PhD degree in Molecular Medicine

(Curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

Dissection of the mechanisms controlling high constitutive activity of housekeeping and tissue-specific cis-regulatory elements

Alessia Curina
(R10332)

Supervisor: Dr. Gioacchino Natoli

Co-Supervisor: Dr. Serena Ghisletti

European Institute of Oncology (IEO), Milan

Table of contents

Figure index

Table index

List of abbreviations

ATAC	Assay for Transposase-Accessible Chromatin
BMDM	Bone Marrow Derived Macrophages
BRE	TFIIB Recognition Element
ChIP	Chromatin Immuno Precipitation
ChIP-seq	ChIP-sequencing
CpGi	CpG island
Da	Dalton
DCE	Downstream Core Element
DPE	Downstream Promoter Element
DRE	TFIID Recognition Element
DTIE	Downstream Transcription Initiation Element
FBS	Fetal Bovine Serum
FE	Fold Enrichment
FT	Flow-Through
GTFs	General Transcription Factors
H3K4me1	Histone 3 Lysine 4 mono methylation
H3K4me3	Histone 3 Lysine 4 tri methylation
H3K27me3	Histone 3 Lysine 27 tri methylation
H3K27ac	Histone 3 Lysine 27 acetylation
HK	Housekeeping genes
IFNγ	Interferon gamma
INOS	Intrinsic Nucleosome Occupancy Score
Inr	Initiatior
LDTFs	Lineage Determined Transcription Factors
LPS	Lipopolysaccharide
MTE	Motif Ten Element
Mut	Mutated
NFR	Nucleosome Free Region
Nt	Nucleotide
PD	Pull-down
PIC	Pre Initiation Complex
PWM	Position Weight Matrix
qPCR	Quantitative polymerase chain reaction
RNA pol II	RNA-Polymerase II

SDTF	Signal Determined Transcription Factors
SDS	Sodium dodecyl sulfate
SDS-page	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TAFs	TBP-Associated Factors
TBP	TATA-Binding Protein
TCT	Polypyrimidine initiator motif
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TRFs	TBP-Related Factors
TSS	Transcription Start Site
XCPE1 and 2	X Core Promoter Element 1 and 2
WT	Wild type

1. Abstract

The genetic information is identical within the organism but the mechanisms by which different cell types achieve specialized functions interpreting the same set of instructions is not completely understood. It is now increasingly accepted that the combination of different genomic elements, both promoters and enhancers, favors the recruitment of different TFs, which in turn promotes the assembly of different pre-initiation complexes, guaranteeing heterogeneity in transcriptional outputs across different tissues. Nevertheless, the cis-regulatory elements and the transcriptional rules that control and maintain the expression of constitutively active genes are still poorly characterized. Specifically, whether the constitutive activity of promoters and enhancers relies on entirely distinct or instead shared regulators is unknown. By dissecting the *cis*-regulatory repertoire of macrophages, we found that the ELF subfamily of ETS proteins selectively bound within 60 bp from the transcription start sites of highly active housekeeping genes. ELFs also bound constitutively active, but not poised macrophage-specific enhancers and promoters. The role of ELFs in promoting constitutive transcription is suggested by multiple evidences: ELF sites enabled transcriptional activation by endogenous and minimal synthetic promoters; ELF recruitment was stabilized by the transcriptional machinery, and ELF proteins mediated recruitment of transcriptional and chromatin regulators to core promoters. These data indicate that a distinct subfamily of ETS proteins imparts high transcriptional activity to a broad range of housekeeping and tissue-specific *cis*-regulatory elements, which is consistent with the role of an ETS family ancestor in core promoter regulation in a lower eukaryote.

2. Introduction

2.1 Generality on the transcription process

A fundamental issue in biology is how cell-type-specific gene expression program is carried out by RNA polymerase II (RNA pol II) across different tissues.

Transcriptional control of gene expression is a highly coordinated process that is regulated in time and space by the interplay of sequence-specific transcription factors (TFs), the transcriptional machinery, co-activators and co-repressors (Kadonaga, 2004). Nucleosome occupancy and epigenetic modifications, such as histone modifications, DNA methylation and RNA interference are other players of gene transcription since they alter the chromatin accessibility and consequently the TFs binding. These types of modifications occur both on genes and on regulatory elements, therefore the control of gene expression is not circumscribed at gene levels but involves also distal regulatory regions.

Transcription of eukaryotic protein-coding genes begins with the binding of TFs to genomic regulatory elements, which in mammalian genomes are located right upstream and downstream of the transcription start site (TSS) both at gene promoters and at distal regulatory elements, specifically enhancers (Bulger and Groudine, 2011).

2.1.1 Assembly of the transcriptional machinery

Gene transcription is a well-controlled multi-step process that requires the pre-initiation complex (PIC) assembly to the core promoter, the region necessary to direct accurate transcription initiation (Sandelin et al., 2007). The PIC formation is triggered by the interaction of TFs and their cognate binding sites at proximal and distal regions and necessitates the presence of general transcription factors (GTFs): TFIIB, TFIID, TFIIE,

TFIIF, TFIIH and TFIIA (Reese, 2003). The multi-subunit TFIID is composed by TATA binding protein (TBP) and other 14 TBP-associated factors (TAFs), it interacts with TATA box, the initiator sequence and the downstream promoter element (DPE) (Hoey et al., 1990; Verrijzer et al., 1995). This is a largely simplified view that derives from initial mechanistic studies performed on TATA box promoter. In 1991 Pugh and Tjian proposed two distinct roles for TBP and TAFs: while the former is involved in the PIC assembly, the latter integrate regulatory signals (Pugh and Tjian, 1991). Nevertheless, evidences suggest that TAFs are not simply auxiliary proteins but conversely they confer stability to the TFIID complex, for example TAF4 rather than TBP stabilizes the complex. Indeed, Wright et al., reported that TAF4, TAF5, TAF6, TAF9 and TAF12 form the stable TFIID core sub-complex whereas TAF1, TAF2, TAF11, and TBP are peripheral subunits (Wright et al., 2006).

As anticipated before, the classical view of unique transcriptional machinery required to transcribe all genes was discredited by the existence of TATA-less promoters and by the discovery of tissue-specific TAFs and TBP-related factors, such as TRF1, which can recognize both TATA-containing and TATA-less promoters (Muller et al., 2007). This finding raises the possibility that the basal transcriptional machinery is not identical across different cell types and that different TFIID complexes can nucleate on distinct core promoters (Haberle and Lenhard, 2016). Therefore, it is likely that the stable TFIID core complex can interact with combination of other TAFs and TBP generating a large variety of structurally and functionally different PICs (Fig. 1) (Wright et al., 2006).

Despite the variability in TFIID complex composition, its mechanism of action is identical in different promoters recognizing core promoter elements. Indeed, TFIID causes DNA bending through a 90° angle and recruits the other GTFs. TFIID binding is stabilized by TFIIB that subsequently recruits RNA pol II, together with TFIIF (Sainsbury, et al., 2015). Alternatively to this classic stepwise PIC assembly, TFIID can directly interact with Mediator Complex favoring the recruitment of the others PIC components (Allen and

Taaties, 2015).

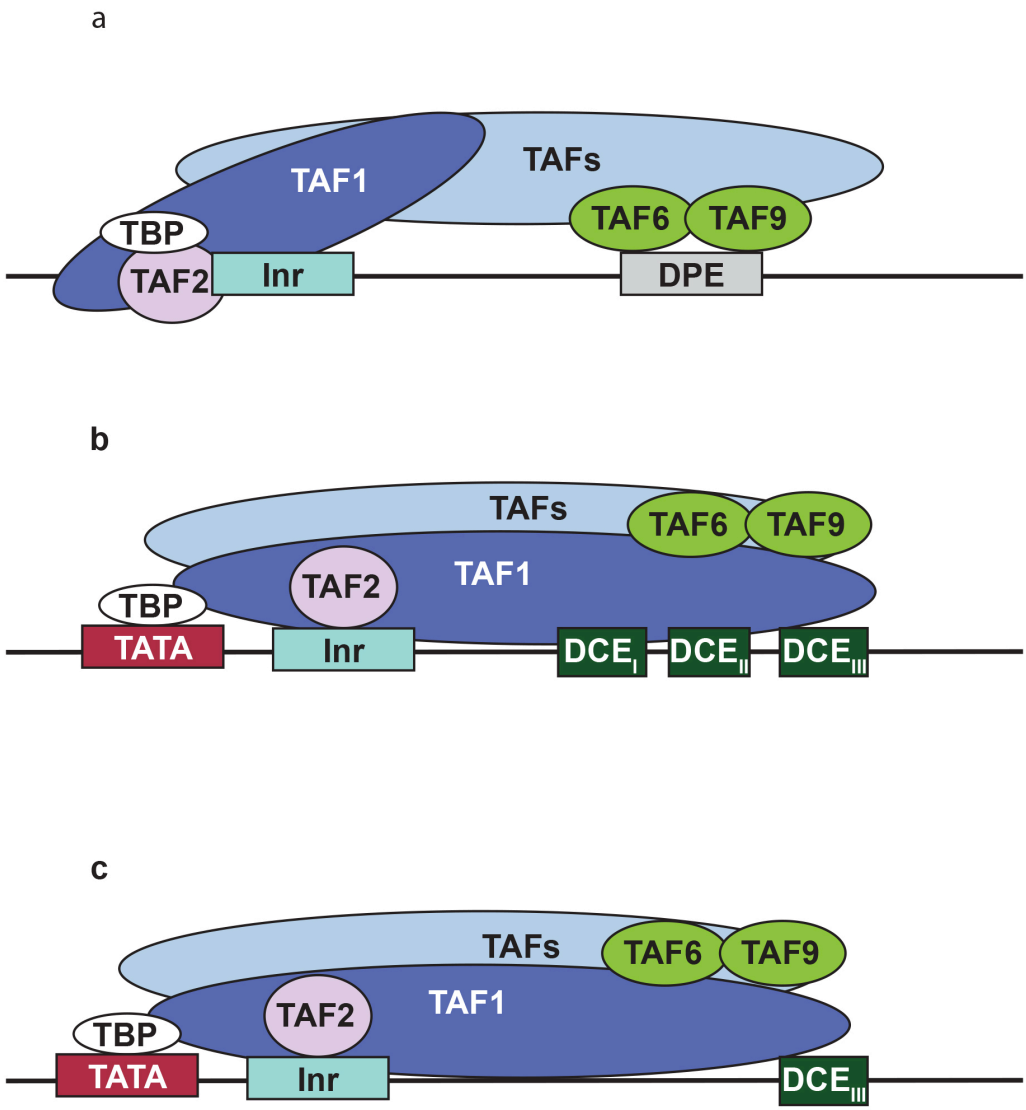


Figure 1. Schematic representation of TFIID recruitment on different promoters.
In all three examples, different TAFs are involved in TFIID positioning on different promoters.
(a) The TFIID complex recognizes the Inr and DPE sequences via TAF1/TAF2 and TAF6/TAF9 respectively. (b-c) TFIID complex is recruited through the interaction between TAF1 and DCE subelements (DCE_I, DCE_{II} and DCE_{III}). Adapted from (Lee et al., 2005)

TFIIB could assume different conformations depending on the presence or absence of TFIIB recognize elements (BRE) in the promoter. These conformational changes are responsible for the TFIIB ability to stabilize TBP to DNA (Reese, 2003).

The PIC formation concludes with the recruitment of TFIIE and TFIIH, which catalyzes ATP-dependent melting of the promoter at the TSS, generating an open region of about 15

base pairs, and allowing the progression into the elongation phase of transcription (Fazal et al., 2015). TFIIA is not a basal GTFs since it has the ability to potentiate the magnitude of the transcription aiding TBP binding (Hieb et al., 2007).

In *in vitro* transcription assay, the PIC complex is sufficient for basal transcription but it is not sufficient to increase or in general to vary transcription rates in response to TFs. Regulated transcription in fact requires molecular bridges between TFs and the basic transcriptional machinery, the most relevant one being Mediator (Malik and Roeder, 2010).

2.1.2 Core promoter elements

As already mentioned, the core promoter is the platform in which the transcription machinery is assembled. It can be constituted by one or more conserved DNA sequence elements, which are not universal and are not essential for promoter function. The mammalian promoters can be characterized by the combination of different motifs or by the lack of known core promoter elements (Juven-Gershon et al., 2008) and, once that the RNA pol II initiation complex is recruited to cis regulatory elements, transcription can initiate at a variety of different positions (Carninci et al., 2006; Frith et al., 2008; Kadonaga, 2012; Forrest et al., 2014). TATA box, initiatior (Inr), TFIIB recognition element (BRE), DPE and downstream core element (DCE) are the most well known elements that are shared by different protein-coding genes and that have positional constraints in relation to the TSS (Kadonaga, 2012). Conversely, motif ten elements (MTE), x core promoter element 1 and 2 (XCPE1 and XCPE2) are additional elements that occur at lower frequency (Fig. 2) (Danino et al., 2015).

TATA box. In metazoan the TATA box (Lifton et al., 1978), the first core promoter element identified, has a consensus of TATAWAAR that is constrained to a narrow window in which the 5' T is located 34-28bp upstream the +1 A (or +1 G) of the initiatior

sequence (Juven-Gershon et al., 2008). The TATA-TSS spacing is relevant to achieve transcription, generally the optimal position is within -31 and -30bp from TSS, whereas TATA motifs located nearly to the TSS (downstream -28bp) are non-functional (Ponjavic et al., 2006). TATA box is found out in less then 15% of metazoan promoters (Haberle and Lenhard, 2016), mainly upstream the TSS of tissue specific and inducible genes.

Initiation. The Inr sequence (Smale and Baltimore, 1989) is a highly degenerate sequence whose only essential component is the Pyrimidine / Purine dinucleotide at the -1 / +1 position, which represents the preferred context for transcription initiation by RNA Pol II in mammalian genomes (Carninci et al., 2006; Frith et al., 2008). The Inr sequence can independently direct accurate transcription initiation interacting with two subunits of the TFIID complex (TAF1 and TAF2) (Haberle and Lenhard, 2016). The pyrimidine-purine dinucleotide shows strong conservation and plays a key role for the transcription event (Carninci et al., 2006). Deletion of the dinucleotide of the human ankyrin 1 promoter causes the abolition of TFIID binding and the reduction of promoter activity (Gallagher et al., 2005).

BRE. Upstream and downstream the TATA box are located two BRE motifs (BREu and BREd) (Deng and Roberts, 2005; Lagrange et al., 1998), which are involved in the TFIIB recruitment (Kadonaga, 2012) and, depending on the context, can influence transcription both in positive or negative manner (Danino et al., 2015). TFIIB mutation causes a shift in the initiation of transcription suggesting that it could be involved in the RNA pol II positioning (Deng and Roberts, 2007).

DPE. DPE is another conserved sequence located +28,+30 bp downstream the TSS that functions cooperatively with the Inr sequence recognizing TAF6 and TAF9 (Burke and Kadonaga, 1997). Studies revealed that the distance between DPE and the Inr sequence is critical for the TFIID binding and basal transcription activity. TFIID binding, and consequently the transcriptional activity, is reduced by insertion of deletion of single nucleotide between the DPE and Inr sequence (Kutach and Kadonaga, 2000). It is possible

that DPE is involved in pausing the RNA pol II since its location and consensus sequence is similar to those of a Drosophila sequence, the "Pause Button", which is associated with stalled polymerase (Kadonaga, 2012).

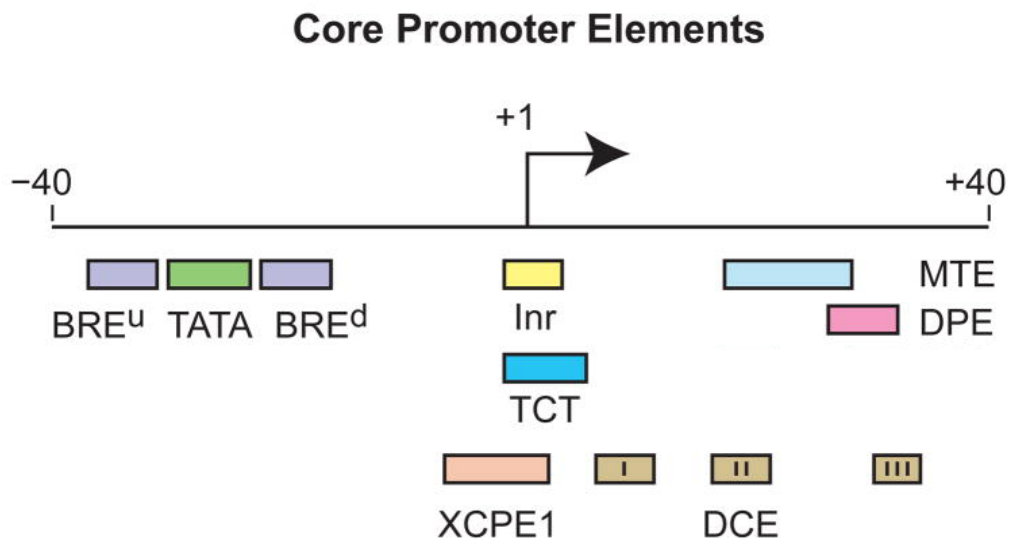


Figure 2. Schematic overview of known core promoter elements.
Depiction of some core promoter elements, which are found in only a fraction of all promoters (there are not universal core promoter motifs). Adapted from (Kadonaga, 2012).

Others core promoter elements. An additional downstream element that has a critical spacing requirement with the Inr sequence is the ***MTE*** motif, which is located +18 to +27 bp from the TSS (Danino et al., 2015).

The ***DCE*** element consists of three sub-elements located +6 to +11, +16 to +21 and +30 to +34 downstream to the TSS, it generally occurs with the TATA box (Juven-Gershon et al., 2008). Whereas DPE and MTE interact with TAF6 and TAF9, DCE has the peculiarity to recognize TAF1 subunit.

The ***XCPE1*** and ***XCPE2*** are two additional core promoter elements that are found in human genes with low frequency (approximately 1%), they are both localized around the TSS but they differ for the ability to drive transcription; whereas the XCPE2 is able to initiate gene expression without additional TFs, XCPE1 acts in conjunction with co-activators (Danino et al., 2015).

Another motif that encompasses the TSS is the polypyrimidine initiator motif (***TCT***), it has

a sequence similar to the Inr sequence but it is enriched and functional in the promoters of ribosomal protein genes and of proteins involved in the regulation of translation (Hariharan and Perry, 1990; Parry et al., 2010). However, even if the Inr sequence and TCT motif seem to be related, TCT cannot substitute the Inr sequence for basal transcription activity (Kadonaga, 2012) since it does not interact with the canonical TBP/TFIID but rather with the TBP-related factor 2 (TRF2) (Wang et al., 2014).

Recently, in the Dikstein lab a new downstream core element ***DTIE*** (Downstream Transcription Initiation Element) was identified in the miR-22 TATA-less and Inr-less promoter. DTIE is strictly localized, from +21 to +25 bp downstream the TSS, and it seems to control the precise TSS positioning and promoter strength cooperating with an upstream element DUCE, located between -111, -137 bp from the miR-22 TSS (Marbach-Bar et al., 2016). Therefore, this work suggests an unprecedented complexity of the transcription initiation, namely the TSS selection does not depend only on the core promoter elements but instead also distal regions can be involved in the positioning of the transcriptional machinery.

ETS site is located within 10-20 bp from the TSS

Studies unveiled that also the core promoter context is relevant for the transcription by specific activators. In mouse, for example, Elf-1 protein is able to activate selectively Inr-containing promoters with or without TATA box but not a TATA-only promoter (Ernst et al., 1996; Garraway et al., 1996). Similarly, Caudal factor in Drosophila activates preferentially DPE containing promoter but it does not activate TATA promoter with the upstream BRE motif (Juven-Gershon et al., 2008; Shir-Shapira et al., 2015). Moreover, Barbash and colleagues showed that core promoter elements have their unique but dispensable transcription activity. They introduced transgenes derived from the MHC class I gene carrying mutations in distinct core promoter elements into different mice. The promoter of MHC class I gene, ubiquitously expressed but subjected to tissue-specific

regulation, is characterized by four motifs (CAAT, TATAA-like, Inr and Sp1BS). Single mutation of each motif revealed that they are not necessary for the promoter activity *in vivo* but that each of them alters the wild-type patterns of gene expression in different tissues. However, it is not known if this is a characteristic of few genes or is a more widespread feature (Barbash et al., 2013).

Genome-wide analysis, performed examining the distribution of 8-mers respect the TSS on Drosophila and human core promoters, uncovered that some sequences are not randomly localized (FitzGerald et al., 2006). Fitzgerald et al. reported that 9 sequences are predominantly associated with human core promoters, raising the possibility that they could be involved in transcription. 7 of these sequences are known TF binding sites (CCAAT, SP1, USF, CREB, TATA, NRF-1 and ETS), the other two motifs are the KOZAK sequence, which is located at 3' of TSS, and an author-defined "CLUS1" sequence which doesn't match with already known TFBS. Additionally, they found out that the ETS motif is one of the most common non-directional site (namely the transcription can occur on both DNA strands) usually located within 10-20 bp from the TSS of housekeeping genes (HK) (FitzGerald et al., 2004). In contrast, the TATA box, which is stranded specific, is more represented in the promoters of tissue-specific and highly regulated genes. Supporting this finding there is also the evidence that TATA box never occurs in the ETS containing promoters, suggesting their involvement in totally different roles (FitzGerald et al., 2004). Therefore, the core promoter architecture varies among genes guaranteeing a huge variability in the protein-coding gene expression across tissues. Despite the huge variability in the core promoter composition, the recruitment of the transcriptional machinery is maintained and accomplished by TFIID. Conversely from what was thought in the past, namely that there was only one universal and highly conserved pre-initiation complex (Thomas and Chiang, 2006), nowadays it is accepted that different TFIID subunits recognize different core promoter elements. Additionally, TBP-related factors (TRFs) and cell-type specific TAFs enhance the variability in the

transcriptional machinery complexes recognizing different promoters. For example, in TRF2 knockdown mice the transcription of multiple post-meiotic and testes-specific genes is impaired causing defects in the spermatogenesis (Zhang et al., 2001; Martianov et al., 2001).

However, the mechanism of transcription initiation from TATA-less promoters or from promoters that does not contain any known core promoter elements remains to be determined. TF binding sites (TFBS) for Specificity protein 1 (Sp1) and Nuclear factor Y(NF-Y) have been found to be present in TATA-less and CpG island promoters (Pugh and Tjian, 1991; Yang et al., 2007). Moreover, few transcription factors have been shown to initiate transcription themselves, like the basic helix-loop-helix/leucine zipper (bHLH/ZIP) family of proteins and the ETS family members (Zhang et al., 2013).

2.1.3 Dissection of different promoter types

The variability of the core promoter influences two crucial aspects of the transcription event: the RNA pol II positioning, which in turn is critical for the transcription initiation (Kadonaga, 2012), and the assembly of different components of the transcriptional machinery on distinct promoters, which impacts on the interaction between enhancers and promoters and thus on gene regulation (Juven-Gershon et al., 2008). Different type of promoters can be identified considering PIC recruitment, function of associated gene, CpG content and nucleosome positions.

TSS pattern and PIC recruitment. Promoters can be characterized by focused or dispersed TSSs depending on the RNA pol II positioning. In a canonical focused promoter, TFIID interacts with TATA box, Inr sequence and DPE and recruits in a well-defined position the transcriptional machinery. In this scenario there is a single TSS, or a cluster of start sites in a short region of few nucleotides, approximately 30 bp downstream the TATA box. By contrast, dispersed promoters are characterized by multiple TSSs across a region

of up to 150 bp. Different scenarios have been hypothesized to explain dispersed promoters. It is likely that different PICs can bind to slightly different sites, and thus dispersed promoters are depicted by the absence of a strong core promoter in which the transcriptional machinery recruitment depends on the interaction between TFs and Mediator proteins. Therefore, since promoters have multiple TF binding sites, the RNA pol II can be recruited on different positions, and the transcription can start in different sites, depending on which TF interacts with Mediator Complex (Sandelin et al., 2007; Goodrich and Tjian, 2010).

Otherwise, it is possible that one PIC can slide to different sites to start separate initiation events, or that distinct TSS represent single TSS specific to each cell, that appear as dispersed in a cell population (Kawaji et al., 2006).

Function of associated gene. It is thought that focused promoters regulate the expression of tissue-specific and inducible genes which required a tightly and controlled regulation, whereas housekeeping genes, that are ubiquitously expressed and required a less stringent regulation, are characterized by dispersed promoters which guarantee a steady level of transcription across cell types. However, also developmental genes are associated with dispersed promoters (Carninci et al., 2006).

CpG content. In mammals dispersed promoters, relative to focused promoters, are characterized by CpG islands (CpGi), regions of around 1000 bp characterized by high C G content (Deaton and Bird, 2011). These promoters, enriched by CpG dinucleotides can be reversibly methylated, resulting in an alteration of the transcriptional state (Pugh and Venters, 2016). Nevertheless, genome-wide analysis of Drosophila promoters, which are CpG islands deficient, revealed that there is not a link between dispersed promoters and CpG island, whereas the presence of different core promoter elements can be used to discriminate focus from dispersed promoters. TATA-box, Inr, DPE and MTE, which have a defined position relative to the TSS, are commonly found in focused promoters (Haberle and Lenhard, 2016).

Nucleosome positioning. Focused and dispersed promoters are also distinguished by the nucleosome positioning (Bhatt et al., 2012; Ramirez-Carrozzi et al., 2009). Whereas dispersed promoters exhibit well positioned nucleosome, a higher degree of chromatin architecture, and therefore the requirement of nucleosome remodeling complex, focused promoters have a less stringent organization (Rach et al., 2011). In both cases promoters are characterized by the presence of a nucleosome free region (NFR), as suggested by the occurrence of DNase I hypersensitive site, which is flanked by two nucleosomes in a well-defined or less stringent position depending on the promoter types (Hesselberth et al., 2009; Rach et al., 2011; Lenhard et al., 2012).

Therefore, promoters can be classified in three different types depending on the TSS patterns, the nucleosome positions, and their histone modifications, and the function of the associated gene (Lenhard et al., 2012).

Type I promoters are focused promoters with a sharp TSS, they are devoid of CpG islands and have core promoter elements in a well-defined distance from the TSS (Roider et al., 2009). They are characterized by low-ordered nucleosomes and regulate tissue specific genes (Rach et al., 2011).

Type II promoters have multiple TSS dispersed in a NFR, which is flanked by two nucleosomes marked with H3K4me3, they are associated with weak core promoter elements and in mammals have a single CpG island (Lenhard et al., 2012).

Type III promoters, similarly to type II, are characterized by multiple TSS and well-positioned nucleosomes, but conversely from type II, have multiple CpG islands and nucleosomes are associated with repressive marks. The presence of polycomb group proteins on type III promoters is required to ensure a tight regulation of developmental genes (Hoskins et al., 2011).

Another layer of complexity in the promoter identification comes out with the discoveries that many genes can be transcribed with multiple (or alternative) promoters (Davuluri et al., 2008) and can use different TSSs in oocytes and in somatic cells (Haberle et al., 2014).

Regarding the first point, the transcription is carried out in a context, or tissue-specific, manner by more than one promoters generating complexity and diversity in the mammalian transcriptome (Sandelin et al., 2007). It is thought that different combinations of core promoter elements, and therefore the recruitment of distinct transcriptional machineries, could modulate the transcription of multiple promoters (Haberle and Lenhard, 2016). The usage of different TSSs in the oocytes or somatic cells of zebrafish developing embryos seems due to the usage of different core promoter elements. Maternal-specific TSS selection is driven by the oocyte-enriched transcription nucleating factor (Tbpl2) which recognizes the W-box motif, an A/T-rich region located 30 upstream the TSS. Conversely, the zygotic TSS selection is related to the stable position of the first nucleosome, which generates an area within which transcription can start (Haberle et al., 2014). All these evidences suggest that the variability of core promoter composition has a striking role in the regulatory mechanisms that govern expression. Therefore, transcription depends on both the combined activity of TFs, expressed in spatial and temporal manner, and on the core promoter elements.

2.2 Macrophage: a model to study gene regulation

Macrophages are essential components of innate immunity that play a key role in host defense mechanisms against invading pathogens, in responding to injury and in maintaining the homeostasis of the body (Qiao et al., 2013).

To quickly adapt to environmental changes caused by infections or tissue injuries, macrophages have the ability to alter their gene expression profile. Only few hours of bacterial lipopolysaccharide (LPS) stimulation are sufficient to generate a massive reorganization of chromatin and transcription, which results in the alteration of the expression of several hundred genes. Therefore, macrophages constitute a dynamic system that has been largely characterized.

Cell-type specific gene expression program is controlled in spatiotemporal manner by mutual interplay of promoters and regulatory elements, such as regulatory promoters and distal-promoter regions, which mediate RNA pol II regulation and recruitment of TFs (Zhou et al., 2011).

Different histone modifications can mediate gene-specific transcriptional regulation allowing TFs binding to their target site and changing DNA accessibility to the transcriptional machinery (Zaret and Carroll, 2011). Histone marks change the higher-order chromatin state, passing from an open and transcriptional active region (euchromatin) to an highly compact and transcriptional silent chromatin (heterochromatin) or vice versa. Therefore, since these two states depend on chromatin modifications, histone marks allow defining a priori if a gene is active or silent (Kouzarides, 2007; Baylin and Schuebel, 2007).

Studies of regulatory elements in macrophages and B-cells have indicated that genes activated by identical stimuli differ extensively among the two cell types, even though induction of these genes depends on TFs (like NF-kB and AP-1 family members, for inflammatory stimuli) that are ubiquitously expressed. This ‘‘context dependence’’ seems to depend mainly on the existence of a cell type specific repertoire of functional cis-regulatory elements, particularly enhancers (Heinz et al., 2015).

2.2.1 Transcriptional regulatory repertoire

The regulatory promoter, located upstream and downstream the core promoter (around -250 to +250bp from TSS), is a region bound by several TFs, which, together with other TFs bound to the distal regions, interpret the genetic regulatory information and coordinate the RNA pol II activity (Butler and Kadonaga, 2002).

Activators are proteins that interact with regulatory promoters increasing the rapidity of PIC formation or, generally, by promoting the elongation and the reinitiation. The activator

functions are modulated by coactivators, group of factors recruited throughout protein-protein interaction, which regulate positively or negatively activators and modify the chromatin (Maston et al., 2006). Among the coregulators, there are enzymes involved in histone tail modifications, such as the histone acetyltransferases p300/CREB-binding protein (CBP), which seem to be necessary for the opening of chromatin and consequently for PIC recruitment (Magnani et al., 2011).

Enhancers can be localized several hundred kilobase pairs upstream or downstream of a promoter, inside an intron or even beyond the 3' end of the gene eliciting their functions in distance and orientation independent manner respect to the promoter (Maston et al., 2006; Bulger and Groudine, 2011). Enhancers, like regulatory promoters, contain several TFs binding sites, which allow the simultaneous binding of TFs that, in turn, leads to the recruitment of coactivators and generate a cell type specific proteins complex. This higher order multicomponent complex is responsible for the RNA pol II recruitment and for the modification of the histone tails in nucleosomes (Panne, 2008).

Depending on the type of modification harbored by enhancer-associated nucleosome, it is possible to classify enhancers as constitutive, poised, latent or repressed (Fig. 3). For example in macrophages all classes are characterized by both H3K4 monomethylation (H3K4me1) and PU.1 binding, the macrophages master regulator, except for the latent enhancers that, prior stimulation, lack histone marks and are unbound by TFs. However, after lipopolysaccharide (LPS) stimulation, latent enhancers display H3K4me1, H3K27ac and PU.1 binding (Ostuni et al., 2013). Constitutively and repressed enhancers are additionally characterized by the presence of H3K27 acetylation (H3K27ac). The difference between these two classes resides in their behavior after LPS stimulation, H3K27ac in constitutive enhancers can increase (constitutive activated) or can be unaffected (constitutive non-activated), whereas in repressed enhancer can be lost or strongly reduced. Regarding poised enhancers, they are marked by H3K4me1 and PU.1 binding and are devoid of H3K27ac, thus they are primed but not active and require the

presence of additional signals to be fully activated. Nevertheless, after stimulation the majority of them are unaffected (poised not-activated) whereas, a small subset can be marked by H3K27ac (poised activated) (Ostuni et al., 2013). Furthermore, poised enhancers can be associated also with the chromatin repressive mark (H3K27me3) (Heinz et al., 2015). One of the most interesting enhancers features is cell type specificity. Studying the chromatin landscape, it has been noted that each cell type utilizes only a small fraction of all possible enhancers. It has been estimated that the total number of enhancers is around 1 million but the combinatorial activity of lineage determining TFs (LDTFs) selects 20000-40000 cell-type specific enhancers (Bulger and Groudine, 2011; Glass and Natoli, 2016). Therefore, each cell type has its own repertoire of cis-regulatory regions, which is critical for determining the cell's identity and the ability to respond to internal and external signals.

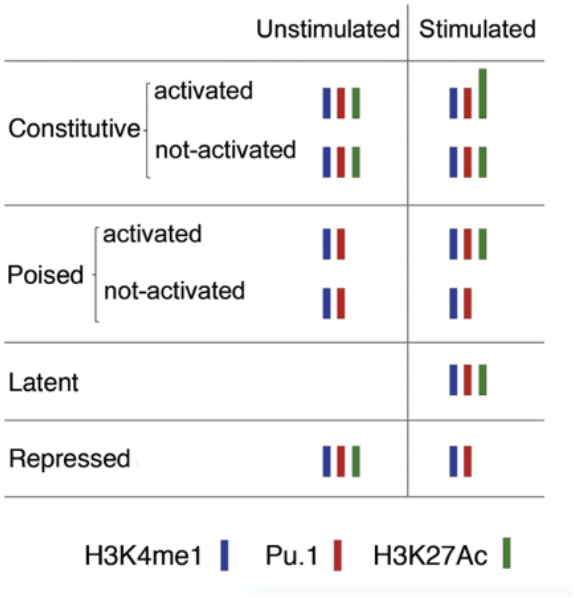


Figure 3. Classification of macrophages enhancers.
Classification of macrophage enhancers based on histone modification and PU.1 recruitment in unstimulated or LPS-stimulated macrophages. Adapted from (Ostuni et al., 2013).

This implicates that enhancers can shape the signal-depended gene expressions (Ghisletti et al., 2010; Heinz et al., 2015). Each cell type responds in a different way to an identical stimulus because the broadly expressed TFs, activated by the stimulus, bind the unique repertoire of enhancers resulting in a transcriptional response that is cell type specific

(Heinz et al., 2015). This occurs with two different mechanisms that can arise at the same time in the same cell type. In one case signal-dependent TFs (SDTFs) bind enhancers already selected by the LDTFs, therefore in absence of LDTFs, SDTFs is not able to recognize and to bind the regulatory regions. Otherwise, SDTFs can cooperate with LDTFs in the unveiling of latent enhancers (Fig. 4) (Kaikkonen et al., 2013; Ostuni et al., 2013). It seems that enhancers act as platform that receives information from multiple signalling cascades and than generates response acting on certain promoters (Maston et al., 2006). In the past, different mechanisms of enhancer action have been proposed: a "looping" model, the most reliable, and an alternative or "oozing" model. In the looping model enhancers and promoters interact through the DNA bending while in the alternative model a complex, that recognizes enhancers, flows through the DNA to research the promoter (Bulger and Groudine, 2011).

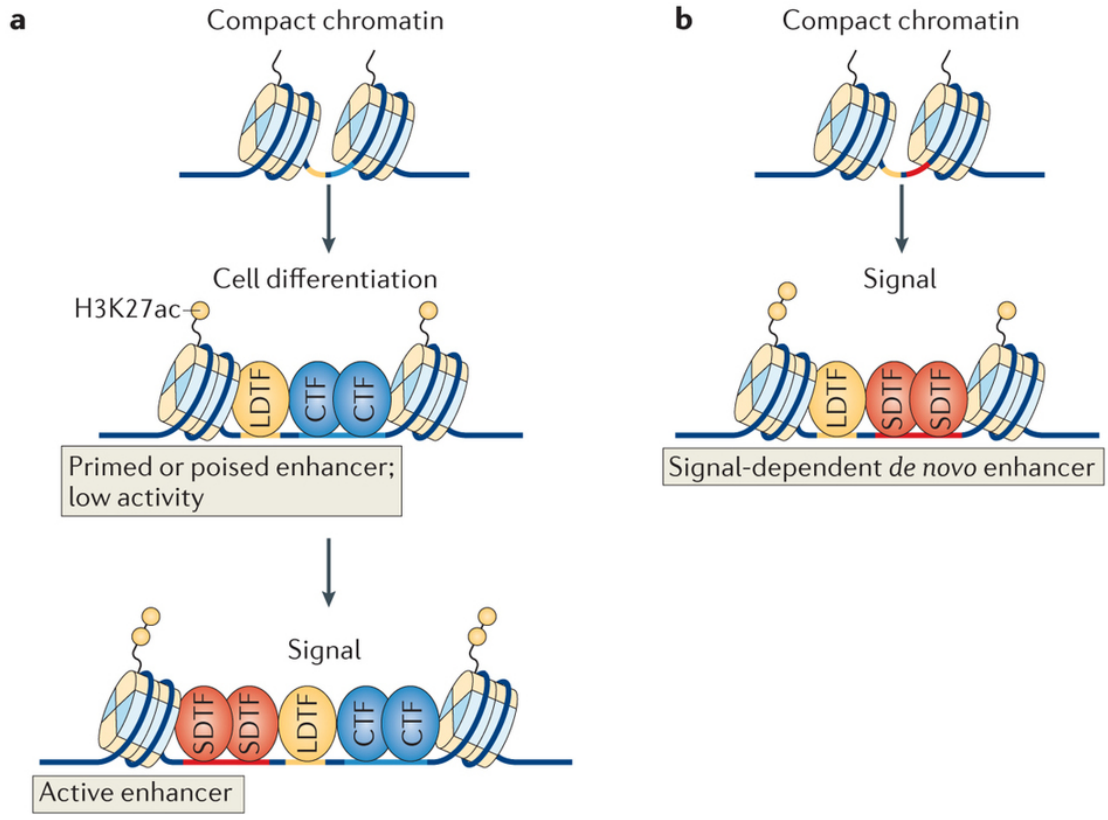


Figure 4. Cell type-specific enhancers' selection and activation.
(a) Enhancers' selection followed by the interaction of LDTFs and collaborating transcription factors (CTFs) during differentiation and enhancers activation accomplished by SDTFs. (b) Latent enhancers' selection mediated by the cooperation between LDTFs and SDTFs. (Heinz et al., 2015)

The chromosome conformation capture, and their high-throughput derivates, confirms the existence of the looping model in which the looping structure is stabilized by the transcription factors loaded both on promoters and enhancers and it is associated with paused RNA pol II prior to gene activation (Ghavi-Helm et al., 2014).

Chromosome conformation approaches detected correlation between enhancers and promoters, on average active promoters are associated with 4-5 distal elements and, vice versa, enhancers are found to contact approximately two promoters (van Arensbergen et al., 2014).

As anticipated before, the diversity in core promoter elements and therefore the recruitment of different components of the transcriptional machinery influences the interaction between enhancers and promoters. Zabidi et al. reported that in Drosophila enhancers can exhibit specificity for TATA versus DPE core promoter motif, resulting in distinct transcriptional properties, namely regulating tissue-specific or developmentally regulated genes (Zabidi et al., 2015). Therefore, the core promoter is not merely a platform on which the transcription machinery assembles. Rather, it has been proposed the existence of two subsets of enhancers that interact specifically with core promoter of housekeeping genes or developmental genes, suggesting the core promoter involvement in gene regulation. It has been proposed that this interaction could be due to biochemical compatibilities of proteins assembled both on promoters and enhancers (Zabidi et al., 2015) and to the chromatin architecture, such as the presence of topological associated domains (TADs) and the presence of insulator elements (van Arensbergen et al., 2014).

2.2.2 Transcription Factors recruitment

The commitment of regulatory regions in a cell-type depends on the chromatin status that can act as a barrier for most TFs. The TF recruitment relies on the presence of pioneer TFs and on the chromatin context such as the nucleosome position, the epigenetic modifications and the three dimensional chromatin organization. All these elements control the TFs binding in spatial and temporal manner.

Generally promoters, regardless for the cell types, are nucleosome free regions and thus the transcription factor binding is not prevented. On the contrary, the enhancer accessibility is highly variable and depends on the cell-type, for this reason only in certain cells a specific enhancer can be bound by TFs (Magnani et al., 2011).

TFs recognize a sequence of 6-12 bp, however, the binding specificity is dictated by only 4-6 nucleotides. Since TFs can bind DNA as homo- or heterodimers and the binding site of a single TF can occur in different orientation and spacing relative to other TFBS, the cooperative binding between different TFs increase the variability of DNA motifs that are recognized by single TFs. Recently, with the CAP-SELEX technique it has been discovered that TFs have different specificity for the sequence recognized when they act alone or together with other TFs. The cooperative binding decreases the stringency of the specificity meaning that proteins belonging to the same family (for example class I, II and III ETS proteins), which recognized different 5' flank sequences, could bind to the same DNA motifs thanks to the presence of the other partner (TEAD4) which mask the difference in binding specificity (Jolma et al., 2015).

Therefore, the cooperative binding between different TFs increases the number of regulatory regions that can be bound and, depending on which TFs is expressed in spatial and temporal manner, it is possible to control gene expression in different cell types (Rodda et al., 2005; Panne et al., 2007; De Val et al., 2008). However, as it is inferred by the existence of many potential target sites which are not bound by TFs, the merely

presence of the binding site is not a sufficient requisite for the TF binding. For instance, PU.1 can potentially bind, with different stringency, between 650000 and 1,4 milion site (Heinz et al., 2013; Barozzi et al., 2014), however, PU.1 ChIP-seq experiment showed that only around 45000 sites are bound in macrophages (Heinz et al., 2010). In addition to the nucleotide sequence and the cooperative DNA-binding of TFs, other features influence the TFs binding such as nucleosome occupancy and the DNA methylation (Slattery et al., 2014).

Several studies unveiled the existence of a dynamic equilibrium where sequences switch to be bound by nucleosomes or by TFs revealing that nucleosomes and TFs compete for the same DNA sequences (Mirny, 2010). According to this model, Segal group found out that p53 binds preferentially regions with high intrinsic nucleosome occupancy score (INOS); in the absence of p53 these regions are occupied by nucleosome but after p53 activation the nucleosome occupancy is reduced (Lidor Nili et al., 2010). Additionally He et al. reported that two TFs, which bind to DNA independently, are able to evict nucleosome if their binding site is within 150 bps from each other (He et al., 2013).

Barozzi et al., demonstrated that TFs recruitment and nucleosome depositions are controlled by overlapping DNA sequence and shape features. In macrophages, and in other PU.1 expressing cells, PU.1 binding occurs on sites that in PU.1-not expressing cells are occupied by nucleosomes. In contrast, PU.1 consensus sites that are not bound by PU.1 are devoid of nucleosomes. Accordingly to these results, PU.1 expression in PU.1-lacking cells is sufficient to generate nucleosome free regions (Ghisletti et al., 2010). Therefore, considering the CG content, the 3D shape and the binding of other TFs it is possible to predict at genome scale both PU.1 binding and nucleosome deposition (Barozzi et al., 2014). Distal and TSS-proximal regulatory elements bound by Pu.1 show different nucleosome occupancy due to the differences in their sequences. The abundance of cytosine and guanosine dinucleotide positively impacts on the nucleosome assembly, until a specific point beyond which the extremely high C-G content is not compatible with

efficient DNA bending around histones and thus it impedes the nucleosome deposition (Glass and Natoli 2016). This hints that the nucleosome depletion at TSS- proximal regions is not dependent on Pu.1 binding but rather on the high G+C content which interferes with nucleosome assembly (Barozzi et al., 2014). All these results suggest that the cell type-specific transcription depends on the interplay between nucleosomes and TFs binding.

As anticipated before, DNA methylation is an epigenetic modification that interferes with TFs binding. Indeed, this modification, as already mentioned, has been considered one of the causes of gene silencing which is accomplished through the recruitment of repressor proteins and the alteration of TFs accessibility to their recognition site (Nan et al., 1998). Generally, the variation in TFs recruitment concerns proteins that contain the CG di-nucleotide on their cognate site. For instance ETS, NF-kB, AP-2 and Myc proteins belong to the group of methylation-sensitive proteins since their binding is impaired after DNA methylation. However, it exists also methylation-insensitive proteins, such as Sp1, CTF and YY1, which are recruited despite the presence of methyl group in their cognate site. The existence of methylation-sensitive and insensitive proteins suggests that the lack of recruitment of the methyl-sensitive proteins does not depend on the mask of the binding site mediated by the methyl group but rather should rely on other mechanisms which selectively repress some genes and not others. This is further confirmed by recent evidence regarding the CTCF binding and the DNA methylation, which suggests that CTCF occupancy is impeded by the methyl group only on a subset of sites (Maurano et al., 2015). However, the mechanism by which proteins recognize their functional binding sites has not been completely understood and recent findings speculate that the TFs can interpret both the base and shape readout (Slattery et al., 2014).

2.3 ETS proteins

The ETS proteins are one of the largest families of winged helix-loop-helix DNA binding proteins that consist of approximately 30 members expressed in many different cell types. They are downstream effectors of many pathways and they control various cellular functions throughout development and adulthood. They have been reported to be involved in cellular proliferation, differentiation, hematopoiesis, apoptosis, metastasis, tissue remodeling, angiogenesis and transformation (Charlot et al., 2010).

The common feature of ETS proteins is the 85 amino acids, highly conserved DNA-binding domain, namely the ETS domain, which recognizes a conserved core GGAA/T motif. The binding specificity of each ETS protein depends on the nucleotides flanking the conserved bases (Gallant and Gilkeson, 2006).

Some ETS proteins have a reduced binding affinity due to the presence of two inhibitory regions, the inhibitory module and the serine rich region, that flank the DNA binding domain (Hollenhorst et al., 2011).

Genome-wide analysis showed both specific and redundant ETS proteins occupancy. Indeed different ETS factors can bind identical site, generally located on the proximal promoter of the housekeeping genes, in different cell-types suggesting a functional redundancy of some ETS proteins. Instead, specific binding sites are low affinity sites flanked by recognition sites for other TFs that in turn stabilize the ETS binding. ChIP-seq studies performed in human T-cell lines showed that ETS1, ELF1 and GABPa have more frequently redundant, than specific binding that is characterized by degenerate site located far from TSS of cell-specific genes (Hollenhorst et al., 2007). It is thought that redundancy could be a strategy to guarantee stable expression of housekeeping gene, independently on the expression level or the identity of ETS factors in different cell lines.

One intriguing question about the ETS family, regards the binding specificity. Even if around half of the ETS proteins are expressed in any cell types and display highly

conserved DNA binding, each ETS protein can accomplish different functions. It is thought that additional domains that flank the ETS domain, post-translational modification and the interaction with other proteins can confer the tissue specific role of ETS proteins (Hollenhorst et al., 2007).

High-throughput micro-well based DNA binding assay and protein-binding microarrays showed that the ETS-binding profiles cluster into four distinct classes (I, II, III and IV) and identified amino-acid residues that guarantee binding specificity (Wei et al., 2010).

In the high-throughput micro-well TF DNA-binding specificity assay the DNA binding domain of each ETS protein was cloned as fusion protein together with *Renilla* luciferase gene. Subsequently, the fusion protein was incubated with biotinylated oligonucleotide containing sequence with high affinity to all known ETS factors in the presence of non-biotinylated double-stranded competitor oligonucleotides containing every possible single base substitution to the consensus binding site. The biotinylated oligonucleotides were then captured on a streptavidin plate and are incubated with luciferase substrate. Reduction in luciferase activity revealed that the non-biotinylated sequences have a higher affinity for the TF-*Renilla* fusion protein compare to the biotinylated oligonucleotide (Hallikas and Taipale, 2006; Wei et al., 2010).

Whereas in the protein binding microarray, the binding specificity was measured on a microarray that contains all possible DNA sequence variants of a 10-mers sequence. Precisely, 44000 single-stranded features were attached to a microarray slide, each feature contains 26 distinct overlapping 10-mers. The binding specificity was measured acquiring the fluorescence intensity of the studied protein expressed with a glutathione S-transferase epitope tag (Berger et al., 2006).

Class I, the largest group, contains 15 members and includes ERG, ETV1, ETV4 and FLI1, the cancer-associated ETS factors. The other classes (II, III and IV) include 8, 3 and 1 member(s). The class II is further subdivided into two, IIa and IIb based on the binding specificity. This classification is due to the different set of amino acid residues which form

the DNA binding domain and, therefore, which interact with the nucleic acids.

A substitution of a single amino acid is sufficient to shift a factor from a class to another. For example, if a leucine of ELF4-DBD is mutated with phenylalanine or tyrosine, ELF4 can be considered a member of class I (Wei et al., 2010). The ETS proteins binding mode can be cooperative or not, as they can bind their respective site alone or within a ternary complex (Gallant and Gilkeson, 2006). The ETS proteins regulate, both in positive and in negative manner, the transcription through DNA binding. Several ETS factors can act either as activators and repressor depending on the type of promoter and they can be regulated by the interaction with other factors or by post-translational modifications that control their subcellular localization, DNA-binding activity or stability (Charlot et al., 2010). Loss-of-function studies showed that ETS proteins have specific functions in development. For example TFs that belong to class I, IIb and III are involved in different stages of hematopoiesis (which in part reflects their different expression patterns) (Wei et al., 2010).

Below it is reported a brief overview on the role of some ETS factors.

2.3.1 GABP

GABP (GA binding protein) has the peculiarity of being the only tetrameric factor, it is composed by two dimers of two subunits (α and β), the α subunit contains the DNA binding domain and the OST domain, which is implicated in co-factors recruitment, whereas the β subunit contains the transcriptional activation domain, the nuclear localization signal and it confers stability to the protein-DNA interaction (Rosmarin et al., 2004). The oxidation of cysteine residues contained in the dimerization domains and in the DNA binding domain of GABP α is critical for its transcriptional activity (Wu et al., 2013).

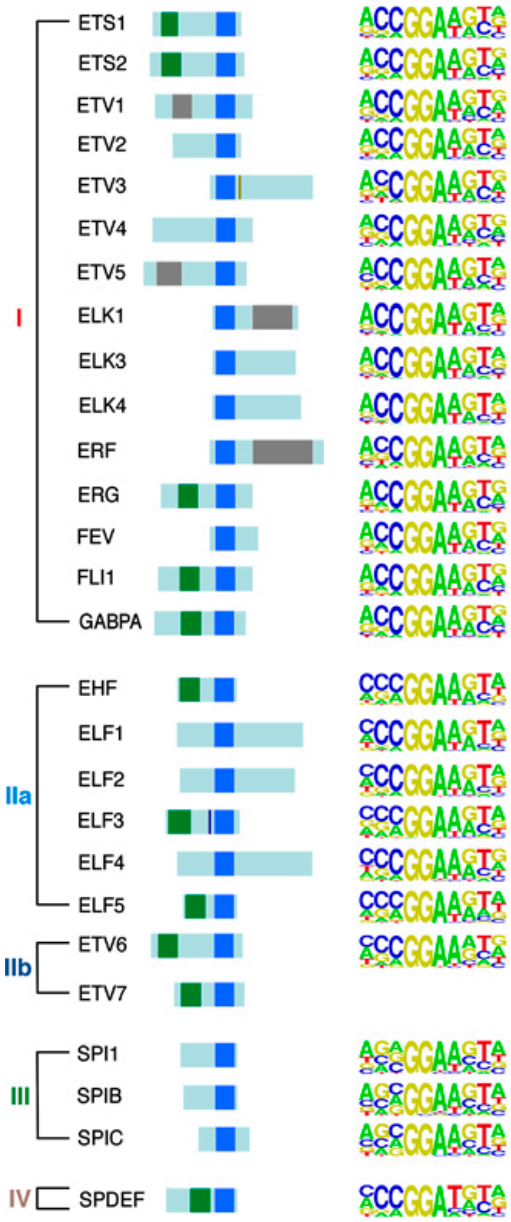


Figure 5. Structural organizations and binding specificities of mammalian ETS TFs. (Left) Schematic representation of the domain structures of the respective full-length proteins. ETS domain is in blue, pointed domain is in green, Proline-rich domain is in grey. The second column shows mouse ETS-binding profiles. Adapted from (Wei et al., 2010).

Regulating the expression of genes necessary for cell adhesion, clearance of pathogens and genes involved in the inflammatory response, GABP seems to be necessary for development and function of myeloid cells. GABP is also a transcriptional regulator of several cytokines and cytokine receptors involved in the growth and function of hematopoietic cells such as IL-2 and IL-4. Additionally, GABP is a key regulator of

essential housekeeping genes as suggest by its involvement in the regulation of ribosomal and mitochondrial genes and by its role in cell growth and antioxidant defense. Recently, it has been discovered that the GABP regulates the expression of Yap protein, which is critical for cell survival and cell-cycle progression (Rosmarin et al., 2004; Wu et al., 2013) Therefore, GABP is a key regulator of both lineage restricted and housekeeping genes and GABP α knock out mice show early embryonic lethality confirming GABP requirement for fundamental cellular process such as cellular differentiation, maturation and activation (Wu et al., 2013). Despite its role in regulating housekeeping genes, GABP is not ubiquitously expressed but it is abundant in hematopoietic cells, specifically myeloid cells and in a wide variety of cell types such as liver, muscle, testis and brain. Instead, the lineage-restricted gene regulation seems to be mediated by the interaction with other TFs, such as Sp1 and PU.1, and with co-factors such as p300 (Rosmarin et al., 2004).

2.3.2 FLI 1

FLI1 (Friend leukemia insertion 1) is initially identified as an oncogene, it controls vasculogenesis, cell cycle, apoptosis and has several roles in embryonic development.

It is expressed primarily in both endothelial and hematopoietic cells regulating the development of hematopoietic stem cells, megakaryocytes, myelomonocytes, erythrocytes and NK cells. FLI1 regulates the expression of vascular homeostasis genes and of several megakaryocyte-specific genes. FLI1 knockout mice, as GABP knockout mice, show embryonic lethality, they are characterized by absence of megakaryocytes and aberrant vasculogenesis, which implies FLI1 involvement in embryonic development (Hart et al, 2000). Aberrant expression of FLI1 is detected in hematopoietic and epithelial tumors and, in humans, the chromosomal translocation of the FLI1 DNA binding domain to the EWS locus on chromosome 22 leads to Ewings sarcoma. Depending on the context, Fli1 may act both as oncogene or tumor suppressor. In hematopoietic tumors, such as erythroleukemia,

FLI1 is highly expressed and it acts as oncogene regulating genes involved in proliferation, survival and differentiation. Related to this, human breast cancers are characterized by reduced level of FLI1 mRNA, the loss of FLI1 correlates with accelerate tumor growth and increase in the expression of proteins associated with cancer progression (i.e. Ki67) and in the number of lung metastases (Scheiber et al., 2014).

FLI1 is also involved in autoimmune disease and, likewise in tumor, both its overexpression and down-regulation can lead to diseases: transgenic mice overexpressing FLI1 show systemic lupus erythematosus-like sintoms, on the contrary the downregulation of FLI1 seems to increase the pathogenicity of the Systemic sclerosis or scleroderma,

2.3.3 ELF1

ELF1 (ETS like factor 1) exists in two-forms: 98 kDa in the nucleus and 80 kDa in the cytoplasm (Juang et al., 2002). This difference depends on post-translational modifications since ELF1 can be phosphorylated and glycosylated at the same time. These modifications are responsible for subcellular localization, protein-protein interactions and protein-DNA interactions. ELF1 is bound to retinoblastoma protein in the cytoplasm while, in the nucleus it interacts with target promoters (Charlot et al., 2010). ELF1 is highly expressed in a variety of endothelial cells, in both embryonic and adult lymphoid tissues and during thymocyte development its level remains similar in all subset of maturing thymocytes and T cells. Interestingly, ELF1 is present at high levels also in the epithelial cells of oral mucosa, in the central nervous system, in the lung, in the gastrointestinal tract, in urinary tracts and in the skin. Liver and testis of adult mouse present lower levels of ELF1 (Bassuk et al., 1998). Like most of the ETS factors, it is implicated in hematopoiesis, angiogenesis and in the development of NK and NKT cells (Choi et al., 2011). It activates the expression of many T cell genes such as the TCR chain while in B cells it cooperates with members of the AP1 family to activate the 3' IgH enhancer in response to stimulation with

IgM (probably it is involved in class-switch recombination) (Gallant and Gilkeson, 2006). ELF1 also allows the expression of the granulocyte-macrophage colony stimulating factor (GM-CSF) gene, the interleukin-2 receptor alpha subunit gene and the CD4 gene. ELF1 binding sites have been identified in the transcriptional regulatory elements of multiple genes involved in T cell, B cell and macrophage development and activation (Bassuk et al., 1998). Additionally, ELF1 seems to be a potential key regulator of expression of key hematopoietic genes, including Scl, Fli1, Lyl1, Runx1, Lmo2 and PU.1; it is interesting to note that for the last gene the activation is reciprocal, as PU.1 is a major regulator of ELF1. ELF1 downregulation is indispensable for the erythroid differentiation (Calero-Nieto et al., 2010).

2.3.4 ELF4

ELF4 (E74 Like ETS Transcription Factor 4) is an ETS factor closely related to ELF1. It is involved in cell cycle regulation and, together with ELF1, it is implicated in NK and NKT cell development. Studies showed its involvement, acting at promoter level, in the expression of IL-8, a chemotactic factor for neutrophils, basophils, eosinophils and, probably, for lymphocytes. It has also a tumor-suppressor role (Gallant and Gilkeson, 2006). This latter function is carried out through the activation of the tumor suppressor KLF4, which induces cell-cycle arrest in naïve $CD8^+$ T cells. Therefore, ELF4 is a negative regulator of proliferation of naive $CD8^+$ T cells, which do not respond to weak self-MHC signals. With regards to the role in NK cells, ELF4 prepares them to a rapid response after stimulation, activating perforin gene expression. Finally ELF4 is also involved in proliferation of B cells, ovarian epithelial cancer cell lines and HSCs (hematopoietic stem cells) (Yamada et al., 2009).

ELF2, ELF3 and ELF5 are other ELF1 paralogues; ELF2 is expressed at low levels in macrophages, whereas ELF3 expression is transiently induced by pro-inflammatory cytokines. Instead, ELF5 is not expressed in macrophages (Grall et al., 2003).

Even if ELF1 and ELF4 seem closely related, knockout experiments display different roles for the two proteins. While ELF1 is implicated only in the development and function of NKT cells (ELF1-/- mice produce less amount of cytokines upon antigene stimulation) (Choi et al., 2011), ELF4 knockout mice present defect both in NK and NK-T cell differentiation, resulting in a decreased amount of the two cell-types respect the WT mice. Conversely from what occurs in ELF1 null mice, the lack of ELF4 does not impair the cytokines secretion from NK-T cells whereas it impairs the secretion of IFNγ and the expression of perforin gene in NK cells (Lacorazza et al., 2002).

2.3.5 PU.1

PU.1 *(Purine-rich box 1)* is characterized by three domains: the trans-activating, the PEST and the DNA-binding domain. The former contains a proline-rich region that allows intramolecular bonds and interaction with other proteins. The second, called PEST for the presence of proline-, glutamic acid-, serine- and threonine-rich regions, is a target for peptidases allowing the rapid turnover of PU.1. This domain is also necessary for the conformational changes and the interaction with other proteins due to the presence of several serine-phosphorylation sites. The phosphorylation of serine 148 is required for the interaction of PU.1 with other proteins in B cells. This modification is also necessary for PU.1 roles in macrophages proliferation (phosphorylation of serines 41 and 45), as suggested by mutagenesis experiments (Celada et al., 1996). Finally, the DNA binding domain recognizes the sequence 5'-AG(A/C/G)GGAAG-3' (PU box). PU.1 is expressed at really high levels during early stages of hematopoietic development. The amount of PU.1 is critical to establish progenitor differentiation into B cells (low level of PU.1) or

monocytes (high level) (Gallant and Gilkeson, 2006). For normal development of megakaryocites, erythroid progenitors, B and T cell progenitors PU.1 levels are low, whereas to induce and then maintain macrophage differentiation PU.1 must be constantly expressed at high levels (Rosenbauer et al., 2006; Kueh et al., 2013). PU.1 deficient mice die of severe septicemia 48h after birth since they have a defective development of macrophages, T and B cells. Neutrophils also require the presence of PU.1 for the secondary granule components expression since PU.1 deficiency affects the ability to ingest and kill bacteria (Gallant and Gilkeson, 2006).

2.3.5.1 PU.1: the master myeloid regulator

Macrophages originate from circulating peripheral-blood mononuclear cells (PBMCs) that are released from bone marrow to blood where they circulate for several days before entering into the tissues and replenishing the tissue macrophage populations. The differentiation of lymphoid and myeloid lineages from the common precursor relies on the enhancers' selection (Romanoski et al., 2015). As already mentioned, the combinatorial activity of LDTFs, which are able to bind compacted chromatin, controls the enhancers' selection, namely they convert inactive enhancers in to primed, poised or active (Natoli, 2010; Calo and Wysocka, 2013)**.** PU.1 binds different regulatory regions in macrophages and in B cells interacting with LDTFs specific for the two cell lineages. In macrophages the cis-regulatory repertoire is defined by the interplay of PU.1 with CCAAT/enhancer-binding proteins (C/EBPs) and AP1, whereas in B cells PU.1 binding occurs in proximity (<100 bp) to the EBF1, E2A and OCT binding sites (Heinz et al., 2010). These dissimilarities in the enhancer repertoire justify also the capability of different cell types to respond in distinct way to an identical stimulus. Indeed, LPS activates, in both macrophages and B-cells, NF-kB, a broadly expressed SDTF, nevertheless, due to the cell-type-specific enhancers, NF-kB activation results in a different pattern of gene expression

(Glass and Natoli 2016).

The enhancer's selection, mainly mediated by PU.1, depends on the role of both internal and external signals and can explain the heterogeneity of macrophage populations (Romanoski et al., 2015). The Langerhans cells that populate the epidermis, osteoclasts, alveolar macrophages, microglia in the central nervous system, metallophilic macrophages that sample the circulation with unknown function, and the Kupffer cells in the liver are just some examples of macrophage populations which differs for their enhancer's repertoire (Gordon and Taylor, 2005). Therefore, the ability of macrophages to differentiate themselves in each tissue is due to the environmental inputs received from the surrounding cells and to the intrinsic differentiation pathways (Lawrence and Natoli, 2011). In different macrophage subtypes, as already seen for different cell types, PU.1 primes a series of potential enhancers but, depending on the stimuli received, only a part of them can be activated and consequently can trigger a gene program that is specific for each tissue (Fig. 6). For example, PU.1 primes a set of enhancers that can respond to retinoic acid or TGFβ and that, consequently, become active only in milieu such as peritoneal cavity or brain in which these two stimuli are respectively present (Romanoski et al., 2015). Additionally, two recent works revealed the role of environment in shaping the cis-regulatory regions independently of cell origin (Lavin et al., 2014; Romanoski et al., 2015). The macrophage plasticity does not reflect only their capability to differentiate in different phenotypes depending on the stimulus, but it also embraces the possibility of inter-conversion between different macrophage subtypes when cells are exposed to a new microenvironment. Similarly, tissue-specific macrophages, when transplanted to a different anatomic location, are subjected to the reprograming of the gene expression (Romanoski et al., 2015). This suggests that macrophage differentiation is not an irreversible process; conversely microenvironment is able to shape and to reprogram the chromatin landscape (Lavin et al., 2014). PU.1 might act as a pioneer factor since it promotes the deposition of H3K4me1 and generates open regions of DNA accessible for

other TFs. Indeed, ChIP-seq analysis revealed that the genomic distribution of PU.1 is widespread in macrophages and it correlates with almost all genomic enhancers marked by H3K4me1 (Ghisletti et al., 2010). Pioneer factors are able to bind DNA when other factors cannot, such as in the presence of condensed chromatin. They can open chromatin, modify nucleosomes position, enable intrinsic cooperative binding and recruit chromatin modifiers and co-regulators (Zaret and Carroll, 2011).

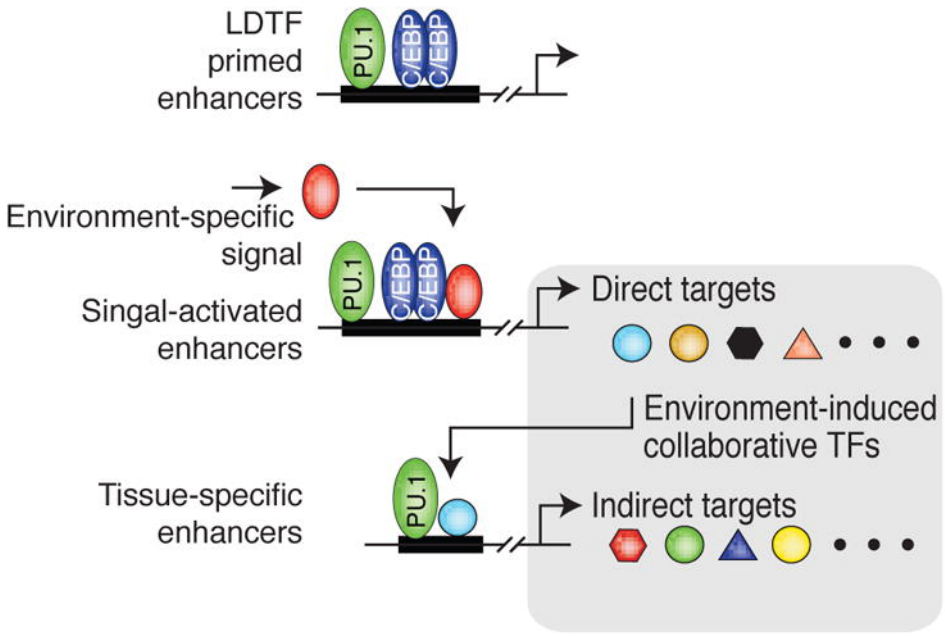


Figure 6. Schematic representation of enhancers' selection upon environmental signal.
Primed enhancers are activated following environmental signal leading to the expression of environment-induced genes, which include also collaborative TFs. Subsequently, tissue-specific enhancers are unveiled by the interaction between collaborative TFs and LDTFs. The combined activity of direct and indirect targets allows determining the identity of macrophages in a particular environment (Link et al., 2015).

Thus, since PU.1 is able to reprogram the chromatin landscape through epigenetic modifications and maintain enhancers in an open chromatin conformation, it can be considered a pioneer factor even if the biochemical evidences are still lacking (Magnani et al., 2011). However, studies unveiled that the cooperative binding between PU.1 and other TFs is required to open chromatin of certain regions. As inferred by ChIP-seq experiments performed in macrophages, the Interferon regulatory factor-8 (IRF8) is another known partner of PU.1, besides C/EBP and AP-1 family members. All together these factors control myeloid-cell development (Ghisletti et al., 2010). PU.1 and C/EBP binding,

occurring at the same distal regions, is frequently reciprocally dependent. Mutagenesis experiments, in which the C/EBP recognition motif carries mutation that abolish the C/EBP binding, demonstrated that PU.1 binding to its nearby intact cognate site is affected by the absence of C/EBP. Likewise mutation in the PU.1 binding motif causes impairment both of PU.1 binding and of C/EBP binding to its intact binding site (Heinz et al., 2013). Additionally it has been suggested that in immune cells, a subset of IFN-stimulated response element (ISREs) called the ETS/IRF response element, is activated by IRF-8 in cooperation with PU.1 (Tamura et al., 2005). Indeed cooperative interaction between IRF8 and PU.1 in basal condition is required for the expression of a subset of genes crucial for macrophage activity (Mancino et al., 2015). This suggests that the cooperative interaction between two TFs is required to compete with histones. Therefore, it is likely that PU.1, C/EBP and other macrophage LDTFs act to define the specific repertoire of enhancers (Heinz et al., 2013). However, in the majority of cases, PU.1 alone is able to bind DNA and to allow the binding of other proteins, which are not able to contact DNA independently (Gallant and Gilkeson, 2006). Therefore, PU.1 is the myeloid master regulator that establishes and maintains the chromatin landscape of macrophages.

3. Aim

Cis-regulatory elements (promoters and enhancers) capable of initiating transcription contain both motifs recognized by sequence-specific TFs, which collaboratively recruit the RNA Pol II initiation complex, and core promoters, namely DNA sequences that enable initiation of RNA synthesis (Kadonaga, 2012).Transcriptionally competent *cis*-regulatory elements differ not only because of the TF binding motifs they contain, but also in core promoter composition and architecture. Such diversity probably reflects the specific and unique regulatory requirements of each individual gene (Tong et al., 2016) and is consistent with the observation that core promoter elements, in addition to enable initiation, also contribute to gene regulation (Zabidi et al., 2015).

However, the identity of factors able to bind to these regions and thus to regulate gene expression remains to be entirely determined. Specifically, the basic mechanism by which the core promoter of constitutive active genes, lacking strong core promoter elements and characterized by multiple weak start sites, are able to maintain a steady level of transcription is not well understood.

We set out to mechanistically dissect those *cis*-regulatory elements that in terminally differentiated cells are endowed with constitutive transcriptional activity. We used a well characterized system, primary mouse macrophages, in which the class III ETS family member Sfpi1/PU.1 determines H3K4me1 deposition and nucleosome depletion at the vast majority of macrophage-specific enhancers as well as at a fraction of the promoters of macrophage-specific genes (Barozzi et al., 2014; Ghisletti et al., 2010; Heinz et al., 2010; Ostuni et al., 2013).

We found out that class I and II ETS sites, but not PU.1 site, were overrepresented on TSS-proximal regions of constitutively expressed genes. For its relative position to TSS, the ETS site could be a good candidate as a possible genomic element involved in

promoting transcription. Therefore, this prompted us to deeply investigate the specific role (if any) of ETS proteins in transcriptional activity.

Genomic distributions of ETS family members expressed in macrophages indicated that some ETS proteins (GAPBA, ELF1 and ELF4) but not others (FLI1 and PU.1) had a striking binding preference for TSS-adjacent regions.

Interestingly, the ability of ETS proteins to bind to core promoters was not simply related to DNA binding preferences, but instead was related to the ability to recruit the general transcription machinery. By using functional, genetic and proteomic approaches we determined a crucial role of ELF proteins in controlling the constitutive transcriptional activity of a broad panel of distal and gene-proximal *cis*-regulatory elements. Overall, these data indicate that a distinct subset of ETS proteins imparts high transcriptional activity to a broad range of both housekeeping and tissue-specific cis-regulatory elements.

4. Materials and methods

4.1 Cell growth and treatment conditions

Macrophages were derived from bone marrows of C57/BL6 mice (Harlan) and plated in 10 cm dishes for 6 days in 10 ml of BM-medium (Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 20% low-endotoxin fetal bovine serum (FBS), 30% L929 conditioned medium, 1% glutamine, 1% penicillin-streptomicin, 0.5% sodium pyruvate and 0.1% β- mercaptoethanol). RAW 264.7 (mouse macrophages), Hepa 1-6 (mouse hepatoma cells) and L-Wnt-3A (mouse fibroblasts) were purchased from the American Type Culture Collection (ATCC). RAW 264.7 and Hepa 1-6 cells were grown in DMEM supplemented with FBS (North and South America origin respectively), 1% penicillin-streptomicin and 1% glutamine, whereas L-Wnt-3A cells were grown in DMEM supplemented with FBS (South America origin), 1% penicillin-streptomicin, 1% glutamine and 400 μg/ml G418.

Stimulations were carried out at day 6 with LPS from E.Coli serotype EH100 (Alexis) at 100ng/ml and with a-amanitin (Sigma) at 10ng/ml.

4.2 Chromatin Immuno Precipitation coupled with sequencing

Macrophages (8×10^7 for FLI1 and ELF1, 4×10^7 for GABPA and ELF4 or 5×10^6 for PU.1) and RAW264.7 cells (8×10^7 for RNA Pol II and 2×10^7 for H3K27ac) were fixed for 10 minutes with 1% of formaldehyde and were lysed in RIPA buffer (140 mM NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 1% Triton x-100, 0,1% SDS, 0,1% DOC). After chromatin shearing by sonication, lysates were incubated with 10 μg of the following antibodies: ELF1 (Santa Cruz sc-631x), ELF4 (homemade), FLI1 (Santa Cruz sc-356x), GABPA (Santa Cruz sc-22810), RNA Pol II (Santa Cruz sc-889x) and 3 μg of PU.1 antibody (home made) and H3K27ac (Abcam ab4729). Antibodies were prebound overnight to 75 μl of G protein-coupled para-magnetic beads (Dynabeads) in PBS/BSA

0.5%. Beads were then added to lysates (the preclearing step was omitted), and incubation was allowed to proceed overnight. Beads were washed six times in a modified RIPA buffer (50 mM HEPES (pH 7.6), 500 mM LiCl, 1 mM EDTA, 1% NP-40, and 0.7% Na-deoxycholate) and once in TE containing 50 mM NaCl. DNA was eluted in TE containing 2% SDS and crosslinks reversed by incubation overnight at 65°C. DNA was then purified by Qiaquick columns (QIAGEN) and quantified with QuBit (Thermo FIscher). For validation by ChIP-QPCR, purified DNA was used for amplification on an Applied Biosystems 7500 Fast Real-time PCR system (with Applied Biosystem Sybr-green). ChIP DNA was prepared for HiSeq2000 sequencing following standard protocols.

4.3 Plasmid construction

Endogenous promoters were cloned into pGL3-basic vector (Promega) between NheI and KpnI sites. The genomic regions cloned, with the distance of the ETS site from the TSS, are shown in Table 1. All the primers used to clone the regions of interest contain two restriction enzyme sites and the SP1-INR sequence (5'-CCCGCCCATCTTG-3', the -1 / +1 Py / Pu dinucleotide is underlined) (Weis and Reinberg, 1997). Primers were designed with sticky ends, considering the restriction enzymes cut, in order to avoid the digestion step. 20 μM of primers were annealed in NEB buffer 2 at 95°C for 4 minutes and then were slowly cooled to room temperature over the period of several hours. 2 μl of annealed primers were ligated into 20 ng of pGL3-basic vector using T4 ligase enzyme (NEB) at 16°C for 4-20 hr.

Name	**Genomic position**			**TSS distance (pb)**
P1	chr7	12922248	12922309	40
P2	chr3	137581390	137581490	33
P3	chr6	83028284	83028384	38
P4	chr15	78971617	78971717	45
P5	chr4	116270265	116270365	24
P6	chr19	38129430	38129530	25
P7	chr7	74367394	74367494	42
P8	chr2	152673600	152673700	45

P9	chr17	22099418	22099518	24
P10	chr5	110425440	110425540	19
P11	chr7	52151349	52151449	15
P12	chr19	5041310	5041410	12
P13	chr2	69723654	69723754	53
P14	chr12	86621368	86621468	59
P15	chr2	4933837	4933937	36
P16	chr11	57985063	57985163	62
P17	chr11	43247161	43247261	10
P18	chr11	19018959	19019018	46
P19	chr16	8637679	8637738	12
P20	chr7	141447311	141447370	21
P21	chrX	12673669	12673728	30
P22	chrX	7841797	7841856	14
P23	chr13	67755135	67755194	47
P24	chr13	67269069	67269128	20
P25	chr12	8208048	8208107	19
P26	chr10	128821592	128821651	49
P27	chr1	37997316	37997375	47
P28	chr12	72085798	72085857	21
P29	chr4	41723138	41723197	42
P30	chr11	17211859	17211918	48
P31	chr3	86999342	86999401	45
P32	chr2	120404078	120404140	10
P33	chr13	18717233	18717292	14
P34	chr2	36049399	36049458	10
P35	chrX	140956690	140956749	31
P36	chr4	43023228	43023287	51
P37	chr2	164804882	164804941	46
P38	chr18	37644205	37644264	26
P39	chr4	86612050	86612109	38
P40	chr6	142345619	142345678	13
P41	chrX	8145818	8145877	20
P42	chr7	7121446	7121505	57
P43	chr2	26902880	26902939	16
P44	chr1	57995912	57995971	31
P45	chr2	129129745	129129804	59
P46	chr9	110656444	110656503	47
P47	chr12	86311789	86311848	16
P48	chr11	98026651	98026710	43
P49	chr2	14970898	14970987	23
P50	chr10	89732296	89732359	10
P51	chr14	12284199	12284258	21

Table 1. Wild-type sequences cloned in pGL3 basic for luciferase assay.

The loss-of-function mutants of the ETS site (GGAA > CCAA), ELF-to-PU.1 mutant site (CCCGGAAGT > AGAGGAAGT) and the enhancer regions were cloned using the same strategy of the wild-type primers. ETS site mutated primers are shown in Table 2, ELF-to-PU.1 mutated primers are listed in Table 3, whereas enhancer regions are reported in Table 4.

Name	Primer Sequence
P1_F	CGCGCGTGCGTCTCTACCTCTGA**CCCCCAAGTT**TTCGCTTTGCATAAAAGCGCTAGTCCGCTGCCCCGCCCATCTTGG
P1_R	CTAGCCAAGATGGGCGGGGCAGCGGACTAGCGCTTTTATGCAAAGCGAAAACTTGGGGGTCAGAGGTAGAGACGCACGCGCGGTAC
P2_F	CCGCGCGCGCAATTTCCGCCCGGGTCTC**CCCCCAAGTT**CCGAGTCCCGGTACTGGAGGGTGCCCGCCCATCTTGG
P2_R	CTAGCCAAGATGGGCGGGCACCCTCCAGTACCGGGACTCGGAACTTGGGGGGAGACCCGGGCGGAAATTGCGCGCGCGGGTAC
P3_F	CAGTCCCGCCCCGCAGGCGGT**ACCCCAAGTG**GCGGGCTGGGATCAGCCTTTAAGATGGCGTCCCGCCCATCTTGG
P3_R	CTAGCCAAGATGGGCGGGACGCCATCTTAAAGGCTGATCCCAGCCCGCCACTTGGGGTACCGCCTGCGGGGCGGGACTGGTAC
P4_F	CGGTCGGCCGCTGCTT**CCCCCAAGTA**ACTGCGCCTGGGTCACGGCGCAAGCGCAAACTACGCCCGCCCATCTTGG
P4_R	CTAGCCAAGATGGGCGGGCGTAGTTTGCGCTTGCGCCGTGACCCAGGCGCAGTTACTTGGGGGAAGCAGCGGCCGACCGGTAC
P5_F	CGGCAGTGGAATTCTGTTACCTGCCCCTGCGGTTCAA**CCCCCAAACG**CTGGCTGCAGAGATCCCGCCCATCTTGG
P5_R	CTAGCCAAGATGGGCGGGATCTCTGCAGCCAGCGTTTGGGGGTTGAACCGCAGGGGCAGGTAACAGAATTCCACTGCCGGTAC
P6_F	CGTGACGTGACCCGCCCTACAGCCGCGGGATTCAAACT**CCCCCAAGCG**GTGTCCTGGTTGATACCCGCCCATCTTGG
P6_R	CTAGCCAAGATGGGCGGGTATCAACCAGGACACCGCTTGGGGGAGTTTGAATCCCGCGGCTGTAGGGCGGGTCACGTCACGGTAC
P7_F	CGTTGGCTCTTAGTCCCCGCC**CCCCCAAGAC**CTGCTTTTTGCAGCCAATCGGCGGGCGCGGTGCCCGCCCATCTTGG
P7_R	CTAGCCAAGATGGGCGGGCACCGCGCCCGCCGATTGGCTGCAAAAAGCAGGTCTTGGGGGGGCGGGGACTAAGAGCCAACGGTAC
P8_F	CCTTGTCCCGCCTCCC**AACTTGGGGT**GTCGCTGAATTTTCTGAGGCCAGTCAAAGGGCAGGGTCCCGCCCATCTTGG
P8_R	CTAGCCAAGATGGGCGGGACCCTGCCCTTTGACTGGCCTCAGAAAATTCAGCGACACCCCAAGTTGGGAGGCGGGACAAGGGTAC
P9_F	CTCTTTGGAGCGGCTCCCCTAGCCCCGCCCAGGACACCA**CGGTTGGGGT**CACGTGACAGCGCTCCCGCCCATCTTGG
P9_R	CTAGCCAAGATGGGCGGGAGCGCTGTCACGTGACCCCAACCGTGGTGTCCTGGGCGGGGCTAGGGGAGCCGCTCCAAAGAGGTAC
P10_F	CCAGAGCCTCTACGCGGACCTGCCAGTCAGAGGGGAAGGCGG**GTTTTGGGGA**TACCTTGGGCCCCCGCCCATCTTGG

P10_R	CTAGCCAAGATGGGCGGGGGCCCAAGGTATCCCCAAAACCCGCC TTCCCCTCTGACTGGCAGGTCCGCGTAGAGGCTCTGGGTAC
P11_F	CCGCTTGGCAAGCAATACTTGTCGCCATCTTTGTTAGGGTCATGA T**GACTTGGGGC**CGGGGCGCCCGCCCATCTTGG
P11_R	CTAGCCAAGATGGGCGGGCGCCCCGGCCCCAAGTCATCATGACC CTAACAAAGATGGCGACAAGTATTGCTTGCCAAGCGGGTAC
P12_F	CCACACGGAAGCGTCGACACCTCCTGGAGAAAGAGAATCCTGCG CAGCA**CTGTTGGGGA**GATCCCCGCCCATCTTGG
P12_R	CTAGCCAAGATGGGCGGGGATCTCCCCAACAGTGCTGCGCAGGA TTCTCTTTCTCCAGGAGGTGTCGACGCTTCCGTGTGGGTAC
P13_F	CCCGCAGTGCCCCAAGTACCGACACGCATGCACGCCCTCTGCCG GGCGCCGAGC**TGCAACCCCC**GCCCATCTTGG
P13_R	CTAGCCAAGATGGGCGGGGGTTGCAGCTCGGCGCCCGGCAGAGG GCGTGCATGCGTGTCGGTACTTGGGGCACTGCGGGGTAC
P14_F	CT**TCCCCAAGTT**TAGAGCTCGGCTTCTCCAACGCTGCCCTGCAC GCCCATCACCGCGATTTCCCGCCCATCTTGG
P14_R	CTAGCCAAGATGGGCGGGAAATCGCGGTGATGGGCGTGCAGGG CAGCGTTGGAGAAGCCGAGCTCTAAACTTGGGGAAGGTAC
P15_F	CTTTGTGGAGGACGCTCTGATTGGCTCC**ACCCCAACTC**TTGGAC CACACAGATCCTTGGCTCCCGCCCATCTTGG
P15_R	CTAGCCAAGATGGGCGGGAGCCAAGGATCTGTGTGGTCCAAGAG TTGGGGTGGAGCCAATCAGAGCGTCCTCCACAAAGGTAC
P16_F	C**CTTTTGGGGTCCCCCAAGCC**GGCCTCGGACTGCGCATGCGTA TAGTCAGCCGGCGTTTGATCCCCGCCCATCTTGG
P16_R	CTAGCCAAGATGGGCGGGGATCAAACGCCGGCTGACTATACGCA TGCGCAGTCCGAGGCCGGCTTGGGGGACCCCAAAAGGGTAC
P17_F	CGTGCCGATCT**CCCCCAAGTT**ATCTCTGCGCA**GCCCCAAGAA**G GCAGGGAA**CCCCCAAGTA**CCCGCCCATCTTGG
P17_R	CTAGCCAAGATGGGCGGGTACTTGGGGGTTCCCTGCCTTCTTGGG GCTGCGCAGAGATAACTTGGGGGAGATCGGCACGGTAC

Table 2. ETS mutated sequences cloned in pGL3 basic for luciferase assay.
The ETS site is highlighted in bold.

Name	Primer sequence
P1_PU.1F	CGCGCGTGCGTCTCTACCTCTGA**AGAGGAAGTT**TTCGCTTTG CATAAAAGCGCTAGTCCGCTGCCCCGCCCATCTTGG
P1_PU.1R	CTAGCCAAGATGGGCGGGGCAGCGGACTAGCGCTTTTATGC AAAGCGAAAACTTCCTCTTCAGAGGTAGAGACGCACGCGCG GTAC
P2_PU.1F	CCGCGCGCGCAATTTCCGCCCGGGTCTC**AGAGGAAGTT**CCG AGTCCCGGTACTGGAGGGTGCCCGCCCATCTTGG

P2_PU.1R	CTAGCCAAGATGGGCGGGCACCCTCCAGTACCGGGACTCGG AACTTCCTCTGAGACCCGGGCGGAAATTGCGCGCGCGGGTA C
P3_PU.1F	CAGTCCCGCCCCGCAGGCGGT**AGAGGAAGTG**GCGGGCTGG GATCAGCCTTTAAGATGGCGTCCCGCCCATCTTGG
P3_PU.1R	CTAGCCAAGATGGGCGGGACGCCATCTTAAAGGCTGATCCC AGCCCGCCACTTCCTCTACCGCCTGCGGGGCGGGACTGGTAC
P4_PU.1F	CGGTCGGCCGC**TACTTCCTCT**AAGTAACTGCGCCTGGGTCA CGGCGCAAGCGCAAACTACGCCCGCCCATCTTGG
P4_PU.1R	CTAGCCAAGATGGGCGGGCGTAGTTTGCGCTTGCGCCGTGAC CCAGGCGCAGTTACTTAGAGGAAGTAGCGGCCGACCGGTAC
P7_PU.1F	CGTTGGCTCTTAGTCCCCGCC**AGAGGAAGTC**CTGCTTTTTGC AGCCAATCGGCGGGCGCGGTGCCCGCCCATCTTGG
P7_PU.1R	CTAGCCAAGATGGGCGGGCACCGCGCCCGCCGATTGGCTGC AAAAAGCAGGACTTCCTCTGGCGGGGACTAAGAGCCAACGG TAC
P8_PU.1F	CCTTGTCCCGCCTCCC**AACTTCCTCT**GTCGCTGAATTTTCTG AGGCCAGTCAAAGGGCAGGGTCCCGCCCATCTTGG
P8_PU.1R	CTAGCCAAGATGGGCGGGACCCTGCCCTTTGACTGGCCTCAG AAAATTCAGCGACAGAGGAAGTTGGGAGGCGGGACAAGGGT AC
P9_PU.1F	CTCTTTGGAGCGGCTCCCCTAGCCCCGCCCAGGACACCA**CAC TTCCTCT**CACGTGACAGCGCTCCCGCCCATCTTGG
P9_PU.1R	CTAGCCAAGATGGGCGGGAGCGCTGTCACGTGAGAGGAAGT GTGGTGTCCTGGGCGGGGCTAGGGGAGCCGCTCCAAAGAGG TAC
P11_PU.1F	CCGCTTGGCAAGCAATACTTGTCGCCATCTTTGTTAGGGTCA TGAT**GACTTCCTCT**CGGGGCGCCCGCCCATCTTGG
P11_PU.1R	CTAGCCAAGATGGGCGGGCGCCCCGAGAGGAAGTCATCATG ACCCTAACAAAGATGGCGACAAGTATTGCTTGCCAAGCGGG TAC
P16_PU.1F	C**CTTTTCCTCTAGAGGAAGTC**GGCCTCGGACTGCGCATGCG TATAGTCAGCCGGCGTTTGATCCCCGCCCATCTTGG
P16_PU.1R	CTAGCCAAGATGGGCGGGGATCAAACGCCGGCTGACTATAC GCATGCGCAGTCCGAGGCCGACTTCCTCTACCGGAAAAGGG TAC
P17_PU.1F	CGTGCCGATCT**AGAGGAAGTT**ATCTCTGCGCA**AGAGGAAGT A**GGCAGGGAA**AGAGGAAGTA**CCCGCCCATCTTGG
P17_PU.1R	CTAGCCAAGATGGGCGGGTACTTCCTCTTTCCCTGCCTACTTC CTCTTGCGCAGAGATAACTTCCTCTAGATCGGCACGGTAC

Table 3. Elf-to-PU.1 sequences cloned in pGL3 basic for luciferase assay.
The PU.1 site is highlighted in bold.

Name	Primer sequence
E1_F	CGTTAACTGTGACAGTGT**CACTTCCGGG**ACTTCCTGGAGGCCTC TGGGAGACACAGAGCTGTCATGTGGGATCCCCCGCCCATCTTGG

E1_R	CTAGCCAAGATGGGCGGGGGATCCCACATGACAGCTCTGTGTCT CCCAGAGGCCTCCAGGAAGTCCCGGAAGTGACACTGTCACAGTT AACGGTAC
E2_F	CGTTAACTTCCATCTGCTGAGAC**CACTTCCGGG**CTCCCAGAGTG AACAGAAGCTTGGCTTCCTGTGAGGGATCCCCCGCCCATCTTGG
E2_R	CTAGCCAAGATGGGCGGGGGATCCCTCACAGGAAGCCAAGCTTC TGTTCACTCTGGGAGCCCGGAAGTGGTCTCAGCAGATGGAAGTT AACGGTAC
E3_F	CGTTAACCAGACGCAGTAGCA**CCCGGAAGTG**AAGTCCGGACTC CTCTCCTCTTTCTTCCTGCTGGAGGGGATCCCCCGCCCATCTTGG
E3_R	CTAGCCAAGATGGGCGGGGGATCCCCTCCAGCAGGAAGAAAGAG GAGAGGAGTCCGGACTTCACTTCCGGGTGCTACTGCGTCTGGTTA ACGGTAC
E4_F	CGTTAACCCCAATTCTGTGGAAATC**CCCGGAAGCT**GGCAGCTGC ATCTCTACCCGCTCTCCTTTGGTTGGATCCCCCGCCCATCTTGG
E4_R	CTAGCCAAGATGGGCGGGGGATCCAACCAAAGGAGAGCGGGTA GAGATGCAGCTGCCAGCTTCCGGGGATTTCCACAGAATTGGGGTT AACGGTAC
E5_F	CGTTAACTCAAGAGGAAGTGTGCCACCTACTGACTCCCAGAAAC **CACTTCCGGT**GGCTGTTTGGTTTAGGATCCCCCGCCCATCTTGG
E5_R	CTAGCCAAGATGGGCGGGGGATCCTAAACCAAACAGCCACCGGA AGTGGTTTCTGGGAGTCAGTAGGTGGCACACTTCCTCTTGAGTTA ACGGTAC
E6_F	CGTTAACTCAGATCTGCCCTGTCCTC**TGCTTCCGGT**ATCTCTATT CCACTTCTGCTTTCTCCTCTAAGGGATCCCCCGCCCATCTTGG
E6_R	CTAGCCAAGATGGGCGGGGGATCCCTTAGAGGAGAAAGCAGAA GTGGAATAGAGAT**ACCGGAAGCA**GAGGACAGGGCAGATCTGAG TTAACGGTAC
E7_F	CGTTAACAAAT**ACCGGAAGCC**CAGCACAGCTAGTTCTTGTGTTT GCTCACAGAAAACAGGAACAGAAGGGATCCCCCGCCCATCTTGG
E7_R	CTAGCCAAGATGGGCGGGGGATCCCTTCTGTTCCTGTTTTCTGTG AGCAAACACAAGAACTAGCTGTGCTGGGCTTCCGGTATTTGTTAA CGGTAC
E8_F	CGTTAACACAGAAGCCTCCGTGGAGT**TCCTTCCGGT**AGGGCTGG TGTTCCTTTTCCAGAGCGTTTCCTGGATCCCCCGCCCATCTTGG
E8_R	CTAGCCAAGATGGGCGGGGGATCCAGGAAACGCTCTGGAAAAGG AACACCAGCCCTACCGGAAGGAACTCCACGGAGGCTTCTGTGTT AACGGTAC
E9_F	CGTTAACAGAGGAAGTACACTCATAGCTGCCTGGGGCTGGGAGT TGTAGAGAA**ACCGGAAGGG**AGCAGGGATCCCCCGCCCATCTTG G
E9_R	CTAGCCAAGATGGGCGGGGGATCCCTGCTCCCTTCCGGTTTCTCT ACAACTCCCAGCCCCAGGCAGCTATGAGTGTACTTCCTCTGTTAA CGGTAC
E10_F	CGTTAACCAAAGAGGAAGTGCTGACGAGAGA**ACCGGAAGTG**AC AGGCCTAAGCCTGATGCAGTTTTGCGGATCCCCCGCCCATCTTGG
E10_R	CTAGCCAAGATGGGCGGGGGATCCGCAAAACTGCATCAGGCTTA GGCCTGTCACTTCCGGTTCTCTCGTCAGCACTTCCTCTTTGGTTAA CGGTAC
E11_F	CGTTAACAGTGCCCCTGGTGTTAATGTCAGTGAAGGACAGA**GCG GAAGTG**TCCTGTTTGTTTTAGCTGGGATCCCCCGCCCATCTTGG

E11_R	CTAGCCAAGATGGGCGGGGGATCCCAGCTAAAACAAACAGGACA CTTCCGCTCTGTCCTTCACTGACATTAACACCAGGGGCACTGTTA ACGGTAC
E12_F	CGTTAACCTTGTAATCATGAACTAGACCTTC**CCCGGAAGTA**CTG ACTTCCTCCCAGCACACATCCTGAGGATCCCCCGCCCATCTTGG
E12_R	CTAGCCAAGATGGGCGGGGGATCCTCAGGATGTGTGCTGGGAGG AAGTCAGTACTTCCGGGGAAGGTCTAGTTCATGATTACAAGGTTA ACGGTAC
E13_F	CGTTAACAACCCTCAG**CACTTCCGCT**TTTTTTCTTATATAGTTGT GGCAACCATGTTGTGCCTGCTGGGGATCCCCCGCCCATCTTGG
E13_R	CTAGCCAAGATGGGCGGGGGATCCCCAGCAGGCACAACATGGTT GCCACAACTATATAAGAAAAAAAGCGGAAGTGCTGAGGGTTGTT AACGGTAC
E14_F	CGTTAACCCGTTCTTCCTCCGCCCAGCCTGGCAACCCCCGCTC**CA CTTCCGCT**TCCAGGAGTCTGGCCGGATCCCCCGCCCATCTTGG
E14_R	CTAGCCAAGATGGGCGGGGGATCCGGCCAGACTCCTGGAAGCGG AAGTGGAGCGGGGGTTGCCAGGCTGGGCGGAGGAAGAACGGGT TAACGGTAC
E15_F	CGTTAACTAAGT**CACTTCCGCT**TCTTGAGCTACAGGAAGTGTGA GGTCAGAGATGTTTGGTCCTGTTTGGATCCCCCGCCCATCTTGG
E15_R	CTAGCCAAGATGGGCGGGGGATCCAAACAGGACCAAACATCTCT GACCTCACACTTCCTGTAGCTCAAGAAGCGGAAGTGACTTAGTTA ACGGTAC

Table 4. Enhancer sequences cloned in pGL3 basic for luciferase assay.
The Elf site is highlighted in bold.

Wild type and Inr-Sp1 mutated synthetic core promoters were cloned in pGL3-basic vector between NheI and NcoI sites, sequences are listed in Table 5 (wild type) and Table 6 (Inr-Sp1 mutated).

Name	Primer Sequence
Synt_1_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGTAAGC TTGGCATTCCGGTACTGTTGGTAAACCCGCCCATCTTGC
Synt_1_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATGC CAAGCTTACTTAGATCGCAGATCTCGAGCCCGGATCGCG
Synt_2_F	CTAGCCCCGGAAGTGCGATCCGGGCTCGAGATCTGCGATC TAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAACCCGCC CATCTTGC
Synt_2_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATGC CAAGCTTACTTAGATCGCAGATCTCGAGCCCGGATCGCAC TTCCGGGG
Synt_3_F	CTAGCGCGATCCGGCCCGGAAGTGCTCGAGATCTGCGATC TAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAACCCGCC CATCTTGC
Synt_3_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATGC CAAGCTTACTTAGATCGCAGATCTCGAGCACTTCCGGGCC GGATCGCG
Synt_4_F	CTAGCGCGATCCGGGCTCGAGATCTCCCGGAAGTGCGATC TAAGTAAGCTTGGCATTCCGGTACTGTTGGTAAACCCGCC

	CATCTTGC
Synt_4_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATGC CAAGCTTACTTAGATCGCACTTCCGGGAGATCTCGAGCCC GGATCGCG
Synt_5_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGCCCGG AAGTTAAGCTTGGCATTCCGGTACTGTTGGTAAACCCGCC CATCTTGC
Synt_5_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATGC CAAGCTTAACTTCCGGGCTTAGATCGCAGATCTCGAGCCC GGATCGCG
Synt_6_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGTAAGC TTGGCCCCGGAAGTATTCCGGTACTGTTGGTAAACCCGCC CATCTTGC
Synt_6_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATAC TTCCGGGGCCAAGCTTACTTAGATCGCAGATCTCGAGCCC GGATCGCG
Synt_7_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGTAAGC TTGGCATTCCGGTACTCCCGGAAGTGTTGGTAAACCCGCC CATCTTGC
Synt_7_R	CATGGCAAGATGGGCGGGTTTACCAACACTTCCGGGAGTA CCGGAATGCCAAGCTTACTTAGATCGCAGATCTCGAGCCC GGATCGCG
Synt_8_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGTAAGC TTGGCATTCCGGTACTGTTGGTAAACCCGGAAGTCCCGCC CATCTTGC
Synt_8_R	CATGGCAAGATGGGCGGGACTTCCGGGTTTACCAACAGTA CCGGAATGCCAAGCTTACTTAGATCGCAGATCTCGAGCCC GGATCGCG
Synt_PU.1_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCTAAGTAAGC TTGGCAGAGGAAGTATTCCGGTACTGTTGGTAAACCCGCC CATCTTGC
Synt_PU.1_R	CATGGCAAGATGGGCGGGTTTACCAACAGTACCGGAATAC TTCCTCTGCCAAGCTTACTTAGATCGCAGATCTCGAGCCCG GATCGCG

Table 5. Wild-type synthetic promoter sequences cloned in pGL3 basic for luciferase assay.

Name	**Primer Sequence**
Synt_1 Mut Inr_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCCGCCCTTCTTGC
Synt_1 Mut Inr_R	CATGGCAAGAAGGGCGGGTTTACCAACAGTAC CGGAATGCCAAGCTTACTTAGATCGCAGATCT CGAGCCCGGATCGCG
Synt_1 w/o Sp1_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCATCTTGC
Synt_1 w/o Sp1_R	CATGGCAAGATGGTTTACCAACAGTACCGGAA TGCCAAGCTTACTTAGATCGCAGATCTCGAGC CCGGATCGCG
Synt_1 w/o Sp1 Mut Inr_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCTTCTTGC

Synt_1 w/o Sp1 Mut Inr_R	CATGGCAAGAAGGTTTACCAACAGTACCGGAA TGCCAAGCTTACTTAGATCGCAGATCTCGAGC CCGGATCGCG
Synt_5 Mut Inr_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCCGGAAGTCCCGCCCTTCTTGC
Synt_5 Mut Inr_R	CATGGCAAGAAGGGCGGGACTTCCGGGTTTAC CAACAGTACCGGAATGCCAAGCTTACTTAGAT CGCAGATCTCGAGCCCGGATCGCG
Synt_5 w/o Sp1_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCCGGAAGTCCATCTTGC
Synt_5 w/o Sp1_R	CATGGCAAGATGGACTTCCGGGTTTACCAACA GTACCGGAATGCCAAGCTTACTTAGATCGCAG ATCTCGAGCCCGGATCGCG
Synt_5 w/o Sp1 Mut Inr_F	CTAGCGCGATCCGGGCTCGAGATCTGCGATCT AAGTAAGCTTGGCATTCCGGTACTGTTGGTAA ACCCGGAAGTCCTTCTTGC
Synt_5 w/o Sp1 Mut Inr_R	CATGGCAAGAAGGACTTCCGGGTTTACCAACA GTACCGGAATGCCAAGCTTACTTAGATCGCAG ATCTCGAGCCCGGATCGCG

Table 6. Inr-Sp1 mutated synthetic promoter sequences cloned in pGL3 basic for luciferase assay.

4.4 Transient Transfections and Luciferase Assays

Transient transfections were performed using Lipofectamine Plus (Invitrogen) according to the manufacturer's protocol.

RAW264.7 cells were transiently transfected in a 24-well format (150 x 10^3 cells/well) with 150 ng of empty vector or vectors containing the specified genomic region, together with 50 ng of pRL-TK-renilla vector (Promega). Hepa 1-6 and L-Wnt-3A cells were transiently transfected in a 48-well format (30 x 10^3 cells/well) with 50 ng of promoter vector and 25 ng of the pRL-TK-renilla vector. The luciferase assay (Dual-Glo Luciferase Assay System, Promega) was performed twenty-four hours after transfection. Values were normalized on the Renilla signal and expressed as fold change relative to the empty vector.

4.5 Western blot

Macrophage cells were harvested and washed once in 1x PBS, then lysed in Ripa buffer (10mM Tris-HCl pH 8.0, 1mM EDTA,140 mM NaCl, 10% Triton x-100, 10% SDS, 10% DOC). 30 ug of protein were used to analyze the level of expression of RNApolII and ELF4 through sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The gel was transferred to a PVDF transfer membrane and blocked with 0,5% BSA in 1x TBST. Rpb1 (Santa Cruz sc-899), ELF4 (home made), INTS10 (Origene, TA337360), NELFB (Proteintech, 16418-1-AP), Chd1 (Novus, NB100-60411) and anti-vinculin (Sigma) antibody were revealed with the ECL western blotting detection reagents (Amersham), according to manufacturer's protocol.

4.6 Nuclear Extract for pull-down experiment

RAW 264.7 cells were lysated with a 2-step nuclear extraction protocol. Cells were harvested and washed once in 1x PBS, then lysed in Buffer A (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 0,1% NP40, 10% glycerol), incubated for 5 minutes at 4°C and centrifuged 5 minutes at 4°C, 3000 RPM (3 times the volume of cell pellet was used to resuspend cells). The surnatant was kept as cytosol fraction. The pellet was resuspended in Buffer B (250 mM NaCl, 50 mM Tris-HCl pH 8.0, 0,5 mM EDTA, 0,5 mM EGTA, 0,2% NP40), incubated by rotation 10 minutes at 4°C and centrifuged 5 minutes at 4°C, 13000 RPM. The surnatant was collected as "first nuclear extract". The obtained pellet was resuspended in Buffer C (420 mM NaCl, 20mM Hepes KOH pH 7.9, 20% glycerol, 2 mM $MgCl_2$, 0,2 mM EDTA, 0,1% NP40), incubated 1 hr by rotation at 4°C and centrifuged 1 hr at 33000 RPM. The surnatant was collected as "second nuclear extract" and joined to the "first nuclear extract". This combined nuclear extract was used for the pull-down (PD) experiment.

4.7 Pulldown experiment

160 µl of myone C1 beads (invitrogen) were washed twice with 500 µl of DNA Binding Buffer (DBB containing 1 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0,05% NP40). 50 pmol of wild-type or mutated DNA were conjugated with the beads by rotation for 1 hr at 4°C. The primers used to amplify a region of 240 bp containing the ETS site are shown below (Table 7). Bead-immobilized DNAs were then washed twice with 500 µl DBB and twice with 500 µl of Protein Binding Buffer (PBB containing 150 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 0,5% NP40, 0,2 µg/µl salmon sperm). 400 µg of nuclear extract form RAW 264.7 cells (see below) were added to respuspended beads, pull-down were performed in a final volume of 600 µl in PBB and incubated by rotation for 2 hr at 4°C. 30 µl of flow-through (FT) from pull-down with wild-type and mutated DNA were collected. Beads were washed twice using 500 µl ice-cold PBB, resuspended in 30 µl of NuPage-LDS (Invitrogen) and boiled 5 minutes at 95°C. 10 µg of Raw nuclear extract, 3 µl of samples and 30 µl of FT were loaded on SDS-PAGE gel for western blot detection, whereas 27 µl were loaded on precast gel (NuPage Novex 4-12%, 1,5 mm, Invitrogen).

Name	Primer Sequence
Pull-down_F	GCGGTACCTTCGTGAGACCCTGGATTCT
Pull-down_R	GCGCTAGCCGCTACTGCCTTTGGTCACT
Pull-down_Mut_F	CGGCTGCCTTGCTCCCCCAAGTGGAGGGTTTACAC
Pull-down_Mut_R	GTGTAAACCCTCCACTTGGGGGAGCAAGGCAGCCG
Desthiobiot	TEG/GTGCAGGTGCCAGAACATTT

Table 7. Primers for pull-down bait cloning.
The first fourth primers were used to clone wild-type and mutated Elf site in to pGL3-basic vector. Desthiobiot and pull-down_R primers were used to amplify the wild type and mutated DNA bait from the pGL3-basic vector.

4.8 In-Gel protein digestion

Gel-separated proteins were sliced, de-stained in 50% v/v acetonitrile Ammonium Bicarbonate (AmBic) 50 mM, reduced with 10 mM DTT in 50 mM AmBic for one hour at 56°C and subsequently alkylated with 55 mM iodoacetamide in 50 mM AmBic for 45 min at RT in dark. After each step samples were dehydrated with 100% ethanol and quickly

‘‘dried’’ in a centrifugal evaporator (SpeedVac). Subsequently, gel pieces were washed extensively with 50 mM AmBic and digested with 12.5 ng mL1 trypsin (Promega V5113) in 50 mM AmBic overnight at 37°C. Digested peptides were extracted with extraction buffer (3% TFA, 30% ACN) and 100% ACN. Lyophilized samples were desalted and concentrated on C18 Stage Tips. The elution was carried out with a highly organic solvent (80% ACN) followed by lyophilisation. Prior to LC-MS/MS analysis, samples were re-suspended in 1% TFA in ddH2O.

4.9 Liquid chromatography and tandem mass spectrometry (LC-MSMS)

Peptide mixtures were separated by online nano-flow liquid chromatography using an EASY-nLC™ 1000 system (Thermo Fisher Scientific, Odense, Denmark) directly connected to a QExactive instrument (Thermo Fisher Scientific) through a nanoelectrospray ion source. The nano LC system was operated in a one-column set-up with a 25 cm analytical column (75 μm inner diameter, 350 μm outer diameter) packed with C18 resin (ReproSil, Pur C18AQ 1.9 μm, Dr. Maisch, Germany). Solvent A was 0.1% FA and 5% ACN in ddH2O and solvent B was 80% ACN with 0.1% FA. Samples were injected in an aqueous 1% TFA solution at a flow rate of 500 nl/min. Peptides were separated with a gradient of 0–30% ACN solvent B over 73 min, followed by a gradient 30-60% ACN in 10 min and 60-95% in 1 min at a flow rate of 250 nl/min in the EASY-nLC 1000 system. The Q Exactive instrument was operated in the data-dependent mode (DDA) to automatically switch between full scan MS and MSMS acquisition. Survey full scan MS spectra (from m/z 300-1650) were analyzed in the Orbitrap detector with resolution R=60,000 at m/z 200. The fifteen most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 3e6 and fragmented by Higher Energy Collision Dissociation (HCD), with a normalized collision energy setting of 27%. The maximum allowed ion accumulation times were 20 ms for full scans and 65 ms for MSMS

and the target value for MSMS was set to 1e5 (R=15,000 at m/z 200). The dynamic exclusion and the isolation window were set to 20s and 1.4 Da, respectively. Standard mass spectrometric conditions for all experiments were: spray voltage, 2.4 kV; no sheath and auxiliary gas flow.

4.10 Assay for Transposase-Accessible Chromatin (ATAC-seq)

The original ATACseq protocol (Buenrostro et al., 2013) was modified according to Lara-Astiaso et al. (Lara-Astiaso et al., 2014). First of all, Tn5 transposase was loaded with sequencing adapters (0,125 µl sequencing adapters, 0,4 µl glycerol 100%, 1,2 µl dialysis buffer, 3,6 µl Tn5 transposase) in a single tube reaction for 1 h at room temperature (Dialysis buffer: 100 mM Hepes KOH pH 7.2, 0.2 M NaCl , 0.2 mM EDTA, 2 mM DTT, 0.2% Triton, 20% glycerol). 50 X 10^3 cells were pelleted by centrifugation and re-suspended in 25 µl of cold lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Igepal CA-630). Nuclei were pelleted by centrifugation for 20 min at 500g, 4 ºC. The supernatant was discarded and nuclei were re-suspended in 25 µl of reaction buffer containing 1 µl of Tn5 transposase (made in house) and 5x transposase buffer (50mM Tris-HCl, pH 8.4 and 25mM $MgCl_2$). The reaction was incubated at 37ºC for one hour. Then 5 µl of clean up buffer (900mM NaCl, 300mM EDTA), 2ul of 5% SDS and 2 µl of Proteinase K (20 µg/µl) (New England Biolabs) were added and incubated for 30 min at 40 ºC. Tagmented DNA was isolated using SPRI beads (2x). For library amplification, two sequential PCR reactions were performed in order to enrich small tagmented DNA fragments. We used 2 µl of indexing primers and KAPA HiFi HotStart ready mix. PCR conditions are reported below). After the first PCR, fragments smaller than 600 bp were isolated using SPRI cleanup and a second PCR was performed using the same conditions in order to obtain the final library. DNA concentration was measured with a Qubit fluorometer (Life Technologies) and library quality was analyzed using a Bioanalyzer (Agilent Technologies). Libraries where sequenced on an Illumina HiSeq 2000.

1° PCR conditions:

72°C for 5’
98°C for 2’
98°C for 20’’ } 9 cycles
63°C for 30’’
72°C for 1’
4°C for ∞

2° PCR conditions:

72°C for 5’
98°C for 2’
98°C for 20’’ } n cycles*
63°C for 30’’
72°C for 1’
4°C for ∞

* The number of cycles is determined by the quantity of DNA obtained with the first PCR. If it is less than 0.9 ng/ul, 6 cycles are required; otherwise, if it is more than 0.9 ng/ul, 5 cycles are sufficient.

4.11 CRISPR/Cas9 genome editing

Single-guide sequences specific to the ETS site of Cep55, Fuz, Acyp1 and Pcgf1 core promoters (showed in Table 8) were designed using the CRISPR design tool (http://tools.genome-engineering.org) and cloned into lentiCRISPRv2 (Sanjana et al., 2014). After RAW264.7 infection, single cells were seeded in 96-well plates by dilution and expanded. Clones were first screened evaluating gene expression by QPCR (expression primers are listed in Table 9 and positive clones were subjected to Sanger sequencing.

Name	Primer sequence
Fuz_F	(phospho)CACCGGTTAGGGTCATGATGACTTC
Fuz_R	(phospho)AAACGAAGTCATCATGACCCTAACC
Cep55_F	(phospho)CACCGGGGATTCAAACTCCCGGAAG
Cep55_R	(phospho)AAACCTTCCGGGAGTTTGAATCCCC
Acyp1_F	(phospho)CACCGGAAGCCGAGCTCTAAACTTC
Acyp1_R	(phospho)AAACGAAGTTTAGAGCTCGGCTTCC

Pcgf1_F	(phospho)CACCGTGATCCCAGCCCGCCACTTC
Pcgf1_R	(phospho)AAACGAAGTGGCGGGCTGGGATCAC

Table 8. Genome editing primers.
Primers used to clone the single-guide sequences.

Name	Primer sequence
Fuz_F	GACCCAGTGTGTGGACTGTG
Fuz_R	ACCACTCGGCCTGATACAAG
Cep55_F	CGCCAGAATATGCAGCATCA
Cep55_R	TGGCTGCTCTGTGATGGTAA
Acyp1_F	TCAGGCTGAGGGTAAAAAGC
Acyp1_R	ATGAAGCGCACCTTGGAG
Pcgf1_F	CTCAAACTGGATCGGGTCAT
Pcgf1_R	CTGTCTAAGCCTCGGGACTG

Table 9. qPCR primers
Primers used to check the level of mRNA expression.

4.12 Computational methods

4.12.1 Motif enrichment analysis

In order to identify over-represented motifs corresponding to known TF-binding sites, Pscan (Zambelli et al., 2009) was performed with 3041 models (position weight matrices, PWM) collected from public repositories and literature: HOCOMOCO database v10 (Kulakovskiy et al., 2016), JASPAR database (Mathelier et al., 2016; Diaferia et al., 2016; Jolma et al., 2015). Foreground: nascent expressed transcripts (FPKM $\geq$ 2) from bone marrow macrophages were used. Refseq TSSs were used as background.

4.12.2 ChIP-seq data analysis

Quality control checks on raw sequence data were performed with FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapter trimming and quality filtering were carried out with Trimmomatic v0.32 (Bolger et al., 2014). Reads were mapped to the mm10 (GRCm38) mouse reference genome using Bowtie2 v2.2.4 (Langmead and Salzberg, 2012). Alignments with low mapping quality (MAPQ<10) and PCR duplicates were removed with Samtools v0.1.19 (Li et al., 2009). Peak calling was

performed using MACS2 2.1.0 (Zhang et al., 2008). Each ChIP was compared to input DNA. Enriched regions ("peaks") were filtered by q-value ≤ 0.01 and fold enrichment ≥ 4. Peaks corresponding to blacklisted regions identified by the ENCODE project (Consortium, 2012) were discarded. In order to visualize the raw profiles on the UCSC Genome Browser, BedGraph files generated by MACS2 were converted to BigWig files (Kent et al., 2010) with the wigToBigWig v4 tool of the UCSC Genome Browser Utilities (Kuhn et al., 2013). Tag counts values were transformed in RPM (reads per million) for visualization.

4.12.3 Peaks genomic distribution

Gene Interval Notator (GIN) (Cesaroni et al., 2008) was used to annotate all peaks over mm10 Ensembl genes release v83 (Cunningham et al., 2015). GIN was run with priority set to “gene” and “-20000” as promoter definition.

4.12.4 Histogram of distances between ChIP-seq peaks to the nearest TSS

Transcription start sites were downloaded from the Ensembl mm10 release v83 (Cunningham et al., 2015). Only peaks with at least one TSS within -10/+10 kb were considered. Distances between peaks summit and the nearest transcription start sites were obtained using Bedtools closest v2.19.1 (Quinlan, 2014). Each bin in the histogram represents a 50 bp region.

4.12.5 De novo motif discovery

We ran MEME v4.11.0 (Bailey et al., 2009) considering a window of +/- 100 nucleotides surrounding the summit of the top 3000 peaks (as determined by MACS2 fold enrichment). The following parameters were used: -dna -mod zoops -evt 0.01 -nmotifs 10 -minw 6 -maxw 14 –revcomp.

4.12.6 Functional enrichment analysis of ChIP-Seq enriched regions using GREAT.

For each sample, the top 3000 peaks (as determined by MACS2 fold enrichment) were analyzed by GREAT v3.0.0 (McLean et al., 2010) with default parameters and selecting the whole mm10 genome as background. For each analysis, the top 10 "Biological Processes" according to the best Binomial False Discovery Rate (FDR Binom) were selected, and values were transformed to positive scores as -10*log10 of FDR Binom. Score values were subject to unsupervised hierarchical clustering (method=average; distance=Pearson correlation). To avoid any bias due to the outliers, a saturation procedure was performed: scores exceeding the 80th percentile were set to this value.

4.12.7 Beeswarm plot of gene expression

Gene expression data from different cell types and tissues were obtained from published microarray dataset (Tippmann et al., 2012). Only genes expressed in at least one sample were considered (log of intensity ≥ 7). Transcription start sites were downloaded from the Eukaryotic Promoter Database (EPDnew Mouse v002) (Dreos et al., 2013) and genomic coordinates were converted to the mm10 Mouse genome release with the liftOver program included in the UCSC Genome Browser Utilities.

4.12.8 RNA-seq data analysis

Quality control checks on raw sequence data was performed with FastQC v0.11.2. After quality filtering according to the Illumina pipeline, reads were aligned to the mm10 (GRCm38) mouse reference genome with TopHat v2.1.0 (Kim et al., 2013) and Bowtie v2.2.4 (Langmead and Salzberg, 2012) using the list of Refseq genes downloaded from UCSC Table Browser (Kuhn et al., 2013). Transcript abundance and differentially expressed genes were called using Cufflinks and Cuffdiff 2.2.1 (Trapnell et al., 2013). Reads from chrM and known rRNA or tRNA were discarded (data from UCSC Table Browser). During transcript quantification we used options –N (which specifies for upper-

quartile normalization) and -u (which allows a better weighting of multi-mapping reads). Tracks for the UCSC Genome Browser were generated from bam files with Bedtools genomecov v2.19.1 (Quinlan, 2014) and wigToBigWig v4 (Kuhn et al., 2013). Tag counts values were transformed in RPM.

4.12.9 Coverage profiles of ChIP-Seq data

Tag counts were obtained from BedGraph files produced by MACS2. Input counts from the same genomic regions were subtracted and final values were normalized from 0 to 1.

4.12.10 Relative enrichment plot

FIMO (Grant et al., 2011) was run over the entire mm9 genomic sequence looking for matches (p-value <= 1e-4) to published position weight matrices (PWMs) for the core promoter elements TATA box and Initiator (Inr)(Jin et al., 2006) as well as for the TF ELF4 (Wei et al., 2010). These maps were used to estimate the spatial density of these sequence motifs with respect to the annotated TSSs of EDPnew genes. On the other hand, ChIP-seq coverage for ELF4 and polII binding was instead calculated starting from the raw reads. These were consistently extended to 200 bps. PCR duplicates were filtered. Coverage was estimated using custom C++ scripts. A sliding window of 10 bp was applied across a 10 kbp region, centered on the considered TSSs. The relative density of the signal from each dataset (either motif or ChIP-seq prediction) was then estimated using the entire 10 kbp region. R was used to generate plots.

4.12.11 Scatterplot of ChIP-Seq regions

The number of reads for each region was normalized based on the sequencing depth of the least sequenced sample. Counts were normalized based on the size of the region in kbp and log2-transformed. Each dot was colored accordingly to the enrichment between the two

samples (fold change ≥ 2) of the corresponding region (Red: enriched in treated; Blue: enriched in untreated; Grey: no enrichment).

4.12.12 Scatterplot of motif occurrences in ChIP-Seq peaks

Using the position weight matrix (PWM) of SPI1 (SPI1_HUMAN.H10MO.A) and ELF4 (ELF1_HUMAN.H10MO.A) obtained from the HOCOMOCO database (Kulakovskiy et al., 2016), the genomic regions of interest were scanned using FIMO (version included in MEME Suite v4.11.10) (Grant et al., 2011) with a p-value threshold of 1e-03. For each region, the best matches (by lowest p-value) for each of the two PWMs were selected and values were transformed to positive scores (-10*log10 of p-values). Each dot represents a genomic region. Darker colors indicate higher density.

4.12.13 ATAC-seq data analysis

Quality control checks on raw sequence data were performed with FastQC v0.11.2. Adapter trimming and quality filtering was carried out with Trimmomatic v0.32 (Bolger et al., 2014). Reads were mapped to the mm10 (GRCm38) reference mouse genome using Bowtie2 v2.2.4 (Langmead and Salzberg, 2012). Alignments with low mapping quality (MAPQ<10), PCR duplicates and reads aligned to chrM were removed with Samtools v0.1.19 (Li et al., 2009). Reads aligning to the forward strand were offset by +4 bp, and reads aligning to the reverse strand were offset −5 bp as described in Buenrostro et al. 2013 (Buenrostro et al., 2013). Peak calling was performed using MACS2 2.1.0 (Zhang et al., 2008). In order to visualize the raw profiles on the UCSC Genome Browser (Kuhn et al., 2013), BedGraph files generated by MACS2 were converted to BigWig files (Kent et al., 2010) with wigToBigWig v4 (Kuhn et al., 2013). Tag counts values were transformed in RPM.

4.12.14 Protein identification by MaxQuant software and data analysis

The mass spectrometric raw data were analyzed with the MaxQuant software (version 1.5.2.8) (http://www.maxquant.org/downloads.htm)(Cox and Mann, 2008), using the Andromeda search engine (Cox et al., 2011). A false discovery rate (FDR) of 1% for proteins and peptides was applied, and a minimum peptide length of 6 amino acids was required. The MOUSE 1401 database (51195 entries) was used for peptide identification. Enzyme specificity was set to Trypsin and maximum of two missed cleavages were allowed. The main search was performed with a mass tolerance of 6 ppm. Cysteine carbamidomethylation (Cys+57.021464 Da) was searched as fixed modification, whereas N-acetylation of protein (N-term, +42.010565 Da) and oxidized Methionine (+15.994915 Da) were searched as variable modifications. Peptide and protein identifications were performed automatically with MaxQuant using default settings. Additional option for Match between runs and LFQ were selected.

Label-free DNA oligo pull-down experiments were analyzed using the Perseus (Tyanova et al., 2016) version 1.4. Contaminants, hits to the reverse database (reversed) and proteins identified only by site were filtered out. LFQ intensity values were logarithmzed and then two groups (wild type (WT) and mutated (Mut), with three biological replicates, were generated. Protein identifications were filtered to require at least two valid values in at least one group. To enable statistical analysis, empty values were imputed with random values from the normal peptide intensity distribution, whose mean and standard deviation were chosen to best simulate low abundance values close to noise level (width=0.3, down shift=1.8). Interacting proteins were identified by performing one-side t-test, comparing the WT oligo DNA pull-down versus the Mutate oligo DNA pull-down protein intensities. The p-values and the t-test differences were plotted against each other in the volcano plot. Proteins with p-value ≤ 0.05 were selected as putative interactors (Figure 30)

5. Results

5.1 A subset of ETS proteins associated with the TSS of housekeeping genes.

To identify the TFs that control constitutive gene expression in bone marrow-derived macrophages (BMDM), we determined the TF consensus DNA binding motifs that are statistically overrepresented in the TSS of all genes highly expressed in basal conditions (**Table 10**). In addition to some GC-rich motifs that likely depend on the abundance of CpG island-containing promoters in this set of genes, this analysis retrieved ($p \leq$ 3.6E-246) an ETS motif different from the one recognized by the myeloid specific ETS protein PU.1 (Ghisletti et al., 2010; Heinz et al., 2010). Therefore, a distinct subset of ETS proteins may be selectively associated with the TSS of constitutively active genes.

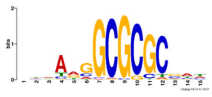
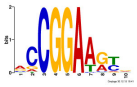
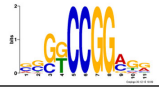
TF Family	PWM	*p*-value
E2F family		4.6^{-261}
ETS family		3.6^{-246}
NRF		1.3^{-238}
Multiple dispersed Zn fingers		1.2^{-230}
CpG-binding proteins		4.9^{-230}

Table 10. Statistical overrepresentation of TF binding sites in macrophages. Pscan analysis was performed on constitutively expressed genes (RNA-seq dataset of untreated macrophages from Liv et al., 2015) versus all TSS.

Therefore we used ChIP-seq to comparatively explore the genomic distribution of ETS proteins highly expressed in macrophages and that based on current annotations (Wei et al., 2010) display distinct DNA-binding specificities, namely FLI1 and GABPA (class I),

ELF1 and ELF4 (class II), and PU.1 (class III). Some obvious trends were apparent irrespective of the threshold used.

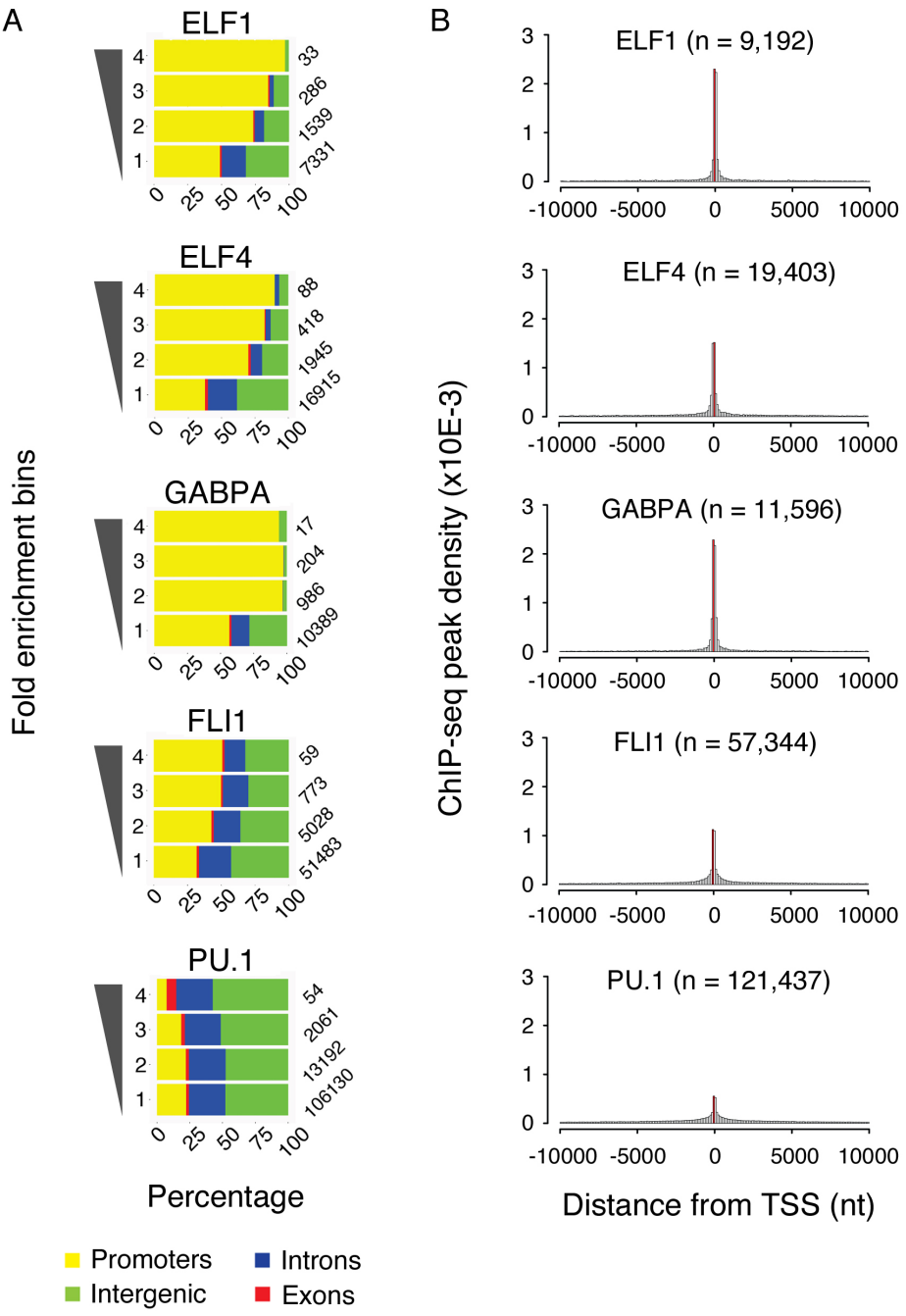


Figure 7. Genomic distribution of ETS family transcription factors.
A) Genomic distribution of ELF1, ELF4, GABPa, FLI1 and PU.1 based on ChIP-seq data.
All peaks were divided into quartiles of progressively higher MACS2 fold enrichment (FE). Peak number is indicated on the right of each bin.
B) TSS-centered analysis of the genomic distribution of ETS family members.

Specifically, ELF1, ELF4 and GABPA showed a significant preference for TSS-proximal regions (TSS ± 2.5 kb), which was increasingly more evident when considering peaks with

a progressively higher score (fold enrichment bins 1 to 4 in **Figure 7A**). Conversely, PU.1 showed a strong preference for TSS-distal (inter- and intra-genic) regions, while FLI1 displayed an intermediate behavior, with *ca.* 50% of the binding events called at intermediate and high stringency being associated with gene promoters **(Figure 7A)**. The genomic distribution of the ETS proteins analyzed in genomic windows of +/-10 kb surrounding the TSS is shown in **Figure 7B**.

A *de novo* motif discovery analysis **(Figure 8)** showed that the TSS-associated ETS proteins (the class I GABPA and the class II ELF1 and ELF4) recognize a virtually identical motif in which the central 5'-GGAA-3' core is preceded by a CC dinucleotide. FLI1, even if previously assigned to the same class as GABPA, showed a distinct binding preference (with an ACA motif preceding the central core), which likely accounts for its different genomic distribution. The PU.1 motif is the most divergent one and is characterized by an extended purine-rich 5' end.

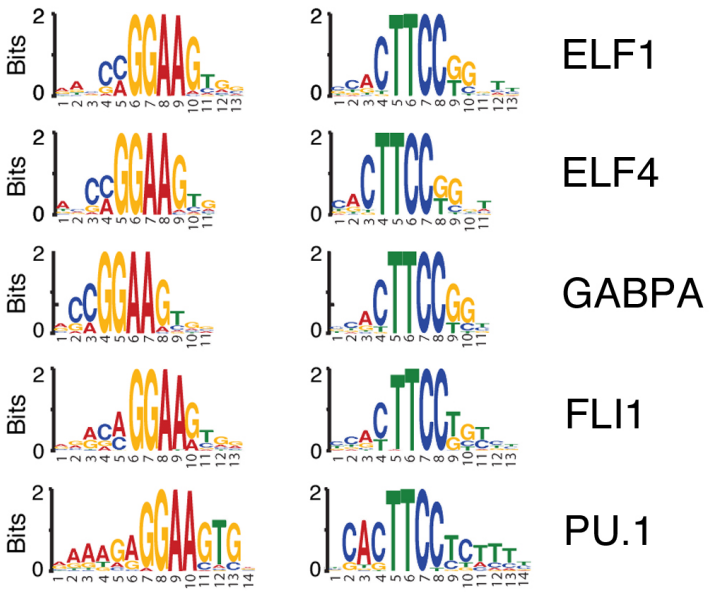


Figure 8. ETS motifs (PWMs) identified by de novo motif discovery analysis (MEME).
the top ranking motifs based on the E values (obtained using MEME) are shown: 4.9e-3395 (FLI1), 1.1e-2686 (GABPA), 1.2e-3443 (ELF1), 2.8e-3285 (ELF4) and 1.4e-5156 (PU.1).

Overall, some ETS proteins (GAPBA, ELF1 and ELF4) have a significant binding preference for TSS-proximal regions that correlated with distinct DNA-binding specificities that were only partially recapitulated by previous *in vitro* studies.

To determine whether differences in genomic distributions reflect distinct functional specificities of the ETS proteins analyzed, we used the GREAT tool (McLean et al., 2010). GREAT links sets of genomic regions to putative biological functions based on the functional annotations of the nearby genes, with a score that is based on the region-gene distance and therefore the probability of correct assignment.

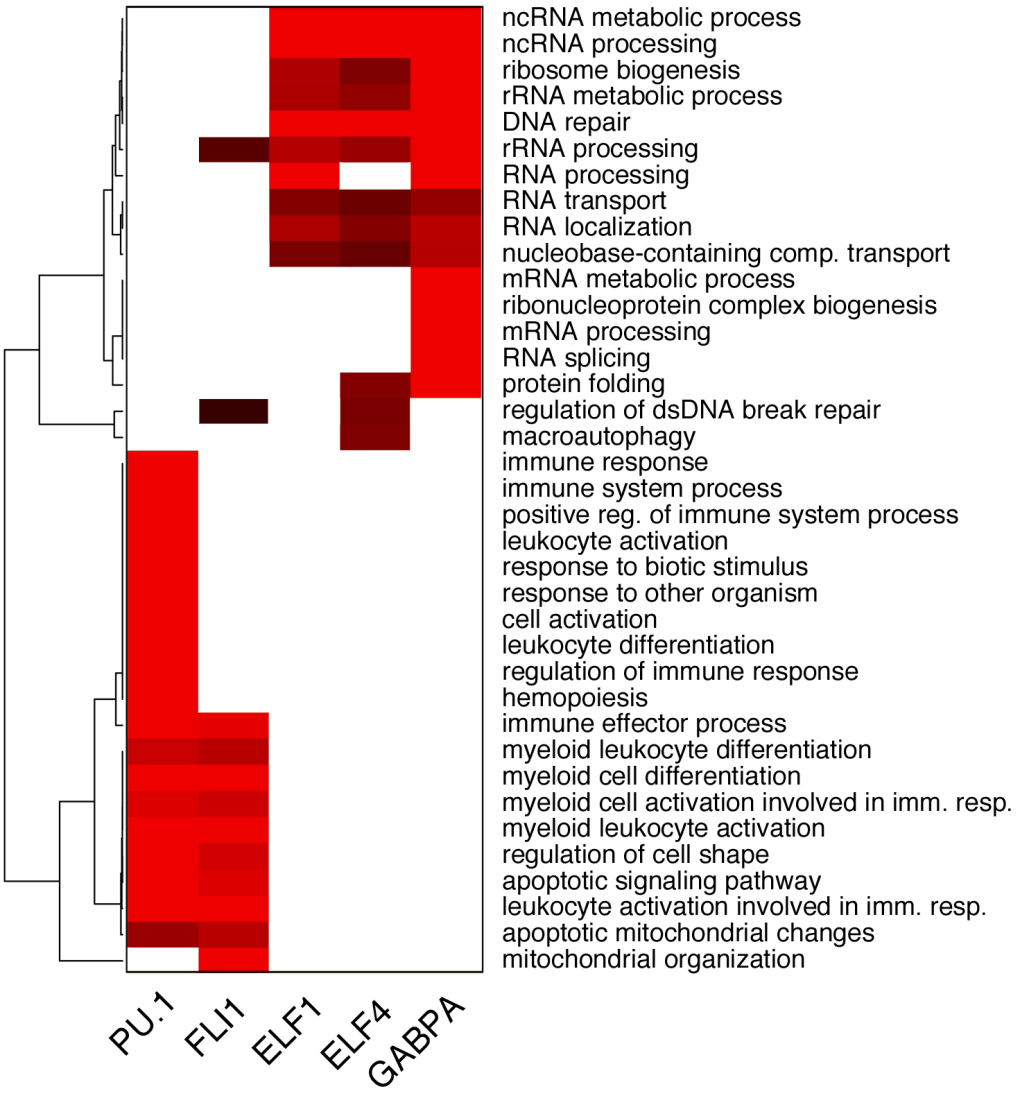


Figure 9. ETS proteins target genes belong to different GO categories.
GO categories associated with different ETS family members as inferred from a GREAT analysis on the individual ChIP-seq data sets.

When considering the genomic regions bound by ELF1, ELF4 and GABPA, GREAT retrieved ontology terms related to housekeeping cellular functions (such as DNA repair, RNA metabolism, ribosome biogenesis and protein folding), while ontology terms associated with PU.1 and FLI1 clustered separately and were mainly related to myeloid cell differentiation and functions (**Figure 9**). These results are in keeping with the known role of PU.1 and FLI1 in macrophage development and indicate that ELF proteins and GABPA are likely involved in housekeeping gene regulation.

Since genes with constitutive expression in macrophages include cell type-specific genes and genes with housekeeping functions and broad expression across tissues, we tested whether the genes bound by ELFs are expressed in different tissues. To this aim, the expression of ELF4-positive genes was evaluated in almost one hundred different tissues using publicly available data sets (Tippmann et al., 2012). In all tissues considered (**Figure 10, upper panel**), and even more so in hematopoietic cells (**Figure 10, bottom panel**), ELF-associated genes were expressed at significantly higher levels than the ELF-negative ones, indicating that irrespective of the cell type, ELF binding correlates with high and constitutive gene expression.

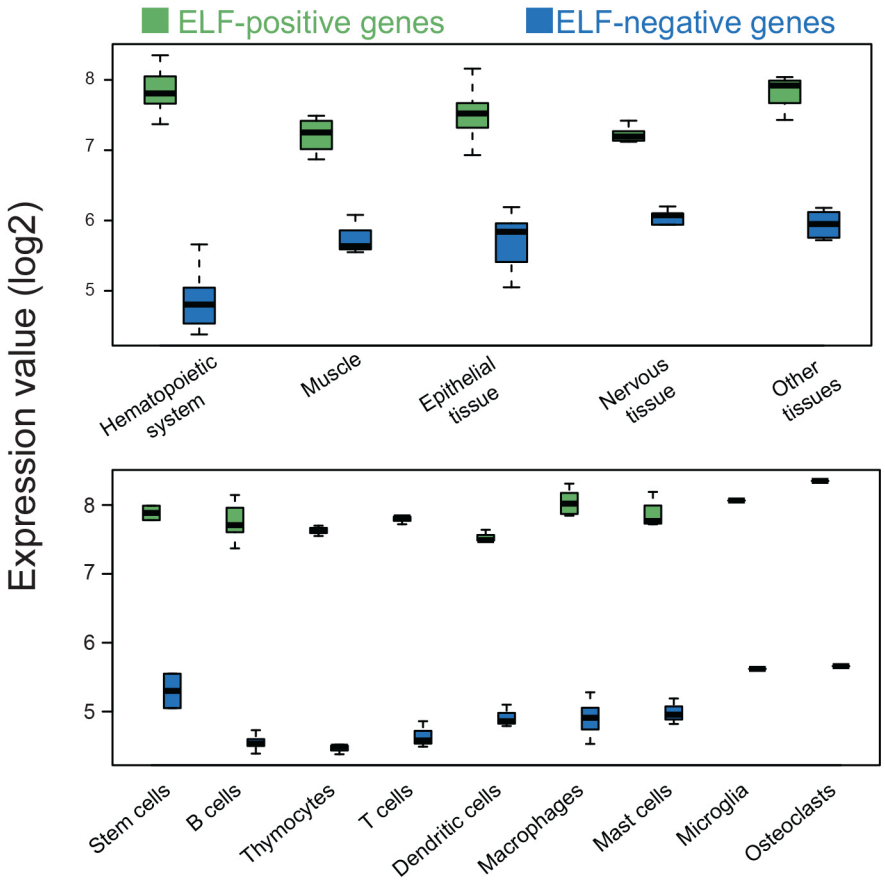


Figure 10. ELF4 association with the TSS of genes with high expression across tissues.
Log2 expression intensity of genes bound (green box plots) or not bound (blue box plots) by ELF4 in different tissues. Black line represents the median of expression. Cells of the hematopoietic system are shown in the lower panel.

Finally, we used a restrictive window of 250pb upstream of TSS to determine the subset of ELF4-bound genes expressed at a basal state in macrophages (Austenaa et al., 2015). 81% (2524 out of 3117) of ELF4 peaks were associated with the promoters of expressed genes. Representative snapshots are shown in **Figure 11**.

Overall, these results indicate that a subset of ETS proteins is associated with TSS of highly expressed genes, including genes that are broadly expressed across tissues.

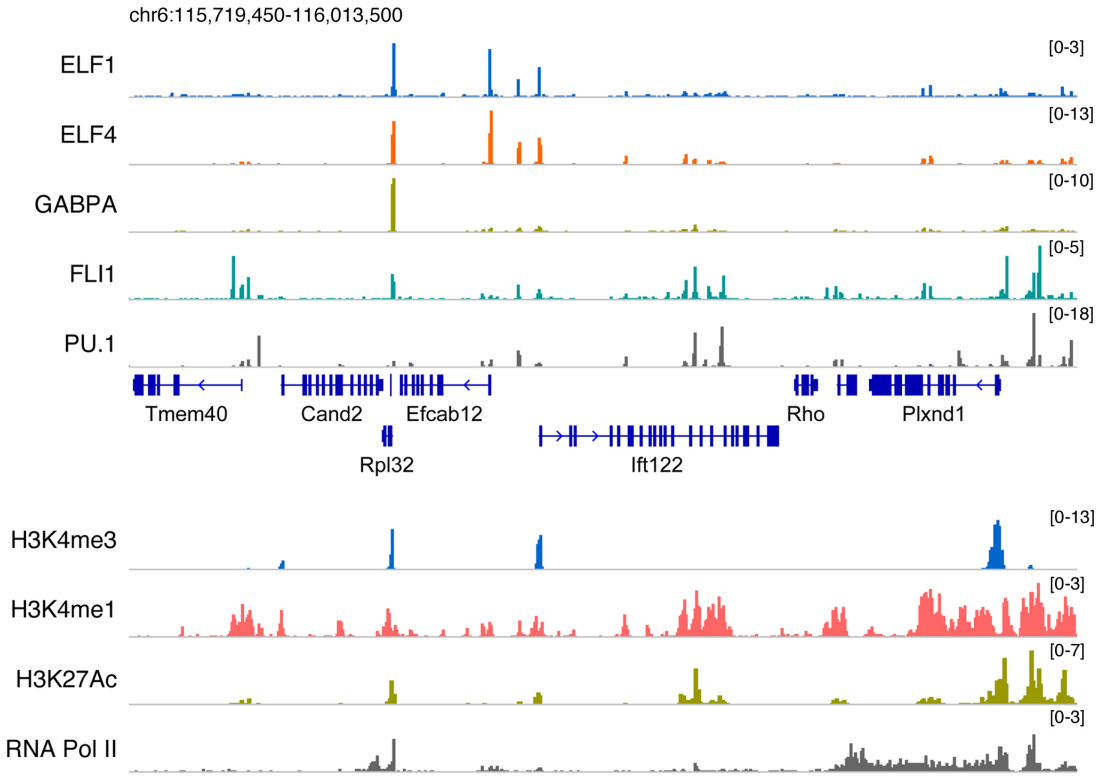


Figure 11. Representative ChIP-seq snapshot.
A representative ChIP-seq snapshot of chr.6 showing the distinct genomic distribution of ETS family TFs and their relationship with histone marks.

5.2 Association of ELF proteins with the macrophage cis-regulatory repertoire.

We next characterized the features of the TSS-proximal regions associated with the promoter-biased ETS proteins (ELFs and GABPA). EDPnew genes were divided in two groups based on their association with ELF4 (being ELF4 genomic distribution extensively overlapping the one of ELF1, **Figure 12**) and then analyzed for the occurrence of the TATA box and Initiator (INR) motifs, the presence of a CpG island and RNA Pol II binding (**Figure 13**).

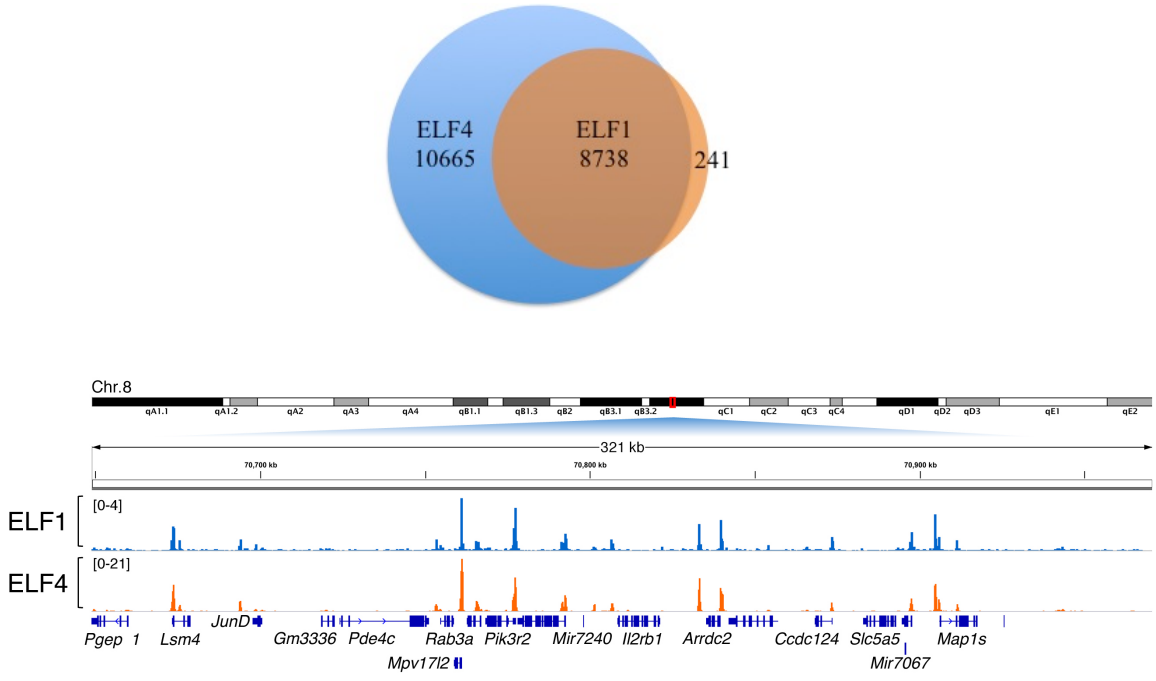


Figure 12. Overlap between the genomic distribution of ELF1 and ELF4.
The Venn diagram on the upper panel shows the overlap between ELF4 (light blue) and ELF1 (orange) genomic occupancy in unstimulated macrophages. The snapshot on the lower panel shows a representative 321 kb genomic region on chromosome 8 with multiple ELF1 and ELF4 peaks.

Consistent with previous data (FitzGerald et al., 2004; FitzGerald et al., 2006), ETS sites mainly occurred nearby TSS, mapping within 50 nt from annotated TSS. The TATA box motif was depleted (but not completely excluded) from the TSS of ELF-bound genes while the INR motif showed the opposite trend, being almost absent at ELF-negative TSS. The ELF-positive group was dramatically enriched for CpG islands (CpGi) and, consistent with the gene expression data shown above, was characterized by much higher constitutive levels of RNA Pol II.

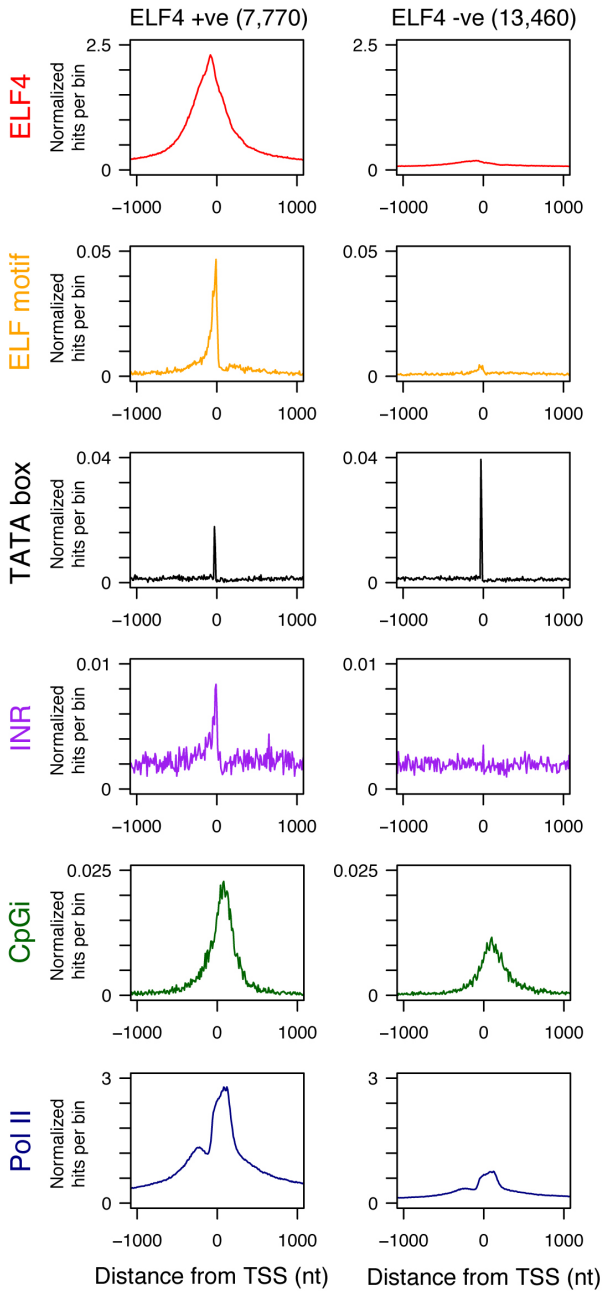


Figure 13. Preferential binding of ELF proteins to active and GC rich TSS.
The TSS (+/- 1000 nt) of EDPnew genes were divided in two groups based on ELF4 binding, as indicated. The profiles of ELF4 ChIP-seq peaks and motifs are shown together with the profiles of the TATA box and INR motifs, CpG islands and Pol II ChIP-seq signal.

An additional and most obvious difference between the two groups became evident when analyzing the organization of nucleosomes around the TSS (Barozzi et al., 2014). ELF binding was almost invariably associated with intense nucleosome depletion upstream of the TSS and strong nucleosome phasing downstream of it, with a particularly prominent +1 nucleosome (**Figure 14**). Differences in nucleosome organization correlate well with the different prevalence of CpG islands in the two groups, since a very high G+C content

disfavors nucleosome assembly (Barozzi et al., 2014; Fenouil et al., 2012). Overall, ELF proteins showed a preference for GC-rich and INR-positive promoters characterized by well-defined nucleosome-depleted areas upstream of the TSS.

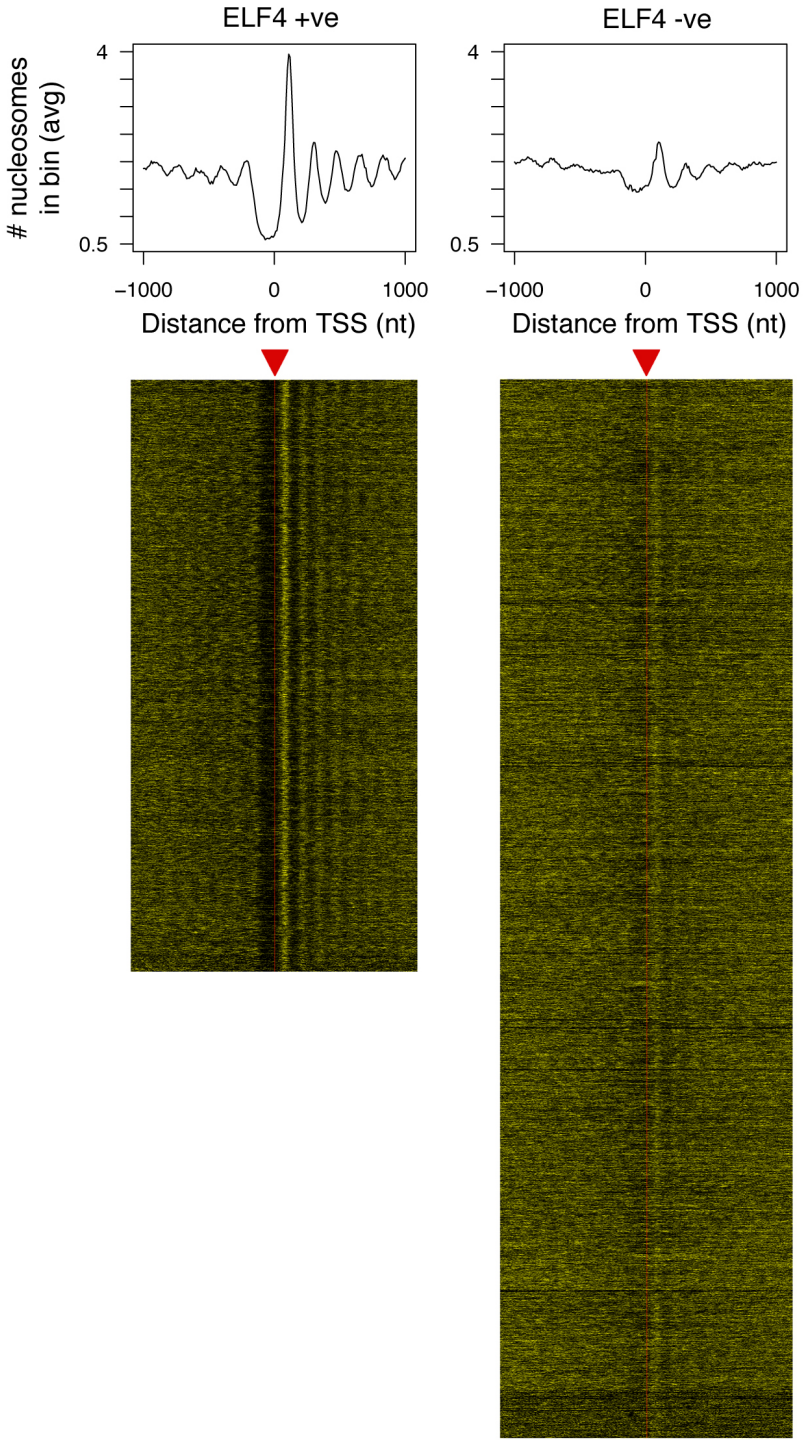


Figure 14. Preferential binding of ELF proteins to nucleosome-depleted TSS.
Nucleosome profiles at ELF-positive (left) and ELF-negative (right) TSS. Upper panels show the distribution of the midpoints of nucleosomal sequencing fragments centered on the annotated TSS of RefSeq genes. The same information is shown below as heatmaps.

To expand the analysis of the relationship between ELF proteins and the activity of *cis*-regulatory elements, we focused on the genomic regions associated with PU.1, the master regulator of the myeloid lineage, which binds and regulates most macrophage-specific enhancers and a fraction of the macrophage-specific promoters (Ghisletti et al., 2010; Heinz et al., 2010; Mancino et al., 2015). We first used PU.1 binding to map genome-wide the macrophage-specific *cis*-regulatory repertoire and then we divided TSS-distal and proximal PU.1-bound regions based on their association with ELFs. At TSS-distal regions, ELF binding was associated with levels of H3K27Ac, H3K4me1, RNA Pol II and even PU.1 significantly higher than those observed at ELF-negative *cis*-regulatory elements (**Figure 15A,** upper panels).

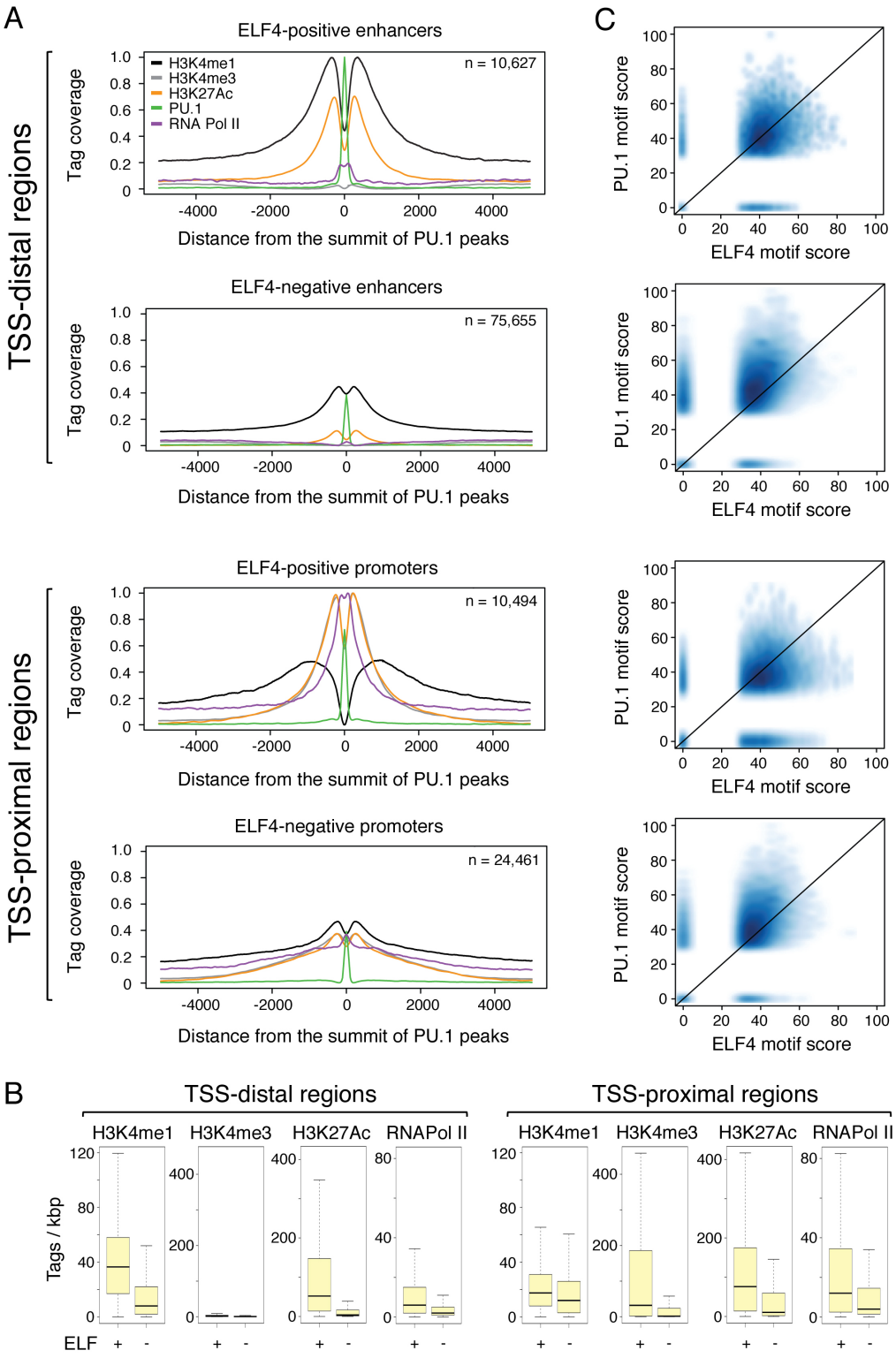


Figure 15. ELF binding is associated with constitutive activity at macrophage PU.1-bound enhancers and promoters.

A) Cumulative distributions showing H3K4me1 (black), H3K4me3 (grey), H3K27Ac (yellow), PU.1 (green) and RNA Pol II (purple) levels of ELF4 positive (ELF4+ve) or negative (ELF4-ve) genomic regions bound by PU.1. Data are shown at both distal (top panels) and TSS-proximal (bottom panels) regions. Plots are centered on the PU.1 signal and normalized from 0 to 1. **B**) Box plots showing read density (tags/kb) for the indicated ChIP-seq at TSS-distal (left) and TSS-proximal (right) *cis*-regulatory regions identified by PU.1 binding. Genomic regions were divided into ELF4 positive and ELF4-negative.

C) Smoothed scatter plots of PU.1 and ELF normalized motif scores in the same genomic regions are shown on the right. Motif scores were measured using FIMO.

The correlation between ELF4 binding and marks of high activity was similarly detected at PU.1-bound TSS-proximal regions (**Figure 15A,** bottom panels), where H3K4me3, H3K27Ac and RNA Pol II showed higher signals at ELF4-positive than at ELF4-negative promoters.

To gain further mechanistic insight into the relationship between ELF binding and underlying sequence features, we scanned the regions in the four groups of regulatory elements shown above for the presence of ELF and PU.1 motifs. ELF and PU.1 sites tended to co-occur in the PU.1+ve / ELF4+ve regions while the ELF motif was either absent or low-scoring in the ELF4-ve regions (**Figure 15C**), indicating that the regions bound by both TFs have a distinct motif composition that correlated with a high activity of the underlying *cis*-regulatory element.

The box plots in **Figure 15B** provide a more quantitative description of the same data and show the strong correlation between ELF binding and indicators of transcriptional activity.

5.3 Release of ELF proteins from promoters correlates with transcriptional shutdown.

Stimulation of macrophages with inflammatory agonists such as LPS results in the transcriptional activation of hundreds of genes as well as widespread gene repression (Glass and Natoli, 2016). Therefore, we investigated whether a relationship exists between LPS-induced transcriptional changes and ELF binding at gene promoters. LPS stimulation for 4 hours resulted in a general reduction in ELF4 occupancy with only a small fraction of regions showing increased binding at either TSS (**Figure 16A,B**) or elsewhere in the genome (**Figure 16C,D,**).

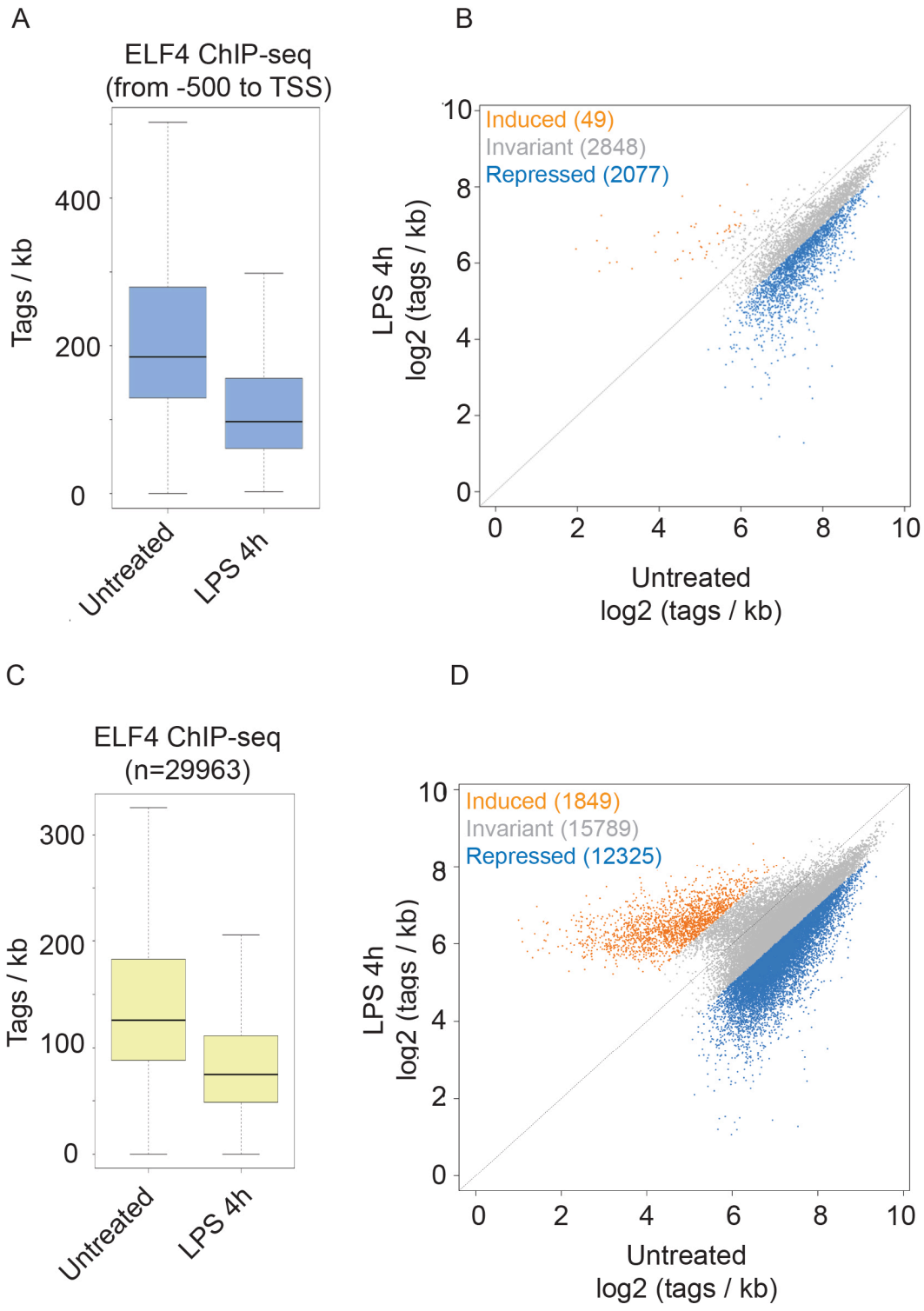


Figure 16. Changes in ELF4 genomic distribution upon LPS stimulation correlated with transcriptional changes.
A) The box plot shows changes in ELF4 occupancy at TSS in response to LPS stimulation. **B)** Scatter plot showing changes in genomic occupancy of ELF4 at TSS after LPS stimulation. Invariant, induced and repressed peaks are indicated.
C) Box plot showing ELF4 ChIP-seq tag density (reads/kb) in unstimulated and LPS-stimulated macrophages. **D)** Scatter plot showing changes in ELF4 occupancy at individual genomic regions after LPS stimulation. Peaks that were significantly increased (1,849) or reduced (12,325) in response to stimulation are highlighted as orange and blue dots, respectively.

To determine whether dynamic variations in ELF4 occupancy correlated with gene expression changes, we generated datasets of chromatin-associated, nascent transcripts from unstimulated and LPS-stimulated (4h) macrophages. In **Figure 17A** (left panel) genes whose transcription significantly changed in response to LPS stimulation (FDR $\leq$ 0.01) were ordered, with the most repressed genes on the left side and the most induced ones on the right. The smoothed scatter plot in the right panel shows ELF4 binding at the corresponding TSS. The overall trend of the plot and the polynomial regression fit (red line) indicate a significant albeit imperfect correlation between changes in transcriptional activity and ELF4 occupancy, with ELF4 release correlating with transcriptional repression and *de novo* ELF4 binding with gene activation. We also correlated genome-wide changes in histone acetylation (H3K27Ac) induced by LPS with changes in ELF4 occupancy. Also in this analysis we detected an imperfect correlation between loss of histone acetylation and ELF4 release on the one hand, and gain in acetylation and gain in ELF4 occupancy on the other (**Figure 17B**).

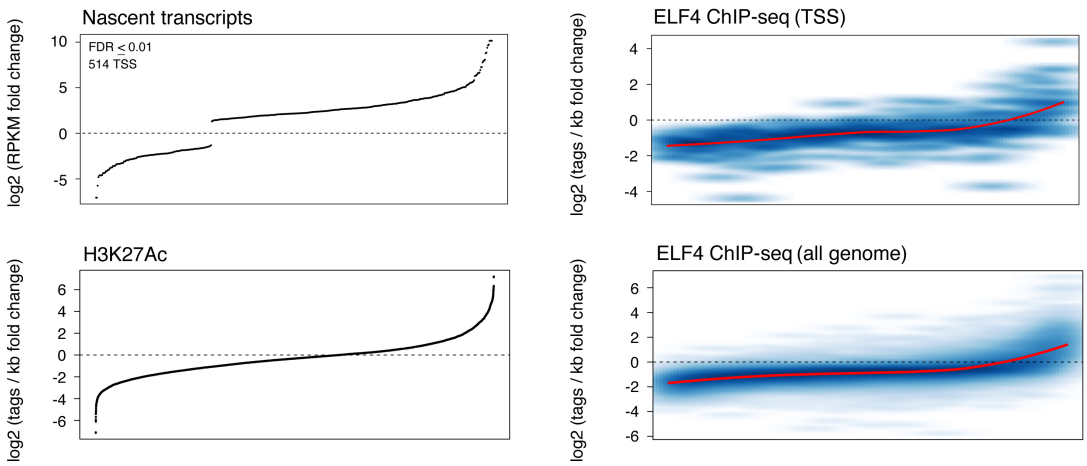


Figure 17. Correlation between transcription changes and ELF4 occupancy.
Upper panel. In the left panel the genes whose transcription was significantly (FDR≤0.01) reduced or increased in response to LPS stimulation (based on nascent transcripts) are ordered from left to right. The smoothed scatter plot in the right panel shows changes in ELF4 occupancy at the TSS (from -500 to 0) of the same genes. The red line represents a polynomial regression fit. **Lower panel.** Correlation between H3K27Ac changes and ELF4 occupancy in LPS-stimulated macrophages. In the left panel all genomic regions whose acetylation was significantly reduced or increased in response to LPS stimulation are ordered from left to right. The changes in ELF4 occupancy at the same regions are shown on the right. Red line: polynomial regression fit.

Figure 18 includes two representative genomic snapshots showing ELF4 release from the TSS of a repressed gene (*Cep55*) and ELF4 recruitment to the TSS and an upstream transcribed enhancer of an LPS-activated gene (*Ccl5*).

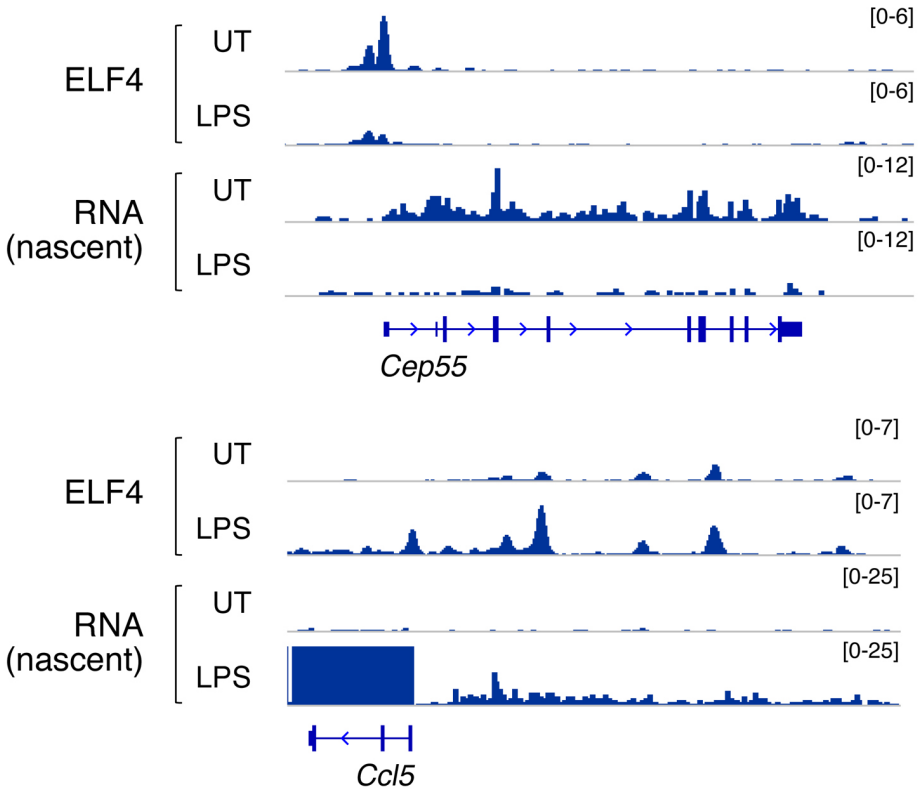


Figure 18. Representative snapshots.
Two representative snapshots showing the correlation between ELF binding and transcription at one LPS-repressed gene (*Cep55*) and one LPS-activated gene (*Ccl5*). The sequence upstream of *Ccl5* is a broad enhancer transcribed upon activation.

Taken together with the data shown above, these results indicate that recruitment of ELF TFs tightly correlates with the recruitment of the transcriptional machinery at a broad panel of *cis*-regulatory elements, in both basal and stimulated conditions.

5.4 Functional activity of ETS sites in minimal promoters.

The strong correlation between ELF binding and constitutive transcriptional activity of TSS-proximal and distal *cis*-regulatory elements, as well as the vicinity of ETS sites to mapped TSS, prompted us to explore a direct role of ETS proteins in transcriptional activation. Attempts to simultaneously deplete all three ELF proteins expressed in

macrophages (ELF1, ELF2 and ELF4) were not successful and we had to resort to alternative strategies. We initially tested a large panel of endogenous core promoters consisting in short sequences of fixed length, namely 60 bp from mapped TSS. The promoters tested were selected based on the following criteria: *i*) presence of the canonical class II ETS site within 60 nt from the mapped TSS; *ii*) absence of a TATA box (see **Figure 13**); *iii*) binding by ELF4 as determined by ChIP-seq. In a first set of experiments, all sequences were cloned upstream a common sequence including a single SP1 site and an INR sequence (Weis and Reinberg, 1997) in a luciferase vector (pGL3-basic) devoid of either promoter or enhancer sequences and thus with very low-to undetectable basal activity (**Figure 19** and **Table 2**).

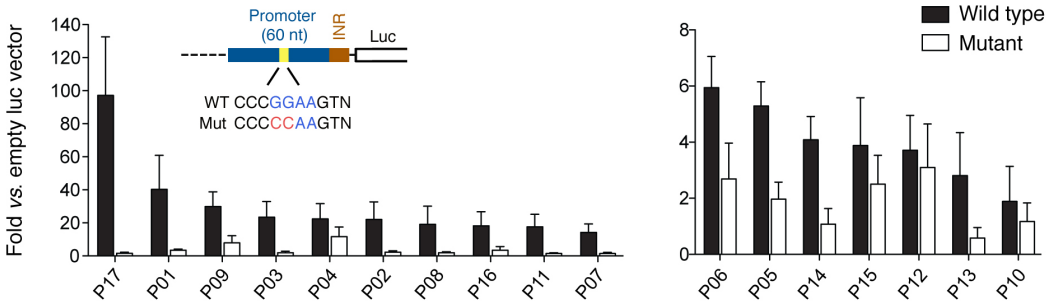


Figure 19. A critical role of ELF-type ETS motifs in the constitutive activity of endogenous promoters. Endogenous core promoters (60 nt from annotated TSS) bound by ELF4 in vivo and bearing a canonical ELF site were cloned into a promoter-less vector (pGL3-basic, Promega) just upstream a cammon canonical initiatior sequence (INR) as indicated. Reporters were transfected in a macrophage cell line (Raw264.7). For each promoter a mutation of the ELF site was generated (white bars). Data are shown as fold enrichment over promoter-less vector. Error bars: SD of at least three independent biological replicates.

Importantly, the SP1-INR combination alone was ineffective at stimulating luciferase expression (**Figure 20**) while a synthetic promoter in which the SP1 or the INR or both motifs were eliminated (**Table 6**), retained ETS-dependent (albeit lower) transcriptional activity. Therefore, the ETS site can efficiently promote transcription in the absence of other TF DNA binding sites or canonical core promoter elements.

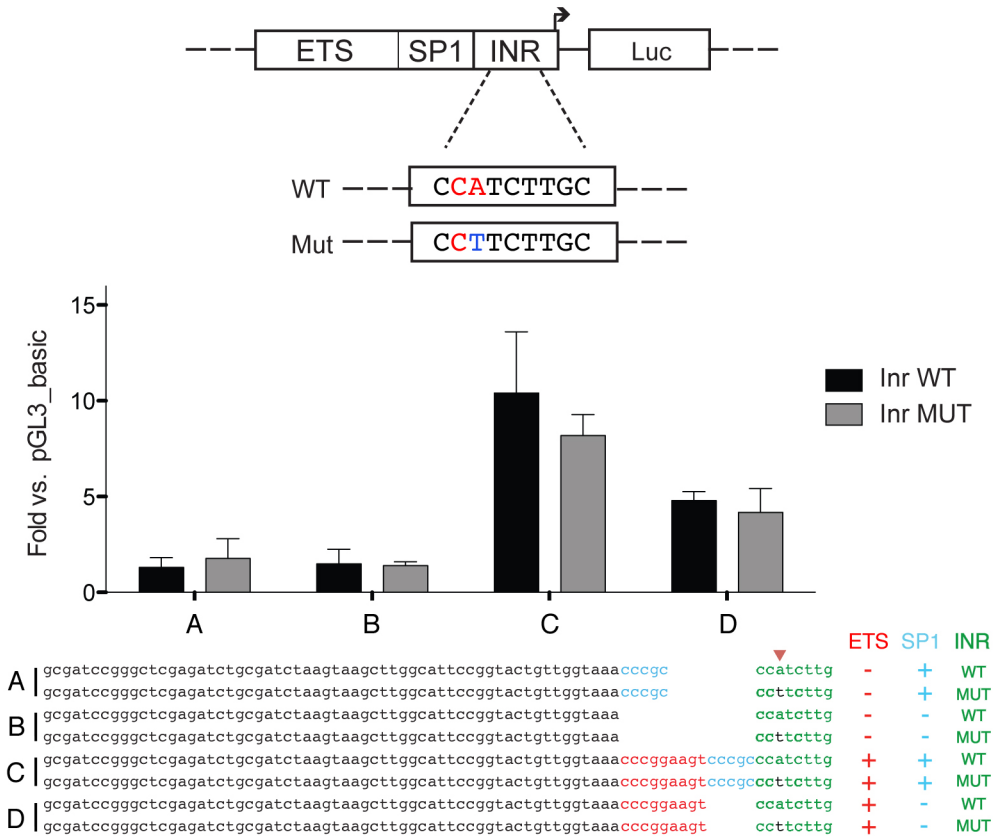


Figure 20. Dissection of the role of INR and SP1 motifs in ELF-dependent transcriptional activation. The schematic diagram above shows the organization of individual sequence elements in the reporter vector. Fold induction over the empty vector are shown in the histogram. The sequence of the plasmids used is shown at the bottom.

Upon transfection in a macrophage cell line (Raw264.7), all the promoters tested stimulated the expression of the reporter gene (**Figure 19**). In 13/17 cases, the transcriptional activity of the promoter was almost completely dependent on the ETS site, since a mutation destroying two nucleotides in its core abrogated luciferase expression (**Figure 19**, white bars). The same core promoters were not only capable to activate transcription in two other non-hematopoietic cell types (hepatocytes and fibroblasts) (**Figure 21**), but they also showed comparable relative strength in the three different cell types. Altogether, these data indicate that binding of ETS proteins close to the TSS imparts constitutive activity to minimal promoters.

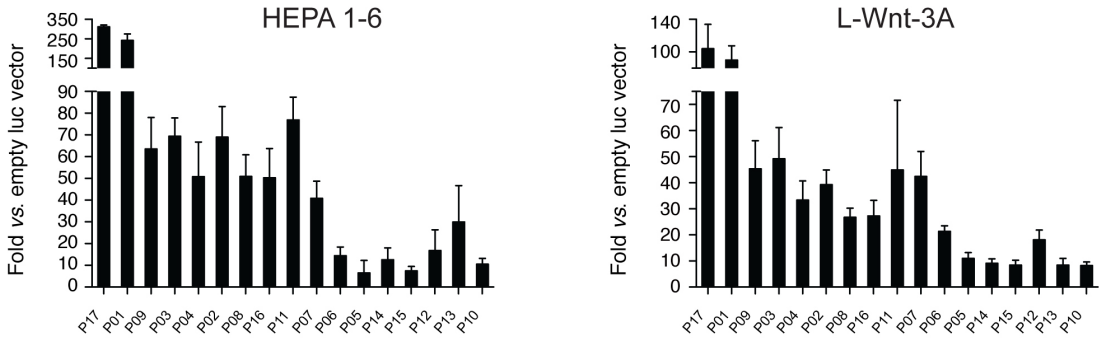


Figure 21. Involvement of ELF-type ETS motifs in the constitutive activity of endogenous promoters in different cell lines. The same ELF-site positive core promoters sequence as in **Fig.19** were tested in hepatocytes (HEPA 1-6 cell line) and in fibroblast (L-Wnt-3A).

Importantly, core promoter activity strictly required an ELF-type ETS site because its replacement with a PU.1 site (**Figure 22**) in nearly all cases (8 / 10) abrogated transcriptional activity.

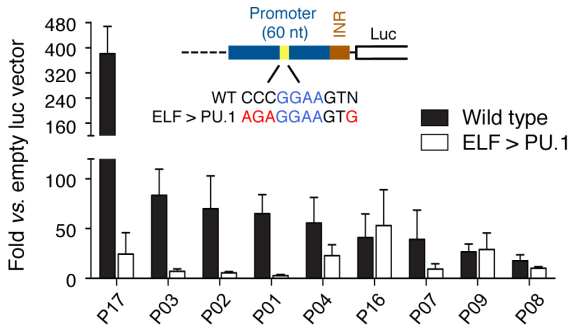


Figure 22. PU.1-type ETS motif drives low expression of endogenous promoters.
The ELF-site of a panel of core promoter sequences shown in **Fig.19** was mutated to PU.1 site (white bars) and the effects on luciferase activity were measured. Error bars: SD of at least three independent biological replicates.

Since ELFs also bound to active enhancers *via* their specific motifs (**Figure 15**) we investigated whether ELF site-containing enhancers were similarly capable of activating transcription. 60 bp-long DNA fragments corresponding to ELF4-positive enhancers were cloned (**Table 3**) and tested as described above: all enhancers tested were able to activate transcription of the reporter gene (**Figure 23**).

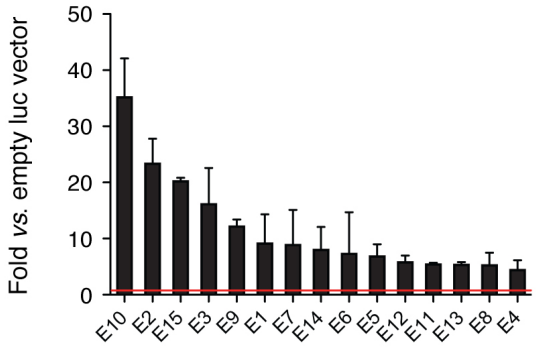


Figure 23. Transcriptional activity of ELF4-positive enhancers.
Enhancers bound by ELF4 and containing an ELF site were tested for their capability to act as core promoters in macrophages. 60 nt regions (+/- 30 bp from ETS summit) were cloned and tested as described in **Fig.19**. Red line indicated a 1 fold enrichment over pGL3-basic vector.

Given the data above, it was critical to determine if ELF sites in endogenous promoters in their own native genomic context are similarly required to activate gene transcription. Therefore we resorted to CRISPR/Cas9-mediated genome editing, taking advantage of the fact that the core ETS site (NGGAA) contains the PAM sequence of SpCas9 (NGG). We selected five genes containing a TSS-proximal ETS site and generated individual clones that were sequenced to identify small and mechanistically informative mutations. For two genes (*Cep55* and *Fuz*) we obtained several clones in which both alleles were properly mutated. In all cases, mutations affecting the ETS site (even if removing a single nucleotide) almost completely abrogated the expression of the adjacent gene (**Figure 24**). Overall, a subset of ETS sites bound by ELF proteins was necessary and sufficient to impart high and constitutive transcriptional activity to a broad panel of *cis*-regulatory elements.

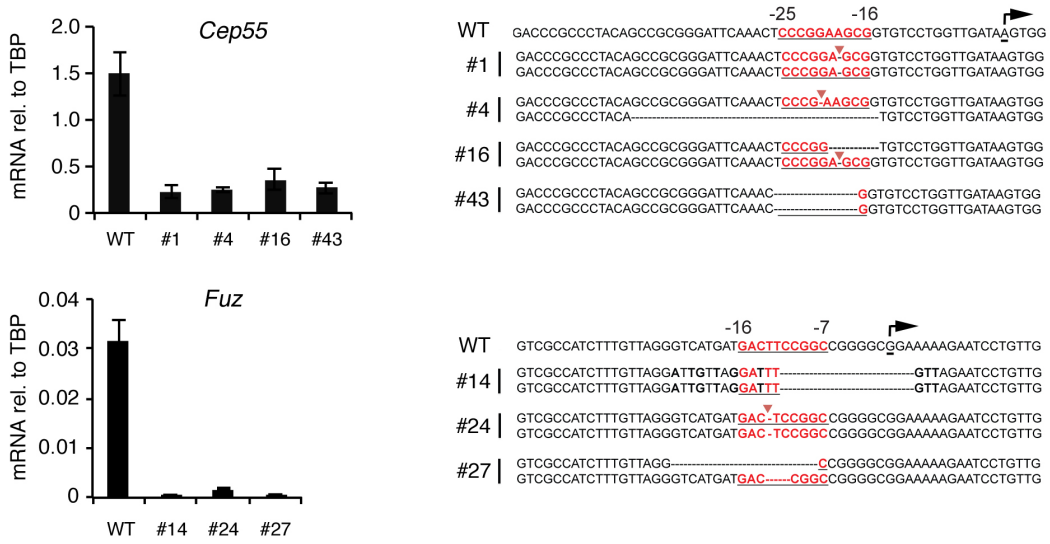


Figure 24. ELF-type ETS sites involvement in the transcriptional activity in the native genomic context.
CRISPR/Cas9 was used to edit genomic sequences spanning the TSS-proximal ELF site at two genes constitutively bound by ELF4. Individual Raw246.7 clones were isolated and sequenced to characterize the mutant sequences at each allele of the two genes. The histogram on the left shows mRNA expression by RT-qPCR in wild type cells (transduced with Cas9 without sgRNA) and in a selected panel of clones bearing informative mutations sapanning the ELF site. The sequence of the mutant clones is shown on the right. The ELF site is shown in red and underlined.

5.5 Mutual interactions between ELF proteins and the transcriptional machinery.

Since the ETS sites recognized by ELFs and GABPA are commonly found within 50 nt (and even more frequently within 20 nt) from mapped TSS (FitzGerald et al., 2004; FitzGerald et al., 2006), we determined the impact on transcriptional activation of the distance between the ETS site and the INR sequence. First, ELF motif-containing promoters (n = 51) were divided into 5 groups based on the distance between the ETS site and the TSS (listed in **Table 1**). 60 bp fragments were then cloned upstream of a common SP1-INR motif as above (**Figure 19**) and tested for luciferase activity. The median transcriptional activity of these fragments correlated with the distance of the ETS site from the TSS, with a shorter distance being generally associated with a higher transcriptional activity (**Figure 25A**). To analyze the effects of the distance between the ETS site and the INR in an identical sequence context, we created a synthetic 60 nt-long core promoter based on a transcriptionally inert backbone from a bacterial plasmid, in which we inserted

an ELF type ETS site at variable distances from the INR (**Table 5**). The combination of an ETS site and the INR was sufficient to activate luciferase expression in this minimal promoter context (**Figure 25B**). Moreover, also the synthetic promoter showed a clear (albeit imperfect) trend where a smaller distance of the ETS site from the INR was associated with higher activity (**Figure 25B**). Also in this synthetic promoter the replacement of the ELF/GABPA site with a PU.1 site caused a loss of activity.

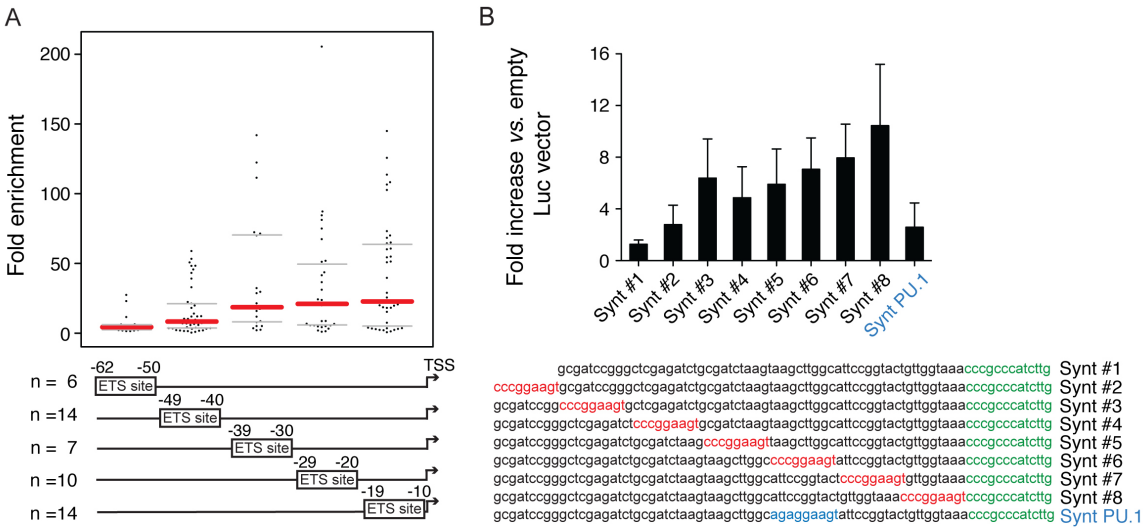


Figure 25 Correlation between ELF-type site and its distance from TSS.
A) Endogenous ELF site-positive core promoters (n=51) as described in Fig.19 were divided into 5 groups depending in the distance from TSS and tested for luciferase activity in macrophages. Each dot of the beeswarm plot represents one biological replicate out of three independent experiments. The central red bars indicate the median, the grey bars the 1st and 3rd quartile. Data are shown as fold enrichment over promoter-less vector (pGL3-basic). **B)** Synthetic sequences containing an ELF-site (red) or a PU.1 site (blue) at different distances from a common initiatior sequence (INR, green) were tested for luciferase activity in macrophages. Data are shown as fold enrichment over promoter-less vector (pGL3-basic). Error bars represent +/-SD of at least three independent biological replicates.

The data shown above hint at the occurrence of close interactions between ETS site-bound proteins and the transcriptional machinery. Therefore we tested if the transcriptional machinery, once recruited to promoters, stabilizes the association of ETS proteins with their DNA-binding site. To this aim we depleted the large RNA Pol II subunit (Rbp1) with an extended (5h) treatment with a-amanitin (**Figure 26A**) and we generated ELF4, GABPA and PU.1 ChIP-seq datasets (**Figure 27**).

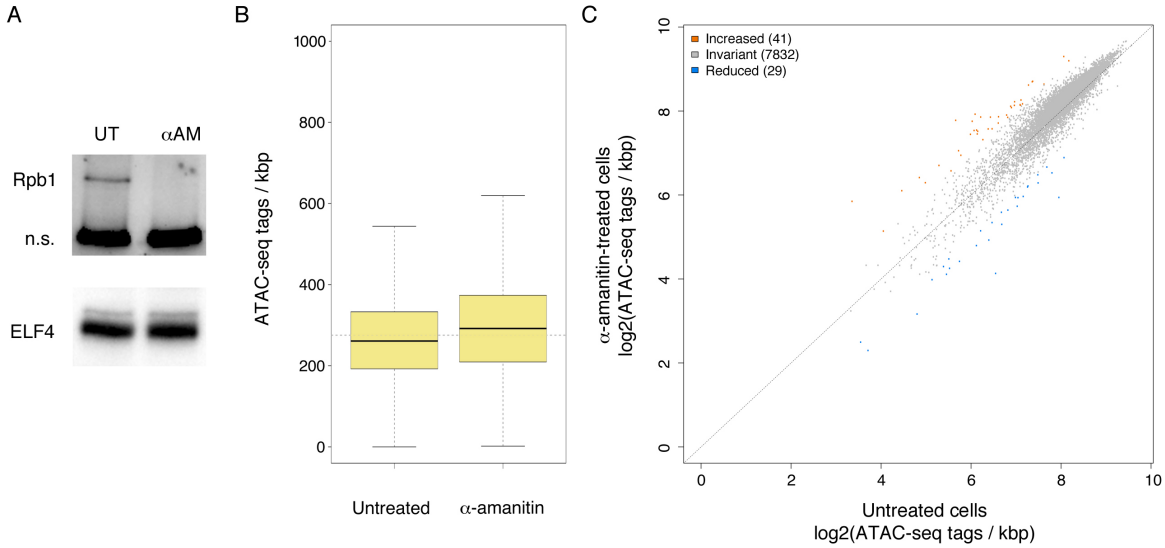


Figure 26. Effects of alpha-amanitin treatment on ELF4 genomic occupancy.
A) Rpb1 and ELF4 western blots in untreated and a-amanitin-treated macrophages. n.s.: non-specific band. **B**) Box plot showing ATAC-seq read density (tags/kb) in untreated and a-amanitin-treated macrophages. **C**) The scatter plot shows ATAC-seq read density in untreated and a-amanitin treated macrophages. Unaffected peaks are indicated as grey dots.

While RNA Pol II depletion did not greatly impact PU.1 and GABPA association with TSS-proximal regions, it nearly completely (7745 / 7902 peaks, 98%) abrogated ELF4 binding (**Figure 27**) without affecting its abundance (**Figure 26A**). The effects of a-amanitin on ELF4 binding were not associated with differences in promoter accessibility, as indicated by an ATAC-seq analysis (**Figure 26B and C**).

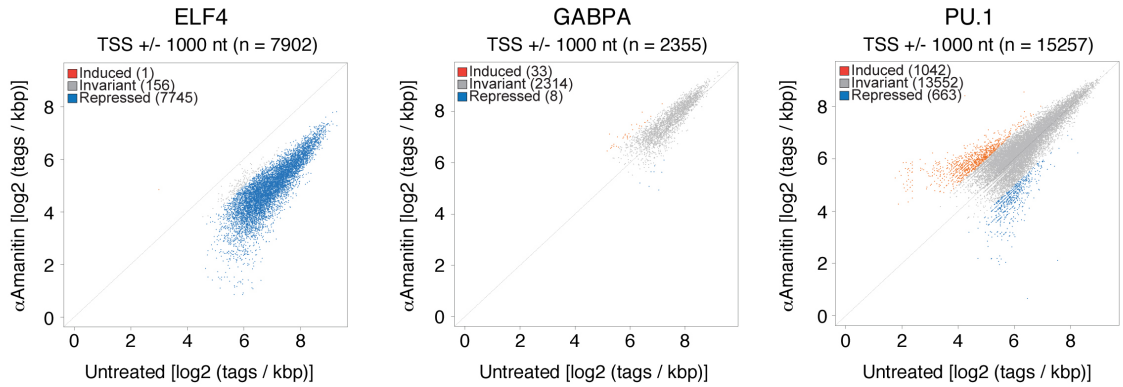


Figure 27. Relationship between TSS-proximal ELF4 binding and the transcriptional machinery.
Scatter plot indicating genomic regions (+/- 1000 bp from TSS) bound by ELF4 (left panel), GABPa (middle panel) and PU.1 (right panel) as determined by ChIP-seq in macrophages treated with alpha-amanitin (5h, 10ng/mL) relative to untreated cells (UT). Blue dots indicate regions where ELF4 signal is reduced upon alpha-amanintin treatment.

These data show that binding of ELF4 (but not GABPA or PU.1) to active TSS-proximal regions is stabilized by RNA Pol II, which is likely an indirect evidence of a close physical interaction between components of the transcriptional machinery and DNA-bound ELF proteins. A representative snapshot is shown in **Figure 28**.

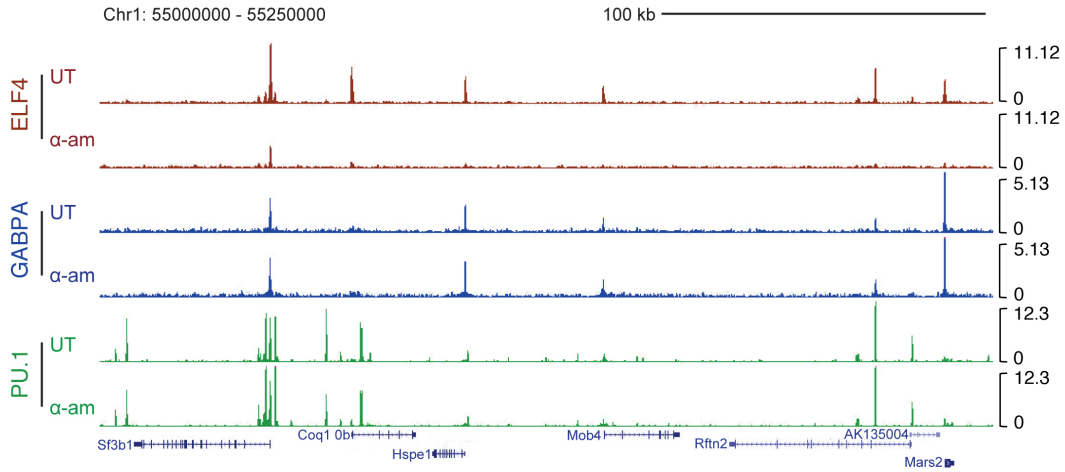


Figure 28. Representative snapshot.
A representative genomic region showing ELF4 binding loss upon alpha-amanitin treatment, while GABPa and PU.1 binding are not affected.

5.6 ETS-dependent recruitment of transcriptional and chromatin regulators to core promoters.

To obtain insight into the mechanism of action of ELF proteins in transcriptional activation, we used a DNA affinity purification approach coupled to mass spectrometry analysis. Briefly, we terminally labeled with biotin a 240 nt DNA fragment (-150/+50 relative to the annotated TSS) corresponding to the *Scamp2* gene promoter, which contains a canonical ELF site and is efficiently bound *in vivo* (**Figure 29**).

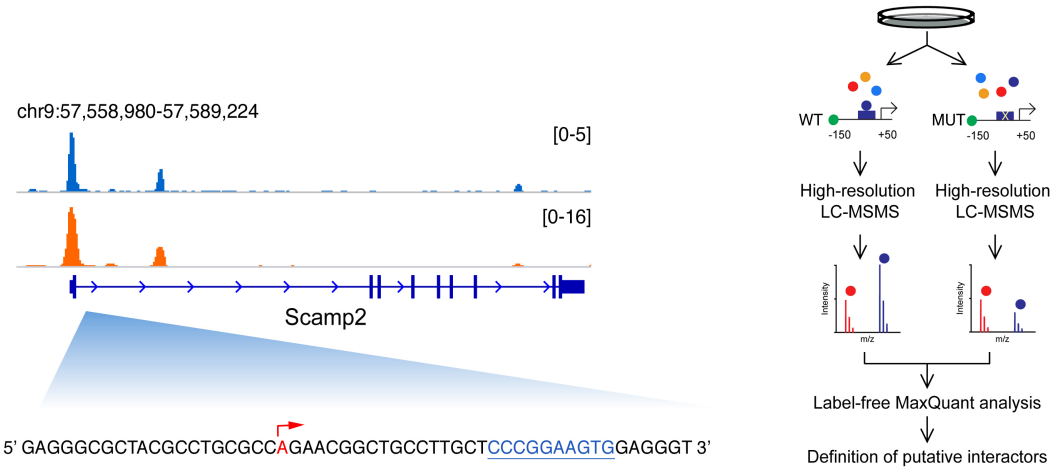


Figure 29. Snapshot and experimental scheme.
On the left is shown the snapshot of the *Scamp2* gene showing TSS-proximal ELF1 and ELF4 binding. The TSS (red) and the ELF site (light blue, underlined) are highlighted in the DNA sequence below. On the right is shown the experimental scheme of the DNA-mediated pull-down experiment using baits containing either a wild-type (WT) or a mutated (Mut) ETS motif, in a label-free experiment. Unlabeled nuclear extract was incubated in parallel with biotin-conjugated oligos, either wild-type (left) or mutated (right) in the ETS site. Proteins from the two individual experiments were analyzed separately by high-resolution LC-MSMS. Proteins in each sample were quantified with label-free algorithm contained in the MaxQuant software suite, whereby the intensity of corresponding peptides are compared between individual spectra. Specific interactors (i.e. blue line) show a higher LFQ intensity value in the WT than in the Mut pull-down, where binding is reduced or abrogated due to the mutation in the ETS site. Peptide peaks from background proteins (i.e. red line), instead, have equal LFQ intensity in the two experiments, with no effect of the mutation.

As a control, we generated a probe with a point mutation in the ELF site. Triplicate experiments were set up in which wild type and mutant probes were separately incubated with a nuclear lysate. Biotinylated DNA fragments were retrieved with streptavidin paramagnetic beads and extensively washed before analyzing bound proteins by label-free mass spectrometry (**Figure 29**).

We identified 1353 proteins, the majority (1149/1353, 84.9%) of which were equally pulled down using the wild type or the ETS sites mutant probe. The 204 proteins specifically enriched in the pull-down with the wild type probe were considered putative ELF interactors (**Figure 30**).

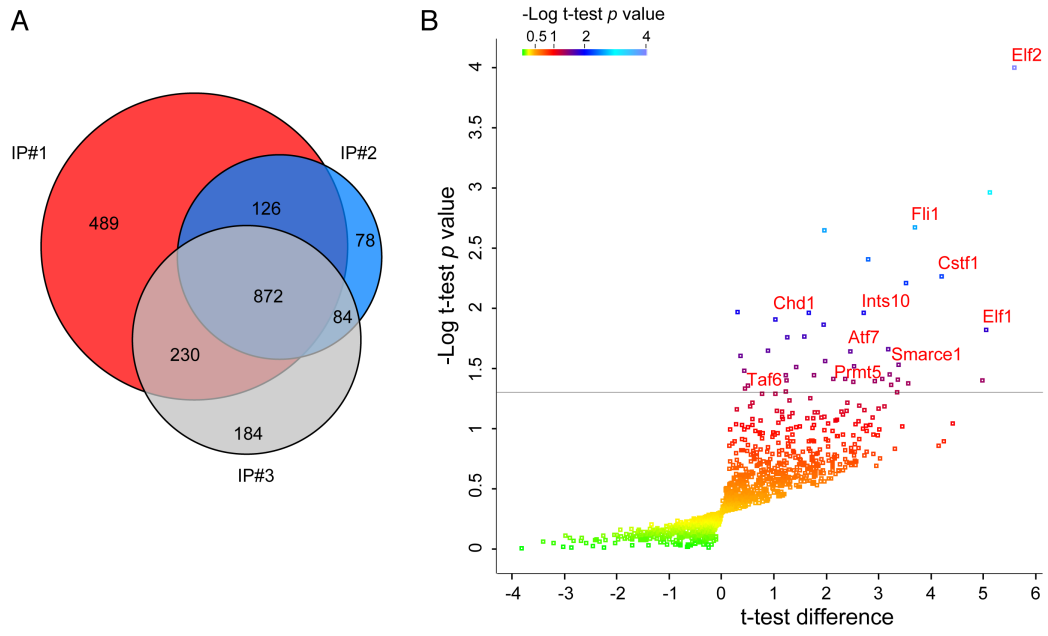


Figure 30. Identification of proteins recruited to core promoters in an ELF-dependent manner.
A) Overlap of proteins identified -with at least two peptides (peptides >1) of which at least one unique (unique >0)- among the 3 pull-down experiments carried out with wild type (WT) biotinylated oligo DNA. **B)** Volcano plot of ELF interactors; the proteins significantly enriched in the WT oligo DNA relative to the mutated oligo DNA pull-downs are above the grey line (one-side t-test p-value ≤0.05). Some transcriptional co-regulator proteins are highlighted.

In addition to the retrieval of ETS proteins (ELFs and FLI1), we identified a number of proteins involved in different aspects of transcriptional regulation (**Figure 30** and **31)**, including chromatin remodelers (BAF57 and CHD1) and proteins affecting RNA Pol II processivity (NELFB, the integrator complex subunit INTS10, PRMT5 and CSTF1/2). NELFB and Integrator complex subunits interact to control RNA Pol II release from promoter proximal pausing (Gardini et al., 2014; Stadelmayer et al., 2014) and enable productive transcriptional elongation. CHD1 was previously shown to bind the TSS of highly active genes and to remove the nucleosomal barrier downstream of the TSS, thus enabling RNA Pol II promoter escape (Skene et al., 2014) and maintaining high-level transcription (Guzman-Ayala et al., 2015).

Gene category and name	Ratio Value			Statistical significance
	exp 1	exp 2	exp 3	
Chromatin remodelers				
Smarce1/Baf57	∞	nd	∞	*p*=0.036
Ruvbl2/Pontin	6.31	∞	∞	n.s.
Brd2	0.89	∞	4.76	n.s.
Chd1	1.47	2.73	2.09	*p*=0.012
Chromatin modifiers				
Prmt5	∞	nd	∞	*p*=0.041
Prmt1	1.61	∞	∞	n.s.
Ash2l	0.42	∞	7.46	n.s.
Transcriptional machinery				
Polr2b/RPB2	0.78	∞	∞	n.s.
Med17	0.41	∞	∞	n.s.
Taf6	1.04	1.84	1.47	*p*=0.044
Transcription pausing, elongation termination				
Nelfb	∞	nd	∞	n.s.
Ints10	∞	∞	∞	*p*=0.011
Cstf1	∞	∞	∞	*p*=0.005
RNA export				
Nup98	2.42	∞	∞	n.s.
Transcription factors				
Fli1	∞	∞	∞	*p*=0.002
Elf2	∞	∞	∞	*p*=0.0001
Atf7	∞	nd	∞	*p*=0.023
Elf1	4.89	∞	∞	*p*=0.015
Tcf12	∞	∞	1.45	n.s.

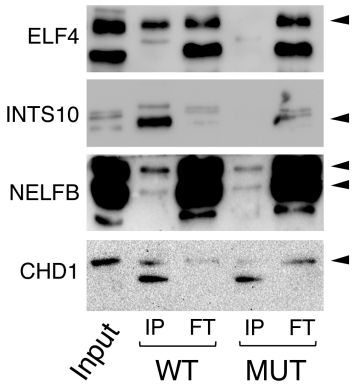


Figure 31. ELF-mediated recruitment of transcriptional and chromatin regulators to core promoters. Upper panel. Shortlist of putative ELF interactors. The ratio value (calculated dividing the LFQ intensity of each proteins in WT oligo DNA PD over the mutated oligo DNA PD) is reported for each protein in all biological replicates. ∞ indicates ratio where the LFQ intensity of a protein was equal to 0 in the mutated oligo DNA PD; nd: not detected protein. Proteins significantly enriched are indicated with their respective p-values. The other proteins indicated were not statistically significant but present in at least two out of three biological experiments in the Top 30% of each independent LFQ proteins ratio distribution. Lower panel. Western blot analysis of selected proteins pulled down in the affinity purification experiments with wild-type (WT) and mutated (MUT) baits. Arrowheads indicate specific bands. PD: Pull-down, FT: Flow-through.

We first validated some of the affinity-isolated proteins by western blot in independent experiments. ELF4 was used as a specific control and in fact it was selectively pulled down with the wild type probe (**Figure 31**). NELFB, INTS10 and CHD1 were relatively

enriched in affinity purifications with the wild type probe (**Figure 31**), with ratios that were overall consistent with those observed in the mass spectrometry analysis. In particular, recruitment of INTS10 was very selectively dependent on an intact ELF site.

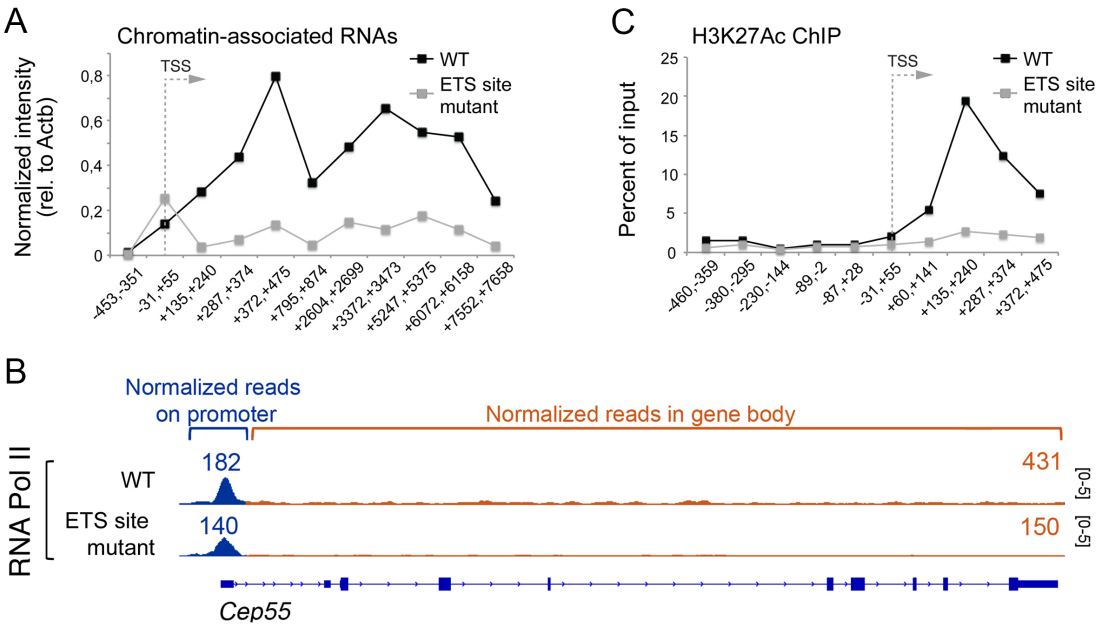


Figure 32. ELFs absence impact on nascent RNA expression, RNA Pol II recruitment and H3K27ac deposition.
A) Nascent (chromatin-associated) transcripts at the *Cep55* gene in wild type and ETS site mutant Raw264.7 cells. Numbers (relative to the mapped TSS) correspond to the PCR amplicon. The grey line represents the TSS. **B)** Snapshot of the *Cep55* gene from a total RNA Pol II ChIP-seq in wild type and ETS site mutant Raw264.7 cells. The normalized reads corresponding to the paused (promoter-proximal) RNA Pol II peak and to the gene body, are indicated. **C)** Histone H3K27 acetylation at the *Cep55* gene in cells with a wild type or mutant ETS site. The grey line represents the TSS.

The ELF-dependent association of proteins controlling promoter escape of RNA Pol II (CHD1) and transcriptional elongation (NELFB and Integrator) prompted us to further characterize the functional consequences of ETS site editing at the *Cep55* gene (**Figure 24**). Since we found that nascent RNAs were down-regulated all over the gene body (**Figure 32A**), we performed an RNA Pol II ChIP-seq to discriminate between elongation and initiation defects. This analysis revealed a similar occupancy of the TSS-proximal region by the paused RNA Pol II regardless of ETS site mutation, but a 2.9-fold reduction in the number of intragenic reads (**Figure 32B**), indicating a defective entry of RNA Pol II inside the coding region. Remarkably, histone acetylation downstream of the TSS was almost completely abrogated in ETS site-mutant cells (**Figure 32C**).

6. Discussion

The functional specificity of the thousands of *cis*-regulatory elements that control the transcriptional output of higher eukaryotic cells is determined by the characteristic combination and arrangement of TF motifs that eventually establish the unique profile of activity of each one of them (Farley et al., 2015). The focus of this study was the repertoire of constitutively active enhancers and promoters, with the objective of identifying the possible existence of shared molecular determinants of their high activity. The data we obtained suggest that irrespective of their role in housekeeping or tissue-specific transcription, a large panel of constitutively active promoters and enhancers may rely on a more limited number of broadly used TFs (such as the ELF proteins) than could in principle be expected based on their functional divergence.

ETS proteins are a large family of TFs whose first identifiable ancestor is IBP39, a protein that in the primitive eukaryote *Trichomonas vaginalis* binds both the Inr core promoter motif and RNA Pol II (Schumacher et al., 2003). Therefore ETS proteins may have originally evolved as molecular bridges between core promoters and the transcriptional machinery, thus directly enabling early steps in transcription. The data reported in this study suggest that while some ETS proteins (such as PU.1 and FLI1) acquired in mammals tissue-specific functions mainly related to enhancer specification (Ghisletti et al., 2010; Heinz et al., 2010; Lichtinger et al., 2012), some others (such as ELFs and GABPA) may have retained ancestral functions related to core promoter regulation and specifically to the maintenance of the high activity of a subset of transcriptionally competent enhancers and promoters. Such functional differences among ETS family members are reflected by the fine specificity of the DNA motifs recognized by each one of them, which results in completely different genomic distribution profiles, with

enhancer-biased and promoter-biased ETS proteins at the two opposite ends of the spectrum.

The tight link between ELF proteins and efficient transcription is underscored by the fact that the relatively small fraction of bound enhancers was also the one with by far the strongest activity (as inferred by both histone acetylation and RNA Pol II levels). The close proximity of ELF-specific ETS sites to the TSS and the negative impact of RNA Pol II depletion on ELF occupancy, are both consistent with the occurrence of a tight interaction between DNA bound ELFs and the transcriptional machinery. However, the precise step(s) at which promoter-associated ETS proteins act to control transcription will require additional investigation. Our proteomic and genomic data hint at a role of ELFs in the control of transcriptional elongation rather than initiation. Specifically, RNA Pol II density in the body of a gene with a mutated TSS-proximal ETS site was strongly reduced while loading at the TSS was not or was only marginally affected. The identification of NELF and an Integrator complex subunit as ELF interactors is consistent with a direct impact of TSS-proximal ELFs on promoter escape by RNA Pol II. Moreover, recruitment of the chromatin remodeler CHD1, which removes intragenic, promoter-proximal nucleosomes (Skene et al., 2014) may critically contribute to transcriptional elongation at genes (such as those bound by ELFs, **Figure 14**) with a prominent +1 nucleosome. However, additional mechanisms linking ELFs to transcriptional control and specifically to RNA Pol II recruitment and initiation cannot be ruled out, particularly considering that unstable and weak interactions may have been overlooked by the affinity purification strategy we used. In fact, the sufficiency at activating transcription of a minimal synthetic promoter containing only an ELF site and no strong core promoter elements points to direct interactions with the transcriptional machinery that may by themselves be sufficient to recruit RNA Pol II and initiate transcription.

The general principle that can be drawn from these data is that the cooptation of a limited number of TFs (such as the ELFs) capable of strongly facilitating transcription may

represent a transversal strategy broadly adopted across cell types to equip *cis*-regulatory elements with disparate functional roles and specificity with the ability to efficiently promote transcription. Conversely, the absence of the same motifs in a *cis*-regulatory region may represent a pre-requisite for its tighter regulation in response to specific micro-environmental or developmental cues. Overall, when considering cell type-specific *cis*-regulatory elements, the emerging paradigm is that they are generated by a combination of motifs recognized by tissue-restricted TFs (commonly endowed with the ability to displace nucleosomes and generate accessible chromatin)(Glass and Natoli, 2016; Zaret and Carroll, 2011) and motifs for TFs that impart specific functional properties such as inducible (e.g. NF-kB, AP-1) or constitutive (such as ELFs) activity to that specific element. Finally, the data shown in our study contribute to provide a mechanistic framework for previous observations linking ubiquitously or broadly expressed TFs to critical tissue-restricted functions (Gilmour et al., 2014).

References

Allen, B.L. and Taatjs, D.J. (2015). The Mediator complex: a central integrator of transcription. Nature Reviews Volume 16, 155-166

Austenaa, L.M., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B.A., de Pretis, S.*, et al.* (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. Molecular Cell *60*, 460-474.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Barbash, Z.S., Weissman, J.D., Campbell, J.A., Jr., Mu, J., and Singer, D.S. (2013). Major histocompatibility complex class I core promoter elements are not essential for transcription in vivo. Molecular and cellular biology *33*, 4395-4407.

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Molecular Cell *54*, 844-857.

Bassuk, A.G., Barton, K.P., Anandappa, R.T., Lu, M.M., and Leiden, J.M. (1998). Expression pattern of the Ets-related transcription factor Elf-1. Mol Med *4*, 392-401.

Baylin, S.B., and Schuebel, K.E. (2007). Genomic biology: the epigenomic era opens. Nature *448*, 548-549.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., 3rd, and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nature biotechnology *24*, 1429-1435.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279-290.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev *11*, 3020-3031.

Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev *16*, 2583-2592.

Calero-Nieto, F.J., Wood, A.D., Wilson, N.K., Kinston, S., Landry, J.R., and Gottgens, B. (2010). Transcriptional regulation of Elf-1: locus-wide analysis reveals four distinct promoters, a tissue-specific enhancer, control by PU.1 and the importance of Elf-1 downregulation for erythroid maturation. Nucleic acids research *38*, 6363-6374.

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

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Celada, A., Borras, F.E., Soler, C., Lloberas, J., Klemsz, M., van Beveren, C., McKercher, S., and Maki, R.A. (1996). The transcription factor PU.1 is involved in macrophage proliferation. J Exp Med *184*, 61-69.

Cesaroni, M., Cittaro, D., Brozzi, A., Pelicci, P.G., and Luzi, L. (2008). CARPET: a web-based package for the analysis of ChIP-chip and expression tiling data. Bioinformatics *24*, 2918-2920.

Charlot, C., Dubois-Pot, H., Serchov, T., Tourrette, Y., and Wasylyk, B. (2010). A review of post-translational modifications and subcellular localization of Ets transcription factors: possible connection with cancer and involvement in the hypoxic response. Methods Mol Biol *647*, 3-30.

Choi, H.J., Geng, Y., Cho, H., Li, S., Giri, P.K., Felio, K., and Wang, C.R. (2011). Differential requirements for the Ets transcription factor Elf-1 in the development of NKT cells and NK cells. Blood *117*, 1880-1887.

Consortium, E. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res *10*, 1794-1805.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., *et al.* (2015). Ensembl 2015. Nucleic Acids Res *43*, D662-669.

Daeton, A.M. and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Danino, Y.M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. Biochim Biophys Acta *1849*, 1116-1131.

Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H. (2008). The functional consequences of alternative promoter use in mammalian genomes. Trends Genet *24*, 167-177.

De Val, S., Chi, N.C., Meadows, S.M., Minovitsky, S., Anderson, J.P., Harris, I.S., Ehlers, M.L., Agarwal, P., Visel, A., Xu, S.M., *et al.* (2008). Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. Cell *135*, 1053-1064.

Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes & development *19*, 2418-2423.

Deng, W., and Roberts, S.G. (2007). TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma *116*, 417-429.

Diaferia, G.R., Balestrieri, C., Prosperini, E., Nicoli, P., Spaggiari, P., Zerbi, A., and Natoli, G. (2016). Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. Embo J *35*, 595-617.

Dreos, R., Ambrosini, G., Cavin Perier, R., and Bucher, P. (2013). EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. Nucleic Acids Res *41*, D157-164.

Ernst, P., Hahm, K., Trinh, L., Davis, J.N., Roussel, M.F., Turck, C.W., and Smale, S.T. (1996). A potential role for Elf-1 in terminal transferase gene regulation. Molecular and cellular biology *16*, 6121-6131.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325-328.

Fazal, F.M., Meng, C.A., Murakami, K., Kornberg, R.D., and Block, S.M. (2015). Real-time observation of the initiation of RNA polymerase II transcription. Nature *525*, 274-277.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I.*, et al.* (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res *22*, 2399-2408.

FitzGerald, P.C., Shlyakhtenko, A., Mir, A.A., and Vinson, C. (2004). Clustering of DNA sequences in human promoters. Genome Res *14*, 1562-1574.

FitzGerald, P.C., Sturgill, D., Shyakhtenko, A., Oliver, B., and Vinson, C. (2006). Comparative genomics of Drosophila and human core promoters. Genome Biol *7*, R53.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M.*, et al.* (2014). A promoter-level mammalian expression atlas. Nature *507*, 462-470.

Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and Sandelin, A. (2008). A code for transcription initiation in mammalian genomes. Genome research *18*, 1-12.

Gallagher, P.G., Nilson, D.G., Wong, C., Weisbein, J.L., Garrett-Beal, L.J., Eber, S.W., and Bodine, D.M. (2005). A dinucleotide deletion in the ankyrin promoter alters gene expression, transcription initiation and TFIID complex formation in hereditary spherocytosis. Hum Mol Genet *14*, 2501-2509.

Gallant, S., and Gilkeson, G. (2006). ETS transcription factors and regulation of immunity. Arch Immunol Ther Exp (Warsz) *54*, 149-163.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. Molecular Cell *56*, 128-139.

Garraway, I.P., Semple, K., and Smale, S.T. (1996). Transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene requires a specific core promoter structure. Proc Natl Acad Sci U S A *93*, 4336-4341.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96-100.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L.*, et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

Gilmour, J., Assi, S.A., Jaegle, U., Kulu, D., van de Werken, H., Clarke, D., Westhead, D.R., Philipsen, S., and Bonifer, C. (2014). A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification. Development *141*, 2391-2401.

Glass, C.K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. Nature immunology *17*, 26-33.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet *11*, 549-558.

Gordon, S., and Taylor, P.R. (2005). Monocyte and macrophage heterogeneity. Nat Rev Immunol *5*, 953-964.

Grall, F., Gu, X., Tan, L., Cho, J.Y., Inan, M.S., Pettit, A.R., Thamrongsak, U., Choy, B.K., Manning, C., Akbarali, Y.*, et al.* (2003). Responses to the proinflammatory cytokines interleukin-1 and tumor necrosis factor alpha in cells derived from rheumatoid synovium and other joint tissues involve nuclear factor kappaB-mediated induction of the Ets transcription factor ESE-1. Arthritis and rheumatism *48*, 1249-1260.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Guzman-Ayala, M., Sachs, M., Koh, F.M., Onodera, C., Bulut-Karslioglu, A., Lin, C.J., Wong, P., Nitta, R., Song, J.S., and Ramalho-Santos, M. (2015). Chd1 is essential for the high transcriptional output and rapid growth of the mouse epiblast. Development *142*, 118-127.

Haberle, V., and Lenhard, B. (2016). Promoter architectures and developmental gene regulation. Semin Cell Dev Biol *57*, 11-23.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M.*, et al.* (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381-385.

Hallikas, O., and Taipale, J. (2006). High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. Nature protocols *1*, 215-222.

Hariharan, N., and Perry, R.P. (1990). Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. Proc Natl Acad Sci U S A *87*, 1526-1530.

Hart, A., Melet, F., Grossfeld, P., Chien, K., Jones, C., Tunnacliffe, A., Favier, R., and Bernstein, A. (2000). *Fli-1* is required for murine vascular and megakaryocytic development and is hemizygously deleted in patiens with thrombocytopenia. Immunity, Vol. 13, 167-177.

He, X., Chatterjee, R., John, S., Bravo, H., Sathyanarayana, B.K., Biddie, S.C., FitzGerald, P.C., Stamatoyannopoulos, J.A., Hager, G.L., and Vinson, C. (2013). Contribution of nucleosome binding preferences and co-occurring DNA sequences to transcription factor binding. BMC Genomics *14*, 428.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell *38*, 576-589.

Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. Nature *503*, 487-492.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144-154.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S.*, et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods *6*, 283-289.

Hieb, A.R., Halsey, W.A., Betterton, M.D., Perkins, T.T., Kugel, J.F., and Goodrich, J.A. (2007). TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. Journal of molecular biology *372*, 619-632.

Hoey, T., Dynlacht, B.D., Peterson, M.G., Pugh, B.F., and Tjian, R. (1990). Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. Cell *61*, 1179-1186.

Hollenhorst, P.C., McIntosh, L.P., and Graves, B.J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem *80*, 437-471.

Hollenhorst, P.C., Shah, A.A., Hopkins, C., and Graves, B.J. (2007). Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev *21*, 1882-1894.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H.*, et al.* (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome research *21*, 182-192.

Jin, V.X., Singer, G.A., Agosto-Perez, F.J., Liyanarachchi, S., and Davuluri, R.V. (2006). Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics *7*, 114.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384-388.

Juang, Y.T., Solomou, E.E., Rellahan, B., and Tsokos, G.C. (2002). Phosphorylation and O-linked glycosylation of Elf-1 leads to its translocation to the nucleus and binding to the promoter of the TCR zeta-chain. Journal of immunology *168*, 2865-2871.

Juven-Gershon, T., Hsu, J.Y., and Kadonaga, J.T. (2008). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes Dev *22*, 2823-2830.

Kadonaga, J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247-257.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol *1*, 40-51.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C.*, et al.* (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Molecular cell *51*, 310-325.

Kawaji, H., Frith, M.C., Katayama, S., Sandelin, A., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. (2006). Dynamic usage of transcription start sites within core promoters. Genome Biol *7*, R118.

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

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kueh, H.Y., Champhekar, A., Nutt, S.L., Elowitz, M.B., and Rothenberg, E.V. (2013). Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. Science *341*, 670-673.

Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief Bioinform *14*, 144-161.

Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Soboleva, A.V., Kasianov, A.S., Ashoor, H., Ba-Alawi, W., Bajic, V.B., Medvedeva, Y.A., Kolpakov, F.A.*, et al.* (2016). HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res *44*, D116-125.

Kutach, A.K., and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Molecular and cellular biology *20*, 4754-4764.

Lacorazza, H.D., Miyazaki, Y., Di Cristofano, A., Deblasio, A., Hedvat, C., Zhang, J., Cordon-Cardo, C., Mao, S., Pandolfi, P.P., and Nimer, S.D. (2002). The ETS protein MEF plays a critical role in perforin gene expression and the development of natural killer and NK-T cells. Immunity *17*, 437-449.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes & development *12*, 34-44.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S.*, et al.* (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943-949.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell *159*, 1312-1326.

Lawrence, T., and Natoli, G. (2011). Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nat Rev Immunol *11*, 750-761.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Molecular and cellular biology *25*, 9674-9686.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lichtinger, M., Ingram, R., Hannah, R., Muller, D., Clarke, D., Assi, S.A., Lie, A.L.M., Noailles, L., Vijayabaskar, M.S., Wu, M.*, et al.* (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. Embo J *31*, 4318-4333.

Lidor Nili, E., Field, Y., Lubling, Y., Widom, J., Oren, M., and Segal, E. (2010). p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. Genome research *20*, 1361-1368.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harbor symposia on quantitative biology *42 Pt 2*, 1047-1051.

Link, V.M., Gosselin, D., and Glass, C.K. (2015). Mechanisms Underlying the Selection and Function of Macrophage-Specific Enhancers. Cold Spring Harbor symposia on quantitative biology *80*, 213-221.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends Genet *27*, 465-474.

Malik, S. and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nature Reviews Volume 11, 761-772.

Mancino, A., Termanini, A., Barozzi, I., Ghisletti, S., Ostuni, R., Prosperini, E., Ozato, K., and Natoli, G. (2015). A dual cis-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. Genes Dev *29*, 394-408.

Marbach-Bar, N., Bahat, A., Ashkenazi, S., Golan-Mashiach, M., Haimov, O., Wu, S.Y., Chiang, C.M., Puzio-Kuter, A., Hirshfield, K.M., Levine, A.J.*, et al.* (2016). DTIE, a novel core promoter element that directs start site selection in TATA-less genes. Nucleic acids research *44*, 1080-1094.

Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. Molecular cell *7*, 509-515.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet *7*, 29-59.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R.*, et al.* (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res *44*, D110-115.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell reports *12*, 1184-1195.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proc Natl Acad Sci U S A *107*, 22534-22539.

Muller, F., Demeny, M.A., and Tora, L. (2007). New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. The Journal of biological chemistry *282*, 14685-14689.

Nan, X., Cross, S., and Bird, A. (1998). Gene silencing by methyl-CpG-binding proteins. Novartis Found Symp *214*, 6-16; discussion 16-21, 46-50.

Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. Immunity *33*, 12-24.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157-171.

Panne, D. (2008). The enhanceosome. Curr Opin Struct Biol *18*, 236-242.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. Cell *129*, 1111-1123.

Parry, T.J., Theisen, J.W., Hsu, J.Y., Wang, Y.L., Corcoran, D.L., Eustice, M., Ohler, U., and Kadonaga, J.T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev *24*, 2013-2018.

Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol *7*, R78.

Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev *5*, 1935-1945.

Pugh, B.F., and Venters, B.J. (2016). Genomic Organization of Human Transcription Initiation Complexes. PLoS One *11*, e0149339.

Qiao, Y., Giannopoulou, E.G., Chan, C.H., Park, S.H., Gong, S., Chen, J., Hu, X., Elemento, O., and Ivashkiv, L.B. (2013). Synergistic activation of inflammatory cytokine genes by interferon-gamma-induced chromatin remodeling and toll-like receptor signaling. Immunity *39*, 454-469.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics *47*, 11 12 11-34.

Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS genetics *7*, e1001274.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114-128.

Reese, J.C. (2003). Basal transcription factors. Curr Opin Genet Dev *13*, 114-118.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. The Journal of biological chemistry *280*, 24731-24737.

Roider, H.G., Lenhard, B., Kanhere, A., Haas, S.A., and Vingron, M. (2009). CpG-depleted promoters harbor tissue-specific transcription factor binding signals--implications for motif overrepresentation analyses. Nucleic acids research *37*, 6305-6315.

Romanoski, C.E., Link, V.M., Heinz, S., and Glass, C.K. (2015). Exploiting genomics and natural genetic variation to decode macrophage enhancers. Trends in immunology *36*, 507-518.

Rosenbauer, F., Owens, B.M., Yu, L., Tumang, J.R., Steidl, U., Kutok, J.L., Clayton, L.K., Wagner, K., Scheller, M., Iwasaki, H*., et al.* (2006). Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat Genet *38*, 27-37.

Rosmarin, A.G., Resendes, K.K., Yang, Z., McMillan, J.N., and Fleming, S.L. (2004). GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein-protein interactions. Blood cells, molecules & diseases *32*, 143-154.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nature Reviews Volume 16, 129-143.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods *11*, 783-784.

Scheiber, M.N., Watson, P.M., Rumboldt, T., Stanley, C., Wilson, R.C., Findlay, V.J., Anderson, P.E., and Watson, D.K. (2014). FLI1 expression is correlated with breast cancer cellular growth, migration, and invasion and altered gene expression. Neoplasia *16*, 801-813.

Schumacher, M.A., Lau, A.O., and Johnson, P.J. (2003). Structural basis of core promoter recognition in a primitive eukaryote. Cell *115*, 413-424.

Shir-Shapira, H., Sharabany, J., Filderman, M., Ideses, D., Ovadia-Shochat, A., Mannervik, M., and Juven-Gershon, T. (2015). Structure-Function Analysis of the Drosophila melanogaster Caudal Transcription Factor Provides Insights into Core Promoter-preferential Activation. The Journal of biological chemistry *290*, 17293-17305.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife *3*, e02042.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem Sci *39*, 381-399.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103-113.

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S.*, et al.* (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nat Commun *5*, 5531.

Tamura, T., Thotakura, P., Tanaka, T.S., Ko, M.S., and Ozato, K. (2005). Identification of target genes and a unique cis element regulated by IRF-8 in developing macrophages. Blood *106*, 1938-1947.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tippmann, S.C., Ivanek, R., Gaidatzis, D., Scholer, A., Hoerner, L., van Nimwegen, E., Stadler, P.F., Stadler, M.B., and Schubeler, D. (2012). Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. Mol Syst Biol *8*, 593.

Tong, A.J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. Cell *165*, 165-179.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol *31*, 46-53.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods.

van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014). In search of the determinants of enhancer-promoter interaction specificity. Trends Cell Biol *24*, 695-702.

Verrijzer, C.P., Chen, J.L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. Cell *81*, 1115-1125.

Wang, Y.L., Duttke, S.H., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev *28*, 1550-1555.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., *et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. Embo J *29*, 2147-2160.

Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Mol Cell Biol *17*, 2973-2984.

Wright, K.J., Marr, M.T., 2nd, and Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proc Natl Acad Sci U S A *103*, 12347-12352.

Wu, H., Xiao, Y., Zhang, S., Ji, S., Wei, L., Fan, F., Geng, J., Tian, J., Sun, X., Qin, F., *et al.* (2013). The Ets transcription factor GABP is a component of the hippo pathway essential for growth and antioxidant defense. Cell reports *3*, 1663-1677.

Yamada, T., Park, C.S., Mamonkin, M., and Lacorazza, H.D. (2009). Transcription factor ELF4 controls the proliferation and homing of CD8+ T cells via the Kruppel-like factors KLF4 and KLF2. Nat Immunol *10*, 618-626.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52-65.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556-559.

Zambelli, F., Pesole, G., and Pavesi, G. (2009). Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res *37*, W247-252.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev *25*, 2227-2241.

Zhang, D., Penttila, T.L., Morris, P.L., Teichmann, M., and Roeder, R.G. (2001). Spermiogenesis deficiency in mice lacking the Trf2 gene. Science *292*, 1153-1155.

Zhang, L., Yu, H., Wang, P., Ding, Q., and Wang, Z. (2013). Screening of transcription factors with transcriptional initiation activity. Gene *531*, 64-70.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7-18.