



HAL
open science

Epromoters are new players in the regulatory landscape with potential pleiotropic roles

Juliette Malfait, Jing Wan, Salvatore Spicuglia

► **To cite this version:**

Juliette Malfait, Jing Wan, Salvatore Spicuglia. Epromoters are new players in the regulatory landscape with potential pleiotropic roles. *BioEssays*, 2023, 10.1002/bies.202300012 . hal-04200857

HAL Id: hal-04200857

<https://hal.science/hal-04200857v1>

Submitted on 8 Sep 2023

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

HYPOTHESES

Insights & Perspectives

Epromoters are new players in the regulatory landscape with potential pleiotropic roles

Juliette Malfait^{1,2} | Jing Wan^{1,2} | Salvatore Spicuglia^{1,2} 

¹Aix-Marseille University, Inserm, TAGC, UMR1090, Marseille, France

²Equipe Labellisée Ligue Contre le Cancer, LIGUE, Marseille, France

Correspondence

Salvatore Spicuglia, Aix-Marseille University, Inserm, TAGC, Marseille, UMR1090, France.
Email: salvatore.spicuglia@inserm.fr

Funding information

Institut National de la Santé et de la Recherche Médicale (INSERM); Ligue contre le Cancer (Equipe Labellisée Ligue 2023); ANR, Grant/Award Numbers: ANR-18-CE12-0019, ANR-17-CE12-0035; Bettencourt Schueller Foundation (Prix coup d'élan pour la recherche française); French Ministry of Education and Aix-Marseille University; Horizon 2020 research and innovation program, Grant/Award Number: 860002

Abstract

Precise spatiotemporal control of gene expression during normal development and cell differentiation is achieved by the combined action of proximal (promoters) and distal (enhancers) *cis*-regulatory elements. Recent studies have reported that a subset of promoters, termed Epromoters, works also as enhancers to regulate distal genes. This new paradigm opened novel questions regarding the complexity of our genome and raises the possibility that genetic variation within Epromoters has pleiotropic effects on various physiological and pathological traits by differentially impacting multiple proximal and distal genes. Here, we discuss the different observations pointing to an important role of Epromoters in the regulatory landscape and summarize the evidence supporting a pleiotropic impact of these elements in disease. We further hypothesize that Epromoter might represent a major contributor to phenotypic variation and disease.

KEYWORDS

diseases, enhancer, epromoter, gene regulation, pleiotropy, promoter, variants

INTRODUCTION

In higher eukaryotes, gene transcription is regulated through the involvement of regulatory elements that are located near the transcription start site (TSS), called promoters, and those that are located far from TSS, called enhancers. This classical definition implies that enhancers activate gene expression at a distance while promoters induce local gene expression. Although these two elements are distinguishable by their genomics and epigenomics characteristics, a strict dichotomy between *cis*-regulatory elements is being challenged by the broad mechanistic similarities between promoters and enhancers.^[1,2] On one hand, active enhancers are able to initiate transcription and recruit general transcription factors as promoters do. On the other hand, several lines of evidence have shown that a subset of coding-gene promoters is able to function as bona fide enhancers (hereafter named Epromoters), as detailed below.

Given the potential regulation of proximal and distal genes by Epromoters, we hypothesized that genetic variation or mutation at Epromoters might play an important role in physiological and pathological traits. In this review, we first describe the different studies

supporting the physiological relevance of genomic regulation by Epromoters. We then discuss the intrinsic features that might drive the enhancer and promoter activity of Epromoters and whether these are shared or specific properties as compared with typical enhancers and promoters. Finally, we provide current observations supporting the hypothesis of an important and pleiotropic role of Epromoter variation on the ontogeny of different diseases.

SHORT AND LONG-RANGE GENE REGULATION BY EPROMOTERS

Dissection of *cis*-regulatory elements is classically based on gene reporter assays where the tested DNA regions are either placed immediately before the reporter gene (to assess for promoter activity) or placed upstream or downstream of a basic promoter (to assess for enhancer activity). Surprisingly, many of the early characterized enhancers in the 80s and 90s overlapped the promoter of inducible genes,^[2,3] including the first identified enhancer which corresponded to the promoter of the simian virus 40 (SV40) early gene.^[4] Currently,

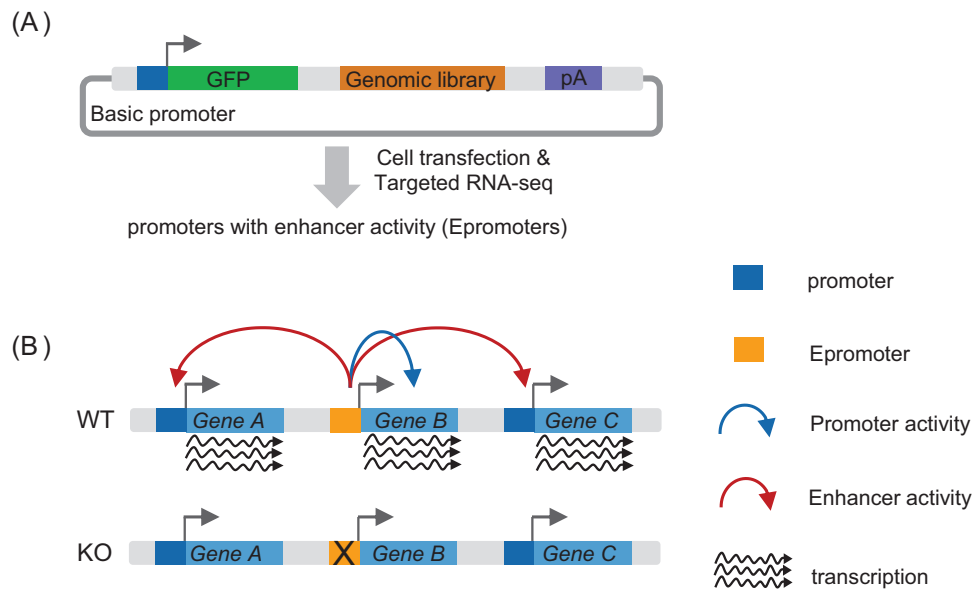


FIGURE 1 Proximal and distal gene regulation by Epromoters. (A) Example of high-throughput reporter assays to assess enhancer activity of genomic regions. A genomic library is cloned downstream of a basic promoter and a GFP reporter gene. With this construct, enhancers will activate their own transcription. The transcripts are then quantified to assess the enhancer activity. In particular, this type of strategy allows identifying of promoter elements with intrinsic enhancer activity (i.e., Epromoters). pA: Poly-adenylation. (B) In wild-type (WT) cells, the Epromoter (yellow) regulates a proximal gene (promoter activity; blue arrow) and distal genes (enhancer activity; red arrows). In Epromoter-deleted cells (KO: knock out), both the proximal and distal gene expressions are impaired.

powerful techniques incorporating high-throughput sequencing into reporter assays enable systematic and straightforward quantification of enhancer activity of *cis*-regulatory elements. Two similar high-throughput reporter assays have been widely used in recent years: Massively Parallel Reporter Assay (MPRA) and Self-Transcribing Active Regulatory Region sequencing (STARR-seq).^[5] One striking observation of these episomal reporter assays, when assessing large genomic regions from *Drosophila* to different mammal cell types, is that many promoters display enhancer activity^[6–15] (Figure 1A). In particular, by assessing all human core promoters of coding genes by STARR-seq, we found that ~3% of promoters exhibited enhancer activity in a given cell line.^[8]

The presence of enhancer activity in some promoters, when tested in episomal reporter assays, does not necessarily mean that they could regulate other promoters in their endogenous context. Thus, whether gene promoters may function as bona fide enhancers by controlling distal gene expression is a critical issue. Several independent studies using mouse transgenics or CRISPR/Cas9 genome editing have demonstrated that the deletion or mutation of some promoters reduces the expression of a distally located gene,^[8,16–19] implying they function as enhancers in their natural context (Figure 1B). So far, around 20 Epromoters have been validated experimentally at their endogenous loci by the different genome editing approaches in mouse or human cell lines (detailed in ref.[2]). High-throughput mapping of regulatory DNA by CRISPR-based screens also found evidence of distal gene regulation by gene promoters.^[20–23] In particular, the repression of 30 (~8%) promoters out of 359 tested by CRISPR inactivation screen, resulted in reduced expression of a distal gene,^[22] supporting the idea that a

substantial amount of promoters display enhancer function at their endogenous loci. Besides coding-gene promoters, promoters of long non-coding RNAs (lncRNAs) have also been shown to display enhancer activity independently of the transcript itself.^[16,24,25] Nevertheless, in such cases, it is challenging to determine whether the tested regulatory element is a distal enhancer associated with a long transcript or rather the promoter of a “functional” lncRNA, which indirectly controls the expression of the neighbor gene.

Further evidence of Epromoter function comes from the analysis of 3D and genetic interactions. In addition to the enhancer-promoter interactions, the analyses of capture Hi-C and similarly derived 3C-based methods have shown that promoter-promoter (P-P) interactions are highly frequent.^[26–32] Similarly, studies of expression quantitative trait loci (eQTLs) have revealed enrichment for genetic variants laying within gene promoters and associated with the regulation of distal genes.^[8,17,29,33,34] Noticeably, some promoters interacting with other promoters, either at the 3D or genetic levels, indeed displayed enhancer activity.^[8,17,23,27,33,35] An important outcome of these observations is that Epromoters often regulate several distal genes (in addition to the proximal one), including clusters of inducible genes,^[18,32,36] suggesting that Epromoters might function as regulatory hubs for the coordinated regulation of gene clusters.

The impact on distal gene expression caused by the deletion or mutation of a promoter might, in principle, be caused by different mechanisms that are independent of a direct enhancer-like function. For instance, a given promoter might regulate distal genes by trans-effects, involving either the transcript itself or the protein-coding gene. However, re-expression of the *FAF2*^[8] or *OAS3*^[18] genes

associated with Epromoters did not rescue the expression of distal genes perturbed by the deletion of these Epromoters. Similarly, genetic dissection combining promoter deletion and the introduction of polyadenylation signals also provided evidence of direct enhancer-like functions.^[16] Another possibility is related to the notion of transcription factories, whereby transcriptional hubs are known to contain many genes and their promoters, and are thought to share limited resources for their expression.^[36–38] Here, the deletion of a promoter within the hub could affect the expression of other distal genes within that hub, without necessarily acting as an enhancer. Although this type of mechanism might be at play in some instances, we have demonstrated that in the case of the aforementioned *FAF2* and *OAS3* Epromoters, the deletion of the promoters associated with the distally regulated genes did not impact the expression of the Epromoter-associated genes, supporting the directionality of the regulation by these Epromoters.^[8,18] Alternatively, the predicted Epromoter might be involved in a complex 3D organization, for example, involving the sequestering of another distal cis-regulatory element and, when deleted, it could indirectly affect the expression of a distal gene without acting as an enhancer for that gene. To date, few examples have unambiguously demonstrated the direct enhancer function of Epromoters, and whether the majority of Epromoters have a direct impact on distal gene expression or work in combination with the aforementioned mechanisms, will need further investigation.

All in all, the discovery that a subset of promoters also functions as bona fide enhancers, has important implications for our understanding of complex gene regulation in normal development and offers a rationale for the frequent repurposing of promoters and enhancers during mammalian evolution.^[1,39,40] Additionally, they raise the intriguing possibility that sequence variation found within Epromoters may have an impact on diseases or physiological traits by directly impacting distal gene expression or changing their relative promoter and enhancer activities. A concomitant question is to understand which intrinsic features drive the specific enhancer and promoter function of Epromoters.

EPROMOTERS SHARE ENHANCER AND PROMOTER PROPERTIES

Although typical enhancers and promoters are logically distinguished by their relative location with respect to the TSS of genes that they regulate, their shared architectural properties have suggested a unifying model of gene regulation by cis-regulatory elements.^[1,2,41–43] Alike promoters, enhancers recruit RNA-Polymerase II (Pol II) and General Transcription Factors (GTF), and transcribe non-coding RNAs (eRNAs).^[41,44–48] Promoters and enhancers are both demarcated by divergent transcription initiation, surrounded by a well-positioned array of nucleosomes, and enriched in core promoter elements.^[41,45,48–51] However, transcripts generated by enhancers are generally bidirectional and less stable. Being generally depleted in CpG islands, enhancers recruit master regulators like CpG-poor promoters,^[48] while some developmental enhancers require a prox-

imal CpG island to function.^[52] Histone post-translational modifications have been used to discriminate between enhancers and promoters.^[53–55] For instance, gene promoters typically exhibit trimethylation of the histone H3 Lys4 (H3K4me3), while enhancers were found to be enriched in monomethylated H4K4 (H3K4me1) and acetylated H4K27 (H3K27ac). As a consequence, the presence of H3K27ac along with high levels of H3K4me1 and low H3K4me3 is commonly used as a proxy for active enhancers.^[54] However, the level of H3K4me3 is positively correlated with the enhancer strength and eRNA level,^[9,41,45,47,48,56,57] and, therefore, the presence of H3K4me3 is fully compatible with the enhancer activity. Thus, the relative enrichment in epigenetic modifications might simply indicate differences in transcriptional levels between the two types of elements, rather than reflecting mutually exclusive functions.

A main intrinsic difference between enhancers and promoters relates to the composition of transcription factor binding sites (TFBS). High promoter activity is associated with a high density of overlapping binding sites for different TFs, in particular, ubiquitously expressed ones, while enhancers are less constrained.^[1] In addition to binding site complexity, the type of TF that binds a cis-regulatory element might influence the relative enhancer or promoter activity.^[58] For instance, the binding of AP1 and NFY is associated with enhancer activity, whereas the binding of CREB, ETS and SP1 is preferentially associated with promoter activity.^[13,59–62] This suggests that the nature of bound TFs might directly contribute to the enhancer and promoter properties of cis-regulatory elements.

Given the aforementioned duality of cis-regulatory elements, current models do not propose that promoter and enhancer activities are mutually exclusive, rather different regulatory elements might accommodate different proportions of these activities^[1] (Figure 2). In such scenarios, Epromoters represent remarkable examples of cis-regulatory elements that share functional and architectural properties with both types of cis-regulatory elements.^[1,2]

High-throughput reporter assays have allowed a systematic comparison of cis-regulatory elements independently of their relative proximity with TSS.^[5] These studies have revealed specific properties of Epromoters that distinguish them from typical enhancers and promoters. When compared to distal enhancers, Epromoters mainly differ by the type of associated TFBS. While distal enhancers are preferentially associated with TFBS for developmental and tissue-specific transcription factors (TFs), Epromoters appear to be associated with TFBS for ubiquitous or inducible TFs.^[7,10,12,15,18] When compared to typical promoters, Epromoters differ by the higher level of unstable bidirectional transcripts,^[8,45,50,51,63] association with co-activators, such as p300^[8] and more frequent P-P interactions.^[8,29] Moreover, Epromoters appear to be associated with a higher density of TF binding and a higher quality of binding sites.^[8,18] The above results are consistent with the fact that Epromoters are preferentially associated with housekeeping and stress response genes,^[7,8,10,12,15] including interferon-response genes.^[8,12,18,64] For instance, we found that Epromoters are significantly associated with the induction of gene clusters during the inflammatory response and that induced Epromoters are characterized by a higher density and quality of Interferon-Stimulated

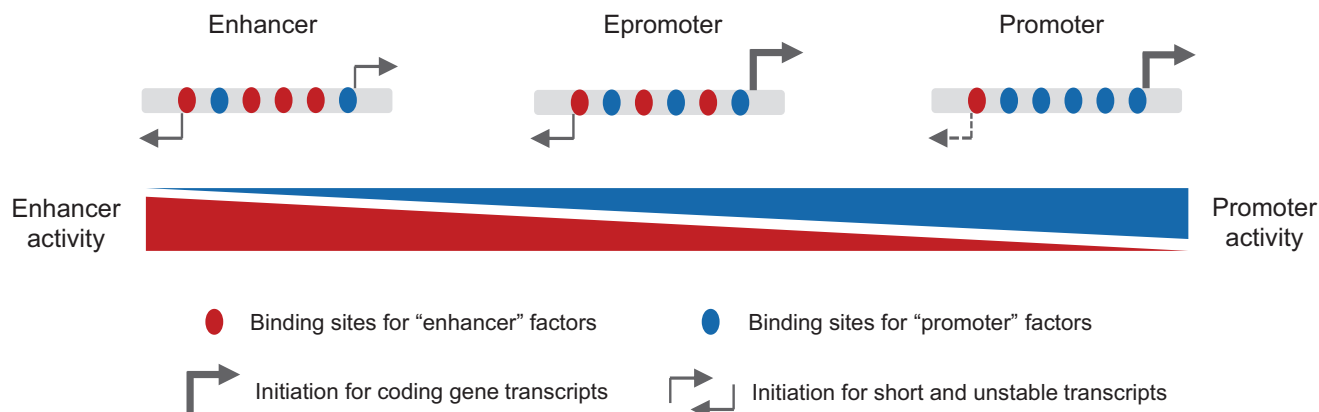


FIGURE 2 A general model defining cis-regulatory function. Regulatory elements are composed of different ranges of binding sites for transcription factors associated with either enhancer (red) or promoter (blue) properties. Enhancers are mostly composed of binding sites associated with enhancer activity and initiate bidirectional transcription of short and unstable transcripts. Promoters are principally composed of binding sites associated with promoter activity and initiate strong unidirectional transcription towards the coding gene. Epromoters share structural features of both promoters and enhancers.

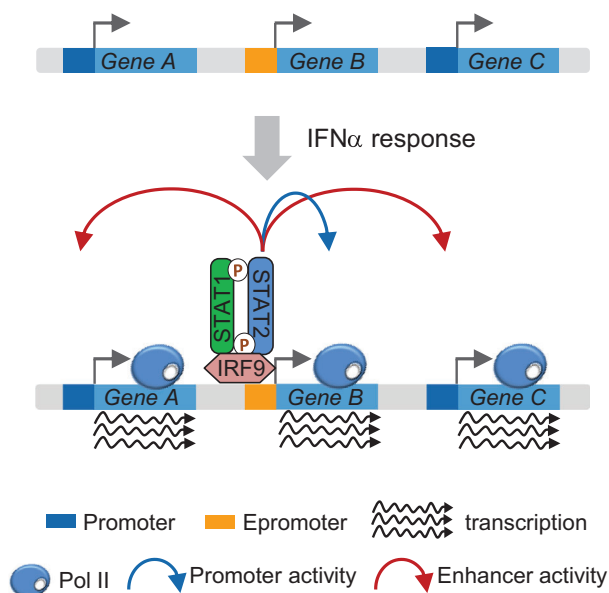


FIGURE 3 Epromoters might function as regulatory hubs for the coordinated induction of gene clusters. During the interferon response, clusters of induced genes are frequently associated with Epromoters. In this context, the Epromoter recruits the key interferon-response TFs STAT1, STAT2 and IRF9, and simultaneously regulates co-induced genes within the same cluster. Legend is as in Figure 1.

Response Elements (ISRE), as compared with typically induced promoters, which, in turn, results in the Epromoter-specific recruitment of STAT1/2 and IRF TFs^[18] (Figure 3). Moreover, inhibition of interferon signaling in HeLa cells drastically reduced the number of active Epromoters without affecting distal enhancers.^[12,18,64] These results suggest that at least a subset of Epromoters plays an essential role in the coordination of rapid gene induction upon cellular response to

intra- and extra-cellular signals which might require a high efficiency of TF recruitment.

We speculate that Epromoters represent a combination of the two types of cis-regulatory elements, thus combining features that are associated with enhancer and promoter activities within an enhancer-promoter continuum of cis-regulatory elements (Figure 2). This intermediated position implies that Epromoters might display a higher density and complexity of TFBS because it has to accommodate the binding of TFs for both enhancer and promoter functions. In this scenario, typical promoters are enriched in binding sites for TF conferring promoter activity and enhancers enriched in binding sites for TF conferring enhancer activity, while Epromoters will be enriched for both types of binding sites leading to a higher density of TFBS. An alternative model will imply that Epromoters are associated with a unique combination of TFBS providing specific enhancer-promoter features. Future works should apply systematic assess the contribution of TFBS and associated transcription factors to enhancer and promoter activity in combination with machine learning models to understand the molecular features that determine the intrinsic promoter and enhancer potentials of cis-regulatory elements, and in particular of Epromoters.^[1,10,65] This, in turn, might help to better predict the impact of mutations or natural variants of Epromoters that might affect either proximal or distal gene regulation.

GENETIC VARIATION AT EPROMOTERS MIGHT HAVE PLEIOTROPIC ROLES

There is growing evidence that a wide range of human diseases is influenced by dysfunctions of cis-regulatory elements caused by genetic, structural, or epigenetic mechanisms.^[66] These processes frequently underpin the susceptibility to common diseases but can be also directly involved in cancer or Mendelian diseases. The advent of genome-wide association studies (GWASs) in the past decade has been one of

the great endeavors in genomic research toward identifying genetic variants associated with candidate genes for common diseases. The majority of these genetic variants are found in non-coding regions, therefore are likely to be involved in regulatory mechanisms controlling gene expression.^[67–69] However, a major challenge in interpreting the impact of genetic mutation or variation in disease is to identify the targets that are impacted by the genomic alteration, which are not necessarily the closest genes and might have confounding features.^[70] Despite this, most studies select the closest gene to the associated GWAS variant to establish possible causal mechanisms, namely when the variant lies in the vicinity of a TSS or within an intronic region. However, this assumption has been shown to be biased in examples like the *FTO*^[71] and *TCF7L2*^[72] loci, in which obesity and type 2 diabetes GWAS data, respectively, were interpreted to implicate the nearest genes, while 3D epigenomics and functional follow-up showed that the disease variants reside in elements that regulate distant genes. In a similar way, it might be envisioned that GWAS variants lying within Epromoters might regulate the expression of distal disease-causing genes.

The discovery of Epromoters thus opens a new paradigm in the study of regulatory variants as a mutation in a promoter could potentially influence the expression of several genes or change the relative ratio of promoter versus enhancer activity. Thus, resulting in a variety of potential changes in the relative expression of neighboring genes (Figure 4A). In addition, it is plausible that the same *cis*-regulatory element displays preferential promoter activity in some tissues while displaying increased enhancer activity in other tissues, depending on the expressed combination of TFs and the epigenetic context.^[8,33,73] For instance, Leung et al. found frequent examples of dynamic epigenetic switches where active promoters in one tissue displayed a histone modification signature of enhancers in other tissues/cell types.^[73] Similarly, Chandra et al. found that a substantial number of promoter-promoter interactions involved transcriptionally inactive genes, suggesting that non-transcribing promoters may function as active enhancers for distal genes.^[33]

The complex regulation by Epromoters might therefore have two predicted consequences. On one hand, there might be a general underestimation of the impact of Epromoter variation in disease because the causal gene might not be the closest one and therefore the link between genotype and phenotype might be missed in many case studies. On the other hand, as Epromoters potentially control several genes at the same time and efficiently recruit key TFs, mutations in these regulatory elements are expected to have a stronger pathological impact, as compared to typical promoters. This might result from the regulation of multiple genes either involved in the same (additive or synergistic effects) or different (pleiotropy) pathways (Figure 4B). Here, pleiotropy refers to a single *cis*-regulatory element affecting more than one trait independently.^[74] Pleiotropy could be due to the perturbation of a single gene playing multiple functions in different tissues^[75,76] or the regulation of multiple genes in the same or different tissues.^[77,78] Several genomic features, such as a higher number of regulated genes and more abundance and diversity of encoded TFBS, are indicative of increasing variant pleiotropy.^[78–80] Notably, these two features are

readily associated with both Epromoters and typical enhancers. Thus, it is fair to hypothesize that genetic alterations affecting Epromoters are likely to be involved in disease, as previously suggested,^[2,18,27] and that Epromoters might have a stronger impact on the regulation of disease-associated genes, as compared with typical promoters. Although the hypothesis is not fully validated yet, several lines of observations support this assumption, as detailed below.

There are several pieces of genetic evidence indicating that promoter variants affecting distal genes are physiologically relevant. One way to connect GWAS-reported genetic variants with effects on gene function is to associate the genetic polymorphisms with eQTLs.^[81] eQTLs with a higher probability to directly impact gene expression variation tend to be found in open chromatin regions, such as promoters and enhancers,^[82] supporting the hypothesis of a possible effect through changes on regulatory mechanisms. Studies of natural genetic variation through eQTLs thus provide important insights into the mechanisms of specific diseases and gene control, and can point to the possible gene regulatory function of specific sequences based on their allelic associations with gene expression.^[66]

Several studies have observed significant enrichment for eQTLs located within gene promoters which are associated with the regulation of distal genes,^[8,17,29,33,83] pointing out to Epromoter-like regulation. A study, integrating predicted allelic variation in TF binding affinity in human lymphoblastic cell lines with their putative target genes inferred from Promoter Capture Hi-C, observed that a large proportion of regulatory variants associated with distal gene expression localized to the promoter regions of other genes, supporting the notion of Epromoters.^[29] Interestingly, some of these variants were co-associated with the expression of both proximal and distal genes, while others were uniquely associated with distal genes. Using a set of Epromoters identified by STARR-seq, we observed that eQTLs lying within Epromoters are more likely to be associated with the expression of a distal gene as compared to other promoters and tend to have stronger effects on distal gene expression.^[8] By analyzing promoter-centered long-range chromatin interactions in the human genome, Jung et al.^[17] and Chandra et al.^[33] found that P-P interactions were significantly enriched in eQTLs where a genetic variant in one of the interacting promoters was associated with the expression of the other interacting promoter. In these three studies, CRISPR-mediated gene editing recapitulated the predicted function of the promoter-associated eQTL variants in the regulation of distal gene expression.^[8,17,33] A similar study found that the majority of genetic variation affecting TF binding at Epromoters in Lymphoblastoid cell lines was associated with the expression of distal genes alone, independently of whether the proximal gene was transcriptionally active.^[29] While it is difficult to ensure that all promoter-distal eQTLs are bona fide distal regulators, we noticed that taking into consideration functional assays for enhancer activity (STARR-seq) allows to significantly enrich for promoter-eQTLs regulating distal genes with a higher probability of perturbing TF binding affinity.^[8] Taken together, these findings support the functional significance of long-range transcriptional regulation by Epromoters and imply that regulatory variants within these elements may have both independent

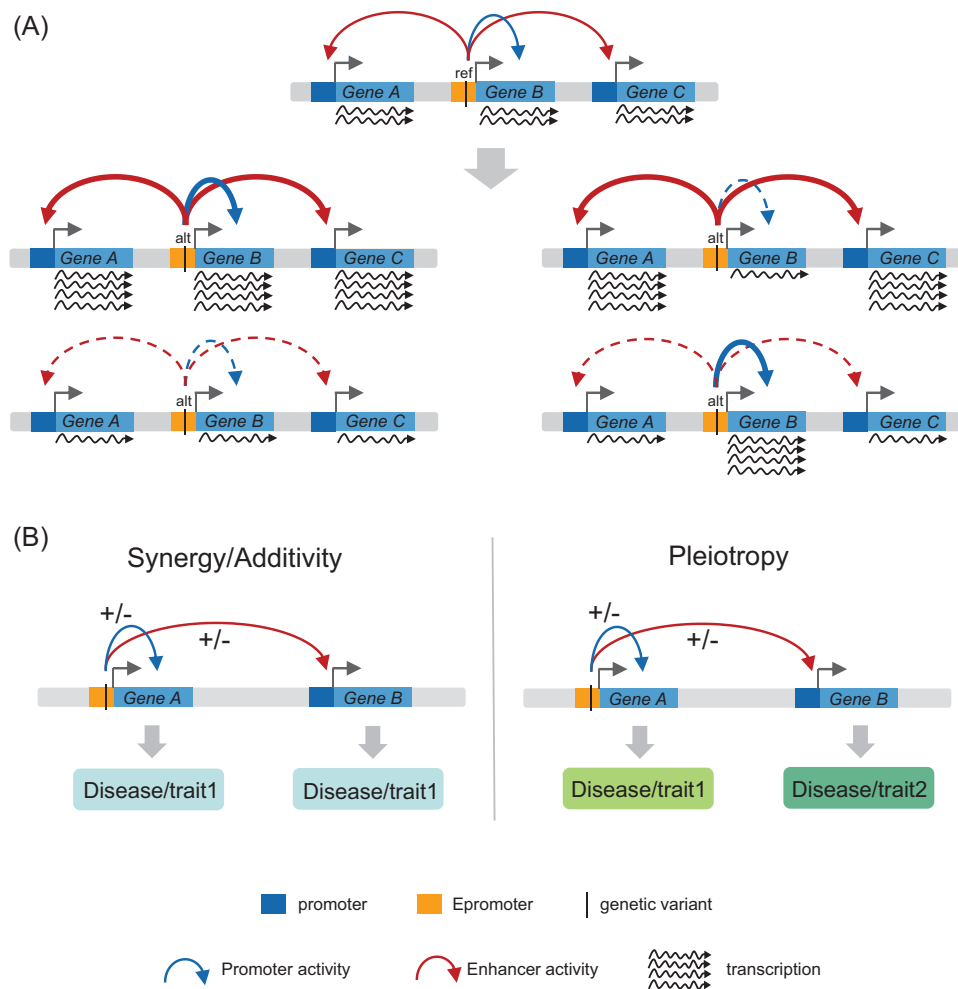


FIGURE 4 Effects of Epromoter variation on gene regulation. (A) Different potential impacts of Epromoter genetic variation on proximal and distal gene expression. In the reference (ref) haplotype, proximal and distal genes transcription are regulated by the Epromoter. In the alternative (alt) haplotypes, the promoter (blue arrows) and enhancer (red arrows) activities could increase (thicker arrows) or decrease (thinner and dashed arrows), resulting in up- or down-regulation of the associated genes. (B) Genetic variants at Epromoters might result in either synergistic/additive (all affected genes are involved in the same disease/trait) or pleiotropic (each affected gene is involved in a different disease/trait) effects.

and shared effects on the expression of their proximal and distal target genes.

Besides the global genomic evidence, several specific examples are pointing toward the relationship between disrupted Epromoters and variants associated with a variety of diseases, including autoimmunity (Crohn's disease, Lupus),^[30,33,34,83,84] cardiovascular diseases,^[19,85,86] diabetes,^[2,35,87] infection diseases,^[88,89] and cancer^[90-93] (Table 1). However, to our best knowledge only in four cases, have the link between the disease-associated variant and the Epromoter function been experimentally validated^[84,88,91,92] (Figure 5). In the first case, the alternative variant of a promoter-overlapping SNP associated with prostate cancer changes the relative affinity for two transcription factors resulting in promoter-enhancer switching and the corresponding increase of the expression of two distal transcripts directly involved in cancer progression (Figure 5A).^[91,92] In the second case, the promoter of the *BAZ2B* gene was identified as Epromoter based on STARR-seq, while CRISPR-mediated deletion resulted in decreased expression of the *MARCHF7* gene located 95 kb

away.^[8] The *BAZ2B* promoter overlaps with an SNP in eQTL with *MARCHF7*^[8] and is associated with hypothyroidism (Ref.[94]; unpublished observation). Haplotype replacement by CRISPR-mediated homology recombination resulted in reduced expression of *MARCHF7*, but not *BAZ2B*,^[8] suggesting it is the distal gene regulation of the identified Epromoter's SNP that is involved in the disease (Figure 5B). In the third case, haplotype-specific chromatin looping implicating genetic variants associated with Systemic Lupus Erythematosus (SLE) revealed that the alternative haplotype laying within the promoter of *BLK* gene decrease the promoter activity while increasing the long-range interaction with the *FAM167A* promoter resulting in the up-regulation of *FAM167A* expression (Figure 5C).^[84] In the last case, a haplotype of five genetic variants associated with severe Malaria and laying within the internal promoter of the *ATP2B4* gene was found to switch the relative promoter and enhancer activity by luciferase reporter assays resulting in the increased expression of the long *ATP2B4* isoform initiated from the upstream promoter (Figure 5D).^[88]

TABLE 1 List of diseases associated variants in Epromoters.

Disease	Epromoter-containing variant ^a	Affected gene(s)	Evidence ^b	Ref.
Prostate cancer	PCAT19-short isoform	PCAT19-long & short isoforms, CEACAM21	P-P interaction, enhancer & promoter reporter assays, CRISPR-Cas9 genome editing, CRISPR interference/activation	[91, 92]
Hypothyroidism	BAZ2B	MARCHF7	P-P interaction, CRISPR-Cas9 genome editing, eQTL	[8]
Systemic lupus erythematosus	BLK	BLK, FAM167A	P-P interaction, CRISPR interference,	[84]
Severe malaria	ATP2B4-short isoform	ATP2B4-long & short isoforms	Enhancer & promoter reporter assays, CRISPR/Cas9 genome editing	[88]
Cardiovascular diseases	Nppb	Nppa	P-P interaction, Mouse transgenic models	[19, 86]
Type 2 diabetes	ARAP1	PDE2A	CapSTARR-seq, promoter reporter assays, eQTL	[2, 87]
Rheumatoid arthritis	CCR6	RNASET2	P-P interaction, eQTL	[33]
Systemic lupus erythematosus	TREH	CXCR5	P-P interaction, eQTL	[34]
Crohn disease	SMAD3	SMAD3, AAGAB	P-P interaction, eQTL	[30, 99]
Coronary artery disease	CDKN2B	IFNA2	P-P interaction	[85]
Multiple cancers	TERT	CLPTM1L	Somatic mutations, correlation with gene expression	[90]
Type 2 diabetes	INS	SYT8	P-P interaction, siRNA	[35]

^aName of the proximal gene associated with the Epromoter.

^bNature of the evidence suggesting an Epromoter-type of regulation.

In addition to small genetic variants, other types of genomic alterations involving enhancer repositioning (a phenomenon also named “enhancer hijacking”) by chromosomal translocations, genomic rearrangements, or insulator disruption, have been described as common molecular mechanisms resulting in disease-related gene deregulation, including overexpression of oncogenes.^[66] Epromoter hijacking could likely impact diseases through related mechanisms. For instance, in T-acute lymphoblastic leukemia, a large intergenic deletion replaces the *TAL1* locus into the vicinity of the *CMPK1* promoter, which displays features of an active enhancer, resulting in *TAL1* oncogenic expression.^[93] More generally, a study, using randomly integrated reporter constructs, found that chromosomal contacts with endogenous promoters

of housekeeping genes is required for the expression of the reporter gene,^[95] supporting the idea whereby genomic repositioning due to structural variations might result in gene expression deregulation by Epromoter hijacking.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Current results point to an important role of Epromoters in the regulatory landscape. These findings also open up the possibility that disease-associated variants or developmental traits lying within a

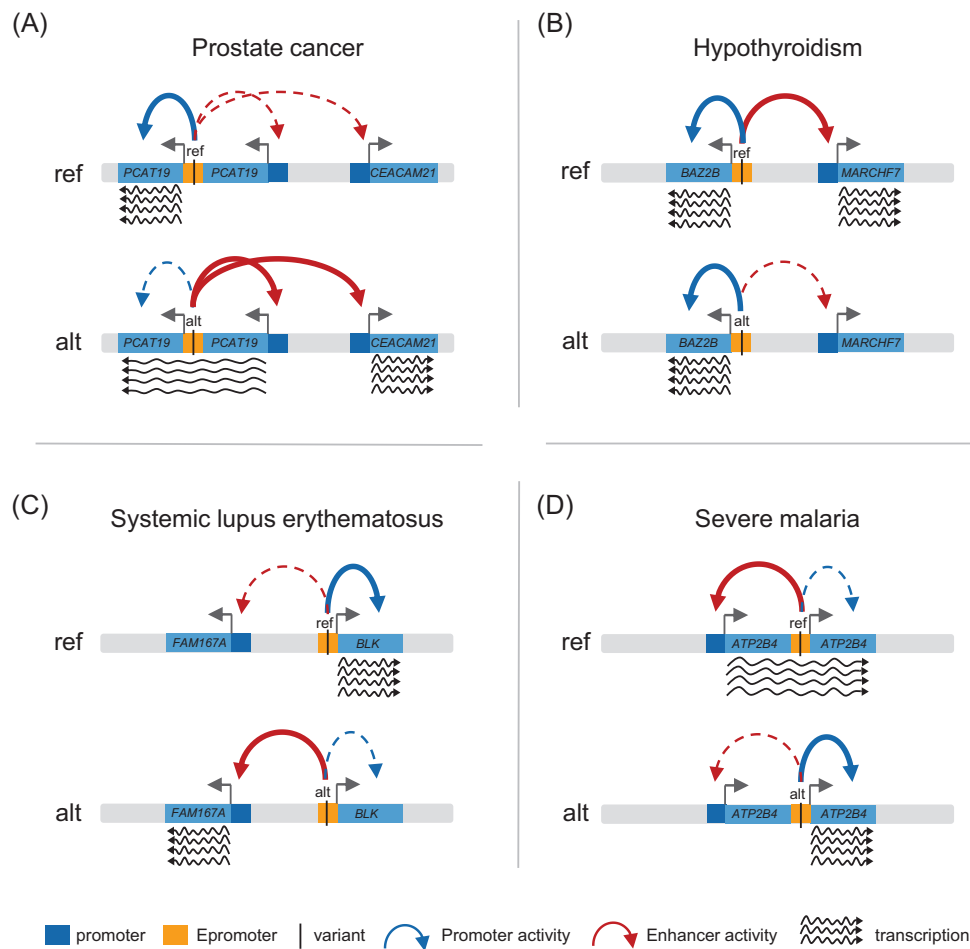


FIGURE 5 Examples of Epromoter variation associated with diseases. (A) The variant rs11672691 is associated with prostate cancer and locate within the internal *PCAT19* promoter. The alternative variant switches the relative promoter and enhancer activity resulting in up-regulation of the most upstream *PCAT19* promoter and the distal gene *CEACAM21*. (B) The variant rs1046496 is associated with hypothyroidism and locate within the *BAZ2B* promoter. The alternative variant decreases the transcription of the *MARCHF7* gene. (C) The variant rs922483 is associated with systemic lupus erythematosus and locate within the *BLK* promoter. The alternative variant decreases the transcription of the *BLK* gene while increasing the expression of the *FAM167A* gene. (D) A haplotype of five variants containing the lead variant rs10900585 is associated with severe malaria and is located within the internal promoter of *ATP2B4*. The alternative variant switches the relative promoter and enhancer activity resulting in up-regulation of the most upstream *ATP2B4* promoter.

subset of promoters directly impact distal gene expression. Indeed, the recent observations support the hypothesis whereby Epromoters have a pleiotropic effect on diseases by perturbing the expression of several genes at the same time. Future works, including additional experimental settings where the Epromoter function is assessed in their endogenous loci, should tell us how commonly promoters are used as distal enhancers and better describe the physiological contexts where they are at play. Equally important will be to assess the physiological relevance of genes regulated by Epromoters. Are both proximal and distal gene deregulation directly involved in diseases? In particular, given the preponderant role of Epromoters in the regulation of interferon-response genes, it is expected that they might have an important impact on the etiology of inflammatory diseases.

Although several works have provided clear examples of natural genetic variation within promoters affecting the expression of a distal gene or an isoform regulated by a distal promoter, more

systematic studies are required to ascertain whether Epromoters indeed play a pleiotropic role in disease. While sequence-based models perform well in predicting how the impact of genetic variants in promoters affects local gene expression, they still perform very low to predict distal gene effects.^[96] Therefore, we suggest that promoter-associated variants should systematically be tested for their proximal and distal effects. One possibility might be to simultaneously test the same DNA fragment in high-throughput reporter assays designed to assess the enhancer and the promoter activity in parallel.^[13,50] This type of approach will help to elucidate whether enhancer and promoter activity of Epromoters are generally correlated amongst different cell types/tissues or whether they exhibit tissue-specific context. Similarly, it will help to assess whether genetic variants generally influence the global regulatory activity of Epromoters or rather affect their relative function to primarily work as a promoter or as an enhancer.

A central goal of biology is to decipher the cis-regulatory code that governs when and how much each gene is transcribed in a given genome and cellular state.^[97] Two major advancements provide a paradigm shift in our capacity to integrate mechanistically informed, quantitative models of transcriptional regulation toward cracking the cis-regulatory code. On one side, the development of high-throughput reporter assays allows systematic and quantitative measurements of cis-regulatory activity. On the other side, the ability of recent deep learning models to learn the most relevant features from genomic data and to interpret and extract the features (e.g., DNA sequence) that underlie the predictions. These interpretable rules can then be used to decode the regulatory “syntax” or “grammar,” providing detailed information about the arrangement of TFBS, including their number, order, orientation and spacing. Several examples, disentangling cis-regulatory functions by combining high-throughput reporter assays with deep learning methods, have provided remarkable results.^[62,65,98] Similar approaches aiming to dissect the intrinsic DNA features required for promoter and enhancer activities should be applied to the study of Epromoters. As Epromoters share features of both enhancer and promoter functions, a clear disentangling of both activities should lead to a better definition of the molecular bases governing gene regulation. In turn, understanding the genetic features driving proximal and distal activities will allow a better prediction of the impact of genetic variants with a pleiotropic effect on disease.

ACKNOWLEDGMENTS

Work in the laboratory of Salvatore Spicuglia was supported by recurrent funding from Institut National de la Santé et de la Recherche Médicale (INSERM) and Aix-Marseille University, and by specific grants from Ligue contre le Cancer (Equipe Labellisée Ligue 2023), ANR (ANR-18-CE12-0019 and ANR-17-CE12-0035) and Bettencourt Schueller Foundation (Prix coup d'élan pour la recherche française). Juliette Malfait is supported by a fellowship from the French Ministry of Education and Aix-Marseille University. Jing Wan is a fellow of the EU-funded Innovative Training Network “Molecular Basis of Human Enhanceropathies” and received funding from the Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement no. 860002.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Salvatore Spicuglia  <https://orcid.org/0000-0002-8101-7108>

REFERENCES

- Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 21, 71–87.
- Medina-Rivera, A., Santiago-Algarra, D., Puthier, D., & Spicuglia, S. (2018). Widespread enhancer activity from core promoters. *Trends in Biochemical Sciences*, 43, 452–468.
- Schaffner, W. (2015). Enhancers, enhancers – from their discovery to today's universe of transcription enhancers. *Biological Chemistry*, 396, 311–327.
- Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27, 299–308.
- Gasparini, M., Tome, J. M., & Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics*, 21, 292–310.
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339, 1074–1077.
- Barakat, T. S., Halbritter, F., Zhang, M., Rendeiro, A. F., Perenthaler, E., Bock, C., & Chambers, I. (2018). Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell*, 23, 276–288.e8.
- Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., Alomairi, J., Martin, D., Torres, M., Fernandez, N., Soler, E., Van Helden, J., Puthier, D., & Spicuglia, S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics*, 49, 1073–1081.
- Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T. S., & Kellis, M. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nature Biotechnology*, 34, 1180–1190.
- Glaser, L. V., Steiger, M., Fuchs, A., Chung, H-R., Vingron, M., & Meijnsing, S. H. (2021). Assessing genome-wide dynamic changes in enhancer activity during early mESC differentiation by FAIRE-STARR-seq. *Nucleic Acids Research*, 49(21), 12178–12195.
- Liu, Y., Yu, S., Dhiman, V. K., Brunetti, T., Eckart, H., & White, K. P. (2017). Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biology*, 18, 219.
- Muerdter, F., Boryń, Ł. M., Woodfin, A. R., Neumayr, C., Rath, M., Zabidi, M. A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R. R., Schernhuber, K., Arnold, C. D., & Stark, A. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. *Nature Methods*, 15, 141–149.
- Nguyen, T. A., Jones, R. D., Snavely, A. R., Pfenning, A. R., Kirchner, R., Hemberg, M., & Gray, J. M. (2016). High-throughput functional comparison of promoter and enhancer activities. *Genome Research*, 26, 1023–1033.
- Wang, X., He, L., Goggin, S. M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., & Kellis, M. (2018). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nature Communications*, 9, 5380.
- Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518, 556–559.
- Engreitt, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539, 452–455.
- Jung, I., Schmitt, A., Diaoy, Y., Lee, A. J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliah, E., Barr, C. L., Li, B., Kuan, S., Kim, D., & Ren, B. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*, 51, 1442–1449.
- Santiago-Algarra, D., Souaid, C., Singh, H., Dao, L. T. M., Hussain, S., Medina-Rivera, A., Ramirez-Navarro, L., Castro-Mondragon, J. A., Sadouni, N., Charbonnier, G., & Spicuglia, S. (2021). Epromoters

- function as a hub to recruit key transcription factors required for the inflammatory response. *Nature Communications*, *12*, 6660.
19. Sergeeva, I. A., Hooijkaas, I. B., Ruijter, J. M., Van Der Made, I., De Groot, N. E., Van De Werken, H. J. G., Creemers, E. E., & Christoffels, V. M. (2016). Identification of a regulatory domain controlling the *Nppa-Nppb* gene cluster during heart development and stress. *Development*, *143*(12), 2135–2146.
 20. Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, *51*, 1664–1669.
 21. Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., Syed, T., Emons, B. J. M., Gifford, D. K., & Sherwood, R. I. (2016). High-throughput mapping of regulatory DNA. *Nature Biotechnology*, *34*, 167–174.
 22. Gasperini, M., Hill, A. J., Mcfaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., Jackson, D., Leith, A., Schreiber, J., Noble, W. S., Trapnell, C., Ahituv, N., & Shendure, J. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*, *176*, 377–390.e19.e19.
 23. Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K. C., Huang, H., Liu, T., Marina, R. J., Jung, I., Shen, Y., Guan, K. L., & Ren, B. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nature Methods*, *14*, 629–635.
 24. Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. A., Sloane-Stanley, J. A., MCGowan, S. J., De Gobbi, M., Hosseini, M., Vernimmen, D., Brown, J. M., Gray, N. E., Collavin, L., Gibbons, R. J., Flint, J., Taylor, S., Buckle, V. J., Milne, T. A., Wood, W. G., & Higgs, D. R. (2012). Intragenic enhancers act as alternative promoters. *Molecular Cell*, *45*, 447–458.
 25. Paralkar, V. R., Taborda, C. C., Huang, P., Yao, Y. u., Kossenkov, A. V., Prasad, R., Luan, J., Davies, J. O. J., Hughes, J. R., Hardison, R. C., Blobel, G. A., & Weiss, M. J. (2016). Unlinking an lncRNA from its associated cis element. *Molecular Cell*, *62*, 104–110.
 26. Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., & Flicek, P. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*, *167*, 1369–1384.e19.e19.
 27. Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J., Sim, H. S., Peh, S. Q., Mulawadi, F. H., Ong, C. T., Orlov, Y. L., Hong, S., Zhang, Z., Landt, S., Raha, D., & Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, *148*, 84–98.
 28. Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., Leproust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*, 598–606.
 29. Mitchelmore, J., Grinberg, N. F., Wallace, C., & Spivakov, M. (2020). Functional effects of variation in transcription factor binding highlight long-range gene regulation by e-promoters. *Nucleic Acids Research*, *48*, 2866–2879.
 30. Mumbach, M. R., Satpathy, A. T., Boyle, E. A., Dai, C., Gowen, B. G., Cho, S. W., Nguyen, M. L., Rubin, A. J., Granja, J. M., Kazane, K. R., Wei, Y., Nguyen, T., Greenside, P. G., Corces, M. R., Tycck, J., Simeonov, D. R., Suliman, N., Li, R., Xu, J., & Chang, H. Y. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature Genetics*, *49*, 1602–1612.
 31. Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S. W., Dimitrova, E., Dimond, A., Edelman, L. B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., & Fraser, P. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research*, *25*, 582–597.
 32. Wen, J., Lagler, T. M., Sun, Q., Yang, Y., Chen, J., Harigaya, Y., Sankaran, V. G., Hu, M., Reiner, A. P., Raffield, L. M., & Li, Y. (2022). Super interactive promoters provide insight into cell type-specific regulatory networks in blood lineage cell types. *Plos Genetics*, *18*, e1009984.
 33. Chandra, V., Bhattacharyya, S., Schmiel, B. J., Madrigal, A., Gonzalez-Colin, C., Fotsing, S., Crinklaw, A., Seumo, G., Mohammadi, P., Kronenberg, M., Peters, B., Ay, F., & Vijayanand, P. (2021). Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nature Genetics*, *53*, 110–119.
 34. Su, C., Johnson, M. E., Torres, A., Thomas, R. M., Manduchi, E., Sharma, P., Mehra, P., Le Coz, C., Leonard, M. E., Lu, S., Hodge, K. M., Chesi, A., Pippin, J., Romberg, N., Grant, S. F. A., & Wells, A. D. (2020). Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nature Communications*, *11*, 3294.
 35. Xu, Z., Wei, G., Chepelev, I., Zhao, K., & Felsenfeld, G. (2011). Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. *Nature Structural & Molecular Biology*, *18*, 372–378.
 36. Zhu, I., Song, W., Ovcharenko, I., & Landsman, D. (2021). A model of active transcription hubs that unifies the roles of active promoters and enhancers. *Nucleic Acids Research*, *49*, 4493–4505.
 37. Feuerborn, A., & Cook, P. R. (2015). Why the activity of a gene depends on its neighbors. *Trends in Genetics*, *31*, 483–490.
 38. Ueyehara, C. M., & Apostolou, E. (2023). 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Reports*, *42*(4), 112068.
 39. Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M., & Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. *Nature Communications*, *9*, 4066.
 40. Wu, X., & Sharp, P. A. (2013). Divergent transcription: A driving force for new gene origination? *Cell*, *155*, 990–996.
 41. Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*, 1311–1320.
 42. Kim, T. K., & Shiekhatar, R. (2015). Architectural and functional commonalities between enhancers and promoters. *Cell*, *162*, 948–959.
 43. Tippens, N. D., Vihervaara, A., & Lis, J. T. (2018). Enhancer transcription: What, where, when, and why? *Genes & Development*, *32*, 1–3.
 44. De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C. L., & Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *Plos Biology*, *8*, e1000384.
 45. Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., Lavender, C. A., Fargo, D. C., & Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes & Development*, *32*, 26–41.
 46. Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*, 182–187.
 47. Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., Spicuglia, S., De La Chapelle, A. L., Heidemann, M., Hintermair, C., Eick, D., Gut, I., Ferrier, P., & Andrau, J.-C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology*, *18*, 956–963.
 48. Consortium, T. F., Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., & Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*, 455–461.

49. He, H. H., Meyer, C. A., Shin, H., Bailey, S. T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., Mieczkowski, P., Lieb, J. D., Zhao, K., Brown, M., & Liu, X. S. (2010). Nucleosome dynamics define transcriptional enhancers. *Nature Genetics*, *42*, 343–347.
50. Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I. E., Males, M., Viales, R. R., & Furlong, E. E. M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes & Development*, *32*, 42–57.
51. Rennie, S., Dalby, M., Lloret-Llinares, M., Bakoulis, S., Vaagensø, C. D., Jensen, T. H., & Andersson, R. (2018). Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities. *Nucleic Acids Research*, *46*(11), 5455–5469.
52. Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faúlí, M., Bleckwehl, T., Asenjo, H. G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., Van Ijcken, W. F. J., Landeira, D., & Rada-Iglesias, A. (2021). Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. *Nature Genetics*, *53*, 1036–1049.
53. Creighton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21931–21936.
54. Heintzman, N. D., & Ren, B. (2009). Finding distal regulatory elements in the human genome. *Current Opinion in Genetics & Development*, *19*, 541–549.
55. Rada-Iglesias, A., Bajpai, R., Swigut, T., Bruggmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*, 279–283.
56. Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J. C., Ferrier, P., & Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers: Epigenetic signature of active enhancers. *The EMBO Journal*, *30*, 4198–4210.
57. Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T. M., Fernandez, N., Ballester, B., Andrau, J. C., & Spicuglia, S. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nature Communications*, *6*, 69 05.
58. Seipel, K., Georgiev, O., & Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *Embo Journal*, *11*, 4961–4968.
59. Colbran, L. L., Chen, L., & Capra, J. A. (2019). Sequence characteristics distinguish transcribed enhancers from promoters and predict their breadth of activity. *Genetics*, *211*, 1205–1217.
60. Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T. S., & Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Research*, *23*, 800–811.
61. Kwasniewski, J. C., Fiore, C., Chaudhari, H. G., & Cohen, B. A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Research*, *24*, 1595–1602.
62. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C. O., Cramer, P., Kivioja, T., & Taipale, J. (2022). Sequence determinants of human gene regulatory elements. *Nature Genetics*, *54*, 283–294.
63. Scruggs, B. S., Gilchrist, D. A., Nechaev, S., Muse, G. W., Burkholder, A., Fargo, D. C., & Adelman, K. (2015). Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Molecular Cell*, *58*, 1101–1112.
64. Dao, L. T. M., & Spicuglia, S. (2018). Transcriptional regulation by promoters with enhancer function. *Transcription*, *9*, 307–314.
65. De Almeida, B. P., Reiter, F., Pagani, M., & Stark, A. (2022). DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nature Genetics*, *54*, 613–624.
66. Zaugg, J. B., Sahlén, P., Andersson, R., Alberich-Jorda, M., De Laat, W., Deplancke, B., Ferrer, J., Mandrup, S., Natoli, G., Plewczynski, D., Rada-Iglesias, A., & Spicuglia, S. (2022). Current challenges in understanding the role of enhancers in disease. *Nature Structural & Molecular Biology*, *29*(12), 1148–1158.
67. Deplancke, B., Alpern, D., & Gardeux, V. (2016). The genetics of transcription factor DNA binding variation. *Cell*, *166*, 538–554.
68. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorf, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Research*, *45*, D896–D901.
69. Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., ... Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, *337*, 1190–1195.
70. Brandes, N., Weissbrod, O., & Linnal, M. (2022). Open problems in human trait genetics. *Genome Biology*, *23*, 131.
71. Smemo, S., Tena, J. J., Kim, K.-H., Gamazon, E. R., Sakabe, N. J., Gómez-Marín, C., Aneas, I., Credidio, F. L., Sobreira, D. R., Wasserman, N. F., Lee, J. H., Puviondran, V., Tam, D., Shen, M., Son, J. E., Vakili, N. A., Sung, H.-K., Naranjo, S., Acemel, R. D., ... Nóbrega, M. A. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, *507*, 371–375.
72. Xia, Q., Chesi, A., Manduchi, E., Johnston, B. T., Lu, S., Leonard, M. E., Parlin, U. W., Rappaport, E. F., Huang, P., Wells, A. D., Blobel, G. A., Johnson, M. E., & Grant, S. F. A. (2016). The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5. *Diabetologia*, *59*, 2360–2368.
73. Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A. Y., Yen, C.-A., Lin, S., Lin, Y., Qiu, Y., Xie, W., Yue, F., Hariharan, M., Ray, P., Kuan, S., Edsall, L., Yang, H., Chi, N. C., Zhang, M. Q., ... Ren, B. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*, *518*, 350–354.
74. Cano-Gamez, E., & Trynka, G. (2020). From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Frontiers in Genetics*, *11*, 424.
75. Gupta, R. M., Hadaya, J., Trehan, A., Zekavat, S. M., Roselli, C., Klarin, D., Emdin, C. A., Hilvering, C. R. E., Bianchi, V., Mueller, C., Khera, A. V., Ryan, R. J. H., Engreitz, J. M., Issner, R., Shores, N., Epstein, C. B., De Laat, W., Brown, J. D., Schnabel, R. B., ... Kathiresan, S. (2017). A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell*, *170*, 522–533.e15.e15.
76. Sinnott-Armstrong, N., Sousa, I. S., Laber, S., Rendina-Ruedy, E., Nitter Dankel, S. E., Ferreira, T., Mellgren, G., Karasik, D., Rivas, M., Pritchard, J., Guntur, A. R., Cox, R. D., Lindgren, C. M., Hauner, H., Sallari, R., Rosen, C. J., Hsu, Y.-H., Lander, E. S., Kiel, D. P., & Clausnitzer, M. (2021). A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell Metabolism*, *33*, 615–628.e13.e13.
77. Sobreira, D. R., & Nóbrega, M. A. (2021). Regulatory landscapes of *Nppa* and *Nppb*. *Circulation Research*, *128*, 130–132.
78. Joslin, A. C., Sobreira, D. R., Hansen, G. T., Sakabe, N. J., Aneas, I., Montefiori, L. E., Farris, K. M., Gu, J., Lehman, D. M., Ober, C., He, X., & Nóbrega, M. A. (2021). A functional genomics pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci. *Nature Communications*, *12*, 5253.
79. Ribeiro, D. M., Rubinacci, S., Ramisch, A., Hofmeister, R. J., Dermitzakis, E. T., & Delaneau, O. (2021). The molecular basis, genetic control and

- pleiotropic effects of local gene co-expression. *Nature Communications*, 12, 4842.
80. Singh, D., & Yi, S. V. (2021). Enhancer pleiotropy, gene expression, and the architecture of human enhancer–gene interactions. *Molecular Biology and Evolution*, 38, 3898–3909.
 81. Flynn, E. D., & Lappalainen, T. (2022). Functional characterization of genetic variant effects on expression. *Annual Review of Biomedical Data Science*, 5, 119–139.
 82. GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature*, 550, 204–213.
 83. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Xu, M., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., Rhie, S. K., Manakongtreecheep, K., Zhou, H., & PsychENCODE Consortium. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, 362, eaat8464.
 84. Saint Just Ribeiro, M., Tripathi, P., Namjou, B., Harley, J. B., & Chepelev, I. (2022). Haplotype-specific chromatin looping reveals genetic interactions of regulatory regions modulating gene expression in 8p23.1. *Frontiers in Genetics*, 13, 1008582.
 85. Li, W., Wong, W. H., & Jiang, R. (2019). DeepTACT: Predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Research*, 47, e60–e60.
 86. Man, J., Barnett, P., & Christoffels, V. M. (2018). Structure and function of the Nppa–Nppb cluster locus during heart development and disease. *Cellular and Molecular Life Sciences*, 75, 1435–1444.
 87. Kulzer, J. R., Stitzel, M. L., Morken, M. A., Huyghe, J. R., Fuchsberger, C., Kuusisto, J., Laakso, M., Boehnke, M., Collins, F. S., & Mohlke, K. L. (2014). A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *The American Journal of Human Genetics*, 94, 186–197.
 88. Nisar, S., Torres, M., Thiam, A., Pouvelle, B., Rosier, F., Gallardo, F., Ka, O., Mbengue, B., Diallo, R. N., Brosseau, L., Spicuglia, S., Dieye, A., Marquet, S., & Rihet, P. (2022). Identification of ATP2B4 regulatory element containing functional genetic variants associated with severe malaria. *International Journal of Molecular Sciences*, 23, 4849.
 89. Stikker, B. S., Stik, G., Van Ouwkerk, A. F., Trap, L., Spicuglia, S., Hendriks, R. W., & Stadhouders, R. (2022). Severe COVID-19-associated variants linked to chemokine receptor gene control in monocytes and macrophages. *Genome Biology*, 23, 96.
 90. Fredriksson, N. J., Ny, L., Nilsson, J. A., & Larsson, E. (2014). Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature Genetics*, 46, 1258–1263.
 91. Gao, P., Xia, J.-H., Sipeky, C., Dong, X.-M., Zhang, Q., Yang, Y., Zhang, P., Cruz, S. P., Zhang, K., Zhu, J., Lee, H.-M., Suleman, S., Giannareas, N., Liu, S., Tammela, T. L. J., Auvinen, A., Wang, X., Huang, Q., Wang, L., ... Wei, G.-H. (2018). Biology and clinical implications of the 19q13 aggressive prostate cancer susceptibility locus. *Cell*, 174, 576–589.e18.e18.
 92. Hua, J. T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., Lu, J., Zhou, S., Wang, M., Li, H., Larson, N. B., McDonnell, S. K., Patel, P. S., Liang, Y., Yao, C. Q., Van Der Kwast, T., Lupien, M., Feng, F. Y., Zoubeidi, A., ... He, H. H. (2018). Risk SNP-mediated promoter-enhancer switching drives prostate cancer through lncRNA PCAT19. *Cell*, 174, 564–575.e18.e18.
 93. Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A., Reddy, J., Borges-Rivera, D., Lee, T. I., Jaenisch, R., Porteus, M. H., Dekker, J., & Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351, 1454–1458.
 94. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., Macarthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., ... Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Research*, 51, D977–D985.
 95. Corrales, M., Rosado, A., Cortini, R., Van Arensbergen, J., Van Steensel, B., & Filion, G. J. (2017). Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome Research*, 27, 1153–1161.
 96. Karollus, A., Mauermeier, T., & Gagneur, J. (2023). Current sequence-based models capture gene expression determinants in promoters but mostly ignore distal enhancers. *Genome Biology*, 24, 56.
 97. Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular Cell*, 83, 373–392.
 98. Movva, R., Greenside, P., Marinov, G. K., Nair, S., Shrikumar, A., & Kundaje, A. (2019). Deciphering regulatory DNA sequences and non-coding genetic variants using neural network models of massively parallel reporter assays. *PLoS ONE*, 14, e0218073.
 99. Wang, Y., He, H., Liyanarachchi, S., Genutis, L. K., Li, W., Yu, L., Phay, J. E., Shen, R., Brock, P., & De La Chapelle, A. (2018). The role of SMAD3 in the genetic predisposition to papillary thyroid carcinoma. *Genetics in Medicine*, 20, 927–935.

How to cite this article: Malfait, J., Wan, J., & Spicuglia, S. (2023). Epromoters are new players in the regulatory landscape with potential pleiotropic roles. *BioEssays*, e2300012. <https://doi.org/10.1002/bies.202300012>