



HAL
open science

Widespread Enhancer Activity from Core Promoters

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

► **To cite this version:**

Alejandra Medina-Rivera, David Santiago-Algarra, Denis Puthier, Salvatore Spicuglia. Widespread Enhancer Activity from Core Promoters. Trends in Biochemical Sciences, 2018, 43 (6), pp.452 - 468. 10.1016/j.tibs.2018.03.004 . hal-01808594

HAL Id: hal-01808594

<https://amu.hal.science/hal-01808594>

Submitted on 15 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **Title:** Wide-spread enhancer activity from core promoters

2
3 Alejandra Medina-Rivera¹, David Santiago-Algarra^{2,3}, Denis Puthier^{2,3}, Salvatore Spicuglia^{2,3}

4
5 ¹Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional
6 Autónoma de México, Juriquilla, Mexico

7 ²Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France.

8 ³Equipe Labéllisée Ligue Contre le Cancer

9
10 **Correspondence:** Salvatore.spicuglia@inserm.fr (S. Spicuglia)

11
12 **Keywords:** Enhancer, Promoters, ePromoters, gene regulation, reporter-assays, CRISPR/Cas9

13
14 **Glossary:**

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

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

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

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

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

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

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

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

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

35 **SNP:** Single Nucleotide Polymorphism

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

38
39 **Abstract**

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

49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96

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 promoter 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.

97

98 It is worth noting that many of the early characterised enhancers are located close to, or overlapping
99 with, the promoter region of inducible genes, such as metallothioneins, histones of early cleavage
100 stages, viral immediate-early genes (from some papovaviruses, cytomegaloviruses and retroviruses),
101 heat-shock genes and the antiviral interferon genes [24] (**Table 2**). A characteristic example is the
102 *IFN β* enhancer, which is one of the most well-studied enhancers [26]. Although located immediately
103 upstream of the *IFN β* gene, it can also function as a classical enhancer element conferring virus
104 infection-dependent activation of heterologous promoters, even when it is placed kilobases away from
105 the targeted promoter [27, 28]. Interestingly, the enhancer activity of the *IFN β* promoter depends on
106 loop formation mediated by critical sequence-specific transcription factors bound to the regulatory
107 sequences [29]. A more recent study reported that a promoter located upstream of the adeno-
108 associated virus type 2 (AAV2) genome also display liver-specific enhancer activity, a finding that
109 might explain the pathogenic association between AAV2 integration events and human hepatocellular
110 carcinoma through insertional dysregulation of cancer driver genes via enhancer-mediated effects
111 [30].

112

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

119

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

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

129

130 Besides the expected interactions between distal enhancers and promoters of target genes, several
131 observations have led to the notion that promoters participate in long-range regulation of distal genes
132 through promoter-promoter (P-P) interactions. Different 3C-based methods such as 3C carbon copy
133 (5C) [33], Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) [34-36],
134 promoter capture Hi-C (CHi-C) [37-39] or HiChIP [40] have revealed extensive P-P interactions. In
135 fact, based on promoter capture Hi-C approaches, P-P interactions represent ~30% of all promoter-
136 centered interactions [41], suggesting that this particular type of multigene regulatory networks is
137 common in mammalian cells.

138

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

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

149

150 **III. High-throughput reporter assays highlight frequent enhancer activity from promoter** 151 **elements**

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

160

161 Using STARR-seq, Zabidi *et al.* screened the whole fly genome with the use of different core
162 promoters from either ubiquitously expressed housekeeping genes or developmentally regulated and
163 cell-type-specific genes [44]. They found that promoter-proximal enhancers mainly regulate
164 promoters of housekeeping genes, while promoters of developmental genes required distally located
165 enhancers. Several independent studies in mammals also reported widespread enhancer activity from
166 TSS-proximal regions. Ernst *et al.* assessed the enhancer activity of a large selection of DNase I
167 hypersensitivity sites (DHSs) across several human cell lines and found that a significant subset of
168 active enhancers overlap the TSS of genes [51]. Nguyen *et al.* performed a functional comparison of a
169 subset of promoters and enhancers in mouse neurons using an integrative MPRA approach [45].
170 Interestingly, gene promoters and distal regulatory regions generated similar enhancer activity. By
171 performing STARR-seq on enriched targets, we found that TSS-proximal and distal DHSs were
172 similarly enriched for active enhancers [46]. Further systematic assessment of all human core
173 promoters of coding genes demonstrated that 2-3% of promoters displayed enhancer activity in a
174 given cell line [46], this type of promoters were denoted ePromoters. Consistent with these results,
175 two recent whole genome STARR-seq studies performed in human cancer cell lines, LNCaP and
176 HeLa, found that between 650 and 1000 of functionally identified enhancers overlapped a TSS [47,
177 48], representing 1% and 6% of all active enhancers detected in the respective cell lines.

178

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

184

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

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

192

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

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

201

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

215

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

226

227 The CRISPR/Cas9 approach has been implemented to assess enhancer function within large genomic
228 regions surrounding a given gene of interest [42, 56]. In these studies, a reporter gene introduced at
229 the place of the target gene is used to monitor gene expression. Then, a tiling single guide RNA
230 (sgRNA) library covering the surrounding genomic regions is screened to identify deleted regions
231 with potential enhancer elements. Interestingly, two independent studies performing such screens of
232 *cis*-regulatory elements also found that the expression of some genes is controlled, at least partially,
233 by distal gene promoters [60, 61] (**Table 3**). In particular, interrogation of a 2 Mb genomic region
234 surrounding the *POUF5F1* locus, using a high-throughput tiling-deletion strategy in human
235 embryonic stem cells identified 45 sequences regulating *POUF5F1* expression in *cis* [60]. Of these,
236 17 sequences corresponded to promoters of functionally unrelated genes. Interestingly, 14 out of 17

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

240

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

247

248 **V. Features of enhancer-like promoters**

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

260

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

277

278 The advent of high-throughput sequencing has allowed to map transcription initiation with an
279 unprecedented sensitivity and resolution [5]. This has revealed that cis-regulatory elements are
280 commonly associated with transcriptional initiation sites flanking the regulatory sequences (**Figure**
281 **2**). Promoters can be associated with either unidirectional or bidirectional transcription, in the latter
282 the signal intensity being biased towards the sense of the gene. Enhancers produce RNAs (eRNA) *in*
283 *vivo* [8, 9, 11] with an initiation and chromatin architecture similar to that of promoters [7, 10, 12, 64].
284 In particular, enhancers have been shown to generally produce bidirectional unstable transcripts with

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

287

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

308

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

325

326 As it could be expected, enhancer-like promoters interact with the promoters of regulated genes [40,
327 46, 60]. Moreover the frequency of P-P interactions is higher when the interaction involves at least
328 one enhancer-like promoter [46]. This suggest that one of the properties defining enhancer-like
329 promoter might be to favour P-P interactions, likely by recruiting key transcription factors such as
330 ZNF143 or YY1, which are two factors involved in looping [67, 68] and enriched at enhancer-like
331 promoters [46]. However, in a given cell type, the number of promoters involved in P-P interactions
332 surpass the number of enhancer-like promoters that can be found in the same cells [46]. It is therefore

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

337

338 **VI. Promoter-centered transcription factories**

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

357

358 As mentioned above, the widespread occurrence of P-P interactions suggests that promoter-centered
359 chromatin structure contribute to the 3D organisation of the genome and has provided a structural
360 framework for the postulated transcription factories [34]. Indeed, the P-P interactions appear to define
361 a subset of co-regulated promoters sharing genomic and structural regulatory properties, which may
362 be critical for stabilizing the local 3D interactions and the activity of transcription factories. For
363 instances, compared to the interactions between enhancer and promoters, the P-P interactions form a
364 higher order chromatin structure involving many loci, have highly coordinated expression, and are
365 more resistant to external changes [34, 37, 38, 74-76]. In these promoter-centered transcription
366 factories, promoter-interacting multigene clusters might represent topological units of transcriptional
367 coordination where co-regulated genes might come to close vicinity by P-P interactions, resulting in
368 an optimal stoichiometry of chromatin factors required for modulation of gene expression (**Figure**
369 **3A**). The interacting regions can be established or maintained by chromatin bridging proteins such as
370 cohesins and CTCF, which are enriched at the interacting promoters [35, 41, 77].

371

372 Given the overall contribution of enhancer-like promoters to the regulation of neighbour genes [40,
373 46, 58, 60] as well as the intrinsic features described in the previous section (frequently involved in
374 P-P interactions; high density of transcription factor binding, etc), it is tempting to speculate that this
375 type of promoters might play a key role within the transcription factories (**Figure 3B**). In this model,
376 the enhancer-like promoters could either facilitate the assembly or maintenance of the transcription
377 factories by tightening the P-P interactions or bring specific transcriptional regulators required for the
378 regulation of the neighbour genes. In any case, it will be essential to investigate the specific
379 contribution of enhancer-like promoters to the functioning of transcription factories.

380

381 VII. Genetic variation within promoters influence distal gene regulation

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

393
394 While GWAS-reported genetic variants are not easily connected to effects on gene function, genetic
395 polymorphisms can be associated with gene expression variability, these variants are known as
396 expression quantitative trait loci (eQTLs). eQTLs with the higher probability to be causal of gene
397 expression variation, tend to be located in open chromatin regions, such as promoters and enhancers
398 [83], supporting the hypothesis of a possible effect through changes in gene expression regulatory
399 mechanisms. Using the set of enhancer-like promoters (ePromoters) defined in Dao et al. [46], we
400 observed that it is more likely to find an eQTL associated with the expression of a distal gene within
401 an ePromoter as compared to other promoters. Given the functional characteristics of eQTLs it is
402 possible to use the reported effect (beta value) of the eQTL as a proxy of the effect a variant could
403 have on its putative target genes. eQTLs lying within ePromoters tend to have stronger effects on
404 distal gene expression than those in other promoters. Moreover, eQTLs potentially affecting
405 transcription factor binding within ePromoters were biased toward having a positive effect on distal
406 gene expression. Specifically, allelic replacement using CRISPR/Cas9 homologous recombination
407 (**Box 3**) of the reference eQTL allele of two of these ePromoters recapitulated the regulatory function
408 of the eQTL variant in the regulation of distal gene expression.

409
410 Several examples from the literature might point toward the relation between disease-associated
411 variants and disrupted regulatory mechanisms. The Type 2 Diabetes associated variant rs11603334
412 lies within the *ARAPI* promoter and affects PAX6/PAX4 binding in human pancreatic islets [84]. The
413 *ARAPI* promoter displayed enhancer activity in STARR-seq assays [46], and the rs11603334 variant
414 is reported in the Genotype-Tissue Expression (GTEx; <http://www.gtexportal.org>) database as an
415 eQTL affecting both *ARAPI* and *PDE2* genes, the latter was already suspected by Kulzer et al. to be
416 of possible relevance for Type 2 Diabetes. The *NPPB-NPPA* cluster is associated with several
417 cardiovascular diseases and multiple GWAS variants have been reported within the *NPPB* promoter
418 [85-87]. Functional analysis of double-reporter transgenic mice revealed that the *Nppb* promoter is
419 required for heart hypertrophy-induced *Nppa* expression [88], raising the possibility that the causal
420 mechanism of *NPPB*-promoter variants might be due to dys-regulation of both *NPPB* and *NPPA*
421 mRNAs. Mumbach *et al.* [40] integrated 3D genome wide interaction maps in primary human cells to
422 identify regulatory connectomes linking intergenic mutations to target genes. One of the identified
423 interactions mapped to the rs56375023 and rs17293632 variants associated with Crohn's Disease and
424 lying within a *SMAD3* alternative promoter. Interestingly, this *SMAD3* promoter interacts with
425 another, more upstream, *SMAD3* promoter as well as the *AAGAB* promoter, while functional
426 association was supported by eQTL data.

427

428 Besides genetic variants, other types of genomic alterations such as enhancer hijacking by
429 chromosomal translocation, genomic rearrangement or insulator disruption, are common molecular
430 mechanisms resulting in disease-related gene deregulation, including overexpression of oncogenes
431 [89, 90]. It is likely expected that enhancer-like promoters could impact on disease through related
432 mechanisms. Integrating information about enhancer-like promoters (e.g. using high-throughput
433 reporter assays) along with 3D interaction data, eQTL and disease-associated variants (e.g. GWAS)
434 might led to the discovery of disease-associated regulation by distal promoters (**Figure 1C**).

435

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

443

444 **Concluding Remarks**

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

458

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

466

467 **Acknowledgments**

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

474

475

476 **Box 1. Genome topology and 3C-based approaches**

477 Interacting genomic regions can be identified by chromosome conformation capture (3C) and its
478 derivative methods, which involve cross-linking distal interacting DNA pieces, proximity ligation and
479 sequencing to map the interactions ([32] and references therein). Variations of 3C can focus on
480 interactions for a small number of genomic bait regions (4C), interactions within specific genomic
481 domains (5C), or analyse the whole set of chromosomal interactions within a cell population (Hi-C).
482 Since the HiC technique requires very high sequencing coverage, alternative methods have been
483 developed allowing exploration of the contacts of a subset of genomic regions, with higher resolution
484 at the same cost. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [34] or
485 HiChIP [40] consider only those interactions that are mediated by a protein of interest by pulling
486 down by chromatin immunoprecipitation only the interacting fragments that include this protein.
487 Other capture approaches have been developed that enable selective enrichment for genome-wide
488 interactions involving, on one end, specific regions of interest; these included capture Hi-C (ChI-C)
489 [37, 38] and HiCap [74]. In these later approaches, promoter elements or DNase hypersensitive sites
490 are generally captured using sequence-specific beads, thus providing a comprehensive view of
491 genomic regions interacting with *cis*-regulatory elements. A major finding of these studies is that the
492 genome contains regions that are defined by high levels of chromatin interactions occurring within a
493 domain, interspersed with genomic regions with fewer interactions. These regions are generally
494 referred to as topologically associating domains (TADs), and studies have shown that their borders
495 are conserved across mammalian cell types and even across mammalian species [32].

496 **Box 2. High-throughput reporter assays**

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

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

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

528 sgRNA flanking the regulatory region of interest (e.g. [46, 58]). II. The CRISPR-mediated
529 mutagenesis permits to create single base mutations by the homologous recombination (HR) repair
530 system using a sgRNA targeting the cis-regulatory element [94] and a donor template containing the
531 mutation. III. Genomic tile-deletion screening using multiple pair of sgRNA to identify cis-regulatory
532 elements of any gene fused with a reporter marker, such as the GFP (e.g. [60]).
533

534

535 **Figure Legends**

536

537 **Figure 1. Role of enhancer-like promoters in gene regulation.** **A)** The enhancer-like promoter (red)
538 interacts with one or more distal promoters (green) and activates the expression of neighbour genes
539 (**top**). A given gene might be regulated by several enhancer-like promoters located in the
540 neighbourhood (**middle**). Promoters of LncRNAs (purple) can also have enhancer-like activity and
541 positively regulate the expression of a nearby gene (**bottom**). **B)** The enhancer and promoter activities
542 of enhancer-like promoters could be dissociated (inverse correlation); in this case the same regulatory
543 element displays enhancer activity in one cell type and promoter activity in another cell type. On the
544 other hand, the enhancer and promoter activities could be linked (positive correlation); in this case the
545 enhancer-like promoter exhibits both enhancer and promoter activities in the same cell type. The later
546 model might results in the coordinated regulation of neighbour genes upon stress or cell-type specific
547 signalling. **C)** Genetic variants (e.g. eQTL or GWAS SNP) lying within an enhancer-like promoter
548 might influence the expression of neighbour genes. It is plausible that the physiological impact (trait
549 or disease) of the variant could rely on the deregulation of a distal gene.
550

551

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

560

561 **Figure 3. Model of enhancer-like promoters and gene regulation.** **A)** Chromatin interactions place
562 promoters in close physically proximity (transcription factories), facilitating the recruitment of
563 transcription factors and RNAPII necessary for the transcription of their associated genes. **B)** The
564 presence of an enhancer-like promoter (ePromoter) inside the transcription factory could favor the
565 recruitment of high levels of transcription factors and RNAPII.
566

567

568

569

570

569 **Tables**570 **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 |

571

572

573

574

575

576
577

Table 2. Individual examples of enhancer activity from promoter elements

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

580
581
582

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 | <i>Nprl3</i> (α -globin) | [55] |
| Co-regulated genes | CRISPR deletion and pAS insertion | mESC | <i>Bendr</i> ¹ ; <i>Slc30a9</i> (<i>Bend4</i>) <i>Snhg17</i> ¹ (<i>Snhg11</i>) <i>Linc1405</i> ¹ (<i>Eomes</i>) <i>Gpr19</i> (<i>Cdkn1b</i>) | [58] |
| Transgenic reporter | Reporter assay | Mouse cardiomyocytes | <i>Nppb</i> (<i>Nppa</i>) | [88] |
| CRISPR screening (MERA) | None | mESC | <i>Lrrc2</i> (<i>Tdgf1</i>) | [61] |
| Co-regulated genes | CRISPR deletion pAS Insertion | G1E | <i>Lockd</i> (<i>Cdkn1b</i>) | [59] |
| Reporter assay (CapSTARR-seq) | CRISPR deletion | HeLa K562 | <i>FAF2</i> (<i>RNF44</i>) <i>TAGLN2</i> (<i>PIGM</i> ; <i>PEA15</i>) <i>CSDE1</i> (<i>BCAS2</i> ; <i>SIKE1</i>) <i>BAZ2B</i> (<i>MARCH7</i>) <i>YPEL4</i> (<i>UBE2L6</i>) <i>METTL21A</i> (<i>CCNYL1</i>) | [46] |
| CRISPR screening (CREST-seq) | CRISPR deletion | hESC | 17 promoters (<i>POU5F1</i>) | [60] |
| HiChIP | CRISPRa | Jurkat | <i>CLEC2D</i> ; <i>CLEC2B</i> (<i>CD69</i>) | [40] |

583 ¹ Promoters of LncRNAs are underlined

584
585

586 **Outstanding Questions**

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

600

601

602 **Highlights**

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

608

609 **References**

610

611

612 **1** Vernimmen, D. and Bickmore, W.A. (2015) The Hierarchy of Transcriptional Activation:
613 From Enhancer to Promoter. *Trends in genetics : TIG* 31, 696-708

614 **2** Chatterjee, S. and Ahituv, N. (2017) Gene Regulatory Elements, Major Drivers of Human
615 Disease. *Annual review of genomics and human genetics*

616 **3** Kim, T.K. and Shiekhatar, R. (2015) Architectural and Functional Commonalities between
617 Enhancers and Promoters. *Cell* 162, 948-959

618 **4** Andersson, R. (2015) Promoter or enhancer, what's the difference? Deconstruction of
619 established distinctions and presentation of a unifying model. *BioEssays : news and reviews*
620 *in molecular, cellular and developmental biology* 37, 314-323

621 **5** Andersson, R., *et al.* (2015) A unified architecture of transcriptional regulatory elements.
622 *Trends in genetics : TIG* 31, 426-433

623 **6** Tippens, N.D., *et al.* (2018) Enhancer transcription: what, where, when, and why? *Genes*
624 *Dev* 32, 1-3

625 **7** Andersson, R., *et al.* (2014) An atlas of active enhancers across human cell types and
626 tissues. *Nature* 507, 455-461

627 **8** De Santa, F., *et al.* (2010) A large fraction of extragenic RNA pol II transcription sites
628 overlap enhancers. *PLoS biology* 8, e1000384

629 **9** Kim, T.K., *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers.
630 *Nature* 465, 182-187

631 **10** Core, L.J., *et al.* (2014) Analysis of nascent RNA identifies a unified architecture of
632 initiation regions at mammalian promoters and enhancers. *Nature genetics* 46, 1311-1320

633 **11** Koch, F., *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-

634 specific enhancers and promoters. *Nature structural & molecular biology* 18, 956-963
635 **12** Henriques, T., *et al.* (2018) Widespread transcriptional pausing and elongation control at
636 enhancers. *Genes Dev* 32, 26-41
637 **13** He, H.H., *et al.* (2010) Nucleosome dynamics define transcriptional enhancers. *Nature*
638 *genetics* 42, 343-347
639 **14** Heintzman, N.D., *et al.* (2009) Histone modifications at human enhancers reflect global
640 cell-type-specific gene expression. *Nature* 459, 108-112
641 **15** Creyghton, M.P., *et al.* (2010) Histone H3K27ac separates active from poised enhancers
642 and predicts developmental state. *Proc Natl Acad Sci USA*,
643 **16** Rada-Iglesias, A., *et al.* (2011) A unique chromatin signature uncovers early
644 developmental enhancers in humans. *Nature* 470, 279-283
645 **17** Heintzman, N.D. and Ren, B. (2009) Finding distal regulatory elements in the human
646 genome. *Curr Opin Genet Dev* 19, 541-549
647 **18** Wang, Z., *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in
648 the human genome. *Nature genetics* 40, 897-903
649 **19** Ernst, J., *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human
650 cell types. *Nature* 473, 43-49
651 **20** Pekowska, A., *et al.* (2011) H3K4 tri-methylation provides an epigenetic signature of
652 active enhancers. *The EMBO journal* 30, 4198-4210
653 **21** Vanhille, L., *et al.* (2015) High-throughput and quantitative assessment of enhancer
654 activity in mammals by CapStarr-seq. *Nat Commun* 6, 6905
655 **22** Catarino, R.R., *et al.* (2017) Promoting transcription over long distances. *Nature genetics*
656 49, 972-973
657 **23** Marriott, S.J. and Brady, J.N. (1989) Enhancer function in viral and cellular gene
658 regulation. *Biochimica et biophysica acta* 989, 97-110
659 **24** Schaffner, W. (2015) Enhancers, enhancers - from their discovery to today's universe of
660 transcription enhancers. *Biol Chem* 396, 311-327
661 **25** Banerji, J., *et al.* (1981) Expression of a beta-globin gene is enhanced by remote SV40
662 DNA sequences. *Cell* 27, 299-308
663 **26** Thanos, D. and Maniatis, T. (1995) Virus induction of human IFN- γ gene expression
664 requires the assembly of an enhanceosome. *Cell* 83, 1091-1100
665 **27** Fan, C.M. and Maniatis, T. (1989) Two different virus-inducible elements are required for
666 human beta-interferon gene regulation. *The EMBO journal* 8, 101-110
667 **28** Goodbourn, S., *et al.* (1985) Human beta-interferon gene expression is regulated by an
668 inducible enhancer element. *Cell* 41, 509-520
669 **29** Nolis, I.K., *et al.* (2009) Transcription factors mediate long-range enhancer-promoter
670 interactions. *Proc Natl Acad Sci U S A* 106, 20222-20227
671 **30** Logan, G.J., *et al.* (2017) Identification of liver-specific enhancer-promoter activity in the
672 3' untranslated region of the wild-type AAV2 genome. *Nature genetics* 49, 1267-1273
673 **31** Mishra, A. and Hawkins, R.D. (2017) Three-dimensional genome architecture and
674 emerging technologies: looping in disease. *Genome medicine* 9, 87
675 **32** Sati, S. and Cavalli, G. (2017) Chromosome conformation capture technologies and their
676 impact in understanding genome function. *Chromosoma* 126, 33-44
677 **33** Sanyal, A., *et al.* (2012) The long-range interaction landscape of gene promoters. *Nature*
678 489, 109-113
679 **34** Li, G., *et al.* (2012) Extensive promoter-centered chromatin interactions provide a
680 topological basis for transcription regulation. *Cell* 148, 84-98
681 **35** Handoko, L., *et al.* (2011) CTCF-mediated functional chromatin interactome in
682 pluripotent cells. *Nature genetics* 43, 630-638

683 **36** Kieffer-Kwon, K.R., *et al.* (2013) Interactome maps of mouse gene regulatory domains
684 reveal basic principles of transcriptional regulation. *Cell* 155, 1507-1520

685 **37** Mifsud, B., *et al.* (2015) Mapping long-range promoter contacts in human cells with high-
686 resolution capture Hi-C. *Nature genetics* 47, 598-606

687 **38** Schoenfelder, S., *et al.* (2015) The pluripotent regulatory circuitry connecting promoters
688 to their long-range interacting elements. *Genome research* 25, 582-597

689 **39** Javierre, B.M., *et al.* (2016) Lineage-Specific Genome Architecture Links Enhancers and
690 Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369-1384 e1319

691 **40** Mumbach, M.R., *et al.* (2017) Enhancer connectome in primary human cells identifies
692 target genes of disease-associated DNA elements. *Nature genetics* 49, 1602-1612

693 **41** Pancaldi, V., *et al.* (2016) Integrating epigenomic data and 3D genomic structure with a
694 new measure of chromatin assortativity. *Genome Biol* 17, 152

695 **42** Santiago-Algarra, D., *et al.* (2017) Recent advances in high-throughput approaches to
696 dissect enhancer function. *F1000Res* 6, 939

697 **43** Arnold, C.D., *et al.* (2013) Genome-wide quantitative enhancer activity maps identified
698 by STARR-seq. *Science* 339, 1074-1077

699 **44** Zabidi, M.A., *et al.* (2015) Enhancer-core-promoter specificity separates developmental
700 and housekeeping gene regulation. *Nature* 518, 556-559

701 **45** Nguyen, T.A., *et al.* (2016) High-throughput functional comparison of promoter and
702 enhancer activities. *Genome research* 26, 1023-1033

703 **46** Dao, L.T.M., *et al.* (2017) Genome-wide characterization of mammalian promoters with
704 distal enhancer functions. *Nature genetics* 49, 1073-1081

705 **47** Liu, Y., *et al.* (2017) Functional assessment of human enhancer activities using whole-
706 genome STARR-sequencing. *Genome Biol* 18, 219

707 **48** Muerdter, F., *et al.* (2018) Resolving systematic errors in widely used enhancer activity
708 assays in human cells. *Nat Methods* 15, 141-149

709 **49** Barakat, T.S., *et al.* (2017) Functional dissection of the enhancer repertoire in human
710 embryonic stem cells. *bioRxiv*

711 **50** Wang, X., *et al.* (2017) High-resolution genome-wide functional dissection of
712 transcriptional regulatory regions in human. *bioRxiv*

713 **51** Ernst, J., *et al.* (2016) Genome-scale high-resolution mapping of activating and repressive
714 nucleotides in regulatory regions. *Nat Biotechnol* 34, 1180-1190

715 **52** Shen, S.Q., *et al.* (2016) Massively parallel cis-regulatory analysis in the mammalian
716 central nervous system. *Genome research* 26, 238-255

717 **53** Murtha, M., *et al.* (2014) FIREWACH: high-throughput functional detection of
718 transcriptional regulatory modules in mammalian cells. *Nat Methods* 11, 559-565

719 **54** Inoue, F., *et al.* (2017) A systematic comparison reveals substantial differences in
720 chromosomal versus episomal encoding of enhancer activity. *Genome research* 27, 38-52

721 **55** Kowalczyk, M.S., *et al.* (2012) Intragenic enhancers act as alternative promoters.
722 *Molecular cell* 45, 447-458

723 **56** Montalbano, A., *et al.* (2017) High-Throughput Approaches to Pinpoint Function within
724 the Noncoding Genome. *Molecular cell* 68, 44-59

725 **57** Lo, A. and Qi, L. (2017) Genetic and epigenetic control of gene expression by CRISPR-
726 Cas systems. *F1000Res* 6

727 **58** Engreitz, J.M., *et al.* (2016) Local regulation of gene expression by lncRNA promoters,
728 transcription and splicing. *Nature*

729 **59** Paralkar, V.R., *et al.* (2016) Unlinking an lncRNA from Its Associated cis Element.
730 *Molecular cell* 62, 104-110

731 **60** Diao, Y., *et al.* (2017) A tiling-deletion-based genetic screen for cis-regulatory element

732 identification in mammalian cells. *Nat Methods* 14, 629-635

733 **61** Rajagopal, N., *et al.* (2016) High-throughput mapping of regulatory DNA. *Nat Biotechnol*

734 34, 167-174

735 **62** Corrales, M., *et al.* (2017) Clustering of *Drosophila* housekeeping promoters facilitates

736 their expression. *Genome research* 27, 1153-1161

737 **63** Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory

738 modules in animals. *Nature reviews. Genetics* 13, 469-483

739 **64** Mikhaylichenko, O., *et al.* (2018) The degree of enhancer or promoter activity is reflected

740 by the levels and directionality of eRNA transcription. *Genes Dev* 32, 42-57

741 **65** Scruggs, B.S., *et al.* (2015) Bidirectional Transcription Arises from Two Distinct Hubs of

742 Transcription Factor Binding and Active Chromatin. *Molecular cell* 58, 1101-1112

743 **66** Leung, D., *et al.* (2015) Integrative analysis of haplotype-resolved epigenomes across

744 human tissues. *Nature* 518, 350-354

745 **67** Whalen, S., *et al.* (2016) Enhancer-promoter interactions are encoded by complex

746 genomic signatures on looping chromatin. *Nature genetics* 48, 488-496

747 **68** Weintraub, A.S., *et al.* (2017) YY1 Is a Structural Regulator of Enhancer-Promoter

748 Loops. *Cell* 171, 1573-1588 e1528

749 **69** Rieder, D., *et al.* (2012) Transcription factories. *Front Genet* 3, 221

750 **70** Feuerborn, A. and Cook, P.R. (2015) Why the activity of a gene depends on its neighbors.

751 *Trends in genetics : TIG* 31, 483-490

752 **71** Papantonis, A., *et al.* (2010) Active RNA polymerases: mobile or immobile molecular

753 machines? *PLoS biology* 8, e1000419

754 **72** Fanucchi, S., *et al.* (2013) Chromosomal contact permits transcription between

755 coregulated genes. *Cell* 155, 606-620

756 **73** Hnisz, D., *et al.* (2017) A Phase Separation Model for Transcriptional Control. *Cell* 169,

757 13-23

758 **74** Sahlen, P., *et al.* (2015) Genome-wide mapping of promoter-anchored interactions with

759 close to single-enhancer resolution. *Genome Biol* 16, 156

760 **75** Zhu, Y., *et al.* (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat*

761 *Commun* 7, 10812

762 **76** Barbieri, M., *et al.* (2017) Active and poised promoter states drive folding of the extended

763 HoxB locus in mouse embryonic stem cells. *Nature structural & molecular biology* 24, 515-

764 524

765 **77** Merkenschlager, M. (2010) Cohesin: a global player in chromosome biology with local

766 ties to gene regulation. *Curr Opin Genet Dev* 20, 555-561

767 **78** MacArthur, J., *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide

768 association studies (GWAS Catalog). *Nucleic acids research* 45, D896-D901

769 **79** Maurano, M.T., *et al.* (2012) Systematic localization of common disease-associated

770 variation in regulatory DNA. *Science* 337, 1190-1195

771 **80** Deplancke, B., *et al.* (2016) The Genetics of Transcription Factor DNA Binding

772 Variation. *Cell* 166, 538-554

773 **81** Claussnitzer, M., *et al.* (2015) FTO Obesity Variant Circuitry and Adipocyte Browning in

774 Humans. *The New England journal of medicine* 373, 895-907

775 **82** Gupta, R.M., *et al.* (2017) A Genetic Variant Associated with Five Vascular Diseases Is a

776 Distal Regulator of Endothelin-1 Gene Expression. *Cell* 170, 522-533 e515

777 **83** Battle, A., *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*

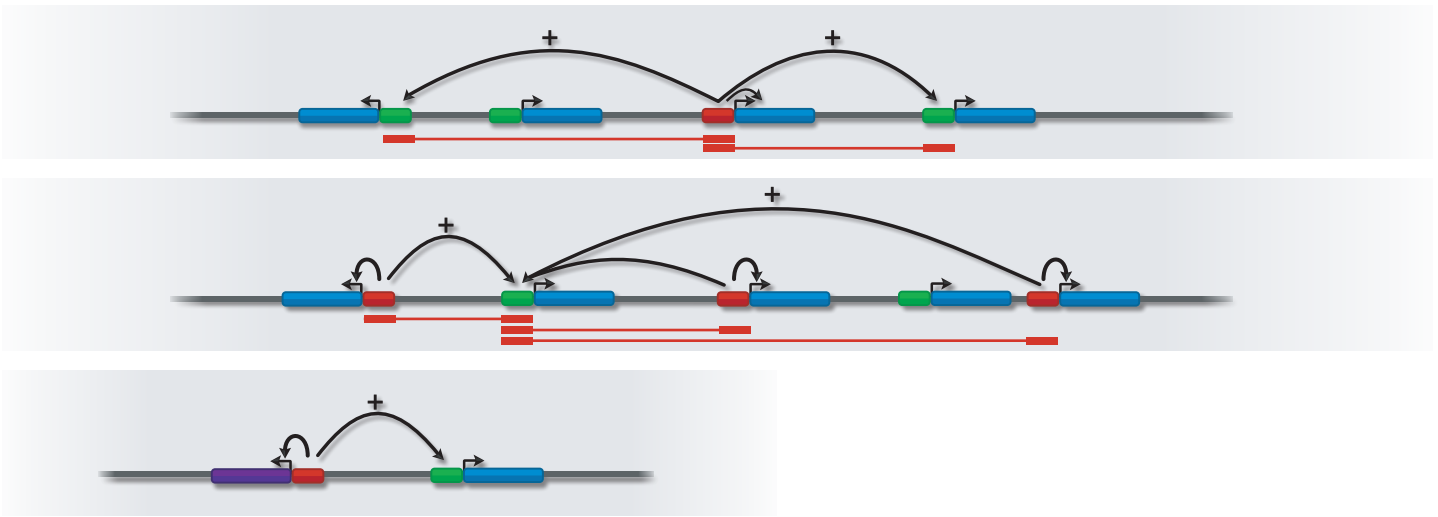
778 550, 204-213

779 **84** Kulzer, J.R., *et al.* (2014) A common functional regulatory variant at a type 2 diabetes

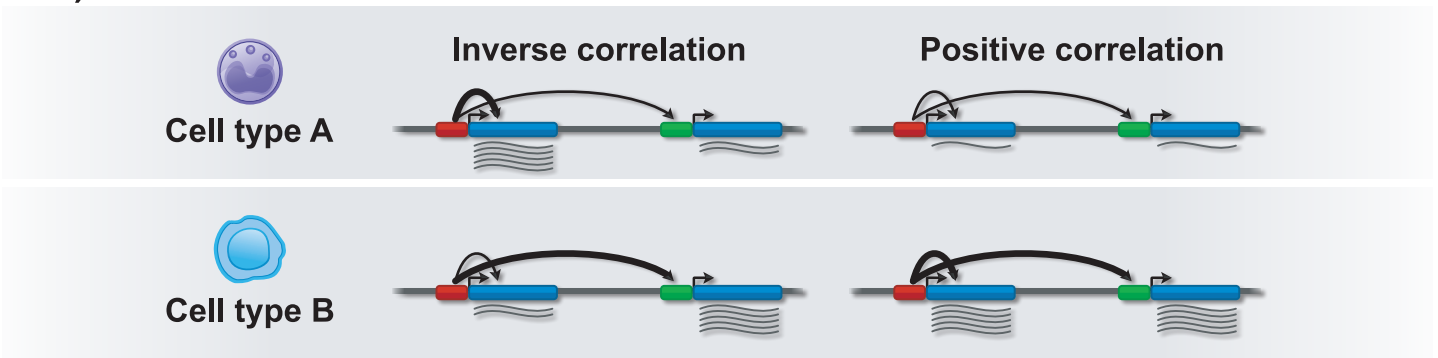
780 locus upregulates ARAP1 expression in the pancreatic beta cell. *Am J Hum Genet* 94, 186-

781 197
782 **85** Del Greco, M.F., *et al.* (2011) Genome-wide association analysis and fine mapping of
783 NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB
784 gene cluster. *Hum Mol Genet* 20, 1660-1671
785 **86** Ellis, K.L., *et al.* (2011) Association of genetic variation in the natriuretic peptide system
786 with cardiovascular outcomes. *Journal of molecular and cellular cardiology* 50, 695-701
787 **87** Fox, A.A., *et al.* (2009) Natriuretic peptide system gene variants are associated with
788 ventricular dysfunction after coronary artery bypass grafting. *Anesthesiology* 110, 738-747
789 **88** Sergeeva, I.A., *et al.* (2016) Identification of a regulatory domain controlling the Nppa-
790 Nppb gene cluster during heart development and stress. *Development* 143, 2135-2146
791 **89** Bradner, J.E., *et al.* (2017) Transcriptional Addiction in Cancer. *Cell* 168, 629-643
792 **90** Smith, E. and Shilatifard, A. (2014) Enhancer biology and enhanceropathies. *Nature*
793 *structural & molecular biology* 21, 210-219
794 **91** Boyle, M.I., *et al.* (2015) Cornelia de Lange syndrome. *Clin Genet* 88, 1-12
795 **92** Boudaoud, I., *et al.* (2017) Connected Gene Communities Underlie Transcriptional
796 Changes in Cornelia de Lange Syndrome. *Genetics* 207, 139-151
797 **93** Muerdter, F., *et al.* (2015) STARR-seq - Principles and applications. *Genomics*
798 **94** Canver, M.C., *et al.* (2015) BCL11A enhancer dissection by Cas9-mediated in situ
799 saturating mutagenesis. *Nature* 527, 192-197
800 **95** Benoist, C. and Chambon, P. (1981) In vivo sequence requirements of the SV40 early
801 promoter region. *Nature* 290, 304-310
802 **96** Boshart, M., *et al.* (1985) A very strong enhancer is located upstream of an immediate
803 early gene of human cytomegalovirus. *Cell* 41, 521-530
804 **97** Bienz, M. and Pelham, H.R. (1986) Heat shock regulatory elements function as an
805 inducible enhancer in the *Xenopus hsp70* gene and when linked to a heterologous promoter.
806 *Cell* 45, 753-760
807 **98** Deschamps, J., *et al.* (1985) Identification of a transcriptional enhancer element upstream
808 from the proto-oncogene *fos*. *Science* 230, 1174-1177
809 **99** Serfling, E., *et al.* (1985) Metal-dependent SV40 viruses containing inducible enhancers
810 from the upstream region of metallothionein genes. *The EMBO journal* 4, 3851-3859
811 **100** Grosschedl, R. and Birnstiel, M.L. (1982) Delimitation of far upstream sequences
812 required for maximal in vitro transcription of an H2A histone gene. *Proc Natl Acad Sci U S A*
813 79, 297-301
814
815

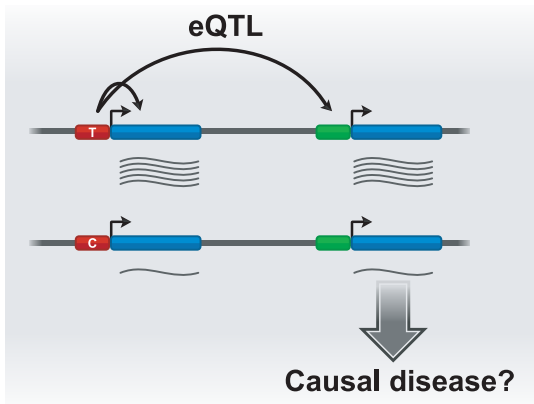
A)



B)



C)



- Promoter
- Enhancer-like promoter (ePromoter)
- Coding gene
- LncRNA
- P-P Interaction
- mRNA

Figure 1

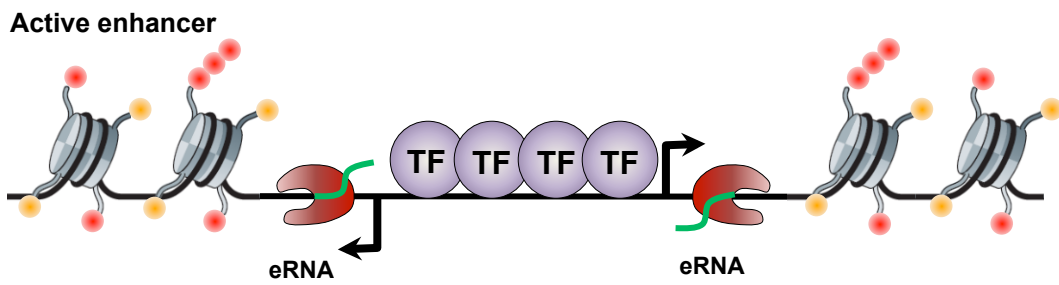
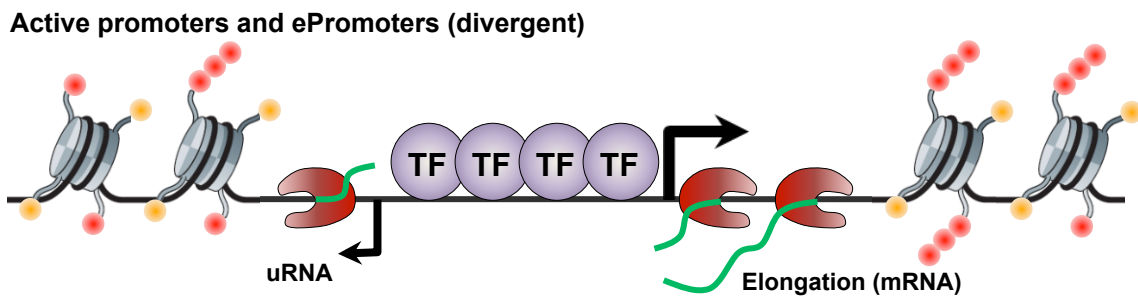
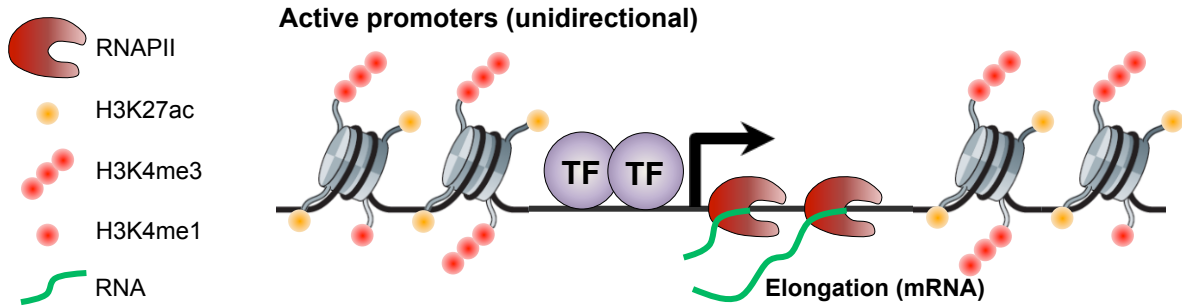


Figure 2

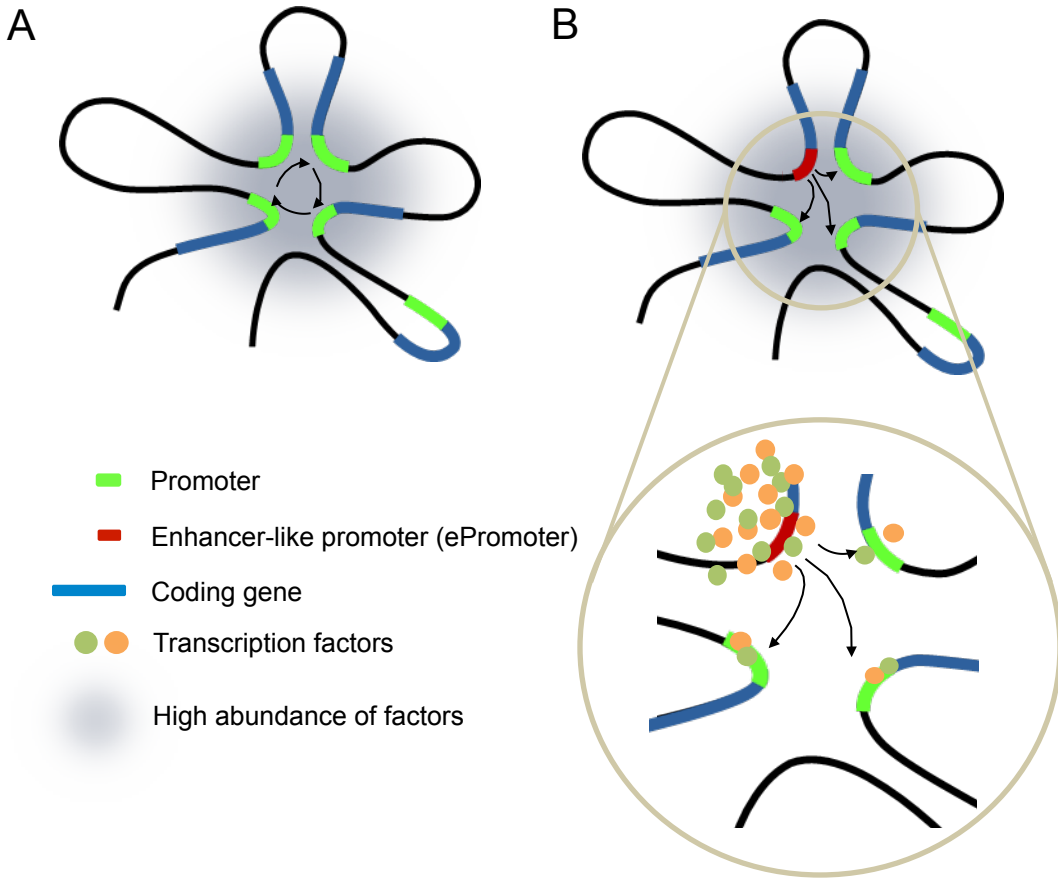


Figure 3

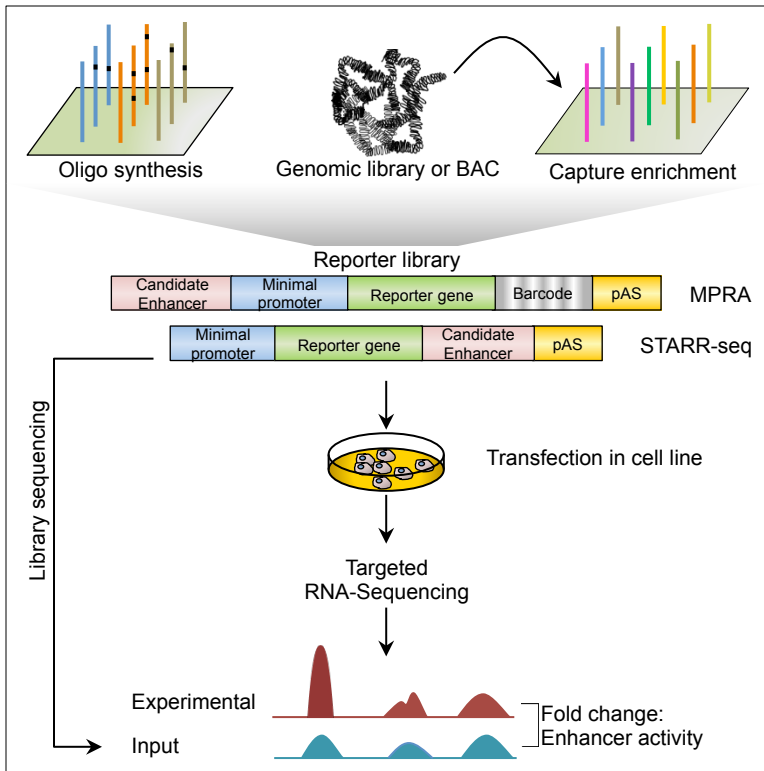


Figure Box 2

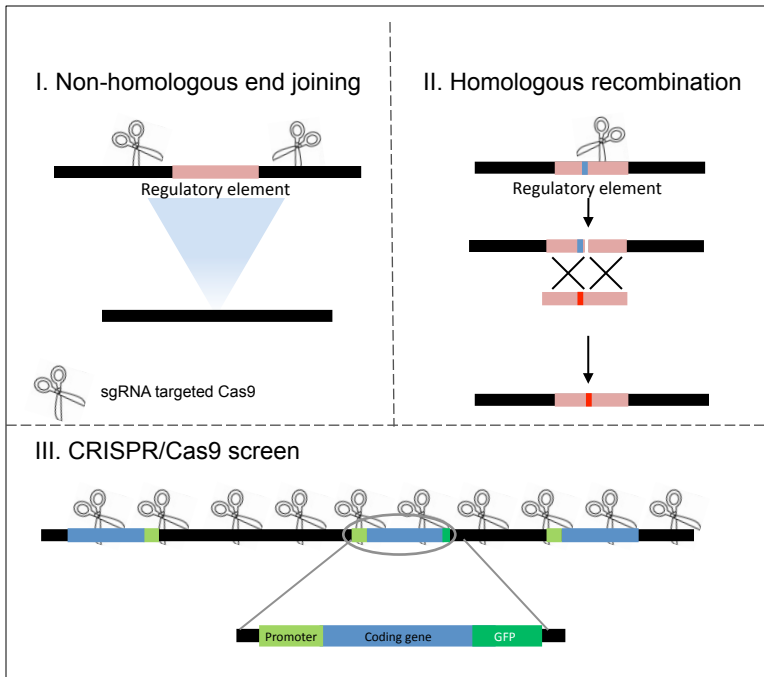


Figure Box 3