REGULATORY ELEMENTS

Transcriptional enhancers: from properties to genome-wide predictions

Daria Shlyueva, Gerald Stampfel* and Alexander Stark*

Abstract | Cellular development, morphology and function are governed by precise patterns of gene expression. These are established by the coordinated action of genomic regulatory elements known as enhancers or *cis*-regulatory modules. More than 30 years after the initial discovery of enhancers, many of their properties have been elucidated; however, despite major efforts, we only have an incomplete picture of enhancers in animal genomes. In this Review, we discuss how properties of enhancer sequences and chromatin are used to predict enhancers in genome-wide studies. We also cover recently developed high-throughput methods that allow the direct testing and identification of enhancers on the basis of their activity. Finally, we discuss recent technological advances and current challenges in the field of regulatory genomics.

During animal development, a single cell — the fertilized egg — gives rise to a multitude of different cell types and organs. These acquire different morphologies and functions by expressing different sets of genes. The initial step in gene expression is the transcription of the genomic DNA of the gene into RNA by RNA polymerase II (Pol II). The genomic sequence in the immediate vicinity of the transcription start site (TSS), which is also known as the core promoter, is sufficient to assemble the Pol II machinery. However, transcription is often weak in the absence of regulatory DNA regions that are more distant from the TSS; these regions are called enhancers or *cis*-regulatory modules (CRMs).

More than 30 years ago, the first enhancer was identified as a 72-bp sequence of the SV40 virus genome, which could enhance the transcription of a reporter gene in HeLa cells by several hundred fold[1]. Soon after this, cellular enhancers were found in animal genomes[2]; since then, many enhancers have been described, and their biochemical and functional properties have been extensively studied (FIG. 1).

Enhancer sequences contain short DNA motifs that act as binding sites for sequence-specific transcription factors (FIG. 1a). These proteins recruit co-activators and co-repressors such that the combined regulatory cues of all bound factors determine the activity of the enhancer. In addition, enhancer activity has been shown to correlate with certain properties of chromatin (FIG. 2): active enhancers are typically devoid of nucleosomes — the structural units of chromatin — such that the DNA is accessible and can be bound by transcription factors (FIG. 2a). Furthermore, nucleosomes in the vicinity of active enhancers typically contain histones with characteristic post-translational modifications, such as histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac), at their amino termini (FIG. 2b).

A functional hallmark of enhancers that has been repeatedly demonstrated is that they act seemingly independently of the distance and orientation to their target genes, and can function at large distances of several hundred kilobases or even megabases by looping[3] (FIG. 1b,c). In addition, enhancers maintain their functions independently of the sequence context (for example, when placed into heterologous reporter constructs). Finally, enhancers are modular, and they contribute additively and partly redundantly to the overall expression pattern of their target genes. This property can be recapitulated in reporter assays, as combining multiple sequences in an *in vivo* assay often results in patterns of expression that reflect their combined activity (reviewed in REF. 4).

Understanding enhancers is currently an area of great interest, as there is an increasing appreciation of their importance not only in developmental gene expression but also in evolution and disease[5–7]. However, our knowledge of these elements has remained surprisingly incomplete. The vast majority of enhancers in any animal genome and their spatiotemporal activities are unknown[8] as suggested by the small number of gene

Research Institute of Molecular Pathology (IMP), 1030 Vienna, Austria. Correspondence to A.S. e-mail: stark@starklab.org
**These authors contributed equally to this work.*
doi:10.1038/nrg3682
Published online 11 March 2014

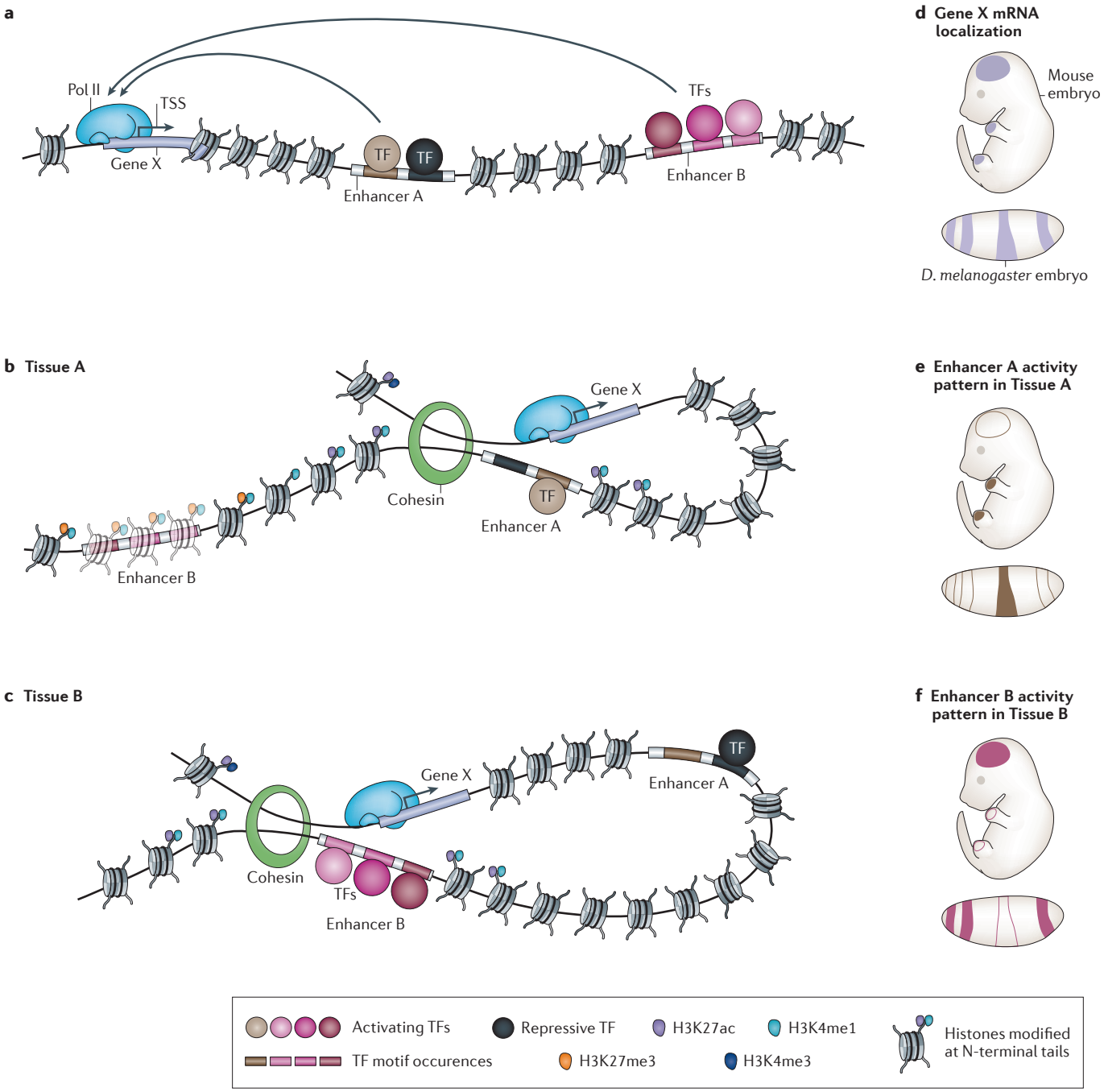


Figure 1 | **Enhancers and their features.** **a** | Enhancers are distinct genomic regions (or the DNA sequences thereof) that contain binding site sequences for transcription factors (TFs) and that can upregulate (that is, enhance) the transcription of a target gene from its transcription start site (TSS). Along the linear genomic DNA sequence, enhancers can be located at any distance from their target genes, which makes their identification challenging. **b**,**c** | In a given tissue, active enhancers (Enhancer A in part **b** or Enhancer B in part **c**) are bound by activating TFs and are brought into proximity of their respective target promoters by looping, which is thought to be mediated by cohesin and other protein complexes. Moreover, active and inactive gene regulatory elements are marked by various biochemical features: active promoters and enhancers are characterized by a depletion of nucleosomes, which is the structural unit of eukaryotic chromatin. Nucleosomes that flank active enhancers show specific histone modifications, for example, histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac). Inactive enhancers might be silenced by different mechanisms, such as by the Polycomb protein-associated repressive H3K27me3 mark (part **b**) or by repressive TF binding (part **c**). **d–f** | Complex patterns of gene expression result from the additive action of different enhancers with cell-type- or tissue-specific activities. The schematic mouse and *Drosophila melanogaster* embryos are drawn after REFS 7,43. Pol II, RNA polymerase II.

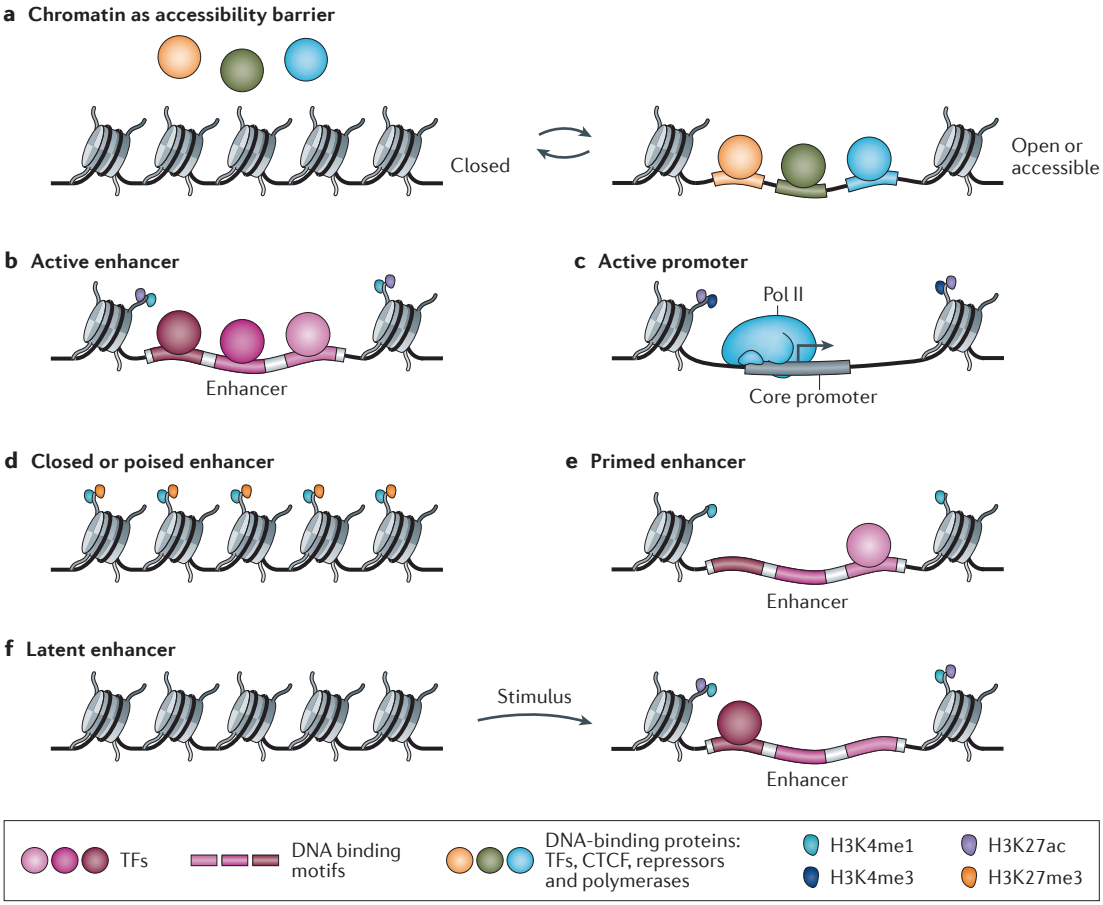


Figure 2 | **Chromatin accessibility and histone marks at regulatory elements.** **a** | Chromatin is shown as a 'gatekeeper' for transcription factor (TF) binding and enhancer activity. Densely positioned nucleosomes can restrict access for transcription factors (both activators and repressors), CCCTC-binding factor (CTCF), RNA polymerase II (Pol II) and other proteins. Accessible (that is, nucleosome-free) regions can be bound by these proteins, which define and mediate the identity of a region (for example, active enhancers, repressors or core promoters). The transition from 'open' to 'closed' chromatin, and vice versa, is determined by regulatory proteins, including pioneer transcription factors. Insulator proteins (for example, CTCF) and other architectural proteins also bind to open regions, and they make up a substantial proportion of sites that are accessible across multiple cell types (for example, 10%[73]). **b** | Histones that flank active enhancers are often marked by histone H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated at lysine 4 (H3K4me1). **c** | Active promoters (depicted as Pol II bound) are flanked by nucleosomes with H3K27ac and H3K4me3 modifications. **d** | Some closed or poised enhancer regions can bear the active H3K4me1 and the repressive Polycomb protein-associated H3K27me3 marks. Interestingly, some of these elements can function as active enhancers in episomal assays or when integrated into the genome of a differentiated cell type[72]. **e** | Enhancers that are not yet active but that are primed for activation either at a later developmental time point or in response to external stimuli can be pre-marked by H3K4me1. **f** | Latent enhancers are located in closed chromatin and are not pre-marked by known histone modifications. However, in the presence of external stimuli the DNA becomes accessible, and flanking nucleosomes acquire H3K4me1 and H3K27ac marks.

expression patterns that have been linked to specific enhancers[9,10], the many disease-associated single-nucleotide polymorphisms in non-coding regions of unknown function[11] and the large number of genomic regions with enhancer-like chromatin features[12,13]. Moreover, enhancers and their activity states cannot be reliably predicted from their DNA sequences or from chromatin features, nor can the important parts of the sequence of an enhancer be easily identified, which precludes the prediction of functional consequences of mutations in these sequences.

In this Review, we describe how the known properties of enhancers are used by a range of methods to predict enhancers. Only in the past few years has the development of high-throughput sequencing technologies enabled genome-wide studies of enhancer-associated chromatin features and enhancer activities. Furthermore, new methods to synthesize long DNA fragments or to manipulate genomes *in vivo* allow unprecedented functional tests to be carried out. We discuss the different methods and pay particular attention to the specific enhancer characteristics that form

Box 1 | **How to find motifs**

Transcription factor binding motifs are degenerate sequence patterns (for example, consensus sequences or position weight matrices[153]) that summarize the DNA binding preference of transcription factors[153]. Several computational and experimental approaches have been developed to determine motifs for specific transcription factors or to identify putative regulatory motifs for unknown factors (reviewed in REFS 8,153). After these motifs have been defined, they can be used to predict functional motif instances or regulatory elements in long DNA sequences, including entire genomes (reviewed in REFS 8,24).

Computational approaches

Computational methods that derive motifs from DNA sequences typically make use of either sequence over-representation (that is, enrichment) in putative regulatory regions[153,154] or high average evolutionary sequence conservation of all motif occurrences[155]. Proximal promoter regions upstream of co-expressed genes have frequently been used as surrogates for regulatory regions [154]. For example, motifs for the transcription factor Zelda (also known as Vielfaltig) were initially discovered as conserved and over-represented sequences in the regions upstream of pre-cellular blastoderm genes in *Drosophila melanogaster*[156,157]. Approaches for finding over-represented motifs can be applied to any set of sequences that share some functional or biochemical similarity, for example, regions that act as enhancers in the same tissue and regions bound by the same transcription factor *in vivo*[35,36,120]. Computational algorithms for the discovery of motifs by over-representation include k-mer enrichment, Gibbs sampling and expectation minimization (reviewed in REF. 153).

Comparative genomics approaches are based on the selective conservation of functionally important sequence elements between closely related species, which proved to be a powerful tool for discovering motifs even in the absence of functional data sets[155] (reviewed in REF. 24).

Experimental approaches

The DNA binding specificities of known transcription factors are intensely studied using *in vivo* methods such as chromatin immunoprecipitation followed by deep sequencing (ChIP–seq)[35,36], as well as *in vitro* methods such as protein-binding microarrays (PBMs)[158], systematic evolution of ligands by exponential enrichment (SELEX)[159], yeast-one-hybrid (Y1H)[160] and bacterial-one-hybrid (B1H)[161]. These methods all assay a single transcription factor for binding to a complex library that usually contains thousands of different DNA fragments. PBMs are highly dense arrays that are spotted with thousands of DNA molecules, which can be probed simultaneously for transcription factor binding; SELEX enriches transcription factor-bound sequences by repeated rounds of selection and sequencing; and Y1H and B1H assays convert transcription factor binding into a survival or growth advantage such that complex fragment libraries can be screened and transcription factor-bound sequences can be retrieved from transformant colonies. The binding preferences found by different studies are then systematically collected in motif databases such as TRANSFAC[162], JASPAR[163] and UniPROBE[164].

the basis of each method, as well as advantages and limitations of each approach. We also highlight how the combination of complementary methods that are based on different enhancer properties increases the ability to predict functional enhancers and how this can lead to novel and interesting insights into enhancer biology.

Predictions using motifs and conservation

The genetic information of transcriptional enhancers is read by a large range of protein factors. These transcription factors specifically recognize and bind to short sequences that are typically 6–10 bp long, and they often allow different nucleotides at some of the binding site positions (that is, at 'degenerate' positions). The binding preference of a transcription factor can be summarized as a consensus sequence known as a transcription factor binding motif[14,15], which allows the computational scanning of long DNA sequences, including entire genomes, for transcription factor motif matches. Indeed, known enhancers contain such motif matches — often clusters thereof — and their mutation or the ablation of the respective transcription factors can abolish enhancer activities (reviewed in REF. 4).

The apparent conceptual simplicity of individual transcription factors binding to their preferred binding sequences to regulate gene expression made it an attractive goal to predict enhancers in entire genome sequences through the computational matching of transcription factor binding motifs. Available approaches either identify genomic regions that are enriched for transcription factor motif matches[16] or seek individual matches that are highly conserved across species[17,18]. Some methods also use both enrichment and conservation, and/or identify only regions in which motif matches occur in specific combinations or in a particular order or arrangement[19–22] (reviewed in REFS 8,23,24).

In general, these methods rely either on known transcription factor binding motifs or on motifs that have been discovered *de novo* from training data. To characterize the DNA binding preferences of transcription factors, numerous methods have recently been developed (BOX 1) and applied, for example, to about half of all annotated transcription factors in *Drosophila melanogaster* and humans[25–27].

However, the relationship between transcription factor motif matches, transcription factor binding and enhancer activity is far from simple and is not fully understood. Short motifs frequently match to genomic or even random DNA sequences (for example, each 6-bp long motif would be expected to occur every 4^6 bp = 4,096 bp), and only a small proportion of all matches in a genome are typically bound by the corresponding transcription factor *in vivo*[28]. Moreover, transcription factor binding can be context specific and depend on other proteins[28,29].

Similarly, even conserved motif matches are not necessarily bound by a transcription factor in a particular cell type of interest, nor are they necessarily functioning as active enhancers[17,28]. Finally, compensatory turnover of transcription factor binding sequences in conserved enhancers and the emergence of novel enhancers indicate that approaches that rely on sequence conservation might only identify a proportion of enhancers. For example, functional enhancers do not need to show any detectable sequence conservation[30] even if their functions are conserved (reviewed in REF. 31).

More recent computational methods use machine-learning approaches to identify characteristic DNA sequence features in experimentally determined enhancers and use these features to predict novel enhancers[32–34]. Such approaches can be highly successful when combined with additional data, such as conservation or the expression of flanking genes[32–34] (see REF. 24 for a comparison of methods and validation rates).

As the function of motifs is to recruit transcription factors to DNA, which in turn recruit cofactors, alternative approaches for finding enhancers assess the actual *in vivo* binding of these regulatory proteins in the genome.

Transcription factor binding motif
(Also known as transcription factor sequence motif, transcription factor motif and transcription factor binding site motif.) A degenerate short (6–10-bp) DNA sequence pattern that summarizes the DNA sequence binding preference of a transcription factor. These motifs are usually represented either as consensus sequences in IUPAC code or by position weight matrices.

Predictions from regulator binding

The strict dependence of enhancer activity on the binding of transcription factors makes it attractive to use genome-wide methods that determine *in vivo* transcription factor binding sites for the prediction of active enhancers. The most frequently used methods are chromatin immunoprecipitation followed by deep sequencing (ChIP–seq)[35,36], and its variants ChIP-exo[37] and DNA adenine methyltransferase (Dam) identification (DamID)[38].

In ChIP–seq, transcription factors are covalently linked to their *in vivo* binding sites by chemical crosslinking, the chromatin is then sheared, and DNA fragments corresponding to the binding sites are co-precipitated with transcription factor-specific antibodies and determined by sequencing and computational analyses (FIG. 3a). ChIP-exo includes an additional step of exonuclease digestion that trims the DNA fragments to allow the identification of transcription factor binding sites at high resolution[37]. In DamID, a DNA-binding protein of interest, such as a transcription factor, is fused to the bacterial enzyme Dam. Dam then specifically methylates adenines at GATC sequences in the vicinity of the binding sites of the fusion protein — a modification that does not naturally occur in eukaryotes. DamID can be used if antibodies are not available; it is able to detect transient or indirect interactions but has a comparatively low resolution. Below, we focus mainly on conventional ChIP–seq, as it is the most frequently used method for assaying protein–DNA interactions *in vivo*.

For a typical transcription factor, ChIP–seq predicts thousands of *in vivo* binding sites in animal genomes, predominantly in promoters, introns and intergenic regions[39]. ChIP–seq experiments have also revealed that transcription factors show dynamic patterns of binding throughout development, which suggests the existence of stage-specific and cell-type-specific regulatory targets[28,40].

ChIP–seq typically successfully recovers the direct regulatory target enhancers (that is, the functional binding sites) of a transcription factor; in other words, it has a low false-negative rate. For example, ChIP recovered essentially all previously known Twist-dependent mesodermal enhancers in *D. melanogaster* embryos[40,41]. However, the converse is not true: whereas transcription factor binding to the sites detected by ChIP–seq can usually be confirmed (for example, by ChIP followed by quantitative PCR (ChIP–qPCR)), many of them do not correspond to functional enhancers[42–44], which suggests that transcription factor binding can occur without affecting the transcription of any gene.

It is unclear why a major proportion of binding sites are non-functional, but there are several non-exclusive explanations. First, it is well established that enhancers are activated by combinations of transcription factors (reviewed in REF. 39) such that the binding of only one or a few transcription factors is possibly insufficient to activate transcription. In addition, transcription factors have a general affinity for DNA[45], such that they bind — potentially at lower levels or more transiently — to accessible DNA that does not match their motifs, even outside functional contexts. Formaldehyde crosslinking during ChIP can capture such transient and potentially non-functional interactions, even for non-DNA-binding proteins[46], especially if applied for an extended time[47]. Functionality might therefore be better captured by measuring the kinetics of transcription factor binding rather than by inferring it from ChIP alone[48]. Finally, transcription factors might bind to enhancers indirectly through interactions with other transcription factors, such that the respective *in vivo* binding sites, especially those bound by many different transcription factors, do not always reflect sequence-specific transcription factor binding or tissue-specific enhancer activity[43,49].

A related strategy based on ChIP–seq is the identification of enhancers by the *in vivo* binding sites of transcriptional cofactors, such as the histone acetyltransferase p300 (REF. 50). Cofactors typically do not bind to DNA directly but are recruited by transcription factors and carry out various biochemical activities, which ultimately lead to the activation or repression of transcription. Binding of the co-activator protein p300 has, for example, been used to predict enhancers in mice and humans[50–52], and 58–82% of the p300-binding sites have been reported to function as enhancers[30,50,53,54]. In addition, the systematic combination of cofactor binding profiles allowed the segmentation of entire genomes into regions of different chromatin states[55–57], some of which correspond to active regulatory regions.

In summary, the determination of the *in vivo* binding sites of transcription factors and cofactors is a powerful method for identifying their target enhancers and for providing insights into combinatorial regulation. However, for the identification of active enhancers, the prevalence of apparently non-functional or neutral binding constitutes an important shortcoming. This issue might be overcome by assessing additional features of active enhancers, such as chromatin accessibility and histone modifications.

Predictions using chromatin accessibility

Experiments in the 1970s found differences in the degree of compaction of eukaryotic DNA when comparing genomic regions containing expressed genes with transcriptionally silent regions[58,59]. We now know that active enhancers are depleted of nucleosomes (that is, they are 'open') such that the DNA is accessible (FIG. 2a). This stereotypical chromatin structure is used for the identification of regulatory regions by enzymatic cleavage of accessible DNA using DNase I or micrococcal nuclease (MNase) coupled to deep sequencing (DNase-seq[60] (FIG. 3b) and MNase-seq[61], respectively). High-resolution DNase-seq can even resolve small transcription factor footprints within larger nucleosome-free regions of DNA that are protected by transcription factor binding[62]. An alternative to DNase-seq — formaldehyde-assisted identification of regulatory elements followed by deep sequencing (FAIRE–seq)[63] — determines open regions by depleting histone-bound 'closed' DNA after chemical crosslinking with formaldehyde.

Transcription factor motif matches
(Also known as transcription factor motif instances, transcription factor motif occurrences and transcription factor binding sequences). Specific genomic sequences or positions that match transcription factor binding motifs and thus constitute potential transcription factor binding sites. These are also sometimes called transcription factor binding sites, although we prefer to reserve this term for experimentally determined ones.

Position weight matrices
(Also known as position-specific weight matrices or position-specific scoring matrices). Matrices that provide the frequencies at which individual nucleotides are found at the positions of the transcription factor binding motif.

Transcription factor binding sites
Genomic locations of transcription factor binding, typically *in vivo*. These sites can be determined experimentally (for example, using chromatin immunoprecipitation (ChIP)). ChIP experiments typically reveal that these binding sites and transcription factor motif matches often, but not always, coincide.

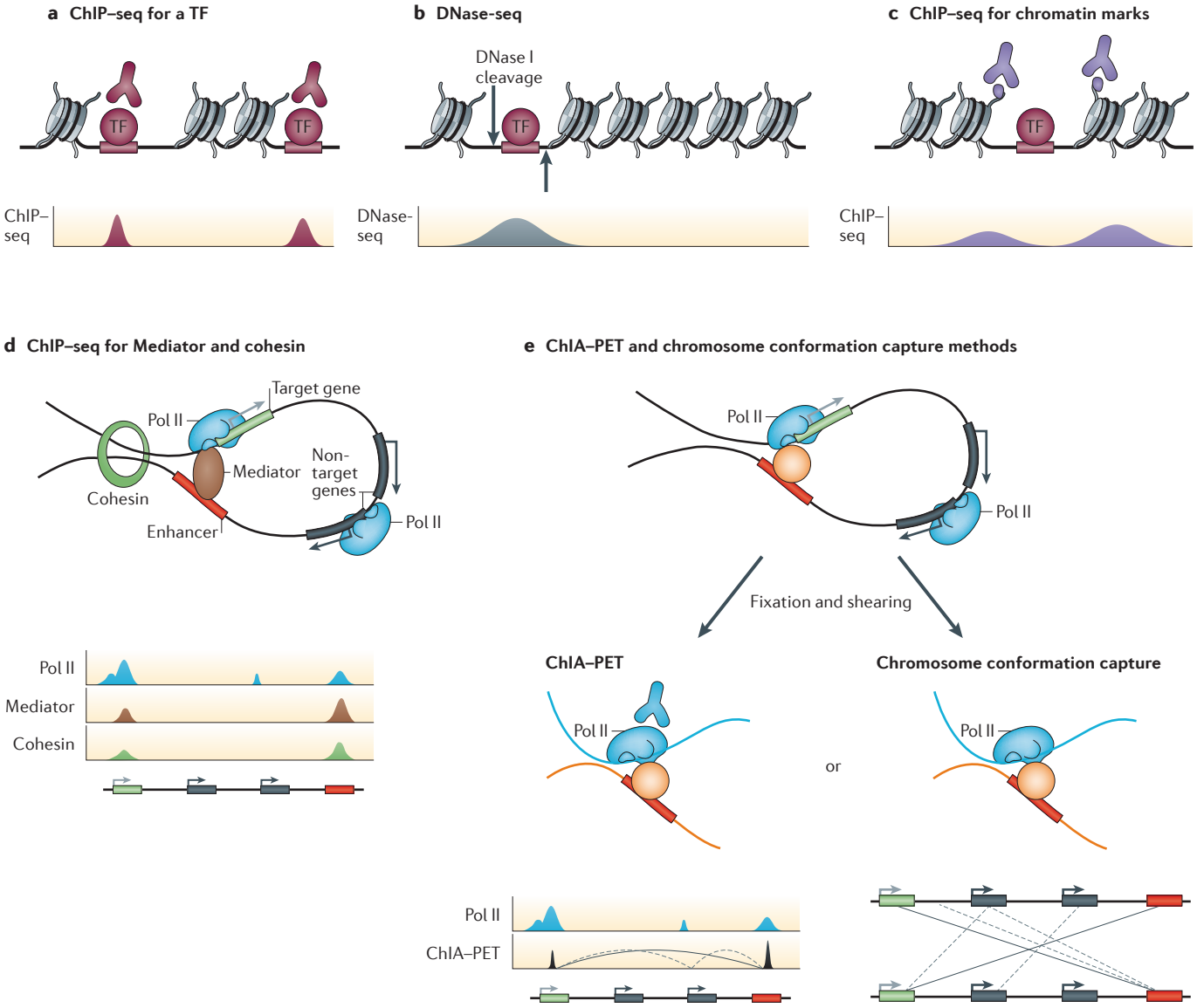


Figure 3 | **Genomic methods for predicting enhancers through the detection of transcription factor binding, 'open' chromatin, chromatin marks, or long-range contacts.** The principles of the different methods (top panel of each part) and the corresponding data output (such as deep sequencing read density) that is used for regulatory element identification (bottom panel of each part) are shown. **a** | Chromatin immunoprecipitation followed by deep sequencing (ChIP–seq) uses antibodies to determine the location of transcription factor (TF) binding sites genome wide. Although enhancers are bound by TFs, not all TF binding sites correspond to functional enhancers. **b** | Active enhancers and other regulatory elements are depleted of nucleosomes such that the DNA is accessible. Such regions can be detected by DNase I or micrococcal nuclease (MNase) digestion followed by deep sequencing (DNase-seq or MNase-seq, respectively). **c** | Nucleosomes that flank active enhancers bear characteristic histone modifications that can be detected by ChIP–seq using specific antibodies. **d** | Enhancers are brought into close proximity of their respective target promoters through the formation of chromatin loops, which are thought to be established by cohesin and Mediator complexes. ChIP–seq can detect the contact points of cohesin and Mediator at promoters and enhancers, and has been used to predict enhancers. **e** | Chromatin interaction analysis with paired-end tag sequencing (ChIA–PET) and chromosome conformation capture (3C)-based methods preserve and detect spatial contacts by crosslinking, DNA fragmentation, DNA fragment ligation and deep sequencing. ChIA–PET includes a ChIP step to enrich for complexes that contain a specific protein, such as RNA polymerase II (Pol II). In contrast to ChIP–seq (part **d**), both ChIA–PET and 3C-based methods detect not only the contact points but also the pairwise connections between these points. The thin, solid lines indicate that pairwise connections between spatial contact points are captured in ChIA–PET and 3C-based methods. For 3C-based methods a schematic output of a chromosome conformation capture carbon copy (5C) or Hi-C experiment is shown; this method probes all interactions between defined genomic loci for their spatial proximity and physical contacts, which is similar to ChIA–PET in that it might (solid lines) or might not (dashed lines) correspond to regulatory interactions.

Chromatin is a dynamic structure that can allow and restrict transcription factor binding; that is, it is able to act as a 'gatekeeper' of regulatory regions (FIG. 2a). For example, DNA accessibility differs between cell types and changes dynamically both during development and in response to external stimuli (reviewed in REF. 64). Such changes in DNA accessibility are believed to be mediated by regulatory proteins, including a special class of transcription factors called pioneer factors (for example, forkhead box protein A1 (FOXA1; also known as HNF3α))[65]. These are thought to bind to nucleosomal DNA and recruit additional factors to evict the nucleosome and open up the region, thereby making the DNA accessible to other transcription factors (reviewed in REF. 65). Alternatively, transcription factors might open DNA by competing with nucleosomes for DNA binding[66,67].

DNase-seq or FAIRE–seq can thus be used to determine accessible chromatin and predict enhancers independently of any given transcription factor (for example, even for cells for which important transcription factors are unknown). The information they provide is thus broader than, but cannot be regarded as entirely independent of, that from ChIP–seq, as transcription factor binding and DNA accessibility are highly correlated[68,69]. Although active enhancers are predominantly found in regions of accessible chromatin, not all accessible regions correspond to active enhancers (FIG. 2b–f). For example, promoter regions around TSSs are often accessible invariantly across different cell types[70,71]. In addition, proteins that regulate other aspects of gene expression or chromosomal biology also bind to the DNA at nucleosome-free regions. These include insulator proteins (for example, Centrosomal protein 190 kDa (Cp190) in flies or CCCTC-binding factor (CTCF) in flies and mammals), which have been found enriched at accessible regions that do not function as active enhancers[72,73]. Finally, accessible regions can also be actively repressed or kept inactive by the binding of repressive transcription factors, which are prevalent in development[74] (FIG. 1c). Indeed, enhancers with cell-type-specific activity were found to be open in another cell type in which they were not active[72]. Moreover, some regions are primed for activity at later developmental time points (FIG. 2e; reviewed in REF. 65). An interesting example of priming was observed for the *lsy-6* microRNA locus in *Caenorhabditis elegans*, which requires priming before it is decompacted and competent to respond to regulatory cues later in development[75].

Although not all accessible regions correspond to active enhancers, combining DNase-seq with additional assays allows the study of diverse types of genomic regulatory elements, including insulators, silencers and elements with putative novel types of functions. These cannot be found by directed means if the factors involved are unknown. However, for the identification of enhancers, additional information, such as the biochemical properties of histone proteins, are frequently used.

Predictions from histone modifications

Active genomic enhancers are devoid of nucleosomes (see above), but the histones in the flanking nucleosomes often carry characteristic post-translational modifications (reviewed in REFS 64,76). Interestingly, genome-wide mapping of histone modifications revealed highly stereotypical patterns with different marks enriched at active enhancers and promoters, and at transcriptionally silent or repressed regions (FIGS 2b,c,3c). These marks therefore allow the genome-wide prediction of promoters and enhancers[51,52,70,77,78]. Promoters are typically marked by H3K4me3 and enhancers by H3K4me1, and both are additionally marked by H3K27ac upon activation[52,72,78] (FIG. 2b,c). By contrast, H3K9me3 typically labels transcriptionally silent heterochromatic regions[79], and silent or repressed promoters and enhancers are often occupied by nucleosomes that contain the H3K27me3 mark, which has been linked to Polycomb repression (FIG. 2d; reviewed in REF. 80). It is interesting that H3K27me3 marks both promoters and enhancers, sometimes across large domains, and cannot co-occur with H3K27ac on the same histone.

The prediction of enhancers using histone marks is now widely used, for example, in the annotation of genome-wide functional elements by individual groups and international consortia[12,52,78,81–84], and it agrees well with enhancer activity assays[51,72,78].

The systematic genome-wide study of histone modifications has revealed several surprising insights into transcriptional regulation. This includes bivalent chromatin[85] at poised enhancers[52] (FIG. 2d) that carry histone modifications associated with both active and repressed states (that is, H3K4me1 and H3K27me3). More recently, latent enhancers have been defined as genomic regions that are not labelled by any enhancer-associated marks (that is, H3K4me1 and H3K27ac) but that acquire these active marks and transcription factor binding upon stimulation of cellular signalling pathways[86] (FIG. 2f).

Intriguingly, H3K79me3 — a modification that is found in the body of actively transcribed genes[87] — has also been observed at active enhancers and, together with Pol II binding, was a good predictor of activity[78]. This might reflect physical contacts between enhancers and promoters that are fixed during crosslinking or could imply that enhancers are transcribed themselves. Indeed, both nascent transcription[88,89] (which is detected by global run-on sequencing (GRO-seq)) and enhancer-derived RNAs (eRNAs)[90,91] were observed at regions marked by enhancer-associated chromatin modifications, and this transcription correlated with the activation of nearby genes[90,91]. However, the extent to which eRNAs help in the *de novo* identification of enhancers is less clear: the transcription of genes might confound the analysis of introns, which frequently contain enhancers. Furthermore, intergenic transcripts might correspond to novel genes rather than to eRNAs and enhancers[88,89]. Finally, the biological function of eRNAs remains unclear, and there is no consensus about whether they mediate specific biochemical functions[92], whether they are involved in or required for the establishment and/or maintenance of

Insulator
A chromatin element that acts as a barrier against the influence of positive signals (from enhancers) or negative signals (from silencers and heterochromatin).

Silencers
DNA sequences that cause reduced expression of their target gene (or genes).

Global run-on sequencing
(GRO-seq). A genome-wide method that maps the position and amount of transcriptionally engaged RNA polymerase II.

accessible chromatin[93], or whether they represent non-consequential events caused by the accidental association of Pol II with active enhancers[94,95].

Given the widespread use of histone modifications to predict enhancers, it is interesting that there is no consensus about which marks should be used. For example, such predictions often require the presence of H3K4me1 and H3K27ac, low levels of H3K4me3 and the absence of H3K27me3 (REFS 51,52). The combination of H3K4me1, H3K27ac and H3K79me3 has been reported to be a good predictor for *D. melanogaster* developmental enhancers[78]. Finally, approaches such as ChromHMM are able to combine the signals of dozens of histone modifications to segment the genome into different chromatin states, including enhancer- and promoter-like states[96]. Generally, the rules used to predict enhancers seem to be driven by the availability of data sets and their prediction performances rather than guided by mechanistic models; the rules also often incorporate data sets on DNA accessibility and on cofactor or Pol II binding.

This variable use of different chromatin features for enhancer prediction probably stems from two related reasons. First, none of the known histone modifications correlates perfectly with enhancer activity, and even combinations of marks are not perfect predictors[72,78]. In addition, many active enhancers seem to lack characteristic marks; for example, 41% of mesodermal enhancers in *D. melanogaster* embryos were not predicted by the presence of H3K27ac[78]. Second, the functional roles for most of the chromatin modifications associated with active enhancers are unknown. There is no evidence that histone marks such as H3K4me1 or H3K27ac are sufficient, necessary or even mechanistically involved in transcription[97,98]. Moreover, both marks have fairly high turnover rates, and it is thus likely that neither is inherited across mitosis nor is instructive[99]. Similarly, H3K4me3 cannot maintain transcription in the absence of an activating transcription factor and is rapidly lost from a previously active promoter region when the activating transcription factor is removed[99]. It will be interesting to learn about the functional roles of such marks and their combinations, and how these will improve enhancer predictions, particularly together with additional marks that are being discovered on active enhancers (reviewed in REF. 100).

Predictions using enhancer–promoter interactions

Above, we discuss methods to predict or identify genomic regions that act as enhancers by assessing enhancer-associated sequence or chromatin features and binding of transcription factors or cofactors.

The fact that enhancers are brought into close spatial proximity of their target promoters has also been exploited by two types of enhancer prediction methods. First, the DNA contact points of the cofactors that mediate enhancer–promoter or enhancer–polymerase interactions can be determined by ChIP–seq. For example, both cohesin and Mediator have been shown to be involved in bringing enhancers and promoters together[101,102], and their ChIP profiles can provide information about the location of enhancers[103] (FIG. 3d).

Another class of methods directly assesses the spatial proximity and physical contact between enhancers and the polymerase machinery at the core promoters of their target genes. Both intrachromosomal and interchromosomal physical contacts (that is, spatial proximity) can be probed using the recently developed chromosome conformation capture (3C) assay and its variants circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C) (FIG. 3e) and Hi-C (reviewed in REF. 104). In these assays, chromosomal contacts are fixed with formaldehyde, the linear genomic DNA is sheared, and spatially proximal DNA fragments are ligated. The resulting chimeric DNA molecules contain fragments that are not close together in the linear genomic sequence but that reflect long-range spatial contacts, which can be identified by deep sequencing (FIG. 3e). A variant method, chromatin interaction analysis with paired-end tag sequencing (ChIA–PET)[105], essentially couples 3C-based methods to ChIP, thereby probing chromosomal contacts that involve certain protein factors. As transcription involves promoter-bound Pol II, one application of ChIA–PET uses Pol II-specific antibodies to identify transcriptional enhancers and their target genes[106] (FIG. 3e).

Even though 3C-based methods are partly being used to predict enhancers and their target genes, they were originally developed to probe the spatial organization of the genome (that is, to detect physical contacts rather than regulatory interactions). When applying these methods to study gene regulation, it is therefore important to carefully consider their properties. The resolution of spatial contacts is often low, typically on the order of kilobases or tens of kilobases (in the case of Hi-C, up to megabases[107]), even though the resolution improves with greater sequencing depth[108,109]. Furthermore, contacts between regions that are adjacent in the linear genomic sequence (that is, less than ~10 kb, where many enhancers are found relative to their target promoters) are difficult to detect reliably because of high background at close distances. In addition, as 3C-based methods are based on formaldehyde crosslinking, they might suffer from the problems discussed above[46,47], such as preferential crosslinks between open genomic regions, which have indeed been observed[107]. Finally, and most importantly, 3C-based methods assess spatial proximity or physical contacts, which do not necessarily reflect functional regulatory relationships (reviewed in REFS 110,111). Indeed, 3C-based methods predict frequent contacts both between many regions on single chromosomes and across chromosomes[107,109,112]. This suggests that many of these contacts are unlikely to participate in gene regulation[110] because it would be difficult to reconcile their range and frequency with prior studies, which typically found distinct enhancers to be necessary and sufficient for gene expression and development (reviewed in REF. 4). The necessity and sufficiency of individual enhancers for gene expression in defined cell types or tissues is also supported by several known diseases that are related to specific enhancer mutations[113–115] (reviewed in REF. 7), by developmental phenotypes that are caused by enhancer disruption during mutational

Chromatin interaction analysis with paired-end tag sequencing
(ChIA–PET). A high-throughput method based on a combination of chromatin immunoprecipitation (ChIP) and chromatin proximity ligation assays to predict long-range chromatin interactions that are mediated by either RNA polymerase II or transcription factors.

Table 1 | **Overview of enhancer testing and screening methods***

Methods	Throughput	Model system	Plasmid or integrated	Reporter set-up‡	Quantification
Image-based methods[53,121,123] (reviewed in REF. 4)	Up to thousands	Mouse and fly; *in vivo*	Random or site-specific integration		Staining intensity
Enhancer-FACS-seq[124]	Hundreds	Fly; *in vivo*	Site-specific integration		Fluorescence intensity and sequencing
Sharon *et al.*[129]	Thousands	Yeast	Plasmid		Fluorescence intensity
CRE-seq[165]	Thousands	Yeast	Site-specific integration		Sequencing of transcribed barcode
Nam *et al.*[125]	Hundreds	Sea urchin; *in vivo*	Random integration		Sequencing of transcribed barcode
MPRA[126],CRE-seq[128] and MPFD[127]	Thousands	Human and mouse; *in vitro*	Plasmid		Sequencing of transcribed barcode
STARR-seq[72]	Millions	Human and fly; *in vitro*	Plasmid		Sequencing of transcribed enhancer

CRE-seq, *cis*-regulatory element analysis by sequencing; FACS, fluorescence-activated cell sorting; MPFD, massively parallel functional dissection; MPRA, massively parallel reporter assay; STARR-seq, self-transcribing active regulatory region sequencing. *This table provides an overview of image- and deep-sequencing-based methods with their respective throughput and model system: sea urchin[125], yeast[129,165], mouse, human or fly[72,124,126–128]. Important properties of such assays include whether the candidate fragment is tested ectopically on a plasmid or after genomic integration, the reporter set-up with respect to the relative position of the candidate fragments or barcodes, and whether and how enhancer activity is quantified. ‡Tested fragments are shown in blue. Reporter genes are shown in beige; fluorescent or otherwise visually detectable reporter genes are shown in yellow. Barcodes are represented by red bars, and regions used for identifying fragments in a multiplexed sample are denoted by grey lines.

analyses (reviewed in REFS 4,116) and by results from the specific targeting of individual enhancers by repressors[117]. It will be important to reconcile these results and determine how many and which chromatin contacts correspond to causal regulatory relationships[110,111], which is a crucial question that has only recently started to gain attention[109,118,119].

Functional testing of enhancers

We have so far discussed methods that use known properties of enhancer sequences, enhancer-associated chromatin features or chromosomal interactions to predict enhancers. DNA sequences can also be directly tested for their ability to activate or enhance transcription from a minimal core promoter. In fact, this activity — which is independent of the sequence context of the enhancer and can be assessed in heterologous reporter systems — has been the defining property of enhancers[1] and is used as the 'gold standard' to evaluate enhancer predictions[50–52,120]. Enhancer tests in whole developing embryos (for example, in flies, nematodes, zebrafish and mice) generally have readouts based on images, whereas tests in cell culture typically either use luciferase or directly measure the abundance of reporter transcripts (for example, by deep sequencing) (TABLE 1).

Image-based enhancer testing. Enhancers are frequently tested in developing embryos by placing the candidate DNA sequences upstream of a minimal promoter and a reporter gene. The enhancer activity is measured by determining the abundance and localization of the reporter transcript (for example, by *in situ* hybridization). Alternatively, the reporter gene is detected at the protein level, such as by enzymatic activities (for example, luciferase or β-galactosidase (which is encoded by *lacZ*)), fluorescence (for example, GFP) or specific antibodies. Many promoter and enhancer candidates have so far been tested in transgenic *D. melanogaster*[43,121], *C. elegans*[122] and mouse embryos[123]. Although such *in vivo* approaches provide a rich readout (such as activity patterns) throughout an entire embryo, they require the generation of transgenic animals. Therefore, they are not suitable for genome-wide enhancer screening and have mainly been used to test predicted enhancers[50,52,120].

Parallel enhancer testing using integrated reporters. Enhancer-FACS-seq (eFS) has recently been developed in *D. melanogaster* to test candidate enhancers and to provide information about their activities in selected cell types or tissues. Similar to the approaches above, enhancer candidates are tested in transgenic animals, but the generation of these transgenes has been simplified and DNA fragments can be tested in parallel using deep sequencing[124]. For this, the candidate fragments are cloned upstream of a minimal promoter and the GFP gene, and the resulting reporter construct is integrated into a defined site in the fly genome. Transgenic animals are selected and crossed to cell-type-specific reporter lines in batches, which allows the isolation of GFP-positive cells by fluorescence-activated cell sorting

(FACS). The abundance of the respective candidate sequences in the genomic DNA of GFP-positive and GFP-negative cells is quantified using deep sequencing and compared. By the use of a second, tissue-specific fluorescent marker (for example, mesodermal twi-CD2), the screen can be targeted to specific cell types or tissues.

Parallel enhancer testing using ectopic reporters. Various methods have recently been developed to assess enhancer activities using plasmid-based systems (that is, ectopically). These methods use deep sequencing and DNA barcodes (TABLE 1), and they circumvent, to various extents, the need to individually clone and test each candidate one at a time, which allows higher throughput of hundreds[125] or thousands[126–129] of candidates in parallel. Each enhancer candidate is placed upstream of a minimal promoter and a reporter gene that contains a heterologous barcode. The reporter plasmids are introduced into cells in batches, and the barcode-containing reporter transcripts are isolated and quantified using deep sequencing. The number of deep sequencing reads per barcode reflects the abundance of the respective reporter RNA and thus the activity of the corresponding candidate enhancer. The unique association of each candidate with a defined barcode has been achieved through individual cloning[125] or random cloning followed by a deep sequencing-based assignment step[127], or by chemically synthesizing DNA oligonucleotides that contain both the candidates and the barcodes on single molecules[126,128,129]. Chemical oligonucleotide synthesis restricts the length of the candidates to its current limits of less than 300 bp and to a few thousand pre-defined candidates. This allows the versatile testing of essentially arbitrary sequence variants of known enhancers, for example, to assess the importance of each nucleotide or all transcription factor binding motifs systematically[126,127]. However, it does not scale to genome-wide screens.

Genome-wide functional enhancer screening. Self-transcribing active regulatory region sequencing (STARR-seq) makes use of the fact that enhancers can work independently of their position and orientation relative to the TSS, and places the candidate enhancers downstream of the minimal promoter into the reporter gene[72]. Active candidates enhance their own transcription, and their strength can be quantitatively assessed by the abundance of their transcripts among cellular RNA. This direct coupling of enhancer activity to its sequence in *cis* avoids the use of heterologous barcodes and means that millions of candidate fragments of variable lengths and from arbitrary sources of input DNA — including genomic DNA or bacterial artificial chromosomes that have been randomly fragmented by sonication — can be cloned and screened in batches. STARR-seq is therefore unique in that it allows genome-wide screens of enhancer activity.

The plasmid-based ectopic assays achieve their high throughput by directly introducing reporter plasmids into the cells of interest. Although this allows the testing of highly complex candidate libraries, it currently limits such assays to cell types that can be efficiently transduced and means that some aspects of developmental gene regulation at the chromatin level are not assessed. By contrast, the image-based methods and eFS assess the enhancer activities of the candidates within the genome of a developing animal, which limits library complexity and the number of candidates that can be tested in parallel. However, the integrated candidates undergo development and cell differentiation within a chromosomal context, which might affect their chromatin states, making it both interesting and instructive to compare the results of both types of methods (BOX 2).

Potential of high-throughput enhancer testing methods. The recent development of high-throughput methods introduced here allows various important questions to be answered. Foremost, the genome-wide identification of enhancers that is directly based on their activity and that is independent of chromatin features[72] will reveal the relationship between enhancer function and chromatin, which allows the reassessment of individual chromatin features and the prediction methods that use them. It will further shed light on the role of chromatin in gene regulation during cell lineage differentiation and development (BOX 2). Furthermore, we expect that the high number of functionally characterized enhancers will enable powerful statistical sequence analyses that can reveal the relationship between enhancer sequence elements (for example, transcription factor motif matches) and enhancer function. Such analyses will also benefit enormously from the ability to test enhancer variants[126,127] or synthetic motif arrangements systematically[130,131] in a high-throughput manner. We also anticipate that further technological improvements of the library delivery methods will make it possible to assess enhancer activity both *in vivo* and *ex vivo* in any tissue, thus making such massively parallel screens a method of choice to study *cis*-regulatory sequences.

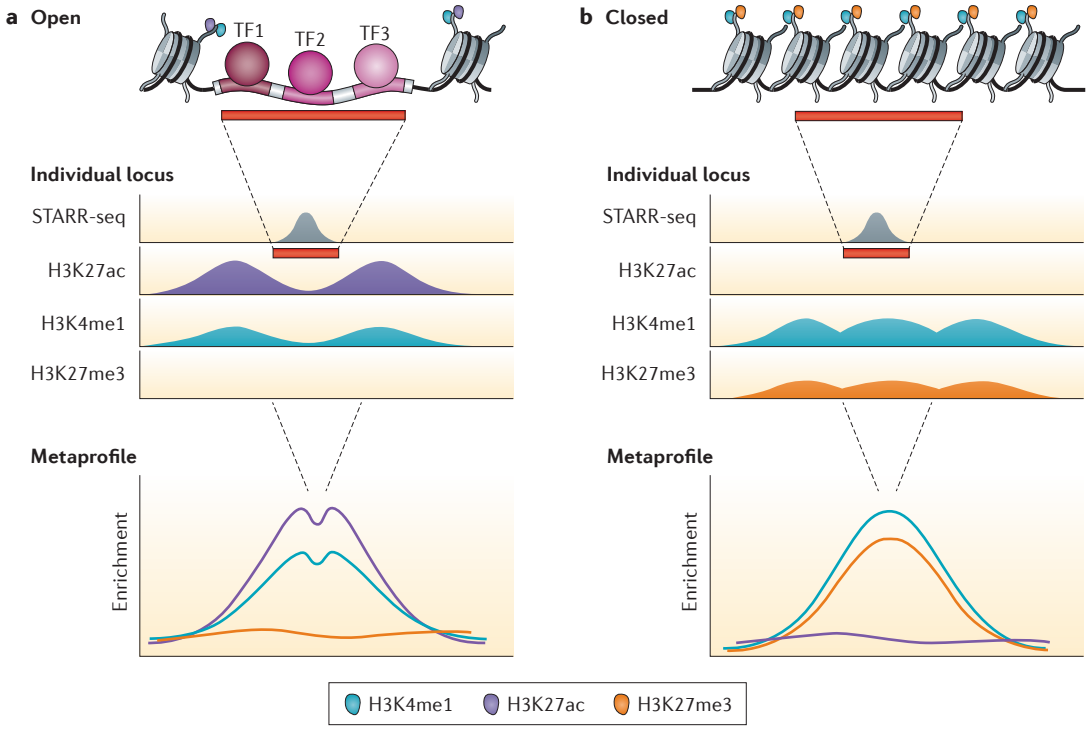
Conclusions and future directions

The advent of deep sequencing has enabled the development of a large variety of methods to study gene expression, regulatory elements and chromatin architecture across entire genomes, which has led to unprecedented insights over the past few years. In this Review, we have covered different approaches to predict transcriptional enhancers or to assess their activities, including the analyses of sequence motifs, *in vivo* binding sites of transcription factors and cofactors, characteristic chromatin features and histone modifications, as well as massively parallel enhancer assays.

The growing number of functionally characterized enhancers will enable powerful systematic analyses of the sequences of these enhancers and provide important insights into how regulatory function is encoded in the DNA sequence. Hypotheses and specific predictions can in turn be tested by the newly developed enhancer assays in a high-throughput manner (for example, using defined synthetic DNA molecules)[126,128,129]. Such cycles of enhancer characterization, prediction, validation and

Barcodes
Short and typically artificially designed DNA sequences that are used to uniquely identify DNA constructs (for example, those expressing short hairpin RNAs or reporter genes) or cell lines (for example, yeast knockouts). The uniqueness of the barcodes allows screening or testing in parallel using pooling.

Box 2 | **Characteristics of ectopic enhancer assays**



Comparison of enhancer activities measured in ectopic assays shows that such assays can assess the enhancer strengths of DNA sequences even if they are silenced in their endogenous genomic loci. For example, in *Drosophila melanogaster* S2 cells, sequences that functioned as active enhancers in ectopic assays — as determined by self-transcribing active regulatory region sequencing (STARR-seq) — could be separated into two groups on the basis of whether the corresponding genomic loci were accessible to DNase; that is, whether these loci are 'open' (see the figure, part **a**) or 'closed' (REF. 72) (see the figure, part **b**). Open enhancers correlated with the classical 'active marks': histone H3 acetylated at lysine 27 (H3K27ac) and histone H3 monomethylated at lysine 4 (H3K4me1). Note that the characteristic 'dip' in the average enrichment profiles is due to nucleosome depletion at the 'summit'. Closed enhancers are found in genomic regions that show the 'silencing mark' H3K27me3 but, importantly, they were also marked by H3K4me1. This suggests that closed enhancers have the potential to activate transcription and that they are recognized as enhancers in their endogenous contexts yet seem to be actively repressed. This silencing presumably involves Polycomb group proteins as suggested by the H3K27me3 marks.

Importantly, the discrepancy between the activity of a sequence in its endogenous genomic context and that in a reporter plasmid cannot be explained by genomic integration per se: results from plasmid-based assays agreed well with results from genomically integrated reporters in both yeast[129] and insect cells[72]. Thus, the silencing might be initiated during cellular differentiation in the course of development, or it could depend on sequences outside the enhancer itself. It will be interesting to reveal the cause and mechanism of this silencing.

refinement of hypotheses will increase our understanding of gene regulation during animal development and physiology.

Improvements of existing methods. In addition, we expect further improvements to these assays, such as their increasing applicability in a tissue- or cell-type-specific manner *in vivo*. This has already been established for ChIP–seq[78,132] and medium-throughput enhancer assays[124], and it is likely to be further developed for many of the genome-wide assays discussed above. A parallel development has been the down-scaling of methods — such as high-throughput RNA sequencing (RNA-seq), ChIP–seq, the mapping of open chromatin (for example, by the assay for transposase-accessible chromatin using sequencing (ATAC-seq)) and Hi-C — for their application to few cells or even to individual cells[133–136]. We also expect additional advances and modifications of existing high-throughput enhancer assays.

Manipulation of endogenous enhancer activity. More recently, the approaches for identifying enhancers have been complemented by methods to manipulate enhancers *in situ*, such as transcription activator-like effectors (TALEs) and the clustered regularly interspaced short palindromic repeat (CRISPR)–Cas9 system, both of which allow the recruitment of transcription factors[117,137] and cofactors[138,139] to any desired position in a genome (FIG. 4A). For example, enhancers and promoters can be activated or repressed[117,137], and histone modifications can be established or removed in a controllable and reversible manner to assess their regulatory effect on

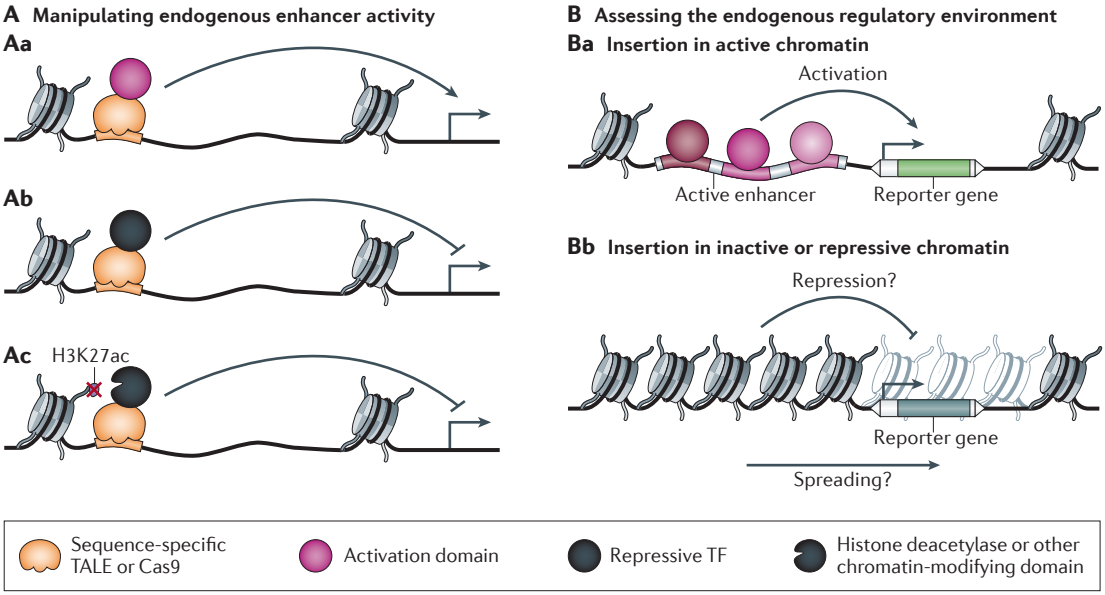


Figure 4 | **Novel approaches to study and manipulate endogenous regulatory activities. A** | Recent technological advances allow the modulation of gene expression at endogenous genomic loci. DNA-binding transcription activator-like effectors (TALEs) or clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) proteins can be designed to target essentially arbitrary DNA sequences. Through these domains, transcriptional activators (part **Aa**), repressive transcription factors (TFs)[117,137,139] (part **Ab**) or chromatin-modifying domains[138,139] (part **Ac**) can be recruited to genomic regulatory regions to modulate the transcription of target genes. Alternatively, both systems can be used to recruit endonucleases to specific genomic loci to introduce mutations or other sequence alterations[142]. **B** | Transcriptional reporters that are integrated into the genome allow the assessment of gene regulatory activities at the specific integrations sites and are conceptually similar to enhancer trap approaches. Reporters integrated near active enhancers are upregulated (part **Ba**), whereas those near silenced or repressive regions can be downregulated (part **Bb**). Recent methods have readouts based on images (for example, genome regulatory organization mapping with integrated transposons (GROMIT)[143]) or deep sequencing, which allows the parallel assessment of a large number of of integration sites (for example, thousands of reporters integrated in parallel (TRIP)[145]).

transcription[138,139]. In addition to the functional manipulation of enhancers, related methods based on TALE nucleases (TALENs)[140] and CRISPR–Cas9[141,142] also allow the direct editing of the genomic DNA sequence. Collectively, such *in situ* approaches will allow the 'fine-tuned' manipulation of enhancers, and we expect such techniques to be widely used in the future.

Assessing the endogenous regulatory environment. Several methods have recently been developed or adapted to assess gene regulatory activities and the influence of the genomic environment at endogenous genomic loci. Similar to enhancer activity assays, these include image- and deep sequencing-based methods. For example, genome regulatory organization mapping with integrated transposons (GROMIT)[143] assesses the tissue-specific expression of a randomly integrated transcriptional sensor and is similar to enhancer traps[144]. Thousands of reporters integrated in parallel (TRIP) uses retroviral integration, barcode-containing reporter transcripts and deep sequencing to evaluate the influence of the extended genomic context on enhancer or promoter activity[145] (FIG. 4B). Both types of methods are complementary to enhancer assays: they assess the net regulatory activity at specific genomic loci irrespective of the genomic elements that contribute to this activity, whereas enhancer assays typically measure the activity of defined candidate sequences regardless of their endogenous context or regulatory target.

The methodological and conceptual progress discussed here occurs at a time when it is becoming increasingly clear that sequence changes in regulatory genomic regions are often causal for various diseases[146], including cancer[114,115]. Interestingly, studying enhancers in the context of cancer also informs us about basic enhancer features, such as the existence of broad regions that are strongly bound by Mediator and that are located near highly expressed genes. In such regions, which have been termed super enhancers[103], several closely spaced enhancers might act synergistically to strongly activate gene transcription, and it will be interesting to understand how multiple enhancers communicate with each other and with their respective promoters to achieve high expression levels. Cofactor occupancy at super enhancers has been shown to be particularly sensitive to cofactor inhibition (for example, by BET bromodomain inhibitors[147,148]), which may underlie the sensitivity of acute myeloid leukaemia[149] and multiple myeloma[150] to these agents. The ability to interpret the functional implications of non-coding sequence changes and the functional roles

of transcription factors and cofactors will be crucial for understanding and treating disease, for example, by the inhibition of chromatin-modifying cofactors[151,152].

In summary, we believe that knowledge gained by studying enhancers on a genome-wide level, together with the careful dissection of individual enhancers, will further improve our understanding of transcriptional regulation in animals over the next few years. We are looking forward to seeing how the currently available approaches and future ones will be used to answer the remaining long-standing open questions and challenges. In our opinion, the three main questions are: how is regulatory information encoded in the four-letter 'alphabet' of enhancer sequences? Are enhancer–promoter relationships specified in the genomic DNA sequence, and if so, how is this achieved? And what are the individual functional roles of the key players (that is, the different transcription factors and cofactors, and histone modifications)? These are exciting times to study transcriptional regulation.

1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
This paper reports the first sequence that can increase transcription levels from a given promoter, defines the term enhancer and describes many enhancer properties.
2. Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. **33**, 729–740 (1983).
3. Amano, T. *et al.* Chromosomal dynamics at the *shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
4. Arnone, M. I. & Davidson, E. H. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**, 1851–1864 (1997).
5. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).
6. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
7. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
8. Yáñez-Cuna, J. O., Kvon, E. Z. & Stark, A. Deciphering the transcriptional *cis*-regulatory code. *Trends Genet.* **29**, 11–22 (2013).
9. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, research0088-0088.14 (2002).
10. Richardson, L. *et al.* EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res.* **38**, D703–D709 (2010).
11. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
12. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
13. The modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
14. Tjian, R. The binding site on SV40 DNA for a T antigen-related protein. *Cell* **13**, 165–179 (1978).
15. Giniger, E., Varnum, S. M. & Ptashne, M. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**, 767–774 (1985).
16. 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).
17. Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets using 12 *Drosophila* genomes. *Genome Res.* **17**, 1919–1931 (2007).
18. Del Bene, F. *et al. In vivo* validation of a computationally predicted conserved *Ath5* target gene set. *PLoS Genet.* **3**, 1661–1671 (2007).
19. Hallikas, O. *et al.* Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47–59 (2006).
20. Sinha, S., van Nimwegen, E. & Siggia, E. D. A probabilistic method to detect regulatory modules. *Bioinformatics* **19**, i292–i301 (2003).
21. Herrmann, C., Van de Sande, B., Potier, D. & Aerts, S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and *cis*-regulatory modules. *Nucleic Acids Res.* **40**, e114 (2012).
22. Warner, J. B. *et al.* Systematic identification of mammalian regulatory motifs' target genes and functions. *Nature Methods* **5**, 347–353 (2008).
23. Aerts, S. Computational strategies for the genome-wide identification of *cis*-regulatory elements and transcriptional targets. *Curr. Top. Dev. Biol.* **98**, 121–145 (2012).
24. Hardison, R. C. & Taylor, J. Genomic approaches towards finding *cis*-regulatory modules in animals. *Nature Rev. Genet.* **13**, 469–483 (2012).
25. Wei, G.-H. *et al.* Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO J.* **29**, 2147–2160 (2010).
26. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
27. Noyes, M. B. *et al.* Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277–1289 (2008).
28. Yanez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D. & Stark, A. Uncovering *cis*-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res.* **22**, 2018–2030 (2012).
This paper shows that transcription factor binding can be predicted by cell-type-specific combinations of transcription factor binding sequences for different partner transcription factors, which are shared across many binding sites.
29. Slattery, M. *et al.* Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
30. Blow, M. J. *et al.* ChIP–seq identification of weakly conserved heart enhancers. *Nature Genet.* **42**, 818–822 (2010).
31. Meireles-Filho, A. C. A. & Stark, A. Comparative genomics of gene regulation-conservation and divergence of *cis*-regulatory information. *Curr. Opin. Genet. Dev.* **19**, 565–570 (2009).
32. Kantorovitz, M. R. *et al.* Motif-blind, genome-wide discovery of *cis*-regulatory modules in *Drosophila* and mouse. *Dev. Cell* **17**, 568–579 (2009).
33. Narlikar, L. *et al.* Genome-wide discovery of human heart enhancers. *Genome Res.* **20**, 381–392 (2010).
34. Burzynski, G. M. *et al.* Systematic elucidation and *in vivo* validation of sequences enriched in hindbrain transcriptional control. *Genome Res.* **22**, 2278–2289 (2012).
35. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
36. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
37. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
38. van Steensel, B. & Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotech.* **18**, 424–428 (2000).
39. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nature Rev. Genet.* **13**, 613–626 (2012).
40. Sandmann, T. *et al.* A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436–449 (2007).
41. Zeitlinger, J. *et al.* Whole-genome ChIP–chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev.* **21**, 385–390 (2007).
42. Li, X.-Y. *et al.* Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27 (2008).
43. Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J. & Stark, A. HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev.* **26**, 908–913 (2012).
44. Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl Acad. Sci.* **109**, 21330–21335 (2012).
45. Hammar, P. *et al.* The *lac* repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
46. Teytelman, L., Thurtle, D. M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl Acad. Sci.* **110**, 18602–18607 (2013).
This study shows that ChIP assays can lead to false-positive binding sites for transcription factors or even for non-DNA binding proteins (such as GFP), thus cautioning the interpretation of this widely used technique.
47. Poorey, K. *et al.* Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342**, 369–372 (2013).
48. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2013).
49. Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
50. Visel, A. *et al.* ChIP–seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
This paper shows that p300 binding in the murine forebrain, hindbrain and limb can predict tissue-specific enhancers.
51. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
This study shows that human promoters and enhancers are marked by characteristic combinations of histone modifications that are predictive.
52. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
53. Visel, A. *et al.* A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).
54. May, D. *et al.* Large-scale discovery of enhancers from human heart tissue. *Nature Genet.* **44**, 89–93 (2012).
55. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
56. van Bemmel, J. G. *et al.* A network model of the molecular organization of chromatin in *Drosophila*. *Mol. Cell* **49**, 759–771 (2013).
57. Ram, O. *et al.* Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**, 1628–1639 (2011).
58. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).

59. Axel, R., Cedar, H. & Felsenfeld, G. Synthesis of globin ribonucleic acid from duck-reticulocyte chromatin *in vitro*. *Proc. Natl. Acad. Sci. USA* **70**, 2029–2032 (1973).
60. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
61. Yuan, G.-C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
62. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **488**, 83–90 (2012).
63. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
64. Bell, O., Tiwari, V. K., Thomä, N. H. & Schübeler, D. Determinants and dynamics of genome accessibility. *Nature Rev. Genet.***12**, 554–564 (2011).
65. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
66. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl Acad. Sci.* **107**, 22534–22539 (2010).
67. Miller, J. A. & Widom, J. Collaborative competition mechanism for gene activation *in vivo*. *Mol. Cell. Biol.* **23**, 1623–1632 (2003).
68. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
69. Kaplan, T. *et al.* Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet.* **7**, e1001290 (2011).
70. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
71. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **488**, 75–82 (2013).
72. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
This paper introduces a method that allows the genome-wide identification of enhancers on the direct basis of their activity and that finds 'closed enhancers', which are silenced endogenously presumably at the chromatin level.
73. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).
74. Gray, S. & Levine, M. Transcriptional repression in development. *Curr. Opin. Cell Biol.* **8**, 358–364 (1996).
75. Cochella, L. & Hobert, O. Embryonic priming of a miRNA locus predetermines postmitotic neuronal left/right asymmetry in *C. elegans*. *Cell* **151**, 1229–1242 (2012).
76. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
77. Roh, T.-Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).
78. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genet.* **44**, 148–156 (2012).
This study couples ChIP–seq with nuclear sorting to allow the cell-type-specific investigation of chromatin features.
79. Peters, A. H. F. M. *et al.* Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nature Genet.* **30**, 77–80 (2002).
80. Simon, J. A. & Kingston, R. E. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nature Rev. Mol. Cell. Biol.* **10**, 697–708 (2009).
81. Wamstad, J. A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
82. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
83. Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2012).
84. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
85. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
86. Ostuni, R. *et al.* Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
87. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
88. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**, 622–634 (2011).
89. Core, L. J. *et al.* Defining the status of RNA polymerase at promoters. *Cell Rep.* **2**, 1025–1035 (2012).
90. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
91. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
92. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
93. Natoli, G. & Andrau, J.-C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
94. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nature Cell Biol.* **10**, 1106–1113 (2008).
95. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
96. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215–216 (2012).
97. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22**, 2253–2257 (2012).
98. Pengelly, A. R., Copur, O., Jackle, H., Herzig, A. & Muller, J. A. Histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**, 698–699 (2013).
References 97 and 98 investigate the importance of histone modifications for gene transcription by mutating H3K4 and H3K27. H3-K27R mutants led to the derepression of Polycomb target genes but was otherwise compatible with gene transcription, as were H3K4 mutants that could not be methylated.
99. Hathaway, N. A. *et al.* Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012).
100. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
101. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
102. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
103. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
104. van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. *Nature Biotech.* **28**, 1089–1095 (2010).
105. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **461**, 58–64 (2009).
106. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
107. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
108. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
109. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
This paper significantly improves the resolution of Hi-C experiments and provides bulk evidence that implicates the interactions in gene expression.
110. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
111. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
112. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2013).
113. Lettice, L. A. 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).
114. Sur, I. K. *et al.* Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
115. Huang, F. W. *et al.* Highly recurrent *TERT* promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
References 114 and 115 describe defined mutations in transcriptional regulatory regions (that is, promoters and enhancers) that are causally linked to the deregulation of MYC and telomerase reverse transcriptase (TERT), and cancer.
116. Zeitlinger, J. & Stark, A. Developmental gene regulation in the era of genomics. *Dev. Biol.* **339**, 230–239 (2010).
117. Crocker, J. & Stern, D. L. TALE-mediated modulation of transcriptional enhancers *in vivo*. *Nature Methods* **10**, 762–767 (2013).
This study recruits transcriptional activators and repressors to specific enhancers in *D. melanogaster* using TALE fusion proteins and thereby modulate target gene expression.
118. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306–310 (2013).
119. Kieffer-Kwon, K.-R. *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520 (2013).
120. Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. M. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462**, 65–70 (2009).
121. Manning, L. *et al.* A resource for manipulating gene expression and analyzing *cis*-regulatory modules in the *Drosophila* CNS. *Cell Rep.* **2**, 1002–1013 (2012).
122. Dupuy, D. A. First version of the *Caenorhabditis elegans* promoterome. *Genome Res.* **14**, 2169–2175 (2004).
123. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA enhancer browser — a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
124. Gisselbrecht, S. S. *et al.* Highly parallel assays of tissue-specific enhancers in whole *Drosophila* embryos. *Nature Methods* **10**, 774–780 (2013).
This paper introduces enhancer-FACS-seq to parallelize *in vivo* enhancer testing in *D. melanogaster*.
125. Nam, J. & Davidson, E. H. Barcoded DNA-tag reporters for multiplex *cis*-regulatory analysis. *PLoS ONE* **7**, e35934 (2012).
126. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotech.* **30**, 271–277 (2012).
127. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nature Biotech.* **30**, 265–270 (2012).
128. 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.* **109**, 19498–19503 (2012).
129. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotech.* **30**, 521–530 (2012).
130. Smith, R. P. *et al.* Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature Genet.* **45**, 1021–1028 (2013).
131. Gertz, J., Siggia, E. D. & Cohen, B. A. Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2008).
132. Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* **22**, 766–777 (2012).

133. Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP–seq. *Nature Protoc.* **6**, 1656–1668 (2011).
134. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
135. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213–1218 (2013).
136. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
137. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
138. Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nature Biotech.* **31**, 1133–1136 (2013).
139. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
140. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotech.* **30**, 460–465 (2012).
141. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
This study shows that the bacterial Cas9 protein uses dual RNAs for sequence-specific DNA targeting and cleavage, and highlights the potential of the CRISPR–Cas9 system for genome editing.
142. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
143. Ruf, S. *et al.* Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genet.* **43**, 379–386 (2011).
144. Mollereau, B. *et al.* A green fluorescent protein enhancer trap screen in *Drosophila* photoreceptor cells. *Mech. Dev.* **93**, 151–160 (2000).
145. Akhtar, W. *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**, 914–927 (2013).
146. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
147. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
This paper defines super enhancers as exceptionally long genomic regions that are strongly bound by cofactors. The proximity to some oncogenes and the loss of bromodomain-containing protein 4 (BRD4) binding upon inhibition draws the attention of a broader medical community to such enhancers.
148. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27**, 2648–2662 (2013).
149. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).
This study reports that inhibition of the broadly expressed transcriptional co-activator BRD4 has a specific effect on acute myeloid leukaemia cells.
150. Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
151. Knutson, S. K. *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nature Chem. Biol.* **8**, 890–896 (2012).
152. Herrmann, H. *et al.* Small-molecule inhibition of BRD4 as a new potent approach to eliminate leukemic stem- and progenitor cells in acute myeloid leukemia AML. *Oncotarget* **3**, 1588–1599 (2012).
153. Stormo, G. D. & Zhao, Y. Determining the specificity of protein–DNA interactions. *Nature Rev. Genet.* **11**, 751–760 (2010).
154. Roth, F. P., Hughes, J. D., Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotech.* **16**, 939–945 (1998).
155. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
156. Bosch, J. R., Benavides, J. A. & Cline, T. W. The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development* **133**, 1967–1977 (2006).
157. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**, 400–403 (2008).
158. Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M. Quantifying DNA–protein interactions by double-stranded DNA arrays. *Nature Biotech.* **17**, 573–577 (1999).
159. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
160. Li, J. J. & Herskowitz, I. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* **262**, 1870–1874 (1993).
161. Meng, X., Brodsky, M. H. & Wolfe, S. A. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature Biotech.* **23**, 988–994 (2005).
162. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
163. Portales-Casamar, E. *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38**, D105–D110 (2010).
164. Newburger, D. E. & Bulyk, M. L. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **37**, D77–D82 (2009).
165. Mogno, I., Kwasnieski, J. C. & Cohen, B. A. Massively parallel synthetic promoter assays reveal the *in vivo* effects of binding site variants. *Genome Res.* **23**, 1908–1915 (2013).

Acknowledgements
The authors thank O. Bell (Institute of Molecular Biotechnology (IMBA), Vienna, Austria) and members of A.S's group (Research Institute of Molecular Pathology (IMP), Vienna, Austria) for discussions. We apologize to all colleagues whose work could not be discussed or referenced owing to space limitations. D.S. is supported by a European Research Council (ERC) Starting grant (no. 242922) awarded to A.S., and A.S.'s group by the Austrian Science Fund (FWF): F4303-B09. Basic research at the IMP is supported by Boehringer Ingelheim GmbH.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
REDfly: http://redfly.ccr.buffalo.edu/
Vienna Tiles Browser: http://enhancers.starklab.org/
VISTA: http://enhancer.lbl.gov/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF