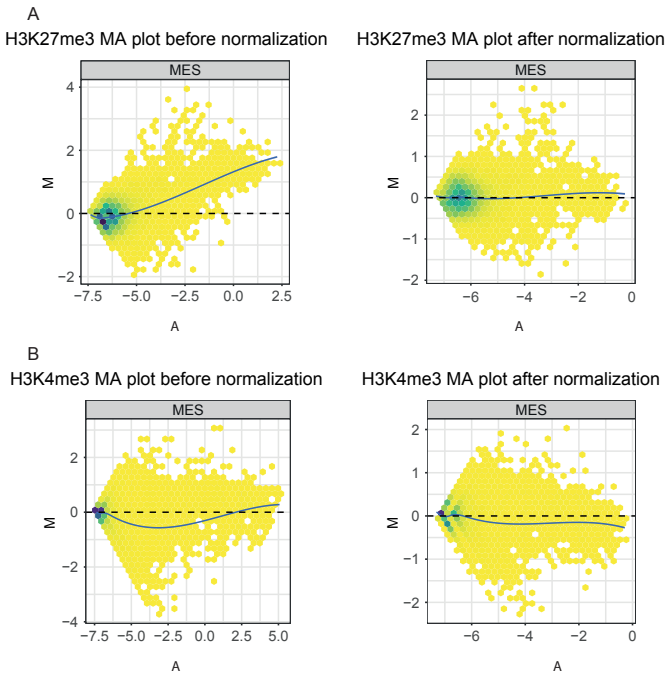

Figure R1.14: Performance of LOESS normalization in the ChIP-seq data. MA plots before and after normalization of MES against mESCs. M represents the $\log_2$ ratio of the intensities of the two samples, and A is the $\log_2$ of the average intensity. Intensity corresponds to normalized count of reads by total number of reads of the ChIP-seq samples of (A) H3K27me3, and (B) H3K4me3.

Once the validness of our approach was evaluated on expression data, we then ran the same normalization method on the ChIP-seq samples for H3K27me3, H3K4me3, H3K4me1, and H3K36me3 at the same time points. Examples of MA plots of MES against mESC before and after LOESS normalization over the full set of bins of 2 Kb at chromosome 19 are shown in Figure R1.14A for H3K27me3, and in Figure R1.14B for H3K4me3.

Similar to expression data, we evaluated our normalization method for the H3K4me3 and H3K27me3 ChIP-seq levels across differentiation on two different sets of genomic regions: the whole collection of bins of 2 Kb in which the genome is segmented, and the coordinates of our collection of BPs. We hypothesized that global ChIP-seq levels of the whole genome would become balanced irrespectively of the particular histone modification analyzed, while BPs should present a different pattern for H3K4me3 and H3K27me3 (e.g., increase and decrease of signal along differentiation time points, respectively, as a subset of the bivalent genes becomes activated during time). Indeed, after LOESS normalization, the levels of H3K4me3 along the whole genome were balanced (Figure R1.15A), whereas the same histone mark at BPs presented a clear pattern of increase during differentiation (Figure R1.15B). Notably, for H3K27me3, we observed the same balance in the levels along the whole genome after LOESS normalization (Figure R1.15C), while BPs exhibited a

pattern of decreased signal across differentiation, in contrast to that observed for H3K4me3 (Figure R1.15D). In all cases, therefore, there is a substantial improvement after applying this novel normalization approach. This suggests that the analysis solely based on data before normalization would be in many cases misleading.

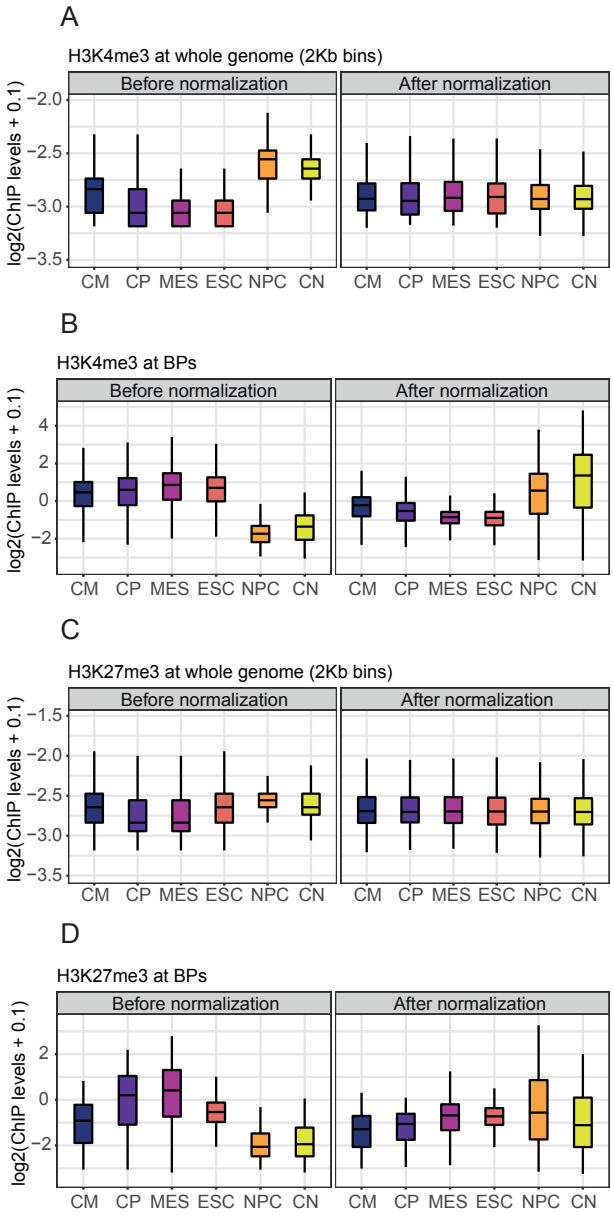

Figure R1.15: ChIP-seq data before and after LOESS normalization. (A) Raw and normalized H3K4me3 ChIP-seq signal levels at all 2-Kb bins of the genome. (B) Raw and normalized H3K4me3 ChIP-seq signal levels at 3,344 BPs. (C, D), same as A and B, respectively, but for H3K27me3. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors.

1.10 Poised enhancers and bivalent promoters are good predictors of gene expression during differentiation

After normalizing the data on expression and histone modifications across differentiation, we next generated predictive models using PEs and BPs for each differentiation time point. We used the Hi-C–top interactions involving PEs, BPs, and target genes in ESCs (1,846 PEs and 1,382 BPs associated with 1,434 target genes). From this dataset, a total of 1,063 protein-coding genes were used in the analysis. As the number of genes is smaller than in the previous gene sets, we decided to build the models at each cell type from the full set of genes to be evaluated in the rest of the differentiation time points. This approach has the advantage of allowing us to check whether the relationship between gene expression and histone marks in PEs is universal. This would be true if a model trained in a specific cellular context has a good performance in predicting gene expression in another one. We hypothesized that, as shown previously for promoters and gene bodies [22-24, 27], there is a universal relationship between gene expression and histone modifications at PEs.

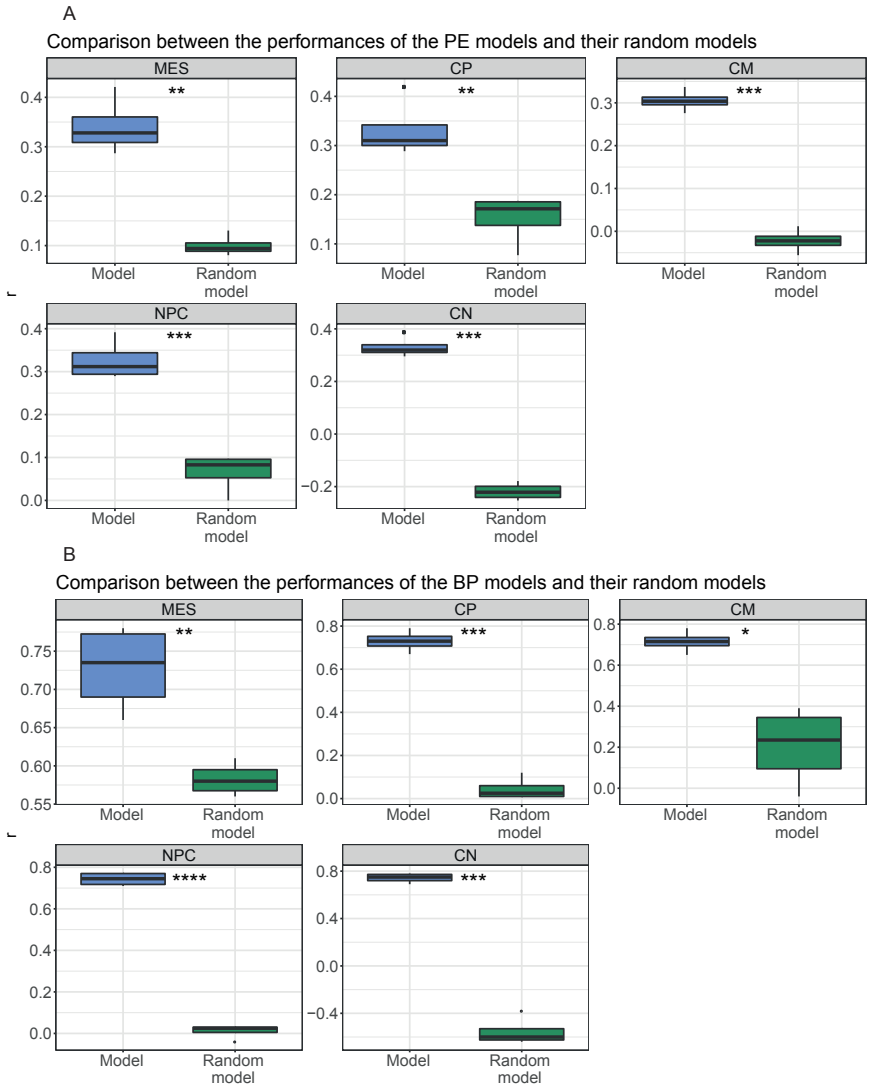

Figure R1.16: PE and BP models trained in differentiation time points. (A) Performance of each differentiation PE model on the rest of the differentiation time points as compared to performance over random models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Significance was assessed using a paired Student's t-test between the performance of the models and the performance of the random models paired by the differentiation test set (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05). CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors. (B) Same as in A, but for BP.

As a control, we randomized expression data and calculated predictive models for each time point. Next, we evaluated the performance of the randomized models on each differentiation dataset. We observed that predictive models for PEs and BPs obtained a significantly higher performance than randomized controls (Figure R1.16A, and R1.16B). Surprisingly, all PE models achieved the best performance in cardiomyocytes (Table R1.1), suggesting that cardiomyocyte expression is easier to predict than the expression set of other time points. Moreover, all PE models had similar performances at each time point (Table R1.1). These observations are also true for the BP models (Table R1.2). Taken together, our results indicate that there is a universal quantitative relationship between gene expression and histone modifications at PEs and BPs across cardiac and neural differentiation.

Table R1.1: Performance of each PE differentiation model at every differentiation time point

	MES	CP	CM	NPC	CN
MES model	-	0.34	0.42	0.32	0.29
CP model	0.3	-	0.42	0.32	0.29
CM model	0.3	0.34	-	0.31	0.28
NPC model	0.3	0.33	0.39	-	0.29
CN model	0.3	0.32	0.39	0.32	-

For each time point of cardiac (MES/CP/CM) and neural (NPC/CN) PE models (rows), the performance of the PE predictive models is shown for each cell type (columns). The performance values are represented as Pearson's correlation (*r*) between the measured expression and the predicted one. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors.

Table R1.2: Performance of each BP differentiation model at every differentiation time point

	MES	CP	CM	NPC	CN
MES model	-	0.78	0.77	0.66	0.77
CP model	0.74	-	0.79	0.67	0.72
CM model	0.72	0.78	-	0.65	0.71
NPC model	0.71	0.77	0.77	-	0.72
CN model	0.73	0.77	0.78	0.69	-

For each time point of cardiac (MES/CP/CM) and neural (NPC/CN) BP models (rows), the performance of the BP predictive models is shown for each cell type (columns). The performance values are represented as Pearson's correlation (*r*) between the measured expression and the predicted one. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors.

In order to confirm the predictive capacity of distal PEs (>5 Kb from a TSS, a total of 2,287 PEs), we elaborated new distal PE models (Figure R1.17). Indeed, distal PEs maintained the predictive capacity of the original model. In this case, 486 protein-coding genes were included in the modelling.

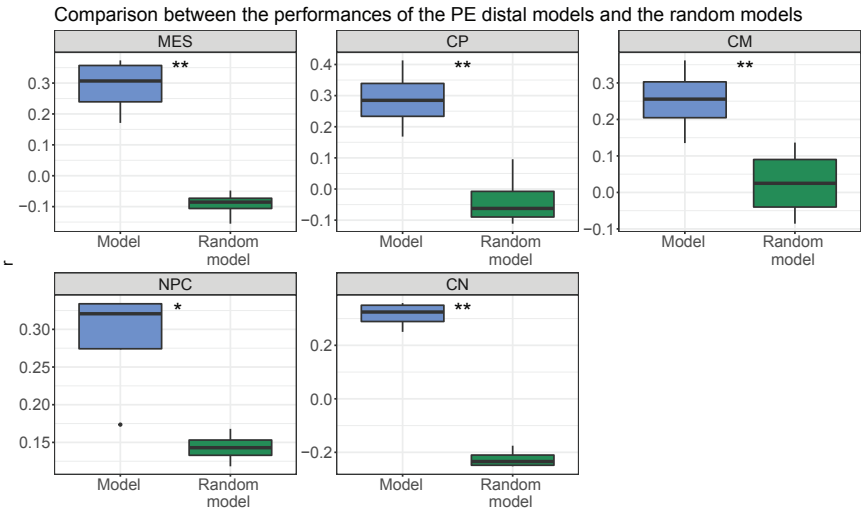

Figure R1.17: distal PE models trained using differentiation time points. Performance of each differentiation BP model on the rest of the differentiation time points as compared to the performance over the random models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Significance was assessed using a paired Student's t-test of the performance of the models or of the random models paired by a differentiation test set (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05). CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors.

Finally, we assessed the variable importance of the PE models for identifying differences in the contribution of each histone modification to the predictive models in different cellular contexts (Figure R1.18A). Strikingly, we observed that, in general, the two most important variables are H3K27me3 and H3K36me3. H3K36me3 was the most important histone modification for cardiac differentiation, whereas H3K27me3 was the most important for neural differentiation. In general, H3K27ac followed the above-mentioned histone modifications. H3K4me1 had a relatively low relevance to the predictive models, which suggests that it is involved in delimitating the

enhancer regions rather than in contributing to its function. H3K4me3, which was vastly associated with promoter activity, is accordingly the less informative mark for the prediction of gene expression using PEs, suggesting that H3K4me3 is not associated to enhancer activity.

The analysis of the variable importance in the BP models showed that H3K27me3, H3K27ac and, importantly, H3K4me3, were the most informative variables (Figure R1.18B). Our results suggested that the quantitative relationship between histone modifications and gene expression varies according to their location in PEs or BPs. Critically, even though there is a universal quantitative relationship between histone modifications and gene expression, this relationship can vary depending on the cellular context.

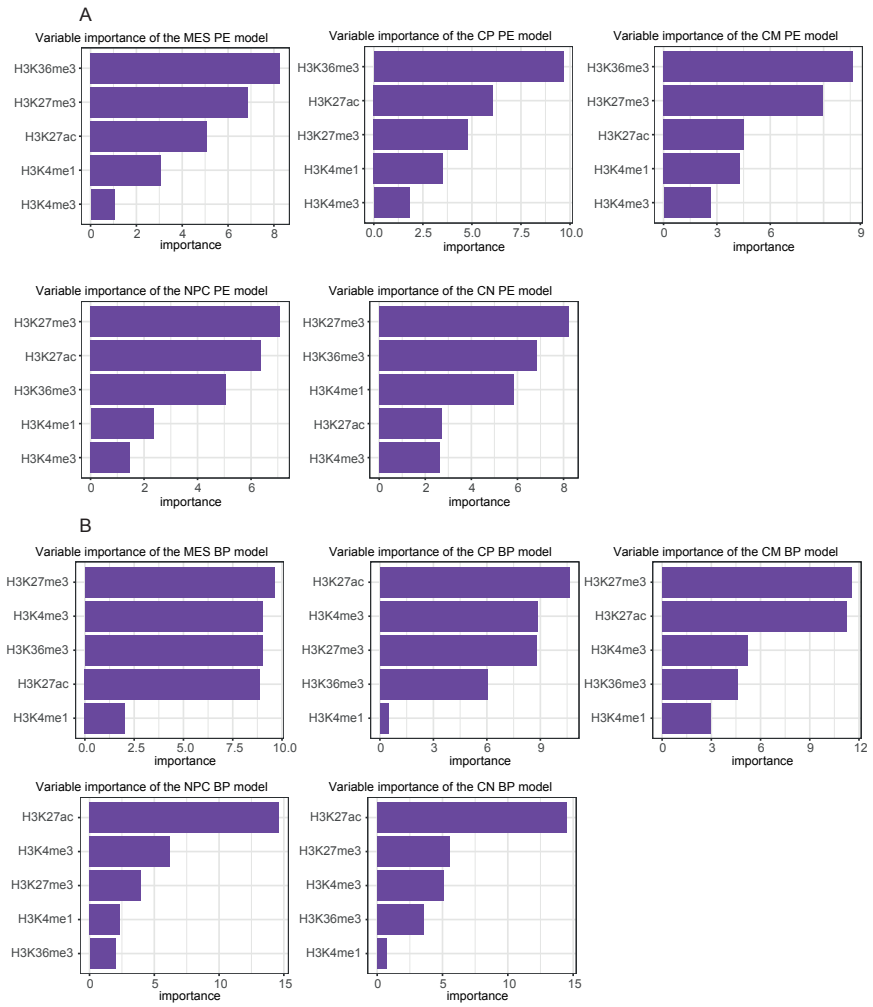

Figure R1.18: Variable importance of the PE and BP differentiation models. (A) Importance of histone modifications for each differentiation PE model. Importance is defined as the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistics for each model parameter. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors. (B) Same as in A, but for BP.

1.11 H3K27me3 is important to predict gene expression from both, intragenic and intergenic poised enhancers

The importance of H3K36me3 in the PE differentiation models might be explained by the fact that almost 60% of the PEs that entered the modelling are located within gene bodies. As H3K36me3 is located in the gene body of active genes [18, 19], intragenic enhancers also become marked when the genes start to be expressed during differentiation. To explore this scenario, we divided PEs into two groups, intragenic or intergenic, and built new PE predictive models. Both, intragenic and intergenic models were capable of predicting gene expression under similar parameters of performance (Figure R1.19).

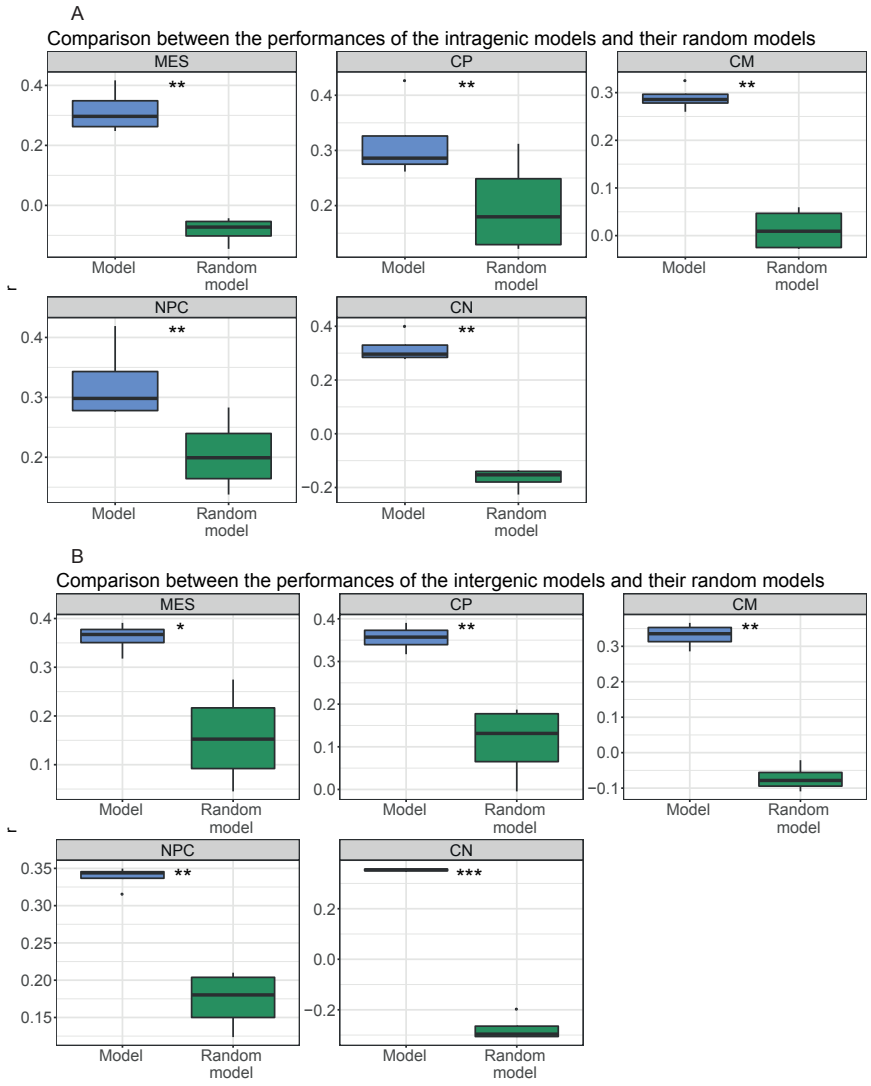

Figure R1.19: Intragenic and intergenic PE models trained using differentiation time points. (A) Performance of each differentiation intragenic model on the rest of the differentiation time points as compared to the performance over the random models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Significance was assessed using a paired Student's t-test of the performance of the models or of the random models paired by a differentiation test set (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05). CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors. (B) Same as in A, but for intergenic PE.

When assessing for variable importance, we observed that, as expected, H3K36me3 maintained its high contribution in the intragenic predictive models (Figure R1.20A). However, H3K36me3 importance was reduced in the intergenic predictive models (Figure R1.20B). On the contrary, H3K27me3 maintained its importance in both, intergenic and intragenic models (Figure R1.20). Thus, H3K27me3, and not H3K36me3, behaves as a truly universal predictor of PE activity.

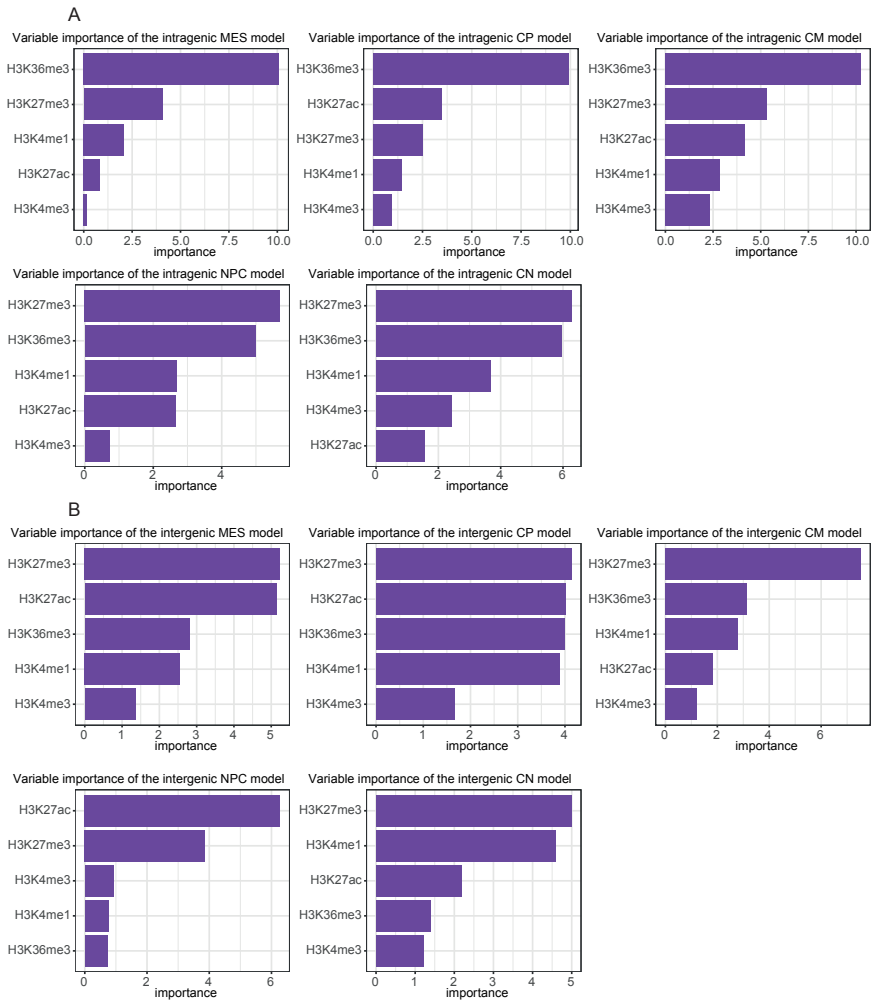

Figure R1.20: Variable importance of the intragenic and intergenic PE differentiation models. (A) Importance of histone modifications for each differentiation intragenic model. Importance is defined as the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistics for each model parameter. CM, cardiomyocytes; CN, cortical neurons; CP, cardio precursors; MES, mesoderm; NPC, neural precursors. (B) Same as in A, but for intergenic PEs.

1.12 Poised enhancers and bivalent promoters are good predictors of gene expression in mouse embryonic tissues

In order to extend our findings from *in vitro* differentiation to *in vivo*, we learnt predictive PE and BP models from mouse developmental stages at different tissues. We downloaded from ENCODE [157] ChIP-seq data of H3K27me3, H3K27ac, H3K36me3, H3K4me3 and H3K4me1 and RNA-seq on mouse embryos: heart tissue from 10.5 embryonic day (Heart10.5), liver tissue from 11.5 embryonic day (Liver11.5), neural tube tissue from 12.5 embryonic day (NeuralTube12.5), kidney tissue from 14.5 embryonic day (Kidney14.5), and lung tissue from 15.5 embryonic day (Lung15.5). We first normalized the ChIP-seq and expression data following the LOESS approach. In here, a total of 1,087 protein-coding genes entered the analysis. We observed that PEs were also predicting gene expression during mouse embryo development (Figure R1.21A). Again, when assessing for variable importance, we found that H3K27me3 was contributing the most, followed by H3K27ac and H3K36me3 (Figure R1.21B).

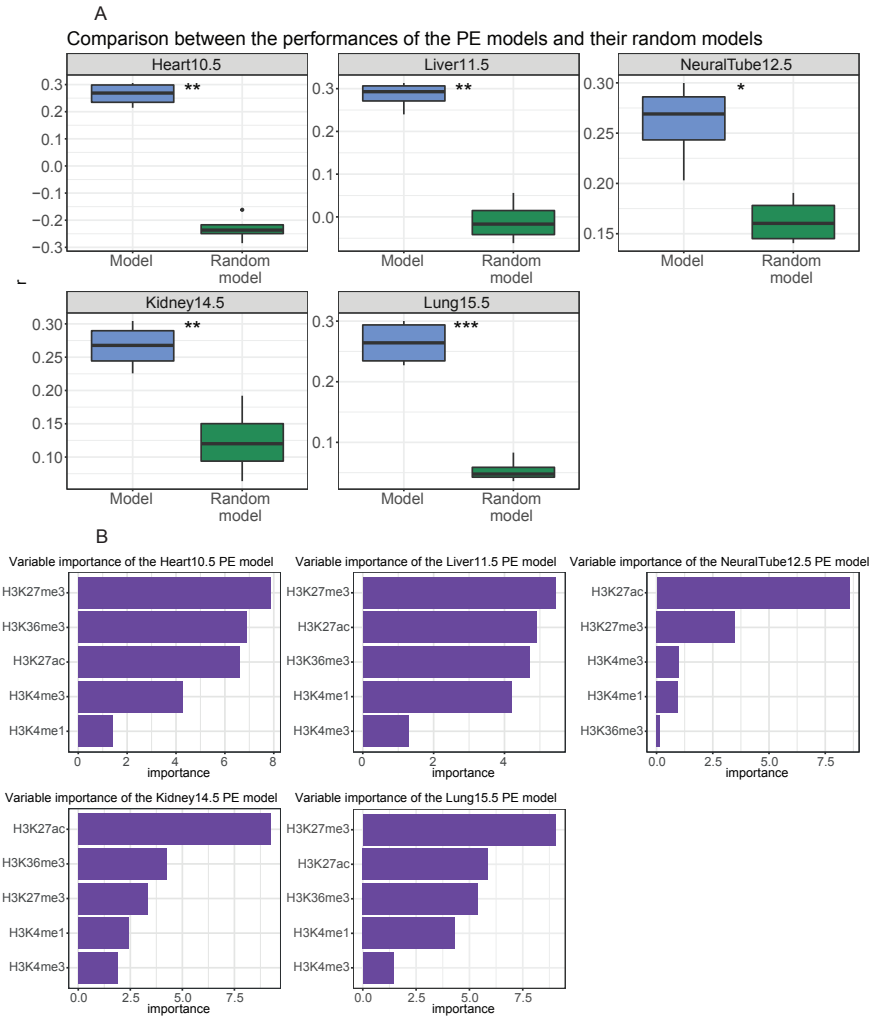

Figure R1.21: PE models trained using developmental stages. (A) Performance of each differentiation PE model on the rest of the developmental stages as compared to the performance over the random models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Significance was assessed using a paired Student's t-test of the performance of the models or of the random models paired by a differentiation test set (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05). (B) Importance of histone modifications for each development PE model. Importance is defined as the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistics for each model parameter. Heart10.5, heart tissue from 10.5 embryonic day; Kidney14.5, kidney tissue from 14.5 embryonic day; Liver11.5, liver tissue from 11.5 embryonic day; Lung15.5, lung tissue from 15.5 embryonic day; NeuralTube12.5, neural tube tissue from 12.5 embryonic day.

Next, we confirmed that BPs were also predicting gene expression during mouse embryo development (Figure R1.22A). In this case, the most predictive variable was H3K27ac followed by H3K27me3 and H3K4me3 (Figure R1.22B). Therefore, we further confirmed that differences in variable importance between PE models and BP models exist, which suggests that different histone modifications relate better to PE and BP function, respectively.

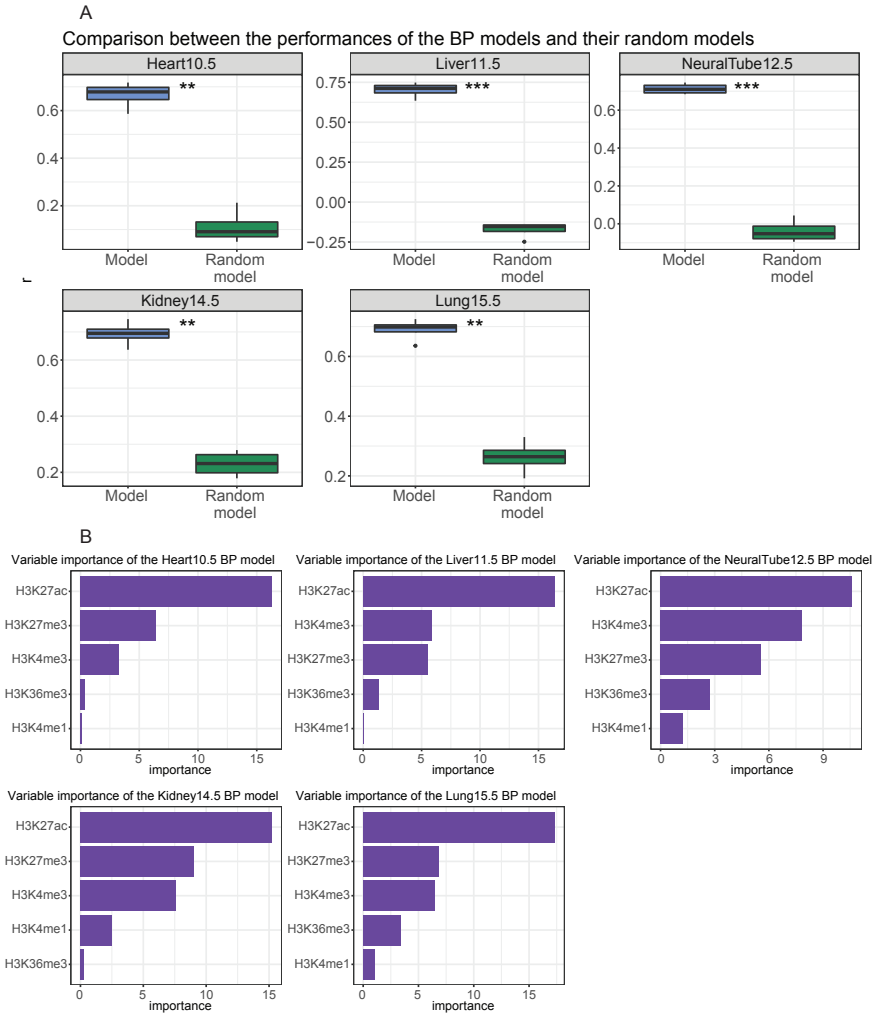

Figure R1.22: BP models trained using developmental stages. (A) Performance of each differentiation BP model on the rest of the developmental stages as compared to the performance over the random models. Performance is represented as Pearson's correlation (r) between predicted expression and measured expression. Significance was assessed using a paired Student's t-test of the performance of the models or of the random models paired by a differentiation test set (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05). (B) Importance of histone modifications for each development BP model. Importance is defined as the contribution of each variable in the linear regression predictive model and corresponds to the absolute value of the t-statistics for each model parameter. Heart10.5, heart tissue from 10.5 embryonic day; Kidney14.5, kidney tissue from 14.5 embryonic day; Liver11.5, liver tissue from 11.5 embryonic day; Lung15.5, lung tissue from 15.5 embryonic day; NeuralTube12.5, neural tube tissue from 12.5 embryonic day.

CHAPTER 2

The results shown in this chapter correspond to objective number 2 of this thesis.

2.1 Definition of the collection of regulatory regions of study

In Chapter 1, we have used our pipeline of chromatin segmentation to identify a total of 3,344 BPs and 9,421 APs in mESCs. Moreover, we have found a total of 1,846 PEs, all of them associated to at least one BP by a Hi-C–top interaction. At the same time, 11,777 AEs have been associated to at least one AP by a Hi-C–top interaction. However, there is a third type of enhancers, called intermediate enhancers (IEs), solely marked by H3K4me1 in absence of H3K27ac and H3K27me3. As historically IEs were grouped in the same category as PEs [65], we decided to characterize separately both types of regions and also in comparison to AEs, APs and BPs. Thus, from our chromatin segmentation model in mESC (Figure R1.2), we considered the intermediate state (state 5) in order to identify IEs. We selected as IEs those segments of state 5 that were surrounded by the unmarked state (state 9), to discard broad peaks of H3K4me1 surrounding active states (Figure R1.4A). Similarly to the identification of PEs and AEs, we also required a minimum length of 600 bp, an overlap with a peak of p300 [124] in at least 1 bp, and that IEs did not overlap with a region ± 500 bp around an existing TSS according to RefSeq [149]. As target genes of IEs have not been unambiguously identified yet, we did not require a Hi-C–top interaction with any type of promoter. In sum, under our pipeline, we found 12,608 IEs in mESCs.

2.2 Histone modification enrichment over all types of regulatory regions

Next, we set out to characterize our collection of regulatory regions to gain insight into the role of each class of enhancers. First of all, we studied the epigenetic landscape of PEs, BPs, AEs, APs and IEs in terms of enrichment for ten different histone modifications: H2Bub, H3K27ac, H3K27me1, H3K27me2, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K79me2 and H4K20me3 (Figure R2.1). We observed that H2Bub is mainly enriched in APs compared to the rest of regulatory regions and to the whole genome. As expected, H3K27ac is mainly enriched in both types of active regions (APs and AEs), whereas H3K27me3 is enriched in the poised ones (BPs and PEs). Moreover, H3K27me1 and H3K27me2 are clearly depleted in PEs and BPs. H3K36me3 and H4K20me3, in average, do not show enrichment in any of the regulatory regions. H3K4me1 is enriched in all types of regulatory regions. H3K4me3 is mainly enriched in both types of promoters, nevertheless some enrichment is observed also at PEs and AEs. Finally, H3K79me2 is clearly depleted in PEs and BPs, while enriched in active regions, mainly APs.

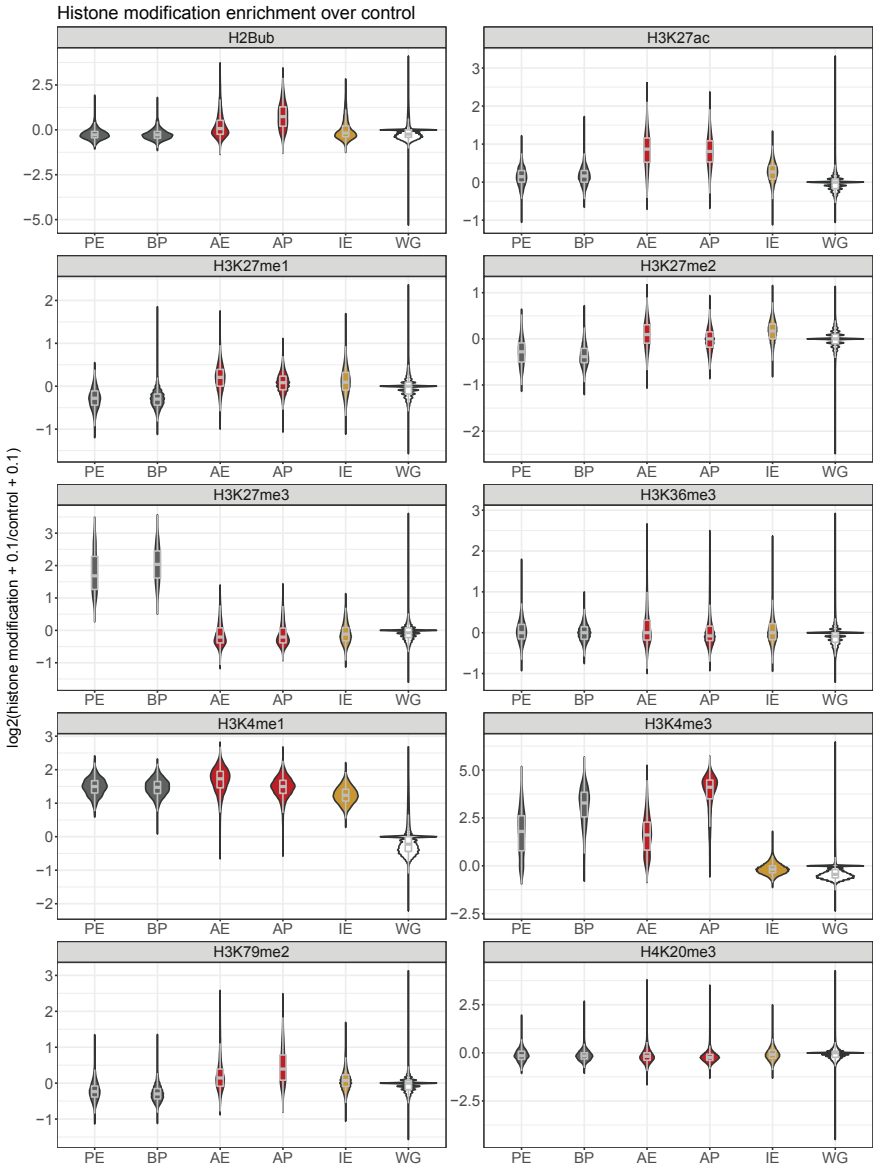

Figure R2.1: Histone modification enrichment at regulatory regions in mESCs. The ratio between the ChIP-seq signal strength of each histone modification over the intensity of the control at each class of regulatory region is shown. ChIP-seq of H3 was used as a control for H3 modifications and the input for H2Bub and H4K20me3. ChIP-seq levels correspond to number of reads normalized by total number of reads. WG (whole genome), enrichment at whole genome divided in bins of 2 Kb (which corresponds to average size of the regulatory regions).

2.3 Polycomb occupies poised enhancers

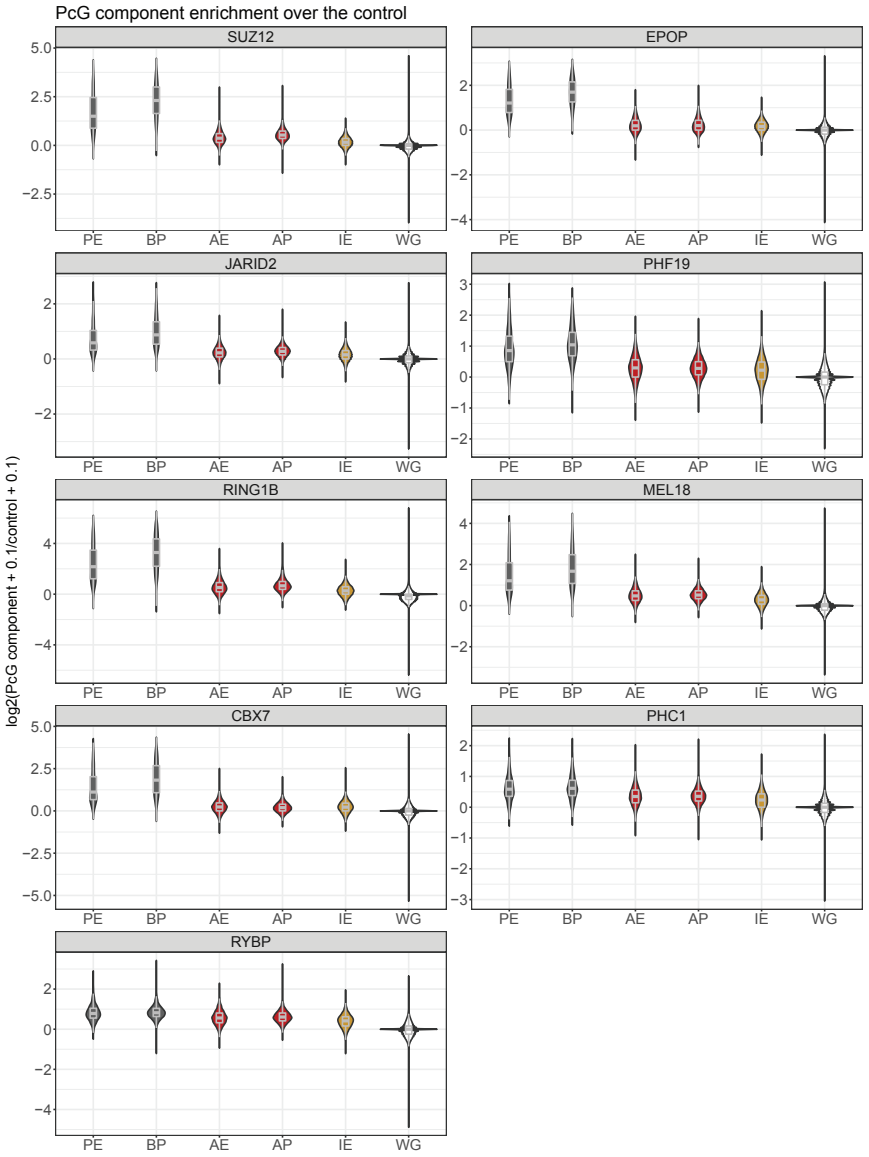

Figure R2.2: PcG component enrichment at regulatory regions in mESCs. The ratio between the ChIP-seq signal strength of distinct PcG components over the intensity of the control (input) at each class of regulatory region is shown. ChIP-seq levels correspond to number of reads normalized by total number of reads. WG (whole genome), enrichment at whole genome divided in bins of 2 Kb (which corresponds to average size of the regulatory regions).

All PRC1 and PRC2 subunits are known to bind BPs. Interestingly, several PRC2 components have already been shown to occupy PEs, such as SUZ12, EZH1, EZH2, JARID2, and PHF19, and also RING1B from PRC1 [124-126]. Indeed, we confirmed PRC2 occupancy and remarkably showed that, besides RING1B, other PRC1 components such as CBX7, MEL18, PHC1 and RYBP are also located at PEs (Figure R2.2) [115, 117, 118, 158]. As expected, PcG subunits are generally depleted from AEs and APs. Interestingly, IEs, which were historically associated to poised states of regulation, are depleted in PcG components too.

Previous work from our lab showed that upon *Phf19* knock-down, there is a decrease in H3K27me3 signal in BPs from mESCs [115]. Here, we confirmed that this decrease happens also in PEs (Figure R2.3A). Moreover, this decrease is significant with a $p < 2.2e\text{-}16$ (Wilcoxon Test) in both types of regions. Interestingly, our lab also documented in another publication that upon *Epop* knock-down, there is an increase in H3K27me3 signal in BPs from mESCs [118]. We again confirmed that this increase happens in both, PEs and BPs (Figure R2.3B). Again, this increase is significant with a $p < 2.2e\text{-}16$ (Wilcoxon Test) in both types of regions.

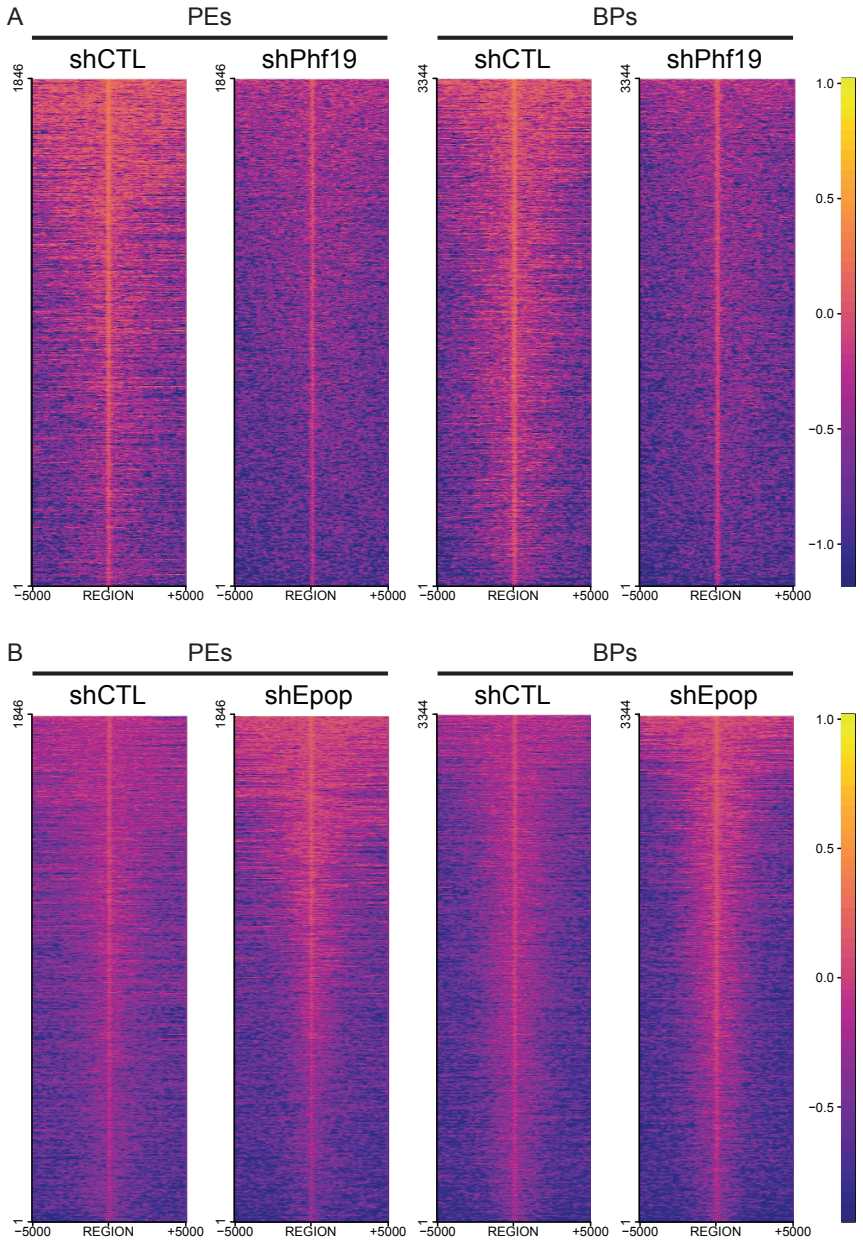

Figure R2.3: Effect of *Phf19* and *Epop* knock-downs over H3K27me3 in mESCs at PEs and BPs. ChIP-seq heatmaps showing H3K27me3 signal on PEs and BPs (centered on the summit) ± 5 Kb. The signal is normalized by the total number of reads. (A) Knock-down of *Phf19* (shPhf19) and its control (shCTR). (B) Knock-down of *Epop* (shEpop) and its control (shCTR).

2.4 Poised enhancers are enriched in epigenetic factors associated to transcriptional activation

As p300 is a factor associated to transcriptional activation used also to identify PEs and IEs by us and others [124], we next studied the enrichment in our collection of regulatory elements of several chromatin remodelers involved in transcriptional activation such as CHD2, CHD7, CHD8 and p400 [159, 160]. Moreover, we also studied the enrichment of architectural proteins such as CTCF –which is involved in the establishment of genome 3D interactions [123, 161]– and YY1 –which is involved in regulating active enhancer-promoter loops [162]– (Figure R2.4). We found that CHD2 is mainly enriched in active regions, whereas CHD7 is also enriched in PEs, BPs and IEs. Interestingly, CHD8 and p400 are mainly enriched in PEs, BPs, AEs and APs, whereas the enrichment in IEs is limited. CTCF is not enriched in any of the regulatory regions. Interestingly, YY1 seems to have certain degree of enrichment in PEs and BPs, besides the expected enrichment in AEs and APs. Thus, PEs seem to be more enriched in epigenetic factors associated to transcriptional activation than IEs.

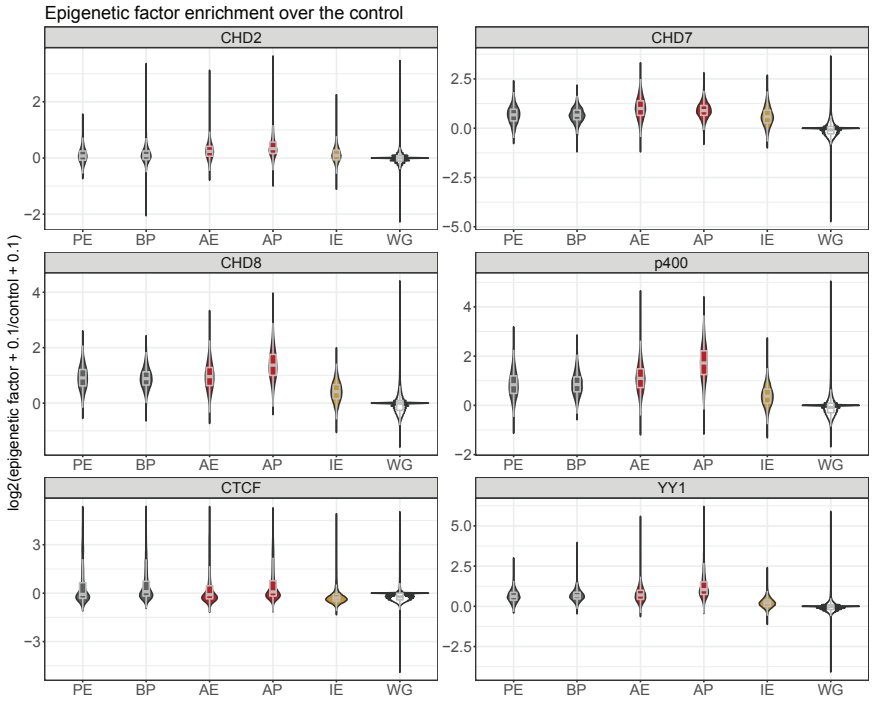

Figure R2.4: Epigenetic factor enrichment at regulatory regions in mESCs. The ratio between the ChIP-seq signal strength of distinct epigenetic factors over the intensity of the control (input) at each class of regulatory region is shown. ChIP-seq levels correspond to number of reads normalized by total number of reads. WG (whole genome), enrichment at whole genome divided in bins of 2 Kb (which corresponds to average size of the regulatory regions).

2.5 Paused RNA polymerase II is enriched at poised enhancers

Paused RNA polymerase II (POL2 S5P) has been identified in the promoters of active genes, but also in BPs [95, 163]. Therefore, we studied the enrichment of POL2 S5P and the elongating RNA polymerase II (POL2 S2P) in all types of

regulatory regions (Figure R2.5) [163]. Interestingly, we found that POL2 S5P is not only enriched in APs and BPs, but also in PEs and AEs. Thus, IEs can be distinguished from the rest of regulatory elements by the absence of POL2 S5P. POL2 S2P was depleted in PEs and BPs, consistent with the repressed state of these regulatory regions.

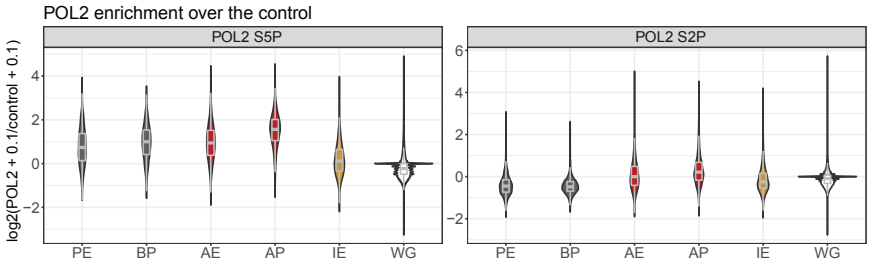

Figure R2.5: POL2 enrichment at regulatory regions in mESCs. The ratio between the ChIP-seq signal strength of POL2 modifications over the intensity of the control (input) at each class of regulatory region is shown. ChIP-seq levels correspond to number of reads normalized by total number of reads. WG (whole genome), enrichment at whole genome divided in bins of 2 Kb (which corresponds to average size of the regulatory regions).

2.6 Poised enhancers are open chromatin regions

Next, we compared which one of the classes of regulatory regions presents a higher degree of accessibility (Figure R2.6). To do this, we quantified their ATAC-seq signal using published data on mESCs [164]. In general, promoters are more accessible than enhancers. Interestingly, PEs and BPs are almost as accessible as AEs and APs, respectively, whereas chromatin over IEs seems more compact. These

observations, together with our previous observations, suggest that PEs are more primed for activation than IEs.

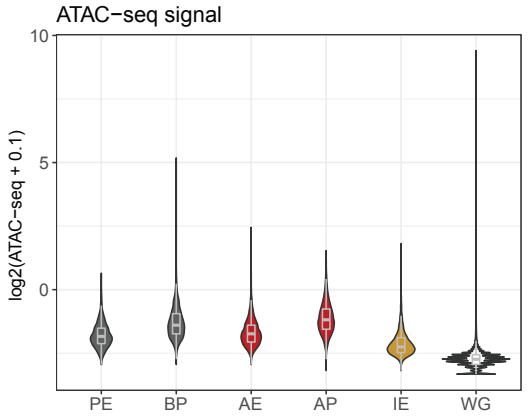

Figure R2.6: ATAC-seq signal at regulatory regions in mESCs. ATAC-seq signal corresponds to number of reads normalized by total number of reads. WG (whole genome), enrichment at whole genome divided in bins of 2 Kb (which corresponds to average size of the regulatory regions).

2.7 Poised enhancers are one of the most conserved regulatory regions

Next, we studied the evolutionary conservation of all types of regulatory regions. We found that in general, poised regions are more conserved than active regions or IEs (Figure R2.7A). Strikingly, PEs display the highest average conservation. Thus, we hypothesized that this could be due to 60% of PEs being intragenic, and therefore, the high conservation could be caused by the overlap with coding sequence. However, we calculated average conservation in both, intergenic and

intragenic PEs separately, and confirmed that even intergenic PEs are one of the most conserved regions (Figure R2.7B). Interestingly, intergenic PEs have an average conservation similar to APs. These findings suggest that PEs, as BPs, have a key role in development.

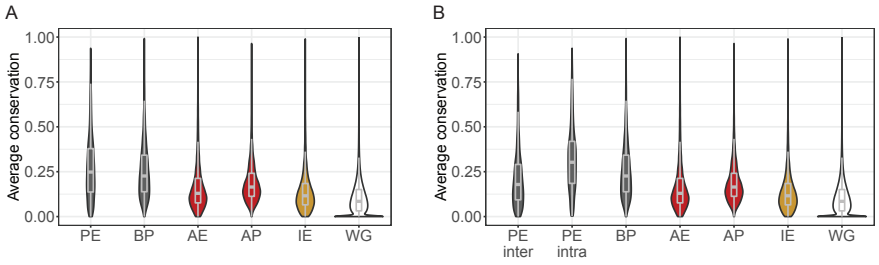

Figure R2.7: Average conservation at regulatory regions. (A) Average conservation at PEs, BPs, AEs, APs, IEs and whole genome (WG) divided in 2 Kb bins (which corresponds to average size of the regulatory regions). (B) Same as in A, but PEs are divided into two groups: intergenic (PE inter) and intragenic (PE intra). Conservation levels were determined using the PhastCons track of the UCSC genome browser [165].

2.8 Poised enhancers colocalize with CpG islands

Recently, it has been shown that CpG islands (CGIs) play a key role in PE function [166]. Thus, we calculated the percentage of each type of regulatory region overlapping with a CGI (Figure R2.8). Our results for PEs are comparable to those previously described [166]. In addition, we found that PEs are the enhancer type that colocalizes most frequently with a CGI.

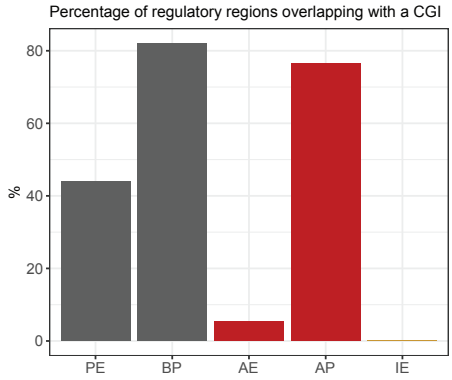

Figure R2.8: Colocalization between regulatory regions and CGIs. Percentage of regions (PEs, BPs, AEs, APs and IEs) overlapping with a CGI by at least 1 bp.

2.9 Identification of poised enhancers becoming active in differentiation

PE activation has generally been related to differentiation towards neural lineage [124-127]. However, recent evidence suggests that this mechanism could be more general [128]. Thus, we aimed to identify PEs switching towards an active state during differentiation towards neural but also cardiac lineages.

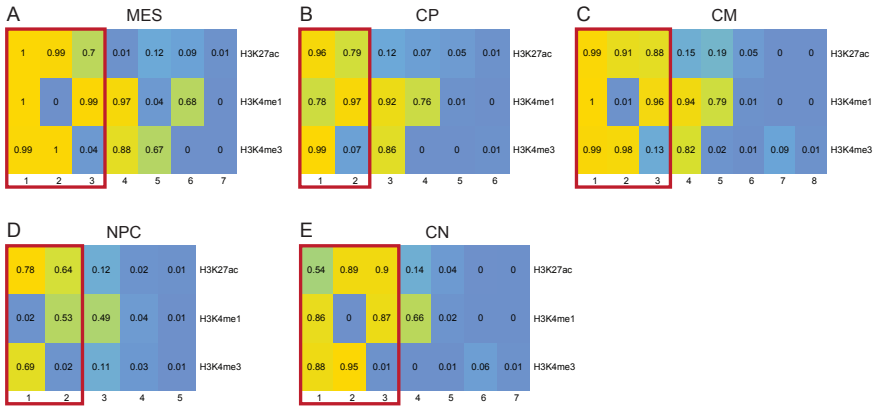

Figure R2.9: Chromatin segmentation models for cardiac and neural lineages. (A) State definition of the chromatin segmentation model of MES. The values represent the probability (from 0 to 1) of finding each histone modification (vertical) in genomic segments of the states (horizontal). The red rectangle denotes states with more than 0.5 probability to have H3K27ac used to identify PEs and BPs becoming active. (B) As in A, but for CPs. (C) As in A, but for CMs. (D) As in A, but for NPCs. (E) As in A, but for CNs.

First of all, we generated chromatin segmentation models of MES, CPs, CMs, NPCs and CNs with the ChromHMM software [49], using ChIP-seq data of H3K27ac, H3K4me1 and H3K4me3 (Figure R2.9). We did not introduce H3K27me3 into the models as we observed colocalization of H3K27me3 and H3K27ac (Figure R2.10), probably caused by population heterogeneity. Therefore, we reasoned this colocalization could mask the gain of H3K27ac in the resulting chromatin segmentation models.

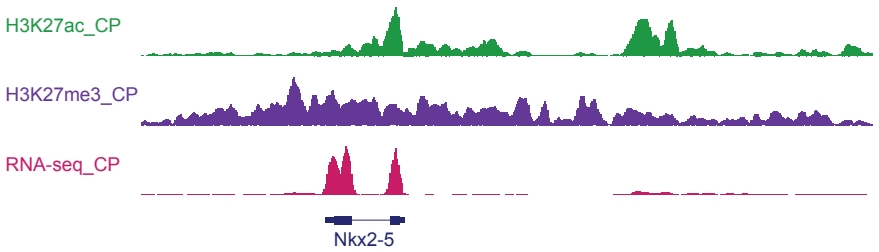

Figure R2.10: Example region of colocalization between H3K27me3 and H3K27ac in CP. The region contains the CP marker gene *Nkx2-5*, which has peaks of H3K27ac and is expressed, although the region seems to be covered by H3K27me3 at the same time.

Next, we selected those states with more than 0.5 probability to have H3K27ac as active states (Figure R2.9). We did not use states enriched in H3K4me3 without H3K27ac as they could be considered to be bivalent. We identified BPs and PEs becoming active during differentiation when they overlap in at least 600 bp with regions of genome covered by active states. The total number of BPs switching towards and active state at each differentiation time point are in Table R2.1, as well as their associated PEs that also become active.

Table R2.1: Number of BPs and associated PEs becoming active during differentiation and their target genes.

	MES	CP	CM	NPC	CN
BP to AP	2,229	1,369	1,796	1,106	2,465
Target genes (BP to AP)	2,248	1,450	1,888	1,179	2,598
PE to AE	561	278	354	160	587
Target genes (PE to AE)	538	244	341	144	590

Number of BPs, associated PEs becoming active at each differentiation time point: MES, CPs, CMs, NPCs and CNs; and number of target genes.

Finally, we confirmed that BPs (Figure R2.11A) and PEs (Figure R2.11B) becoming active decrease their H3K27me3 levels with respect to mESCs (denoted by the negative fold change). Likewise, BPs and PEs that are not switching towards an active state (inactive BPs and PEs) present similar H3K27me3 levels (denoted by a fold change close to 0). Moreover, the differences in H3K27me3 fold change between active and inactive regions are significant (Figure R2.11). Therefore, inactivated PEs retain H3K27me3, whereas the active ones lose it.

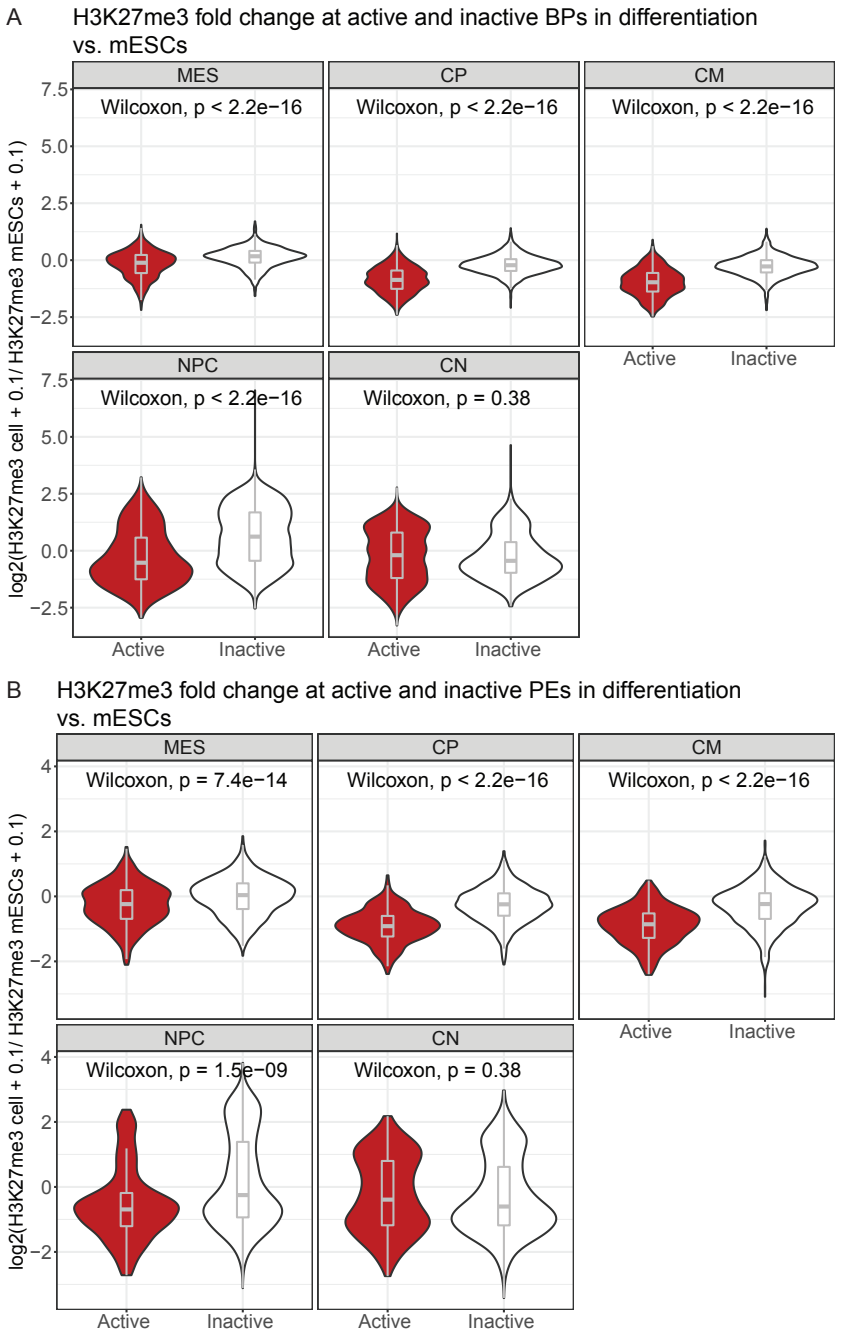

Figure R2.11: H3K27me3 at PEs and BPs during differentiation. (A) Fold change of H3K27me3 between each differentiation time point and mESC at activated BPs (red) and inactive BPs (white). Activated BPs are those that switch towards an active state in each differentiation time point. (B) As for A, but in the PEs.

2.10 Poised enhancer activation is more cell type-specific than bivalent promoter activation

We performed a GO term enrichment analysis on the target genes of activated BPs and PEs at each differentiation time point. As expected, among the most significant categories we found terms related to development and differentiation (Figure R2.12). Interestingly, for the cardiac lineage, the top GO categories related to differentiation and development where strongly associated to cardiac differentiation in the target genes of the activated PEs than in those of the activated BPs (Figure R2.12A-C). This finding suggests that PE activation is more cell type-specific than BP activation.

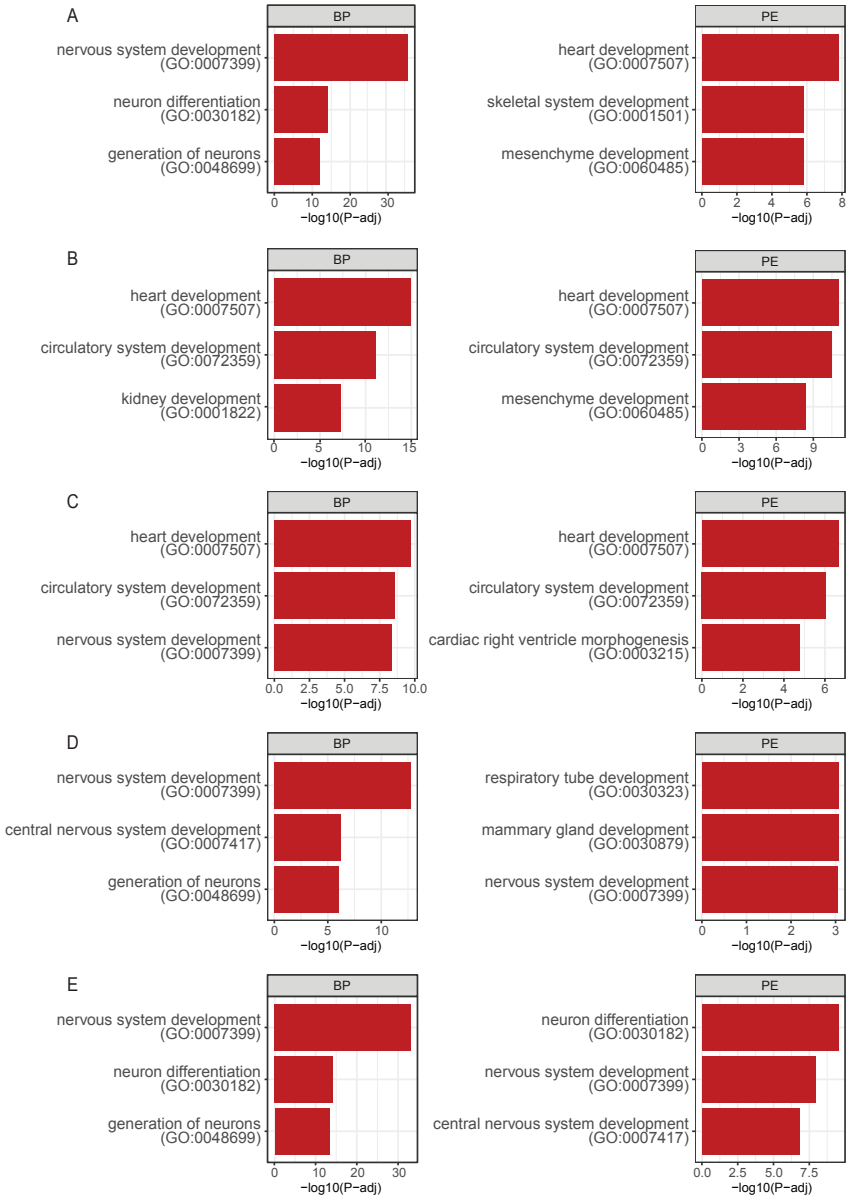

Figure R2.12: Functional analysis of the target genes of the BPs and PEs becoming active during differentiation. (A) Top GO biological process 2018 categories enriched in target genes of activated BPs in MES (left) and activated PEs in MES (right). (B) As for A, but in CP. (C) As for A, but in CM. (D) As for A, but in NPC. (E) As for A, but in CN.

Surprisingly, we found that an important subset of genes activated their BPs at all differentiation time points (Figure R2.13A), while this shared subset of genes was relatively less abundant for PEs (Figure R2.13B). Moreover, we confirmed this observation not only at gene level but also when comparing the activation of the regulatory regions (Figure R2.13C-D). Among the genes activating their BPs in several differentiation time points, we found *Isl1*, whose associated BP is activated at both CMs and CNs, but whose PEs are only activated in CMs (Figure R2.13E). Interestingly, *Isl1* is expressed not only at CMs (11.63 FPKMs), but also at CNs (1.75 FPKMs). Although lower than at CMs –which is expected– *Isl1* expression at CNs is higher than at mESCs (0.51 FPKMs). This observation goes in line with the activation of the BP alone in CNs. Altogether, these results further suggest that PE activation is more cell type-specific than BP activation.

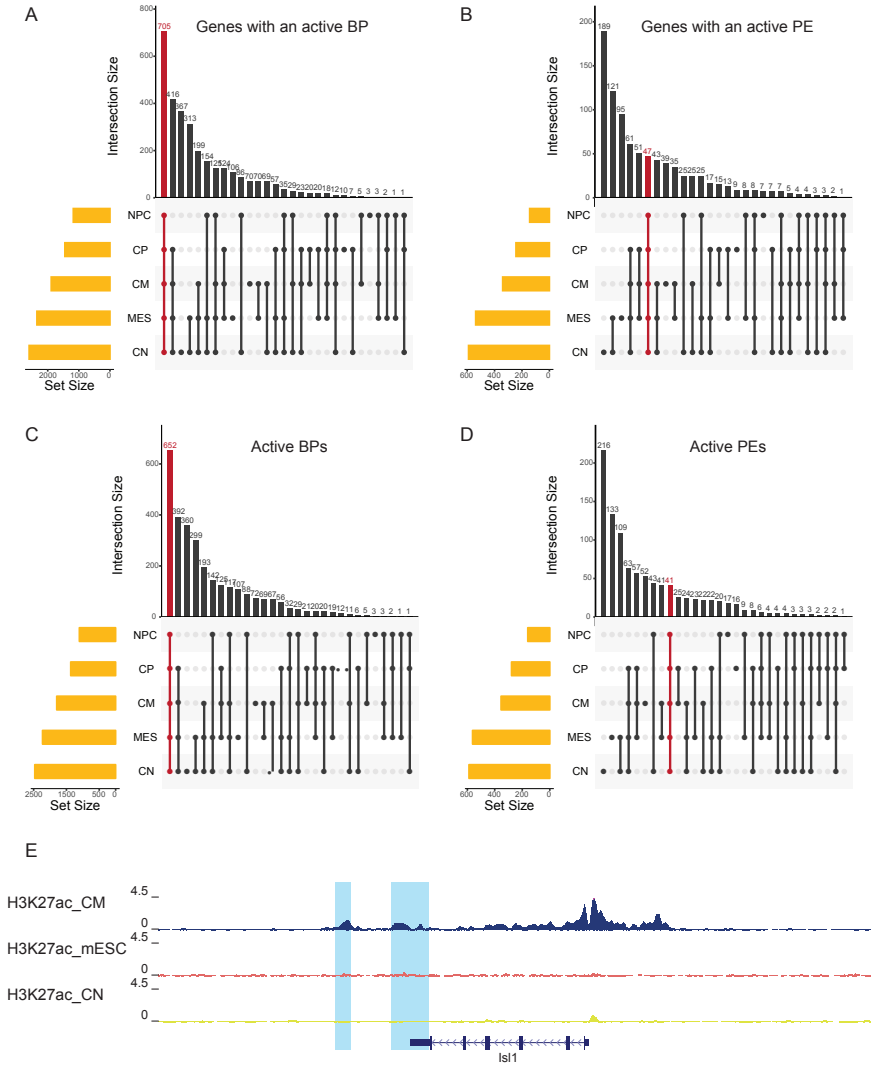

Figure R2.13: PE activation is more cell type-specific than BP activation. (A) Upset plot comparing overlap between the target genes of the activated BPs at each differentiation time point. (B) As for A, but in PEs. (C) Upset plot comparing the overlap between activated BPs at each differentiation time point. (D) As for A, but in PEs. (E) Example of a gene, *Isl1*, whose associated BP is activated in both CMs and CNs, but whose PEs are only activated in CMs. The PEs are highlighted in light blue.

2.11 The increase in expression of poised enhancer target genes is cell type-specific

In order to confirm that what we observed for *Isl1* gene was a general trend, we compared the increase in gene expression at each differentiation time point with respect to mESCs, between those genes for which we reported the activation of only their BP or their BP plus their PE. We confirmed that those genes whose PE (in addition to its BP) was activated have a higher increase in gene expression than those that only have their BP in an active state (Figure R2.14).

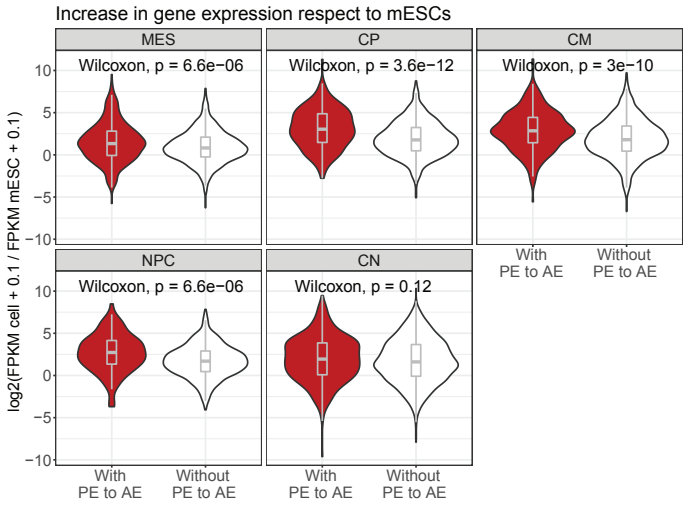

Figure R2.14: Genes activating their PE besides their BP have a higher increase in gene expression with respect to mESCs. Gene expression fold-change between each differentiation time point and mESCs at genes activating their BP alone (white), or both, their BP and their PE (red).

Finally, we wondered whether this higher increase in gene expression when activating the PE is cell type-specific. Thus,

we calculated the gene expression increase, of the target genes of one of the differentiation time points in the rest (Figures R2.15-19). We observed that such differences in the fold change are generally not significant in the rest of the time points, and mostly, they are not significant across the other lineage. Thus, we concluded that gene expression increase by PE activation is specific of the cell type. Interestingly, the activated BP target genes without an activated PE in CNs have a significantly higher increase in expression with respect to mESCs in CMs than those genes that also have an activated PE in CNs (Figure R2.19). This observation goes in line with gene expression increase by PE activation being specific of the cell type.

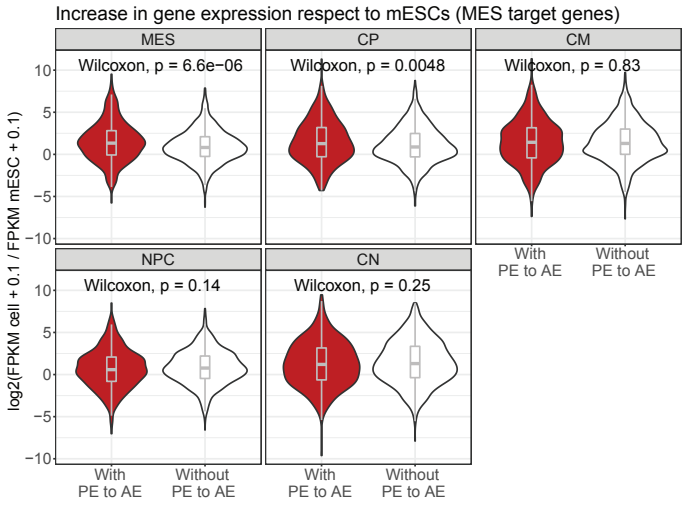

Figure R2.15: The higher gene expression increase observed at MES in its activated PE target genes is not observed in the rest of the differentiation time points. Gene expression fold-change between each differentiation time point and mESCs at MES target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in MES.

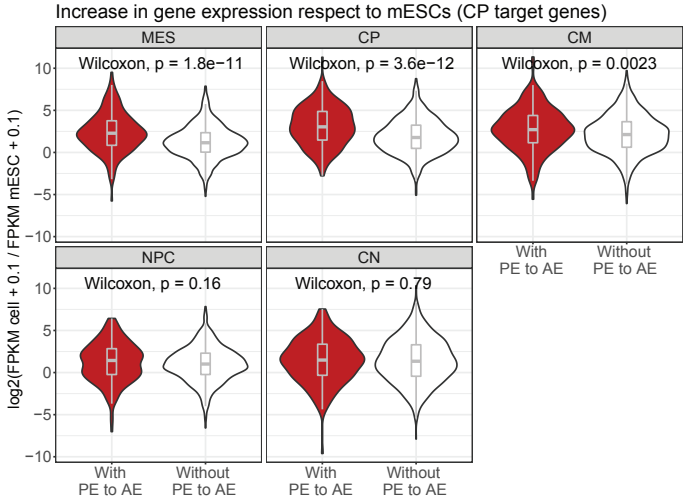

Figure R2.16: The higher gene expression increase observed at CP in its activated PE target genes is not observed in the rest of the differentiation time points. Gene expression fold-change between each differentiation time point and mESCs at CP target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in CPs.

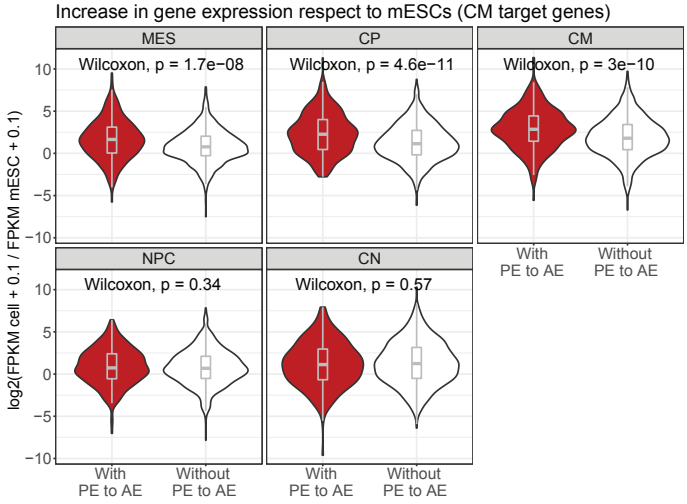

Figure R2.17: The higher gene expression increase observed at CM in its activated PE target genes is not observed in the neural lineage. Gene expression fold-change between each differentiation time point and mESCs at CM target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in CMs.

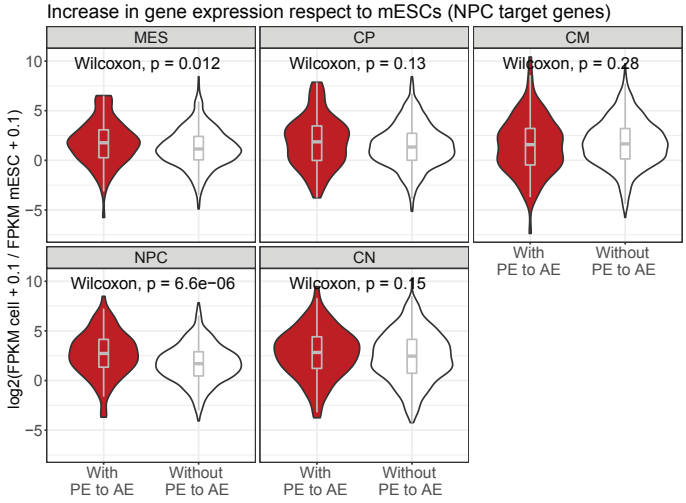

Figure R2.18: The higher gene expression increase observed at NPC in its activated PE target genes is not observed in the rest of the differentiation time points. Gene expression fold-change between each differentiation time point and mESCs at NPC target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in NPCs.

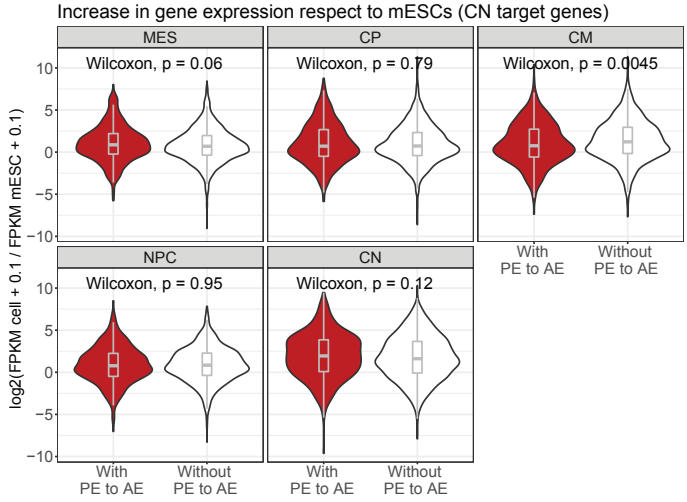

Figure R2.19: The higher gene expression increase observed at CN in its activated PE target genes is not observed in the rest of the differentiation time points. Gene expression fold change between each differentiation time point and mESCs at CN target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in CNs.

2.12 Commonly activated bivalent genes in differentiation have the same activated poised enhancer at all differentiation time points

Enhancers are generally cell type-specific, and even the expression of the same gene can be governed by different enhancers in different cell types [68]. Therefore, we wondered whether the expression of commonly activated bivalent genes (those associated to a BP) is driven by the same PE or by different ones. Surprisingly, we found that common genes with an activated PE generally share the same enhancers (Figure R2.20A). Among the common genes activating the same PE at all differentiation time points we found *Cbx4* (Figure R2.20B), which accordingly, is expressed in MES (1.28 FPKMs), CPs (3.55 FPKMs), CMs (7.10 FPKMs), NPCs (1.51 FPKMs) and CNs (5.20 FPKMs) but not in mESCs (0.1 FPKMs).

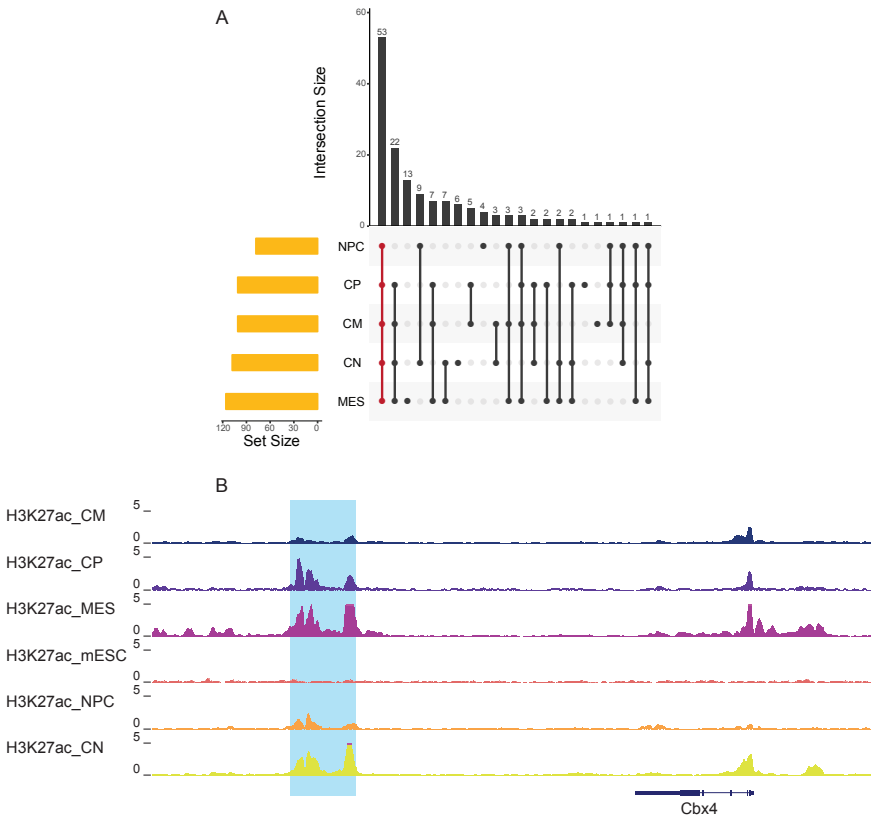

Figure R2.20: Common genes activated at all differentiation time point share the activation of the same PE. (A) Upset plot comparing the overlap between activated PEs at each differentiation time point that govern common genes. (B) Example of a gene, *Cbx4*, which activates the same PE in all the differentiation time points. The PE is highlighted in light blue.

2.13 Identification of poised enhancers becoming active at mouse embryonic tissues

We next aimed to demonstrate that our findings on *in vitro* differentiation are also valid *in vivo*. As for differentiation, we generated chromatin segmentation models of mouse

embryonic tissues at different embryonic days (Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5 and Lung15.5) [157] with the ChromHMM software [49], using ChIP-seq data of H3K27ac, H3K4me1 and H3K4me3 (Figure R2.21).

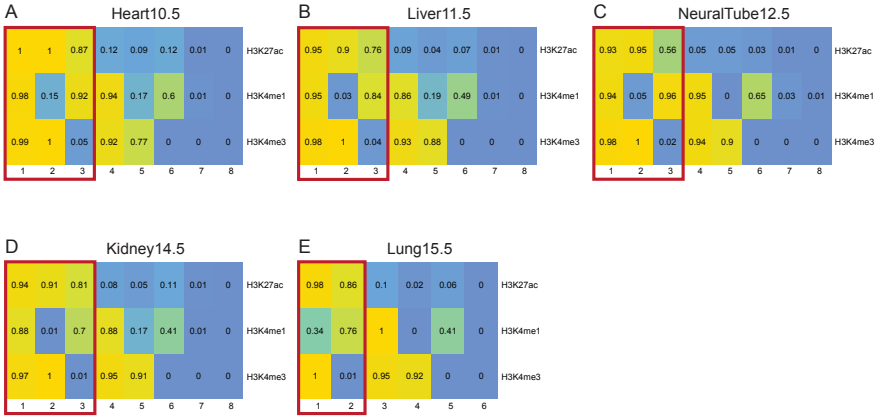

Figure R2.21: Chromatin segmentation models for embryonic tissues. (A) State definition of the chromatin segmentation model of Heart10.5. The values represent the probability (from 0 to 1) of finding each histone modification (vertical) in genomic segments of the states (horizontal). The red rectangle denotes states with more than 0.5 probability to have H3K27ac used to identify PEs and BPs becoming active. (B) As in A, but for Liver11.5. (C) As in A, but for NeuralTube12.5. (D) As in A, but for Kidney14.5. (E) As in A, but for Lung15.5.

Next, we selected those states with more than 0.5 probability to have H3K27ac as active states (Figure R2.21). As before, we identified BPs and PEs becoming active in embryonic tissues when they overlap in at least 600 bp with regions of genome covered by active states. The total number of BPs switching towards and active state at each embryonic tissue are in Table R2.2, as well as their associated PEs that also become active.

Table R2.2: Number of BPs and associated PEs becoming active in embryonic tissues and their target genes.

	Heart10.5	Liver11.5	NeuralTube12.5	Kidney14.5	Lung15.5
BP to AP	2,143	632	828	962	1,006
Target genes (BP to AP)	2,348	692	892	1,028	1,075
PE to AE	561	89	233	213	194
Target genes (PE to AE)	518	87	188	191	171

Number of BPs, associated PEs becoming active at each embryonic tissue: Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5 and Lung15.5; and number of target genes.

Finally, we confirmed that BPs (Figure R2.22A) and PEs (Figure R2.22B) becoming active presented significantly lower H3K27me3 signal than those that do not. Therefore, as in differentiation, inactive PEs retain H3K27me3, whereas the active ones lose it.

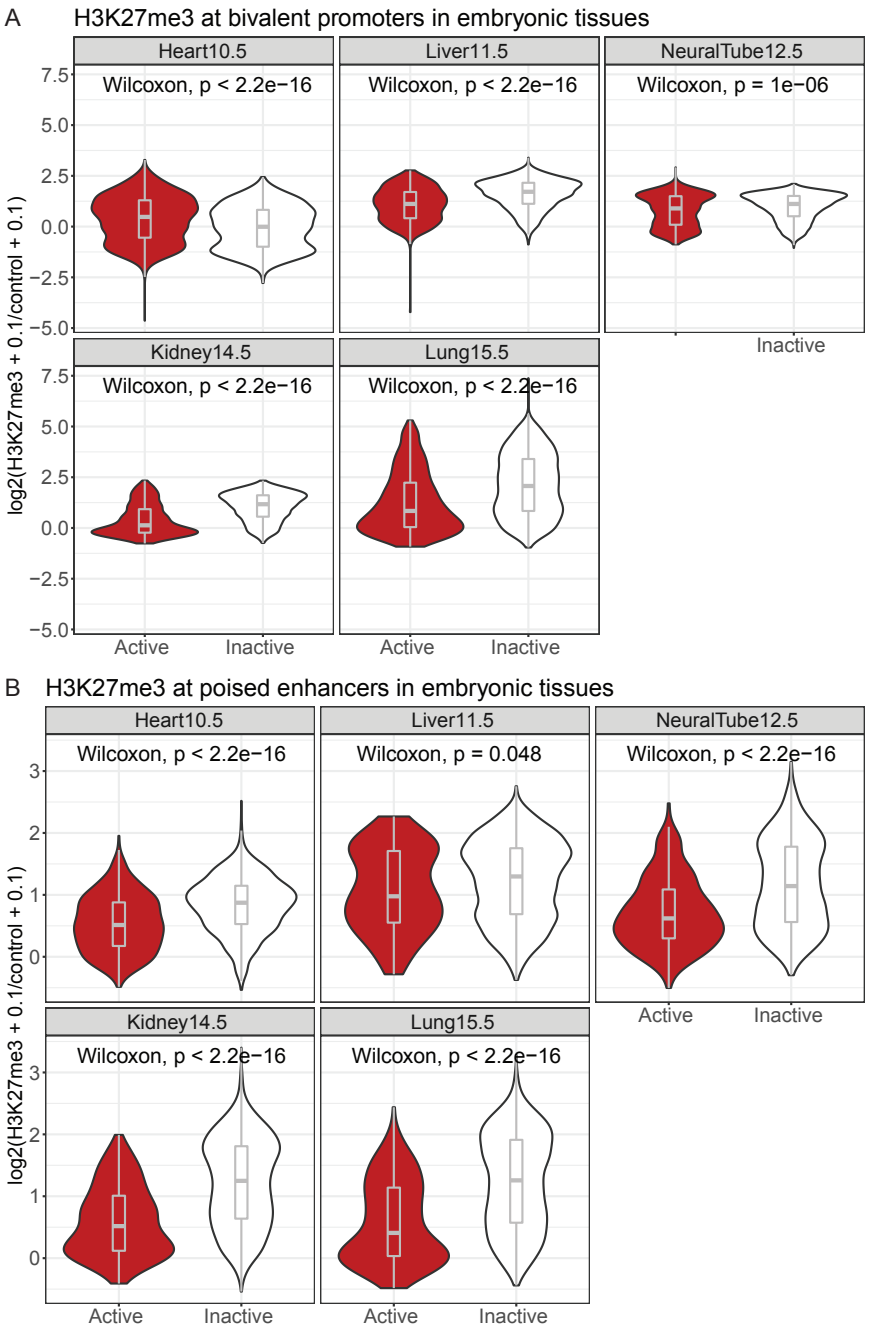

Figure R2.22: H3K27me3 at PEs and BPs in embryonic tissues. (A) Enrichment of H3K27me3 over control (input) in each embryonic tissue at activated BPs (red) and inactive BPs (white). Activated BPs are those that switch towards an active state in each embryonic tissue. (B) As for A, but in the PEs.

2.14 Poised enhancer activation is more tissue-specific than bivalent promoter activation

We performed a GO term enrichment analysis on the target genes of activated BPs and PEs in each embryonic tissue. As expected, among the most significant categories we found terms related to development and differentiation (Figure R2.23). As in differentiation, for some of the tissues such as Heart10.5 and Lung15.5, the top GO categories related to differentiation and development where more related to the specific tissue of study in the target genes of the activated PEs than in those of the activated BPs (Figure R2.23A and E). This finding reinforces again the model of gene regulation in which PE activation is more tissue-specific than BP activation.

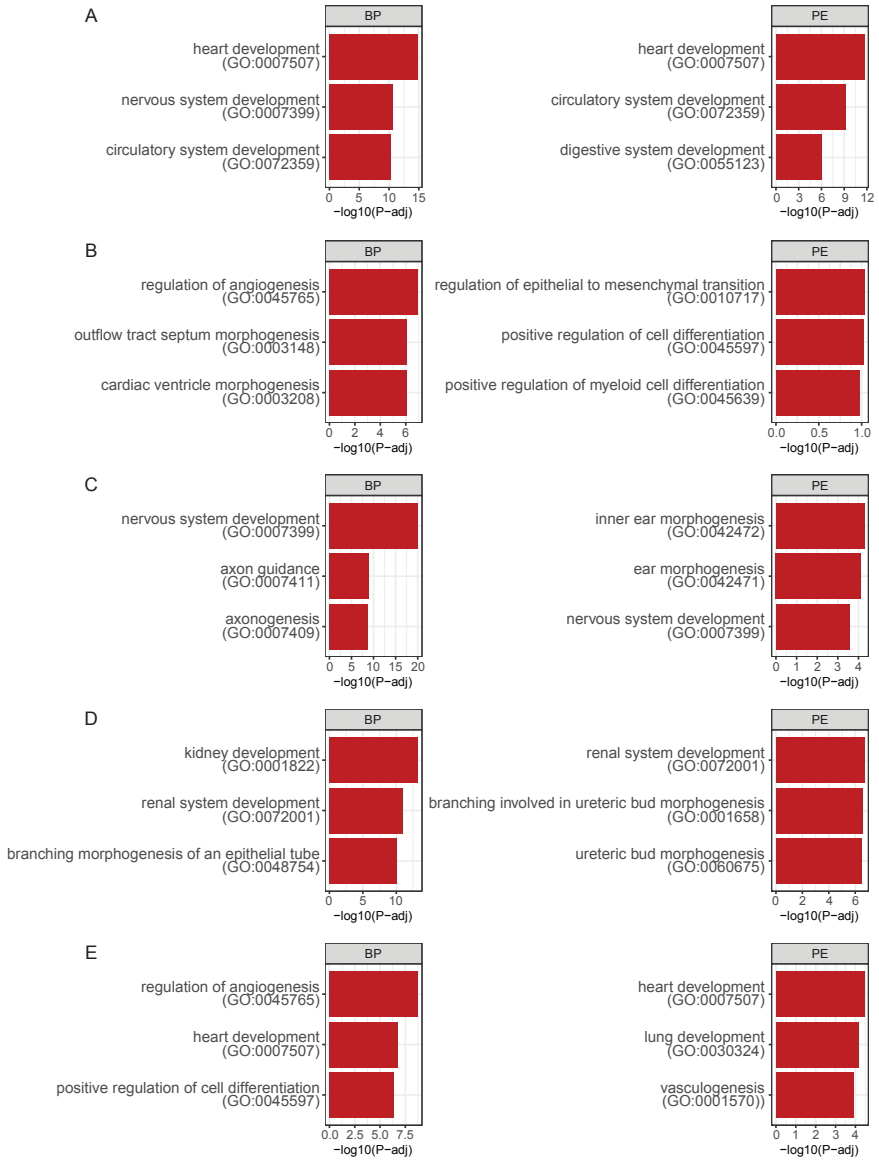

Figure R2.23: Functional analysis of the target genes of the BPs and PEs becoming active during differentiation. (A) Top GO biological process 2018 categories enriched in target genes of activated BPs in Heart10.5 (left) and activated PEs in MES (right). (B) As for A, but in Liver11.5. (C) As for A, but in NeuralTube12.5. (D) As for A, but in Kidney14.5. (E) As for A, but in Lung15.5.

As in differentiation, we observed that an important subset of genes activated their BPs at all embryonic tissues (Figure R2.24A), while this shared subset of genes is relatively less abundant for PEs (Figure R2.24B). Moreover, we confirmed this observation not only at gene level but also when comparing the activation of the regulatory regions (Figure R2.24C-D). These results further indicate that PE mechanisms of activation are more tissue-specific than in the case of BP activation.

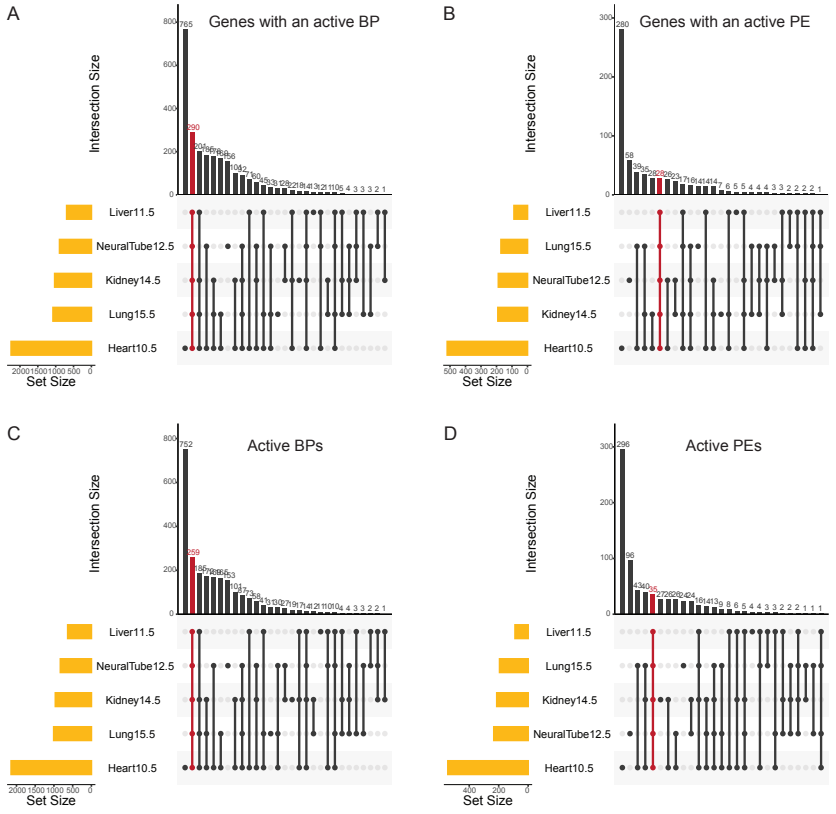

Figure R2.24: Overlap between target genes of activated BPs and PEs at each embryonic tissue. (A) Upset plot comparing overlap between the target genes of the activated BPs at each embryonic tissue. (B) As for A, but in PEs.

2.15 Expression of poised enhancer target genes is tissue-specific

We next compared gene expression between those genes in which only the BP or the BP plus the PE are activated. We found that those genes whose associated PE is activated have higher gene expression than those that only have an activated BP (Figure R2.25).

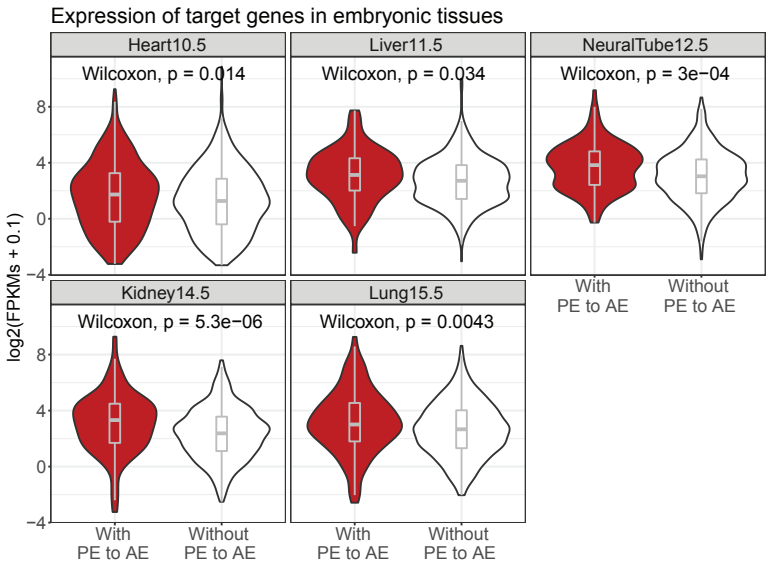

Figure R2.25: Genes activating their PE besides their BP have a higher gene expression. Gene expression in each embryonic tissue at genes activating their BP alone (white), or both, their BP and their PE (red).

Finally, as in differentiation lineages, we wondered whether this higher gene expression when activating the PE is tissue-specific. Thus, we calculated expression of the target genes of one of the embryonic tissues in the rest (Figures R2.26-30).

We observed again that such differences in expression are generally not significant in other tissues. Thus, we concluded that the higher gene expression induced by PE activation is specific of the tissue.

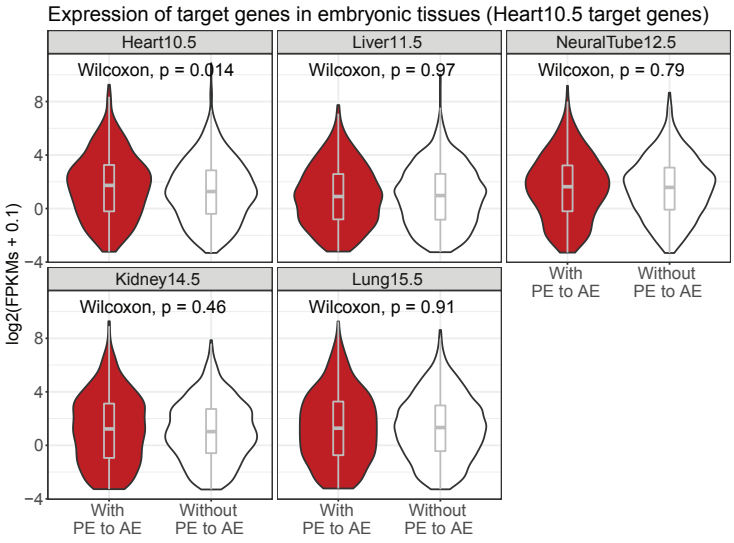

Figure R2.26: The higher gene expression observed at Heart10.5 in its activated PE target genes is not observed in the rest of the embryonic tissues. Gene expression in each embryonic tissue at Heart10.5 target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in Heart10.5.

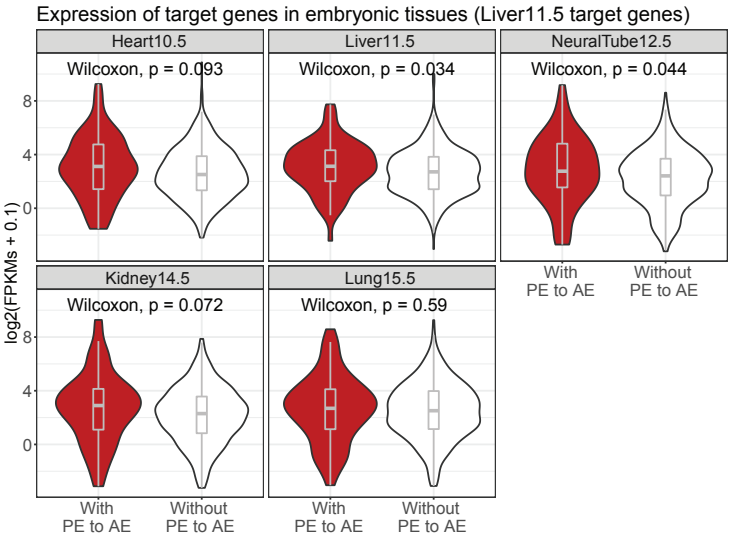

Figure R2.27: The higher gene expression observed at Liver11.5 in its activated PE target genes is not observed in the rest of the embryonic tissues. Gene expression in each embryonic tissue at Liver11.5 target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in Liver11.5.

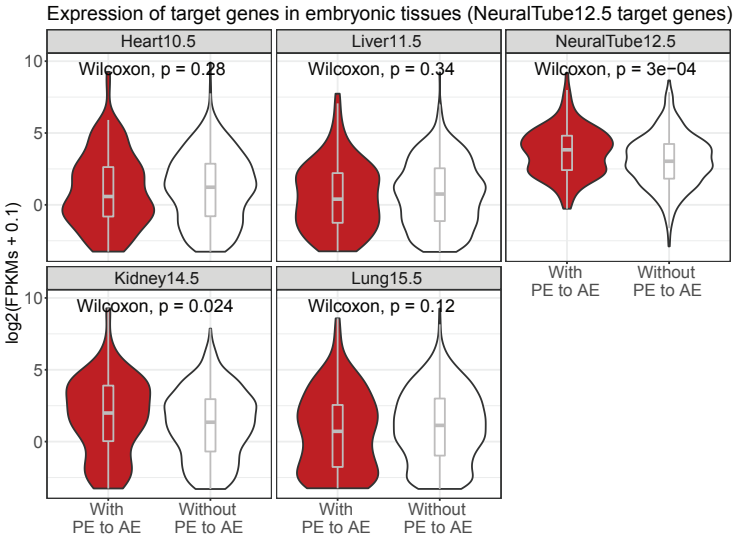

Figure R2.28: The higher gene expression observed at NeuralTube12.5 in its activated PE target genes is not observed in the rest of the embryonic tissues. Gene expression in each embryonic tissue at NeuralTube12.5 target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in NeuralTube12.5.

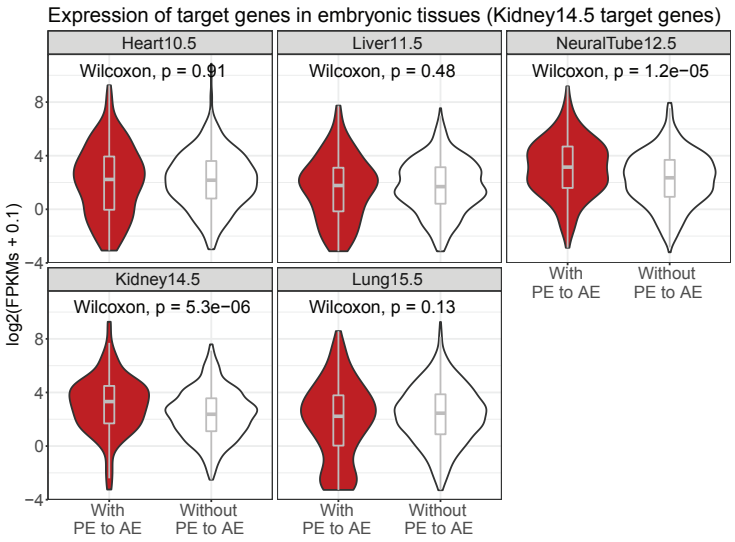

Figure R2.29: The higher gene expression observed at Kidney14.5 in its activated PE target genes is not observed in the rest of the embryonic tissues. Gene expression in each embryonic tissue at Kidney14.5 target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in Kidney14.5.

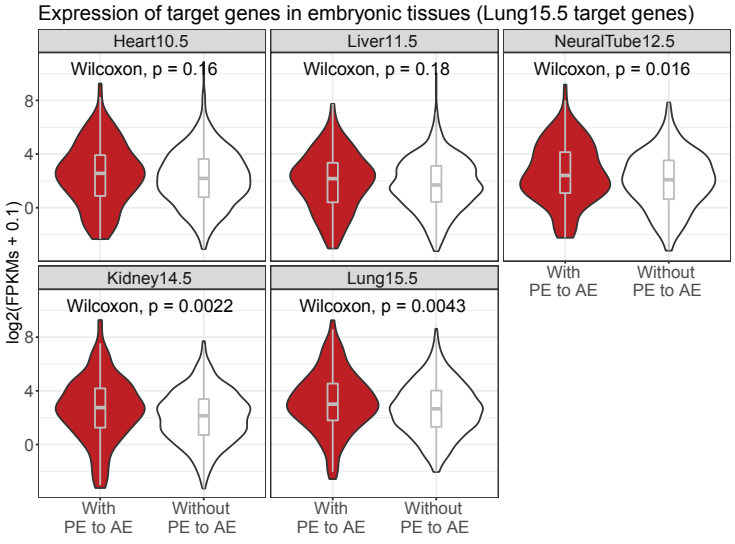

Figure R2.30: The higher gene expression observed at Lung15.5 in its activated PE target genes is not observed in the rest of the embryonic tissues. Gene expression in each embryonic tissue at Lung15.5 target genes. Genes activating their BP alone (white), or both, their BP and their PE (red), in Lung15.5.

2.16 Commonly activated bivalent genes in development share the activation of the same poised enhancer at all embryonic tissues

Similar to our previous results during differentiation, we wondered whether the expression of common genes is driven by the same PE or by different ones. Again, we found that common genes with an activated PE generally share the same enhancers (Figure R2.31).

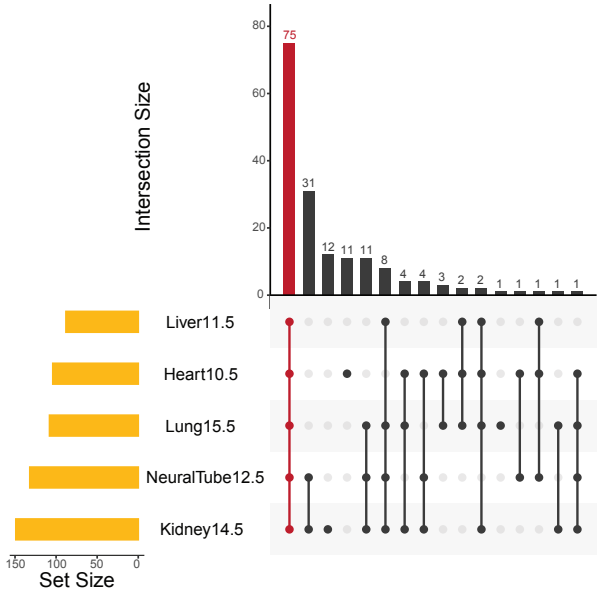

Figure 2.31: Overlap between activated PEs in development associated to common genes. Upset plot comparing the overlap between activated PEs at each embryonic tissue that govern common genes.

CHAPTER 3

The results shown in this chapter correspond to objective number 3 of this thesis.

García-Eguren G.*, **González-Ramírez M.***, Vizán P., Giró O., Vega-Beyhart A., Boswell L., Mora M., Halperin I., Carmona F., Gracia M., Squarcia M., Enseñat J., Vidal O., Di Croce L. and Hanzu F. A. Glucocorticoid-induced fingerprints on visceral adipose tissue transcriptome and epigenome. Submitted to JCI Insight.

3.1 Mouse model of Cushing's syndrome

Glucocorticoids (GCs) play important roles on the metabolism and transcription of many tissues including the adipose tissue through the binding of the GC receptor (GR) [167, 168]. Importantly, the alterations caused by GC overexposure are very critical in the visceral adipose tissue (VAT), as it contains higher GR levels and is considered more metabolically active than subcutaneous adipose tissue [169, 170]. However, the availability of VAT biopsies from Cushing's Syndrome (CS) patients, even more after long-term remission, has difficulted the study of GCs in VAT. Thus, we benefited from a well-established mouse model of reversible hypercortisolism [171, 172] to overcome this limitation (Figure R3.1). Briefly, mice were treated with corticosterone during 5 weeks (CORT_5w group) in their drinking water to induce the systemic effects of active CS. Control mice were treated only with the vehicle (ethanol) in their drinking water also during 5 weeks (VEH_5w group). To mimic long-term remission of CS, treated mice were recovered during 10 more weeks (CORT_15w group) without corticosterone in their drinking water to reduce their corticosterone levels. As a control, ethanol treatment was also stopped during 10 weeks in control mice (VEH_15w group). Mice procedures were performed by Guillermo García-Eguren (Endocrine disorders lab, IDIBAPS, Barcelona).

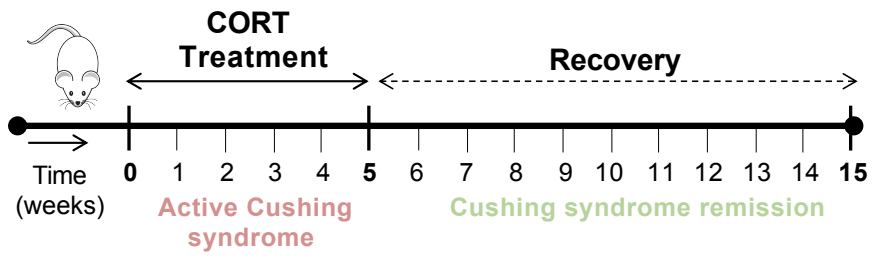

Figure R3.1: Experimental design. Mice treated with corticosterone (CORT) during 5 weeks in their drinking water constitute our animal model of active Cushing syndrome. Treated mice let to recover during 10 weeks without treatment constitute our animal model of Cushing syndrome remission.

In order to study the epigenetic landscape of CS during the active state of the disease and after remission, RNA-seq and ChIP-seq experiments were performed on VAT from all four mice groups. All experiments were performed by Guillermo García-Eguren (Endocrine disorders lab, IDIBAPS, Barcelona) and Pedro Vizán (our lab).

3.2 Correlation between histone modifications and gene expression in mice

We performed ChIP-seq experiments with *Drosophila melanogaster* spike-in for three histone modifications: H3K4me3 and H3K27ac, which are both associated to gene activation, and H3K27me3 which is associated to gene repression. Spike-in normalization consists in the addition of a known amount of chromatin from *Drosophila melanogaster* used to correct for technical biases. To confirm the expected

positive and negative correlations with gene expression, respectively, we classified all genes in each condition into four categories depending on their expression (measured in FPKMs): from 0 to 1, from 1 to 5, from 5 to 50, and more than 50. As expected, we observed that H3K4me3 and H3K27ac ChIP-seq signals were lower in the least expressed category and higher in the most expressed one, while the opposite trend was reported for H3K27me3 ChIP-seq signal (Figure R3.2). These results confirm the validness of our panel of high-throughput experiments in the mouse model of CS. Therefore, we set out to explore the putative epigenetic fingerprint after CS remission.

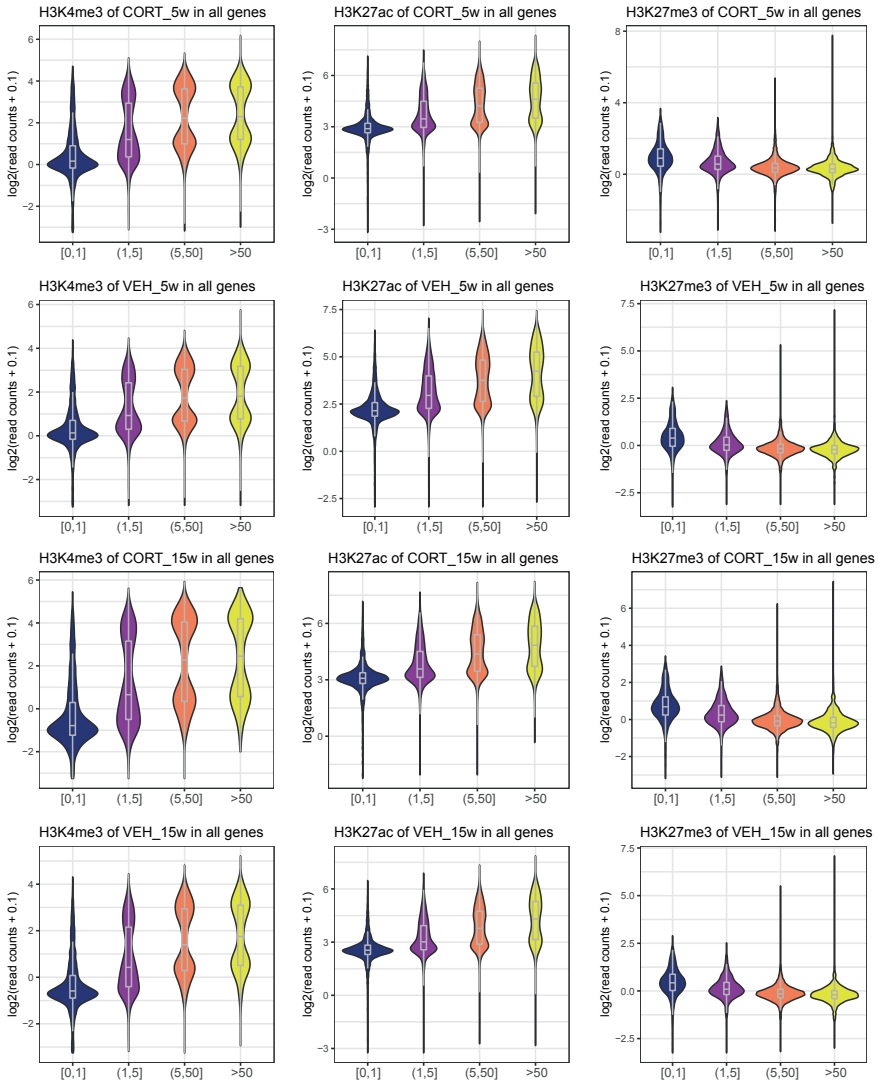

Figure R3.2: Correlation between ChIP-seq and RNA-seq in mice. ChIP-seq signal in genes stratified by expression measured in FPKMs for active CS mice (CORT_5w), control mice of active CS (VEH_15w), CS remission mice (CORT_15w) and control mice of CS remission (VEH_15w). ChIP-seq signal is measured in the region ± 2 Kb around a TSS from RefSeq [149], and calculated as the number of read counts averaged by the length of the region and normalized by the total number of *Drosophila melanogaster* reads, plus a pseudo-count of 0.1. FPKM and ChIP-seq signal values are averaged across all the replicates. For H3K27ac in CORT_15w, only replicate 2 was used.

3.3 Pipeline to obtain consensus peaks for each condition

As we performed two replicates for each ChIP-seq experiment, we were interested in obtaining a list of consensus peaks for each condition. Usually, one uses the overlap between both replicates as the final list of peaks, however, this leads to the loss of true peaks not called by the peak caller in one of the replicates. To circumvent this issue, we decided to design a novel approach. First of all, we constructed a Python script to merge the lists of peaks of both replicates and unify the coordinates. After unifying the coordinates, we performed a differential test with DiffBind [48], between the ChIP-seq replicates and their input, to select those peaks that presented a significant increase of ChIP-seq signal over input (Figure R3.3). We used a $\log_2$(fold change) > 0, *p*-value < 0.05 and false discovery rate (FDR) < 0.1 as threshold. The total number of consensus peaks are in Table R3.1. In the case of H3K27ac ChIP-seq of CORT_15w, we selected as consensus peaks the list of peaks of the second replicate (29,342), as the first replicate has a high background and only reported 1,811 peaks.

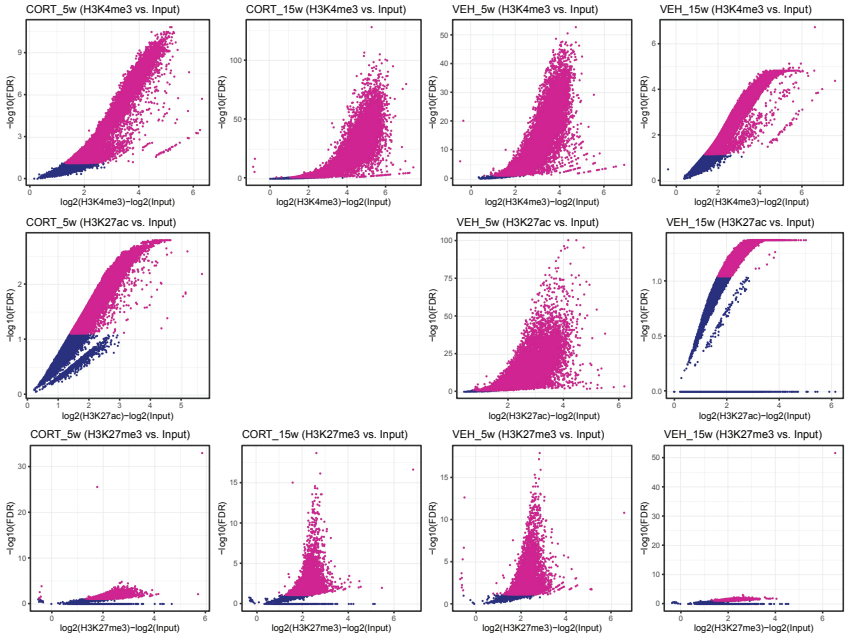

Figure R3.3: Significant peaks over input in mice. Volcano plots, each point represents a ChIP-seq peak; significant peaks are colored in pink, whereas non-significant ones are in blue. FDR, false discovery rate.

Table R3.1: Consensus peaks in mice.

Replicate	Number of peaks	Consensus peaks
H3K4me3_CORT_1_5w	29,375	22,351
H3K4me3_CORT_2_5w	20,985	
H3K27ac_CORT_1_5w	41,591	26,782
H3K27ac_CORT_2_5w	22,166	
H3K27me3_CORT_1_5w	11,091	6,452
H3K27me3_CORT_2_5w	4,431	
H3K4me3_CORT_1_15w	38,729	29,048
H3K4me3_CORT_2_15w	28,813	
H3K27ac_CORT_1_15w	1,811	-
H3K27ac_CORT_2_15w	29,342	
H3K27me3_CORT_1_15w	4,600	5,557
H3K27me3_CORT_2_15w	9,109	
H3K4me3_VEH_1_5w	19,008	18,246
H3K4me3_ VEH _2_5w	18,125	
H3K27ac_ VEH _1_5w	40,228	38,617
H3K27ac_ VEH _2_5w	36,285	
H3K27me3_ VEH _1_5w	8,632	6,476
H3K27me3_ VEH _2_5w	5,915	
H3K4me3_VEH_1_15w	22,997	24,129
H3K4me3_ VEH _2_15w	29,123	
H3K27ac_ VEH _1_15w	20,645	15,356
H3K27ac_ VEH _2_15w	52,109	
H3K27me3_ VEH _1_15w	2,256	4,066
H3K27me3_ VEH _2_15w	8,285	

For each histone modification and condition, the total number of peaks per replicate and the total number of consensus peaks between replicates.

With this approach we were able to overcome some of the limitations of working with ChIP-seq replicates. Indeed, we were able to retain interesting peaks that the peak caller did not call in one of the replicates (Figure R3.4A). Moreover, we discarded peaks that are only present in one of the replicates

(Figure R3.4B), meaning that the peaks are not specific of the treatment but to the individuals from which the chromatin was pulled. Finally, we discarded a significant number of false positives (Figure R3.4C).

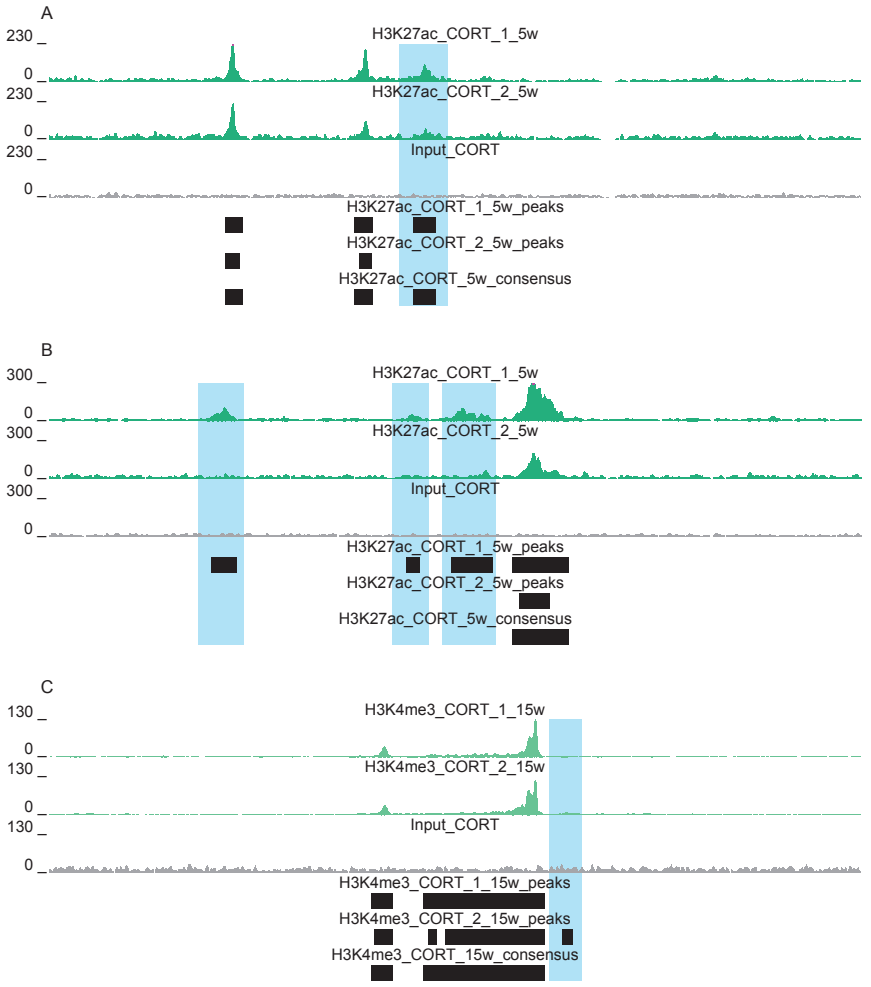

Figure R3.4: Examples of the performance of the pipeline to obtain consensus peaks. (A) Region of mm10 with a peak not called in one of the replicates (chr2-25,445,158-25,527,467). (B) Region of mm10 with specific peaks of one replicate (chr7-152,263,211-152,320,300). (C) Region of mm10 with a peak that is not real but has been called by the peak caller (chr7-149,523,628-149,606,953).

3.4 Chronic hypercortisolism increases histone modification signal genome-wide in mice

After differential peak analysis, CORT_5w mice showed a genome-wide increase in the levels of all three histone modifications (Figure R3.5A). Thus, up to 57% of H3K4me3 peaks had a significant increase in ChIP-seq signal (13,336 peaks out of 23,236 merged peaks of H3K4me3 in CORT_5w and VEH_5w), whereas only four peaks had a significant decrease of signal. Moreover, 35% of H3K27ac peaks had a significant increase in ChIP-seq signal (13,964 peaks out of 39,581 merged peaks of H3K27ac in CORT_5w and VEH_5w), whereas only 177 peaks had a significant decrease of signal. Finally, 28% of H3K27me3 peaks had a significant increase in ChIP-seq signal (2,168 peaks out of 7,792 merged peaks of H3K27me3 in CORT_5w and VEH_5w), whereas none of the peaks had a significant decrease.

Interestingly, CORT_15w mice showed the same tendency (Figure R3.5B). Thus, 51% of H3K4me3 peaks had a significant increase in ChIP-seq signal (15,984 peaks out of 31,037 merged peaks of H3K4me3 in CORT_15w and VEH_15w), whereas only one peak had a significant decrease of signal. Furthermore, 34% of H3K27ac peaks had a significant increase in ChIP-seq signal (9,602 peaks out of 28,322 merged peaks of H3K27ac in CORT_15w and VEH_15w), whereas only five peaks had a significant decrease of signal. Finally, 31% of H3K27me3 peaks had a significant

increase in ChIP-seq signal (1,864 peaks out of 5,992 merged peaks of H3K27me3 in CORT_15w and VEH_15w), whereas none of the peaks had a significant decrease.

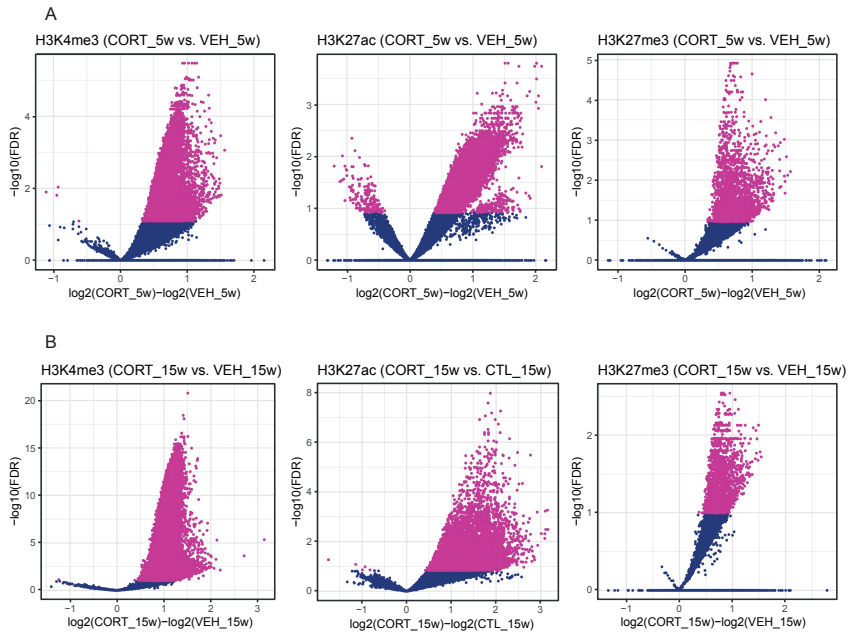

Figure R3.5: Effect on histone modifications of GC overexposure in mice. Volcano plots, each point represents a ChIP-seq peak; significant peaks are colored in pink, whereas non-significant ones are in blue. (A) Changes in H3K4me3, H3K27ac and H3K27me3 in GC-treated mice (CORT_5w) compared to control mice (VEH_5w). (B) Same as in A, but for recovered mice (CORT_15w) compared to control mice (VEH_15w).

Visual inspection of ChIP-seq profiles also showed a signal increase of GC-treated mice compared to their controls at both time points (Figure R3.6). Therefore, our observations suggest that chronic hypercortisolism induces changes in histone marks.

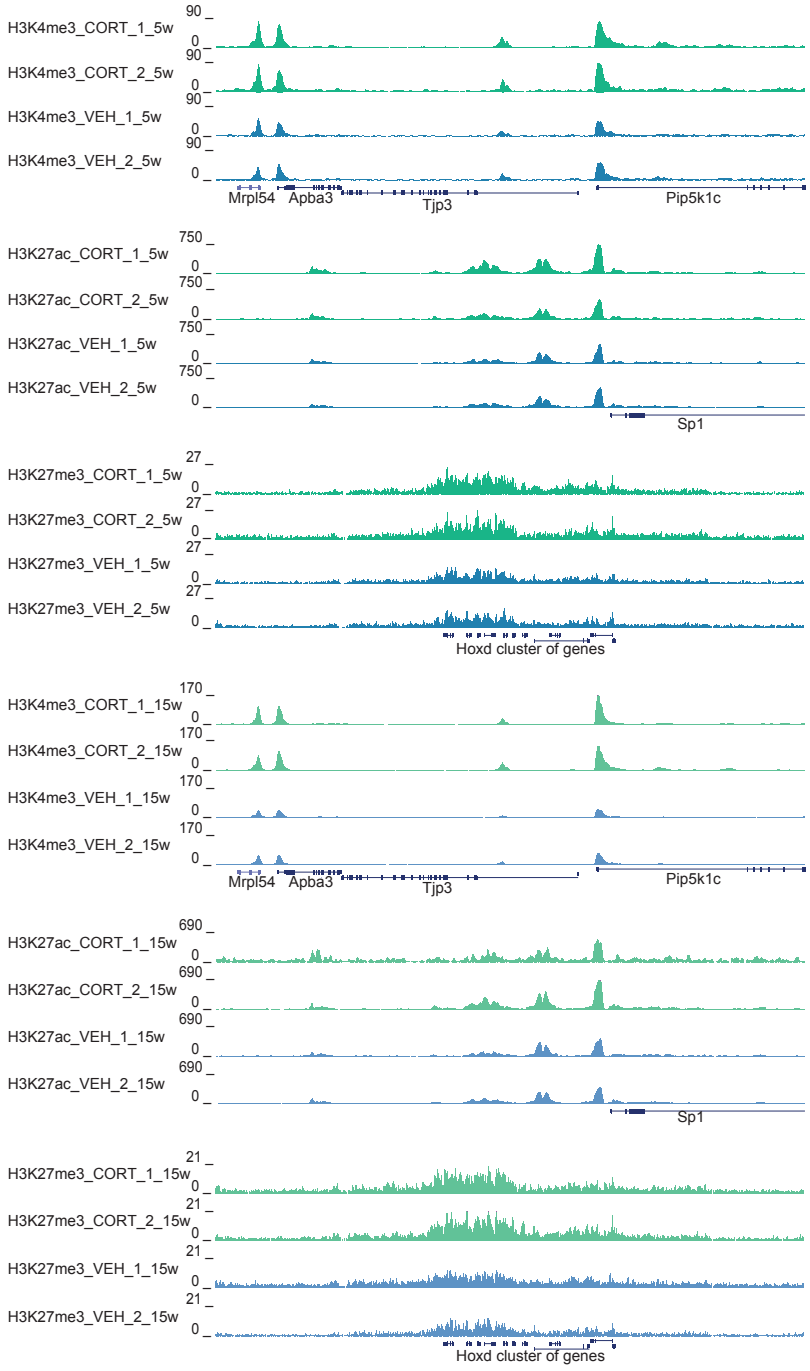

Figure R3.6: Examples of regions increasing histone modifications signal after GC-treatment. ChIP-seq profiles of H3K4me3, H3K27ac and H3K27me3 in GC-treated mice (CORT_5w), after recovery (CORT_15w) and their controls (VEH_5w and VEH_15w, respectively).

However, to confirm that spike-in normalization was working properly and our results were not a consequence of technical issues, we performed the same differential analysis comparing mice controls. As they were not treated, they should have the same epigenetic landscape. Accordingly, there were no significant differences between VEH_5w and VEH_15w mice for any of the histone modifications (Figure R3.7).

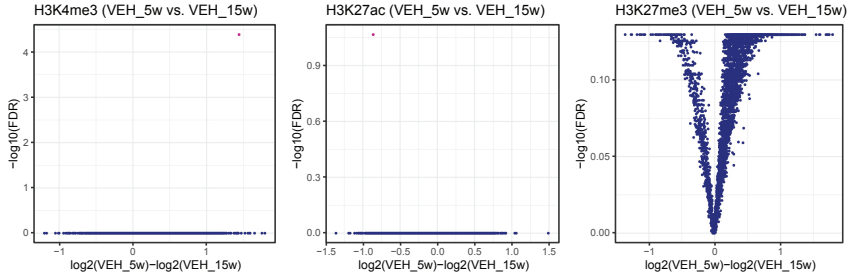

Figure R3.7: There are not significant differences in histone modifications signal between controls. Volcano plots, each point represents a ChIP-seq peak; significant peaks are colored in pink, whereas non-significant ones are in blue.

3.5 The histone modification signal increase in GC-treated mice occurs both at promoters and at putative enhancers

As mentioned in the previous chapters, H3K27ac is known to be located in active enhancers [65]. On the contrary, H3K27me3 is found in poised enhancers [125]. In addition, although more prevalent in active promoters, H3K4me3 has

also been reported in active enhancers [36, 129, 150, 151]. Accordingly, genome distribution of the differential peaks of all three histone modifications showed that the signal increase occurred not only in promoters but also in intergenic and intragenic regions (Figure R3.8). Therefore, these intergenic and intragenic peaks could be marking the location of putative enhancers.

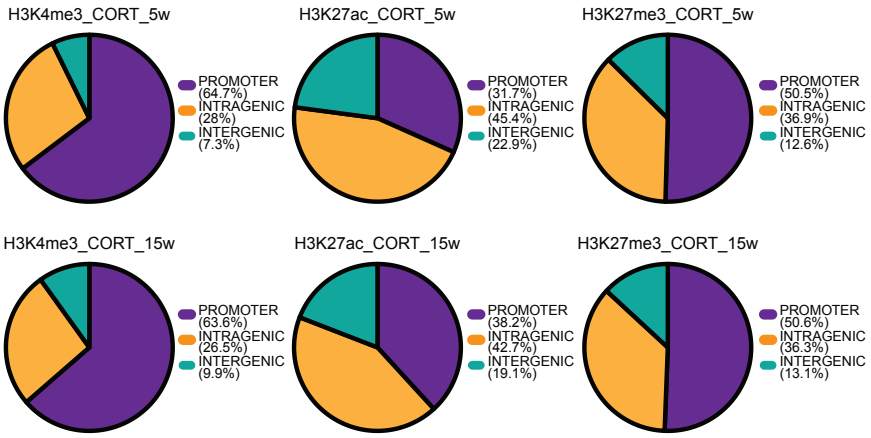

Figure R3.8: Increase of histone modification signals occur not only at promoters but also at intergenic and intragenic regions. Genome distribution of H3K4me3, H3K27ac and H3K27me3 differential peaks in CORT_5w vs. VEH_5w (top) and CORT_15w vs. VEH_15w (bottom). Promoter is considered as the region ± 2 Kb around a TSS from RefSeq catalogue [149].

3.6 The histone modification signal increase occurs at the same regions in active CS and after long-term remission

To confirm that the increase occurred in the same regions in both conditions, we calculated the overlap between the significantly up peaks of each histone modification (Figure R3.9A). Accordingly, all three overlaps were significant with a *p*-value < 2.2e-16 (Fisher's Exact Test). Moreover, at gene level (Figure R3.9B), the overlap was also significant with a *p*-value < 2.2e-16 (Fisher's Exact Test), which confirms that the increase affects the same target genes. Therefore, there is a persistent epigenetic fingerprint after resolution of hypercortisolism.

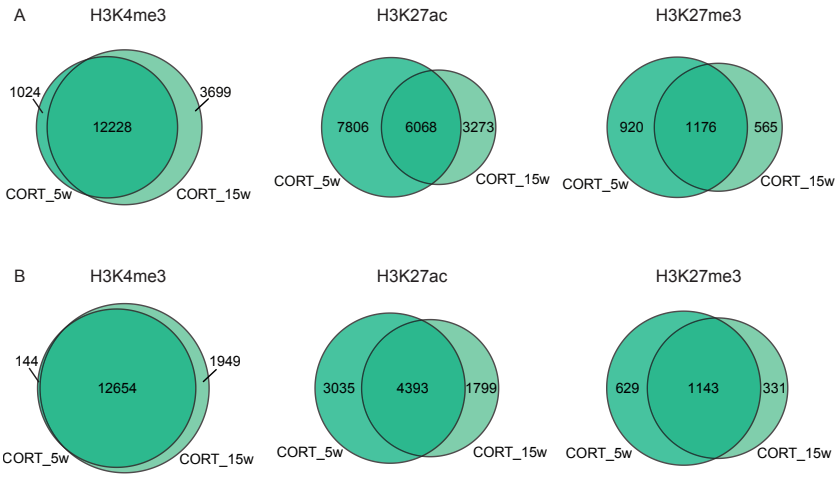

Figure R3.9: Overlaps of significantly up peaks and its target genes between CORT_5w and CORT_15w. (A) Venn diagrams of the overlaps between CORT_5w and CORT_15w significantly up peaks of H3K4me3, H3K27ac and H3K27me3. (B) Same as in A, but for target genes of the significantly up peaks.

We next performed GO term enrichment analysis on the common target genes of the peaks that significantly increased the signal of any of the three histone modifications (Figure R3.10). GO terms of common genes overlapping peaks increasing in H3K4me3 were related to house-keeping categories, which is consistent with the genome-wide increase of the mark that we observe. Among the most significant GO terms in common genes overlapping peaks increasing in H3K27me3, we found categories related to development, as expected for the target genes of this mark. Therefore, in agreement with a generalized increase of H3K27me3 at its target sites. Interestingly, we found that neutrophil activation involved in immune response category was enriched in common target genes overlapping peaks increasing in H3K27ac, which agrees with previous findings that show that GC overexposure causes alterations in inflammatory and immune response [173-175].

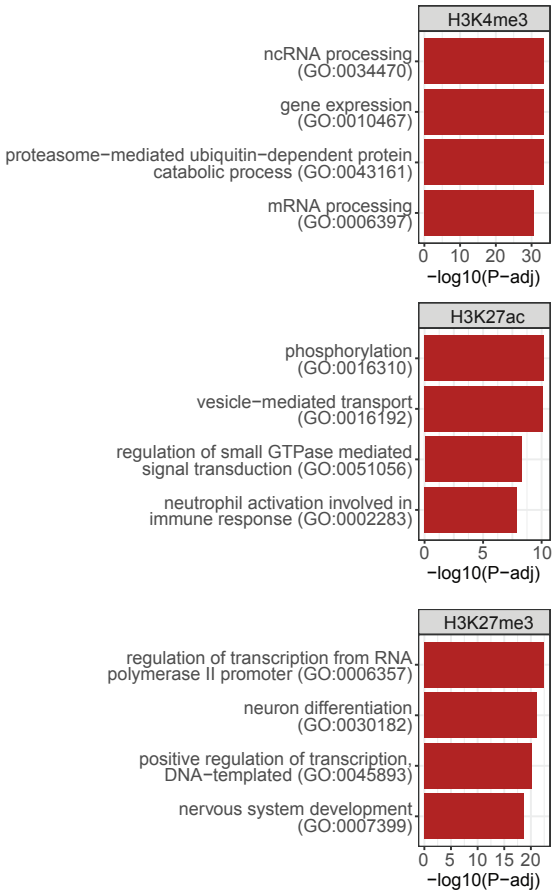

Figure R3.10: Functional analysis of genes with an increased histone modification signal in CS and after remission. Gene ontology biological process (2018 categories) enrichment analysis of the common genes increased in H3K4me3 signal (top), H3K27ac signal (middle) and H3K27me3 signal (bottom) in CORT_5w versus VEH_5w and CORT_15w versus VEH_15w.

3.7 Persistent changes of chronic hypercortisolism in gene expression correlate with persistent changes in histone modifications in mice

Differential analysis of gene expression revealed 4,454 up-regulated genes and 4,154 down-regulated genes in active CS compared to controls, and 1,402 up-regulated genes and 1,679 down-regulated genes after CS remission compared to controls. Among them, 830 up-regulated genes and 961 down-regulated genes were common at both time points. GO term enrichment analysis showed several categories altered by GC treatment (Figure R3.11). Remarkably, we found that neutrophil activation involved in immune response was the top category in the common up-regulated genes, in line with our previous observation for genes increasing H3K27ac at both time points. Thus, there is a persistent transcriptomic signature caused by GC overexposure.

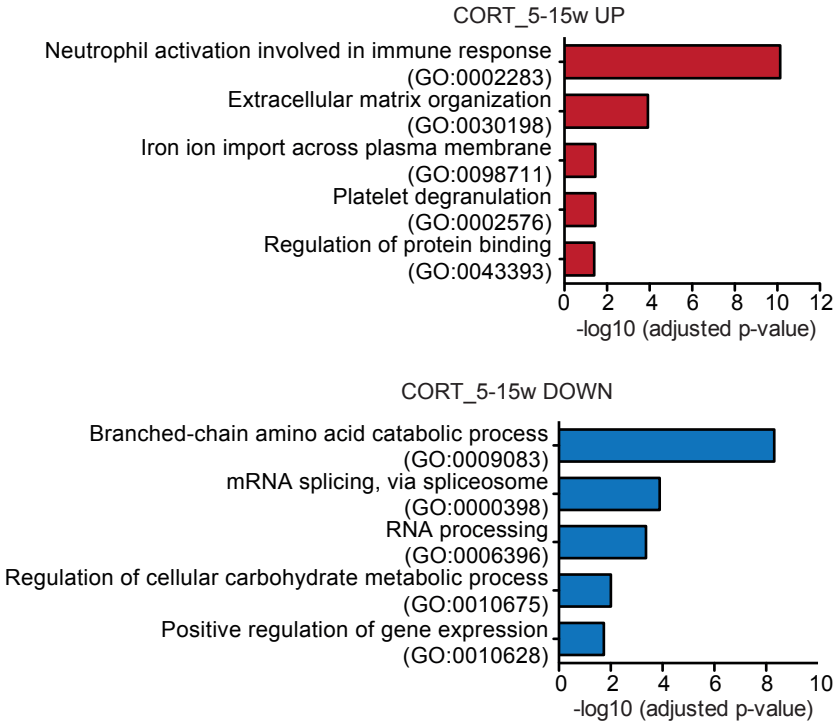

Figure R3.11: Functional analysis of common differential genes in active CS and after remission. Gene ontology biological process (2018 categories) enrichment analysis of the common up-regulated genes (top) and down-regulated genes (bottom) in CORT_5w versus VEH_5w and CORT_15w versus VEH_15w.

Next, we wondered whether there were changes at epigenetic level which could explain the persistent transcriptomic effects of GC overexposure that we observed. Therefore, even though we showed that there is a generalized increase in ChIP-seq signal for all three histone modifications in GC-treated mice compared to controls, we wondered whether these increases were significantly different between the common up-regulated genes and the down-regulated genes for any of the marks. Accordingly, the fold changes of the histone modifications

associated with activation were significantly higher in the up-regulated genes than in the down-regulated genes (Figure R3.12), consistent with the differences in gene expression. On the contrary, the difference was not significant in the case of H3K27me3. These results indicate that persistent changes in expression after GC overexposure can be explained by persistent changes in H3K4me3 and H3K27ac.

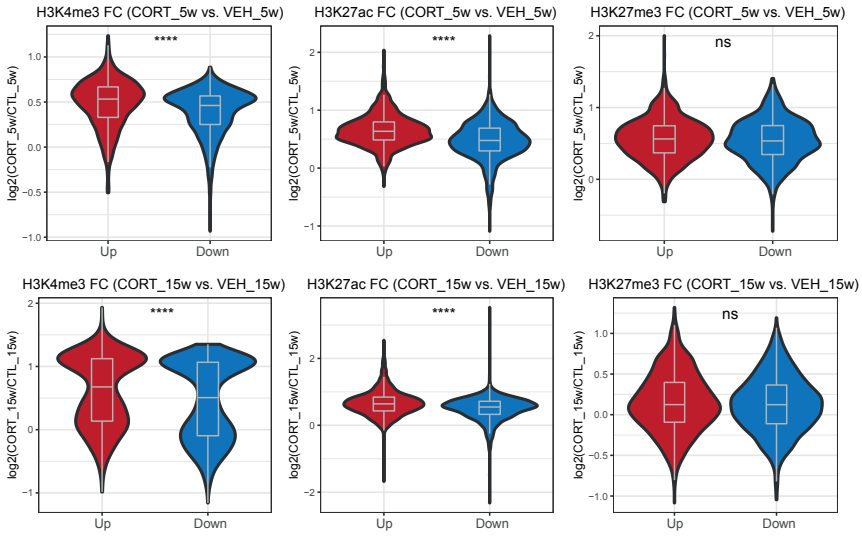

Figure R3.12: Differences in histone modifications between common up-regulated and down-regulated genes in both time points of GC treatment versus controls. Fold change of ChIP-seq signal between active CS versus control mice (top) and CS remission versus control mice (bottom) in common up-regulated genes and common down-regulated genes in CORT_5w versus VEH_5w and CORT_15w versus VEH_15w. ChIP-seq signal is averaged across the replicates, measured in the region ± 2 Kb around a TSS from RefSeq [149], and calculated as the number of read counts averaged by the length of the region and normalized by the total number of drosophila reads, plus a pseudo-count of 0.1. Significance was assessed using a Wilcoxon test (****p-value < 0.0001, ***p-value < 0.001, **p-value < 0.01, *p-value < 0.05, ·p-value < 0.1, ns non-significant). For H3K27ac in CORT_15w, only replicate 2 was used.

3.8 Correlation between histone modifications and gene expression in human patients

Next, we explored whether our observations in the mouse model of CS were also true for human patients. Therefore, we performed RNA-seq and ChIP-seq of H3K4me3, H3K27ac and H3K27me3 for human CS patients and pair-matched controls. In here, ChIP-seq experiments included *Drosophila melanogaster* spike-in, as in mice. All experiments were performed by Guillermo García-Eguren (Endocrine disorders lab, IDIBAPS, Barcelona) and Pedro Vizán (our lab).

As in mice, we observed that all three histone modifications had the expected correlation with gene expression in each human patient. Examples of this correlation can be found in Figure R3.13. These results confirm the validness of our set of data from human patients.

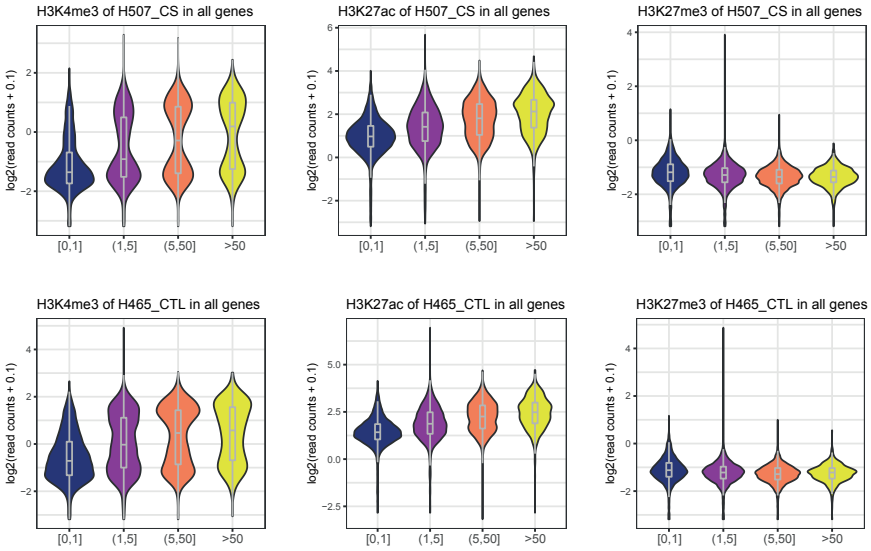

Figure R3.13: Correlation between ChIP-seq and RNA-seq in human patients. ChIP-seq signal in genes stratified by expression measured in FPKMs for active CS patient H507 (top) and control patient H465 (bottom). ChIP-seq signal is measured in the region ± 2 Kb around a TSS from RefSeq [149], and calculated as the number of read counts averaged by the length of the region and normalized by the total number of *Drosophila melanogaster* reads, plus a pseudo-count of 0.1.

3.9 Chronic hypercortisolism alters the epigenetic landscape in CS patients, with an opposite pattern than in mice

In here, we again obtained a list of consensus peaks for each histone modification and group (CS or control) by following our differential analysis between each ChIP-seq of the same group and their input using DiffBind (Figure R3.14). As for human patients we have only four individuals per group, we merged and unified coordinates of peaks with DiffBind, which at the

same time selects only peaks that are at least present in two of the samples. We used a $\log_2$(fold change) > 0, *p*-value < 0.05 and false discovery rate (FDR) < 0.1 as threshold. The total numbers of consensus peaks are reported in Table R3.2.

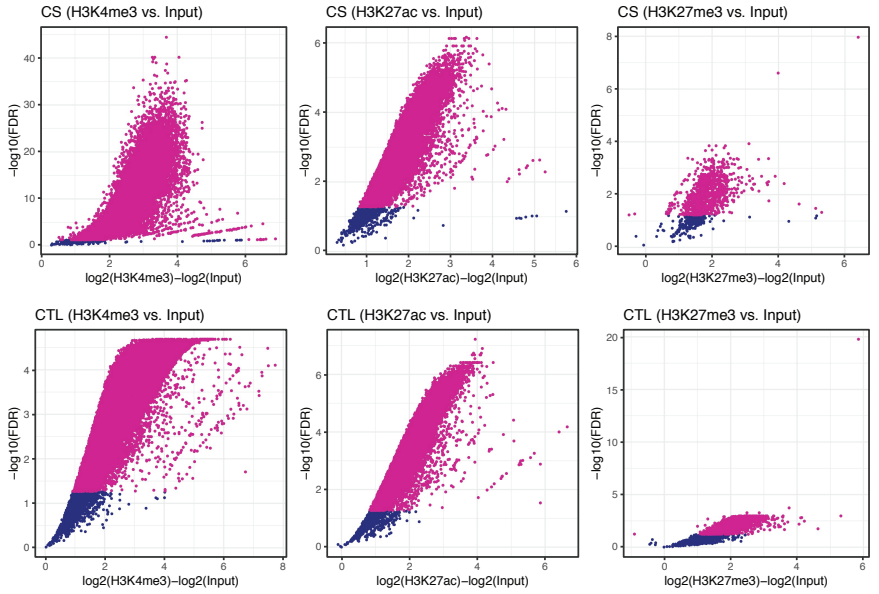

Figure R3.14: Significant peaks over input in human patients. Volcano plots, each point represents a ChIP-seq peak; significant peaks are colored in pink, whereas non-significant ones are in blue. FDR, false discovery rate.

Table R3.2: Consensus peaks in human.

Replicate	Number of peaks	Consensus peaks
H3K4me3_CS_H507	19,942	20,403
H3K4me3_CS_VC02	23,128	
H3K4me3_CS_VC04	35,140	
H3K4me3_CS_VC06	18,649	
H3K27ac_CS_H507	30,722	22,964
H3K27ac_CS_VC02	20,492	
H3K27ac_CS_VC04	19,690	
H3K27ac_CS_VC06	32,927	
H3K27me3_CS_H507	4,227	962
H3K27me3_CS_VC02	2,178	
H3K27me3_CS_VC04	1,499	
H3K27me3_CS_VC06	4,006	
H3K4me3_CTL_H425	19,413	25,782
H3K4me3_CTL_H465	36,098	
H3K4me3_CTL_HCTR10	33,426	
H3K4me3_CTL_HCTR13	29,550	
H3K27ac_CTL_H425	35,627	32,410
H3K27ac_CTL_H465	25,060	
H3K27ac_CTL_HCTR10	40,106	
H3K27ac_CTL_HCTR13	49,022	
H3K27me3_CTL_H425	2,006	3,897
H3K27me3_CTL_H465	3,400	
H3K27me3_CTL_HCTR10	10,303	
H3K27me3_CTL_HCTR13	2,347	

For each histone modification and condition, the total number of peaks per replicate and the total number of consensus peaks between replicates.

Surprisingly, after differential peak analysis between CS patients and controls, we observed the opposite pattern than in mice. Instead of a genome-wide increase of ChIP-seq signal, we observed a genome-wide decrease in CS patients compared to controls (Figure R3.15). Thus, 51% of H3K4me3

peaks had a significant decrease in ChIP-seq signal (13,503 peaks out of 26,467 merged peaks of H3K4me3 in CS and CTL), whereas only one peak had a significant increase of signal. Moreover, 80% of H3K27ac peaks had a significant decrease in ChIP-seq signal (26,671 peaks out of 33,526 merged peaks of H3K27ac in CS and CTL), whereas only one peak had a significant increase of signal. Finally, 83% of H3K27me3 peaks had a significant decrease in ChIP-seq signal (3,291 peaks out of 3,985 merged peaks of H3K27me3 in CS and CTL), whereas only one peak had a significant increase.

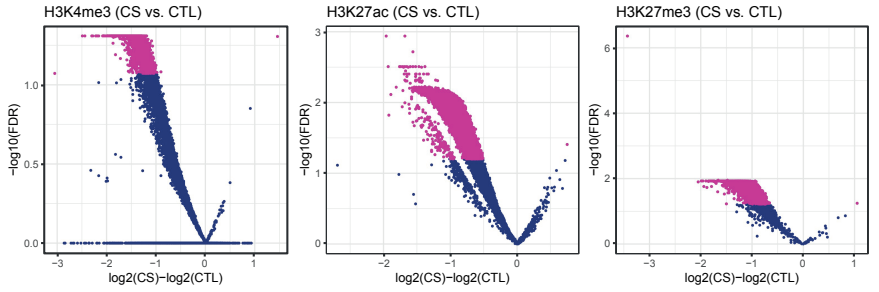

Figure R3.15 Effect on histone modifications of CS in human patients. Volcano plots, each point represents a ChIP-seq peak; significant peaks are colored in pink, whereas non-significant ones are in blue. Changes in H3K4me3, H3K27ac and H3K27me3 in CS patients compared to control patients.

Visual inspection of ChIP-seq profiles also showed a signal decrease (Figure R3.16). Therefore, our observations indicate that CS induces changes in histone modifications in humans with an opposite pattern than in mice.

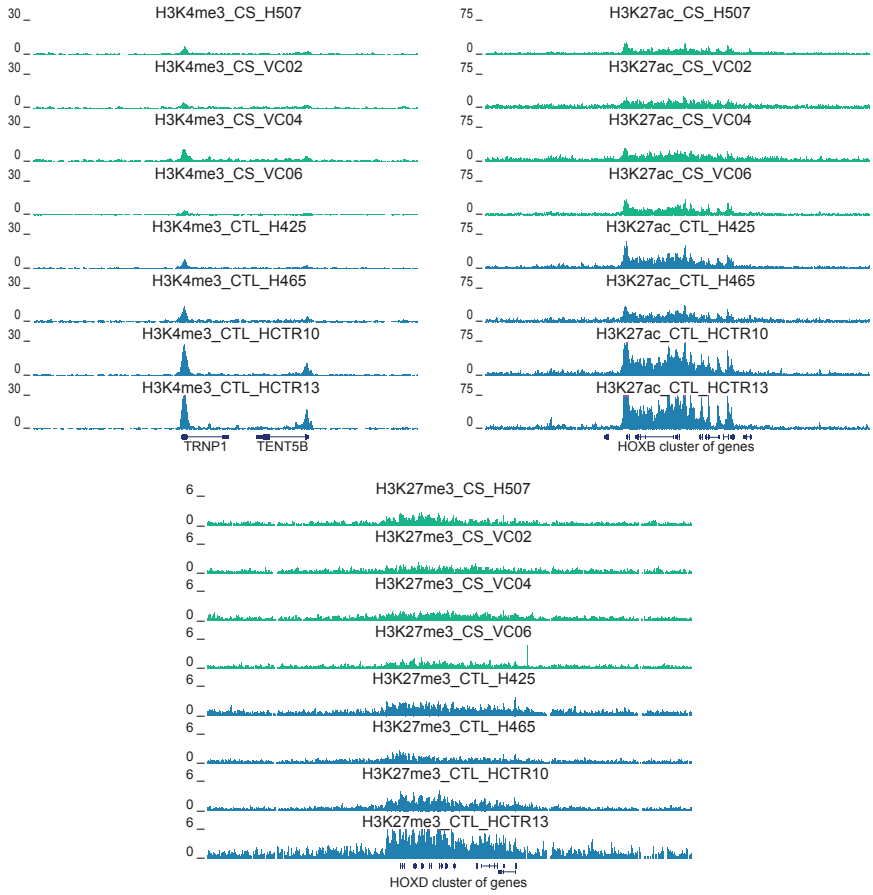

Figure R3.16: Examples of regions decreasing histone modifications signal in CS patients compared to controls. ChIP-seq profiles of H3K4me3, H3K27ac and H3K27me3 in CS patients and their pair-matched controls.

3.10 The histone modification signal decrease in human patients occurs both at promoters and at putative enhancers

We wondered whether this generalized decrease in histone modifications marking was occurring at promoters and also at putative enhancers. Genome distribution of the peaks showing a significant decrease in their signal strength in CS patients compared to controls showed that an important fraction of them was located outside promoters, within intergenic and intragenic regions (Figure R3.17). Therefore, these intergenic and intragenic peaks could be delimiting putative enhancers.

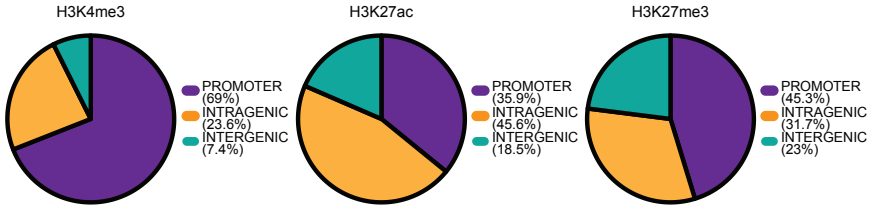

Figure R3.17: Decrease of histone modification signal in human CS patients occurs not only at promoters but also at intergenic and intragenic regions. Genome distribution of H3K4me3, H3K27ac and H3K27me3 differential peaks in human CS patients. Promoter is considered as the region ± 2 Kb around a TSS from RefSeq catalogue [149].

We next performed GO term enrichment analysis on the target genes of the peaks in which the signal of any of the three histone modifications significantly decreased (Figure R3.18). A total of 14,503 genes decreased in H3K4me3 signal, 13,861 genes decreased in H3K27ac signal and 1,998 genes

decreased in H3K27me3 signal. GO terms of genes overlapping peaks decreasing in H3K4me3 or H3K27ac were related to house-keeping categories, which is consistent with the observed genome-wide decrease of both marks. Among the most significant GO terms in genes overlapping a peak decreasing in H3K27me3, we found categories related to development, as expected for the target genes of this mark. Again, this result is consistent with a generalized decrease of H3K27me3 at its target sites.

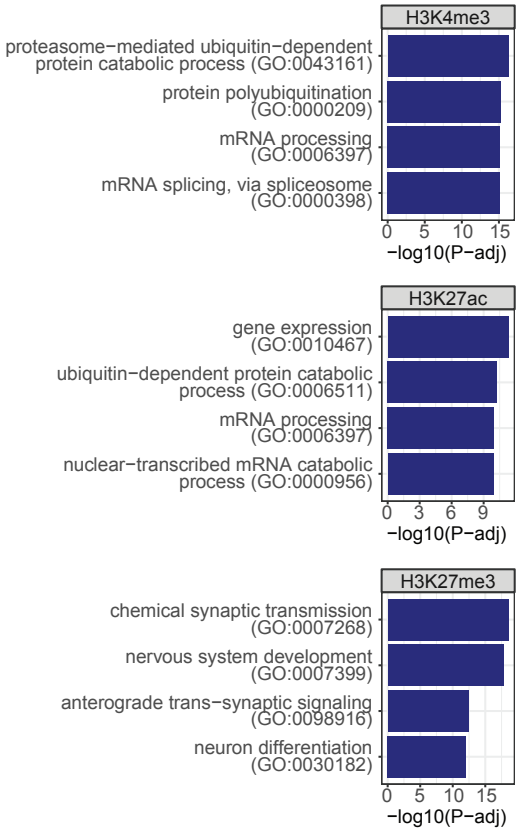

Figure R3.18: Functional analysis of genes with decreased histone modification signal in CS. Gene ontology biological process (2018 categories) enrichment analysis of the genes decreased in H3K4me3 signal (top), H3K27ac signal (middle) and H3K27me3 signal (bottom) in CS patients compared to controls.

3.11 Changes in gene expression correlate with changes in histone modifications in CS patients

We performed differential analysis of gene expression and identified 87 up-regulated genes and 97 down-regulated genes in CS patients compared to human controls. GO enrichment analysis (Figure R3.19) showed that previously reported CS-related categories were enriched [176, 177], such as glucocorticoid biosynthetic process in the up-regulated genes and protein kinase B signaling in the down-regulated genes.

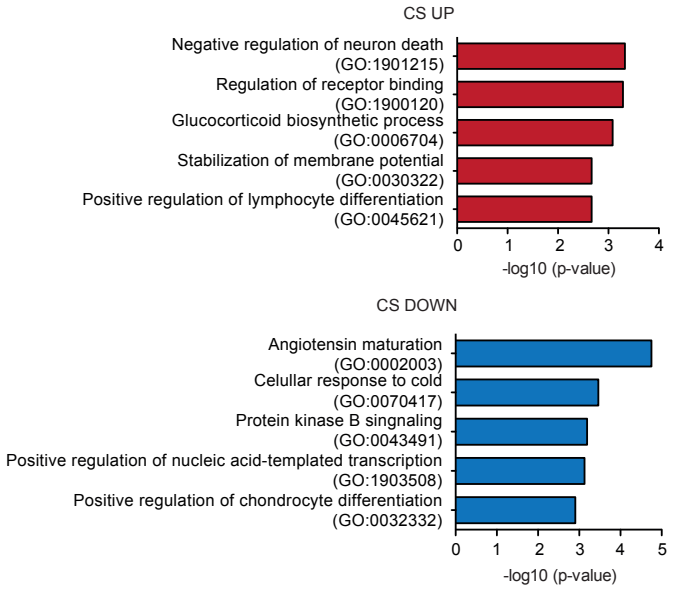

Figure R3.19: Functional analysis of differential genes in CS patients. Gene ontology biological process (2018 categories) enrichment analysis of the up-regulated genes (top) and down-regulated genes (bottom) in CS patients versus control patients.

Opposite to mice, human CS patients showed a generalized decrease in ChIP-seq signal for all three histone modifications. As observed in mice, we hypothesized that the decrease in activation-related histone marks was significantly different between the up-regulated and down-regulated genes in CS patients compared to controls. Certainly, we observed that the fold changes of the histone modifications associated with activation were significantly smaller in the up-regulated genes than in the down-regulated genes, i.e. the decrease in signal is more pronounced in the down-regulated ones (Figure R3.20). On the contrary, the difference was not significant in the case of H3K27me3. As in mice, these observations show that differences in expression can be explained by changes in H3K4me3 and H3K27ac.

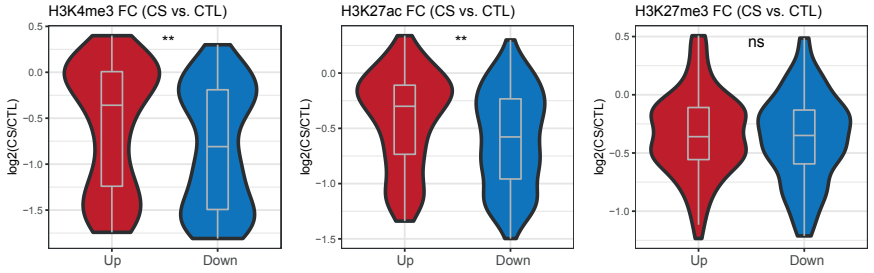

Figure R3.20: Differences in histone modifications between up-regulated and down-regulated genes in CS patients versus controls. Fold change of ChIP-seq signal between active CS versus control in up-regulated genes and down-regulated genes in CS versus controls. ChIP-seq signal is averaged across the individuals, measured in the region ± 2 Kb around a TSS from RefSeq [149], and calculated as the number of read counts averaged by the length of the region and normalized by the total number of drosophila reads, plus a pseudo-count of 0.1. Significance was assessed using a Wilcoxon test (****p-value < 0.0001, ***p-value < 0.001, **p-value < 0.01, *p-value < 0.05, ·p-value < 0.1, ns non-significant).

3.12 Transcriptomic signatures driven by hypercortisolism in human and mice

In order to validate the persistent transcriptomic alterations induced by hypercortisolism in mice also in human, we performed Gene Set Enrichment Analysis (GSEA) on RNA-seq data from human patients using the common orthologous up-regulated genes during active and after CS remission in mice as gene set list (Figure R3.21A). Indeed, the mice signature had also a positive trend in human patients compared to controls. Next, we performed a GO term enrichment analysis on the genes that contributed to this trend (Figure R3.21B). Accordingly, the top categories were related to immune response and to extracellular matrix organization. Thus, these results suggest that, as in the mouse model, immune response and extracellular organization could also be potentially altered in VAT from CS patients after remission.

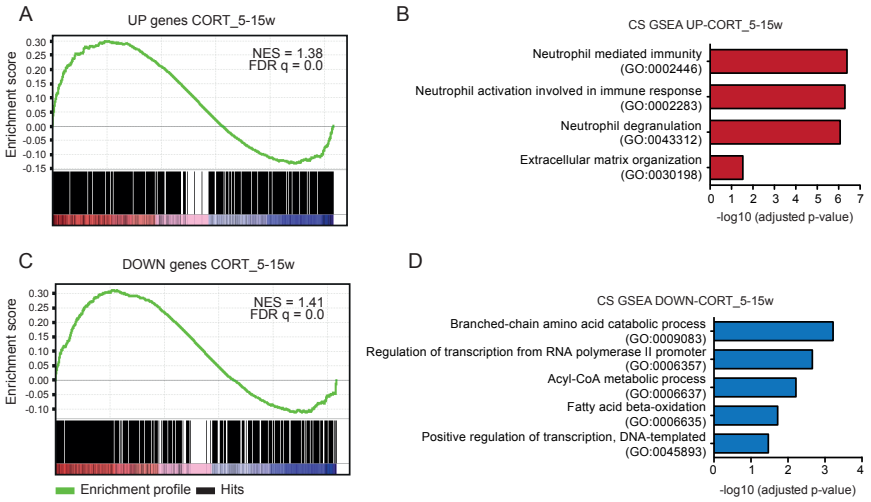

Figure R3.21: Common alterations after GC overexposure in mice and humans. (A) GSEA between active CS humans and their controls using as gene set list the common up-regulated genes in CORT_5w versus VEH_5w and CORT_15w versus VEH_15w. (B) GO term enrichment analysis of the genes that follow a significant tendency in A. (C) GSEA between active CS humans and their controls using as gene set list the common down-regulated genes in CORT_5w versus VEH_5w and CORT_15w versus VEH_15w. (D) GO term enrichment analysis of the genes that follow a significant tendency in C.

Surprisingly, most of the common down-regulated genes during active and after CS remission in mice follow an opposite tendency in the transcriptome of CS patients (Figure R3.21C). Thus, we speculated that this could also be related to the opposite tendency in histone marking. GO enrichment analysis of genes contributing to the positive tendency in human showed that genes related to branched-chain amino acid catabolism process and transcription regulation are altered in an opposite direction between both species (Figure R3.21D). Importantly, within the transcription regulation category, we

identified several core circadian genes (such as *ARNTL*, *NPAS2* and *CLOCK*).

Circadian genes regulate many physiological processes, including metabolism and behavior, through the generation of 24h circadian oscillations in gene expression [178]. As mice and humans are nocturnal and diurnal species, respectively, we speculated that the opposite patterns of expression for these circadian genes could be due to the opposite circadian rhythm between both species. Therefore, mice at night might have the same GC-induced alterations as humans during the day. To assess this, Guillermo Garcia-Eguren (Endocrine disorders lab, IDIBAPS, Barcelona) repeated the experiments in mice but, in this case, VAT samples were obtained at the beginning of the dark phase and the gene expression profile of the main core circadian genes was assessed by qPCR. Interestingly, we observed that chronic hypercortisolism induces a differential gene expression pattern of core circadian genes in humans and mice compared to their controls (Figure R3.22A). In both species, during the active phase, chronic hypercortisolism causes a similar overexpression of circadian activators (*ARNTL*, *CLOCK* and *NPAS2*, naturally enhanced during the active phase) and a more stressed downregulation of the *PER2* repressor (naturally repressed during the active phase), compared to controls. During the rest phase, circadian activators and circadian repressors physiologically switch their expression pattern (from enhanced to repressed and from

repressed to enhanced, respectively). Notably, GC treatment induces an exacerbated repression of circadian activators (ARNTL, CLOCK and NPAS2) during the rest phase. All these results suggest that chronic hypercortisolism induces a larger amplitude oscillation pattern of the core circadian genes in VAT in both species.

Most of these alterations are maintained, to a lesser degree, after CS remission in mice (Figure R3.22B). Importantly, we also observed the long-term down-regulation of *PER1* at both, active and rest phases. This suggests that *PER1* could be the most severely affected by hypercortisolism as we found a loss in its intrinsic oscillation pattern between phases. These observations reveal that hypercortisolism markedly influences circadian gene expression in VAT, not only during active CS, but also after long-term recovery.

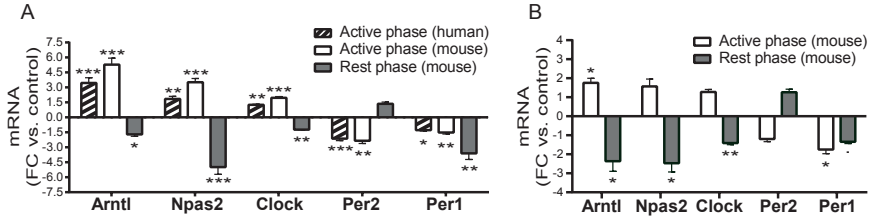

Figure R3.22: Persistent alteration of core circadian genes due to hypercortisolism in mice and human. Active CS (A) or CS-remission (B) differential gene expression analysis of core circadian genes, presented as fold change, during active and rest phase in humans and mice. In the active phase, fold change was calculated using FPKMs from each RNA-seq dataset between active CS patients, CORT_5w and CORT_15w versus their respective controls. In the rest phase, fold change was calculated from qPCR results (n= 4-5 mice/group/time). Data are means ± s.e.m., significance was assessed using a Student's t-test (***p-value < 0.001, **p-value < 0.01, *p-value < 0.05, ·p-value < 0.1).

DISCUSSION

This idea that science needs women is really right on target. The ability to solve complex problems is greatly enriched by having different viewpoints.

Elizabeth Blackburn (1948)

1. mESCs as a biological model to study active and repressed regulatory regions

To construct proper predictive models, we adopted mESCs as our model system, as the full spectrum of active and repressed enhancers can be identified. However, we decided to focus only in AEs and PEs to obtain predictive models, as IEs remain poorly understood, and their target promoters have not been unambiguously identified. In the near future, IEs could be introduced into the modelling to explore their impact on the performance of the predictions and to discover new relationships between histone modifications at enhancers and gene expression. However, this is not a limitation for our differentiation predictive models. Here, enhancers and promoters are required to be in either a poised or a bivalent state in mESCs, but many will transition towards an active, or even intermediate state, along cardiac and neural *in vitro* differentiation, and along *in vivo* embryo development. Therefore, in the subset of PEs during differentiation, we have assessed the dynamics of a complete spectrum of enhancers in cardiac (MES/CPs/CMs) and neural (NPCs/CNs) cells, and throughout developmental tissues (heart, liver, neural tube, kidney and lung).

One could speculate that the mESC enhancer model performance is an artefact of matching the chromatin state of promoters and enhancers (both active or both repressed).

However, as no expression information is introduced in this step, further conclusions are not affected by this matching procedure. Moreover, we have confirmed the predictive capacity of PEs in the differentiation PE models, in which we do not require a coordinated activation of PEs and BPs.

2. Differences in predictive model performance

At the promoter level, the performances of mESCs and differentiation predictive models were very similar ($r = 0.79$ vs. $r \approx 0.75$). The fact that we are limited to exploit a small number of histone modifications in the differentiation models (an issue that will be easily overcome when more ChIP-seq datasets are available) could be the reason for the minimal difference in the promoter models' performance. Indeed, at the enhancer level, the difference in performance between mESCs and the differentiation models was higher ($r = 0.49$ vs. $r > 0.3$). Other factors besides the number of ChIP-seq datasets could explain such a difference: (i) as mentioned before, enhancers and promoters in the mESC model were required to match their chromatin state, which could lead to overrating the performance of the mESC enhancer model; (ii) the enhancer-promoter Hi-C interactions were retrieved from data published on mESCs. Nevertheless, we predicted gene expression in cellular contexts distinct from mESCs; (iii) some genes might be specific for a cell lineage, and their interactions with

enhancers might be lost in the other cell lineages. In these cases, the enhancer and the gene would no longer be related; (iv) enhancer–promoter interactions relevant for early stages of differentiation might be lost once they have served their function. This would imply that the enhancer and its target gene are no longer coordinated. In fact, it has been shown that intensive rearrangement of promoter–enhancer interactions occurs during differentiation, and that these loops become disrupted when their target genes are repressed [52, 123].

3. Association between enhancers and target genes

How to assign enhancers to target genes is still under debate. In this study, we used Hi-C and matched chromatin states to link enhancers to genes and promoters. A recent study evaluated distinct ways of linking genes to enhancers by modelling gene expression and DNase-seq data [83]. The authors showed that expression predictive models using chromatin conformation data, such as Hi-C, performed better than those using other traditional ways of assigning target genes, such as the closest-gene method or by distance. The closest-gene method consists of assigning each enhancer to the nearest TSS. This prevents enhancers from being assigned to two or more genes but does not take into account that one enhancer can regulate the expression of more than

one gene [179, 180]. The distance method consists of assigning an enhancer to all the genes that are closer than a pre-set number of base pairs. We achieved a performance of r = 0.34 by using 1 Mb distance to assign enhancers to promoters. Although this predictive model had lower performance than the models based on Hi-C data (r = 0.38 and r = 0.49 for Hi-C–all and Hi-C–top enhancer models, respectively), it maintained the predictive capacity. This suggests that in absence of Hi-C data, using 1 Mb distance to assign target genes to enhancers performs well.

Recently, two novel computational methods have been developed to properly predict enhancer–gene associations using chromatin capture data (such as Hi-C and Hi-ChIP) and enhancer activity data (such as H3K27ac ChIP-seq and DHS-seq) [72, 73]. For instance, the so-called FOCS inference method provides a map of active enhancer–promoter associations consistent across several cellular contexts, although no cell type-specific associations could be detected [72]. Further, the activity-by-contact method identified cell type-specific associations of active enhancers and genes [73]. However, such a methodology cannot be applied to PEs due to the lack of enhancer activity. Indeed, the capacity of PEs to dynamically predict variable gene expression during differentiation suggests that our approach of assigning target genes to PEs is the most appropriate in this regulatory context.

4. Differences in contribution to gene expression prediction

We have reported differences in the histone modification contribution to the expression predictive models depending on their location in enhancers or promoters. The different contribution of each histone modification suggests that the epigenetic landscape is different in enhancers and promoters. For example, although H3K4me3 has been previously shown to be located in enhancers in particular scenarios [36, 129, 150, 151], our results suggest that its presence in enhancers has little association to gene expression. Therefore, H3K4me3 does not seem to be a good indicator of enhancer activity. In contrast, H3K4me3 proved to be key in predicting gene expression from the differentiation BP models, confirming its relevance in establishing promoter activity. Moreover, whereas H3K36me3 —mostly in intragenic PEs— proved to be important for the differentiation PE models, it showed little contribution to the BP ones.

Even though there is a universal relationship between histone modifications and gene expression, we observed that H3K36me3 is more informative in the cardiac PE models than in the neural PE models. We reached this conclusion thanks to our LOESS normalization approach, which allowed us to reduce biases when comparing heterogeneous datasets (RNA-seq and ChIP-seq), such that our results were not

influenced by the different origin of data. Without such a normalization, the conclusions reached would be wrong. However, it is worth mentioning that we assume constancy of ChIP-seq signal and expression, although they might change in their abundance during differentiation. This problem will be solved in the future with spike-in normalization.

Strikingly, H3K27me3 was found to be the most important histone modification in the majority of predictive models for enhancers and promoters. This suggests that H3K27me3 plays a key role in gene regulation, as it is important for both types of regulatory regions. Our results show that mainly H3K27me3, and also in combination with H3K36me3 and H3K27ac, are sufficient to predict future gene expression from PEs. In any case, the predictive power of our models will benefit in the future from the introduction of other histone modifications into the modelling, which can be extremely useful for identifying unknown quantitative relationships between histone modifications at enhancers and gene expression.

5. Other types of data could be introduced as variables in the predictive models

Other types of information could also be introduced in the modelling in the future. In fact, previous work has modelled gene expression using accessibility data (e.g. DHS-seq) [27,

82, 83], and other types of ChIP-seq samples (e.g. TFs or RNA polymerase II) [23, 24, 26, 81]. It would be also interesting to take advantage of enhancer RNA (eRNA) data to predict gene expression of target genes. Promising results have been obtained in predicting eRNA transcription by modelling GRO-seq and histone modification ChIP-seq at enhancers [181]. All this information at enhancers could be integrated into the modelling to improve the power and, most importantly, to discover new quantitative relationships between gene expression and multiple epigenetic features.

6. Differences between poised and intermediate enhancers

We have observed that the epigenetic landscape of PEs and IEs is different, which could indicate that they have different roles in mESC. Indeed, we have observed that PEs are enriched in factors associated to repression such as PcG, and also in factors associated to activation such as CHD8 and p400. On the contrary, IEs showed little enrichment in all these factors. Importantly, we have found that PEs, and not IEs, are enriched in paused RNA polymerase, and are more accessible than IEs. These findings go in line with previous observations from our lab in which we found that BPs favor an open chromatin architecture in mESCs which allows the proper modulation of developmental gene expression [97]. Thus, PEs

might also participate in this mechanism. All these findings suggest that PEs are indeed primed for future activation, whereas IEs might have a different role. Therefore, we believe that the alternative name for IEs, which is primed enhancers, might be misleading.

We and others have found that PEs colocalize with CGIs [124, 166], which is consistent with PcG recruitment [182] and therefore goes in line with H3K27me3 deposition at these regions. In fact, depletion of the CGIs in PEs reduces H3K27me3 significantly [124]. In line with this PcG recruitment at PEs, we have demonstrated that depletion of PcG subunits in mESCs has the same effect on H3K27me3 at PEs than what was previously described for BPs [115, 118]. One might think that IEs could constitute a primed state prior to PE state. However, IEs do not overlap with CGIs, and therefore further confirms that IEs cannot be primed for future activation by PcG. Thus, it seems that either it exists an alternative regulation to prime IEs for future activation, or IEs have a different role in mESCs.

7. Poised enhancer activation is not exclusive of the neural lineage

Historically, PE activation in ESCs was associated to differentiation towards neural lineage [124-127]. However,

recent evidence suggested that this mechanism of PE activation might be general to other lineages [128]. Indeed, we have been able to identify activated PEs throughout not only neural lineage but also cardiac lineage, and in several embryonic tissues. Thus, we believe that this mechanism is general for all differentiation processes from mESCs and in development. However, further experiments should be performed in order to validate what we observed genome-wide. For example, knock-out of activated PEs in cardiac lineage should be performed in order to study the implications in target gene expression during differentiation from mESCs towards CMs. Another interesting approach would be to activate a specific PE in mESCs and see whether its target gene starts to be expressed. Both validations would benefit from CRISPR-Cas9 genome engineering [57]. Moreover, the interaction between the PE and its BP should be validated by 3C and/or 4C.

8. Model for poised enhancer activation during differentiation

Interestingly, we have found that the mechanisms of PE activation are more specific than those of BP activation. Indeed, we observed a relatively basal activation of BPs at all differentiation time points and at all embryonic tissues. Therefore, we hypothesize a model for PE activation to tightly

control developmental gene expression (Figure D1) in which an important subset of genes has their BPs in an activated state throughout differentiation, but only those truly lineage-specific genes would also have the appropriate PE also activated. Therefore, those genes in which the corresponding PE was triggered would be more expressed than those that were only activated through the activation of the BP. In this regard, we exemplified the case of the *Isl1* gene, which is highly expressed in CMs and accordingly, a clear activation of its PEs is observed, whereas in CNs, it is expressed at lower levels as it only activates its BP. In line with this, *Isl1* has a key role in cardiomyocyte cell fate [183, 184] and its loss of function produces congenital heart defects [185]. Moreover, it seems to be temporally required for sympathetic nervous system development [186]. Therefore, the proper regulation of BP and PE activation might establish different levels of target gene expression specific of different cellular lineages.

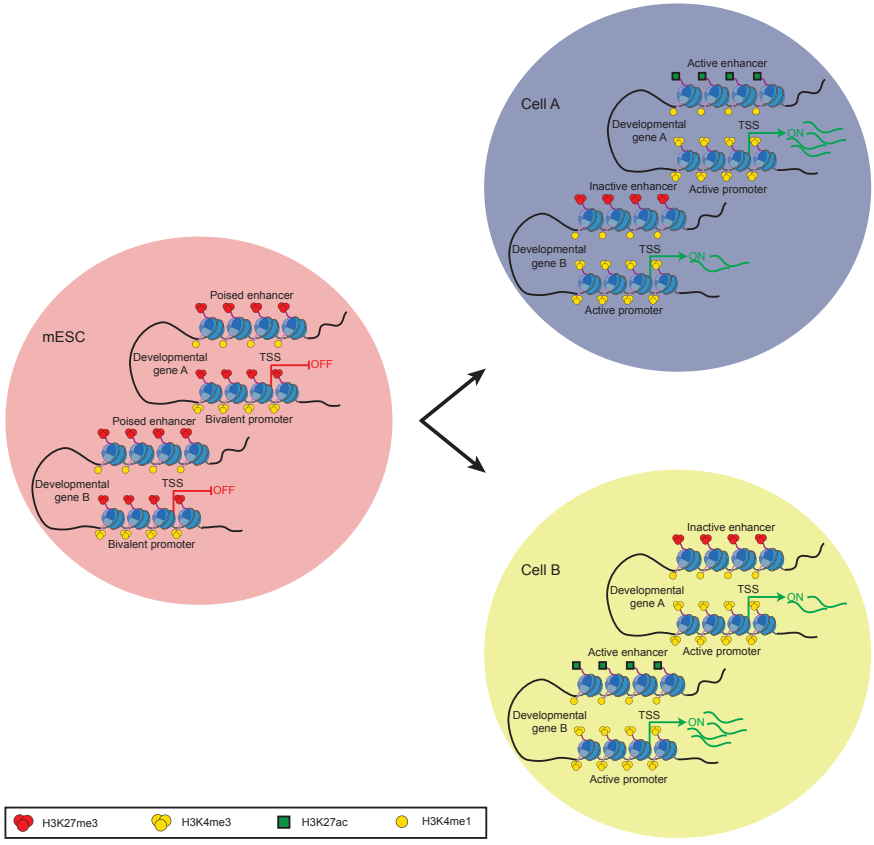

Figure D1: Model for PE activation during differentiation from mESCs. In mESC developmental gene A and B are not expressed. During differentiation, both genes will have their respective BPs in an activated state and the gene will start to be expressed. However, in Cell A, the PE of developmental gene A will become active and therefore its expression will be higher than those of gene B. On the contrary, in Cell B, it will be developmental gene B the one that becomes active, and therefore developmental gene B will be more expressed than developmental gene A.

Interestingly, we have also found that there are some genes that have the same activated PE at all differentiation time points and at all embryonic tissues. It is known that the expression of the same gene can be governed by different enhancers in different cell types [68]. Thus, it seems that this mechanism does not apply to PEs, and it might be specific of

enhancers established *de novo* during development. Indeed, most of the enhancers seem to appear concomitantly with the activation of the gene during development [68, 70]. Therefore, PEs might constitute the less abundant common basic regulatory network needed to establish any differentiation process. On the contrary, *de novo* enhancers might be a second layer of regulation that further supports the specific regulatory program that needs to be activated.

9. Pipeline to obtain consensus peaks among replicates

We have developed a novel pipeline to obtain significant consensus sets of peaks across ChIP-seq replicates. This statistical approach, based on the application of software for the identification of differential peaks, allows us to recover interesting peaks not called in one of the replicates, to discard peaks present only in one replicate and discard false positives. Here, we have tested its performance in broad peaks such as H3K27me3 and sharp peaks such as H3K4me3. Indeed, this novel approach has proved to be useful also in other biological scenarios [187, 188]. Interestingly, in these two publications we have assessed our pipeline not only over histone modifications but also on PcG subunits. Therefore, we believe this approach can become extremely useful for consistently dealing with

replicates of any type of ChIP-seq experiment in a short-term future.

10. Persistent epigenetic and transcriptional signatures after Cushing's syndrome long-term remission

In this thesis we have been able to identify for the first time, transcriptional and epigenetic signatures during GC overexposure common at both, humans and mice. Moreover, these signatures were maintained after resolution of the GC treatment in mice. Notably, the most commonly up-regulated genes during GC overexposure and after resolution were associated to immune response and extracellular matrix organization, both categories previously related with obesity and insulin resistance [189, 190]. In turn, obesity and insulin resistance have previously been linked to GC overexposure [136, 137]. Indeed, these observations confirm the relevance of macrophage infiltration and low-grade pro-inflammatory status in GC-induced VAT dysfunction previously described [141, 174, 175]. Interestingly, the immune response category was not previously described in the unique published RNA-seq dataset from subcutaneous adipose tissue in CS patients [176], which remarks the novelty of our findings.

Importantly, alterations in the levels of both H3K4me3 and H3K27ac seem to explain the differences observed in gene expression in active CS and after remission. These results indicate that chronic hypercortisolism induces an altered epigenetic fingerprint in VAT maintained through time, which could explain the persistent changes in gene expression after remission. Indeed, we have observed that genes related with immune response are among the genes with higher levels of H3K27ac in mice during and after GC overexposure. Thus, these findings open the window to epigenetic therapies for the treatment of the persistent comorbidities and reinforce the importance of a lifelong follow-up of these patients.

11. Alteration of the circadian rhythm by GC overexposure

Surprisingly, most of the commonly down-regulated genes during active and after CS remission in mice followed an opposite tendency in CS patient's transcriptome. Among these genes with an opposite trend, we identified several circadian clock genes, such as *ARNTL*, *NPAS2* and *CLOCK*. As mice and humans are nocturnal and diurnal species, respectively, we speculated that the opposite tendencies of these circadian genes could be explained by the opposite circadian rhythm between both. Confirming this hypothesis, we observed a similar gene expression pattern during the active phase of both

species, which suggested a larger amplitude oscillation of these circadian genes.

Interestingly, we observed a long-term down-regulation of *PER1* in both, active and rest phase. Therefore, *PER1* seems to lose its intrinsic oscillation pattern between phases and may be the trigger of the dysregulation of the circadian feedback loop induced by GCs. Consistent with this hypothesis, a previous report revealed a GC receptor element (GRE) on the *Per1* gene [191], indicating a direct modulation of *Per1* through GC receptor (GR) binding. In line with this, we observed a reduction in the expression of the GR gene (*NR3C1*) due to chronic hypercortisolism. Thus, we postulate that GC overexposure induces GR reduction levels in VAT which impairs the proper circadian induction of *PER1* through GR. Moreover, as *PER1*, *KLF15* is also down regulated in both species and is a target of GR [192, 193]. Interestingly, *Klf15* regulates circadian genes expression in the heart of mice, and in its absence, several non-cycling genes show circadian oscillations [194].

Core clock genes control the expression levels of almost 25% of all the genes in the adipose tissue [195], which could explain why we observe opposite tendencies in gene expression between mouse and human. Interestingly, circadian rhythm impairment has been related with obesity, diabetes, cardiovascular risk, poor sleep and depression [178], and most

of this symptomatology is present in CS patients even after long-term remission [137, 177].

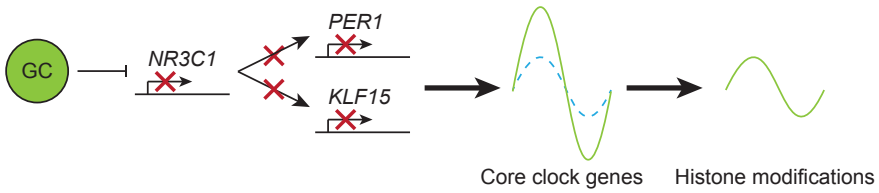

Figure D2: Model of circadian alteration by GCs. GC overexposure down regulates the GR gene (*NRC1*), which in turn down regulates *PER1* and *KLF15* genes. This leads to an exacerbation of the circadian oscillation of the core clock genes (in dashed blue the normal oscillation, in green the altered oscillation by GCs), which causes a circadian oscillation of some histone modifications.

Besides the alteration of the circadian genes, we also observed that the epigenetic landscape is altered in an opposite pattern between mice and CS patients. As human and mouse samples were both taken during the morning (active phase in human and rest phase in mouse), it seems that the opposite epigenetic pattern could be caused by a reverse and exacerbate circadian cycle. In line with this, it is known that the circadian cycle reshapes chromatin in many levels [196, 197], and importantly, oscillations of H3K4me3 and H3K27ac signals have been observed at some genes in the mouse liver [198]. Moreover, GR and core clock proteins are known to interact with chromatin and chromatin-modifying complexes [148, 196], and CLOCK has a histone acetyltransferase (HAT) activity [199]. Therefore, we hypothesize a model (Figure D2) in which GC overexposure causes a down regulation of *PER1* and *KLF15* through the down regulation of the GR, which in turn alters

clock circadian genes by exacerbating their oscillation, which finally causes a genome-wide oscillation at least in some histone modifications. However, in order to confirm the circadian oscillation of the histone modification signals, ChIP-seq experiments of mice during its active phase (at night) would be needed.

CONCLUSIONS

The first thing about empowerment is to understand that you have the right to be involved. The second one is that you have something important to contribute. And the third piece is that you have to take the risk to contribute it.

Mae Jemison (1956)

From the results presented on this thesis we can conclude the following:

1. Enhancers, as well as promoters, are good predictors of gene expression.

2. Associating enhancers to target genes by Hi-C ensures better results than by 1 Mb distance, although results with the latest are still acceptable.

3. H3K27me3 is the most important mark for predicting gene expression from enhancers in mESCs.

4. H3K27me3 is the most important mark for predicting gene expression from promoters in mESCs, although other marks are similarly important.

5. LOESS normalization performs well in RNA-seq and ChIP-seq data from a variety of sources.

6. Poised enhancers and bivalent promoters are good predictors of gene expression at differentiation time points and in mouse embryonic tissues.

7. H3K27me3 is the most universal variable to predict gene expression from intergenic and intragenic poised enhancers.

8. H3K4me1 and H3K4me3 show little importance to predict gene expression from poised enhancers, whereas H3K4me3 is one of the most predictive variables to predict gene expression from bivalent promoters.

9. Different histone modifications relate better to enhancer and promoter function, respectively.

10. Poised enhancers are primed for future activation whereas intermediate enhancers might have a different role in mESCs.

11. Poised enhancers are the most evolutionary conserved type of enhancers in mESCs, and often overlap with CpG islands.

12. Poised enhancer activation is not an exclusive mechanism of neural lineage development.

13. Poised enhancer activation is more cell type-specific than bivalent promoter activation.

14. Target genes of activated poised enhancers are more expressed than those that only have an activated bivalent promoter.

15. Commonly activated bivalent genes at several lineages have the same activated poised enhancer.

16. We have designed a successful pipeline to obtain consensus peaks across ChIP-seq replicates.

17. There is a genome-wide increase of H3K4me3, H3K27ac and H3K27me3 signals caused by chronic hypercortisolism in mice maintained even after long-term resolution.

18. There is a genome-wide decrease of H3K4me3, H3K27ac and H3K27me3 signals caused by chronic hypercortisolism in human CS patients.

19. Genes related to immune response and extracellular matrix organization are up-regulated in both, human CS patients and the mouse model.

20. H3K4me3 and H3K27ac explain the observed transcriptional changes in both, human CS patients and the mouse model.

21. The circadian rhythm is deeply altered during and after chromic hypercortisolism.

22. The circadian rhythm could explain the opposite epigenetic pattern observed between the mice model and human CS patients.

MATERIALS AND METHODS

Estando igual de preparadas, si a la vez vemos que hay menos mujeres arriba, claramente estamos ante una situación de discriminación, no es un problema de competitividad, de que no haya mujeres suficientemente buenas. Claro que las hay. El problema es otro. Los cupos simplemente sirven para asegurar que quien sea que tome la decisión, que normalmente son en su mayoría hombres, considere a las mujeres.

María Blasco Marhuenda (1965)

Cell culture

E14Tg2A mESCs were cultured feeder-free on 15-cm plates coated with 0.1% gelatin. Plates were coated with gelatin for 15 min at 37°C, and then non-bound gelatin was removed. mESCs were cultured with Glasgow minimum essential medium (Sigma) supplemented with β-mercaptoethanol, sodium pyruvate, penicillin–streptomycin, non-essential amino acids, GlutaMAX, 20% fetal bovine serum (Hyclone), and leukemia inhibitory factor (LIF).

By Cecilia Ballaré (our lab), Malte Beringer (former member of the lab), Alexandra Santanach (former member of the lab), and Lluís Morey (former member of the lab).

Animal treatment and design

Six-week-old C57BL/6J male mice were group-housed in a facility with 12-h light-darkness cycle (Light cycle 08:00h to 20:00h). After one week of acclimation, two different groups were established for 5 weeks of treatment. The CORT group was fed *ad libitum* on a chow rodent maintenance diet (Envigo, Huntingdon, UK) and a stable administration in drinking water of 500 µg of free corticosterone (Sigma-Aldrich, Madrid, Spain) per mouse and day dissolved in 100% ethanol (EtOH), as vehicle, up to a final EtOH concentration of 0.66%. Water

consumption was measured twice a week, and, if necessary, solutions were replaced with the proper amount of corticosterone needed to maintain a daily exposure of approximately 500 μg of corticosterone as described previously [172]. The vehicle (VEH) group was fed *ad libitum* on the same chow diet and to control for the effects of dietary EtOH in adiposity, mice were also given regular drinking water with a final EtOH concentration of 0.66%. After 5 weeks of treatment, CORT mice switched to VEH for an additional 10 weeks to assess recovery from the corticosterone treatment. The CORT group underwent a 1-week washout period of corticosterone dose tapering (75–50% to 25–0% of treatment dose), in order to avoid adrenal insufficiency. Body weight, food intake and water intake were measured weekly. At the beginning of the light cycle (9:00 a.m.), epididymal fat pads were removed under fasting conditions at the end of the 5 weeks of treatment and after 10 weeks of recovery. All animal procedures were approved by and conducted in accordance with the Animal Research Committee of the University of Barcelona.

By Guillermo García-Eguren (Endocrine Disorders lab, IDIBAPS, Barcelona).

Patient recruitment

A cross sectional study was performed in patients with active adrenal endogenous CS referred to the Endocrinology Department of Hospital Clínic from Barcelona, between 2015 to 2018 (n = 6). Diagnosis criteria of CS followed the European Society of Endocrinology and the Endocrine Society guidelines [200] through repeated elevated levels of urinary free cortisol, loss of circadian rhythm (elevated free night salivary cortisol) and lack of suppression of cortisol secretion after dexamethasone administration. Localization of the cortisol-secreting tumour was determined by ACTH level and imaging tests. None of the patients had been on steroidogenesis inhibitors treatment. Patients with genetic forms of CS due to adrenal macro/micronodular hyperplasia were excluded. The number of cases during the enrolment period determined the sample size of CS patients. Eligible controls (n = 12) were identified from subjects undergoing elective procedures of abdominal surgery to correct benign conditions in the surgery departments of the Hospital Clínic, Barcelona.

Exclusion criteria in both groups were age above 70 or below 18 years old; acute illness other than CS; pregnancy and previous history of abdominal surgery. Control subjects with history of adrenal incidentaloma, severe chronic diseases other than those related to metabolic syndrome, recent or active malignancy, previous recent use of exogenous GC or

other drugs that could interfere with the hypothalamic-pituitary-adrenal axis were excluded. The study was approved by the Hospital's Ethics Committee and written informed consent was obtained from all participants. Main biochemical parameters of all participants were measured in serum of the patients and controls after overnight fasting employing standard laboratory methods as previously described [201]. Free night salivary cortisol was measured to discard CS in controls. Participants were matched for age, BMI, sex and cardiometabolic phenotype: type 2 diabetes, hypertension (HTN) and dyslipidemia (DLP). Overview of clinical characteristics of participants is displayed in Table M1.

Table M1: Patient characteristics

	Cushing Syndrome (n=6)	**Controls** (n=12)	***P* value**
Sex (male/female)	2/4	2/10	0.569
Age (years)	48.8 ± 11.1	49.4 ± 9.0	0.456
BMI (kg/m^2)	27.6 ± 3.8	27.0 ± 2.3	0.602
HTN (%)	5 (83)	11 (91)	0.985
DLP (%)	3 (50)	6 (50)	1.000
TC (mmol/L)	5.3 ± 0.6	5.2 ± 0.5	0.868
Triglycerides (mmol/L)	1.5 ± 0.6	1.4 ± 0.4	0.650
T2D (%)	2 (33%)	4 (33%)	1.000
Fasting glucose (mmol/L)	6.2 ± 0.7	6.3 ± 0.9	0.861
HOMA-IR	3.8 ± 2.0	2.7 ± 1.4	0.040*
MSC (µg/dL)	23.6 ± 8.4	17.8 ± 3.5	0.001*
UFC (µg /24h)	452 (201-880)	-	-
LNFSC (µg/dL)	6.41 ± 2.3	1.01 ± 0.4	0.001*

BMI: body mass index. HTN: hypertension. DLP: dyslipidemia. TC: total cholesterol, LDL-c: low-density lipoprotein-, HDL-C: high-density lipoprotein-cholesterol. T2D: diabetes mellitus type 2. HOMA-IR: HOMA-IR: [Insulin mUI/L x Glycemia: (mmol/L)/22.5. 24h-UFC: 24h urinary free cortisol, MSC: morning serum cortisol; LNSC: late night free salivary cortisol. Significance was assessed using Student's t-test for continuous data and Fisher's exact test for categorical data (*p-value < 0.05). Data are presented as mean ± S.D.

By Felicia Hanzu (Hospital Clínic and Endocrine Disorders lab, IDIBAPS, Barcelona).

Visceral fat biopsy

Visceral (omental) adipose tissue biopsies from CS patients and matched controls were isolated after overnight fasting, at the beginning of the laparoscopic surgery, between 8:30-10:00 a.m. and immediately transported to the laboratory for analysis. All subsequent procedures were carried out under laminar airflow and sterile conditions. VAT adipose tissue samples were immediately carefully cleaned and approximately 500-1000 mg of this fat graft was snap frozen in liquid nitrogen and stored at -80º C until further processed as described below. All human VAT biopsies were used for RNA-seq experiments (n = 6 for CS patients; n = 12 for control patients). Eight VAT samples from CS patients and matched controls (n = 4/group) were used for ChIP-seq experiments.

By Felicia Hanzu (Hospital Clínic and Endocrine Disorders lab, IDIBAPS, Barcelona).

Chromatin immunoprecipitation

mESC

Cells were grown in 15-cm plates until 70% confluency and crosslinked in 1% formaldehyde in growth medium for 10 min at room temperature in a shaker. To stop fixation, glycine was added to a final concentration of 0.125 M and incubated for 5 min at room temperature. Cells were then washed twice with ice-cold PBS and harvested by gently scrapping plates (on ice) in PBS plus protease inhibitors. Cells from two 15-cm plates were pooled together and centrifuged at 3,400 × *g* at 4°C for 5 min. Cell pellets were frozen at –80°C until use.

Chromatin was prepared by resuspending the crosslinked pellet in 1.3 ml ice cold ChIP buffer [1× volume SDS buffer (100 mM NaCl, 50 mM Tris-HCl pH 8.1, 5 mM EDTA pH 8.0, and 0.5 % SDS) and 0.5 × volume Triton dilution buffer (100 mM NaCl, 100 mM Tris-HCl pH 8.6, 5 mM EDTA pH 8.0, and 5% Triton X-100)] plus proteinase inhibitors. Samples were sonicated 40 cycles (30 seconds on / 30 seconds off) in a Bioruptor Pico (Diagenode) and centrifuged at 16,000 × *g* at 4°C for 20 min to remove the cell debris. To check chromatin size, a 25-µl aliquot was mixed with 175 µl of PBS plus 5 µl of 20 mg/ml proteinase K, and de-crosslinked for 5 h at 65°C. DNA was purified using the QIAquick PCR purification kit

(Qiagen), quantified in Nanodrop, and checked by electrophoresis on a 1.2% agarose gel.

ChIP experiments were performed using 30 µg of chromatin (DNA) and 5 µg of antibody in a final volume of 500 µl ChIP buffer. Aliquots of 5 µl were removed as input material (1%). ChIP samples were incubated overnight at 4°C on rotation, and then Protein A agarose beads (Diagenode) (42 µl per ChIP) were blocked 30 min with 0.05% BSA, washed, and added to the ChIP reaction. Samples were incubated for 2 h at 4°C with rotation. After incubation, beads were washed three times with 1 ml of low-salt buffer (140 mM NaCl, 50 mM HEPES pH 7.5, and 1% Triton X-100) and once with 1 ml high-salt buffer (500 mM NaCl, 50 mM HEPES pH 7.5, and 1% Triton X-100). ChIPed material was eluted from the beads in 200 µl freshly prepared elution buffer (1% SDS, 100 mM NaHCO3) at 65°C in a shaker (1000 rpm) for 1 h. Input samples were also brought to 200 µl with elution buffer. After addition of 8 µl of 5 M NaCl to the eluted chromatin and input samples, samples were de-crosslinked overnight at 65°C. The next day, samples were treated with proteinase K [1 µl of 20 mg/ml Proteinase K, plus 4 µl 0.5 M EDTA, and 8 µl Tris-HCl pH 6.5] for 1 h at 45°C. ChIPed DNA and inputs were purified using the QIAquick PCR purification kit (Qiagen) and eluted in 60 µl. The following antibodies were used in the ChIP experiments: H3K27me3 (Millipore, #07-449); H3K4me3 (Diagenode, C15410003); H3K4me1 (Abcam, ab8895); H3K27Ac (Millipore, #07-360);

H3 (Abcam, Ab1791); H3K36me3 (Abcam, ab9050); H3K27me1 (Active Motif, #61015); H3K27me2 (Cell Signaling, #9728); H3K79me2 (Abcam, ab3594); H2Bub (Cell Signaling, #5546); H4K20me3 (Abcam, ab9053); and PHC1 (Active Motif, #39723).

By Cecilia Ballaré (our lab), Malte Beringer (former member of the lab), Alexandra Santanach (former member of the lab), and Lluís Morey (former member of the lab).

Visceral adipose tissue

Freshly pooled epididymal fat depots from mouse (n = 2-4 mice/pool, 2 pools/group/time point) or frozen individual visceral fat biopsies from patients (n = 4 samples/group) were chopped into small pieces of about 3 mm (McIlwain Tissue Chopper, Redding, CA, USA) and crosslinked in fixing solution (PBS + formaldehyde 0.5%) for 5 minutes in rotation at room temperature as previously described [202]. Briefly, to stop the fixation, glycine was added to a final concentration of 0.125 M and incubated for 5 min at room temperature and under rotation. The tissue was washed twice in cold PBS plus proteinase inhibitor cocktail (PIC) (Thermofisher, MA, USA), centrifuged and the liquid phase discarded with a glass pipette. Then, the tissue was homogenized with Ultra-Turrax (IKA, Staufen, Germany), and centrifuged at 2870xg for 2 mins in

order to precipitate and store the nuclei fraction at -80ºC until use. Nuclei were lysed in immunoprecipitation (IP) buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1 + Protease/phosphatase inhibitors) and sonicated (40 cycles of 30 seconds on/30 seconds off) in a Bioruptor (Diagenode, NJ, USA). ChIP was performed using 30 μg chromatin/ChIP and 5 μg/ChIP of the following antibodies: H3K27me3 (Millipore #07–449), H3K27ac (Millipore #07-360), H3K4me3 (Diagenode # C15410003-50) and control IgG (Abcam #ab172730). ChIP experiments were performed with spike-in control as previously reported [187]. For this, an equal amount of *Drosophila melanogaster* S2 cell chromatin was added to each ChIP reaction (2.5% of the chromatin for all the ChIPs), together with 1 μg of an antibody against a *Drosophila melanogaster* specific histone variant, H2Av (Active Motif, catalog no. 61686).

By Guillermo García-Eguren (Endocrine Disorders lab, IDIBAPS, Barcelona) and Pedro Vizán (our lab).

RNA extraction

Total RNA was extracted from mouse epididymal fat pad or human visceral fat biopsies stored at -80ºC using the RNeasy Lipid Tissue Mini Kit (Qiagen, Madrid, Spain).

By Guillermo García-Eguren (Endocrine Disorders lab, IDIBAPS, Barcelona).

Quantitative real-time PCR

1 μg of mouse RNA was reverse-transcribed into cDNA with a SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific, Barcelona, Spain), according to the manufacturer's instructions. Quantitative real-time PCR was performed using SYBR Green reagents on a 7900HT real-time PCR system (Applied Biosystems, Foster City, CA, USA). Primers are listed in Table M2. mRNA expression levels were normalized to *Tbp* as the internal control, using the cycle threshold (2−ΔΔCt) method.

By Guillermo García-Eguren (Endocrine Disorders lab, IDIBAPS, Barcelona).

Table M2: Primer sequences for quantitative real-time PCR.

Mouse	Forward	Reverse
Arntl	CCAGGGTTTGAAGTTAGAGTCC	TGAAGTCGCTGATGGTTGAG
Npas2	ATGTTCGAGTGGAAAGGAGAC	CAAGTGCATTAAAGGGCTGTG
Clock	TCTCAAGGAAGCACTGGAAAG	CAGTAGGGATCTTTGTCGGTG
Per1	CATGTCTACTTACACCCTGGAG	TGCCTGCTCCGAAATATAGAC
Per2	AGCAGGTTGAGGGCATTAC	TTACAGTGAAAGATGGAGGCC
Tbp	ACCCTTCACCAATGACTCCTATG	ATGATGACTGCAGCAAATCGC

Library preparation and sequencing

Library preparation for ChIP-seq experiments was performed at the UPF/CRG Genomics Unit. Libraries were sequenced using Illumina HiSeq2500 sequencer. Libraries for RNA-seq were prepared from total RNA at the CRG Genomics Unit and sequenced using the Illumina HiSeq2500 sequencer (2x50, v4, HiSeq high output mode).

Input datasets

Raw data of multiple samples from the literature was downloaded and reanalyzed to be included in this PhD thesis. RNA-seq data and ChIP-seq data of SUZ12, EPOP and JARID2 in mESCs; and ChIP-seq of H3K27me3 in knock-down of *Epop* and its knock-down control in mESCs were extracted from a previous publication from our lab (GEO accession number: GSE79606) [118]. ChIP-seq data of RING1B, CBX7 and RYBP in mESCs were obtained from another previous publication from our lab (GEO accession number: GSE42466) [158]. ChIP-seq data of MEL18 in mESCs was extracted from another previous publication from our lab (GEO accession number: GSE67868) [117]. ChIP-seq data of PHF19 in mESCs and ChIP-seq of H3K27me3 in knock-down of *Phf19* and its knock-down control in mESCs were obtained from a previous publication from our lab (GEO accession number: GSE41589)

[115]. ChIP-seq data of p300 in mESCs was obtained via GEO (GEO accession number: GSE89211) [124]. ChIP-seq data of CHD7 in mESC was retrieved from GEO (GEO accession number: GSE22341) [159]. ChIP-seq data of CHD8 and p400 in mESC was extracted from GEO (GEO accession number: GSE64825) [160]. ChIP-seq data of YY1 in mESC was obtained via GEO (GEO accession number: GSE92412) [162]. ChIP-seq data of POL2 S5P and POL2 S2P was extracted via GEO (GEO accession number: GSE34518) [163]. ATAC-seq data of mESC was retrieved from GEO (GEO accession number: GSE85505) [134]. ChIP-seq data of H3K27me3, H3K4me3, H3K27ac, H3K4me1, and H3K36me3, and RNA-seq data of cardiac differentiation (MES, CPs and CMs), were obtained from https://b2b.hci.utah.edu/gnomex/ (accession numbers: 44R and 7R2) [153]. ChIP-seq data of H3K27me3, H3K4me3, H3K27ac, H3K4me1, and H3K36me3, RNA-seq data of neural differentiation (NPCs and CNs), ChIP-seq data of CTCF and Hi-C data of mESCs were retrieved from GEO (GEO accession number: GSE96107) [123]. ChIP-seq data of H3K27me3, H3K4me3, H3K27ac, H3K4me1, and H3K36me3, and RNA-seq of mouse developmental tissues (Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5, Lung15.5) and ChIP-seq of CHD2 in mESC were obtained from ENCODE project [157]. The list of ENCODE accession numbers can be found in Table M3. When replicates were available, pooling was done except for the ChIP-seq samples of H3K4me3 of NPCs

(replicate 1 was used) and H3K27ac of NPCs (replicate 2 was used).

Table M3: List of ENCODE accession numbers

Developmental stage	Experiment	Experiment accession
Heart10.5	H3K27me3 ChIP-seq	ENCSR266JQW
Heart10.5	H3K4me3 ChIP-seq	ENCSR782DEA
Heart10.5	H3K27ac ChIP-seq	ENCSR582SPN
Heart10.5	H3K4me1 ChIP-seq	ENCSR782DGO
Heart10.5	H3K36me3 ChIP-seq	ENCSR328WMV
Heart10.5	RNA-seq	ENCSR049UJU
Liver11.5	H3K27me3 ChIP-seq	ENCSR244KCW
Liver11.5	H3K4me3 ChIP-seq	ENCSR447DOF
Liver11.5	H3K27ac ChIP-seq	ENCSR058DOA
Liver11.5	H3K4me1 ChIP-seq	ENCSR419MSI
Liver11.5	H3K36me3 ChIP-seq	ENCSR932BNP
Liver11.5	RNA-seq	ENCSR284AMY
NeuralTube12.5	H3K27me3 ChIP-seq	ENCSR375RUA
NeuralTube12.5	H3K4me3 ChIP-seq	ENCSR538SRO
NeuralTube12.5	H3K27ac ChIP-seq	ENCSR891SAW
NeuralTube12.5	H3K4me1 ChIP-seq	ENCSR263CKR
NeuralTube12.5	H3K36me3 ChIP-seq	ENCSR094QZC
NeuralTube12.5	RNA-seq	ENCSR508GWZ
Kidney14.5	H3K27me3 ChIP-seq	ENCSR399UVI
Kidney14.5	H3K4me3 ChIP-seq	ENCSR669AQL
Kidney14.5	H3K27ac ChIP-seq	ENCSR057SHA
Kidney14.5	H3K4me1 ChIP-seq	ENCSR196ENU
Kidney14.5	H3K36me3 ChIP-seq	ENCSR425FLT
Kidney14.5	RNA-seq	ENCSR504GEG
Lung15.5	H3K27me3 ChIP-seq	ENCSR861MUP
Lung15.5	H3K4me3 ChIP-seq	ENCSR305GII
Lung15.5	H3K27ac ChIP-seq	ENCSR895BMP
Lung15.5	H3K4me1 ChIP-seq	ENCSR858AUB
Lung15.5	H3K36me3 ChIP-seq	ENCSR776RJR
Lung15.5	RNA-seq	ENCSR457RRW
mESC	CHD2 ChIP-seq	ENCSR531HWD

Accession numbers of the ENCODE data used in this PhD thesis. Heart10.5, heart tissue from 10.5 embryonic day; Kidney14.5, kidney tissue from 14.5 embryonic day; Liver11.5, liver tissue from 11.5 embryonic day; Lung15.5, lung tissue from 15.5 embryonic day; NeuralTube12.5, neural tube tissue from 12.5 embryonic day.

ChIP-seq analysis

mESCs, cardiac differentiation, neural differentiation and mouse embryonic tissues

The sequence reads of ChIP-seq data from mESCs, MES, CPs, CMs, NPCs, CNs, Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5 and Lung15.5 were mapped to the mm10 version of the mouse genome with the BOWTIE software [41], setting the option –m 1, which eliminates multilocus reads that align in more than one region. The ChIP-seq profiles were obtained using the function `buildChIPprofile` from SeqCode [203]. For the p300 ChIP-seq, peak calling against input was performed using MACS [45] with the option --shiftsize 100, which shifts tags to their midpoint.

Visceral adipose tissue

The sequenced reads of ChIP-seq data from mice were mapped to a synthetic genome constituted by the mouse and the *Drosophila melanogaster* chromosomes (mm9 + dm3), and the ones from human were mapped to a synthetic genome constituted by the human and the fruit fly chromosomes (hg19 + dm3). The mapping was performed with the BOWTIE software [41], setting the option -m 1, which eliminates those multilocus-reads which align in more than one region. The

ChIP-seq profiles normalized by the total number of fruit fly spike-in reads [204] were obtained using the function `buildChIPprofile` from SeqCode [203]. ChIP-seq signal represents the count of reads normalized by total number of fruit fly spike-in reads and averaged by the length of the genomic coordinates, and was calculated with `recoverChIPlevels` function from SeqCode [203]. Peak calling of each replicate against its corresponding input was performed using MACS [45] with the option –shift size 100, which shifts tags to their midpoint.

Chromatin segmentation

ChromHMM [49] was used to obtain a chromatin segmentation model for mESCs, MES, CPs, CMs, NPCs, CNs, Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5 and Lung15.5 using the default parameters. The input data for ChromHMM consisted of ChIP-seq experiments of H3K4me3, H3K27ac, and H3K4me1 in all cell types, plus H3K27me3 in the case of mESCs. When available, ChIP-seq of H3 was used as control, otherwise, Input was used. First, the function `BinarizeBam` was used to binarize the input mapped data. Next, the `LearnModel` function was ran to learn different chromatin segmentation models of each cell type or tissue separately, using from 4 to 8 states for all the cases except for mESCs were we used from 4 to 16 states; the models were selected

when they showed the higher number of states with no redundancy.

Unification of ChIP-seq peaks

To obtain a unified list of peaks across replicates of the same experiment, DiffBind [48] was run against input. First, in the case of mice VAT (where two replicates of each ChIP-seq experiment are available), both lists of peaks reported by MACS were pooled and overlapping peaks in at least 1 bp were merged taking the lowest and the highest coordinate. A Python script was developed to perform this first step. In the case of human VAT (where four replicates of each ChIP-seq are available), only peaks which overlap in at least 1 bp with another peak from another replicate were selected, and overlapping peaks in at least 1 bp were merged taking the lowest and the highest coordinate to define a merged peak. This first step in human was done with DiffBind [48] Next, peaks with a significant enrichment of ChIP-seq signal over input with a p-value < 0.05 and an FDR < 0.1, were selected to obtain a unified list of peaks across replicates for the same experiment in each condition (CORT_5w, VEH_5w, CORT_15w, VEH_15w, CS and CTL). In the case of H3K27ac in CORT_15w, all the peaks of the replicate 2 were selected as unified peaks and the peaks of replicate 1 were discarded.

Differential ChIP-seq analysis

To obtain the lists of differential peaks from each comparison (CORT_5w vs. VEH_5w, CORT_15w, vs VEH_15w and CS vs. CTL) we run DiffBind [48], using the total number of *Drosophila melanogaster* reads for each replicate as spike-in normalization factors, and setting the threshold at p-value < 0.05 and a FDR < 0.2. Replicate 1 of H3K27ac in CORT_15w was discarded for the differential analysis. The lists of genes overlapping with differential peaks were obtained using `matchpeaksgenes` function from SeqCode [203], using the option –u 2500, which selects genes that overlap with a peak within +2,500 upstream the transcription start site and the transcription end site from the RefSeq catalogue [149].

RNA-seq analysis

mESC, cardiac differentiation, neural differentiation and mouse embryonic tissues

The pair-end sequence reads of RNA-seq data mESCs, MES, CPs, CMs, NPCs, CNs, Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5 and Lung15.5 were mapped to the mm10 version of the mouse genome with TopHat [205], setting the options --mate-inner-dist 100, which is the expected mean distance

between mate pairs, and -g 1, which eliminates those multilocus reads which align in more than one region. The RNA-seq profiles were obtained using the function `buildChIPprofile` from SeqCode [203]. The FPKMs (fragments per kilobase of transcript per million mapped reads) of each gene in the RefSeq catalogue [149] of the mouse genome were calculated using Cufflinks [206], setting the option --max-bundle-frags 5,000,000, which specifies the maximum genomic length for the bundles.

Visceral adipose tissue

The pair-end sequence reads of RNA-seq data from mice were mapped to the mm9 version of the mouse genome, and the ones from human were mapped to the hg19 version of the human genome with TopHat [205] setting the options --mate-inner-dist 100, which is the expected mean distance between mate pairs, and -g 1, which eliminates those multilocus reads which align in more than one region. The RNA-seq profiles were obtained using the function `buildChIPprofile` from SeqCode [203]. The FPKMs of each gene in the RefSeq catalogue [149] of the mouse genome were calculated using DESeq2 [207].

Differential analysis of gene expression

DESeq2 [207] was used to identify differentially expressed genes in all the comparisons (CORT_5w vs. VEH_5w, CORT_15w vs. VEH_15w and CS vs. CTL). The threshold was set at p-adjusted < 0.1.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) of the pre-ranked lists of genes by DESeq2 stat value was performed with the GSEA software [208].

Hi-C analysis

Hi-C data from mESCs were processed with TADbit [209]. Briefly, sequencing reads were mapped to the reference genome (mm10) by applying a fragment-based strategy, which is dependent on the GEM mapper [44]. Mapped reads were filtered to remove those resulting from unspecified ligations, errors, or experimental artefacts. Specifically, seven different filters were applied using the default parameters in TADbit: self-circles, dangling ends, errors, extra dangling-ends, over-represented, duplicated, and random breaks [209]. After

pooling replicates, Hi-C data were normalized with OneD correction [210] at 5 kb of resolution to remove known biases. Significant Hi-C interactions were called with the `analyzeHiC` function of HOMER software suit [47], binned at 5 kb of resolution, and with the default *p*-value threshold of 0.001.

By Francesca Mugianesi (our lab).

ATAC-seq analysis

The sequence reads of ATAC-seq data from mESCs were mapped to the mm10 version of the mouse genome with the BOWTIE software [41], setting the option –m 1, which eliminates multilocus reads that align in more than one region, and the option -X 2000, which defines the maximum insert size for paired-end alignment. Mitochondrial reads were removed from the resulting map.

Gene expression predictive model

The regression linear models were built to predict gene expression by adjusting the following formula:

$$y_i \sim \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} + \epsilon,$$

where y_i is the $\log_2$ of the FPKMs of gene *i*, with a pseudo count of 0.1. x_{i1} to x_{in} are the $\log_2$-normalized count of reads of each ChIP-seq signal at the defined promoters or enhancers averaged by the length of the region calculated by `recoverChIPlevels` from SeqCode [203], plus a pseudo count of 0.1. β_0 to β_n are the coefficients that we would like to calculate and ε is the error. The predictive models were trained on protein-coding genes. The set of data was randomly divided into two subsets, a training subset with the 80% of entries, and a test subset with the remaining 20% of entries. In the case of differentiation, each of the time points was used as training subsets and then the predictive models were evaluated in the rest. A 10-fold cross-validation was repeated three times to verify that the quantitative relationship between expression and histone modifications was not specific for a subset of the data. The following functions were used: `trainControl` to perform the 10-fold cross-validation, train to train the models, and `varImp` to calculate the variable importance, from the R package `caret` [211]. For models trained on enhancers, genes were introduced into the dataset as many times as the number of associated enhancers they had.

LOESS normalization

The FPKMs of all protein-coding genes and ChIP-seq levels of PEs and BPs were normalized for mESCs, MES, CPs, CMs,

NPCs and CNs; and also, for Heart10.5, Liver11.5, NeuralTube12.5, Kidney14.5, and Lung15.5. To normalize the ChIP-seq levels of H3K27me3, H3K4me3, H3K27ac, H3K4me1, and H3K36me3 on the PEs and BPs, the genome was first divided into 2 Kb bins (note that bin size reflects the average size of PEs and BPs). Next, the count of reads, normalized by total number of reads and averaged by the length, was calculated with `recoverChIPlevels` function from SeqCode [203]. Finally, the normalization parameters were calculated in those bins and applied to the count of reads normalized by the total number of reads and averaged by the length of PEs and BPs. The `normalize.loess` function of the R package `affy` [212] was used to normalize ChIP-seq data and expression data. Genes and bins with a 0 in any columns were discarded, as it was not possible to determine whether it was due to a sequencing error or a real absence of signal.

Conservation

Average conservation ranging between 0 and 1 was calculated with the UCSC genome browser application `bigWigAverageOverBed` [165]. As input we used phastCons60way, which measures the evolutionary block conservation among 59 vertebrates, in bigWig format

downloaded from UCSC genome browser [165]. We used mean0 as values of average conservation.

Overlap with CpG islands

The percentage of regions overlapping in at least 1 bp with a CpG island was calculated by `matchpeaks` function from SeqCode [203]. The coordinates for CpG islands were obtained from UCSC genome browser [165].

Statistical tests

Statistical tests (Exact Binomial Test; Student's t-test; Fisher's Exact Test; and Wilcoxon Test) were performed with R [213].

Plots

Plots (Violin plots, Boxplots, Scatterplots, etc.) were obtained with R package ggplot2 [214]. Color palettes are from R package viridis [215]. P-values were automatically plotted with R package ggpubr [216]. UpSetPlots were obtained with R package UpSetR [217]. Pie charts of genome distribution and

Venn Diagrams were obtained with SeqCode [203]. Genome screen shots were taken from UCSC genome browser [165].

Scripts

All the scripts have been written in Python and/or R, and run in an operative system MAC OS X Mojave, processor 2,8 GHz Intel Core i5, and 16 GB of RAM. Scripts to model gene expression and histone modification signals, and for LOESS normalization were fully written in R [213]. Scripts to bin the genome, to associate enhancers to promoters and genes by Hi-C interactions, to associate enhancers to promoters and genes by distance in the linear genome, and to unify peak coordinates were fully written in Python. The script to calculate state transitions was written in Python and internally calls R to plot the heatmap.

PCA webserver

The PCA webserver is part of the SeqCode web site (http://ldicrocelab.crg.eu/). It has been implemented in PHP, and uses R package factoextra [218] to plot the PCA. The output is provided in PNG and PDF formats.
Collaboration with Enrique Blanco (our lab).

Data availability

ChIP-seq data on histone modifications in mESCs can be found in GEO with the accession code GSE150633. RNA-seq data and ChIP-seq data on histone modifications in mouse VAT is available in GEO under the accession code GSE153934. RNA-seq data and ChIP-seq data on histone modifications in human VAT is available in GEO with the accession code GSE140126.

REFERENCES

1. Kornberg, R.D. and J.O. Thomas, *Chromatin structure; oligomers of the histones.* Science, 1974. **184**(4139): p. 865-8.
2. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.
3. Thomas, J.O. and R.D. Kornberg, *An octamer of histones in chromatin and free in solution.* Proc Natl Acad Sci U S A, 1975. **72**(7): p. 2626-30.
4. Luger, K., et al., *The atomic structure of the nucleosome core particle.* J Biomol Struct Dyn, 2000. **17 Suppl 1**: p. 185-8.
5. Allshire, R.C. and H.D. Madhani, *Ten principles of heterochromatin formation and function.* Nat Rev Mol Cell Biol, 2018. **19**(4): p. 229-244.
6. Yadav, T., J.P. Quivy, and G. Almouzni, *Chromatin plasticity: A versatile landscape that underlies cell fate and identity.* Science, 2018. **361**(6409): p. 1332-1336.
7. Hauer, M.H. and S.M. Gasser, *Chromatin and nucleosome dynamics in DNA damage and repair.* Genes Dev, 2017. **31**(22): p. 2204-2221.
8. Lai, W.K.M. and B.F. Pugh, *Understanding nucleosome dynamics and their links to gene expression and DNA replication.* Nat Rev Mol Cell Biol, 2017. **18**(9): p. 548-562.
9. Waddington, C.H., *The epigenotype. 1942.* Int J Epidemiol, 2012. **41**(1): p. 10-3.

10. Cavalli, G. and E. Heard, *Advances in epigenetics link genetics to the environment and disease.* Nature, 2019. **571**(7766): p. 489-499.
11. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
12. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
13. Holoch, D. and D. Moazed, *RNA-mediated epigenetic regulation of gene expression.* Nat Rev Genet, 2015. **16**(2): p. 71-84.
14. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-11.
15. Bernstein, B.E., et al., *Methylation of histone H3 Lys 4 in coding regions of active genes.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8695-700.
16. Schneider, R., et al., *Histone H3 lysine 4 methylation patterns in higher eukaryotic genes.* Nat Cell Biol, 2004. **6**(1): p. 73-7.
17. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.
18. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional*

elongation by RNA polymerase II. Mol Cell Biol, 2003. **23**(12): p. 4207-18.

19. Bannister, A.J., et al., *Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes.* J Biol Chem, 2005. **280**(18): p. 17732-6.
20. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.
21. Miao, F. and R. Natarajan, *Mapping global histone methylation patterns in the coding regions of human genes.* Mol Cell Biol, 2005. **25**(11): p. 4650-61.
22. Karlic, R., et al., *Histone modification levels are predictive for gene expression.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2926-31.
23. Cheng, C., et al., *A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets.* Genome Biol, 2011. **12**(2): p. R15.
24. Cheng, C. and M. Gerstein, *Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells.* Nucleic Acids Res, 2012. **40**(2): p. 553-68.
25. Wang, C., et al., *Computational inference of mRNA stability from histone modification and transcriptome profiles.* Nucleic Acids Res, 2012. **40**(14): p. 6414-23.

26. Tippmann, S.C., et al., *Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels.* Mol Syst Biol, 2012. **8**: p. 593.
27. Dong, X., et al., *Modeling gene expression using chromatin features in various cellular contexts.* Genome Biol, 2012. **13**(9): p. R53.
28. Zhou, X., et al., *Epigenetic modifications are associated with inter-species gene expression variation in primates.* Genome Biol, 2014. **15**(12): p. 547.
29. Budden, D.M., D.G. Hurley, and E.J. Crampin, *Predictive modelling of gene expression from transcriptional regulatory elements.* Brief Bioinform, 2015. **16**(4): p. 616-28.
30. Budden, D.M. and E.J. Crampin, *Distributed gene expression modelling for exploring variability in epigenetic function.* BMC Bioinformatics, 2016. **17**(1): p. 446.
31. Singh, R., et al., *DeepChrome: deep-learning for predicting gene expression from histone modifications.* Bioinformatics, 2016. **32**(17): p. i639-i648.
32. Read, D.F., et al., *Predicting gene expression in the human malaria parasite Plasmodium falciparum using histone modification, nucleosome positioning, and 3D localization features.* PLoS Comput Biol, 2019. **15**(9): p. e1007329.
33. Gilmour, D.S. and J.T. Lis, *Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on*

specific bacterial genes. Proc Natl Acad Sci U S A, 1984. **81**(14): p. 4275-9.

34. Solomon, M.J., P.L. Larsen, and A. Varshavsky, *Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene.* Cell, 1988. **53**(6): p. 937-47.
35. Johnson, D.S., et al., *Genome-wide mapping of in vivo protein-DNA interactions.* Science, 2007. **316**(5830): p. 1497-502.
36. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
37. Robertson, G., et al., *Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.* Nat Methods, 2007. **4**(8): p. 651-7.
38. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.
39. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nat Rev Genet, 2009. **10**(10): p. 669-80.
40. Lloyd, S.M. and X. Bao, *Pinpointing the Genomic Localizations of Chromatin-Associated Proteins: The Yesterday, Today, and Tomorrow of ChIP-seq*. Curr Protoc Cell Biol, 2019. **84**(1): p. e89.

41. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
42. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
43. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.
44. Marco-Sola, S., et al., *The GEM mapper: fast, accurate and versatile alignment by filtration.* Nat Methods, 2012. **9**(12): p. 1185-8.
45. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
46. Zang, C., et al., *A clustering approach for identification of enriched domains from histone modification ChIP-Seq data.* Bioinformatics, 2009. **25**(15): p. 1952-8.
47. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
48. Ross-Innes, C.S., et al., *Differential oestrogen receptor binding is associated with clinical outcome in breast cancer.* Nature, 2012. **481**(7381): p. 389-93.
49. Ernst, J. and M. Kellis, *ChromHMM: automating chromatin-state discovery and characterization.* Nat Methods, 2012. **9**(3): p. 215-6.

50. Rowley, M.J. and V.G. Corces, *Organizational principles of 3D genome architecture.* Nat Rev Genet, 2018. **19**(12): p. 789-800.
51. Denholtz, M., et al., *Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization.* Cell Stem Cell, 2013. **13**(5): p. 602-16.
52. Freire-Pritchett, P., et al., *Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells.* Elife, 2017. **6**.
53. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
54. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
55. Rowley, M.J., et al., *Evolutionarily Conserved Principles Predict 3D Chromatin Organization.* Mol Cell, 2017. **67**(5): p. 837-852 e7.
56. Rao, S.S.P., et al., *Cohesin Loss Eliminates All Loop Domains.* Cell, 2017. **171**(2): p. 305-320 e24.
57. Gasperini, M., J.M. Tome, and J. Shendure, *Towards a comprehensive catalogue of validated and target-linked human enhancers.* Nat Rev Genet, 2020.
58. Klemm, S.L., Z. Shipony, and W.J. Greenleaf, *Chromatin accessibility and the regulatory epigenome.* Nat Rev Genet, 2019. **20**(4): p. 207-220.

59. Banerji, J., S. Rusconi, and W. Schaffner, *Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.* Cell, 1981. **27**(2 Pt 1): p. 299-308.
60. Stalder, J., et al., *Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I.* Cell, 1980. **20**(2): p. 451-60.
61. Moreau, P., et al., *The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants.* Nucleic Acids Res, 1981. **9**(22): p. 6047-68.
62. Catarino, R.R. and A. Stark, *Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation.* Genes Dev, 2018. **32**(3-4): p. 202-223.
63. Gross, D.S. and W.T. Garrard, *Nuclease hypersensitive sites in chromatin.* Annu Rev Biochem, 1988. **57**: p. 159-97.
64. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
65. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

66. Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.
67. Tippens, N.D., A. Vihervaara, and J.T. Lis, *Enhancer transcription: what, where, when, and why?* Genes Dev, 2018. **32**(1): p. 1-3.
68. Andrey, G. and S. Mundlos, *The three-dimensional genome: regulating gene expression during pluripotency and development.* Development, 2017. **144**(20): p. 3646-3658.
69. Furlong, E.E.M. and M. Levine, *Developmental enhancers and chromosome topology.* Science, 2018. **361**(6409): p. 1341-1345.
70. Robson, M.I., A.R. Ringel, and S. Mundlos, *Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D.* Mol Cell, 2019. **74**(6): p. 1110-1122.
71. Essebier, A., et al., *Bioinformatics approaches to predict target genes from transcription factor binding data.* Methods, 2017. **131**: p. 111-119.
72. Hait, T.A., et al., *FOCS: a novel method for analyzing enhancer and gene activity patterns infers an extensive enhancer-promoter map.* Genome Biol, 2018. **19**(1): p. 56.
73. Fulco, C.P., et al., *Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations.* Nat Genet, 2019. **51**(12): p. 1664-1669.

74. Symmons, O., et al., *Functional and topological characteristics of mammalian regulatory domains.* Genome Res, 2014. **24**(3): p. 390-400.
75. Lupianez, D.G., et al., *Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions.* Cell, 2015. **161**(5): p. 1012-1025.
76. Franke, M., et al., *Formation of new chromatin domains determines pathogenicity of genomic duplications.* Nature, 2016. **538**(7624): p. 265-269.
77. Long, H.K., S.L. Prescott, and J. Wysocka, *Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.* Cell, 2016. **167**(5): p. 1170-1187.
78. Farley, E.K., et al., *Suboptimization of developmental enhancers.* Science, 2015. **350**(6258): p. 325-8.
79. Farley, E.K., K.M. Olson, and M.S. Levine, *Regulatory Principles Governing Tissue Specificity of Developmental Enhancers.* Cold Spring Harb Symp Quant Biol, 2015. **80**: p. 27-32.
80. Farley, E.K., et al., *Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers.* Proc Natl Acad Sci U S A, 2016. **113**(23): p. 6508-13.
81. Ouyang, Z., Q. Zhou, and W.H. Wong, *ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21521-6.

82. Duren, Z., et al., *Modeling gene regulation from paired expression and chromatin accessibility data.* Proc Natl Acad Sci U S A, 2017. **114**(25): p. E4914-E4923.
83. Schmidt, F., F. Kern, and M.H. Schulz, *Integrative prediction of gene expression with chromatin accessibility and conformation data.* Epigenetics Chromatin, 2020. **13**(1): p. 4.
84. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
85. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.
86. Young, R.A., *Control of the embryonic stem cell state.* Cell, 2011. **144**(6): p. 940-54.
87. Brook, F.A. and R.L. Gardner, *The origin and efficient derivation of embryonic stem cells in the mouse.* Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5709-12.
88. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
89. Azuara, V., et al., *Chromatin signatures of pluripotent cell lines.* Nat Cell Biol, 2006. **8**(5): p. 532-8.
90. Zhao, X.D., et al., *Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic*

compartments in human embryonic stem cells. Cell Stem Cell, 2007. **1**(3): p. 286-98.

91. Pan, G., et al., *Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.* Cell Stem Cell, 2007. **1**(3): p. 299-312.
92. Voigt, P., et al., *Asymmetrically modified nucleosomes.* Cell, 2012. **151**(1): p. 181-93.
93. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.
94. Blanco, E., et al., *The Bivalent Genome: Characterization, Structure, and Regulation.* Trends Genet, 2020. **36**(2): p. 118-131.
95. Ferrai, C., et al., *RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation.* Mol Syst Biol, 2017. **13**(10): p. 946.
96. Vieux-Rochas, M., et al., *Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain.* Proc Natl Acad Sci U S A, 2015. **112**(15): p. 4672-7.
97. Mas, G., et al., *Promoter bivalency favors an open chromatin architecture in embryonic stem cells.* Nat Genet, 2018. **50**(10): p. 1452-1462.
98. Roh, T.Y., et al., *The genomic landscape of histone modifications in human T cells.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15782-7.

99. Kinkley, S., et al., *reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells.* Nat Commun, 2016. **7**: p. 12514.
100. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
101. Weiner, A., et al., *Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution.* Nat Biotechnol, 2016. **34**(9): p. 953-61.
102. Sodersten, E., et al., *A comprehensive map coupling histone modifications with gene regulation in adult dopaminergic and serotonergic neurons.* Nat Commun, 2018. **9**(1): p. 1226.
103. Cui, K., et al., *Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.* Cell Stem Cell, 2009. **4**(1): p. 80-93.
104. Schuettengruber, B., et al., *Genome Regulation by Polycomb and Trithorax: 70 Years and Counting.* Cell, 2017. **171**(1): p. 34-57.
105. Denissov, S., et al., *Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant.* Development, 2014. **141**(3): p. 526-37.

106. Lewis, P.H., *New mutants report.* Drosoph. Inf. Serv., 1947. **21**: p. 69.
107. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-70.
108. Whitcomb, S.J., et al., *Polycomb Group proteins: an evolutionary perspective.* Trends Genet, 2007. **23**(10): p. 494-502.
109. Gil, J. and A. O'Loghlen, *PRC1 complex diversity: where is it taking us?* Trends Cell Biol, 2014. **24**(11): p. 632-41.
110. Wang, H., et al., *Role of histone H2A ubiquitination in Polycomb silencing.* Nature, 2004. **431**(7010): p. 873-8.
111. Chammas, P., I. Mocavini, and L. Di Croce, *Engaging chromatin: PRC2 structure meets function.* Br J Cancer, 2020. **122**(3): p. 315-328.
112. Czermin, B., et al., *Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites.* Cell, 2002. **111**(2): p. 185-96.
113. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.
114. Leeb, M. and A. Wutz, *Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells.* J Cell Biol, 2007. **178**(2): p. 219-29.

115. Ballare, C., et al., *Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity.* Nat Struct Mol Biol, 2012. **19**(12): p. 1257-65.
116. Morey, L., et al., *Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells.* Cell Stem Cell, 2012. **10**(1): p. 47-62.
117. Morey, L., et al., *Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms.* Cell Stem Cell, 2015. **17**(3): p. 300-15.
118. Beringer, M., et al., *EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells.* Mol Cell, 2016. **64**(4): p. 645-658.
119. Santanach, A., et al., *The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity.* Nat Commun, 2017. **8**(1): p. 1235.
120. Joshi, O., et al., *Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency.* Cell Stem Cell, 2015. **17**(6): p. 748-757.
121. Schoenfelder, S., et al., *Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome.* Nat Genet, 2015. **47**(10): p. 1179-1186.
122. Kundu, S., et al., *Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation.* Mol Cell, 2017. **65**(3): p. 432-446 e5.

123. Bonev, B., et al., *Multiscale 3D Genome Rewiring during Mouse Neural Development.* Cell, 2017. **171**(3): p. 557-572 e24.
124. Cruz-Molina, S., et al., *PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation.* Cell Stem Cell, 2017. **20**(5): p. 689-705 e9.
125. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.
126. Zentner, G.E., P.J. Tesar, and P.C. Scacheri, *Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.* Genome Res, 2011. **21**(8): p. 1273-83.
127. Rada-Iglesias, A., et al., *Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest.* Cell Stem Cell, 2012. **11**(5): p. 633-48.
128. Ngan, C.Y., et al., *Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development.* Nat Genet, 2020. **52**(3): p. 264-272.
129. Russ, B.E., et al., *Regulation of H3K4me3 at Transcriptional Enhancers Characterizes Acquisition of Virus-Specific CD8(+) T Cell-Lineage-Specific Function.* Cell Rep, 2017. **21**(12): p. 3624-3636.
130. Bonn, S., et al., *Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity*

during embryonic development. Nat Genet, 2012. **44**(2): p. 148-56.

131. Koenecke, N., et al., *Drosophila poised enhancers are generated during tissue patterning with the help of repression.* Genome Res, 2017. **27**(1): p. 64-74.
132. Bailey, M.H., et al., *Comprehensive Characterization of Cancer Driver Genes and Mutations.* Cell, 2018. **174**(4): p. 1034-1035.
133. Helin, K. and D. Dhanak, *Chromatin proteins and modifications as drug targets.* Nature, 2013. **502**(7472): p. 480-8.
134. Topper, M.J., et al., *The emerging role of epigenetic therapeutics in immuno-oncology.* Nat Rev Clin Oncol, 2020. **17**(2): p. 75-90.
135. Newell-Price, J., et al., *Cushing's syndrome.* Lancet, 2006. **367**(9522): p. 1605-17.
136. Lacroix, A., et al., *Cushing's syndrome.* Lancet, 2015. **386**(9996): p. 913-27.
137. Pivonello, R., et al., *Complications of Cushing's syndrome: state of the art.* Lancet Diabetes Endocrinol, 2016. **4**(7): p. 611-29.
138. van Staa, T.P., et al., *Use of oral corticosteroids in the United Kingdom.* QJM, 2000. **93**(2): p. 105-11.
139. Overman, R.A., J.Y. Yeh, and C.L. Deal, *Prevalence of oral glucocorticoid usage in the United States: a general population perspective.* Arthritis Care Res (Hoboken), 2013. **65**(2): p. 294-8.

140. Colao, A., et al., *Persistence of increased cardiovascular risk in patients with Cushing's disease after five years of successful cure.* J Clin Endocrinol Metab, 1999. **84**(8): p. 2664-72.
141. Barahona, M.J., et al., *Persistent body fat mass and inflammatory marker increases after long-term cure of Cushing's syndrome.* J Clin Endocrinol Metab, 2009. **94**(9): p. 3365-71.
142. Geer, E.B., et al., *Body composition and cardiovascular risk markers after remission of Cushing's disease: a prospective study using whole-body MRI.* J Clin Endocrinol Metab, 2012. **97**(5): p. 1702-11.
143. Wagenmakers, M., et al., *Persistent centripetal fat distribution and metabolic abnormalities in patients in long-term remission of Cushing's syndrome.* Clin Endocrinol (Oxf), 2015. **82**(2): p. 180-7.
144. Shah, N., et al., *Proinflammatory cytokines remain elevated despite long-term remission in Cushing's disease: a prospective study.* Clin Endocrinol (Oxf), 2017. **86**(1): p. 68-74.
145. Glad, C.A., et al., *Reduced DNA methylation and psychopathology following endogenous hypercortisolism - a genome-wide study.* Sci Rep, 2017. **7**: p. 44445.
146. Zannas, A.S. and G.P. Chrousos, *Epigenetic programming by stress and glucocorticoids along the*

human lifespan. Mol Psychiatry, 2017. **22**(5): p. 640-646.

147. Lee, H.A., et al., *Histone deacetylase inhibition ameliorates hypertension and hyperglycemia in a model of Cushing's syndrome.* Am J Physiol Endocrinol Metab, 2018. **314**(1): p. E39-E52.
148. Bartlett, A.A., H.E. Lapp, and R.G. Hunter, *Epigenetic Mechanisms of the Glucocorticoid Receptor.* Trends Endocrinol Metab, 2019. **30**(11): p. 807-818.
149. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
150. Pekowska, A., et al., *H3K4 tri-methylation provides an epigenetic signature of active enhancers.* EMBO J, 2011. **30**(20): p. 4198-210.
151. Koch, F. and J.C. Andrau, *Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity.* Transcription, 2011. **2**(6): p. 263-8.
152. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
153. Wamstad, J.A., et al., *Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage.* Cell, 2012. **151**(1): p. 206-20.
154. Dudoit, S.Y., Y. H.; Callow, M. J.; Speed, T. P., *Statistical methods for identifying differentially*

expressed genes in replicated cDNA microarray experiments. Stat Sin, 2002. **12**(1): p. 111-139.

155. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-93.
156. Hounkpe, B.W.C., F.; Lima, F.; de Paula, E. V., *HT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets* bioRxiv, 2019.
157. Consortium, E.P., et al., *Expanded encyclopaedias of DNA elements in the human and mouse genomes.* Nature, 2020. **583**(7818): p. 699-710.
158. Morey, L., et al., *RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells.* Cell Rep, 2013. **3**(1): p. 60-9.
159. Schnetz, M.P., et al., *CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression.* PLoS Genet, 2010. **6**(7): p. e1001023.
160. de Dieuleveult, M., et al., *Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells.* Nature, 2016. **530**(7588): p. 113-6.
161. Ong, C.T. and V.G. Corces, *CTCF: an architectural protein bridging genome topology and function.* Nat Rev Genet, 2014. **15**(4): p. 234-46.

162. Weintraub, A.S., et al., *YY1 Is a Structural Regulator of Enhancer-Promoter Loops.* Cell, 2017. **171**(7): p. 1573-1588 e28.
163. Brookes, E., et al., *Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs.* Cell Stem Cell, 2012. **10**(2): p. 157-70.
164. Acharya, D., et al., *KAT-Independent Gene Regulation by Tip60 Promotes ESC Self-Renewal but Not Pluripotency.* Cell Rep, 2017. **19**(4): p. 671-679.
165. Karolchik, D., A.S. Hinrichs, and W.J. Kent, *The UCSC Genome Browser.* Curr Protoc Bioinformatics, 2007. **Chapter 1**: p. Unit 1 4.
166. Pachano, T., et al., *Orphan CpG islands boost the regulatory activity of poised enhancers and dictate the responsiveness of their target genes.* bioRxiv, 2020.
167. Macfarlane, D.P., S. Forbes, and B.R. Walker, *Glucocorticoids and fatty acid metabolism in humans: fuelling fat redistribution in the metabolic syndrome.* J Endocrinol, 2008. **197**(2): p. 189-204.
168. Ratman, D., et al., *How glucocorticoid receptors modulate the activity of other transcription factors: a scope beyond tethering.* Mol Cell Endocrinol, 2013. **380**(1-2): p. 41-54.
169. Rebuffe-Scrive, M., et al., *Steroid hormone receptors in human adipose tissues.* J Clin Endocrinol Metab, 1990. **71**(5): p. 1215-9.

170. Ibrahim, M.M., *Subcutaneous and visceral adipose tissue: structural and functional differences.* Obes Rev, 2010. **11**(1): p. 11-8.
171. Karatsoreos, I.N., et al., *Endocrine and physiological changes in response to chronic corticosterone: a potential model of the metabolic syndrome in mouse.* Endocrinology, 2010. **151**(5): p. 2117-27.
172. Garcia-Eguren, G., et al., *Chronic hypercortisolism causes more persistent visceral adiposity than HFD-induced obesity.* J Endocrinol, 2019. **242**(2): p. 65-77.
173. Do, T.T.H., et al., *Glucocorticoid-induced insulin resistance is related to macrophage visceral adipose tissue infiltration.* J Steroid Biochem Mol Biol, 2019. **185**: p. 150-162.
174. Lee, I.T., et al., *Active Cushing Disease Is Characterized by Increased Adipose Tissue Macrophage Presence.* J Clin Endocrinol Metab, 2019. **104**(6): p. 2453-2461.
175. Garcia-Eguren, G., et al., *Long-term hypercortisolism induces lipogenesis promoting palmitic acid accumulation and inflammation in visceral adipose tissue compared with HFD-induced obesity.* Am J Physiol Endocrinol Metab, 2020. **318**(6): p. E995-E1003.
176. Hochberg, I., et al., *Gene expression changes in subcutaneous adipose tissue due to Cushing's disease.* J Mol Endocrinol, 2015. **55**(2): p. 81-94.

177. Ferrau, F. and M. Korbonits, *Metabolic comorbidities in Cushing's syndrome.* Eur J Endocrinol, 2015. **173**(4): p. M133-57.
178. Rijo-Ferreira, F. and J.S. Takahashi, *Genomics of circadian rhythms in health and disease.* Genome Med, 2019. **11**(1): p. 82.
179. Chang, T.H., et al., *An enhancer directs differential expression of the linked Mrf4 and Myf5 myogenic regulatory genes in the mouse.* Dev Biol, 2004. **269**(2): p. 595-608.
180. Link, N., et al., *A p53 enhancer region regulates target genes through chromatin conformations in cis and in trans.* Genes Dev, 2013. **27**(22): p. 2433-8.
181. Zhu, Y., et al., *Predicting enhancer transcription and activity from chromatin modifications.* Nucleic Acids Res, 2013. **41**(22): p. 10032-43.
182. Riising, E.M., et al., *Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide.* Mol Cell, 2014. **55**(3): p. 347-60.
183. Cai, C.L., et al., *Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart.* Dev Cell, 2003. **5**(6): p. 877-89.
184. Gao, R., et al., *Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate.* Cell Res, 2019. **29**(6): p. 486-501.

185. Ma, L., et al., *ISL1 loss-of-function mutation contributes to congenital heart defects.* Heart Vessels, 2019. **34**(4): p. 658-668.
186. Zhang, Q., et al., *Temporal requirements for ISL1 in sympathetic neuron proliferation, differentiation, and diversification.* Cell Death Dis, 2018. **9**(2): p. 247.
187. Jain, P., et al., *PHF19 mediated regulation of proliferation and invasiveness in prostate cancer cells.* Elife, 2020. **9**.
188. Sanchez-Molina, S., et al., *RING1B recruits EWSR1-FLI1 and cooperates in the remodeling of chromatin necessary for Ewing sarcoma tumorigenesis.* Sci Adv, 2020. **6**(43).
189. Henegar, C., et al., *Adipose tissue transcriptomic signature highlights the pathological relevance of extracellular matrix in human obesity.* Genome Biol, 2008. **9**(1): p. R14.
190. Zatterale, F., et al., *Chronic Adipose Tissue Inflammation Linking Obesity to Insulin Resistance and Type 2 Diabetes.* Front Physiol, 2019. **10**: p. 1607.
191. Yamamoto, T., et al., *Acute physical stress elevates mouse period1 mRNA expression in mouse peripheral tissues via a glucocorticoid-responsive element.* J Biol Chem, 2005. **280**(51): p. 42036-43.
192. Shimizu, N., et al., *Crosstalk between glucocorticoid receptor and nutritional sensor mTOR in skeletal muscle.* Cell Metab, 2011. **13**(2): p. 170-82.

193. Sasse, S.K., et al., *The glucocorticoid receptor and KLF15 regulate gene expression dynamics and integrate signals through feed-forward circuitry.* Mol Cell Biol, 2013. **33**(11): p. 2104-15.
194. Takeuchi, Y., et al., *KLF15 Enables Rapid Switching between Lipogenesis and Gluconeogenesis during Fasting.* Cell Rep, 2016. **16**(9): p. 2373-86.
195. Loboda, A., et al., *Diurnal variation of the human adipose transcriptome and the link to metabolic disease.* BMC Med Genomics, 2009. **2**: p. 7.
196. Takahashi, J.S., *Transcriptional architecture of the mammalian circadian clock.* Nat Rev Genet, 2017. **18**(3): p. 164-179.
197. Pacheco-Bernal, I., F. Becerril-Perez, and L. Aguilar-Arnal, *Circadian rhythms in the three-dimensional genome: implications of chromatin interactions for cyclic transcription.* Clin Epigenetics, 2019. **11**(1): p. 79.
198. Vollmers, C., et al., *Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome.* Cell Metab, 2012. **16**(6): p. 833-45.
199. Doi, M., J. Hirayama, and P. Sassone-Corsi, *Circadian regulator CLOCK is a histone acetyltransferase.* Cell, 2006. **125**(3): p. 497-508.
200. Nieman, L.K., et al., *The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice*

Guideline. J Clin Endocrinol Metab, 2008. **93**(5): p. 1526-40.

201. Aranda, G., et al., *Translational evidence of prothrombotic and inflammatory endothelial damage in Cushing syndrome after remission.* Clin Endocrinol (Oxf), 2018. **88**(3): p. 415-424.
202. Castellano-Castillo, D., et al., *Chromatin immunoprecipitation improvements for the processing of small frozen pieces of adipose tissue.* PLoS One, 2018. **13**(2): p. e0192314.
203. Blanco, E., M. Gonzalez-Ramirez, and L. Di Croce, *Productive visualization of high-throughput sequencing data using the SeqCode open portable platform* Submitted, 2020.
204. Orlando, D.A., et al., *Quantitative ChIP-Seq normalization reveals global modulation of the epigenome.* Cell Rep, 2014. **9**(3): p. 1163-70.
205. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
206. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.
207. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

208. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
209. Serra, F., et al., *Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors.* PLoS Comput Biol, 2017. **13**(7): p. e1005665.
210. Vidal, E., et al., *OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes.* Nucleic Acids Res, 2018. **46**(8): p. e49.
211. Kuhn, M., *caret: Classification and Regression Training.* R package version 6.0-86. https://CRAN.R-project.org/package=caret, 2020.
212. Gautier, L., et al., *affy--analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-15.
213. RCoreTeam, *R: A Language and Environment for Statistical Computing.* https://www.R-project.org/, 2020.
214. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. isbn:978-3-319-24277-4. https://ggplot2.tidyverse.org, 2016.
215. Garnier, S., *viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1.* https://CRAN.R-project.org/package=viridis, 2018.

216. Kassambara, A., *ggpubr: 'ggplot2' Based Publication Ready Plots.* R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr, 2020.
217. Gehlenborg, N., *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets.* R package version 1.4.0. https://CRAN.R-project.org/package=UpSetR, 2019.
218. Kassambara, A. and F. Mundt, *factoextra: Extract and Visualize the Results of Multivariate Data Analyses.* R package version 1.0.7. https://CRAN.R-project.org/package=factoextra, 2020.

PUBLICATIONS

Research articles resulting from this thesis:

González-Ramírez M., Ballaré C., Mugianesi F., Beringer M., Santanach A., Blanco E. and Di Croce L. Differential contribution to gene expression prediction of histone modifications at enhancers or promoters. In revision at Nucleic Acids Res.

García-Eguren G.*, **González-Ramírez M.***, Vizán P., Giró O., Vega-Beyhart A., Boswell L., Mora M., Halperin I., Carmona F., Gracia M., Squarcia M., Enseñat J., Vidal O., Di Croce L. and Hanzu F. A. Glucocorticoid-induced fingerprints on visceral adipose tissue transcriptome and epigenome. Submitted to JCI Insight.

Research article from a collaboration in the laboratory:

Blanco E., **González-Ramírez M.** and Di Croce L. Productive visualization of high-throughput sequencing data using the SeqCode open portable platform. Submitted to Genome Biol.

Review article:

Blanco E., **González-Ramírez M.**, Alcaine-Colet A., Aranda S. and Di Croce L. The Bivalent Genome: Characterization, Structure, and Regulation. Trends Genet, 2020. 36(2): p. 118-131.

ACKNOWLEDGEMENTS

Primer de tot, a tota aquella persona que considere que li he d'agrair alguna cosa, doncs moltes gràcies.

Muchas gracias a todas las personas que han pasado por el laboratorio, realmente habéis contribuido a tener muy buen ambiente. No os nombro porque seguro que me dejo a alguien y no sería justo, pero muchas gracias de verdad. En especial, a mis supervisores, Enrique y Luciano.

A mon pare i a ma mare: Paco i Carme. I a eixa família que no és família però quasi: Pep i Manuel. Gràcies pel suport i estar sempre al meu costat.

Thanks to the people that started the PhD at the same time as me, my dearest compactos! Especialmente a las guapas por fuera y por dentro: Silvia y Alejandra. Al meu company de pis Marcos. My favorite "guiri" Tobias, and my other favorite "guiri" Artyom. Thank you for the support, all the travels, dinners, parties, conversations, etc. that we have shared together. I hope we continue our friendship!

Moltes gràcies al gueto valencià que ens trobem al si de la Casa València a Barcelona. En especial, a qui va ser la meua porta d'entrada, la Societat Musical del País Valencià a Barcelona "La Valenciana". Al meu estimat Espai País Valencià. I també, al grup de danses "La de la Panxa Pelà".

Gràcies per ajudar-me a no perdre les meues arrels i a gaudir de la cultura popular del meu país.
Moltes gràcies a tots els meus amics i amigues, que tampoc els anomenaré a tots, ja que em deixaria a algú...

Per últim, a Pepe Toni. Ens vam conèixer en el moment en el que jo començava aquesta tesi, i poc a poc ens vam anar fent molt amics, fins a arribar al que tenim ara... Gràcies per no callar ni baix l'aigua, realment m'ha ajudat a desemboirar el cap en els moments més estressants d'aquesta tesi.

APPENDIX

Model	Predictors	Coeficients	P-value
Enhancer-Hi-C-all	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	1.769662;0.036313;-0.022462;0.034875;0.150573;0.82029 4;-0.154638;0.113001;0.192828;-1.626130;0.081335	<2e-16;0.0606;0.3041;0.2219;<2e-16;<2e-16;9.33e-09;<2e-16;<2e-16;<2e-16;2.22e-05
Promoter-Hi-C-all	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	3.42981;-0.35635;1.38463;0.36739;0.71092;-0.35287;0.08980;0.34151;-0.25338;-1.04633;0.30528	<2e-16;1.32e-06;<2e-16;2.91e-05;<2e-16;0.000271;0.368440;<2e-16;0.002894;<2e-16;5.90e-06
Enhancer-Hi-C-top	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	1.34375;0.02848;0.16148;-0.05578;0.07362;0.75660;-0.16639;0.08160;0.11158;-1.75951;0.20185	5.41e-08;0.509958;0.000496;0.354335;0.042874;<2e-16;0.003625;2.36e-08;0.015662;<2e-16;7.68e-07
Promoter-Hi-C-top	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	2.53925;-0.35923;1.16504;0.46248;0.74821;-0.61529;0.03518;0.34958;-0.09401;-1.11801;0.29481	1.26e-07;4.11e-05;<2e-16;3.52e-06;<2e-16;2.54e-08;0.756338;<2e-16;0.344272;<2e-16;0.000122
Enhancer-1Mb	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	1.777184;0.038401;-0.086740;0.045368;0.142935;0.801215;-0.097764;0.129381;0.199824;-1.527967;0.009363	3.73e-09;0.481741;3.66e-07;0.000407;3.68e-08;<2e-16
Enhancer-Hi-C-top_distal	Intercept;H4K20me3;H2Bub;H3K79me2;H3K36me3;H3K27me2;H3K27me1;H3K4me3;H3K4me1;H3K27me3;H3K27ac	1.37103;0.05715;0.1420;0.02684;0.11059;0.6968;-0.17495;0.09613;0.11919;-1.79132;0.16231	1.99e-06;0.246782;0.008710;0.709693;0.011708;<2e-16;0.009309;4.10e-08;0.025154;<2e-16;0.000622
PE-MES	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.87548;0.32947;-0.10706;0.77341;-0.42950;0.65131	<2e-16;0.0021;0.2954;<2e-16;8.59e-12;4.28e-07
PE-CP	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	4.49789;0.41363;-0.14266;1.05564;-0.41013;0.79118	<2e-16;0.00043;0.07172;<2e-16;1.70e-06;1.37e-09
PE-CM	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	4.20736;0.54595;-0.19799;1.16195;-0.77449;0.68414	<2e-16;1.59e-05;0.00837;<2e-16;<2e-16;5.66e-06
PE-NPC	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.11781;0.14783;0.06880;0.44423;-0.25199;0.55242	4.19e-13;0.0187;0.1436;4.10e-07;1.68e-12;2.32e-10
PE-CN	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.92670;0.45550;0.12358;0.65623;-0.36589;0.24901	<2e-16;6.38e-09;0.00888;9.79e-12;2.84e-16;0.00685
PE_distal-MES	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.8547;0.5077;-0.4345;0.3344;-0.3675;1.3149	5.92e-08;0.00613;0.01588;0.05648;0.00151;1.68e-08
PE_distal-CP	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.59968;0.52860;-0.06379;0.80747;-0.69685;0.72327	4.64e-06;0.02133;0.65513;4.39e-05;2.59e-05;0.00458
PE_distal-CM	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.0441;0.41129;-0.01988;0.99875;-0.99453;0.45472	0.000429;0.082664;0.880602;1.49e-07;8.17e-09;0.106193
PE_distal-NPC	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	1.32232;-0.01954;0.01013;0.04849;-0.16286;0.68722	0.0136;0.8752;0.9063;0.7657;0.0332;2.94e-05
PE_distal-CN	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.55576;0.29260;0.01389;0.37789;-0.12767;0.35556	9.34e-06;0.0727;0.8726;0.0266;0.1745;0.0502
BP-MES	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	6.18926;-0.27152;1.23363;1.63188;-0.74628;1.25739	<2e-16;0.0424;<2e-16;<2e-16;<2e-16;<2e-16
BP-CP	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	5.96830;0.07251;0.87269;1.35756;-0.89645;1.30821	<2e-16;0.627;<2e-16;2.36e-09;<2e-16;<2e-16
BP-CM	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	5.87637;0.45771;0.49104;1.17969;-1.21084;1.47382	4.39e-12;0.00313;2.31e-07;4.79e-06;<2e-16;<2e-16
BP-NPC	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.46702;0.19662;0.29683;0.34182;-0.17981;1.37008	5.51e-12;0.0199;7.76e-10;0.0455;8.39e-05;<2e-16
BP-CN	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.82808;0.07358;0.24210;0.74514;-0.30879;1.32765	3.73e-09;0.481741;3.66e-07;0.000407;3.68e-08;<2e-16
PE_intragenic-MES	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.50151;0.27126;-0.02209;1.01565;-0.30737;0.14076	2.84e-15;0.0388;0.8624;<2e-16;4.86e-05;0.4070
PE_intragenic-CP	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	4.36975;0.21512;-0.09455;1.1933;-0.27606;0.66160	<2e-16;0.151912;0.350868;<2e-16;0.011987;0.000544
PE_intragenic-CM	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	4.88009;0.46912;-0.22694;1.24425;-0.5939;0.87209	<2e-16;0.0042;0.0204;<2e-16;1.12e-07;2.86e-05

PE_intragenic-NPC	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	1.73654;0.22209;0.04564;0.50081;-0.26970;0.31045	8.77e-07;0.00741;0.46726;6.51e-07;1.37e-08;0.00782
PE_intragenic-CN	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.70760;0.37268;0.15745;0.66610;-0.37609;0.19218	1.34e-11;0.000253;0.015041;3.27e-09;4.63e-10;0.118102
PE_intergenic-MES	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	4.1239;0.4780;-0.2287;0.9326;-0.5689;1.0571	1.52e-06;0.01085;0.17213;0.00499;1.94e-07;3.17e-07
PE_intergenic-CP	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	6.2644;0.7649;-0.2105;1.6297;-0.5718;0.7494	1.09e-07;0.000108;0.096130;7.11e-05;3.82e-05;6.22e-05
PE_intergenic-CM	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.4950;0.5722;-0.1439;1.1886;-1.0423;0.4047	0.00292;0.00534;0.22221;0.00181;1.07e-13;0.07041
PE_intergenic-NPC	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	1.91952;0.07552;0.06692;0.16711;-0.21193;0.82063	0.004122;0.445558;0.351736;0.463575;0.000117;4.9e-10
PE_intergenic-CN	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.49007;0.56773;0.08360;0.40700;-0.33775;0.30427	0.00466;4.91e-06;0.22802;0.16672;6.98e-07;0.02903
PE-Heart10.5	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.18597;0.23328;-0.32014;0.80373;-0.80904;1.08654	8.90e-16;0.16;1.84e-05;6.22e-12;3.98e-15;4.21e-11
PE-Liver11.5	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.10344;0.57590;-0.09486;0.57254;-0.36270;0.48339	<2e-16;2.72e-05;0.197;2.57e-06;5.53e-08;1.02e-06
PE-NeuralTube12.5	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	2.43303;0.09565;-0.05656;-0.01470;-0.24669;0.62079	3.07e-15;0.340000;0.312097;0.880431;0.000461;<2e-16
PE-Kidney14.5	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.60808;0.28586;-0.09900;0.58582;-0.26159;0.72405	<2e-16;0.016351;0.058716;2.06e-05;0.000836;<2e-16
PE-Lung15.5	Intercept;H3K4me1;H3K4me3;H3K36me3;H3K27me3;H3K27ac	3.43695;0.11042;-0.18592;0.45259;0.13052;0.67831	4.87e-15;0.406519;0.000664;0.000636;0.059553;2.18e-14