

# High constitutive activity of a broad panel of housekeeping and tissue-specific *cis*-regulatory elements depends on a subset of ETS proteins

Alessia Curina,<sup>1,5</sup> Alberto Termanini,<sup>2,5</sup> Iros Barozzi,<sup>1,4</sup> Elena Prosperini,<sup>1</sup> Marta Simonatto,<sup>1</sup> Sara Polletti,<sup>1</sup> Alessio Silvola,<sup>1</sup> Monica Soldi,<sup>1</sup> Liv Austenaa,<sup>1</sup> Tiziana Bonaldi,<sup>1</sup> Serena Ghisletti,<sup>2,6</sup> and Gioacchino Natoli<sup>1,3,6</sup>

<sup>1</sup>Department of Experimental Oncology, European Institute of Oncology (IEO), 20139 Milan, Italy; <sup>2</sup>Humanitas Clinical and Research Center, 20089 Rozzano-Milan, Italy; <sup>3</sup>Humanitas University, 20089 Rozzano-Milan, Italy

Enhancers and promoters that control the transcriptional output of terminally differentiated cells include cell type-specific and broadly active housekeeping elements. Whether the high constitutive activity of these two groups of *cis*-regulatory elements relies on entirely distinct or instead also on 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 base pairs (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 high-level constitutive transcription was suggested by multiple evidence: ELF sites enabled robust 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 suggest that the co-optation of a limited number of highly active transcription factors represents a broadly adopted strategy to equip both cell type-specific and housekeeping *cis*-regulatory elements with the ability to efficiently promote transcription.

[*Keywords*: ETS; enhancers; housekeeping genes; macrophages; transcription]

Supplemental material is available for this article.

Received November 2, 2016; revised version accepted February 3, 2017.

The transcriptional output of mammalian cells depends on the combined activity of thousands of *cis*-regulatory elements, mainly enhancers and promoters, that act as recruitment platforms for cell type-specific and broadly expressed transcription factors (TFs) (Spitz and Furlong 2012; Shlyueva et al. 2014). Depending on the combination of TF-binding motifs, regulatory genomic regions may work in a constitutive or regulated fashion, in the latter case being influenced by developmental or microenvironmental cues (Creyghton et al. 2010; Rada-Iglesias et al. 2011; Kaikkonen et al. 2013; Ostuni et al. 2013).

Defining features of active promoters and enhancers include high levels of histone acetylation and the ability to initiate transcription (De Santa et al. 2010; Kim et al.

2010; Wang et al. 2011). In fact, transcriptionally competent promoters and enhancers have an overall similar organization (Barozzi et al. 2014; Core et al. 2014) consisting of the variable combination of motifs recognized by sequence-specific TFs (which collaboratively recruit the RNA polymerase II [Pol II] initiation complex) and core promoter elements; namely, DNA sequences that enable initiation of RNA synthesis and are therefore located in close proximity to transcription initiation sites (Kadonaga 2012). RNA Pol II core promoter elements such as the TATA box (Lifton et al. 1978; Goldberg 1979), Initiator (Inr) (Smale and Baltimore 1989), and the TFIIB recognition elements (BREs) (Lagrange et al. 1998; Deng and Roberts 2005) were initially identified in classical mechanistic studies and then characterized in a global manner using genome-wide approaches that provided clues to

<sup>4</sup>Present address: Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>These authors contributed equally to this work.

Corresponding authors: [gioacchino.natoli@hunimed.eu](mailto:gioacchino.natoli@hunimed.eu), [serena.ghisletti@humanitasresearch.it](mailto:serena.ghisletti@humanitasresearch.it)

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.293134.116>.

© 2017 Curina et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

their distribution and abundance. The TATA box, which is recognized by the TBP subunit of TFIID, is associated with a well-defined transcription start site (TSS; or a cluster of closely spaced TSSs) located 28–34 base pairs (bp) downstream and is found at a small minority of metazoan promoters controlling highly regulated and tissue-specific genes (Carninci et al. 2006; Juven-Gershon et al. 2006; Ponjavic et al. 2006; Yang et al. 2007). Inr (Smale and Baltimore 1989) is a highly degenerate sequence whose 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).

However, core promoter elements are not the only motifs specifically enriched in close proximity to TSSs. Computational analyses of the DNA sequences in the entire complement of human gene promoters (Xie et al. 2005; FitzGerald et al. 2006) revealed that very few TF consensus sites are specifically overrepresented within 50 bp of mapped TSSs. This finding hints at the possibility that selected sequence-specific TFs (among the hundreds expressed in mammalian cells) may act as broadly used facilitators of early steps in transcription. The noncore promoter motif with the strongest enrichment in the immediate vicinity ( $<20$  bp) of the TSS was the consensus DNA-binding site for ETS proteins, a metazoan-restricted family of TFs that, in mice, includes 28 winged helix-loop-helix DNA-binding proteins involved in a variety of biological processes, including hematopoiesis and angiogenesis (for review, see Hollenhorst et al. 2011). The close proximity of ETS motifs to TSSs suggests their direct contribution to the control of core promoter activity. In line with this notion, the first identifiable ancestor of ETS proteins 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 evolved originally as molecular bridges between core promoters and the transcriptional machinery, thus directly enabling early steps in transcription.

In this study, 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 hematopoietic ETS family member PU.1 maintains the activity of the vast majority of macrophage-specific enhancers and a fraction of the promoters of macrophage-specific genes (Ghisletti et al. 2010; Heinz et al. 2010; Barozzi et al. 2014; Mancino et al. 2015; Goode et al. 2016). We found that ETS sites distinct from those recognized by PU.1 were significantly overrepresented in the TSS-proximal regions of constitutively and highly expressed genes. Consistently, chromatin immunoprecipitation (ChIP) combined with high-throughput sequencing (ChIP-seq) revealed that selected ETS proteins (ELF1, ELF4, and GABPA) but not others (FLI1 and PU.1) had a conspicuous binding preference for regions immediately adjacent to the TSSs of highly expressed genes. Unexpectedly, ELF proteins were also selectively associated with the constitutively acetylated

and transcriptionally active subset of macrophage-specific and PU.1-bound enhancers. 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.

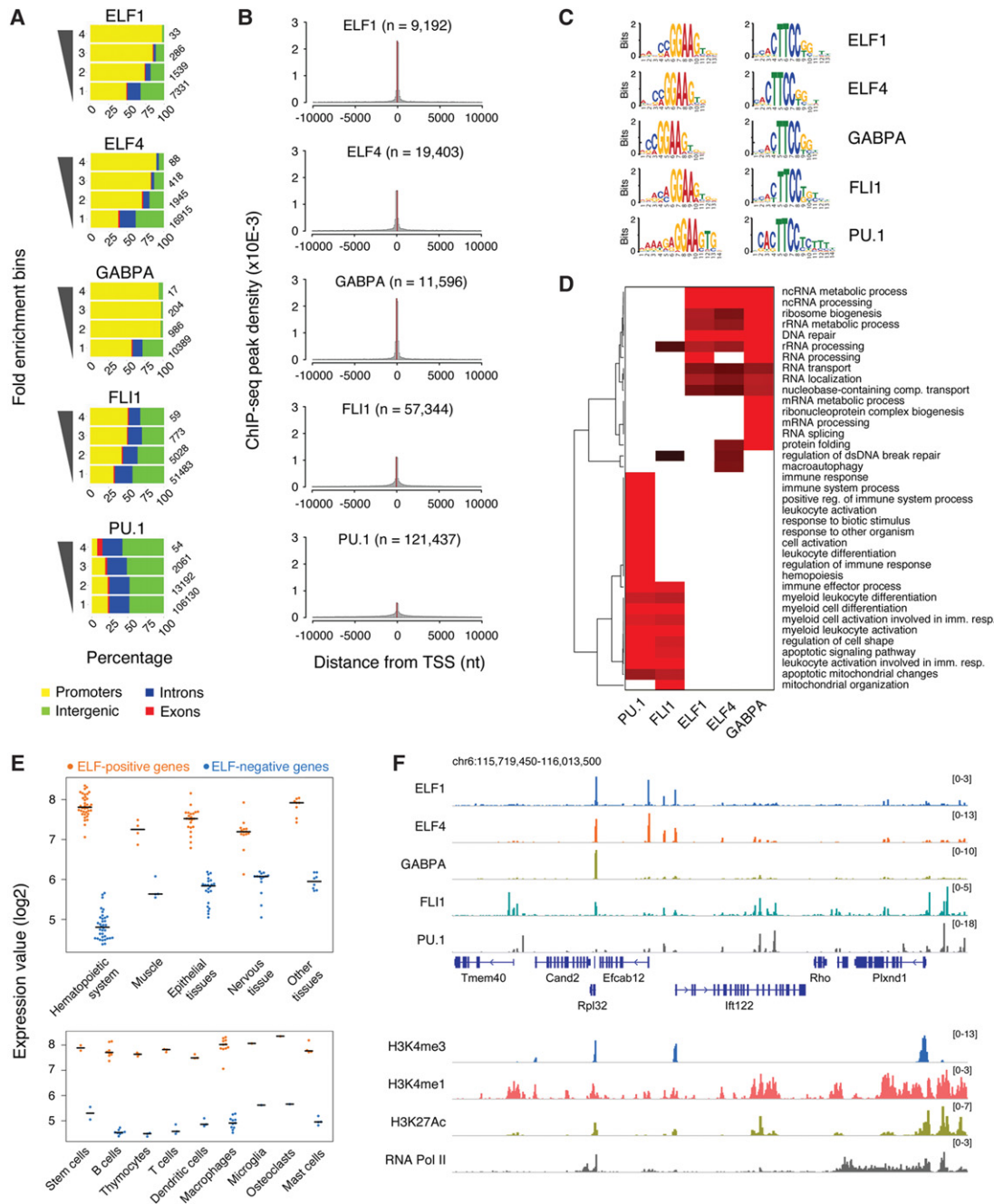
## Results

### *A subset of ETS proteins is associated with the TSSs 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 TSSs of all genes highly expressed in basal conditions (Supplemental Material; Supplemental Table S1). 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 ( $P \leq 3.6 \times 10^{-246}$ ) an ETS motif different from the one recognized by the myeloid-specific ETS protein PU.1 (Ghisletti et al. 2010; Heinz et al. 2010). Whereas all ETS proteins recognize a core GGAA/T motif via a shared 85-amino-acid ETS domain (Gallant and Gilkeson 2006), divergences in amino acid residues involved in DNA recognition determine different affinities for nucleotides 5' and 3' from such cores and define four classes (I–IV) (Wei et al. 2010).

Therefore, we used ChIP-seq to comparatively explore the genomic distribution of ETS proteins that we found to be 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; the only class IV family member, SPDEF, is not expressed in mouse macrophages). Some obvious trends were apparent irrespective of the threshold used. Specifically, ELF1, ELF4, and GABPA showed a significant preference for TSS-proximal regions ( $\text{TSS} \pm 2.5$  kb), which was increasingly more evident when considering peaks with a progressively higher score (Fig. 1A, fold enrichment bins 1–4). Conversely, PU.1 showed a strong preference for TSS-distal (intergenic and intragenic) regions, while FLI1 displayed an intermediate behavior, with  $\sim 50\%$  of the binding events called at intermediate and high stringency being associated with gene promoters (Fig. 1A). The genomic distribution of the ETS proteins analyzed in genomic windows of  $\pm 10$  kb surrounding the TSS is shown in Figure 1B.

A *de novo* motif discovery analysis (Fig. 1C) 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. Overall, some ETS proteins (GABPA, ELF1, and ELF4) have a



**Figure 1.** Genomic distribution of ETS family TFs. (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 fold enrichment (based on MACS2). Peak number is indicated on the right of each bin. (B) TSS-centered analysis of the genomic distribution of ETS family members. (C) ETS motifs (position weight matrices [PWMs]) identified by de novo motif discovery analysis (MEME). (D) Gene ontology (GO) categories associated with different ETS family members, as inferred from a GREAT (genomic regions enrichment of annotations tool) analysis of the individual ChIP-seq data sets. (E) ELF4 association with the TSSs of genes with high expression across tissues. Log<sub>2</sub> expression intensity of genes bound (orange dots) or not bound (blue dots) by ELF4 in different tissues. Each dot represents the average of the gene expression values of each tissue considered. (Black bar) Median. Cells of the hematopoietic system are shown in the bottom panel. (F) A representative snapshot of chromosome 6 showing the distinct genomic distribution of ETS family TFs and their relationship with histone marks.

significant binding preference for TSS-proximal regions that correlated with distinct DNA-binding specificities and were only partially recapitulated by previous in vitro studies. Among the promoter-biased ETS proteins,

GABPA and ELF1 (which bound a similar number of regions and were thus amenable to direct comparison) showed substantial overlap at promoters (4299 peaks) but also bound distinct sites (3278 GABPA-specific and

1031 ELF1-specific peaks) (Supplemental Fig. S1). These data suggest that differences in the composition of the underlying DNA sequence may contribute to the generation of functional specificity between ETS proteins with the ability to recognize an identical motif. Indeed, when directly compared with each other, GABPA-specific peaks showed an overrepresentation of CpG-rich motifs such as E2F and SP1 motifs, while ELF1-specific peaks showed an overrepresentation of the canonical ELF motifs and PU.1-like motifs (Supplemental Table S2). Additional layers of specificity may arise from the differential transcription activation ability of proteins recruited to the same site.

To determine whether differences in genomic distributions reflect distinct functional specificities of the ETS proteins analyzed, we used GREAT (genomic regions enrichment of annotations 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. 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 related mainly to myeloid cell differentiation and functions (Fig. 1D; Supplemental Table S3). 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 the regulation of housekeeping gene promoters.

Since genes with constitutive expression in macrophages include cell type-specific genes and broadly expressed genes with housekeeping functions, we tested whether the genes bound by ELFs are expressed across different tissues. To this aim, the expression of ELF4-positive genes was evaluated in almost 100 different tissues using publicly available data sets (Tippmann et al. 2012). In all tissues considered (Fig. 1E, top panel) and even more so in hematopoietic cells (Fig. 1E, 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.

Finally, we used a restrictive window of 250 bp upstream of the TSS to determine the subset of ELF4-bound genes expressed at a basal state in macrophages (Austena et al. 2015). Eighty-one percent (2524 out of 3117) of ELF4 peaks were associated with the promoters of expressed genes. Representative snapshots are shown in Figure 1F.

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

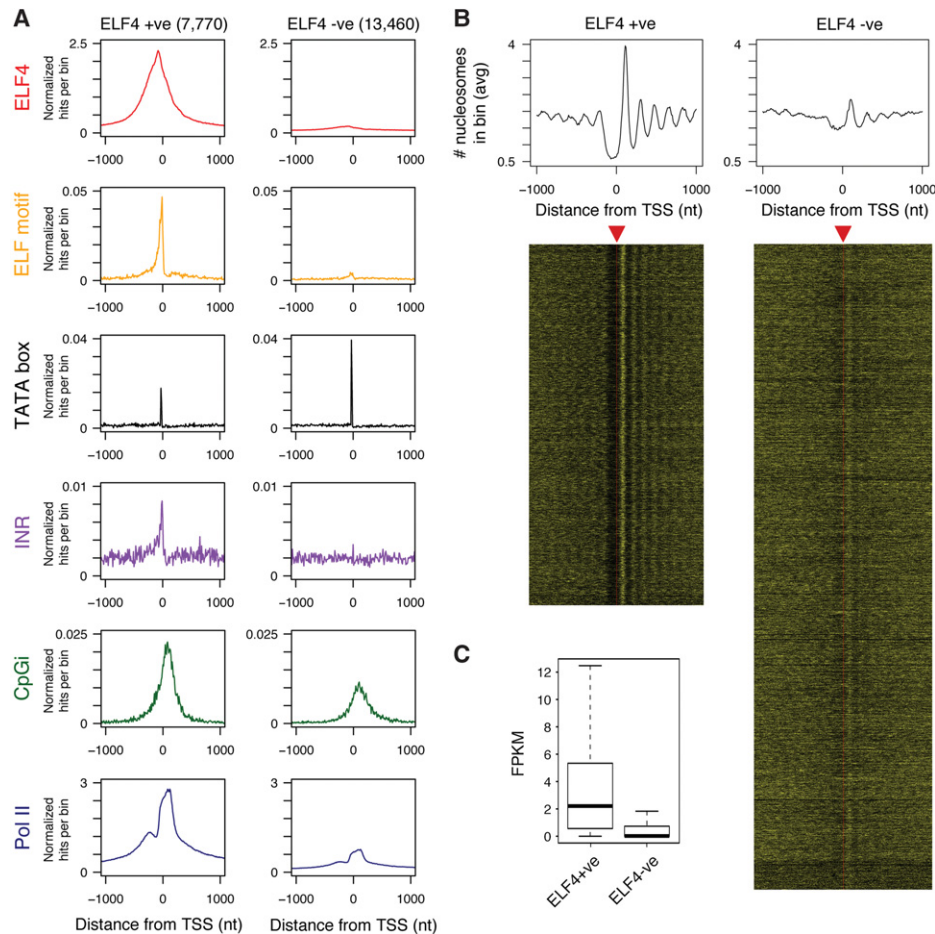
#### *Association of ELF proteins with the macrophage cis-regulatory repertoire*

We next characterized the features of the TSS-proximal regions associated with the promoter-biased ETS proteins

(ELFs and GABPA). RefSeq genes were divided into two groups based on their association with ELF4 (ELF4 genomic distribution extensively overlapping that of ELF1 and ELF4 antibodies is shown in Supplemental Fig. S3) and then analyzed for the occurrence of the TATA-box and INR motifs, the presence of a CpG island, and RNA Pol II binding (Fig. 2A). Consistent with previous data (FitzGerald et al. 2004, 2006), ETS sites occurred mainly near TSSs, mapping within 50–60 nucleotides (nt) from annotated TSSs. The TATA-box motif was depleted (but not completely excluded) from the TSSs of ELF-bound genes, while the INR motif showed the opposite trend, being almost absent at ELF-negative TSSs. The ELF-positive group was enriched for CpG islands and, consistent with the gene expression data shown above, was characterized by much higher constitutive levels of RNA Pol II.

An additional and most obvious difference between the two groups became evident when analyzing the organization of nucleosomes around the TSSs (Barozzi et al. 2014). ELF binding was almost invariably associated with intense nucleosome depletion upstream of the TSS and strong nucleosome phasing downstream from it, with a particularly prominent +1 nucleosome (Fig. 2B). Differences in nucleosome organization correlate with the different prevalence of CpG islands in the two groups, since a very high G + C content disfavors nucleosome assembly (Fenouil et al. 2012; Barozzi et al. 2014). Overall, ELF proteins showed a preference for GC-rich and INR-positive promoters characterized by well-defined nucleosome-depleted areas upstream of the TSS. Finally, when considering newly synthesized nascent transcripts (which reflect the actual transcriptional activity), ELF-associated genes were expressed at higher levels than the ELF-negative ones, whose activity was below the detection threshold in many cases (Fig. 2C). ELF-negative genes were also relatively enriched for tissue-specific genes. Using a mouse tissue-specific list of genes (from Pattern Gene Database version 10), we found that 4597 genes out of 10,610 (43.3%) of the ELF4-negative subsets and 1407 genes out of 6292 (22.4%) of the ELF4-positive subsets were tissue specific ( $P < 2.2 \times 10^{-16}$  using a two-sample test for equality of proportions with continuity correction).

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 divided TSS-distal and TSS-proximal PU.1-bound regions based on their association with ELFs. At TSS-distal regions, ELF binding was associated with significantly higher levels of H3K27ac, H3K4me1, RNA Pol II, and even PU.1 than those observed at ELF-negative *cis*-regulatory elements (Fig. 3A, top panels). GREAT analysis of TSS-distal



**Figure 2.** Preferential binding of ELF proteins to active, GC rich, and nucleosome-depleted TSSs. (A) The TSSs ( $\pm 1000$  nt) of RefSeq genes were divided into 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. (B) Nucleosome profiles at ELF-positive (*left*) and ELF-negative (*right*) TSSs. The *top* panels show the distribution of the midpoints of nucleosomal sequencing fragments centered on the annotated TSSs of RefSeq genes. The same information is shown *below* as heat maps. (C) Expression levels (shown as FPKM [fragments per kilobase per million mapped fragments]) of nascent transcripts in untreated macrophages of ELF4-negative and ELF4-positive genes as identified in A. Nascent transcript data are from a previously published data set (Austena et al. 2015).

PU.1-positive regions indicated that both the ELF-positive and the ELF-4-negative groups were associated with ontology terms related to myeloid cell differentiation and function (Supplemental Table S4), suggesting that these putative enhancers are enriched for macrophage-specific *cis*-regulatory regions.

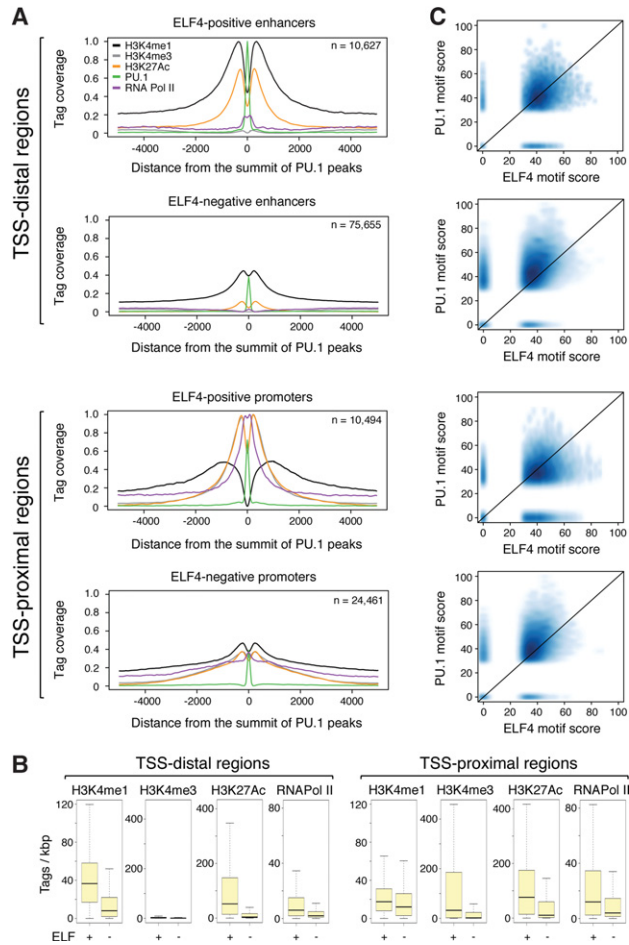
The correlation between ELF4 binding and marks of high activity was similarly detected at PU.1-bound TSS-proximal regions (Fig. 3A, bottom panels), where H3K4me3, H3K27ac, and RNA Pol II showed higher signals at ELF4-positive than at ELF4-negative promoters. The box plots in Figure 3B provide a quantitative description of the same data and show the strong correlation between ELF binding and indicators of transcriptional activity.

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-positive/ELF4-positive regions, while the ELF motif was either absent or low-scoring in the ELF4-negative regions (Fig. 3C), 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.

#### *Release of ELF proteins from promoters correlates with transcriptional shutdown*

Stimulation of macrophages with inflammatory agonists such as lipopolysaccharide (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 h resulted in a general reduction in ELF4 occupancy (Fig. 4A,B,



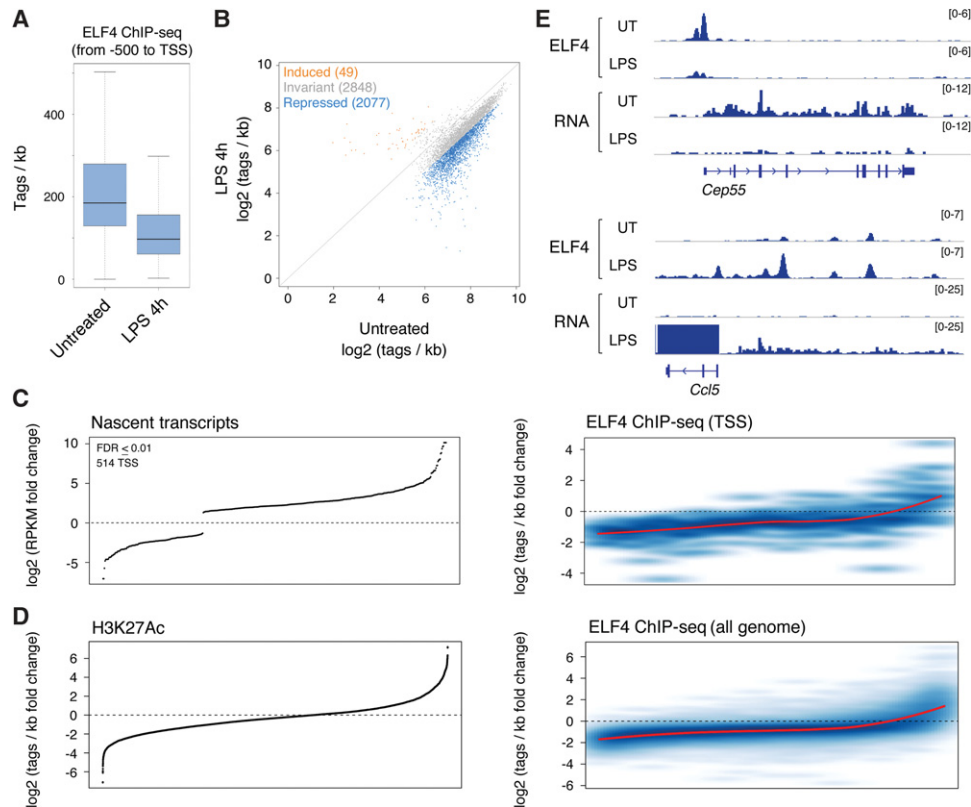
**Figure 3.** ELF binding is associated with constitutive activity at macrophage PU.1-bound enhancers and promoters. (A) Cumulative distributions showing H3K4me1 (black), H3K4me3 (gray), H3K27ac (yellow), PU.1 (green), and RNA Pol II (purple) levels of ELF4-positive or ELF4-negative genomic regions bound by PU.1. Data are shown at both TSS-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 per kilobase) 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. Motif scores were measured using FIMO (find individual motif occurrences) analysis.

data at TSS; Supplemental Fig. S4, global data), with only a small fraction of regions showing increased binding at either the TSS (Fig. 4B) or elsewhere in the genome (Supplemental Fig. S4). While the induced ELF4 peaks in Figure 4B were significantly enriched for gene ontology (GO) categories related to immune cell activation, the repressed ones were associated with GO categories related to house-keeping functions such as DNA repair and RNA metabolism (Supplemental Table S5). To determine whether dynamic variations in ELF4 occupancy correlated with gene expression changes, we generated data sets of chro-

matin-associated nascent transcripts from unstimulated and LPS-stimulated (4 h) macrophages. In Figure 4C (left panel), genes whose transcription significantly changed in response to LPS stimulation (false discovery rate [FDR]  $\leq 0.01$ ) were ordered, with the most repressed genes on the left side and the most induced ones on the right (Supplemental Table S6). The smoothed scatter plot in the right panel of Figure 4C shows ELF4 binding at the corresponding TSS. The overall trend of the plot and the polynomial regression fit (Fig. 4C, 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 correlating with gene activation (Supplemental Table S6). However, there was also a substantial fraction of genes (corresponding to the central area of the plot in Fig. 4C) in which partial ELF release correlated with induction. The promoters of these genes were associated, among others, with motifs recognized by stimulus-inducible TFs such as SRF, IRFs, STATs, and NF- $\kappa$ B, albeit their level of enrichment was lower than that found at the promoters of genes at the right side of the plot (namely, genes with a strong induction and increased ELF recruitment) (Supplemental Table S7). We also correlated genome-wide changes in histone acetylation (H3K27ac) induced by LPS with changes in ELF4 occupancy. Again, we detected an imperfect correlation between loss of histone acetylation and ELF4 release on the one hand and gain in both acetylation and ELF4 occupancy on the other (Fig. 4D). Figure 4E 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*). Taken together, these results indicate that recruitment of ELF TFs tightly correlates with the recruitment of the transcriptional machinery in a broad panel of cis-regulatory elements in both basal and stimulated conditions.

#### Functional activity of ETS sites in minimal promoters

The strong correlation between ELF binding and constitutive transcriptional activity of TSS-proximal and TSS-distal cis-regulatory elements as well as the vicinity of ETS sites to mapped TSSs 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 of short sequences of fixed length; namely, 60 bp from mapped TSSs. The promoters tested were selected based on the following criteria: (1) the presence of the canonical class II ETS site within 60 nt of the mapped TSSs, (2) the absence of a TATA box (see Fig. 2), and (3) binding by ELF4 as determined by ChIP-seq. In the first set of experiments, all sequences were cloned upstream of 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



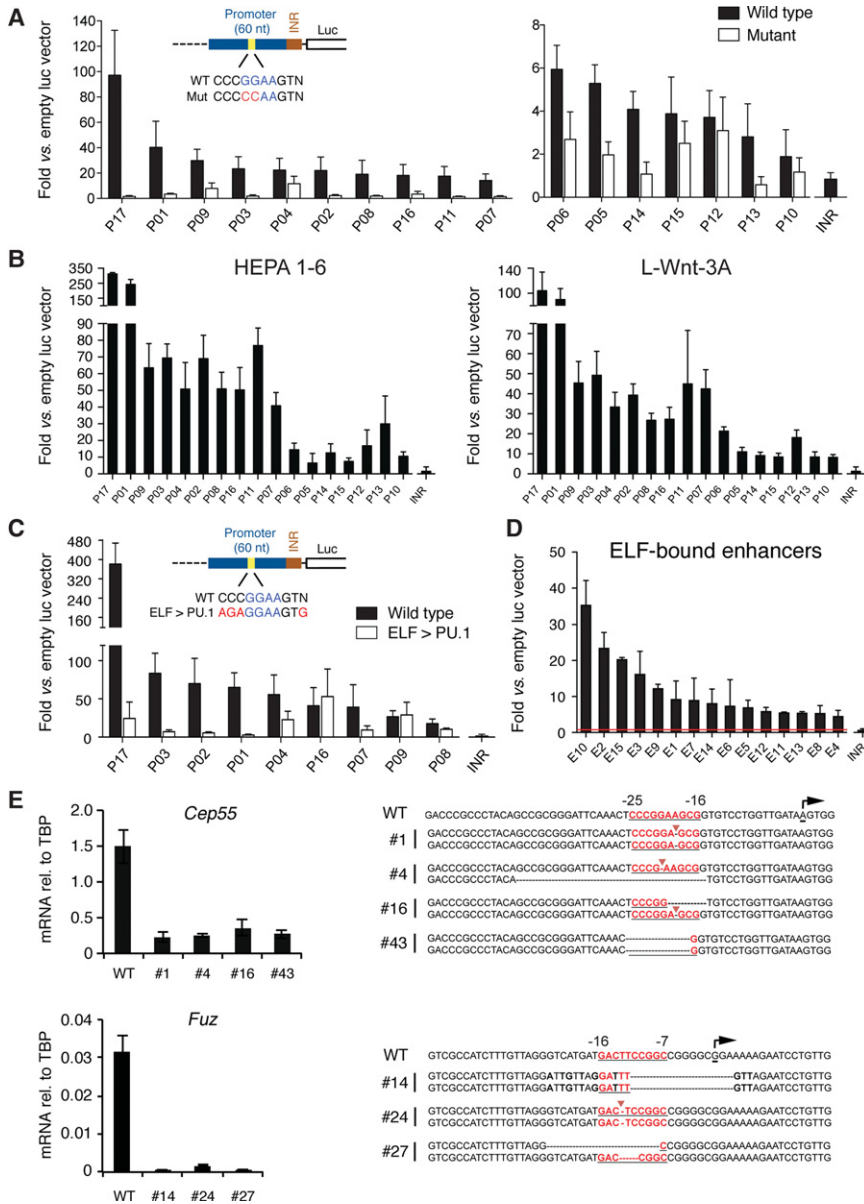
**Figure 4.** Changes in ELF4 genomic distribution upon LPS stimulation correlated with transcriptional changes. (A) The box plot shows changes in ELF4 occupancy at TSSs in response to LPS stimulation. (B) Scatter plot showing changes in genomic occupancy of ELF4 after LPS stimulation. Invariant, induced, and repressed peaks are indicated. (C) Correlation between transcriptional changes and ELF4 occupancy in LPS-stimulated macrophages. In the *left* panel, the genes whose transcription was significantly ( $FDR \leq 0.01$ ) 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 TSSs (from  $-500$  to  $0$ ) of the same genes. The red line represents a polynomial regression fit. (D) 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 at the *right*. (Red line) Polynomial regression fit. (E) 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.

enhancer sequences and thus with very low to undetectable basal activity (Fig. 5A; Supplemental Table S8). Importantly, the SP1–INR combination alone was ineffective at stimulating luciferase expression (Supplemental Fig. S5), while a synthetic promoter in which the SP1, the INR, or both motifs were eliminated 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 (Supplemental Fig. S5). Upon transfection in a macrophage cell line (Raw264.7), all of the promoters tested stimulated the expression of the reporter gene (Fig. 5A). In 13 out of 15 cases, the transcriptional activity of the promoter was almost completely dependent on the ETS site, since a mutation destroying 2 nt in its core abrogated luciferase expression (Fig. 5A, white bars). The same core promoters not only were able to activate transcription in two other nonhematopoietic cell types (hepatocytes and fibroblasts) (Fig. 5B) but also showed comparable relative strength in the three different

cell types. Altogether, these data indicate that binding of ETS proteins close to the TSSs imparts constitutive activity to minimal promoters. Importantly, core promoter activity strictly required an ELF-type ETS site because its replacement with a PU.1 site (Fig. 5C) in nearly all cases (eight out of 10) abrogated transcriptional activity.

Since ELFs also bound to active enhancers via their specific motifs (Fig. 3), we investigated whether ELF site-containing enhancers were similarly capable of activating transcription. Sixty-base-pair-long DNA fragments corresponding to ELF4-positive enhancers were cloned (Supplemental Table S8) and tested as described above: All enhancers tested were able to activate transcription of the reporter gene (Fig. 5D), and, in most cases, this activity was reduced when the ETS site was mutated (Supplemental Fig. S6).

Given the data above, it was critical to determine whether ELF sites in endogenous promoters in their own native genomic context are similarly required to activate gene transcription. Therefore, we resorted to CRISPR/



**Figure 5.** A critical role of ELF-type ETS motifs in the constitutive activity of endogenous promoters. (A) Endogenous core promoters (60 nt from annotated TSS) bound by ELF4 in vivo and bearing a canonical ELF site were cloned into a promoterless vector (pGL3-basic, Promega) just upstream of a common canonical INR sequence, 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 promoterless vector. The activity of a construct containing only an INR motif upstream of the luciferase gene is shown at the right. Error bars represent the SD of at least three independent biological replicates. (B) The same ELF site-positive core promoter sequences as in A were tested in hepatocytes (HEPA1-6 cell line) and fibroblasts (L-Wnt-3A). (C) The ELF site of a panel of core promoter sequences shown in A was mutated to PU.1 sites (white bars), and the effects on luciferase activity were measured. Error bars represent the SD of at least three independent biological replicates. (D) Transcriptional activity of ELF4-positive enhancers. Enhancers bound by ELF4 and containing an ELF site were tested for their ability to act as core promoters in macrophages. Sixty-nucleotide regions ( $\pm 30$  bp from the ELF4 summit) were cloned and tested as described in A. The red line indicates a onefold enrichment over pGL3 basic vector. Data are shown as fold enrichment over promoterless vector (pGL3-basic). Error bars represent the SD of at least three independent biological replicates. (E) Editing of endogenous ELF sites. CRISPR/Cas9 was used to edit genomic sequences spanning the TSS-proximal ELF site at two genes constitutively bound by ELF4. Individual Raw264.7 clones were isolated and sequenced to characterize the mutant sequences at each allele of the two genes. The histogram at the left shows

mRNA expression by RT-qPCR in wild-type cells (transduced with Cas9 without single guide RNA [sgRNA]) and in a selected panel of clones bearing informative mutations spanning the ELF site. The sequence of the mutant clones is shown at the right. The ELF site is shown in red and underlined. The position of the TSS is indicated.

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 (Fig. 5E).

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.

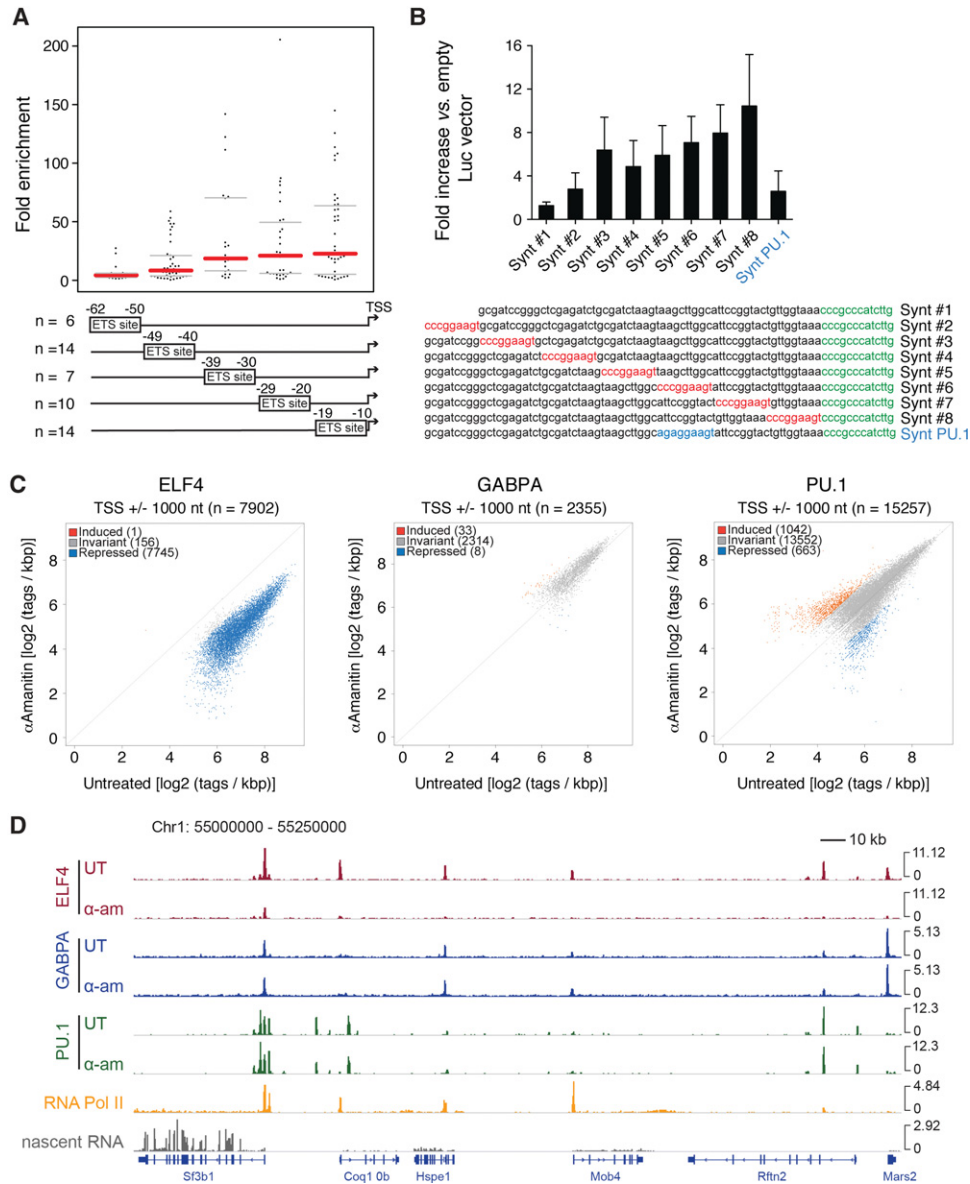
*Mutual interactions between ELF proteins and the transcriptional machinery*

Since the ETS sites recognized by ELFs and GABPA are commonly found within 50 nt (and even more frequently within 20 nt) of mapped TSSs (FitzGerald et al. 2004, 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 five groups based on the distance



between the ETS site and the TSS (listed in Supplemental Table S8). Sixty-base-pair fragments were then cloned upstream of a common SP1–INR motif as above (Fig. 5A) 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 (Fig. 6A). 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



**Figure 6.** The relationship between TSS-proximal ELF4 binding and the transcriptional machinery. (A) Endogenous ELF site-positive core promoters ( $n = 51$ ) (listed in Supplemental Table S8) as described in Figure 5A were divided into five groups depending on the distance from TSSs and tested for luciferase activity in macrophages. Each dot of the bee swarm plot represents one biological replicate out of three independent experiments. The central red bars indicate the median, and the gray bars indicate the first and third quartile. Data are shown as fold enrichment over promoterless vector (pGL3-basic). (B) Synthetic sequences containing an ELF site (red) or a PU.1 site (blue) at different distances from a common INR sequence (green) were tested for luciferase activity in macrophages. Data are shown as fold enrichment over promoterless vector (pGL3-basic). Error bars represent  $\pm$ SD of at least three independent biological replicates. (C) Scatter plot indicating genomic regions ( $\pm 1000$  bp from TSSs) bound by ELF4 (left panel), GABPA (middle panel), and PU.1 (right panel), as determined by ChIP-seq in macrophages treated with 10  $\mu$ g/mL  $\alpha$ -amanitin for 5 h relative to untreated cells (UT). Blue dots indicate regions where the ELF4 signal is reduced upon  $\alpha$ -amanitin treatment. (D) A representative genomic region showing ELF4 binding loss upon  $\alpha$ -amanitin treatment, while GABPA and PU.1 binding were not affected.

distances from the INR. The combination of an ETS site and the INR was sufficient to activate luciferase expression in this minimal promoter context (Fig. 6B). Consistent with the previous experiment, the synthetic promoter also showed a clear (albeit imperfect) trend in which a shorter distance of the ETS site from the INR was associated with higher activity (Fig. 6B), and the replacement of the ELF/GABPA site with a PU.1 site caused a loss of activity.

The distance constraints for optimal transcriptional activity revealed by these experiments hint at the occurrence of close interactions between ETS site-bound proteins and the transcriptional machinery. We therefore tested whether the association of ETS proteins with their DNA-binding site is influenced by the recruitment of the transcriptional machinery. To this aim, we depleted the large RNA Pol II subunit (Rbp1) with an extended (5-h) treatment with  $\alpha$ -amanitin (Supplemental Fig. S7A) and generated ELF4, GABPA, and PU.1 ChIP-seq data sets (Fig. 6C). While RNA Pol II depletion did not greatly impact PU.1 and GABPA association with TSS-proximal regions, it nearly completely (7745 out of 7902 peaks; 98%) abrogated ELF4 binding (Fig. 6C) without affecting its abundance (Supplemental Fig. S7A). The effects of  $\alpha$ -amanitin on ELF4 binding were not associated with differences in promoter accessibility, as indicated by an ATAC-seq (assay for transposase-accessible chromatin [ATAC] with high-throughput sequencing) analysis (Supplemental Fig. S7B,C).

These data show that binding of ELF4 (but not PU.1 or GABPA, which contains two domains absent in ELFs) to active TSS-proximal regions is stabilized by RNA Pol II, which is likely 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 6D.

#### *ETS-dependent recruitment of transcriptional and chromatin regulators to core promoters*

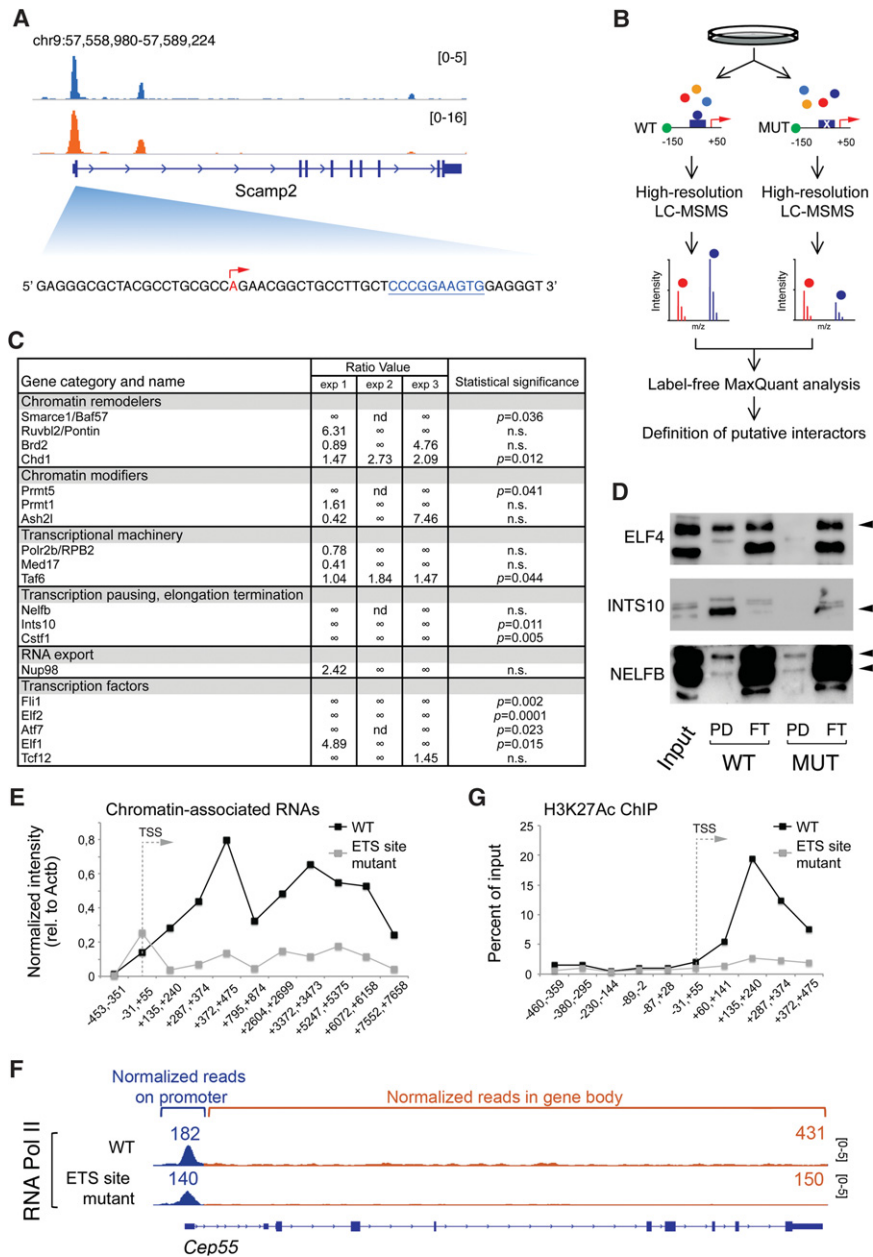
To obtain insight into the mechanism of action of ELF proteins in transcriptional activation, we used a DNA affinity purification approach coupled to mass spectrometry analysis. Briefly, we terminally labeled with biotin a 200-bp DNA fragment (−150/+50 relative to the annotated TSS) corresponding to the *Scamp2* gene promoter, which contains a canonical ELF site and is efficiently bound in vivo (Fig. 7A). 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 (Fig. 7B) were separately incubated with a nuclear lysate in the presence of an excess of nonspecific competitor. Biotinylated DNA fragments were retrieved with streptavidin paramagnetic beads and extensively washed before analyzing bound proteins by label-free mass spectrometry. We identified 1353 proteins, the majority of which (1149 out of 1353; 84.9%) was pulled down equally with the wild-type and ETS site mutant probes. The 204 proteins specifically enriched in the pull-down with the wild-type probe were considered putative ELF interactors (Sup-

plemental Material; Supplemental Fig. S8; Supplemental Table S9). In addition to the retrieval of ETS proteins (ELFs and FLI1) within this group, we identified a number of proteins involved in different aspects of transcriptional regulation (Fig. 7C; Supplemental Fig. S8), 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 shown previously to bind the TSSs of highly active genes and remove the nucleosomal barrier downstream from the TSSs, thus enabling RNA Pol II promoter escape (Skene et al. 2014) and maintaining high-level transcription (Guzman-Ayala et al. 2015). We first validated some of the affinity-isolated proteins by Western blot in independent experiments. ELF4 was used as a specificity control and in fact was selectively pulled down with the wild-type probe (Fig. 7D). Both NELFB and INTS10 were more enriched in affinity purifications with the wild-type probe compared with those with the mutant probe (Fig. 7D), with ratios that, overall, were consistent with those observed in the mass spectrometry analysis. In particular, recruitment of INTS10 was very selectively dependent on an intact ELF site.

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 (Fig. 5E). Since we found that nascent RNAs were down-regulated all over the gene body (Fig. 7E), 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 (Fig. 7F), indicating a defective entry of RNA Pol II inside the coding region. Remarkably, histone acetylation downstream from the TSS was almost completely abrogated in ETS site mutant cells (Fig. 7G).

## 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 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 that we obtained suggest that, irrespective of their tissue-restricted or broad activity across cell types, a large panel of constitutively active promoters and enhancers may rely on a more limited number of TFs (such as the ELF proteins) than could be expected, in principle, based on their functional divergence.



**Figure 7.** ELF-mediated recruitment of transcriptional and chromatin regulators to core promoters. (A) Snapshot of the *Scamp2* gene showing TSS-proximal ELF1 and ELF4 binding. The TSS (red) and the ELF site (light blue and underlined) are highlighted in the DNA sequence below. (B) Experimental scheme of the DNA-mediated pull-down experiment using baits containing either a wild-type (WT) or a mutated (MUT) ETS motif by label-free proteomics. Three experimental replicates were carried out in which a 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 distinct experiments were analyzed separately by high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS). Proteins in each sample were quantified with the label-free algorithm contained in the MaxQuant software suite, in which the intensities of corresponding peptides are compared between individual spectra. Specific interactors (i.e., blue line) show higher LFQ intensity values in the wild-type 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 intensities in the two experiments, with no effect due to the mutation. (C) Short list of putative ELF interactors. The ratio value (calculated by dividing the LFQ intensity of each protein in the wild-type oligo DNA pull-down by the one measured for the same protein in the mutated oligo DNA pull-down) is reported for each protein in all biological replicates. (∞) Ratios where the LFQ intensity of a protein was equal to 0 in the mutated oligo DNA pull-down; (nd) protein not detected. Proteins significantly enriched are indicated with their respective *P*-values. Proteins without the *P*-value were not statistically significant upon LFQ analysis but were within the top 30% of each independent LFQ protein ratio distribution in

at least two out of three biological replicate experiments. (D) 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. (E) 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. (F) 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 the gene body are indicated. (G) Histone H3K27ac at the *Cep55* gene in cells with a wild-type or mutant ETS site.

ETS proteins may have originally appeared in primitive eukaryotes to bridge core promoters and the transcriptional machinery (Schumacher et al. 2003). The data reported in this study suggest that while some ETS proteins (such as PU.1 and FLI1) acquired in mammalian tissue-specific functions related mainly 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 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 of them, which results in completely different genomic distribution profiles, with enhancer-biased and promoter-biased ETS proteins at the 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). Both the close proximity of ELF-specific ETS sites to the TSSs and the negative impact of RNA Pol II depletion on ELF occupancy are consistent with the occurrence of a tight interaction between DNA-bound ELFs and the transcriptional machinery. However, the precise steps 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 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 promoter-proximal nucleosomes (Skene et al. 2014), may critically contribute to transcriptional elongation at genes (such as those bound by ELFs) (Fig. 2) 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 that 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 be sufficient by themselves to recruit RNA Pol II and initiate transcription.

The general principle that can be drawn from these data is that the co-optation 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 prerequisite for its tighter regulation in response to specific microenvironmental 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) (Zaret and Carroll 2011; Glass and Natoli 2016) and motifs for TFs that impart specific functional properties such as inducible (e.g., NF- $\kappa$ B and 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).

## Materials and methods

### Cell culture and reagents

Macrophages were derived from the BM of C57/BL6 mice (Harlan) as described (Austena et al. 2015). The project was approved by the Italian Ministry of Health and performed under the supervision of the Institutional Committee for Animal Welfare. RAW264.7 (mouse macrophages), Hepa1-6 (mouse hepatoma cells), and L-Wnt-3A (mouse fibroblasts) were purchased from American Type Culture Collection (ATCC).  $\alpha$ -Amanitin (Sigma) was used at 10  $\mu$ g/mL. For Western blots, the following antibodies were used: Rpb1 (Santa Cruz Biotechnology, sc-899), vinculin (Sigma), ELF4 (homemade rabbit polyclonal antibody raised against residues 555–655 of the mouse protein), ELF1 (Santa Cruz Biotechnology, sc-631x), INTS10 (Origene, TA337360), and NELFB (Proteintech, 16418-1-AP).

### ChIP-seq

Fixed macrophages ( $8 \times 10^7$  for FLI1 and ELF1,  $4 \times 10^7$  for GABPA and ELF4, or  $5 \times 10^6$  for PU.1) and RAW264.7 cells ( $8 \times 10^7$  for RNA Pol II and  $2 \times 10^7$  for H3K27ac) were lysed in RIPA buffer and, after chromatin shearing by sonication, incubated overnight at 4°C with protein G Dynabeads (Invitrogen) that were previously coupled with 3–10  $\mu$ g of antibody (Austena et al. 2015). The following antibodies were used: ELF1 (Santa Cruz Biotechnology, sc-631x), FLI1 (Santa Cruz Biotechnology, sc-356x), GABPA (Santa Cruz Biotechnology, sc-22810), RNA Pol II (Santa Cruz Biotechnology, sc-899x), and H3K27ac (Abcam, ab4729). The rabbit polyclonal antibodies for PU.1 (Mancino et al. 2015) and ELF4 were generated in-house. ChIP-seq libraries were generated as described (Austena et al. 2015) and sequenced on an Illumina HiSeq 2000 (Supplemental Material).

### Cloning, transfections, and luciferase assays

The genomic regions of interest (Supplemental Table S8) were placed upstream of a characterized SP1–INR sequence (5'-CCCGCCCATCTTG-3'; the -1/+1 Py/Pu dinucleotide is underlined) (Weis and Reinberg 1997) and cloned by KpnI and NheI digestion in the pGL3-basic luciferase reporter vector (Promega). The loss-of-function mutants of the ETS site (GGAA > CCAA), ELF-to-PU.1 mutant site (CCCGGAAGT > AGAGGAAGT), and synthetic core promoter sequences (cloned by NheI and NcoI) are listed in Supplemental Table S8. Transient transfections were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. RAW264.7 cells were transiently transfected in a 24-well format ( $150 \times 10^3$  cells per well) with 150 ng of empty vector or vectors containing the specified genomic region together with 50 ng of pRL-TK-renilla vector (Promega). Hepa1-6 and L-Wnt-3A cells were transiently transfected in a 48-well format ( $30 \times 10^3$  cells per 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 24 h after transfection. Values were normalized on the Renilla signal and expressed as fold change relative to the empty vector.

### ATAC-seq

The original ATAC-seq protocol (Buenrostro et al. 2013) was modified according to Lara-Astiaso et al. (2014). A detailed description of the ATAC-seq experiment is in the Supplemental Material.

*CRISPR/Cas9 genome editing*

Single-guide sequences specific to the ETS site of *Cep55* and *Fuz* core promoters (listed in Supplemental Table S8) 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 listed in Supplemental Table S8). Positive clones were subjected to Sanger sequencing.

*DNA affinity purification and mass spectrometry-based proteomics analysis*

Biotin-conjugated DNA baits (240 bp) corresponding to the *SCAMP2* gene promoter (mm9 coordinates chromosome 9: 57408589–57408808) and its ETS site mutant were generated by PCR using the primers listed in Supplemental Table S8. A detailed description of the pull-down experiment and the subsequent mass spectrometry analysis and label-free protein quantitation to identify specific interactors is in the Supplemental Material.

*Computational methods*

Short reads obtained from Illumina HiSeq 2000 runs were analyzed as described (Austena et al. 2015). Detailed computational methods are described in the Supplemental Material.

*Accession numbers*

Raw data sets are available in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE88702, which comprises ChIP-seq data (GSE88699), ATAC-seq data (GSE88698), and expression data (GSE88700).

**Acknowledgments**

We thank Silvia Monticelli for comments on the manuscript, and Luca Rotta, Thelma Capra, and Salvatore Bianchi (Istituto Europeo di Oncologia and Istituto Italiano di Tecnologia Center for Genomic Sciences) for the preparation and processing of next-generation sequencing libraries. This work was supported by the European Research Council (Advanced ERC grant to G. N.), the Italian Association for Research on Cancer (AIRC grant to G.N.), and the Italian Ministry of Health (Grant Ricerca Finalizzata to G.N.). A.C., A.T., S.G., and G.N. conceived the study. A.C. and S.G. collected data and designed and carried out most experiments with the help of E.P., M.S., S.P., and L.A. Proteomics experiments were acquired by M.S. and A.S. and were analyzed under the supervision of T.B. A.T. analyzed all data sets with help from I.B. S.G. and G.N. wrote the manuscript with input from all authors. S.G. and G.N. supervised. G.N. acquired the funding.

**References**

Austena LM, Barozzi I, Simonatto M, Masella S, Della Chiara G, Ghisletti S, Curina A, de Wit E, Bouwman BA, de Pretis S, et al. 2015. Transcription of mammalian *cis*-regulatory elements is restrained by actively enforced early termination. *Mol Cell* **60**: 460–474.

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

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 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.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Sempke CA, Taylor MS, Engstrom PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–635.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–21936.

Deng W, Roberts SG. 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* **19**: 2418–2423.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**: e1000384.

Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. 2015. Suboptimization of developmental enhancers. *Science* **350**: 325–328.

Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, 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 PC, Shlyakhtenko A, Mir AA, Vinson C. 2004. Clustering of DNA sequences in human promoters. *Genome Res* **14**: 1562–1574.

FitzGerald PC, Sturgill D, Shlyakhtenko A, Oliver B, Vinson C. 2006. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol* **7**: R53.

Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. 2008. A code for transcription initiation in mammalian genomes. *Genome Res* **18**: 1–12.

Gallant S, 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 JM, Wagner EJ, Lazar MA, Shilatfard A, Shiekhhattar R. 2014. Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell* **56**: 128–139.

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

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

Glass CK, Natoli G. 2016. Molecular control of activation and priming in macrophages. *Nat Immunol* **17**: 26–33.

- Goldberg ML. 1979. "Sequence analysis of *Drosophila* histone genes." PhD thesis, Stanford University, San Francisco.
- Goode DK, Obier N, Vijayabaskar MS, Lie ALM, Lilly AJ, Hannah R, Lichtinger M, Batta K, Florkowska M, Patel R, et al. 2016. Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev Cell* **36**: 572–587.
- Guzman-Ayala M, Sachs M, Koh FM, Onodera C, Bulut-Karslioglu A, Lin CJ, Wong P, Nitta R, Song JS, Ramalho-Santos M. 2015. Chd1 is essential for the high transcriptional output and rapid growth of the mouse epiblast. *Development* **142**: 118–127.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576–589.
- Hollenhorst PC, McIntosh LP, Graves BJ. 2011. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* **80**: 437–471.
- Juven-Gershon T, Hsu JY, Kadonaga JT. 2006. Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans* **34**: 1047–1050.
- Kadonaga JT. 2012. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol* **1**: 40–51.
- Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, et al. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* **51**: 310–325.
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187.
- Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* **12**: 34–44.
- Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretzky I, Jaitin DA, 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.
- Lichtinger M, Ingram R, Hannah R, Muller D, Clarke D, Assi SA, Lie ALM, Noailles L, Vijayabaskar MS, Wu M, et al. 2012. RUNX1 reshapes the epigenetic landscape at the onset of hematopoiesis. *EMBO J* **31**: 4318–4333.
- Lifton RP, Goldberg ML, Karp RW, Hogness DS. 1978. The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications. *Cold Spring Harb Symp Quant Biol* **42**: 1047–1051.
- Mancino A, Termanini A, Barozzi I, Ghisletti S, Ostuni R, Prosperini E, Ozato K, Natoli G. 2015. A dual *cis*-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev* **29**: 394–408.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaer BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat Biotechnol* **28**: 495–501.
- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**: 157–171.
- Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A. 2006. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol* **7**: R78.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.
- Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**: 783–784.
- Schumacher MA, Lau AO, Johnson PJ. 2003. Structural basis of core promoter recognition in a primitive eukaryote. *Cell* **115**: 413–424.
- Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**: 272–286.
- Skene PJ, Hernandez AE, Groudine M, Henikoff S. 2014. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife* **3**: e02042.
- Smale ST, Baltimore D. 1989. The 'initiator' as a transcription control element. *Cell* **57**: 103–113.
- Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–626.
- 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.
- Tippmann SC, Ivanek R, Gaidatzis D, Scholer A, Hoerner L, van Nimwegen E, Stadler PF, Stadler MB, Schubeler D. 2012. Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol Syst Biol* **8**: 593.
- Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**: 390–394.
- Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, et al. 2010. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J* **29**: 2147–2160.
- Weis L, 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.
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338–345.
- Yang C, Bolotin E, Jiang T, Sladek FM, 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.
- Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**: 2227–2241.