Widespread Enhancer Activity from Core Promoters

Alejandra Medina-Rivera, David Santiago-Algarra, Denis Puthier, Salvatore Spicuglia

Title: Wide-spread enhancer activity from core promoters

Alejandra Medina-Rivera[1], David Santiago-Algarra[2,3], Denis Puthier[2,3], Salvatore Spicuglia[2,3]

[1]Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Mexico
[2]Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France.
[3]Equipe Labéllisée Ligue Contre le Cancer

Correspondence: Salvatore.spicuglia@inserm.fr (S. Spicuglia)

Keywords: Enhancer, Promoters, ePromoters, gene regulation, reporter-assays, CRISPR/Cas9

Glossary:

Enhancer: regulatory element that activates transcription over large distances and independently of orientation. These *cis*-regulatory elements are generally located distally with respect to the 5' end of genes.

Promoter: regulatory element capable of inducing gene expression. These *cis*-regulatory elements are generally located in close proximity to the 5' end of genes.

Enhancer RNA (eRNA): eRNAs are non-coding RNAs produced by the enhancers. They are generally non-polyadenylated, low in abundance, unspliced, and retained within the nucleus.

Core promoter: short sequence of around 50 bp that serves as a binding platform for the transcriptional machinery consisting of RNA Pol II and is associated General Transcription Factors

Enhancer/promoter activity: this makes reference to any functional experiment that assesses the propensity of a given regulatory element to act as an enhancer or promoter.

ePromoter: define a promoter element that display enhancer activity in a functional experimental setting.

Transcription Start Site (TSS): It defines the nucleotide position of any transcription initiation event. However, it generally refers to the position of the main 5' end of an mRNA.

Transcription factories: describe the discrete sites where transcription occurs in the nucleus. The factories contain RNA polymerase (under active or inactive status) and the necessary transcription factors (activators and repressors) for transcription.

Genome Wide Association Study (GWAS): A GWAS is intended to detect genomic variants that are found to be associated with a trait or disease.

SNP: Single Nucleotide Polymorphism

Expression Quantitative Trait Loci (eQTL): A genetic polymorphisms whose alleles are associated with gene expression variability are known as expression Quantitative Trait Loci (eQTL)

Abstract

Gene expression in higher eukaryotes is precisely regulated in time and space through the interplay between promoters and gene-distal regulatory regions, known as enhancers. The original definition of enhancers implies the ability to activate gene expression remotely, while promoters entail the capability to locally induce gene expression. Despite the conventional distinction between them, promoters and enhancers share many genomic and epigenomic features. One intriguing finding in the gene regulation field comes from the observation that many core promoter regions display enhancer activity. Recent high-throughput reporter assays along with CRISPR/Cas9-related approaches have indicated that this phenomenon is relatively common and might have strong impact in our global understanding of genome organization and gene expression regulation.

Similarities between enhancers and promoters

The regulation of gene transcription in higher eukaryotes is accomplished through the involvement of transcription start site (TSS)-proximal (promoters) and -distal (enhancers) regulatory elements [1, 2]. The classical distinction between enhancers and promoters generally relies on their location with respect to the 5' end of genes and the enrichment of specific histone modifications. From a functional point of view, an enhancer implies the property of activating a distal promoter, independently of location and orientation with respect to the target genes. In contrast, promoters must be able to initiate transcription locally and induce efficient transcription elongation towards the direction of the gene. However, this basic dichotomy of *cis*-regulatory elements has been challenged by broad similarities between genetic and epigenetic properties of promoters and enhancers and has been the topic of several recent reviews [3-6] (summarised in **Table 1**).

Like promoters, active enhancers are bound by RNA-Polymerase II (RNAPII) and General Transcription Factors (GTF), and transcribe non-coding RNAs (eRNAs) [7-12]. Promoters and enhancers are demarcated by divergent transcription initiation and a well-positioned array of surrounding nucleosomes [7, 10, 13]. While enhancers are generally depleted of CpG islands, they recruit master regulators like CpG-poor promoters [7] and are enriched in core promoter elements [10]. Histone modifications have been commonly used to discriminate between enhancers and promoters [14-16]. For instance, enhancers were found to be enriched in monomethylation of histone H3 Lys4 (H3K4me1) and acetylation of histone H3 Lys27 (H3K27ac). In contrast, gene promoters typically exhibit trimethylated H3K4 (H3K4me3). As a consequence, the presence of H3K27ac accompanied by high levels of H3K4me1 and low H3K4me3 have been used as a proxy for active enhancers [17]. However, recent works have demonstrated that the presence of H3K4me3 is fully compatible with enhancer activity [10, 11, 18-20], the level of H3K4me3 being actually positively correlated with the enhancer strength and eRNA level [7, 10, 12, 21]. Thus, the current view postulates that similar regulatory mechanisms are at play at enhancers and promoters, but differences in H3K4 methylation patterns simply reflect differences in transcription levels between the two types of elements.

Besides the shared architectural characteristics between promoterS and enhancers, some promoter elements have been shown to function as enhancers in ectopic enhancer reporter assays and to form long-range contacts with other promoters [4, 22]. However, whether this fraction of promoters could function as distal-acting enhancers *in vivo* has remained unclear. More recently, high-throughput functional screens and *in vivo* genetic experiments have highlighted the commonality and physiological functions of these enhancer-like promoters, also referred as ePromoters (see below). In the present review, we will describe the different evidences for the existence of enhancer-like promoters and discuss whether they might define a new type of regulatory elements, the implications for the understanding of complex gene regulation in normal development and disease, as well as, for the topological organisation of the genome.

I. Initial evidence of enhancer activity from promoters

Initial characterisation of enhancer elements from the early 80's consisted in isolating DNA sequences able to stimulate transcription of a heterologous promoter using episomal reporter assays [23, 24]. For instance, the first identified enhancer by Schaffner and collaborators in 1981 corresponded to the promoter of a Simian Virus 40 (SV40) early gene [25]. They showed that a 72-repeat sequence motif was sufficient to increase expression of ectopic beta-globin gene by 200 fold and to function over long distances in an orientation-independent fashion relative to the beta-globin gene.

It is worth noting that many of the early characterised enhancers are located close to, or overlapping with, the promoter region of inducible genes, such as metallothioneins, histones of early cleavage stages, viral immediate-early genes (from some papovaviruses, cytomegaloviruses and retroviruses), heat-shock genes and the antiviral interferon genes [24] (**Table 2**). A characteristic example is the *IFNb* enhancer, which is one of the most well-studied enhancers [26]. Although located immediately upstream of the *IFNb* gene, it can also function as a classical enhancer element conferring virus infection-dependent activation of heterologous promoters, even when it is placed kilobases away from the targeted promoter [27, 28]. Interestingly, the enhancer activity of the *IFNb* promoter depends on loop formation mediated by critical sequence-specific transcription factors bound to the regulatory sequences [29]. A more recent study reported that a promoter located upstream of the adeno-associated virus type 2 (AAV2) genome also display liver-specific enhancer activity, a finding that might explain the pathogenic association between AAV2 integration events and human hepatocellular carcinoma through insertional dysregulation of cancer driver genes via enhancer-mediated effects [30].

A common characteristic of most of the aforementioned promoters is that they are associated with inducible genes that have to quickly respond to environmental stress, which might take more time or be less efficient with a remote enhancer [24]. These early studies already highlighted that enhancers and promoters are very similar entities with some gene promoters having the intrinsic properties to work as enhancers and raised the possibility that enhancer-like promoters could regulate distal genes in their natural context.

II. Promoter-promoter interactions suggest distal regulation by gene promoters

Mammalian genomes are intricately and dynamically organized into higher-order conformation inside the micron-sized nuclear space [31]. Such three-dimensional (3D) organization of the genome is thought to have a role in the mechanisms of transcription regulation and coordination by mediating dynamic looping between distantly located *cis*-regulatory elements while enabling fine-tuning of gene expression. The development of different molecular methods for capturing the spatial organization of the genome (**Box 1**), such as Chromosome Conformation Capture (3C) and related techniques has provided an unprecedented view of the 3D organization of the genome as well as the spatial resolution of interacting regions [31, 32].

Besides the expected interactions between distal enhancers and promoters of target genes, several observations have led to the notion that promoters participate in long-range regulation of distal genes through promoter-promoter (P-P) interactions. Different 3C-based methods such as 3C carbon copy (5C) [33], Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) [34-36], promoter capture Hi-C (CHi-C) [37-39] or HiChIP [40] have revealed extensive P-P interactions. In fact, based on promoter capture Hi-C approaches, P-P interactions represent ~30% of all promoter-centered interactions [41], suggesting that this particular type of multigene regulatory networks is common in mammalian cells.

In general, promoters contact other promoters with similar expression levels [34, 36, 38], indicating that 3D contacts between promoters are non-random. Therefore, promoter interaction networks may facilitate the coordinated expression control of associated genes and allow for regulatory crosstalk between them. Within this hypothesis it is plausible that a fraction of these P-P interactions represent a more specific regulatory circuitry, whereby a given promoter might regulate the activity of distal neighbour genes. Epigenetic analyses of P-P interactions identified by RNAPII based ChIA-PET

experiments revealed a strong bias toward higher H3K4me1/me3 ratio [34], thus suggesting potential enhancer like-activity for a fraction of interacting promoters. Interestingly, in this study, two promoters involved in P-P interactions were shown to function as enhancers of the other associated promoter by luciferase reporter assays.

III. High-throughput reporter assays highlight frequent enhancer activity from promoter elements

In recent years, various powerful techniques that incorporate high-throughput sequencing into reporter assays have enabled quantitative and straightforward measurements of enhancer activity of thousands of regulatory elements [42] (**Box 2**). In particular, two approaches have been widely used in recent years: Massively Parallel Reporter Assay (MPRA) and Self-Transcribing Active Regulatory Region sequencing (STARR-seq). One interest of high-throughput enhancer assays is the possibility to explore enhancer function without preconceived notions, thus potentially leading to new unforeseen findings. Indeed one intriguing and recurrent observation of several episomal assays is that many core promoter regions display enhancer activity [22, 42-50].

Using STARR-seq, Zabidi *et al.* screened the whole fly genome with the use of different core promoters from either ubiquitously expressed housekeeping genes or developmentally regulated and cell-type-specific genes [44]. They found that promoter-proximal enhancers mainly regulate promoters of housekeeping genes, while promoters of developmental genes required distally located enhancers. Several independent studies in mammals also reported widespread enhancer activity from TSS-proximal regions. Ernst *et al.* assessed the enhancer activity of a large selection of DNAse I hypersensitivity sites (DHSs) across several human cell lines and found that a significant subset of active enhancers overlap the TSS of genes [51]. Nguyen *et al*. performed a functional comparison of a subset of promoters and enhancers in mouse neurons using an integrative MPRA approach [45]. Interestingly, gene promoters and distal regulatory regions generated similar enhancer activity. By performing STARR-seq on enriched targets, we found that TSS-proximal and distal DHSs were similarly enriched for active enhancers [46]. Further systematic assessment of all human core promoters of coding genes demonstrated that 2-3% of promoters displayed enhancer activity in a given cell line [46], this type of promoters were denoted ePromoters. Consistent with these results, two recent whole genome STARR-seq studies performed in human cancer cell lines, LNCaP and HeLa, found that between 650 and 1000 of functionally identified enhancers overlapped a TSS [47, 48], representing 1% and 6% of all active enhancers detected in the respective cell lines.

High-throughput reporter assays have several intrinsic caveats that might over or under-estimate the actual number of promoters with enhancer-like activity [2, 42]. These caveats include, the size of the tested fragments, the heterologous promoters used in the assays, and the fact that candidate enhancers are studied outside their endogenous chromatin context, which is likely required for their *in vivo* function.

Another potential concern is that the enhancer activity in the reporter assays actually reflects intrinsic properties of the promoter (e.g. acting as hotspot for the recruitment of transcription factors), which not necessarily imply enhancer activity *in vivo*. Certainly, an equally valid argument is that episomal reporter assays allow to unbiasedly studying enhancer function independently of any "perturbing" chromatin or genomic context. In any case, it would be interesting to systematically assess enhancer

activity from gene promoters using chromatinized episomal or viral-based high-throughput reporter assays [45, 52-54].

IV. *In vivo* assessment of distal gene regulation by promoter elements

As mentioned above, the fact that that some promoters might display enhancer capacity, when tested in episomal reporter assays, does not necessarily ~~imply~~ implIES that they could influence other promoters *in vivo*. Therefore, a critical issue is whether gene promoters are able to function as *bona-fide* enhancers by regulating distal gene expression in their endogenous context. A pioneer study showed that one enhancer of the α-globin locus located within the intron of the *Nbl1* gene harbours intrinsic promoter activity and induces the expression of a non-coding isoform [55], however, the physiological function of this non-coding transcript remains elusive.

The advent of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing methods allows now to systematically study the role of *cis*-regulatory elements in their endogenous context [56, 57] (**Box 3**). Several independent studies using CRISPR genome editing demonstrated that some promoters function as enhancers in their endogenous context (**Figure 1A**) (**Table 3**). Using a CRISPR/Cas9-based promoter deletion strategy, we showed that selected promoters of coding genes with enhancer activity identified in a human STARR-seq reporter assay (i.e. ePromoters), are indeed involved in *cis*-regulation of distal gene expression in their natural context, therefore functioning as *bona fide* enhancers [46]. These ePromoters were shown to physically interact with the promoters of the regulated genes, in some cases involving several target genes, implying that in these P-P interactions, one promoter acts as an active regulatory element of the other(s). Interestingly, inversion of one of the model promoters still retained significant enhancer activity, suggesting that, like classical distal enhancers, enhancer-like promoters might display orientation independent enhancer activity.

Moreover, Engreitz and col. performed systematic genomic editing of promoters of lncRNAs co-regulated with neighboured coding genes. Out of 12 deleted lncRNA promoters, five resulted in significant reduction in the expression of the associated neighbour gene [58]. Further genetic manipulation of the loci by inserting a polyadenylation site downstream the promoter of the lncRNA, thus blocking transcription without affecting the integrity of the promoter, demonstrated that regulation of the target genes do not require the specific lncRNA transcripts themselves, but instead involves enhancer-like activity of the lncRNA promoters [58]. Another study found similar results for the promoter of a lncRNA located downstream of the *Cdkn1b* gene [59]. Nevertheless, as for the α-globin locus mentioned above, it is difficult to ascertain whether the tested regulatory element is a "functional" promoter of the lncRNA or rather a distal enhancer associated with a long eRNA.

The CRISPR/Cas9 approach has been implemented to assess enhancer function within large genomic regions surrounding a given gene of interest [42, 56]. In these studies, a reporter gene introduced at the place of the target gene is used to monitor gene expression. Then, a tiling single guide RNA (sgRNA) library covering the surrounding genomic regions is screened to identify deleted regions with potential enhancer elements. Interestingly, two independent studies performing such screens of *cis*-regulatory elements also found that the expression of some genes is controlled, at least partially, by distal gene promoters [60, 61] (**Table 3**). In particular, interrogation of a 2 Mb genomic region surrounding the *POUF5F1* locus, using a high-throughput tiling-deletion strategy in human embryonic stem cells identified 45 sequences regulating *POUF5F1* expression in *cis* [60]. Of these, 17 sequences corresponded to promoters of functionally unrelated genes. Interestingly, 14 out of 17

POU5F1-regulating promoters had significant level of chromatin interactions with the *POU5F1* promoter, confirming that enhancer-like activity of promoters require long-range chromatin interactions.

An alternative strategy to assess enhancer activity in the endogenous context is to use a nuclease-deactivated Cas9 (dCas9) fused to an activator or repressor domain to precisely modify gene expression from promoters or distal regulatory elements [57]. By using this approach, another study assessed the functional relevance of two heterologous promoters interacting with the promoter of the T cell inducible gene *CD69* and demonstrated that these distal promoters indeed regulate the expression of *CD69* after T cell activation [40].

V. Features of enhancer-like promoters

It is clear that not all gene promoters display enhancer activity. For instances, in the Engreitz et al. study only a subset of tested promoters had significant enhancer activity [58]. Similarly, in the Dao et al. study, while the ePromoter of the *FAF2* gene is required for the expression of *RNF44* gene, deletion of the *RNF44* promoter did not have any impact on *FAF2* expression [46]. Therefore, what defines enhancer-like promoters and what are the underlying characteristics that entail their enhancer function? First of all, enhancer-like promoters appear to be preferentially associated with housekeeping and stress response genes, including interferon response genes [44, 46, 48, 49]. Consistently, a study in *Drosophila* using random insertion of reporter constructs found that expression of the reporter gene depends on chromosomal contacts with endogenous promoters of housekeeping genes [62], suggesting that promoters of housekeeping genes might influence the expression of neighbour loci.

In comparison to classical promoters and distal enhancers, the enhancer-like promoters (ePromoters) display distinct genomic and epigenomic features. They differ in motif content, transcription factor binding and histone modifications [45, 46, 48]. Indeed, enhancer-like promoters bind higher levels of p300, a cofactor usually associated with active enhancers [17] and display increased ratio of H3K27ac over H3K4me3 [46], this ratio correlating with enhancer activity in different cell lines. Consistent with housekeeping and stress response functions, the enhancer-like promoters are preferentially bound by general inducible transcription factors such as AP1, STAT and ATF/CREB family of transcription factors [45, 46, 48]. High-throughput reporter assays using synthetic sequences with tandem repeats of DNA motifs assessed the intrinsic properties of transcription factor binding sites to display promoter or enhancer activities [45]. The study found that distinct DNA motifs were required for either type of activity. For example, the presence of the AP1 motif resulted in significant enhancer activity, but little promoter activity, while motifs for EGR, CREB, and RFX families of transcription factors generated preferential promoter activity. Thus, it is plausible that within the same regulatory sequence different motifs might provide specific enhancer or promoter functions. Another striking feature of enhancer-like promoters is that they harbour a higher density of distinct motifs and bound transcription factors, key properties shared with distal enhancers [63].

The advent of high-throughput sequencing has allowed to map transcription initiation with an unprecedented sensitivity and resolution [5]. This has revealed that cis-regulatory elements are commonly associated with transcriptional initiation sites flanking the regulatory sequences (**Figure 2**). Promoters can be associated with either unidirectional or bidirectional transcription, in the latter the signal intensity being biased towards the sense of the gene. Enhancers produce RNAs (eRNA) *in vivo* [8, 9, 11] with an initiation and chromatin architecture similar to that of promoters [7, 10, 12, 64]. In particular, enhancers have been shown to generally produce bidirectional unstable transcripts with

no particular orientation bias. While the functional relevance of eRNAs is not fully understood, it is clear that their relative abundance is positively correlated with enhancer activity [7, 12, 64].

In macrophages, promoters highly induced during the immune challenge are characterised by the presence of divergent transcription initiation in which the sense and antisense TSSs are separated by large distances [65]. This in turn correlates with enlarged nucleosome depleted regions and enhancer-like features such as higher transcription factor occupancy, binding of p300 and high level of H3K4me1 and suggest that the (**Figure 2, middle panel**). Thus, the size of the nucleosome-depleted region in bidirectional promoters appears to contribute toward enhancer-like properties. Reminiscent of these findings promoter with enhancer activity are predominantly associated with bidirectional transcription [46]. Similarly, testing gene promoters for enhancer activity in Drosophila embryos revealed that when bidirectionally transcribed, promoters could function as enhancer *in vivo*, while unidirectional promoters generally cannot [64]. Overall, these results point towards an unifying model whereby there is a continuum of *cis*-regulatory activity with some elements acting strictly as either enhancer or promoter, while others function predominantly as an enhancer with weak promoter activity or *vice versa*, yet others can have both strong promoter and enhancer activities [4-6, 10, 64] (**Figure 2**). This spectrum of activities might be highly correlated with the directionality of transcription, which likely reflects the underlying sequence properties. In this context, bidirectional transcription at enhancer-like promoters might provide enlarged nucleosome depleted regions serving as hubs for transcription factor binding and establishment of highly active chromatin to further regulate or enhance proximal and distal gene expression (**Figure 2, middle panel**). This would be particularly relevant in the case of rapid and coordinated regulation of gene expression in response to environmental or intrinsic cellular stimuli.

Another outstanding question is whether promoter and enhancer activities of enhancer-like promoters are correlated (**Figure 1B**). Nguyen et al. compared the enhancer and promoter activities of defined promoter elements using distinct reporter assays. They observed a clear positive correlation between enhancer and promoter activity [45]. Similarly, a recent study developed a transgenic assay in drosophila embryos with dual vectors that simultaneous assesses the elements' ability to function as an enhancer and a promoter *in vivo* [64]. Interestingly, some of the tested promoters harboured concomitant promoter and enhancer activity. Comparison of enhancer activity of Starr-seq defined ePromoters with the expression level of the associated gene (as a proxy of the promoter activity) did not show a strict correlation [46]. However, some of the ePromoters displayed high levels of both promoter and enhancer activity, whereas for others ePromoters both activities were anti-correlated. Consistently, integrative analysis of epigenomes across human tissues revealed that a given genomic region could have epigenetic features of enhancer or promoter in different tissues, suggesting that the type of regulatory activity (i.e. enhancer or promoter) might be tissue-specific [66]. Therefore, it is plausible that depending on the locus, enhancer-like promoters might either coordinate the mRNA expression of clusters of genes (for instances, upon stress response signalling) or display context-dependent enhancer or promoter activities **(Figure 1B)**.

As it could be expected, enhancer-like promoters interact with the promoters of regulated genes [40, 46, 60]. Moreover the frequency of P-P interactions is higher when the interaction involves at least one enhancer-like promoter [46]. This suggest that one of the properties defining enhancer-like promoter might be to favour P-P interactions, likely by recruiting key transcription factors such as ZNF143 or YY1, which are two factors involved in looping [67, 68] and enriched at enhancer-like promoters [46]. However, in a given cell type, the number of promoters involved in P-P interactions surpass the number of enhancer-like promoters that can be found in the same cells [46]. It is therefore

likely that not all P-P interactions require an enhancer-like promoter. Alternatively, it is possible that not all enhancer-like promoters are detected by the enhancer reporter assays. Finally, whether enhancer-like promoters represent a hub of interactions with multiple genes need to be explored in the future.

VI. Promoter-centered transcription factories

The expression of interacting genes within multigene complexes is generally well correlated, suggesting that 3D gene organization contributes to coordination of gene expression programs. Evidence from *in situ* fluorescence studies in the last decade suggests that transcription is not evenly distributed and is instead concentrated within large discrete foci in mammalian nuclei, raising the possibility that genes are organized into ‘‘transcription factories’’ containing RNAPII and other components for transcription [69] **(Figure 3A)**. In the current model of transcription factories, regulatory regions of neighbour genes are clustered together and contribute to the expression of each other by increasing the local concentration of regulatory factors and RNA polymerases which might form non-membrane bound compartments with transcription activating and repressing micro-environments [70]. Such clustering has been reported for NFKB-regulated genes in response to TNF-alpha stimulation [71]. Experimental removal of a gene from the NFKB-dependent multigene complex was shown to directly affect the transcription of its interacting genes, suggesting that co-association of co-regulated genes might contribute to a hierarchy of gene expression control [72]. Building up on the transcription factory model, Hinisz and collaborators recently proposed a phase separation model for transcriptional control, whereby clusters of enhancers and promoters mediate multi-molecular assemblies of protein-nucleic acids complexes providing a general regulatory mechanism to compartmentalize membrane-less nuclear compartments [73]. However, the precise contribution of enhancer-like promoters within these transcription factories is currently unknown.

As mentioned above, the widespread occurrence of P-P interactions suggests that promoter-centered chromatin structure contribute to the 3D organisation of the genome and has provided a structural framework for the postulated transcription factories [34]. Indeed, the P-P interactions appear to define a subset of co-regulated promoters sharing genomic and structural regulatory properties, which may be critical for stabilizing the local 3D interactions and the activity of transcription factories. For instances, compared to the interactions between enhancer and promoters, the P-P interactions form a higher order chromatin structure involving many loci, have highly coordinated expression, and are more resistant to external changes [34, 37, 38, 74-76]. In these promoter-centered transcription factories, promoter-interacting multigene clusters might represent topological units of transcriptional coordination where co-regulated genes might come to close vicinity by P-P interactions, resulting in an optimal stoichiometry of chromatin factors required for modulation of gene expression (**Figure 3A**). The interacting regions can be established or maintained by chromatin bridging proteins such as cohesins and CTCF, which are enriched at the interacting promoters [35, 41, 77].

Given the overall contribution of enhancer-like promoters to the regulation of neighbour genes [40, 46, 58, 60] as well was the intrinsic features described in the previous section (frequently involved in P-P interactions; high density of transcription factor binding, etc), it is tempting to speculate that this type of promoters might play a key role within the transcription factories (**Figure 3B**). In this model, the enhancer-like promoters could either facilitate the assembly or maintenance of the transcription factories by tightening the P-P interactions or bring specific transcriptional regulators required for the regulation of the neighbour genes. In any case, it will be essential to investigate the specific contribution of enhancer-like promoters to the functioning of transcription factories.

VII. Genetic variation within promoters influence distal gene regulation

One of the major endeavours in genomic research in the past decade was the advent of Genome Wide Association Studies (GWAS) in order to identify genetic variants associated with candidate genes for human diseases. Most of these variants are located in non-coding regions [78, 79], hence are more likely to be modifying gene expression regulatory mechanisms [2, 80]. It is possible that genetic variants outside coding regions play a regulatory role, but the target genes of these variants are difficult to identify, in particular when the location of the hit is far away from the neighbouring genes. Regardless of this, most GWAS studies establish plausible causality mechanisms by selecting the closest gene to the associated variant, especially when the variant lies within an intronic region, or in the vicinity of a TSS. However, this assumption has been proven to be biased in several examples (e.g. [81, 82]). In a similar way, it might be envisioned that GWAS variants lying within enhancer-like promoters might regulate the expression of distal disease-causal genes.

While GWAS-reported genetic variants are not easily connected to effects on gene function, genetic polymorphisms can be associated with gene expression variability, these variants are known as expression quantitative trait loci (eQTLs). eQTLs with the higher probability to be causal of gene expression variation, tend to be located in open chromatin regions, such as promoters and enhancers [83], supporting the hypothesis of a possible effect through changes in gene expression regulatory mechanisms. Using the set of enhancer-like promoters (ePromoters) defined in Dao et al. [46], we observed that it is more likely to find an eQTL associated with the expression of a distal gene within an ePromoter as compared to other promoters. Given the functional characteristics of eQTLs it is possible to use the reported effect (beta value) of the eQTL as a proxy of the effect a variant could have on its putative target genes. eQTLs lying within ePromoters tend to have stronger effects on distal gene expression than those in other promoters. Moreover, eQTLs potentially affecting transcription factor binding within ePromoters were biased toward having a positive effect on distal gene expression. Specifically, allelic replacement using CRISPR/Cas9 homologous recombination (**Box 3**) of the reference eQTL allele of two of these ePromoters recapitulated the regulatory function of the eQTL variant in the regulation of distal gene expression.

Several examples from the literature might point toward the relation between disease-associated variants and disrupted regulatory mechanisms. The Type 2 Diabetes associated variant rs11603334 lies within the *ARAP1* promoter and affects PAX6/PAX4 binding in human pancreatic islets [84]. The *ARAP1* promoter displayed enhancer activity in STARR-seq assays [46], and the rs11603334 variant is reported in the Genotype-Tissue Expression (GTEx; http://www.gtexportal.org) database as an eQTL affecting both *ARAP1* and *PDE2* genes, the latter was already suspected by Kulzer et al. to be of possible relevance for Type 2 Diabetes. The *NPPB-NPPA* cluster is associated with several cardiovascular diseases and multiple GWAS variants have been reported within the *NPPB* promoter [85-87]. Functional analysis of double-reporter transgenic mice revealed that the *Nppb* promoter is required for heart hypertrophy-induced *Nppa* expression [88], raising the possibility that the causal mechanism of *NPPB*-promoter variants might be due to dys-regulation of both *NPPB* and *NPPA* mRNAs. Mumbach *et al.* [40] integrated 3D genome wide interaction maps in primary human cells to identify regulatory connectomes linking intergenic mutations to target genes. One of the identified interactions mapped to the rs56375023 and rs17293632 variants associated with Crohn's Disease and lying within a *SMAD3* alternative promoter. Interestingly, this *SMAD3* promoter interacts with another, more upstream, *SMAD3* promoter as well as the *AAGAB* promoter, while functional association was supported by eQTL data.

Besides genetic variants, other types of genomic alterations such as enhancer hijacking by chromosomal translocation, genomic rearrangement or insulator disruption, are common molecular mechanisms resulting in disease-related gene deregulation, including overexpression of oncogenes [89, 90]. It is likely expected that enhancer-like promoters could impact on disease through related mechanisms. Integrating information about enhancer-like promoters (e.g. using high-throughput reporter assays) along with 3D interaction data, eQTL and disease-associated variants (e.g. GWAS) might led to the discovery of disease-associated regulation by distal promoters (**Figure 1C**).

Another way distal promoter regulation might have pathological relevance is by indirect perturbation of genome topology. For instances, Cornelia de Lange syndrome (CdLS) is a complex multisystem developmental disorder caused by mutations in cohesin subunits and regulators [91]. Interestingly, some of the genes deregulated in CdLS are not directly associated with cohesin subunits but are positioned within reach of cohesin-occupied regions through promoter-promoter interactions [92], suggesting that wide gene expression deregulation rely on enhancer-like function of cohesin-bound promoters.

Concluding Remarks

Overall, the reviewed results reveal the commonality and widespread use of promoters as distal enhancers. Furthermore, these finding extend and support the increasing evidences pointing toward a unified model of transcriptional regulation, highlighting broad similarities between enhancers and promoters [3, 4, 6, 10]. Although several of these regulatory elements have been validated *in vivo*, more systematic studies using CRISPR/Cas9-based technology will be needed to assess the actual proportion of promoters functioning as *bona fide* enhancers. For instances, recent developments combining CRISPR/Cas9 screening and single-cell RNA-seq [44], thus enabling high-throughput interrogation of enhancers at single cell resolution and directly linking enhancer function with its target gene(s) might help to provide a more comprehensive view of enhancer-like promoters function in living cells. Whether this phenomenon uncovers non-specific contribution of promoters to gene regulation (e.g. keeping open chromatin structure or a defined 3D topology) or rather a specific enhancer-like activity (defining new types of regulatory elements; i.e. ePromoters), will require further investigations.

These findings also open up the intriguing possibility that developmental traits or disease-associated variants lying within a subset of promoters might directly impact on distal gene expression. While there is already work to be done on the understanding of the molecular mechanisms that govern the enhancer-like activity from promoters in cell type or response specific regulatory systems (see Outstanding Questions), the "ePromoters" concept stresses the fact that the identification of regulatory variant target genes in the context of disease is not a straightforward task, and the door should remain open for new association studies and more complex regulatory networks than previously foreseen.

Acknowledgments

We thank Mauricio Guzmán Araiza for help in figures design. Work in SS's laboratory was supported by recurrent funding from the Inserm and Aix-Marseille University and by ARC (PJA 20151203149), Plan Cancer (P036496) and "*Equipe Labellisée Ligue Contre le Cancer*" grants. A.M.-R.'s laboratory is supported by CONACYT (269449) and "*Programa de Apoyo a Proyectos de investigación e innovación tecnológica"* - Universidad Nacional Autónoma de México (PAPIIT- UNAM) (IA206517) grants. DSA was a fellow of CONACYT-Mexico

Box 1. Genome topology and 3C-based approaches

Interacting genomic regions can be identified by chromosome conformation capture (3C) and its derivative methods, which involve cross-linking distal interacting DNA pieces, proximity ligation and sequencing to map the interactions ([32] and references therein). Variations of 3C can focus on interactions for a small number of genomic bait regions (4C), interactions within specific genomic domains (5C), or analyse the whole set of chromosomal interactions within a cell population (Hi-C). Since the HiC technique requires very high sequencing coverage, alternative methods have been developed allowing exploration of the contacts of a subset of genomic regions, with higher resolution at the same cost. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [34] or HiChIP [40] consider only those interactions that are mediated by a protein of interest by pulling down by chromatin immunoprecipitation only the interacting fragments that include this protein. Other capture approaches have been developed that enable selective enrichment for genome-wide interactions involving, on one end, specific regions of interest; these included capture Hi-C (CHi-C) [37, 38] and HiCap [74]. In these later approaches, promoter elements or DNase hypersensitive sites are generally captured using sequence-specific beads, thus providing a comprehensive view of genomic regions interacting with *cis*-regulatory elements. A major finding of these studies is that the genome contains regions that are defined by high levels of chromatin interactions occurring within a domain, interspersed with genomic regions with fewer interactions. These regions are generally referred to as topologically associating domains (TADs), and studies have shown that their borders are conserved across mammalian cell types and even across mammalian species [32].

Box 2. High-throughput reporter assays

Episomal reporter assays have been widely used to characterize putative regulatory regions. Several high-throughput strategies have been developed, enabling the simultaneous analysis of hundreds of thousands of reporter plasmids at once. These methods can be either qualitative (usually based on cell sorting) or quantitative (based on RNA-seq) and designed to test enhancer or promoter activity. Recent quantitative methods have been developed aiming to characterize enhancers. In particular, two approaches massively parallel reporter assay (MPRA) and self-transcribing active regulatory region sequencing (STARR-seq), have been widely used in recent years. The MPRA method consists of the generation of a library of reporter constructs based on microarray synthesis of DNA sequences (generally, tested sequences are cloned upstream of a basal promoter) and unique sequence tags or barcodes (placed in the 3' UTR of the reporter gene). To increase the sensitivity and reproducibility, several barcodes could be added to any given sequence. The reporter library is then transfected into cell lines of interest and RNA sequencing of the barcodes is performed, thus providing a quantitative readout of the regulatory activity of the tested regions. STARR-seq is a massively parallel reporter assay (reviewed in [93]) aimed to identify and quantify transcriptional enhancers directly based on their activity across whole genomes. In brief, a bulk of DNA fragments from arbitrary sources is cloned downstream of a core promoter and into the 3'UTR of a GFP reporter gene. Once in cellular context, active enhancers will activate the promoter and transcribe themselves resulting in reporter transcripts among cellular RNAs. Thus, each reporter transcript contains the reporter gene and the "barcode" of itself. These reporter transcripts can be isolated separately by targeted PCR and eventually detected by deep sequencing. The main advantage over the classical MPRA is that the tested sequence itself is used as a "barcode", substantially simplifying the whole procedure of quantifying the enhancer activity. Capture-based approaches can be used to enrich for particular region of interest. For recent reviews on these methods, see [2, 42]**.**

Box 3. CRISPR/Cas9 based approached to study *cis*-regulatory elements

Since its discovery, the clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) technology has been widely used for genome editing. This method permit to target genome DNA using a small RNA fragment (referred as single-guide RNA; sgRNA). The Cas9 enzyme recognizes the sgRNA/DNA complex and cuts the DNA, triggering the DNA repair system of the cell. This strategy can help to study the cis-regulatory elements in their natural context: I. Deletion of a cis-regulatory element by non-homologous end joining (NHEJ) repair using two

sgRNA flanking the regulatory region of interest (e.g. [46, 58]). II. The CRISPR-mediated mutagenesis permits to create single base mutations by the homologous recombination (HR) repair system using a sgRNA targeting the cis-regulatory element [94] and a donor template containing the mutation. III. Genomic tile-deletion screening using multiple pair of sgRNA to identify cis-regulatory elements of any gene fused with a reporter marker, such as the GFP (e.g. [60]).

Figure Legends

Figure 1. Role of enhancer-like promoters in gene regulation. A) The enhancer-like promoter (red) interacts with one or more distal promoters (green) and activates the expression of neighbour genes (**top**). A given gene might be regulated by several enhancer-like promoters located in the neighbourhood (**middle**). Promoters of LncRNAs (purple) can also have enhancer-like activity and positively regulate the expression of a nearby gene (**bottom**). **B)** The enhancer and promoter activities of enhancer-like promoters could be dissociated (inverse correlation); in this case the same regulatory element displays enhancer activity in one cell type and promoter activity in another cell type. On the other hand, the enhancer and promoter activities could be linked (positive correlation); in this case the enhancer-like promoter exhibits both enhancer and promoter activities in the same cell type. The later model might results in the coordinated regulation of neighbour genes upon stress or cell-type specific signalling. **C)** Genetic variants (e.g. eQTL or GWAS SNP) lying within an enhancer-like promoter might influence the expression of neighbour genes. It is plausible that the physiological impact (trait or disease) of the variant could rely on the deregulation of a distal gene.

Figure 2. Chromatin structure of active regulatory elements. Unidirectional promoters (top) have a main TSS (arrow) and are associated with high levels of H3K4me3 and H3K27ac. Bidirectional promoters (middle) have two unbalanced TSSs defining a larger promoter region than unidirectional promoters and allow the recruitment of a higher number of transcription factors. They are also associated with H3K4me3 and H3K27ac, but the upstream region is also enriched in H3K4me1. The enhancer-like promoters (ePromoter) belong to this category. uRNA: upstream RNA. Active enhancers (bottom) have two balanced TSSs, produced eRNAs in both direction and are enriched for all three histone marks.

Figure 3. Model of enhancer-like promoters and gene regulation. A) Chromatin interactions place promoters in close physically proximity (transcription factories), facilitating the recruitment of transcription factors and RNAPII necessary for the transcription of their associated genes. B) The presence of an enhancer-like promoter (ePromoter) inside the transcription factory could favor the recruitment of high levels of transcription factors and RNAPII.

Tables

Table 1: Features associated with active promoters and enhancers

Features (Active elements)	Promoter	Enhancer
Intrinsic property	Induce transcription of a heterologous reporter gene	Activate a distal (heterologous) promoter
Transcription initiation	Unidirectional or divergent	Mainly divergent
Ratio between sense and antisense transcripts	Biased towards sense transcription	Equilibrated
Transcription elongation	Produce long polyadenylated transcripts	Some enhancers can produce low levels of polyadenylated transcripts
Histone modifications	H3K27ac (H3K4me1<H3K4me3)	H3K27ac (H3K4me1>H3K4me3)
RNAPII and GTF	Present	Present
GpG islands	Majority	Very rare

576
577 **Table 2.** Individual examples of enhancer activity from promoter elements

Gene	Origin	Size (bp)	Distance from TSS	References
Early gene	Simian Virus 40 (SV40)	196	~200	[25] [95]
Early gene	Cytomegalovirus (CMV)	406	-524 to -118	[96]
Hsp70	Xenopus	160	-260 to -100	[97]
Fos	Human	340	-404 to -64	[98]
hMT-IIA	Human	327	-366 to-39	[99]
Mmt-IA	Mouse	114 155	-187 to -73 -194 to -39	[99]
H2A	Urchin	28	-139 to -111	[100]
IFNb	Human	40	-77 to -37	[28]

Table 3. List of experimentally validated promoters with enhancer-like activity in their natural context.

Strategy	Validation	Cell type or line	Gene associated with the enhancer-like promoter (Target gene)	References
Characterisation of DHS associated with α-globin locus	Knock-out mice	Mouse erythrocytes	*Nprl3* (α-globin)	[55]
Co-regulated genes	CRISPR deletion and pAS insertion	mESC	*Bendr*[1]; *Slc30a9* (*Bend4*) *Snhg17*[1] *(Snhg11)* *Linc1405*[1] *(Eomes)* *Gpr19 (Cdkn1b)*	[58]
Transgenic reporter	Reporter assay	Mouse cardiomyocytes	*Nppb* (*Nppa*)	[88]
CRISPR screening (MERA)	None	mESC	*Lrrc2 (Tdgf1)*	[61]
Co-regulated genes	CRISPR deletion pAS Insertion	G1E	*Lockd (Cdkn1b)*	[59]
Reporter assay (CapSTARR-seq)	CRISPR deletion	HeLa K562	*FAF2 (RNF44)* *TAGLN2 (PIGM*; *PEA15)* *CSDE1 (BCAS2*; *SIKE1)* *BAZ2B (MARCH7)* *YPEL4 (UBE2L6)* *METTL21A (CCNYL1)*	[46]
CRISPR screening (CREST-seq)	CRISPR deletion	hESC	17 promoters *(POU5F1)*	[60]
HiChIP	CRISPRa	Jurkat	*CLEC2D*; *CLEC2B (CD69)*	[40]

[1] Promoters of LncRNAs are underlined

Outstanding Questions

- What are the specific components within the promoter region driving promoter *versus* enhancer activity?
- Are promoter and enhancer activities correlated across different tissues?
- Do ePromoter-promoter interactions rely on similar mechanisms as previously shown for enhancer-promoter interactions?
- Are enhancer-like promoters a hub of P-P interactions?
- Are enhancer-like promoters involved in particular biological processes?
- Is the enhancer activity of promoters dependent on the genomic context?
- Is the regulation by enhancer-like promoters a specific process or rather an unspecific contribution to gene expression within transcription factories?
- Is enhancer activity from promoters evolutionary conserved? Could enhancer-like promoters be associated with evolutionarily new genes originated from distal enhancer elements?
- Finally, what are the contributions of enhancer-like activity of promoters to disease?

Highlights

- Promoters and enhancers share architectural and functional properties.
- When tested on episomal reporters, many promoters display enhancer activity.
- In vivo experiments demonstrated that enhancer like promoters function as *bona fide* enhancers.
- Genetic variants lying in enhancer-like promoters might impact on physiological traits or diseases by altering the expression of distal genes.

References

1 Vernimmen, D. and Bickmore, W.A. (2015) The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends in genetics : TIG* 31, 696-708

2 Chatterjee, S. and Ahituv, N. (2017) Gene Regulatory Elements, Major Drivers of Human Disease. *Annual review of genomics and human genetics*

3 Kim, T.K. and Shiekhattar, R. (2015) Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* 162, 948-959

4 Andersson, R. (2015) Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays : news and reviews in molecular, cellular and developmental biology* 37, 314-323

5 Andersson, R.*, et al.* (2015) A unified architecture of transcriptional regulatory elements. *Trends in genetics : TIG* 31, 426-433

6 Tippens, N.D.*, et al.* (2018) Enhancer transcription: what, where, when, and why? *Genes Dev* 32, 1-3

7 Andersson, R.*, et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461

8 De Santa, F.*, et al.* (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS biology* 8, e1000384

9 Kim, T.K.*, et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187

10 Core, L.J.*, et al.* (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics* 46, 1311-1320

11 Koch, F.*, et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-

specific enhancers and promoters. *Nature structural & molecular biology* 18, 956-963
12 Henriques, T.*, et al.* (2018) Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* 32, 26-41
13 He, H.H.*, et al.* (2010) Nucleosome dynamics define transcriptional enhancers. *Nature genetics* 42, 343-347
14 Heintzman, N.D.*, et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108-112
15 Creyghton, M.P.*, et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA*,
16 Rada-Iglesias, A.*, et al.* (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279-283
17 Heintzman, N.D. and Ren, B. (2009) Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev* 19, 541-549
18 Wang, Z.*, et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* 40, 897-903
19 Ernst, J.*, et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43-49
20 Pekowska, A.*, et al.* (2011) H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO journal* 30, 4198–4210
21 Vanhille, L.*, et al.* (2015) High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6, 6905
22 Catarino, R.R.*, et al.* (2017) Promoting transcription over long distances. *Nature genetics* 49, 972-973
23 Marriott, S.J. and Brady, J.N. (1989) Enhancer function in viral and cellular gene regulation. *Biochimica et biophysica acta* 989, 97-110
24 Schaffner, W. (2015) Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* 396, 311-327
25 Banerji, J.*, et al.* (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299-308
26 Thanos, D. and Maniatis, T. (1995) Virus induction of human IFN· gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091-1100
27 Fan, C.M. and Maniatis, T. (1989) Two different virus-inducible elements are required for human beta-interferon gene regulation. *The EMBO journal* 8, 101-110
28 Goodbourn, S.*, et al.* (1985) Human beta-interferon gene expression is regulated by an inducible enhancer element. *Cell* 41, 509-520
29 Nolis, I.K.*, et al.* (2009) Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A* 106, 20222-20227
30 Logan, G.J.*, et al.* (2017) Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome. *Nature genetics* 49, 1267-1273
31 Mishra, A. and Hawkins, R.D. (2017) Three-dimensional genome architecture and emerging technologies: looping in disease. *Genome medicine* 9, 87
32 Sati, S. and Cavalli, G. (2017) Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* 126, 33-44
33 Sanyal, A.*, et al.* (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109-113
34 Li, G.*, et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84-98
35 Handoko, L.*, et al.* (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics* 43, 630-638

36 Kieffer-Kwon, K.R.*, et al.* (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507-1520
37 Mifsud, B.*, et al.* (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics* 47, 598-606
38 Schoenfelder, S.*, et al.* (2015) The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome research* 25, 582-597
39 Javierre, B.M.*, et al.* (2016) Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369-1384 e1319
40 Mumbach, M.R.*, et al.* (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature genetics* 49, 1602-1612
41 Pancaldi, V.*, et al.* (2016) Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol* 17, 152
42 Santiago-Algarra, D.*, et al.* (2017) Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res* 6, 939
43 Arnold, C.D.*, et al.* (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074-1077
44 Zabidi, M.A.*, et al.* (2015) Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556-559
45 Nguyen, T.A.*, et al.* (2016) High-throughput functional comparison of promoter and enhancer activities. *Genome research* 26, 1023-1033
46 Dao, L.T.M.*, et al.* (2017) Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature genetics* 49, 1073-1081
47 Liu, Y.*, et al.* (2017) Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol* 18, 219
48 Muerdter, F.*, et al.* (2018) Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* 15, 141-149
49 Barakat, T.S.*, et al.* (2017) Functional dissection of the enhancer repertoire in human embryonic stem cells. *bioRxiv*
50 Wang, X.*, et al.* (2017) High-resolution genome-wide functional dissection of transcriptional regulatory regions in human. *bioRxiv*
51 Ernst, J.*, et al.* (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol* 34, 1180-1190
52 Shen, S.Q.*, et al.* (2016) Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome research* 26, 238-255
53 Murtha, M.*, et al.* (2014) FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat Methods* 11, 559-565
54 Inoue, F.*, et al.* (2017) A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome research* 27, 38-52
55 Kowalczyk, M.S.*, et al.* (2012) Intragenic enhancers act as alternative promoters. *Molecular cell* 45, 447-458
56 Montalbano, A.*, et al.* (2017) High-Throughput Approaches to Pinpoint Function within the Noncoding Genome. *Molecular cell* 68, 44-59
57 Lo, A. and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Res* 6
58 Engreitz, J.M.*, et al.* (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*
59 Paralkar, V.R.*, et al.* (2016) Unlinking an lncRNA from Its Associated cis Element. *Molecular cell* 62, 104-110
60 Diao, Y.*, et al.* (2017) A tiling-deletion-based genetic screen for cis-regulatory element

identification in mammalian cells. *Nat Methods* 14, 629-635
61 Rajagopal, N.*, et al.* (2016) High-throughput mapping of regulatory DNA. *Nat Biotechnol* 34, 167-174
62 Corrales, M.*, et al.* (2017) Clustering of Drosophila housekeeping promoters facilitates their expression. *Genome research* 27, 1153-1161
63 Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory modules in animals. *Nature reviews. Genetics* 13, 469-483
64 Mikhaylichenko, O.*, et al.* (2018) The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* 32, 42-57
65 Scruggs, B.S.*, et al.* (2015) Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular cell* 58, 1101-1112
66 Leung, D.*, et al.* (2015) Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 518, 350-354
67 Whalen, S.*, et al.* (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nature genetics* 48, 488-496
68 Weintraub, A.S.*, et al.* (2017) YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171, 1573-1588 e1528
69 Rieder, D.*, et al.* (2012) Transcription factories. *Front Genet* 3, 221
70 Feuerborn, A. and Cook, P.R. (2015) Why the activity of a gene depends on its neighbors. *Trends in genetics : TIG* 31, 483-490
71 Papantonis, A.*, et al.* (2010) Active RNA polymerases: mobile or immobile molecular machines? *PLoS biology* 8, e1000419
72 Fanucchi, S.*, et al.* (2013) Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606-620
73 Hnisz, D.*, et al.* (2017) A Phase Separation Model for Transcriptional Control. *Cell* 169, 13-23
74 Sahlen, P.*, et al.* (2015) Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol* 16, 156
75 Zhu, Y.*, et al.* (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat Commun* 7, 10812
76 Barbieri, M.*, et al.* (2017) Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nature structural & molecular biology* 24, 515-524
77 Merkenschlager, M. (2010) Cohesin: a global player in chromosome biology with local ties to gene regulation. *Curr Opin Genet Dev* 20, 555-561
78 MacArthur, J.*, et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* 45, D896-D901
79 Maurano, M.T.*, et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190-1195
80 Deplancke, B.*, et al.* (2016) The Genetics of Transcription Factor DNA Binding Variation. *Cell* 166, 538-554
81 Claussnitzer, M.*, et al.* (2015) FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *The New England journal of medicine* 373, 895-907
82 Gupta, R.M.*, et al.* (2017) A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* 170, 522-533 e515
83 Battle, A.*, et al.* (2017) Genetic effects on gene expression across human tissues. *Nature* 550, 204-213
84 Kulzer, J.R.*, et al.* (2014) A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am J Hum Genet* 94, 186-

197
85 Del Greco, M.F.*, et al.* (2011) Genome-wide association analysis and fine mapping of NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB gene cluster. *Hum Mol Genet* 20, 1660-1671
86 Ellis, K.L.*, et al.* (2011) Association of genetic variation in the natriuretic peptide system with cardiovascular outcomes. *Journal of molecular and cellular cardiology* 50, 695-701
87 Fox, A.A.*, et al.* (2009) Natriuretic peptide system gene variants are associated with ventricular dysfunction after coronary artery bypass grafting. *Anesthesiology* 110, 738-747
88 Sergeeva, I.A.*, et al.* (2016) Identification of a regulatory domain controlling the Nppa-Nppb gene cluster during heart development and stress. *Development* 143, 2135-2146
89 Bradner, J.E.*, et al.* (2017) Transcriptional Addiction in Cancer. *Cell* 168, 629-643
90 Smith, E. and Shilatifard, A. (2014) Enhancer biology and enhanceropathies. *Nature structural & molecular biology* 21, 210-219
91 Boyle, M.I.*, et al.* (2015) Cornelia de Lange syndrome. *Clin Genet* 88, 1-12
92 Boudaoud, I.*, et al.* (2017) Connected Gene Communities Underlie Transcriptional Changes in Cornelia de Lange Syndrome. *Genetics* 207, 139-151
93 Muerdter, F.*, et al.* (2015) STARR-seq - Principles and applications. *Genomics*
94 Canver, M.C.*, et al.* (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192-197
95 Benoist, C. and Chambon, P. (1981) In vivo sequence requirements of the SV40 early promotor region. *Nature* 290, 304-310
96 Boshart, M.*, et al.* (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41, 521-530
97 Bienz, M. and Pelham, H.R. (1986) Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp70 gene and when linked to a heterologous promoter. *Cell* 45, 753-760
98 Deschamps, J.*, et al.* (1985) Identification of a transcriptional enhancer element upstream from the proto-oncogene fos. *Science* 230, 1174-1177
99 Serfling, E.*, et al.* (1985) Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *The EMBO journal* 4, 3851-3859
100 Grosschedl, R. and Birnstiel, M.L. (1982) Delimitation of far upstream sequences required for maximal in vitro transcription of an H2A histone gene. *Proc Natl Acad Sci U S A* 79, 297-301

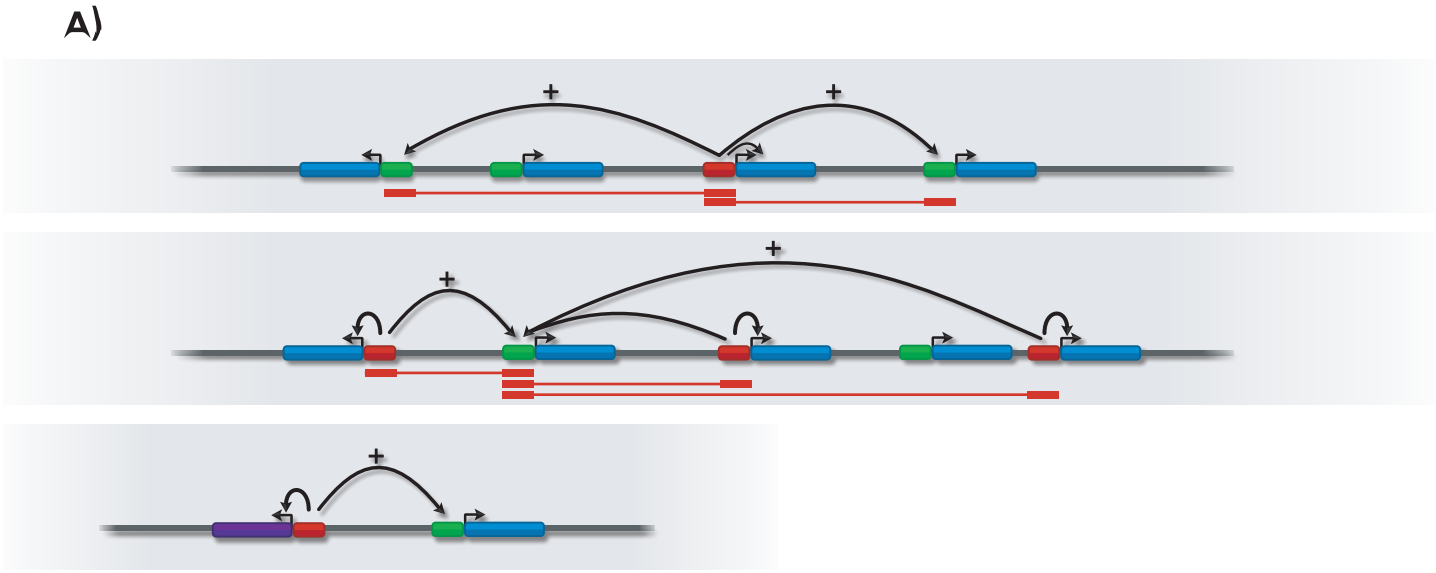
A)
B)
Inverse correlation
Positive correlation
Cell type A
Cell type B
C)
eQTL
T
C
Causal disease?
Promoter
Enhancer-like promoter (ePromoter)
Coding gene
LncRNA
P-P Interaction
mRNA

Figure 1

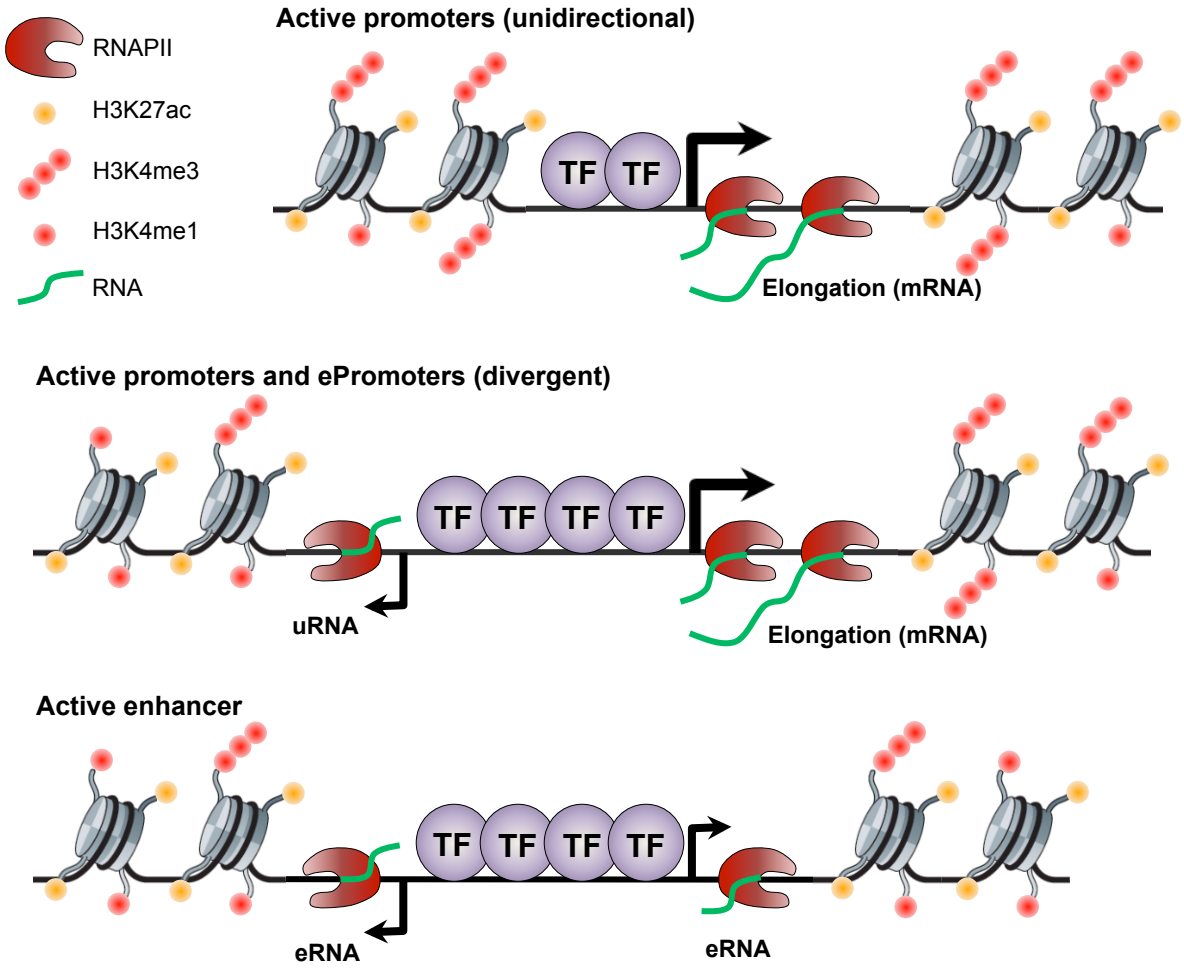
RNAPII
H3K27ac
H3K4me3
H3K4me1
RNA
Active promoters (unidirectional)
TF
TF
Elongation (mRNA)
Active promoters and ePromoters (divergent)
TF
TF
TF
TF
uRNA
Elongation (mRNA)
Active enhancer
TF
TF
TF
TF
eRNA
eRNA

Figure 2

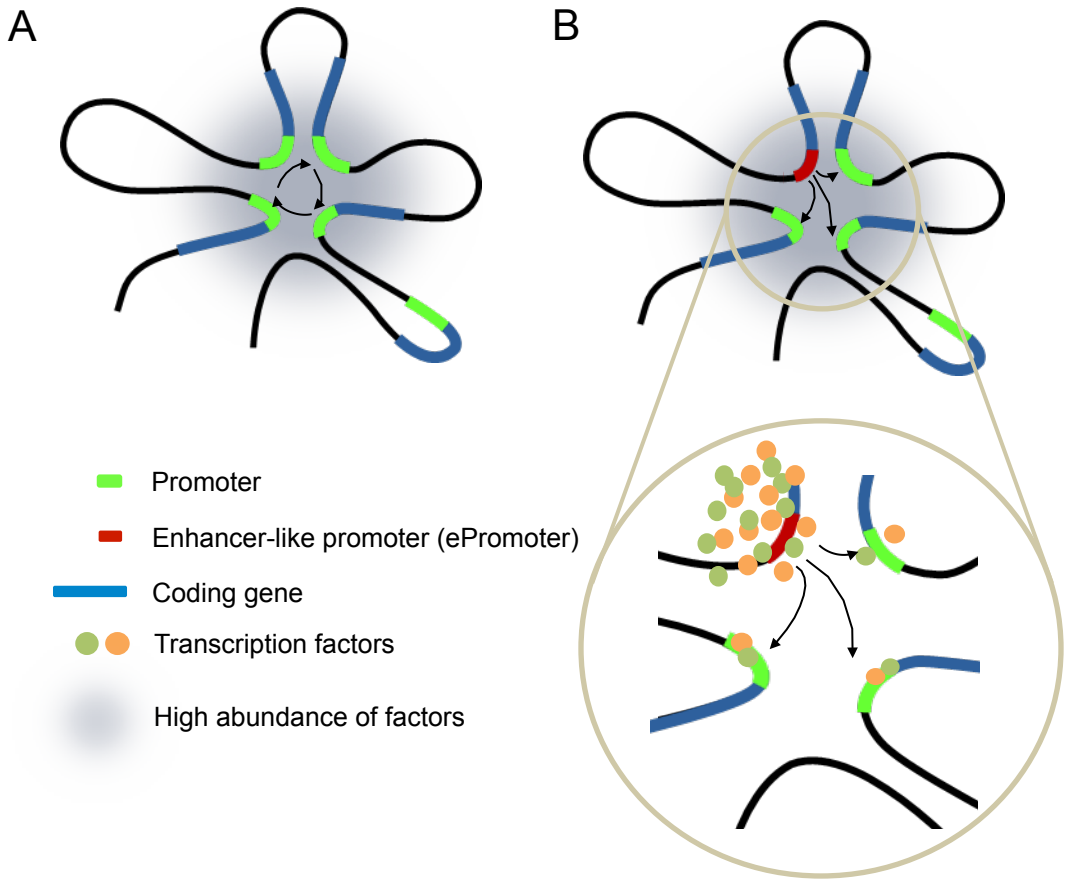
A
B
Promoter
Enhancer-like promoter (ePromoter)
Coding gene
Transcription factors
High abundance of factors

Figure 3

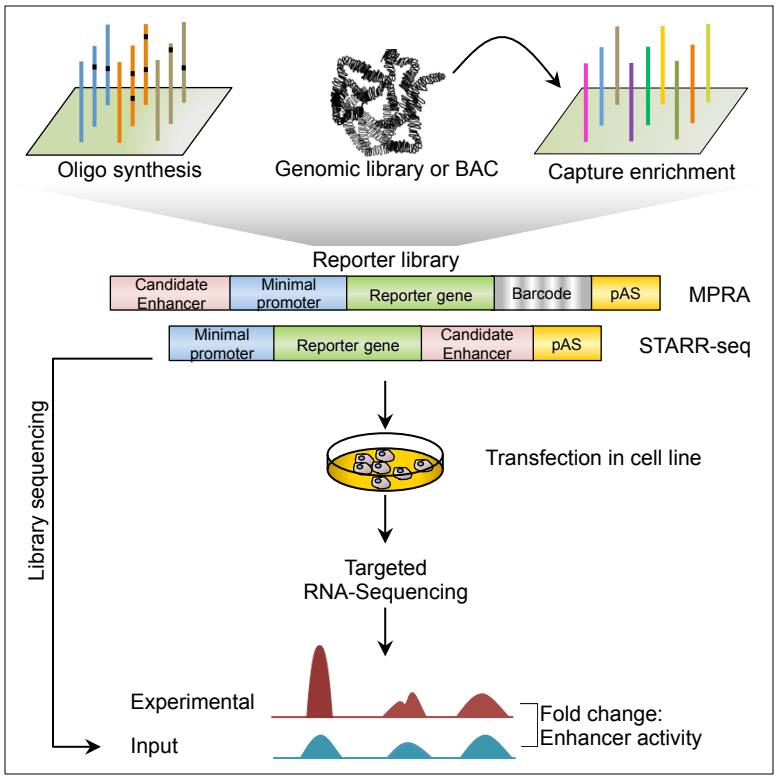
Oligo synthesis
Genomic library or BAC
Capture enrichment
Reporter library
Candidate Enhancer
Minimal promoter
Reporter gene
Barcode
pAS
MPRA
Minimal promoter
Reporter gene
Candidate Enhancer
pAS
STARR-seq
Library sequencing
Transfection in cell line
Targeted RNA-Sequencing
Experimental
Input
Fold change:
Enhancer activity

Figure Box 2

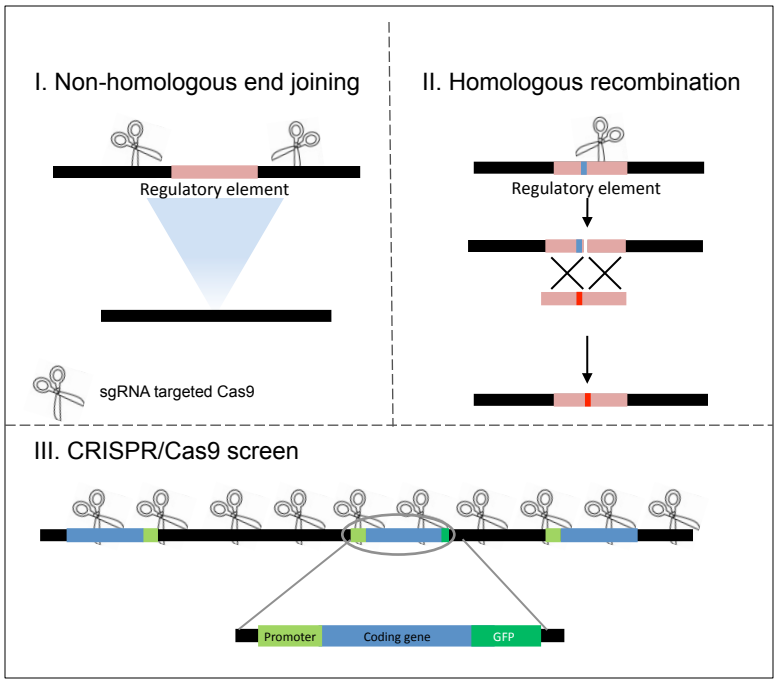
I. Non-homologous end joining
Regulatory element
sgRNA targeted Cas9
II. Homologous recombination
Regulatory element
III. CRISPR/Cas9 screen
Promoter
Coding gene
GFP

Figure Box 3