The hierarchy of transcriptional activation:

from enhancer to promoter

Douglas Vernimmen¹ & Wendy A. Bickmore²

¹The Roslin Institute, Developmental Biology Division, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK. ²MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.

Corresponding author; Douglas Vernimmen
Telephone +44 (0)131 651 9119
Fax: +44 (0)131 651 9105
e-mail: douglas.vernimmen@roslin.ed.ac.uk

Abstract word count: 65

Characters (including space): 26,895

Words count: 4,160

Figures: 2

Tables: 2

References: 165
Abstract

Regulatory elements (enhancers) that are remote from promoters play a critical role in the spatial, temporal and physiological control of gene expression. Studies on specific loci, together with genome-wide approaches, suggest that there may be many common mechanisms involved in enhancer-promoter communication. Here, we discuss the multi-protein complexes that are recruited to enhancers and the hierarchy of events taking place between regulatory elements and promoters.
Glossary

DNase Hypersensitive Site (DHS): Open region in the genome with increased chromatin accessibility to DNaseI that may reflect the occupation by a transcription factor or the disruption of nucleosome structure. DHS form nucleosome free regions (NFR).

Pioneer Transcription Factors: Transcription factors that can bind their target sites at nucleosomal DNA. This facilitates chromatin remodelling and the binding of other transcription factors with the formation of open chromatin regions (DHS) before enhancer/promoter activation. They also have the property of being retained on mitotic chromosomes and thus could serve as “bookmarking” proteins in mitosis.

Relay Transcription Factors: Transcription factors from the same family (e.g. SOX, GATA) relaying each other for the same binding site (exchange model) as they are expressed at different stages during gene priming.

Transcription Start Site (TSS): Nucleotide marking the site of initiation of mRNA transcription.

Enhancer: Regulatory sequence that increases the rate or the probability of transcription of a target gene. An enhancer may lie far away, upstream or downstream from the gene it regulates or may be located in an intron of its target gene or indeed in an intron of another gene.

Locus Control Region (LCR): Genomic region that has the ability to confer physiological levels of tissue-specific expression on a gene linked in cis, independent of the gene’s integration site. A LCR can open silent chromatin.
General Transcription Factor (GTF): Also referred to as basal transcription factors (TFIIA, TFIB, TFID, TFIE, TFIIF and TFIIH) that bind to core promoters.

Pre-initiation complex (PIC): Association between GTF and RNA polymerase II.

PolII Holocomplex: Complex of subunits forming the complete enzymatically active form of PolII.

Paused PolII: After promoter escape, the engaged PolII is stalled at a pause site, waiting for further signals to progress during elongation.

C-terminal domain (CTD): The C-terminal domain (CTD) of the largest subunit of RNA Polymerase II (PolII) consists of an array of repeats of a heptapeptide sequence (52 repeats in mammals). Amino acids in these repeats are targets for post-translational modification such as phosphorylation of serine 5 (Ser5P) - associated with early elongation (or the paused state of PolII) and Ser2P associated with full elongation.

Activator: Trans-acting factor binding a DNA sequence to activate the transcriptional activity of a target gene.

Co-activator: Non-DNA binding protein that associates with an activator and enhances transcription.

Mediator: Large co-activator complex containing 30 subunits in metazoans distributed in three modules: the head, the middle and the tail. Mediator is conserved throughout all eukaryotes.

Integrator: Large co-activator complex containing at least 14 subunits with a total MW over 1MDa. Integrator is restricted to metazoans.
The core promoter

Genes transcribed by RNA Polymerase II (PolII) usually have two distinct families of cis-acting elements: the promoter \[\leq 1 \text{ kb from the transcription start site (TSS)}\] - composed of a core promoter \(^1\) \(^2\) and nearby (proximal) regulatory elements \(^3\) \(^4\), and more remote (distal) cis-regulatory elements (\(\geq 1 \text{ kb from TSS}\)), which can be enhancers, silencers, insulators or locus control regions (LCR) \(^3\). The exact composition of core promoter elements may be a key determinant of enhancer-promoter specificity \(^5\) \(^6\). In mammalian genomes, enhancers are enriched in core promoter elements but are CpG poor whereas promoters are generally CpG rich \(^7\) \(^8\). Beside the CpG content, enhancers and promoters have broad similarities and overlapping functional properties, and have been considered to form a single class of regulatory element \(^9\).

The core promoter represents the docking site for the General Transcription Factors (GTFs), including TFIIA, TFIIIB, TFIIID, TFIIE, TFIIF and TFIIH, which, together with PolII, form the pre-initiation complex (PIC) \(^10\). The PIC is thought to assemble on the core promoter in a specific and sequential order that directs PolII to the nearby TSS \(^10\). However, this is only sufficient to direct low levels of accurately initiated transcription from DNA templates \textit{in vitro}, a process generally referred to as basal transcription.

The first step in PIC assembly is binding of TFIID, a multi-subunit complex consisting of TATA-box-binding protein (TBP) and a set of 14 TBP-associated factors (TAFs) \(^10\). Transcription then proceeds through a series of steps, including promoter melting, clearance and escape, before fully functional PolII elongation is achieved. Alternative
core promoter complexes may help to maintain specific transcriptional programs in terminally differentiated cell types\textsuperscript{11,12,13,14}.

Models of transcription regulation view this as a cycle, in which complete PIC assembly is stimulated only once. After PolII escapes from the promoter, TFIID, TFIIE, TFIIH and the Mediator complex (see glossary) remain on the core promoter; subsequent re-initiation then only requires \textit{de novo} recruitment of sub-complexes comprising Pol II-TFIIF and TFIIB (Reviewed in \textsuperscript{15}). The various steps of PIC assembly on a core promoter can occur with different timings during differentiation. For example, TBP is already bound to the promoters of \textit{\alpha1-AT, HNF-4\alpha, VpreB1} and \textit{\lambda5}, long before differentiation and the transcriptional activation of these genes\textsuperscript{16,17}. Additional transcription factors (TFs) and PolII are recruited later when the genes are transcribed. The one-step recruitment of a (pre-)formed holocomplex (see glossary) at promoters has been also described\textsuperscript{18-21}. However, it is worth noting that the right temporal window to appreciate the dynamics of PIC recruitment is often missing from most studies.

In metazoans, the transition from initiation to productive elongation is another important step that involves several levels of regulation. In a region between 30-60 nucleotides downstream the TSS, PolIII is often found stalled and thus paused\textsuperscript{ed} at this site, awaiting additional signals for full elongation\textsuperscript{22}. The release of paused PolIII is controlled by several TFs such as the negative elongation factor (NELF), the DRB sensitivity-inducing factor (DSIF) and the transcription elongation factor P-TEFb complex (CDK9 and cyclin T). P-TEFb is part of a larger multisubunit complex, called super elongation complex (SEC)\textsuperscript{23}. The CTD of PolIII plays an important role in elongation by its phosphorylation at several residues (see glossary). Recently, a new
multiprotein complex, termed Integrator, has been shown to regulate elongation by recruiting the SEC\textsuperscript{24}.

**Large protein complexes are bound to promoters and enhancers**

Transcription is greatly stimulated by a second class of TFs, termed activators. In general, activators are sequence-specific DNA-binding proteins whose recognition sites are usually present near the core promoter and/or at enhancers. Binding of TFs at these elements usually corresponds to nucleosome free regions (NFRs) characterised by hypersensitivity to digestion by nucleases (DNase Hypersensitive Sites, DHS)\textsuperscript{2, 25, 26}. This open-chromatin structure can be facilitated by chromatin remodelling factors, which are recruited by TFs and modify histones of the nearby nucleosomes.

Binding of activators does not stimulate transcription from chromatinised templates \textit{in vitro}. The search for factors that stimulate activator-dependent transcription led to the identification of co-activators including; Mediator complexes \textsuperscript{27, 28}, CBP \textsuperscript{29}, p300 \textsuperscript{30} and BAF \textsuperscript{31}. TFs recruit co-activators that can then modify chromatin and/or interact with the core transcription machinery.

The large multiprotein Mediator complex can act as a bridge between transcription activators and components of the PIC \textsuperscript{32} (see below). It appears to play important roles in many steps of transcription, including PIC formation and the transition to elongation \textsuperscript{32}. Mediator is over a megadalton (MDa) in size and 30nm in length, with distinct structural modules and a flexible structure that changes in response to the binding of different TFs \textsuperscript{33}. TF binding seems to induce a conformation change in
Mediator that facilitates PolIII binding. Different TFs bind different Mediator subunits, and Mediator complexes that lack a specific subunit can still activate transcription in response to TFs that bind to other subunits. Therefore, among other proteins (e.g. CTCF and cohesin complex) not described in this review, Mediator provides a very important bridge for integrating information coming from different signalling pathways. Mediator might also provide an important binding surface for non-coding RNAs, including eRNAs (see below).

Other co-activators are ATP-dependent chromatin remodelling factors (such as Brahma-associated factor – BAF), or histone acetyltransferases (HAT) – p300/CBP. These can be part of the same complexes. ATP-dependent chromatin-remodelling families form different complexes by a combinatorial assembly of many subunits, to produce biological specificity \(^{34}\). BAF complexes, which belong to the SWI/SNF family of ATPase dependent chromatin remodelling complexes, are involved in the relaxation of higher-order chromatin structures and in nucleosome movement and exchange \(^ {35}\). The p400 SWI/SNF is associated with a HAT (TIP60) in the Tip60/p400 complex that is involved in histone (H2A/H2A.Z) exchange. CREB-binding protein (CBP) and its paralog p300 are co-activator HATs that are found at both promoters and enhancers, and chromatin immunoprecipitation (ChIP) for p300/CBP, together with H3K27ac, is often used to identify active enhancers \(^ {36,37}\). However, this is unlikely to be a universal signature of all active enhancers. Indeed, another class of enhancers, containing H4K16ac and KAT8 (MYSM1) but not p300 and H3K27ac have been recently described in embryonic stem (ES) cells \(^ {38}\). Moreover, HATs also have important non-histone substrates and the role of this in enhancer function is under-studied \(^ {39}\). Other HATs and HAT-containing complexes (SAGA/PCAF) also have co-activator activity \(^ {40}\).
One of the main questions that need to be addressed is at which step during gene activation do various nucleoprotein complexes assemble at distant enhancers, and how do these complexes then contribute to promoter accessibility, PIC recruitment and/or assembly, transcription initiation and transcription elongation? Enhancers have been shown to have a role in: PIC recruitment at target promoters, removing proteasome complexes at promoters, the generation of intrachromosomal loops between regulatory regions, and the regulation of elongation. Enhancers are also involved in the removal of repressive histone modifications, suggesting that they also contribute to the delivery of enzymes that regulate histone modifications.

Below, we compare studies that have been done in a few mammalian loci in enough depth to provide significant mechanistic insight. Together with supportive genome-wide studies, we discuss if there are common principles that govern the regulation of enhancer-driven transcription.

**Sequential recruitment of factors to enhancers and promoters**

It is now well established that genes are primed for expression by the binding of pioneer TFs (see glossary) generating nucleosome free regions at regulatory elements and bookmarking the genome for gene expression at a later stage of differentiation (Figure 1A). Enhancer priming is followed by the replacement or recruitment of additional TFs (namely relay, tether and trigger, see glossary and below), which may be recruited in a sequential order, mirroring a similar phenomenon on core promoters (e.g. PIC assembly).
Pioneer TFs are able to disrupt chromatin structure and bind to their cognate binding sites irrespective of nucleosomes that may be occluding these sequences (Figure 1A)\(^64\)\(^65\), although this is dependent on the context of other TFs that they associate with in a particular cell type\(^66,\)\(^67\). Pioneer factors, together with chromatin remodelling complexes, are therefore involved in generating an NFR to facilitate the binding of other TFs (Figure 1B). Table 1 lists TFs that have been reported to have pioneer activity. The DNA-binding domain (winged-helix DNA-binding domain/forkhead) of TFs such as HNF3 (FoxA) resembles that of linker histones H1 and H5 and therefore could be involved in chromatin opening by altering nucleosome structure\(^68\). The CCAAT Box binding factor, NFY has also a core histone-like structure\(^69\) and has been suggested to be involved in opening chromatin by nucleosome replacement\(^70\)\(^58\) and facilitating the binding of master regulators to enhancers in ES cells\(^71\). It has been suggested that the pioneer activity of PU.1 - a hematopoietic pioneer factor - TF may relate to the tighter DNA-binding of its ETS-domain compared to that of other ETS-family TFs\(^72\). In reprogramming of somatic cells, Oct4, Sox2, and Klf4 act as pioneer factors, binding at closed chromatin sites\(^73\)\(^65\). Importantly, their binding occurs first at distal enhancers during early reprogramming steps (Figure 1A), whereas promoter occupancy is a much later event\(^74\) (Figure 1B). The formation of NFRs at promoter and enhancer occur independently from each other. Enhancer priming by pioneer TFs in specific cell-lineages provides a chromatin landscape that can then direct cell-type-specific responses to TFs that act downstream of generic signaling pathways\(^75\)\(^79\).

<table>
<thead>
<tr>
<th>Pioneer TF</th>
<th>DNA binding domain</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-1</td>
<td>Basic leucine zipper</td>
<td>(^76)</td>
</tr>
<tr>
<td>AP-2γ (TFAP2C)</td>
<td>Basic helix-span-helix</td>
<td>(^80)</td>
</tr>
<tr>
<td>FOXA1 (HNF-3α)</td>
<td>Forkhead</td>
<td>(^81,)(^82)</td>
</tr>
<tr>
<td>Transcription Factor</td>
<td>Domain/Type</td>
<td>References</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>FOXA2 (HNF-3β)</td>
<td>Forkhead</td>
<td>68, 83</td>
</tr>
<tr>
<td>FOXE1</td>
<td>Forkhead</td>
<td>84</td>
</tr>
<tr>
<td>FOXD3</td>
<td>Forkhead</td>
<td>68, 83</td>
</tr>
<tr>
<td>GATA2</td>
<td>2X GATA-type zinc fingers</td>
<td>20</td>
</tr>
<tr>
<td>GATA3</td>
<td>2X GATA-type zinc fingers</td>
<td>85</td>
</tr>
<tr>
<td>GATA4</td>
<td>2X GATA-type zinc fingers</td>
<td>68, 85, 86</td>
</tr>
<tr>
<td>KLF4</td>
<td>3X C2H2-type zinc fingers</td>
<td>73</td>
</tr>
<tr>
<td>NF-Y (CBF)</td>
<td>NF-YA/HAP2</td>
<td>70, 71</td>
</tr>
<tr>
<td>OCT4</td>
<td>POU-specific + POU-Homeodomain</td>
<td>73, 67</td>
</tr>
<tr>
<td>OTX2</td>
<td>Homeodomain</td>
<td>67</td>
</tr>
<tr>
<td>PAX7</td>
<td>Paired + Homeodomain</td>
<td>87</td>
</tr>
<tr>
<td>PBX1</td>
<td>Homeodomain</td>
<td>88</td>
</tr>
<tr>
<td>PU.1</td>
<td>Ets</td>
<td>66, 72, 75</td>
</tr>
<tr>
<td>SOX2</td>
<td>Hmg box</td>
<td>73, 89, 90</td>
</tr>
<tr>
<td>SOX9</td>
<td>Hmg box</td>
<td>91</td>
</tr>
<tr>
<td>TP53</td>
<td>p53</td>
<td>92</td>
</tr>
<tr>
<td>P63</td>
<td>p53</td>
<td>92</td>
</tr>
<tr>
<td>RFX</td>
<td>Rfx-type winged helix</td>
<td>93</td>
</tr>
</tbody>
</table>

Table 1. Pioneer transcription factors involved in DHS formation prior to gene activation.

Pioneer TFs may remain bound throughout the stages of enhancer activation, or they can be replaced by other TFs (exchange model with relay TFs, see glossary)\(^{20, 89, 94}\).

In some cases, PIC recruitment to enhancers has been reported early during enhancer priming\(^{17, 19}\), in other cases this is a late event\(^{21}\). The first situation led to the idea that enhancers act as a docking site for the recruitment of the general transcription machinery (Figure 1C) that would then be subsequently transferred to the promoter (Figure 1D)\(^{95, 96, 97}\). However, many studies have shown that levels of PIC occupancy at enhancers – as judged by ChIP - often appear to be relatively low compared to those at promoters\(^{98}\). This could be explained by i) one PIC being spread across the enhancer sequence, which is larger than a core promoter (Figure 4C)\(^{1-3}\) by the transient nature of several PICs binding to those sequences, or ii) by indirect binding of PICs.
Apparent differences in the timing of recruitment of PIC components to a promoter either before 16, 17, 99, or at the onset of mRNA transcription 18, 21 might be due to the different role(s) attributed to the enhancers, but also to the presence of other important elements located nearby in the proximal promoter. For example, deletion of the Sp1 site at the T-cell receptor beta (TCRβ) promoter, and the CAAT or CACCC boxes at the γ globin promoter, result in failure to detectably recruit TBP at the promoter 100-102, suggesting that these proximal promoter elements are needed to recruit the PIC in order to form a full “promoter complex”.

Transcription from enhancers and promoters

Enhancers are required for transcription of target genes

Genetic ablation is a powerful approach to address how enhancers influence TF and PolII assembly/elongation at promoters. Table 2 summarize the few studies on single loci addressing this, together with the mechanisms of enhancer-promoter communication (see relevant section below). Supporting independent GTFs and PolII recruitment at enhancers and promoters, a few studies have shown that removing a promoter does not affect GTFs or PolII recruitment at the enhancer 21, 103, but removing the enhancer affects GTFs or PolII binding at the promoter 21, 41-44.

Other studies have shown that deletion of enhancers also affects downstream events such as elongation through PolII phosphorylation 18, 48 3, 49. Release of paused PolII might also require additional TFs or additional enhancers 45, 104 (Figure 2C). PolII elongation is regulated by several kinases (cdk7/TFIIH, cdk8/Mediator, and cdk9/p-TEFb) and these are all recruited to genes when expressed 21, 48 and may be delivered through the enhancers 105 106. This is consistent with genome-wide studies.
showing Ser5P \(^{98}\) and Ser -2P \(^{107, 108}\) phosphorylated forms of PolII at active enhancers, supporting the idea that enhancers can deliver an activated PolII to target promoters. Alternatively, other regulators of elongation such as DSIF and FACT \(^{49}\) may be also involved and this idea has been strengthened by recent studies showing the binding of the Integrator at both enhancers and promoters \(^{109}\). Overall, the current model arising from these studies is that the promoter is not needed to recruit PolII at enhancers, but the enhancer is always needed to recruit PolII at the promoter or for downstream events such as elongation. Beside these studies, it is worth noting that the timing of an enhancer deletion might also influence outcomes. Enhancers might be required for the initiation of a transcription event whereas others might be important for the maintenance of transcription. For example, histone modifications controlled from a transiently required enhancer might remain after a conditional deletion \(^{110, 111}\).

**Enhancers are also transcribed**

Many of the scenarios described above appear to blur the distinction between enhancers and promoters. This is further compounded by evidence of transcription and the production of short RNAs at enhancers (eRNAs) \(^{112, 103, 113}\) (Figure 1D). The level of expression of these eRNAs is low but positively correlates with the level of mRNA synthesis at nearby genes \(^{103}\). eRNAs are short and unstable, probably because the absence of downstream exons (5' splice donors), or the presence of other signals, fails to stabilise the production of the transcribed RNA \(^{114}\) and indeed the degradation of eRNAs by the exosome is important to prevent the formation of deleterious RNA/DNA hybrids \(^{115}\). Note that intragenic enhancers can also function as alternative gene promoters – being spliced to downstream exons to produce stable mRNAs \(^{113}\). Conversely, promoters can also work as enhancers \(^{9, 116}\). Thus,
the distinction between enhancers and promoters lies to some extent simply in their genomic context.

The biological relevance of eRNAs remains controversial as it is still unclear if eRNAs are byproducts of transcription or if they have regulatory functions in themselves. Bidirectional eRNA transcripts can be detected at early stages, prior to the appearance of H3K4me1 at enhancers and to the production of mRNA from the target genes. This could be associated with early recruitment of PolII at enhancers by pioneer TFs. Several attempts at elucidating the role of eRNAs have been addressed by knockdown approaches, showing transcription down-regulation from some but not all target promoters. Other studies, using a more robust approach, that remove the promoter of the target gene were also not conclusive: in the absence of the Arc promoter, eRNA synthesis is abolished, suggesting that it is mRNA dependent; whereas no effect was observed in the same type of experiments on the human growth hormone (hGH-N) locus. One study has proposed that eRNAs act as decoy molecules to release NELF from paused PolII at immediate early genes. This is interesting because, this scenario was supported by a recent study showing that the Integrator, is also recruited to enhancers. Integrator is required for the full processing of eRNAs, and depletion of Integrator subunits reduces the production of eRNAs and abolishes enhancer-promoter communication. As Integrator controls the elongation of mRNA transcription of the genes regulated by paused PolII, the role of eRNAs may depend on the context of elongation regulation of as only 50% of genes are regulated by such mechanism.

**Enhancer-promoter communication**
Enhancers can be separated from promoters by distances ranging from a few kilobases to a little over one thousand kilobases \(^{127}\), yet transcriptional regulation requires some kind of communication between these distant elements (Figure 2). It is still unclear what form this communication takes, e.g. what are the molecules that are transmitted between regulatory element and promoter, when this takes place, and whether this is the same for all classes of enhancers.

Historically, a linking model suggested that an activator protein (eg pioneer Figure 1B) first binds the promoter at a proximal sequence and facilitates the recruitment of a second TF to a site located just downstream the former \(^{90, 128, 129}\). This cascade of recruitment occurs until it reaches the core promoter to finally recruit the PIC \(^{21, 99}\).

The tracking model and/or a facilitated tracking model (Figure 2) is described as a mechanism by which enhancer bound proteins move progressively in an unidirectional manner towards the promoter sometimes without leaving the enhancer sequence, and thus results in the formation of a progressive loop that increases its size until it reaches the promoter to form a stable conformation \(^{99, 130-132}\) (Figure 2A).

In this model, histone acetylation and TF complexes are transiently detected in the intervening sequence and precedes transcription. Originally, it was proposed that intergenic transcripts (eRNAs) are just involved in maintaining an open chromatin structure \(^{133}\). Once the gene is expressed, additional transcripts (mostly unidirectional) have been detected across the intervening sequence between enhancers and promoters, which could reflect the tracking of an active PolII \(^{17, 43}\) (Figure 2A).

The looping model implies a direct interaction between two chromosomal regions by looping out the intervening DNA sequence. Various proteins bound at enhancers and promoters have also been proposed to bridge enhancers and promoters together in
as looped chromatin structures. These include TFs, such as TAF3\textsuperscript{134}, GATA1\textsuperscript{135}, EKLF\textsuperscript{136}, Brg1\textsuperscript{137}, Ldb-1\textsuperscript{138}, Mediator\textsuperscript{139}, CTCF\textsuperscript{140}, SATB1\textsuperscript{141} and cohesins\textsuperscript{3, 142-144}. eRNAs have also been proposed to be physically involved in establishing enhancer-promoter ‘looping’\textsuperscript{124}, and involving the Integrator\textsuperscript{109}.

The stiffness of the chromatin fibre might restrict short-range enhancer-promoter interactions, with a minimal estimated length of 10kb for uninterrupted 30nm chromatin fibres and 0.5kb for naked DNA\textsuperscript{145, 146}. NFRs – e.g. created by pioneer TFs - could thus act as hinges, to facilitate chromatin bending and thus the formation of short loops\textsuperscript{146 147}.

Given the larger distance (> 10 kb and up to 100s of kb) separating many enhancers from their target promoter it is difficult to envisage a mechanism in which the intervening chromatin is directly involved in a mechanism of enhancer-promoter communication such as tracking. Therefore tracking mechanisms are likely limited to enhancers that are close (1-10kb) to their target promoters (Figure 1D). Indeed, the two cases where a facilitated-tracking mechanism has been described, involve a moderate enhancer - promoter distance (Table 2)\textsuperscript{99, 132}, in comparison to intrachromosomal looping which has been described for longer enhancer-promoter distances (Table 2)\textsuperscript{47, 135, 148, 149}. In the latter cases: do random collisions suffice to facilitate these interactions, or do enhancers actively “seek” for targets both downstream and upstream with equal frequency? Clustered enhancers such as LCRs may be formed by sequential priming progressing from the most upstream element to those downstream, generating a directionality toward the final target promoter\textsuperscript{130}. An upstream enhancer (MCS-R2) of the α-globin locus, when relocated downstream of the target genes, still requires interactions with the other upstream enhancers for globin transcription\textsuperscript{47} (reviewed in\textsuperscript{58}). A polarity between several
enhancers has also been shown with the β-globin locus 150. Deletion of the MCS-R2 α-globin enhancer decreases TF occupancy from the most upstream enhancer towards the downstream promoter 53, 58 again suggesting a directionality in the signal. The duplicated α-globin genes in most species have similar or identical promoters, and it is the gene closest to the upstream elements, that is usually expressed at the higher level (reviewed in 151). When more than two α-globin genes are present in cis, the additional genes lying downstream are expressed at even lower levels 152. Thus, it is conceivable that these mechanisms might all be used to regulate a single gene: a looping mechanism between enhancers and promoter for long distance interactions; a tracking mechanism between the different genes of the same cluster, and finally, a linking mechanism between the proximal and the core promoter. Studies using 3C technology and its variations have tended to concentrate attention on long-range interactions, and therefore may have distracted from other possible mechanisms. A study showed that latent enhancers induced by a given stimulus were shown to be frequently at a short distance from target genes 153; therefore more studies analysing proximal enhancers are needed to characterise the nature of other mechanisms of enhancer-promoter communication.

Concluding Remarks

BOX1: Outstanding questions

- Why PolII recruitment at enhancers sometimes occurs early, long before transcription, and sometimes late, when transcription occurs? In the first scenario, eRNA production might be important for downstream events.
What is the order of events leading to gene transcription during differentiation or hormonal stimulation, and which feature is cause vs consequence? Molecular dissection of appropriate model loci are required to address these questions.

Do random collisions suffice to facilitate enhancer-promoter interactions, or do enhancers actively “seek” for targets both downstream and upstream with equal frequency?

Does increased enhancer-promoter distance in higher organisms favour particular mechanisms of interactions between these elements, e.g. looping rather than tracking or linking?

What mechanism would a gene use if the intervening DNA sequence is increased, or abolished, or if a ‘linear tracking blocker’ is inserted?

Some genes are regulated by several remote enhancers located at distances that vary from 1Kb to up to 1Mb. There are also genes in clusters that are regulated by the same remote enhancer (e.g. globin genes). Although these genes can be expressed at different stages of development or in different tissues, they are all expressed in the same orientation and their expression level often reduces with increasing distance from the enhancers 44, 151, 152. During differentiation, enhancers are first primed by pioneer TFs 59, and the signal is subsequently replaced by relay TFs (exchange model). Then, it spreads or loops towards the downstream promoter via other TFs. There is thus a hierarchy among these elements involving a sequential recruitment of TFs, generating the polarity of the transcription signal, from the remote
enhancers towards the promoter they regulate. Most of these enhancers have a role as centres of recruitment of PIC. On one hand, the role of the enhancer would be to deliver the PIC to the promoter, and thus explain the enhancer-dependence for PIC recruitment at the promoter. On the other hand, the mechanism(s) that prevent(s) this transfer at the early stages of activation is unclear. Enhancers and promoter seem to communicate by i) physical association and formation of chromosomal loops – in which the intervening DNA sequence would seem to be irrelevant (Looping Model, Figure 2B) or ii) by spreading a signal though the intervening sequence separating enhancer and promoter (Facilitated-Tracking Model) (Figure 2A). Although short distances between these elements are usually found in simpler organisms, the distance has increased in higher organisms. Has this increased distance favoured other mechanisms of interactions between these elements, e.g. looping rather than linking? What mechanism would a gene use if the intervening DNA sequence is increased, or abolished, or if a ‘tracking blocker’ is inserted. Originally, several studies have addressed the role of a tracking blocker using insulator elements (e.g. 154). However, the caveats with such experiments, is that we know now that CTCF bound elements are involved in the 3D organisation of the genome in looped structures. Thus, the use of ‘linear’ tracking blockers such as the lac repressor 155 or TerF terminator 156) would be more appropriate, and only a couple of studies have addressed this 43, 157. Even for very long-range enhancers, these elements are capable of working at very short distances in enhancer reporter and transgene assays 158. There are many empty experimental boxes to be filled in Table 2, but we hope this review will help the research community to complete the puzzle. High-throughput sequencing studies have enabled the genome-wide mapping of putative enhancers in diverse cell types. Now functional analyses are required to provide the
mechanistic insight into how these enhancers work, and this will be facilitated by genome-editing strategies.

**Acknowledgements**

We are very grateful to Roger Patient, Bob Hill and Abdenour Soufi for discussions. We apologise to those whose publications we were unable to cite due to space limitations. The Vernimmen lab benefits from funding by the British Society for Haematology (BSH) and the Roslin Foundation. Douglas Vernimmen is supported by a Chancellor's Fellowship at The University of Edinburgh and The Roslin Institute receives Institute Strategic Grant funding from the BBSRC. The Bickmore lab benefits from funding from the UK Medical Research Council and ERC Advanced Grant 249956.

**Figure 1. Multi-steps model of long-range gene regulation.**  
A. Enhancers are first primed by pioneer transcription factors binding to nucleosomal DNA. B. A nucleosome free region (NFR) is formed at an enhancer - often spanning more than one nucleosome. This provides a broad accessible platform for the recruitment of large protein complexes. A similar process occurs independently at the proximal promoter. The *linking model* suggests that an activator protein first binds the promoter at a proximal sequence and facilitates the recruitment of a second TF to a site located just downstream the former. This cascade of recruitment occurs until it reaches the core promoter. This builds a landing platform for the general transcription machinery to the TSS (angled arrow). C. The enhancer recruits very large protein complexes, including PIC and Mediator. D. The enhancer is now active and is associated with short bi-directional transcripts. Recruitment of PIC at the enhancer can precede that at the promoter, or may happen simultaneously. Proteins and
chromatin structures are drawn approximately to scale. Note that other complexes discussed in the text (e.g. Integrator, BAF, cohesins, etc) are not included for simplification.

**Figure 2. Mechanisms of enhancer-promoter communication.** A. The facilitated-tracking model is described as a mechanism by which enhancer bound proteins move progressively in an unidirectionally manner towards the promoter, sometimes without leaving the enhancer sequence, and thus results in the formation of a progressive loop that increases its size until it reaches the promoter to form a stable conformation (B). In this model, histone acetylation and TF complexes are transiently detected in the intervening sequence and this precedes transcription. The tracking is associated with unidirectional transcripts detected in the intervening DNA sequence. B. The looping model implies a direct interaction between two chromosomal regions with the looping out of the intervening DNA sequence. A looped structure together with PolII recruitment at the promoter does not always correlate with transcription, but rather with paused PolII. C. Transcription elongation occurs after release of paused PolII, at the onset of looping or afterwards. As in Figure 1, proteins and chromatin structures are drawn approximately to scale.
The number of enhancers (hypersensitive sites, HS) and genes they contain are shown. Note that for the human \(\alpha\)-globin and H\(\alpha\) globin genes, the intervening DNA sequence (I) has been targeted by insertion of an insulator (Ins) or a terminator (Ter) element respectively. Deletion of the promoter of the mouse TCR\(E\) gene includes an Sp1 binding site. Abbreviation: HS: Hypersensitive site. √: Available, and x: no study yet performed.

| Locus | Proposed Mechanism | References | Mouse I-A(DRA) | One HS | 2.3 kb | √ | 1 | x | x | x | x | 7 kb | x | x | x | x | 15 kb | Mouse β2-Try | One HS | 4.2 kb | x | x | x | x | x | x | 159 |
|-------|--------------------|------------|----------------|---------|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mouse I-A(DRA) | One HS | 2.3 kb | √ | 1 | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human PS\(\alpha\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Mouse H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |
| Human H\(\alpha\)-\(\beta\) | One HS | 4.2 kb | x | x | x | x | x | x | x | x | x | 7 kb | x | x | x | x | 159 |

**Table 2.** List of loci that have been analysed by deletion (∆) of Enhancers (E) or Promoters (P) and the mechanisms proposed for these interactions.
References


24
52. Seenundun, S., et al. (2010) UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. EMBO J 29, 1401-1411
74. Soufi, A., and Zaret, K.S. (2013) Understanding impediments to cellular conversion to pluripotency by assessing the earliest events in ectopic transcription factor binding to the genome. Cell Cycle 12, 1487-1491
80. Tan, S.K., et al. (2011) AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. EMBO J 30, 2569-2581
85. Shoemaker, J., et al. (2006) GATA-3 directly remodels the IL-10 locus independently of IL-4 in CD4+ T cells. J Immunol 176, 3470-3479


**A** Enhancer priming

**B** Recruitment of large complexes and formation of open regions

**C** Recruitment of PIC and Mediator at the enhancer

**D** Recruitment of PIC and Mediator at the promoter
A Facilitated-tracking model

B Looping model

C Final looped structure

Enhancer

Tracking?

Promoter

Looping?

5Kb

unidirectional intergenic transcripts

Paused Pol II

Full elongation

Release of paused Pol II
The hierarchy of transcriptional activation: from enhancer to promoter

Douglas Vernimmen & Wendy A. Bickmore

Trends Box:

• Enhancers are first primed by pioneer transcription factors.

• Other transcription factors are likely required for subsequent events.

• There is a hierarchy between enhancers and the promoters that they regulate.

• Enhancers and promoters share similar properties, but differ in the characteristics and the abundance of the RNAs that they produce.

• By recruiting the pre-initiation complex and other proteins, enhancers have a role of increasing the concentration of the transcription machinery at target promoters.
The hierarchy of transcriptional activation: from enhancer to promoter

Douglas Vernimmen & Wendy A. Bickmore

BOX1: Outstanding questions

- Why PolII recruitment at enhancers sometimes occurs early, long before transcription, and sometimes late, when transcription occurs? In the first scenario, eRNA production might be important for downstream events.

- What is the order of events leading to gene transcription during differentiation or hormonal stimulation, and which feature is cause vs consequence? Molecular dissection of appropriate model loci are required to address these questions.

- Do random collisions suffice to facilitate enhancer-promoter interactions, or do enhancers actively “seek” for targets both downstream and upstream with equal frequency?

- Does increased enhancer-promoter distance in higher organisms favour particular mechanisms of interactions between these elements, e.g. looping rather than tracking or linking?

- What mechanism would a gene use if the intervening DNA sequence is increased, or abolished, or if a ‘linear tracking blocker’ is inserted?