

Determinants of enhancer and promoter activities of regulatory elements

Robin Andersson¹* and Albin Sandelin^{1,2}*

Abstract | The proper activities of enhancers and gene promoters are essential for coordinated transcription within a cell. Although diverse methodologies have been developed to identify enhancers and promoters, most have tacitly assumed that these elements are distinct. However, studies have unexpectedly shown that regulatory elements may have both enhancer and promoter functions. Here we review these results, focusing on the factors that determine the promoter and/or enhancer activity of regulatory elements. We discuss emerging models that define regulatory elements by accessible DNA and their non-mutually-exclusive abilities to drive transcription initiation (promoter activity) and/or to enhance transcription at other such regions (enhancer activity).

RNA polymerase II (RNAPII). An enzyme that catalyses the transcription of DNA to RNA, including mRNAs and many long non-coding RNAs.

Transcription start site (TSS). The first transcribed genomic nucleotide of a transcript.

General transcription factors (GTFs). Proteins that, together with RNA polymerase II, make up the pre-initiation complex.

Pre-initiation complex
A polypeptide complex consisting of RNA polymerase II and general transcription factors. This forms around the transcription start site and primes RNA polymerase II for transcription.

¹Department of Biology, University of Copenhagen, Copenhagen, Denmark.

²Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, Denmark.

*e-mail: robin@binf.ku.dk,

albin@binf.ku.dk

<https://doi.org/10.1038/s41576-019-0173-8>

Transcriptional regulatory elements in complex genomes are key players in the dynamic usage of the genome during development, cell and tissue homeostasis, responses to external stimuli, and disease. They overlap a sizeable fraction of trait-associated genetic variants in pathologically relevant cell types^{1–3}, and their alterations have been implicated in several congenital diseases and cancers^{4–6}. Understanding the functions of regulatory elements and the determinants thereof is therefore a major challenge in the field.

Classical definitions of activating regulatory elements are focused on two classes: promoters and enhancers, where the first category defines where transcription is initiated, and the other, elements that amplify such transcription initiation^{7–10}. In hindsight, these historical definitions are vague, based on arbitrarily sized regions, and dichotomous. Powered by diverse genomics techniques, the field has made substantial progress in identifying the promoters and enhancers driving important biological processes. While these techniques have made it possible to identify regulatory elements genome-wide, they have also led to the realization that enhancers and promoters share several properties and functions. For example, their chromatin and sequence architectures are remarkably similar^{11–16}, there are several examples of promoters with enhancer activity^{17–20}, and active enhancers are able to drive local transcription initiation at their boundaries by themselves^{3,21,22}, thereby working as promoters. Therefore, the distinction between enhancers and promoters is becoming increasingly unclear and motivates new models that allow regulatory elements to have varying degrees of promoter and enhancer potential.

To build such models, it is necessary to decouple the effects of a regulatory element of interest from those of other regulatory elements and the larger chromatin environment. Several studies have presented methods to measure the promoter or enhancer potential of regulatory elements on a massive scale. From these data, features are emerging that may explain why a regulatory element has promoter and/or enhancer potential.

Here we review recent results on the characteristics of promoters and enhancers and the determinants of their regulatory functions, with a focus on vertebrates. We contrast the new findings with old definitions, suggest an updated model of regulatory elements based on DNA accessibility and enhancer/promoter potential, and outline open questions and challenges in order to further our understanding of transcriptional regulatory elements.

Transcriptional regulatory elements

The regulation of RNA polymerase II (RNAPII) activity has been the subject of decades of research. At the centre of this process lies the selection of the RNAPII transcription start site (TSS), defined as the first transcribed genomic nucleotide of a transcript. The process of TSS selection is influenced by the presence and DNA-binding activities of general transcription factors (GTFs), reviewed in REFS^{23,24}, which together with RNAPII and other factors form the RNAPII pre-initiation complex. The binding of such GTFs to specific DNA sequence elements is typically contained within a size-constrained region around the TSS, called the ‘core promoter’, typically defined as the ±50 base pair (bp) region around the TSS.

TATA box

A T/A-rich sequence that lies upstream (typically 24–30 bp) of transcription start sites, with a role in positioning the pre-initiation complex.

INR

The initiator sequence; a sequence pattern often overlapping transcription start sites.

Transcription factors

(TFs). Sequence-specific DNA-binding proteins with a role in regulating transcription.

The role of the core promoter is thought to be to determine the precise location of the TSS and the direction of transcription, by means of GTFs and RNAPII binding to specific core promoter sequence elements, which are located at specific distances from the TSS (reviewed in REFS^{10,25}).

The most well-known core promoter elements are the TATA box and the initiator (INR) element. The TATA box was initially believed to be a universal element roughly 24–30 bp upstream from each TSS, recognized by the TATA-binding protein (TBP), which is a central part of the pre-initiation complex²⁶. This was also true for the pyrimidine–purine INR element located at the TSS, but characterization of many TSSs using genome-wide methods has shown that core promoter architectures are complex and diverse (reviewed in REFS^{10,25}). For instance, only a small fraction of mammalian core promoters have a clear TATA box²⁷. This may be due to GTFs being complexes of multiple, interchangeable proteins with various sequence preferences²⁸; for example, the GTF TFIID consists of TBP and TBP-associated factors (TAFs)²⁹, allowing for flexible and degenerate core promoter recognition.

While core promoters contain sufficient information for the transcriptional machinery to select TSSs, the rate of RNAPII initiation can be modulated by the integration of other signals. DNA-binding transcription factors (TFs)^{30,31} can directly or indirectly, via the recruitment of co-activators (for example, CBP-P300 (REF.³²), Mediator³³ or the SAGA complex³⁴), influence core promoter RNAPII recruitment, initiation or elongation. Their input may come from either local or distal binding of TFs, with respect to core promoter genomic

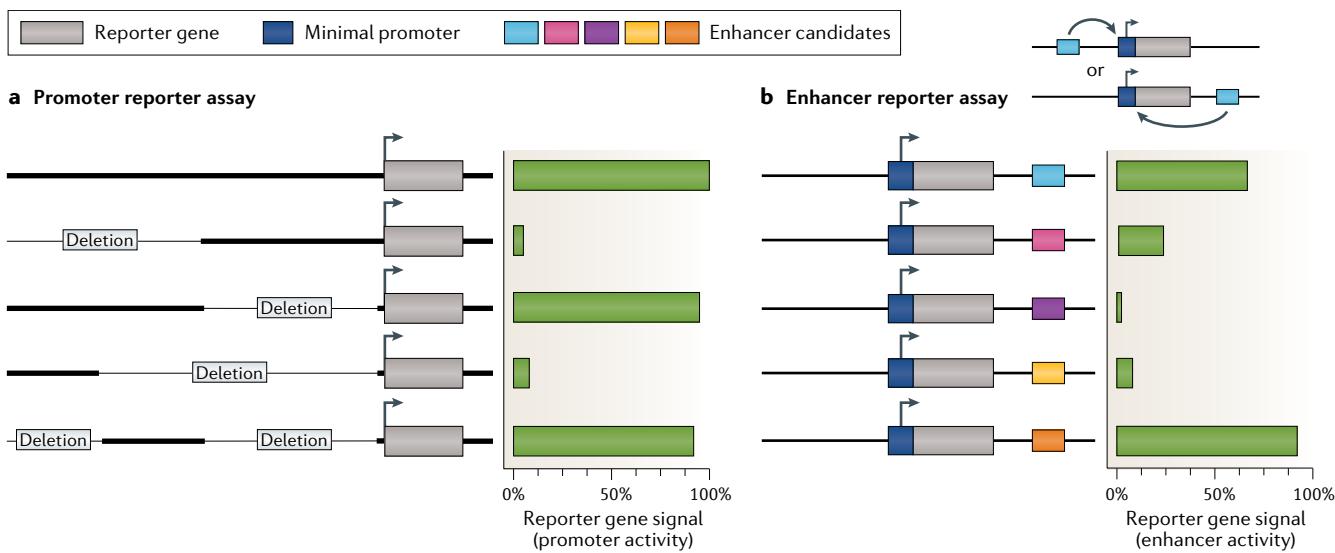
proximity, and often comes about through cooperation between multiple TFs in so-called *cis*-regulatory modules (reviewed in REFS^{35,36}). Because some TFs bind proximally to, but not within, the core promoter region, a larger, arbitrarily sized region around the TSS encompassing the core promoter and this ‘proximal promoter’ region is often referred to as the ‘promoter’, even though only part of this region may be important for the regulation of transcription initiation. Finding out which regions are important within such a promoter region requires exhaustive ‘promoter bashing’, a technique in which selected sub-regions upstream of a TSS are fused with a reporter gene (BOX 1). Therefore, if the term ‘promoter’ refers to a genomic region, it is almost always arbitrarily defined on the basis of the locations of TSSs.

Transcriptional regulation may also be achieved by integrating the input from more distal TF binding events at genomic regions termed ‘enhancers’, some of which can be up to 1 Mb away from the core promoter³⁷. Regulatory input from distal locations is enabled by a favourable folding of the genome that places regulatory elements close in three-dimensional space^{38,39}. Enhancers were first identified in simian virus 70, followed by their identification in mammalian genomes, and they were found to be able to increase reporter gene expression regardless of their orientation or distance from the core promoter^{40–44} (for a historical review of enhancers, see REF.⁴⁵). Most early enhancer work relied on reporter gene assays, in which a candidate enhancer sequence was placed upstream or downstream of a minimal core promoter that was itself upstream of a reporter gene (in many ways similar to the ‘promoter bashing’ discussed above; BOX 1). This technique remains a workhorse in

Box 1 | Overview of promoter and enhancer reporter assays

In a promoter reporter assay (see the figure, part a), the DNA sequence from a genomic region suspected of being able to drive transcription is fused with a reporter gene, encoding a protein that gives a signal when the gene is expressed and the protein translated (for example, green fluorescent protein). A series of sequential deletions of the sequence are made, followed by measurement of the reporter gene activity. This can pinpoint sub-sequences carrying all or much of the capability of driving

transcription initiation within the genomic region. The sequential deletion process is often referred to as ‘promoter bashing’. Enhancer reporter assays work similarly (see the figure, part b) but fuse the DNA sequence of a genomic region suspected of having enhancer activity either upstream or downstream of a minimal promoter that needs enhancer activity in order to transcribe the reporter gene. This makes it possible to test the ability to enhance the minimal promoter for different enhancer candidates.



Box 2 | Identification of gene TSSs and core promoters genome-wide

Because transcription start sites (TSSs) are central to the identification of core promoters, techniques based on the sequencing of RNAs have been highly instrumental in this task. These techniques have been based on either full-length cDNA sequencing (flcDNA-seq)¹³⁵, which is typically based on large-scale Sanger sequencing, or high-throughput sequencing of the first 20–50 nucleotides of such full-length cDNAs (for example, cap analysis of gene expression (CAGE)¹³⁶ and TSS-seq¹³⁷) (see the figure, part a). The latter methods have the advantage that they allow for both identification of the location of TSSs and estimating the abundance of produced RNAs.

A set of complementary techniques that assess the rate of transcription rather than the amount of steady-state RNAs emerged later. These were based on the incorporation of labelled nucleotides into cells and sequencing either the 3' ends (global run-on sequencing (GRO-seq)¹³⁵ and precision nuclear run-on sequencing (PRO-seq)¹³⁸) or the 5' ends (GRO-cap¹³⁹, PRO-cap¹³⁸ and 5'-GRO-seq¹⁴⁰) of nascent RNAs, or on the isolation of RNAs bound by either RNA polymerase II (RNAPII) or chromatin (for example, native elongating transcript sequencing (NET-seq)¹⁴¹, mammalian NET-seq (mNET-seq)¹⁴², 3' end of nascent transcripts sequencing (3'NT-seq)¹⁴³ or start-site-associated RNA sequencing (Start-seq)¹⁴⁴) (see the figure, part b). Technically, methods that sequence the 3' ends of nascent RNAs are not designed for the detection of TSSs, since they may also capture RNAs from elongating RNAPII within gene bodies or stalled RNAPII at splice sites; however, because RNAPII pauses shortly after transcription initiation, the peaks of reads from such methods can be used as proxies for transcription initiation, and thereby promoter detection.

The advantage of nascent-RNA sequencing techniques is that RNA turnover will have limited effect, and therefore transcription, rather than steady-state RNA levels, is measured. Conversely, steady-state techniques often require only small amounts of RNAs and can therefore be used for applications with limited numbers of cells.

B, biotin-labelled dNTP; dNTP, deoxyribose nucleoside triphosphate; m7G, 7-methylguanosine cap.

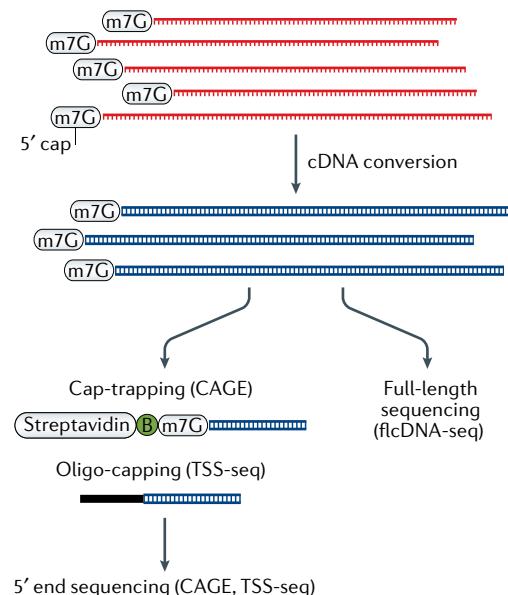
enhancer biology and has lately been scaled up using high-throughput sequencing techniques (see below).

Although various modes of enhancer function have been proposed (reviewed in REFS^{9,46,47}), following their discovery almost four decades ago — including trans-activation, RNAPII transfer, RNAPII promoter pause-release or anchoring TF–RNAPII condensates — almost all definitions of enhancers tacitly assume that enhancers and promoters are placed far apart in the genome and have distinct molecular functions and abilities.

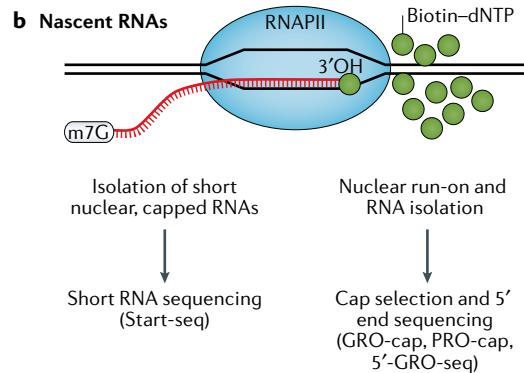
Regulatory element insights from genomics

A slew of techniques based on high-throughput DNA sequencing have enabled the study of enhancers and promoters genome-wide and at a large scale, including their role in diverse biological processes such as development, cell-type differentiation and specialization, and disease^{3,48–51}. Here, we briefly review the main methods for the detection and quantification of enhancers and promoters, which will also serve as a background for

a Steady-state RNAs



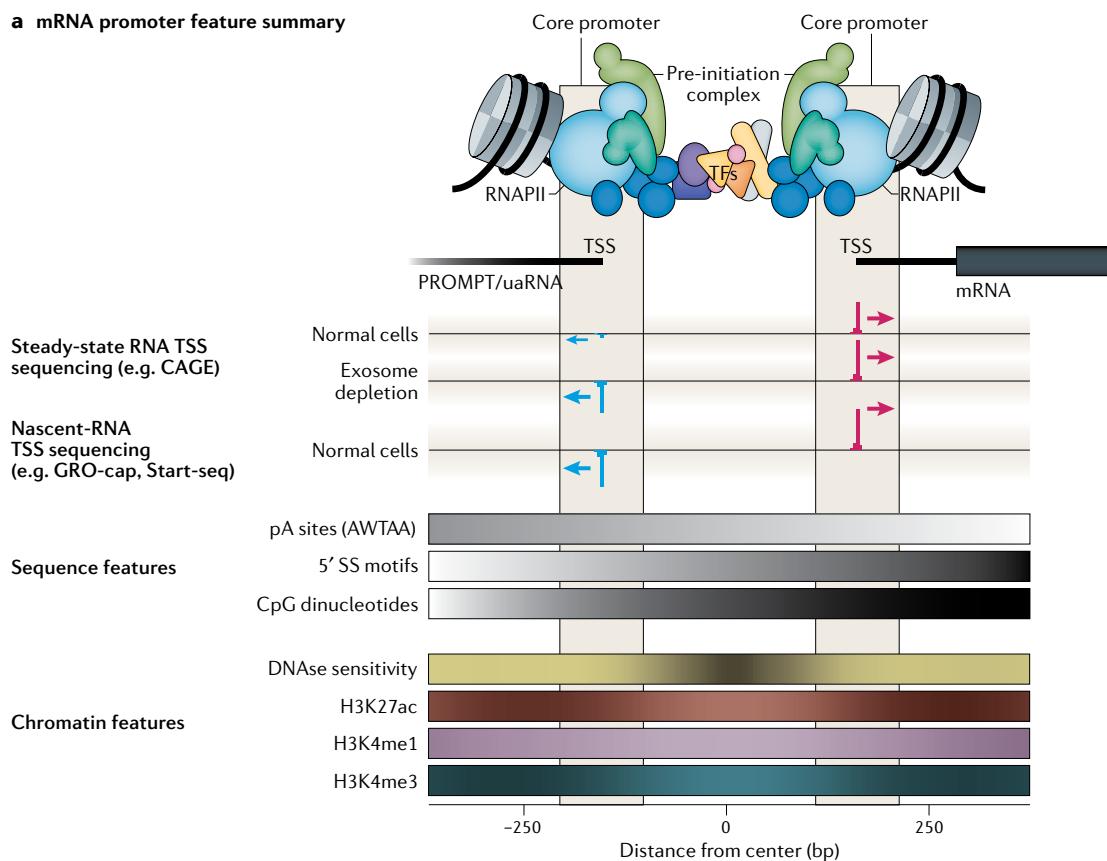
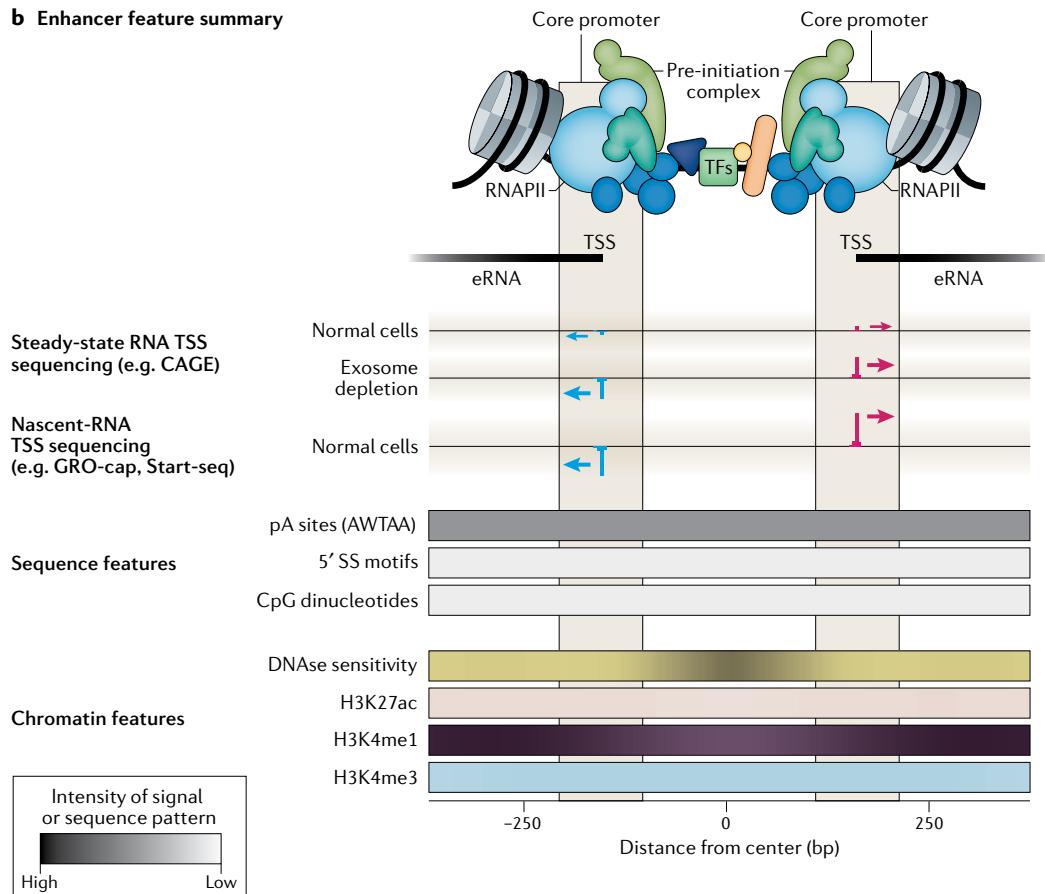
b Nascent RNAs



how our collective understanding of transcriptional regulatory elements has gradually shifted away from the originally clear-cut promoter/enhancer dichotomy.

Various methods developed to identify gene TSSs directly from 5' end sequencing of steady-state or nascent RNAs (BOX 2) have enabled the field to map the core promoter regions for the vast majority of mammalian cells. Together with indirect inference of promoters from stalled RNAPII, derived from 3' end sequencing of nascent RNAs, these techniques have revealed several unexpected features that were not evident from earlier studies based on single-gene reporter assays (FIG. 1a). These features include the following observations. First, RNAPII initiation is often dispersed over a local area, within the same nucleosome-depleted region (NDR), resulting in multiple close-by TSSs with varying initiation frequencies. This gives rise to a spectrum of core promoter TSS distribution ‘shapes’²⁷ (reviewed in REF.¹⁰). Second, most genes have many distinct TSS clusters (termed ‘alternative promoters’)^{27,52,53} arising

Nucleosome-depleted region (NDR). A region depleted of nucleosomes, often identified by DNaseI hypersensitivity, and often carrying regulatory potential.

a mRNA promoter feature summary**b Enhancer feature summary**

◀ Fig. 1 | Features used to distinguish promoters and enhancers. The figure shows a summary of the key features often used to distinguish mRNA gene promoters (part a) and enhancers (part b). At the top of each panel, a schematic of the nucleosome and nucleosome-depleted region (NDR) placements is shown, together with the binding of transcription factors (TFs) within the centre of the NDR and the pre-initiation complex, with RNA polymerase II (RNAPII) positioned at the NDR edges. The second row shows the produced RNAs as black boxes, initiating at the edges of the NDR (enhancer RNAs (eRNAs)) for enhancers, and pairs of promoter-upstream transcripts (PROMPTs; also known as upstream antisense RNAs (uaRNAs)) and mRNAs at mRNA promoters. The remaining rows show typical signal intensity levels from experiments used in enhancer/promoter identification and characterization, as follows. The first two rows show typical signals from techniques detecting transcription start sites (TSSs) by sequencing the 5' ends of steady-state RNAs (for example, by cap analysis of gene expression (CAGE)), with or without depleting the nuclear exosome. Signal intensity (the number of mapped reads) is represented as bar plots, with colour and arrows indicating the direction of transcription. The next row shows typical results from techniques sequencing the 5' ends of nascent RNAs (for example, GRO-cap or Start-seq) in normal cells using the same graphical representation. Note that because some RNAs are subject to exosomal degradation, the signals from steady-state RNAs in cells with depleted exosomes are more similar to those of nascent-RNA techniques in normal cells. Sequence features including poly(A) (pA) sites, 5' splice sites (5' SSs) and CpG dinucleotides are visualized as greyscale bars showing their average pattern density, where black indicates higher density. Importantly, pA site densities at PROMPTs and eRNAs are roughly the same as the genomic average, but are lower downstream of mRNA TSSs. Chromatin features from DNasel hypersensitivity sequencing (DNase-seq) and chromatin immunoprecipitation coupled to sequencing (ChIP-seq), for acetylation of histone H3 lysine 27 residues (H3K27ac), mono-methylation of histone H3 lysine 4 residues (H3K4me1) and H3K4me3 histone modifications, are visualized as coloured bars showing average signal intensity. Historically, promoter activity and gene TSSs have been identified by 5' end RNA/cDNA sequencing methods, whereas enhancers have typically been predicted by the H3K4me1-to-H3K4me3 signal ratio, and later also by the detection of bidirectionally transcribed eRNAs by, for example, the CAGE or GRO-cap method. As discussed in the main text and FIG. 2, most such measurements are directly related to promoter activity and only indirectly related to enhancer activity.

Nascent-RNA techniques

A wide range of methods aimed at capturing RNA as it is being transcribed. See BOX 2.

DNase-seq

A method for identifying accessible regions of the genome, based on DNasel hypersensitivity. See BOX 3.

Promoter-upstream transcript

(PROMPT). A short RNA (also known as upstream antisense RNA (uaRNA)) that is transcribed upstream and on the opposite strand from an mRNA transcription start site and is typically degraded by the nuclear exosome. It has many similarities to an enhancer RNA.

Nuclear exosome

A multi-protein complex responsible for the degradation of RNAs from the 3' end.

Poly(A) sites

Sequence patterns (AT/ATAA) associated with the 3' ends of genes, but that also occur with high frequency in intergenic DNA.

from independent core promoters in separate NDRs, typically separated by several hundred or thousand base pairs, where the choice of an alternative promoter may alter the final protein product⁵⁴. Third, RNAPII pauses downstream of the TSS before entering a state of active elongation^{55–57}. Fourth, and perhaps most unexpectedly, the vast majority of gene TSSs are accompanied by an additional, proximal, upstream TSS on the opposite strand^{55,57,58}. The latter feature, termed ‘divergent transcription’ (reviewed in REFS^{59,60}), arises from a configuration in which each TSS is positioned in independent core promoters close to the edges of nucleosomes flanking an NDR^{12,13}. Generally, the upstream, divergent TSS produces short (<500 bp), unspliced transcripts, termed ‘promoter-upstream transcripts’ (PROMPTs⁵⁸) or ‘upstream antisense RNAs’ (uaRNAs⁵⁵), that are degraded by the nuclear exosome, an exonuclease complex that degrades RNAs from the 3' end⁵⁸. The difference in exosome sensitivity between PROMPTs and mRNAs is, at least partially, due to a depletion of early poly(A) sites and an enrichment of 5' splice sites (5' SSs) downstream of mRNA TSSs, and the opposite patterns downstream of PROMPT TSSs^{61,62}.

In parallel, chromatin immunoprecipitation coupled to either microarrays (ChIP-chip) or high-throughput sequencing (ChIP-seq)⁶³ has made it possible to measure bound TFs and the chromatin state around gene TSSs, in particular in terms of specific post-translational modifications of histones. It was found that transcriptionally

active TSSs are depleted of nucleosomes and have enrichments of tri-methylation of histone H3 lysine 4 residues (H3K4me3) and acetylation of histone H3 lysine 27 residues (H3K27ac) in the first nucleosomes downstream of TSSs^{64–67}. The same techniques also showed that the vast majority of bound TFs around gene TSSs are located within the NDR between the divergent TSS pair^{13,68}, and that most active gene promoters have accumulated levels of stalled or paused RNAPII^{69–71} (FIG. 1a). Techniques based on ChIP and DNA accessibility (reviewed in REFS^{63,72}) were instrumental in early genomics studies of enhancers (BOX 3). The characterization of various histone modifications, in particular methylations and acetylations, around the inferred binding sites of enhancer-associated co-activators or around well-studied enhancers allowed the establishment of rules designed for discriminating enhancers from gene promoters (FIG. 1b). Using these specific patterns, such as a high H3K27ac signal⁷³ and a high H3K4me1-to-H3K4me3 ratio at putative enhancers^{74,75}, quickly became the state of the art and has been the basis for several large consortia (for example, the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics) in genome segmentation algorithms^{76–78} used to classify genomic regions into different classes of transcriptional regulatory elements.

An unexpected finding when combining RNA-sequencing and ChIP-seq techniques was that many enhancer candidates initiated transcription of so-called enhancer RNAs (eRNAs) at the edges of their NDRs in a divergent manner^{21,22}. eRNAs are short, typically unspliced and non-polyadenylated transcripts that are degraded by the exosome, and therefore they share many similarities with the PROMPTs arising from gene promoter NDRs^{3,12} (FIG. 1b). In contrast to gene promoters, where PROMPTs tend to be less transcribed than the mRNA, the divergent transcription associated with candidate enhancers are generally more bidirectionally balanced, and both transcripts are exosome substrates³. We and others have shown that, on the basis of these properties, it is possible to predict enhancers from 5' end RNA-sequencing data alone^{7,9,80}. Furthermore, transcribed enhancer candidates were more likely to validate in reporter assays than non-transcribed enhancer candidates predicted by histone modifications alone³.

Collectively, these findings show that promoter and enhancer candidates, as defined by the respective methodologies above, share some characteristics and abilities but may differ in others. This motivates a critical investigation of the similarities of such regulatory elements.

Similarity and duality of enhancers and promoters

We will here summarize similarities of enhancer and promoters, taking recent genomics data into consideration, to arrive at the conclusion that enhancers and promoters are, in contrast to their historical definitions, unexpectedly similar in terms of overall chromatin, sequence, and core promoter architecture. When active, both are encompassed by NDRs, bound by RNAPII, and initiate transcription divergently at NDR edges^{11,12,14,81}. Whereas protein-coding gene promoters give rise to mRNAs on their sense strands and PROMPTs on their antisense strands, enhancers are transcribed into eRNAs

Box 3 | Identification of enhancers by DNA-binding proteins and DNA accessibility

Chromatin immunoprecipitation (ChIP) coupled to microarrays (ChIP-chip) and ChIP coupled to sequencing (ChIP-seq) techniques have been central for the identification of candidate enhancers, either by targeting known important transcription factors (TFs) of the cells or tissues of interest^{50,145}, or by assessing the enrichment of histone modification patterns genome-wide (see the figure, part a), sometimes in combination with targeting proteins expected to be involved in enhancer activities, such as the P300–CBP histone acetyltransferase^{50,74,84,146–148}. Other approaches have also assessed DNA accessibility⁷², first through DNasel hypersensitivity (DNase-seq)¹⁴⁹ and more lately through the assay for transposase-accessible chromatin (ATAC-seq)¹⁵⁰ (see the figure, part b), as means to predict enhancers. For all methods, the resulting sequenced DNA reads are mapped to a reference genome, where the mapping density can be used to assess the strength of binding or DNA accessibility (see the figure, part c). Through analysis of histone modifications around P300 binding sites and a small handful of well-characterized enhancers available from earlier works and strongly active gene promoters, a pattern emerged in which enhancers displayed higher levels of mono-methylation (H3K4me1) than of tri-methylation (H3K4me3) of histone H3 lysine 4 residues^{74,75}, setting them apart from gene promoters (FIG. 1). It was possible to train computational models based on these patterns to predict the locations of enhancers (and of gene transcription start sites (TSSs)) on the basis of these histone marks alone. Acetylation of histone H3 lysine 27 residues (H3K27ac) was later suggested to distinguish between active and inactive enhancers⁷³, and other histone acetylations were later suggested to have similar or specific predictive performances compared to H3K27ac^{151,152}.

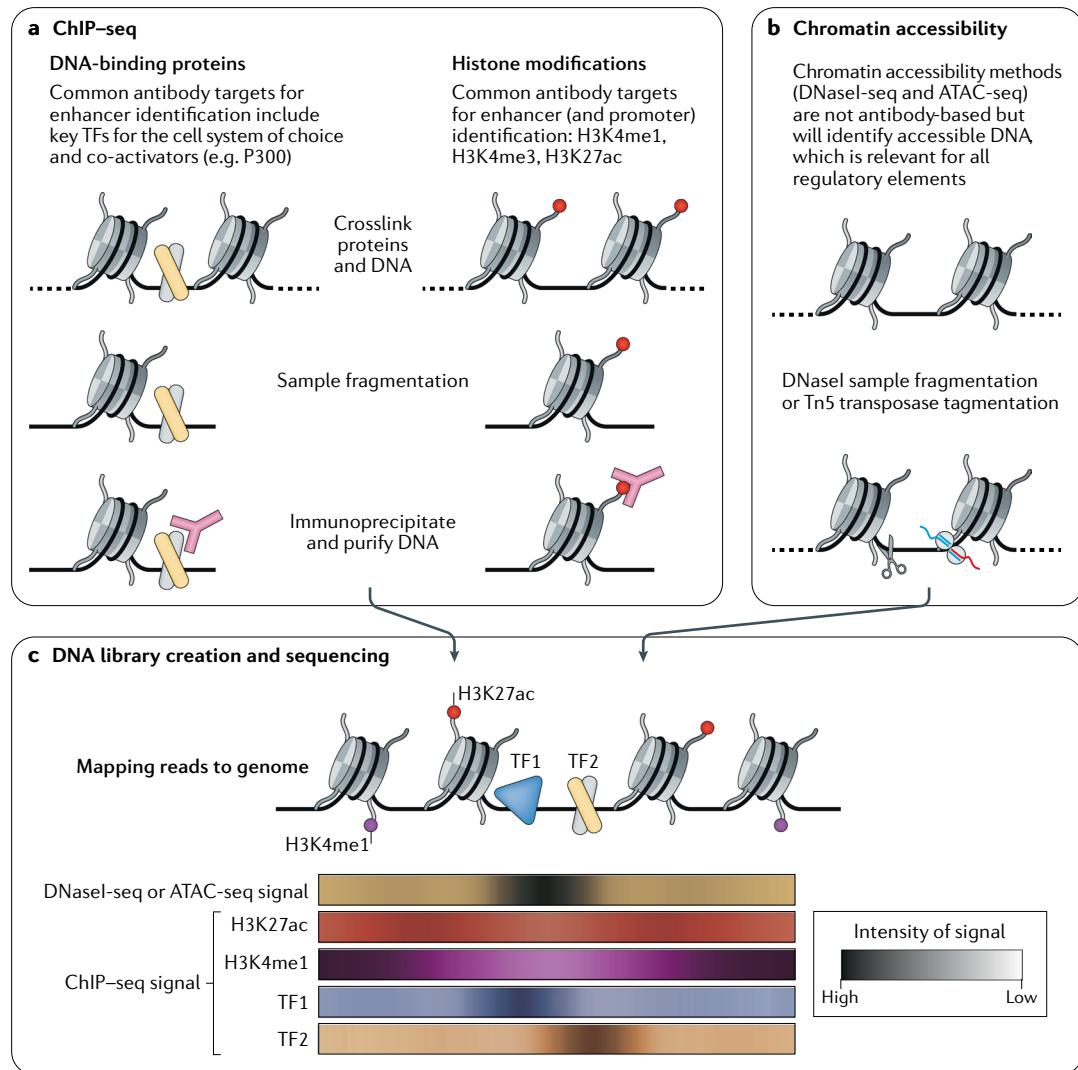


Figure adapted from REF.⁶³, Springer Nature Limited.

ChIP-seq
Chromatin immunoprecipitation coupled to sequencing; a method for finding DNA–protein interactions by combining immunoprecipitation and high-throughput DNA sequencing. See BOX 3.

Enhancer RNAs
(eRNAs). Short RNAs (<500 bp) that are transcribed from enhancers, with many similarities to promoter-upstream transcripts (PROMPTs).

ATAC-seq
Assay for transposase-accessible chromatin; a method for identifying accessible regions of the genome, based on transposase activity.

CpG islands
Genomic sequences that are not depleted of cytosine–phosphate–guanine (CpG) dinucleotides, which would occur by 5-methylcytosine deamination. They often overlap or are near transcription start sites. Most definitions set a minimum length (for example, 200 or 500 bp) and a minimum observed/expected CpG ratio.

in both directions. This means that gene promoters and enhancers share the same promoter architecture¹², both having a pair of divergent TSSs, each with its own core promoter^{3,12,13,15,16} (FIG. 1). This promoter architecture also seems to be shared with respect to sequence preferences: the TSSs often have INR elements and there

is a weak preference for upstream TATA boxes even at eRNA TSSs.

One marked difference between mammalian gene promoters and enhancers is the overall CG content. Roughly 50% of gene promoters overlap CpG islands, whereas almost no enhancers do³ (FIG. 1). Another

apparent difference is the amounts of associated transcriptional activity and RNAs produced^{3,11,12}. While many candidate enhancers can act as autonomous promoters in vitro⁸², active gene TSSs produce on average 17-fold more RNAs (as assessed by cap analysis of gene expression (CAGE)) than candidate enhancer TSSs in HeLa cells in vivo¹², and are associated with 2- to 3-fold more transcriptional activity (as assessed by precision nuclear run-on sequencing (PRO-seq)) in GM12878 cells¹¹. Thus, enhancers initiate less transcription and, on top of that, generate RNAs that are often degraded in the nucleus, thus further depleting the final RNA copy number.

However, it is important to consider that this average difference will be driven by a subset of highly expressed gene TSSs and that the overall chromatin modification and sequence properties of regulatory elements are highly correlated with their transcriptional output. For instance, highly transcriptionally active and/or ubiquitously expressed TSSs are more likely to reside within CG-rich areas⁸³ and are much more likely to be enriched by flanking H3K4me3-modified histones^{84,85}, which are two key features that have been used to distinguish enhancers and promoters (FIG. 1). In fact, many candidate enhancers are associated with H3K4me3 (REFS^{86,87}), and the levels of chromatin modification associated with respective TSSs are highly similar if the difference in transcriptional activity is taken into account^{11,87,88} (FIG. 2). Thus, most of the features used to distinguish enhancers and promoters reflect differences in transcriptional activity and RNA output. Hence, classical enhancer/promoter-discerning histone modifications are not optimal in terms of classifying regulatory elements by their function, as they are likely to reflect only the ability to initiate transcription (promoter activity), not the ability to enhance distal transcription (enhancer activity).

Early attempts at defining enhancers and promoters assumed that they had distinct and separable functions. As we stated above, many regulatory elements with enhancer activity also have promoter activity, because they can initiate local transcription. To make the situation even more complex, recent studies have shown that gene promoters may also have enhancer activity. For instance, 17 out of 45 identified regulatory elements with in vivo enhancer activity in human embryonic stem cells around the *POU5F1* gene locus were annotated gene promoters¹⁷, 3% out of 20,719 tested gene promoters in human K562 cells had strong enhancer activity in vitro¹⁸, and 5 out of 12 tested promoters of long non-coding RNAs acted as enhancers in mouse embryonic stem cells in vivo²⁰. In addition, chromatin conformation data indicate that a sizeable fraction of the distal regions that interact with gene promoters are in fact other gene promoters^{89–91}, suggesting prevalent enhancer activity from gene promoters. Examples include the *INS* gene promoter, which interacts over a long distance with the *SYT8* gene promoter to regulate its expression in human islets⁹², and hundreds of promoter–promoter interactions that likely mediate the effect of expression quantitative trait loci (eQTLs) in human primary blood cells⁹¹.

Thus, as a summary, the regulatory elements previously designated as enhancers typically have promoter

activity, and typically were predicted on the basis of features more related to promoter than to enhancer activity, and many known gene promoters have enhancer activities. This motivates new models and definitions of regulatory elements that will encompass these findings.

An updated model of regulatory elements

Because of the duality of many regulatory elements to act as both promoters and enhancers, we and others have proposed a model that does not assume that promoter and enhancer activities are mutually exclusive, but rather are two capabilities that regulatory elements might have different degrees of^{81,87,88,93,94}. In this model, any NDR that can affect transcription initiation either locally or distally is defined as a regulatory element. Specifically, on the basis of findings from the genomics techniques reviewed above, a regulatory element has the following properties. First, it is centred and mostly contained within one NDR, which is often bound by TFs. Second, it may have a degree of promoter activity: the ability to recruit RNAPII and initiate transcription at either NDR edge. Third, it may also have a degree of enhancer activity: the ability to positively influence transcription initiation at other regulatory elements (FIG. 3a).

The conceptual difference from previous models is that there is no imposed enhancer/promoter dichotomy: a regulatory element may have any combination or degree of promoter and/or enhancer capabilities defining its functional repertoire. This model also deals with some of the historical problems in our conceptions and definitions of promoters and enhancers. First, the promoters in the literature are based on arbitrary lengths from gene TSSs (for example, –300 to +100, or –2,000 to +200), and a common assumption (which is still held by parts of the transcription community) is that bona fide promoters must initiate the transcription of mRNAs or other functional/stable RNAs. In our proposed model above, the lengths of regulatory elements are based on accessible DNA, and the types of RNAs produced and their stability are decoupled from the concept of promoter activity, which only refers to the ability to initiate transcription locally. The latter property makes sense because RNAPII transcription initiation at regulatory elements seems to follow the same rules, regardless of the type of RNAs produced and of whether the regulatory element also has enhancer function. Thus, a regulatory element with strong enhancer activity that initiates eRNA transcription will also be considered to have a degree of promoter activity. Second, in the classical definition, an enhancer must be able to enhance distal transcription initiation regardless of distance and orientation. Why would a regulatory element that is able to enhance transcription initiation at another locus, but that is only effective within a moderate distance or in a specific orientation, not be considered to have enhancer activity? In the framework above, no such constraints are imposed.

Measuring regulatory potential en masse

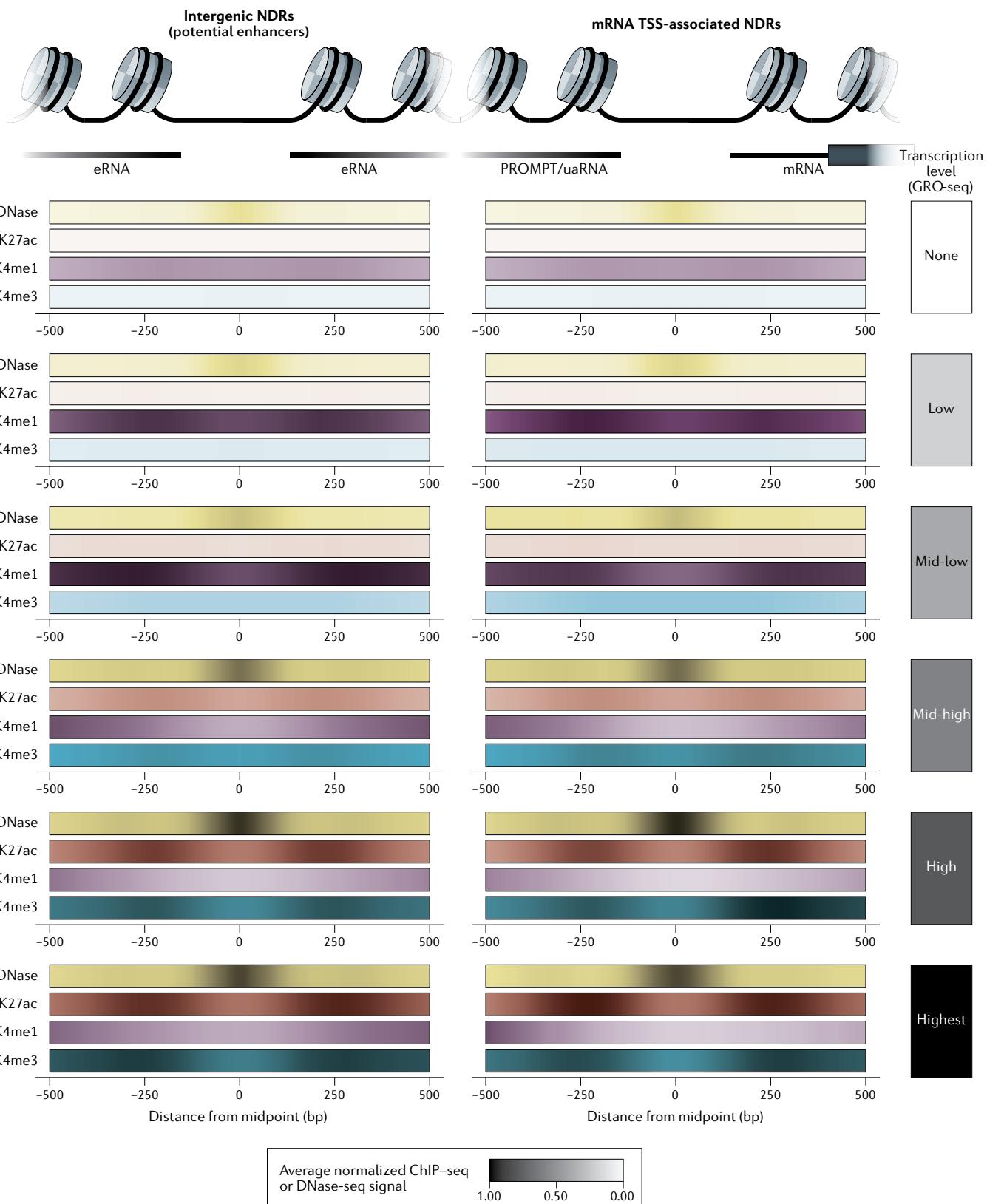
The model framework proposed above is only of conceptual interest if we can measure the intrinsic enhancer and promoter strength of a regulatory element and,

CAGE

Cap analysis of gene expression; a method to identify transcription start sites by sequencing the 5' ends of capped, steady-state RNAs. See BOX 2.

Expression quantitative trait loci

(eQTLs). Regions of DNA in which genetic variation is associated with variability in the expression of one or more genes.



secondly, deduce what drives the respective potential. Conceptually, there are two major, and complementary, experimental avenues towards answering these questions: CRISPR-based *in vivo* methods (BOX 4) and massively parallel reporter assays (MPRAs) (BOX 5).

CRISPR-based *in vivo* genome-editing methods (BOX 4) allow either direct editing, mutagenesis and deletion^{17,19,20,95–98} (CRISPR–Cas9) of regulatory elements with enhancer and/or promoter activity, or indirect interference with (CRISPRi)^{17,99} or activation

◀ Fig. 2 | Chromatin features used to identify and distinguish enhancers and promoters are correlated with the promoter activity of regions, regardless of genomic location. In this analysis, based on chromatin and global run-on sequencing (GRO-seq) data from HeLa cells¹², intergenic nucleosome-depleted regions (NDRs) (potential enhancers), identified from DNase hypersensitive sites (DHSs), and NDRs that overlap annotated mRNA transcription start sites (TSSs) are compared (left and right columns; also shown with the cartoons at the top, as in FIG. 1). The average intensity of chromatin immunoprecipitation coupled to sequencing (ChIP-seq) data for typical histone modifications used to predict enhancers or promoters, and DNaseI hypersensitivity sequencing (DNase-seq) intensity, as a function of location relative to the DNase-seq signal summit positions of DHSs (referred to as the DHS midpoint, 0 on each x-axis) are shown as coloured bars, where colour intensity corresponds to the average signal intensity (same colours as in FIG. 1). The DHSs in each category are split by their total transcription level on both strands (measured by GRO-seq) into six bins (row groups; see the labels on the right). Note that there is no substantial difference between intergenic and mRNA TSS-overlapping chromatin states if the transcription level (and thereby, promoter activity) is taken into account. Therefore, chromatin-based classification of enhancers and promoters is likely to reflect promoter rather than enhancer activity. eRNA, enhancer RNA; H3K4me1, mono-methylation of histone H3 lysine 4 residues; H3K27ac, acetylation of histone H3 lysine 27 residues; PROMPT, promoter-upstream transcript; uaRNA, upstream antisense RNA.

of (CRISPRa)¹⁰⁰ such elements, followed by measuring the resulting RNA output from one or more regions of interest. The main advantage of such methods is that they take the complexity of the genome organization into account, including the local chromatin context and interactions between multiple regulatory elements, because loci are targeted in their endogenous genomic context. There are two main drawbacks. First, it can be difficult to assess the causality and impact of individual regulatory elements in a larger region; for example, two regulatory elements with enhancer activity within the same larger genomic region may be redundant, in the sense that each can compensate for loss of the other. Second, these methods are hard to scale up beyond detailed investigations of one or a few regions around one gene or gene cluster. This means that while such methods can give unprecedented insights for a chosen genomic region, it is challenging to build generic, quantitative and predictive models from such data.

MPRAs (BOX 5) attempt to decouple the intrinsic promoter and/or enhancer potential of single sequences from the regulatory potential provided by their native genomic contexts. In essence, these methods build upon classical enhancer and promoter reporter assays but are, through barcode integration and coupling to high-throughput sequencing, scaled up in order to test thousands of candidate elements. The main advantage, beyond their scale, of plasmid-based MPRA methods is also their in-built disadvantage relative to *in vivo* genome-editing methods: they can assess the enhancer/promoter potential of regulatory elements decoupled from their native context, including local chromatin state, as discussed above. Thus, with such data it is possible to model what intrinsic features of putative regulatory elements drive their promoter or enhancer potential. For this reason, we mainly focus on MPRA-derived features in this Review, although, as we point out below, *in vivo* genome-editing methods and MPRAs are complementary, and such methods/data ideally should be combined.

Massively parallel reporter assays

(MPRAs). Methods that can measure the promoter or enhancer activity of many candidate DNA sequences in parallel. See BOX 5.

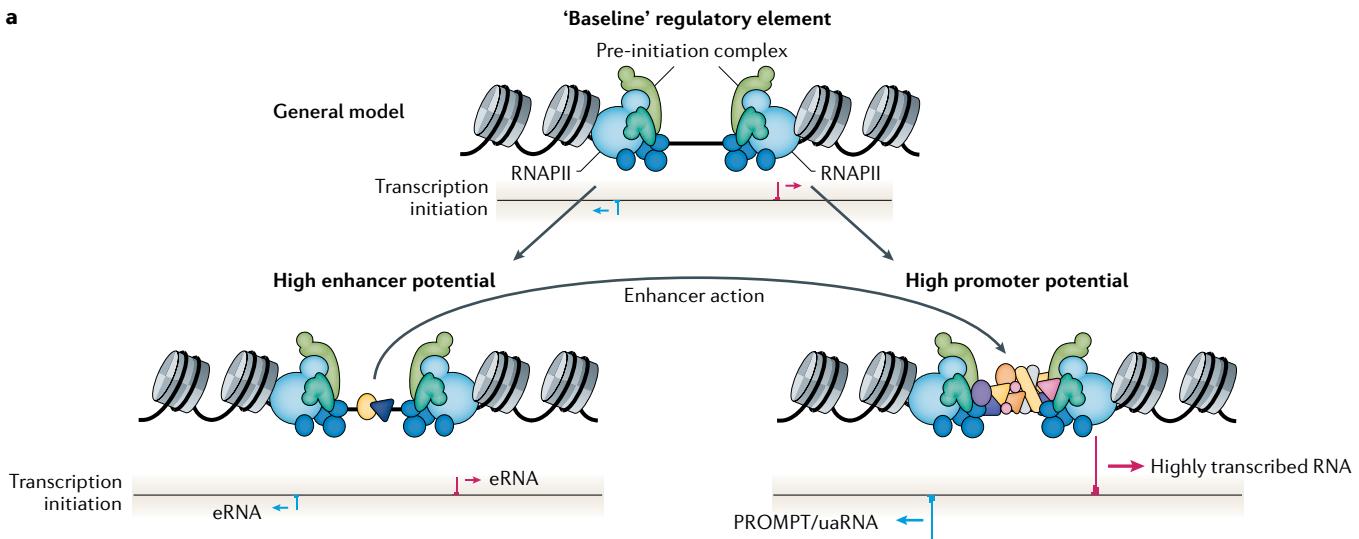
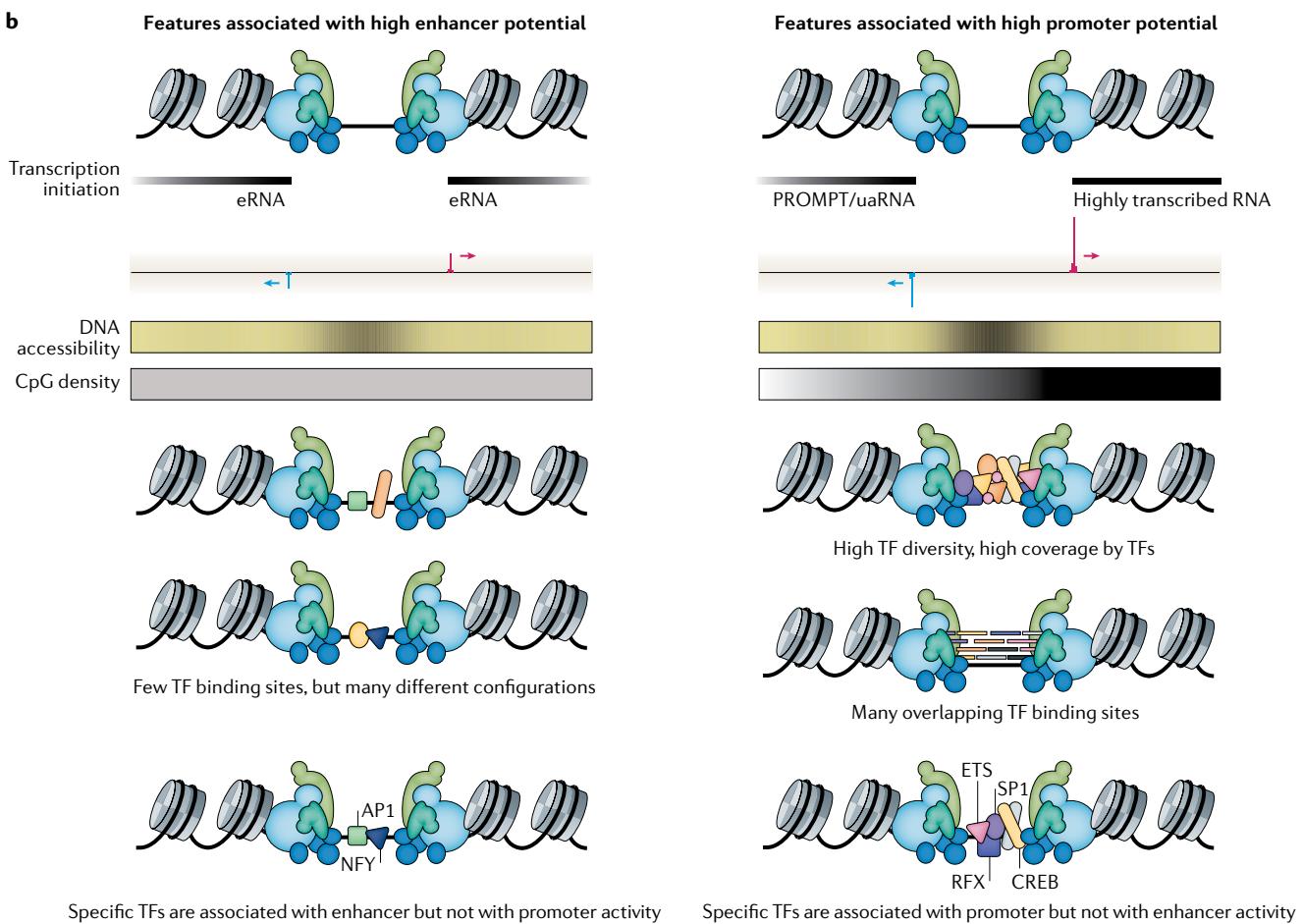
DNase hypersensitive sites (DHSs). Highly accessible genomic regions identified by DNase-seq.

Determinants of regulatory potential

MPRA-based methods have resulted in several new insights. In particular, it has been possible to gain a first view, beyond single loci, of what sequence features drive strong promoter and/or enhancer activity. MPRA approaches have also enabled the study of the relationship between these activities. For example, is it possible for a regulatory element to have both strong enhancer and strong promoter activity? Do all regions with enhancer activity have some promoter activity? Here, we review studies aiming to answer such questions (FIG. 3b).

Nguyen et al.¹⁰¹ used promoter and enhancer MPRA to test both candidate enhancers and gene-associated promoters that displayed increased H3K27ac and CREBPP co-activator binding upon mouse cortical neuron activation. Interestingly, the MPRA readouts demonstrated a moderately positive correlation between enhancer and promoter activity. However, most candidate enhancers had limited promoter activity, and the tested gene promoters tended to have strong promoter activity but, surprisingly, enhancer activity similar to that of the tested candidate enhancers. These results indicate that regulatory elements intrinsically have both enhancer and promoter activity, but that some have features that specifically encode substantially stronger promoter activity.

These findings are reflected in studies using MPRA that have aimed to identify features that are indicative of promoter activity (FIG. 3b, right panel). First, the overall CpG content is positively correlated with promoter but not with enhancer activity¹⁰¹. However, it is probably not the CpG dinucleotides themselves that cause the higher promoter activity, but rather TFs that favour CpG-containing binding site sequences¹⁰¹. Indeed, studies have shown that a subset of TF binding sites or other sequence patterns (either predicted or measured by, for example, ChIP-seq) are associated with promoter activity. It is important to consider that these results are likely to be context-specific to at least some degree, as the repertoire and availability of TFs are different between cells. Weingarten-Gabbay et al.¹⁰² showed, using MPRA, that strong promoter activity is highly associated with the presence of core promoter elements, in particular strong TATA and INR sites with canonical spacing and orientation. This is expected and recapitulates the body of work in small-scale reporter assays. Predicted sites for particular TFs — for example, SP1, ETS, CREB and RFX — have also been found to be associated with high promoter activity^{101–103}. Interestingly, only some TFs — for example, ETS and SP1 — have additive effects, so that the number of predicted binding sites within a region is correlated with promoter activity. On top of this, high promoter activity is associated with having several overlapping predicted binding sites for different TFs — in particular, ubiquitously expressed ones¹⁰³. This is in turn related to general DNA accessibility: regulatory elements within highly accessible regions typically have high promoter activity, and the degree of accessibility is related to the expression levels of bound TFs¹⁰³. Indeed, the introduction of nucleosome-repellent sequences results in overall higher promoter activity¹⁰².

a**b**

• At minimum, low promoter activity

- Low CpG density
- Often lower DNA accessibility
- Few TFs per region, often bound to AT-rich sequences (e.g. AP1, NFY)
- Might have imperfect TF binding
- Many diverse configurations of TFs (and/or recruited co-activators) may lead to strong enhancer activity
- Seldom highly evolutionarily conserved
- May be the 'base state' of regulatory elements

• Often high DNA accessibility, higher density of nucleosome-repellent sequences

- Strong core promoter elements
- High CpG density, but CpG dinucleotide effect is mainly due to TFs binding CpG-rich binding sites
- Many TF binding sites per region: high diversity per site and high TF binding coverage of the NDR, high overlap between binding sites
- Some TFs have a clear promoter-driving effect, including CREB, SP1, ETS and RFX-class TFs
- Only some TFs seem to have an additive effect
- Often evolutionarily conserved

◀ Fig. 3 | General model of regulatory elements and of features associated with promoter and enhancer potential. **a** | A regulatory element is defined as a nucleosome-depleted region (NDR) that can affect transcription initiation either locally or distally (the graphics are as in FIG. 1, but simplified). Regulatory elements may have various degrees of promoter and enhancer activities, and these capabilities are not necessarily associated with regulatory elements in a mutually exclusive manner. **b** | Features associated with high enhancer activity (left) and high promoter activity (right) from massively parallel reporter assays (MPRAs); graphical elements are organized as in FIGS 1 and 2. Brief summaries of the features are shown as cartoons and bullet points; see the main text for further discussion. CpG, cytosine–phosphate–guanine; eRNA, enhancer RNA; PROMPT, promoter-upstream transcript; RNAPII, RNA polymerase II; TF, transcription factor; uaRNA, upstream antisense RNA.

The characterization of selected enhancer candidates through saturation mutagenesis has demonstrated that mutations associated with a change in enhancer activity often fall in TF binding sites^{97,104–107} and affect the activity in a cell-type-specific manner¹⁰⁶, demonstrating the importance of TFs in driving not only promoter activity but also cell-type-specific enhancer activity. As with promoter activity, the presence of binding site sequences for certain TFs, their combinations and the number of such sites are associated with enhancer strength^{108–110}. However, unlike the TFs that have been associated with high promoter activity, mRNA gene promoters that also have high enhancer activity were enriched with predicted binding sites of TFs that favoured low CG content^{3,18,101} (FIG. 3b, left panel). In addition, an MPRA study assessing the core promoter response to various co-activators showed that restricted compatibilities between certain co-activators and core promoters were partly explained by their CG content¹¹¹. Thus, in addition to the core promoter strength and binding site complexity discussed above, differences in enhancer and promoter potential may be influenced by which TFs, or TF-recruited co-activators, bind to a regulatory element. Indeed, the binding of AP1 and NFY is biased towards regulatory elements with enhancer activity, whereas the binding of other TFs — for example, CREB — is biased towards regulatory elements with promoter activity^{101,106,112–114}. Because high CpG content regions are generally nucleosome repelling¹¹⁵, there may be a specific need for binding of certain TFs favouring low CpG sequences, such as AP1 (REF.³), to CpG-poor regulatory elements in order to modulate chromatin accessibility¹¹⁶. These may include so-called pioneer transcription factors that are able to bind nucleosomal DNA and to recruit chromatin remodelers¹¹⁷. Indeed, the activation of candidate enhancers has been linked to proper recruitment of the SWI/SNF (BAF) chromatin remodelling complex¹¹⁸, and genetic variants that both disrupt TF binding sites and affect chromatin accessibility are highly enriched in candidate enhancers¹¹³.

Thus, if differences in CpG content are taken into account, regulatory elements with MPRA-inferred enhancer activity are to a large extent similar to those with low promoter activity¹⁰¹. This suggests that the sequence features encoding enhancer activity also generalize to (low) promoter activity, but not necessarily the other way around, as has been indicated by sequence-based machine learning¹¹⁹. In an interesting parallel in flies, it has been shown that the enhancer strength of regulatory elements depends on clusters of suboptimal TF binding site sequences^{120,121}, and that the optimization

of TF binding affinity leads to increased promoter activity and loss of enhancer activity. Certain exceptions to the relationship between low promoter activity and high enhancer potential exist. For example, annotated mRNA gene promoters that had high enhancer activity were highly expressed^{17,18}, and it remains to be identified what determines such a strong duality. The general lack of distinct features driving enhancer activity, apart from a few biased TFs, may explain why known regions with high enhancer activity and weak promoter activity feature much higher evolutionary turnover in comparison to gene promoters¹²². High promoter activity may require selection for certain features, while the constraints on enhancer activity may be lower. Alternatively, this pattern may mean that there are many more diverse ways to encode enhancer activity, which in turn may be more specific to particular cell types or contexts.

Models explaining enhancer/promoter duality

Given the pervasiveness of regulatory elements with both enhancer and promoter potential, it is worthwhile to speculate on the origin and reason behind this duality. Several models have been suggested that present functional roles of eRNAs or the act of enhancer transcription (reviewed in REFS^{9,123}), thereby providing a reason for promoter activity of regulatory elements with high enhancer activity. Since many eRNAs have low transcription, are subject to nuclear degradation and are transcribed from DNA with high evolutionary turnover, hypotheses concerning eRNA functions that are contingent on high copy numbers and/or sequence-specific interaction are unlikely to be true in the large majority of cases. Functions that are based on the act rather than the product of transcription are more likely, because they are not as affected by the constraints discussed above. For instance, divergent transcription may induce negative supercoiling and thereby affect the chromatin structure and accessibility around a regulatory element, or it may spread permissive histone modifications to flanking nucleosomes through RNAPII carboxy-terminal domain-associated acetyltransferases and methyltransferases, thereby acting as a barrier to spreading repressive marks⁶⁰. Alternatively, divergent transcription may simply be a way of displacing RNAPII from the NDR. An opposing view is that the transcription from regulatory elements with enhancer activity may merely be noise caused by spurious transcription of RNAPII at accessible chromatin¹²⁴ in foci with high RNAPII concentrations. However, candidate regions with enhancer potential primed for subsequent activation are generally accessible but not transcribed⁵¹.

Alternatively, the recruitment and initiation of RNAPII (promoter activity) may instead lead to enhancer activity. In this model⁹³, which bears a resemblance with the hub and TF-RNAPII liquid condensate models of transcriptional regulation⁴⁶, regulatory elements that associate in three-dimensional space, possibly through homotypic attraction³⁹, work in concert to increase or maintain the local concentration of TFs, GTFs, co-activators and RNAPII, which transiently bind, engage and relocate between physically proximal elements. According to this view, having multiple

Pioneer transcription factors

Transcription factors that can directly bind nucleosomal DNA, possibly in compacted chromatin.

Homotypic attraction

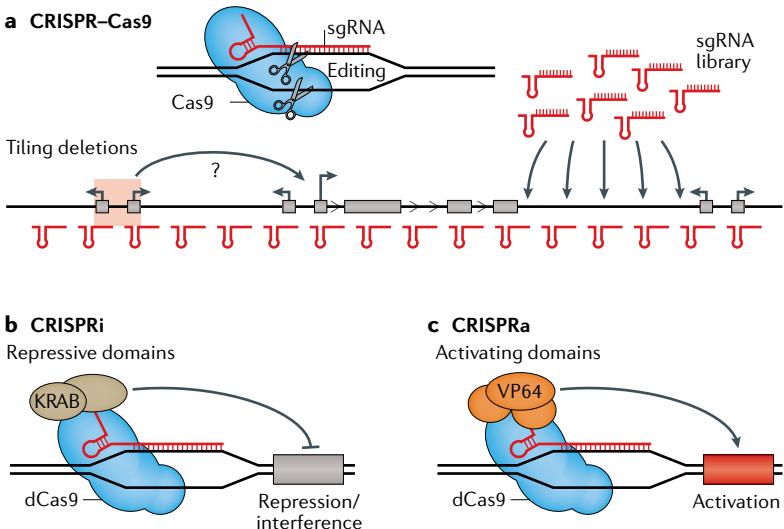
A force driving chromatin with similar characteristics or associated proteins to self-associate. The physical association may be formed by protein bridges or liquid–liquid phase separation, the latter of which involves molecules separating into liquid condensates with specific compositions.

Box 4 | In vivo CRISPR-based methods to modulate regulatory activity

Recent technology developments based on CRISPR systems to target enzymes (for example, the Cas9 nuclease) via single-guide RNAs (sgRNAs) in order to cleave specific strands of DNA¹⁵³ have enabled precise and accurate genetic perturbation of regulatory elements^{17,20,95–97}. In particular, CRISPR–Cas9 has been used to systematically delete regions within loci (tiling deletions) followed by gene expression readouts (see the figure, part a). In such studies, a library of sgRNAs is designed to target tiling genomic regions for deletion. In a given cell, the expression of a pair of these sgRNAs from the library is used to delete the region between the cut sites (the red shaded region in the figure). Assessing the resultant effects on proximal gene activity can implicate functional roles for candidate regulatory elements within the deleted region.

Based on this system, variant assays have been developed. CRISPR interference (CRISPRi) uses sgRNAs to target a nuclease-deficient Cas9 (dCas9) to regions of interest, alone or fused with repressive domains such as the Krüppel-associated box (KRAB) and the SIN3A interacting domain of MAD1 (SID), whereby interference with regulatory activity is achieved through steric blocking and/or repression^{98,99,154,155} (see the figure, part b).

The dCas9 system may further be utilized to activate regulatory elements, termed CRISPR activation (CRISPRa), through fusion of dCas9 with transcriptional activators^{98,100,156}, for example, VP64 (see the figure, part c). The versatility of the CRISPR toolbox thus demonstrates great promise to study native, context-aware transcriptional regulation (for further reading, see REF.¹⁵⁷).



regulatory elements in close three-dimensional proximity increases the probability of TF and RNAPII recruitment to target gene TSSs (FIG. 4). A feature of such a TF- and RNAPII-centric model of cooperation between regulatory elements is that differences in TF binding potential or responsiveness provide an explanation for the observed enhancer or co-activator incompatibility between particular regulatory elements^{11,125}. Furthermore, competition between regulatory elements when recruiting the transcriptional machinery may explain why certain elements have a stronger enhancer potential and weaker promoter potential, or vice versa, but it remains to be investigated why certain TFs allow simultaneous strong enhancer and promoter activities.

As we discussed above, given the high evolutionary turnover observed for most regulatory elements with enhancer activity, it is unlikely that enhancer potential as a generic ability is selected for in most cases, even though there clearly are cases of such elements that are

highly conserved^{126,127}. Rather, the results discussed in this Review point towards a model in which enhancer potential (and intrinsically low promoter potential) is directly encoded by binding sites for TFs and other components of the transcriptional machinery. High promoter potential requires additional constraints and is acquired on top of this basal enhancer potential, which may or may not lead to reduced enhancer potential. We note that in such a model, enhancer activity and low promoter activity is the ground state of activating regulatory elements (FIG. 3a). Transposable elements have lately been recognized as a prominent source of regulatory innovation¹²⁸, and their intrinsic enhancer potential and weak promoter activity may serve as a basis for further changes. In this way, transcribed regulatory elements with enhancer activity may be repurposed into having stronger promoter activity and producing longer and more highly abundant RNAs through changes in NDR-associated and downstream DNA sequences^{14,129,130}.

Conclusions, perspectives and challenges ahead

Although transcriptional regulation is one of the most research-intensive subject areas in biology, only now we can begin to interpret the underlying logic of regulatory elements systematically. This opportunity is enabled by the development and application of high-throughput methods for in vitro and in vivo assessments of regulatory activities. We here list some of the challenges and opportunities ahead in this endeavour.

Much research on transcriptional regulation in recent years has been descriptive — for example, focusing on finding genomic patterns associated with enhancer activity and using these patterns genome-wide to predict ‘maps’ of candidate active enhancers for cells or tissues of interest. While this research is laudable and highly useful, there are clear risks to such approaches, outlined below. Importantly, the number of well-characterized regulatory elements with strong enhancer potential in vertebrates is small, and these may not be the most representative regulatory elements with enhancer potential. Thus, the patterns of features (chromatin modifications, eRNA production etc.) used to predict regulatory elements with enhancer capabilities genome-wide are based on a very small base of training cases. This has led to a situation in which candidate enhancers are predicted by one type or set of data (for example, ChIP-seq) and are then validated by other predictions (for example, those based on divergent transcription, or vice versa). This carries clear risks of ascertainment bias, overfitting, oversimplification and ‘validation creep’¹³¹.

Developments in MPRA and CRISPR-based screening methods of the promoter and enhancer potential of a large number of genomic regions, either as plasmids or within genomes, hold great promise for enlarging our understanding of the features underlying regulatory elements with clear enhancer or promoter activity, and of those covering the spectrum of both activities, with much less ascertainment bias. Combining genomics or transcriptomics approaches with MPRA and/or CRISPR-based screening is likely to open new avenues for building predictive and quantitative models for enhancer and promoter activity.

Ascertainment bias

Drawing general conclusions based on biased sampling of non-representative examples.

Validation creep

Treating predictions as validated entities, in either the same study or subsequent studies. This type of bias is related to ascertainment bias and overfitting.

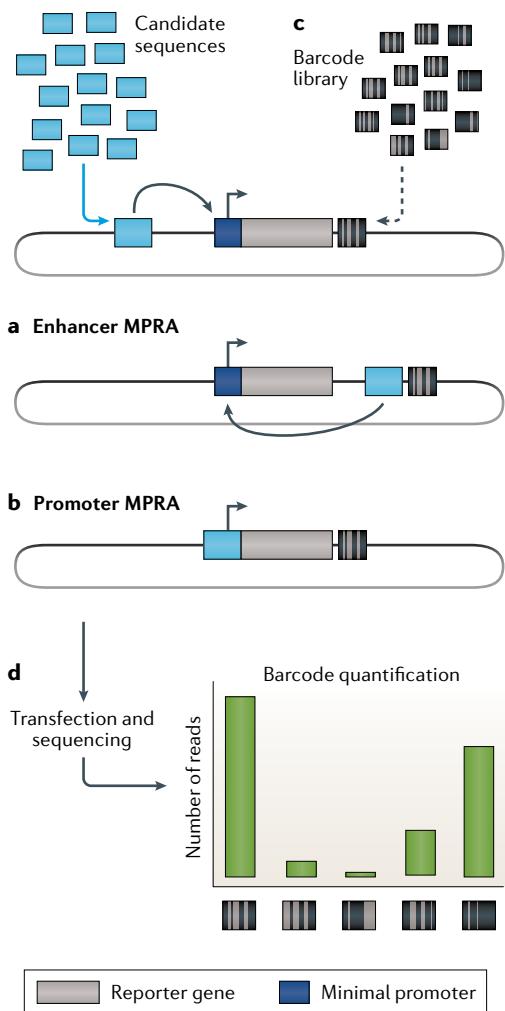
Box 5 | Massively parallel reporter assays to measure enhancer and promoter potential

Massively parallel reporter assays (MPRAs) have adopted the classical vector design of reporter assays, in which candidate regulatory sequences are placed in reporter gene vectors (BOX 1). In the case of testing enhancer potential, the candidate sequence is placed upstream or downstream of a specified minimal promoter (see the figure, part a), and promoter testing is done by placing the candidate region directly in front of the reporter (see the figure, part b), which can be done with or without a fixed enhancer element.

Upscaling from classical reporter assays, which test one sequence at a time, is achieved through high-throughput sequencing and quantification of barcodes (see the figure, parts c and d). These barcodes are typically integrated in the 5' or 3' untranslated region (UTR) of the reporter gene and are uniquely associated with a particular candidate regulatory sequence out of the thousands tested. Thus, the RNA level of the reporter is used as the readout, rather than the amount of produced reporter protein. Although several different MPRA strategies have been developed^{82,104,105,158,159}, including versions in which the enhancer candidate itself serves as a barcode^{18,114,160}, many approaches rely on episomal plasmids and are limited to testing short sequences (<200 bp). Furthermore, the MPRA readout can be affected by the nature of additional regulatory elements beyond those being tested — such as the minimal promoter tested in combination with the candidate enhancer sequences, or the fixed enhancer that is often tested in combination with candidate promoter sequences^{125,161} — which indicates restricted compatibilities between some enhancers and core promoters. Despite this, most enhancer MPRA studies test all candidate enhancer sequences with the same minimal promoter. Lastly, plasmid-based MPRAs are limited, in the sense that they cannot determine the contextual, chromatin-aware potential of regulatory elements. To this end, lentiviral MPRA¹⁵⁹ have been developed that test large numbers of candidate regulatory elements in their native chromatin contexts. This can be viewed as a hybrid between the *in vivo* approaches in BOX 4 and the *in vitro* approaches described here.

In general, we lack MPRA experiments that measure both the enhancer and promoter potentials of the same set of sequences. This will be necessary in order to model and understand the features that determine the respective types of activity. The Nguyen et al.¹⁰¹ study reviewed above is one of the few examples employing this strategy, but it only considered one cell type and a limited number of candidate sequences. Another important limitation of many MPRA studies is that the sequences tested for enhancer activity are fused to the same minimal core promoter. Testing many core promoters is important, because the sequence-encoded enhancer potential of one regulatory element is likely to be dependent on which promoters it can enhance. Thus, an ideal experiment would randomly fuse all proposed regulatory elements, testing the enhancer activity of each one by the promoter activity of another, in all combinations.

It will also be important to find ways to bridge the MPRA methods and CRISPR-based approaches, so that the enhancer and promoter potential inherent in regulatory elements measured by MPRA can be complemented with effects from their surrounding regulatory elements and chromatin contexts, or vice versa, thus capitalizing on the advantages of each method. It is unlikely



that any of these challenges can be addressed by a single research group, but it will require substantial collaborations between research groups focusing on screening methods (MPRA and/or *in vivo* genome-editing techniques), genomics or transcriptomics approaches, and statistical modelling. In other words, we believe this is a key area for cross-pollination of the machine-learning, biophysics, genomics and bioinformatics fields.

Some key outstanding questions to be answered are as follows. First, why can certain TFs drive promoter but not enhancer activity, and why do only some TFs have additive effects? Second, what TFs (or co-activators) and what combinations of binding sites drive promoter and/or enhancer activity in a given cell type? Third, enhancer activity seems to be much more diversely encoded by sequence, suggesting that there are many different ways in which a regulatory element can attain enhancer activity, or that many subtle signals cooperate. But what are those different ways in a given system, and what is the interaction between the signals (if any)? Fourth, what are the mechanisms underlying strong gene promoters that also have strong enhancer activity? Finally, what is the logic behind the compatibilities of regulatory elements (that is, a regulatory element may

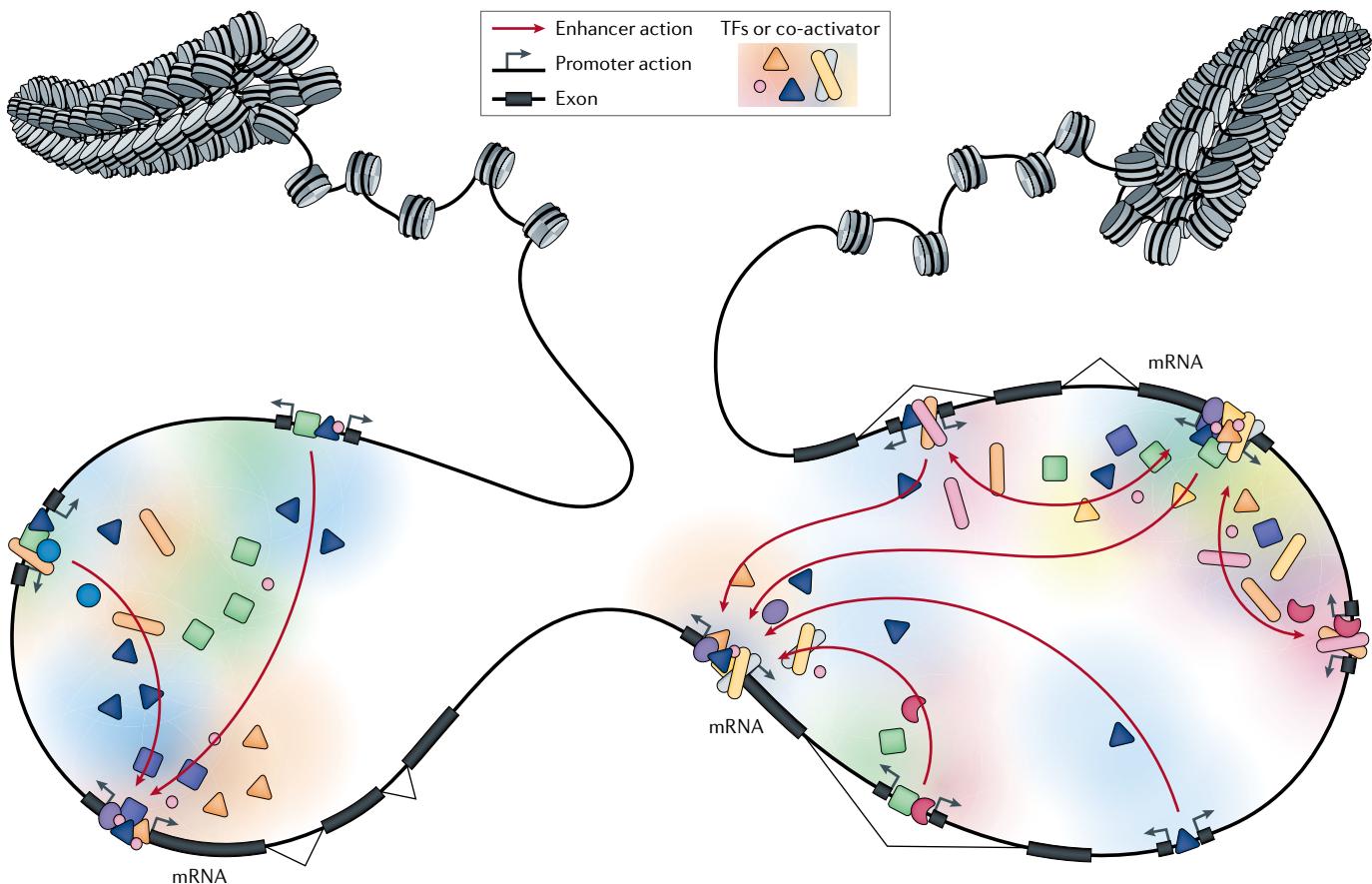


Fig. 4 | A TF and RNAPII-centric cooperative model of transcriptional regulation. In this conceptual model, regulatory elements, with varying degrees of enhancer and promoter activities, work together to increase or maintain the local concentrations of transcription factors (TFs), RNA polymerase II (RNAPII), general transcription factors (GTFs) and co-activators, thereby increasing the probability of the recruitment of TFs or co-activators and RNAPII to target gene transcription start sites. Red arrows indicate enhancer activity, and right-angled dark-grey arrows indicate promoter activity. Thick black lines indicate exons, which may be joined by splicing, as indicated. Coloured blocks indicate TFs or co-activators, where coloured gradients represent their local concentration or probability. RNAPII and GTFs are not shown, but are expected to follow similar patterns. Two regulatory topologies (left and right) with different regulatory complexities and characteristics are illustrated. The left topology features two regions with low promoter and high enhancer

activity, enhancing the transcription of a region with high promoter activity, producing a promoter-upstream transcript (PROMPT)-mRNA pair. The right topology shows two regulatory elements with high promoter activity, producing PROMPT-mRNA pairs. One has dual activity, carrying high enhancer potential, enhancing the transcription of the other PROMPT-mRNA pair. The four remaining regulatory elements have low promoter activity and high enhancer activity. Note that two of these are located within a gene intron; this is common, as ~50% of the human genome is intronic. Transcription of two of these regions is enhanced by the regulatory element with dual potential (to our knowledge, this phenomenon has not been conclusively demonstrated, but it follows from the models proposed). The right topology also demonstrates the notion that regulatory elements with enhancer potential may be able to enhance the transcription of some, but not all, other regulatory elements. Figure adapted from REF.¹⁰, Springer Nature Limited.

enhance the transcription of some, but not all, regulatory elements with promoter activity?

Although we have primarily focused our Review on the TF- and sequence-associated features determining the activating (enhancer and promoter) functions of regulatory elements, the transcriptional activity of genes also depends on additional regulatory events and determinants, including chromatin remodelling, the activity of proximal silencers and insulators, the binding of repressors, the folding and dynamics of the surrounding three-dimensional chromatin topology, and the responsiveness of regulatory elements to developmental or external cues. Many of these factors are likely to be related; for example, CpG-rich regulatory elements are less dependent on chromatin-remodelling events for their regulatory

activity than are AT-rich regulatory elements, which in turn may also affect their ability to be dynamically regulated¹³². Nonetheless, it will be important to analyse and integrate large-scale quantitative measures of such factors with those of enhancer and promoter potential. In the general model we proposed above, only enhancer and promoter capabilities were considered, but the same model could be extended so as to allow for additional functionalities, such as local or distal repressive activity. For instance, a regulatory element with enhancer activity for a gene promoter might act as a silencer or insulator for other promoters in certain contexts¹³³, or to regulate replication timing during mitosis¹³⁴.

Despite the large challenges ahead, we are now in a highly promising and exciting time for research on

transcriptional regulation. Our current knowledge, together with the approaches outlined above, holds the promise of allowing us to finally understand much of the logic underlying transcriptional regulation. This will have profound importance in biological research, medicine and biotechnology, furthering our understanding of sequence variation and evolution as well as regulatory robustness, and enabling the design of regulatory

elements with custom strength, inducibility and robustness for applications in biotechnology and medicine. In other words, we may be in a transition time when the field of transcriptional regulation is moving from describing and classifying regulatory phenomena towards building quantitative biological models of gene regulation.

Published online 11 October 2019

1. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
2. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
3. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014). **This is the largest study to date defining active enhancers from divergent TSSs of eRNAs. It demonstrates exosome-mediated decay of eRNAs and describes transcribed enhancers as being better validated in vitro reporter assays than enhancers predicted from histone modifications.**
4. Chatterjee, S. & Ahituv, N. Gene regulatory elements, major drivers of human disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 45–63 (2017).
5. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).
6. Miguel-Escalada, I., Pasquali, L. & Ferrer, J. Transcriptional enhancers: functional insights and role in human disease. *Curr. Opin. Genet. Dev.* **33**, 71–76 (2015).
7. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
8. Shlyueva, D., Stampfli, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
9. Beagrie, R. A. & Pombo, A. Gene activation by metazoan enhancers: diverse mechanisms stimulate distinct steps of transcription. *BioEssays* **38**, 881–893 (2016).
10. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012). **This study demonstrates a unified divergent architecture of candidate enhancers and promoters and observes a correlation between H3K4me3 and transcription levels.**
11. Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014). **This study demonstrates a unified divergent architecture of candidate enhancers and promoters and observes a correlation between H3K4me3 and transcription levels.**
12. Andersson, R. et al. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.* **5**, 5336 (2014).
13. Scruggs, B. S. et al. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* **58**, 1101–1112 (2015).
14. Chen, Y. et al. Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.* **48**, 984–994 (2016).
15. Koch, F. et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963 (2011).
16. Andersson, R. et al. Human gene promoters are intrinsically bidirectional. *Mol. Cell* **60**, 346–347 (2015).
17. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).
18. Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49**, 1073–1081 (2017).
19. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
20. Engeitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016). **Diao (2017), Dao (2017) and Engeitz (2016) have observed, using in vivo genome editing and in vitro MPRA, gene promoters with high enhancer strengths.**
21. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLOS Biol.* **8**, e1000384 (2010).
22. Kim, T.-K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010). **De Santa (2010) and Kim (2010) are the first studies to observe bidirectional transcription of eRNAs from candidate enhancers at large scale.**
23. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
24. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479 (2003).
25. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 40–51 (2012).
26. Breathnach, R. & Chambon, P. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349–383 (1981).
27. Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
28. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
29. Dynlach, B. D., Hoey, T. & Tjian, R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**, 563–576 (1991).
30. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
31. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
32. Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. *J. Biol. Chem.* **276**, 13505–13508 (2001).
33. Malik, S. & Roeder, R. G. Dynamic regulation of pol II transcription by the mammalian mediator complex. *Trends Biochem. Sci.* **30**, 256–263 (2005).
34. Koutelou, E., Hirsch, C. L. & Dent, S. Y. R. Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* **22**, 374–382 (2010).
35. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
36. Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5**, 276–287 (2004).
37. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
38. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
39. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. *Mol. Cell* **74**, 1110–1122 (2019).
40. Moreau, P. et al. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**, 6047–6068 (1981). **Moreau (1981) and Banerji (1981) are the first studies to discover enhancers through the identification of a 72-bp repeat sequence that could enhance gene transcription in SV40. They further show that such enhancers can act both upstream and downstream of, and at various distances from, a reporter gene.**
41. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981). **Moreau (1981) and Banerji (1981) are the first studies to discover enhancers through the identification of a 72-bp repeat sequence that could enhance gene transcription in SV40. They further show that such enhancers can act both upstream and downstream of, and at various distances from, a reporter gene.**
42. Benoist, C. & Chambon, P. In vivo sequence requirements of the SV40 early promoter region. *Nature* **290**, 304–310 (1981).
43. Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729–740 (1983).
44. Neuberger, M. S. Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. *EMBO J.* **2**, 1373–1378 (1983).
45. Schaffner, W. Enhancers, enhancers — from their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).
46. Furlong, E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
47. Catarino, R. R. & Stark, A. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* **32**, 202–223 (2018).
48. Forrest, A. R. R. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
49. Consortium, R. E. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
50. Bernstein, B. E. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
51. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
52. Zavolan, M., van Nimwegen, E. & Gaasterland, T. Splice variation in mouse full-length cDNAs identified by mapping to the mouse genome. *Genome Res.* **12**, 1377–1385 (2002).
53. Landry, J.-R., Mager, D. L. & Wilhelm, B. T. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet.* **19**, 640–648 (2003).
54. Valen, E. et al. Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.* **19**, 255–265 (2009).
55. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
56. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
57. Seila, A. C. et al. Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
58. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008). **Core (2008), Seila (2008) and Preker (2008) are the first studies to independently observe divergent transcription from mammalian gene promoters.**
59. Jensen, T. H., Jacquier, A. & Libri, D. Dealing with pervasive transcription. *Mol. Cell* **52**, 473–484 (2013).
60. Seila, A. C., Core, L. J., Lis, J. T. & Sharp, P. A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564 (2009).
61. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
62. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
63. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).
64. Bernstein, B. E. et al. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. USA* **99**, 8695–8700 (2002).

65. Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
66. Schübeler, D. et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**, 1263–1271 (2004).
67. Roh, T., Ngau, W. C., Cui, K., Landsman, D. & Zhao, K. High-resolution genome-wide mapping of histone modifications. *Nat. Biotechnol.* **22**, 1013–1016 (2004).
68. Song, L. et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757–1767 (2011).
69. Kim, T. H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
70. Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
71. Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
72. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
73. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
74. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
75. Robertson, A. G. et al. Genome-wide relationship between histone H3 lysine 4 mono- and trimethylation and transcription factor binding. *Genome Res.* **18**, 1906–1917 (2008). **Heintzman (2007) and Robertson (2008) identify a signature of high H3K4me1 and low H3K4me3 at candidate enhancers that could be used to discern enhancers from promoters and predict their locations genome-wide.**
76. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
77. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).
78. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
79. Melgar, M. F., Collins, F. S. & Sethupathy, P. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol.* **12**, R113 (2011).
80. Wu, H. et al. Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLOS Genet.* **10**, e1004610 (2014).
81. Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**, 426–433 (2015).
82. van Arensbergen, J. et al. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* **35**, 145–153 (2017).
83. Landolin, J. M. et al. Sequence features that drive human promoter function and tissue specificity. *Genome Res.* **20**, 890–898 (2010).
84. Barski, A. et al. High-resolution profiling of histone methylation in the human genome. *Cell* **129**, 823–837 (2007).
85. Soares, L. M. et al. Determinants of histone H3K4 methylation patterns. *Mol. Cell* **68**, 773–785.e6 (2017).
86. Pekowska, A. et al. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* **30**, 4198–4210 (2011).
87. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
88. Rennie, S. et al. Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities. *Nucleic Acids Res.* **46**, 5455–5469 (2018).
89. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
90. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
91. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
92. Xu, Z., Wei, G., Chepelev, I., Zhao, K. & Felsenfeld, G. Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. *Nat. Struct. Mol. Biol.* **18**, 372–378 (2011).
93. Andersson, R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* **37**, 314–323 (2015).
94. Mikhaylichenko, O. et al. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).
95. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
96. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR–Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
97. Sanjana, N. E. et al. High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
98. Klann, T. S. et al. CRISPR–Cas9 genome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561–568 (2017).
99. Carleton, J. B., Berrett, K. C. & Gertz, J. Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor α -bound enhancers. *Cell Syst.* **5**, 333–344.e5 (2017).
100. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
101. Nguyen, T. A. et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res.* **26**, 1023–1033 (2016). **This is the first study to systematically compare the promoter and enhancer potentials of a large number of candidate sequences using MPRA.**
102. Weingarten-Gabbay, S. et al. Systematic interrogation of human promoters. *Genome Res.* **29**, 171–183 (2019).
103. Mattioli, K. et al. High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Res.* **29**, 344–355 (2019).
104. Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* **30**, 265–270 (2012).
105. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
106. Kheradpour, P. et al. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* **23**, 800–811 (2013).
107. Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl Acad. Sci. USA* **109**, 19498–19503 (2012).
108. Smith, R. P. et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* **45**, 1021–1028 (2013).
109. Grossman, S. R. et al. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc. Natl Acad. Sci. USA* **114**, E1291–E1300 (2017).
110. Shen, S. Q. et al. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* **26**, 238–255 (2016).
111. Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
112. Kwasnieski, J. C., Fiore, C., Chaudhari, H. G. & Cohen, B. A. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* **24**, 1595–1602 (2014).
113. Chaudhari, H. G. & Cohen, B. A. Local sequence features that influence AP-1-regulatory activity. *Genome Res.* **28**, 171–181 (2018).
114. Muerdter, F. et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* **15**, 141–149 (2018).
115. Fenouil, R. et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
116. Biddie, S. C. et al. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43**, 145–155 (2011).
117. Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).
118. Nakayama, R. T. et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat. Genet.* **49**, 1613–1623 (2017).
119. Colbran, L. L., Chen, L. & Capra, J. A. Sequence characteristics distinguish transcribed enhancers from promoters and predict their breadth of activity. *Genetics* **211**, 1205–1217 (2019).
120. Farley, E. K. et al. Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).
121. Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. & Levine, M. S. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl Acad. Sci. USA* **113**, 6508–6513 (2016).
122. Cheng, Y. et al. Principles of regulatory information conservation between mouse and human. *Nature* **515**, 371–375 (2014).
123. Kim, T.-K. & Shiekhattar, R. Architectural and functional commonalities between enhancers and promoters. *Cell* **162**, 948–959 (2015).
124. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* **14**, 103–105 (2007).
125. Zabidi, M. A. et al. Enhancer–core–promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
126. Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499.e15 (2018).
127. Pennacchio, L. A. et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
128. Kidwell, M. & Lisch, D. Transposable elements and host genome evolution. *Trends Ecol. Evol.* **15**, 95–99 (2000).
129. Wu, X. & Sharp, P. A. Divergent transcription: a driving force for new gene origination? *Cell* **155**, 990–996 (2013).
130. Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M. & Kaessmann, H. Repurposing of promoters and enhancers during mammalian evolution. *Nat. Commun.* **9**, 4066 (2018).
131. Halfon, M. S. Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases. *Trends Genet.* **35**, 93–103 (2019).
132. Ramirez-Carrozzi, V. R. et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).
133. Kolovos, P., Knoch, T. A., Grosfeld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epinetics Chromatin* **5**, 1 (2012).
134. Sima, J. et al. Identifying cis elements for spatiotemporal control of mammalian DNA replication. *Cell* **176**, 816–830.e18 (2019).
135. Kawai, J. et al. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685–690 (2001).
136. Takahashi, H., Lassmann, T., Murata, M. & Carninci, P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* **7**, 542–561 (2012).
137. Yamashita, R. et al. Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. *Genome Res.* **21**, 775–789 (2011).
138. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
139. Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T. & Meyer, B. J. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife* **2**, e00808 (2013).
140. Lam, M. T. Y. et al. Rev-Erb α repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
141. Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).

142. Nojima, T. et al. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
143. Weber, C. M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).
144. Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
145. Berman, B. P. et al. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl Acad. Sci. USA* **99**, 757–762 (2002).
146. May, D. et al. Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **44**, 89–93 (2011).
147. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
148. Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
149. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
150. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).
151. Pradeepa, M. M. et al. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* **48**, 681–686 (2016).
152. Rajagopal, N. et al. Distinct and predictive histone lysine acetylation patterns at promoters, enhancers, and gene bodies. *G3* **4**, 2051–2063 (2014).
153. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
154. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
155. Fulco, C. P. et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
156. Matharu, N. et al. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science* **363**, eaau0629 (2019).
157. Lopes, R., Korkmaz, G. & Agami, R. Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**, 597–604 (2016).
158. Patwardhan, R. P. et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
159. Inoue, F. et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* **27**, 38–52 (2017).
160. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
161. Arnold, C. D. et al. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol.* **35**, 136–144 (2017).

Acknowledgements

R.A. was supported by the Independent Research Fund Denmark (6108-00038B), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (StG no. 638173), and the Novo Nordisk Foundation (NNF18OC0052570). A.S. was supported by the ERC under the European Union's Horizon 2020 research and innovation programme (MSCA ITN pHioniC), the Lundbeck Foundation, the Danish Cancer Society, the Danish Council for Independent Research and Innovation Fund Denmark. We thank L. van Duin, M. Wu and A. Thieffry for critical comments on the manuscript.

Author contributions

Both authors contributed to all aspects of the article.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Genetics thanks K. Adelman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.