Towards a comprehensive catalogue of validated and target-linked human enhancers

Molly Gasperini[1], *Jacob M. Tome*[1] *and Jay Shendure*[1,2,3,4]*

Abstract | The human gene catalogue is essentially complete, but we lack an equivalently vetted inventory of bona fide human enhancers. Hundreds of thousands of candidate enhancers have been nominated via biochemical annotations; however, only a handful of these have been validated and confidently linked to their target genes. Here we review emerging technologies for discovering, characterizing and validating human enhancers at scale. We furthermore propose a new framework for operationally defining enhancers that accommodates the heterogeneous and complementary results that are emerging from reporter assays, biochemical measurements and CRISPR screens.

Transcription factors
(TFs). Proteins that bind DNA, typically consisting of specific DNA sequences or motifs, and contribute to the regulation of RNA transcription.

Open chromatin
A nucleosome-loose packaging state of DNA that is permissive for transcription-factor binding.

Episomal reporter vector
Plasmid DNA that can be synthetically delivered, is autonomous from genomic DNA and includes a reporter gene, typically downstream of a candidate regulatory element (for example, an enhancer adjacent to a minimal promoter).

[1]*Department of Genome Sciences, University of Washington, Seattle, WA, USA.*

[2]*Brotman Baty Institute for Precision Medicine, Seattle, WA, USA.*

[3]*Allen Discovery Center for Cell Lineage, Seattle, WA, USA.*

[4]*Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.*

**e-mail: shendure@uw.edu*

https://doi.org/10.1038/s41576-019-0209-0

The human genome is currently believed to harbour from hundreds of thousands to millions of enhancers — stretches of DNA that bind transcription factors (TFs) and enhance the expression of genes encoded in *cis*. Collectively, enhancers are thought to play a principal role in orchestrating the fantastically complex programme of gene expression that underlies human development and homeostasis. Although most causal genetic variants for Mendelian disorders fall in protein-coding regions, the heritable component of common disease risk distributes largely to non-coding regions and appears to be particularly enriched in enhancers that are specific to disease-relevant cell types. This observation has heightened interest in both annotating and understanding human enhancers. However, despite the clear importance of human enhancers to both basic and disease biology, there is a tremendous amount that we still do not understand about their repertoire, including where they reside, how they work and what genes they mediate their effects through.

This situation does not arise from a lack of effort. Rather, our understanding of the core characteristics of enhancers, based largely on a few paradigmatic examples, is being challenged by studies that suggest a more heterogeneous landscape. New data types — for example, data based on massively parallel reporter assays (MPRAs) or genome editing — are further complicating the picture, particularly because biochemical annotations and the functional data are not always in agreement. As a consequence, the field lacks a clear framework for identifying enhancers, with different subfields (for example, biochemistry, genomics and so forth) using different definitions and criteria, even though we are all ostensibly studying the same underlying biological phenomenon. These challenges are critical to resolve and also represent an excellent opportunity to gain further insight into the nature of enhancers, as well as into the landscape and heterogeneity of gene regulatory mechanisms across the human genome.

Here we present a survey of emerging technologies for discovering, characterizing and validating enhancers at scale. We begin with a history of the concept of an enhancer and its evolving operational definition. We then review contemporary and emerging technologies for characterizing enhancers at scale. Next we propose a set of evidentiary standards for considering a candidate enhancer as being strongly, moderately or weakly supported. Finally, we look forwards and highlight the key challenges in the field.

A brief history of the concept of an enhancer

The term 'enhancer' first appeared in the context of molecular biology in 1981 (BOX 1). By this point in time, gene expression was already thought to be regulated by proteins[1] that bound DNA[2]. But why do these proteins bind to specific locations, and how does their binding control gene expression? In eukaryotic systems, in addition to a primary sequence itself, chromatin accessibility was suspected to have a role[3,4], and distal, cell type-specific regions of open chromatin had already been identified far from genes' promoters[5]. However, these distal sites had not yet been shown to affect gene expression.

In 1981, these concepts culminated in the first demonstration of a non-coding DNA sequence that 'enhanced' the expression of a gene encoded in *cis*, in a manner that was distinct from transcriptional activation mediated by promoters[6,7]. Specifically, on an episomal reporter vector, a non-coding region of the simian virus 40 (SV40) genome increased expression at a distance remote from the reporter gene's promoter

Box 1 | **A history of operational definitions of enhancers**

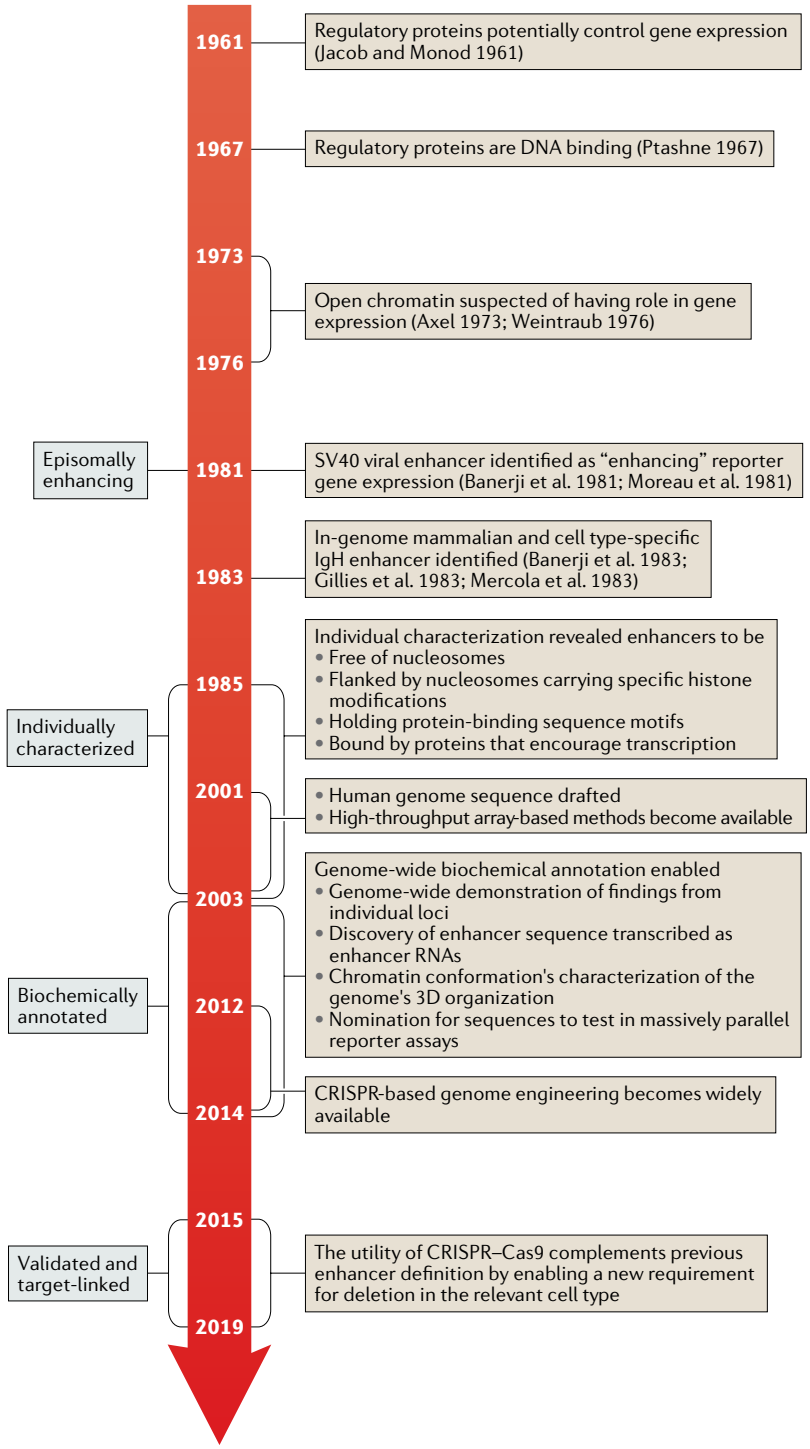
Operational definitions of an enhancer — that is, the practical criteria by which enhancers are distinguished from other sequences — have varied over time. This box summarizes the emergence of different operational definitions of enhancers (see the timeline figure).

'Episomally enhancing'
The first reported enhancer sequence was described as a non-coding sequence that could enhance the expression of a *cis*-encoded reporter gene (FIG. 1b). The enhancer demonstrated activity from a number of locations on the same plasmid, both upstream and downstream of the promoter and in either orientation[6,7].

'Individually characterized'
New enhancers were discovered and experimentally characterized on a one-by-one basis. Shared features of such enhancers included that they were free of nucleosomes; were flanked by nucleosomes with transcriptional-activity-associated histone modifications; contained sequence motifs for transcription factors (TFs) and were bound by these TFs; and likely were accessing target promoters by looping in 3D space. Highly expressed genes located within the vicinity of the enhancer and exhibiting similar cell-type specificity were inferred to represent the target gene[224,225].

'Biochemically annotated'
Biochemical features associated with enhancers were measured genome-wide in selected cell types and tissues. These were used to annotate and define cell type-specific enhancers on a genome-wide basis, generally without demonstration of enhancing activity. Enhancers were additionally found to be transcribed (enhancer RNAs) and to be enriched for 3D proximity to putative target promoters. Massively parallel reporter assays (MPRAs) began to enable the scalable generation of supporting functional data similar to the original enhancer-defining work of Banerji et al.[7].

'Validated and target-linked' (proposed)
With the emergence of CRISPR–Cas9 genome engineering, we propose that to reach the highest level of support as an enhancer, distally located elements should meet three criteria: first, deletion from its native genomic context results in altered expression of a potential target gene; second, evidence for a *cis*-acting mechanism; and last, one line of orthogonal evidence that the underlying sequence is an enhancer (in the form of either a reporter assay or biochemical annotation) (FIG. 4).



and independent of the enhancing region's orientation. From this experiment came the original definition of an enhancer, which is still widely quoted today: "the transcriptional enhancer element could act in either orientation at many positions . . . [even] downstream from the transcription initiation site"[7]. A few years later, using a similar in vitro method, the first endogenous, mammalian, cell type-specific enhancer was identified within the IgH locus[8–10]. A few years after that, endogenous regulatory sequences were shown to have in vivo activity, enhancing the expression of the cancer-inducing large T antigen in a cell type-specific manner[11].

The genomic characteristics of typical enhancers were further fleshed out over the ensuing decade, and a few general principles emerged. First, enhancers are free of nucleosomes, as measured by hypersensitivity to DNase I[12,13], but are flanked by nucleosomes with specific, transcription-associated histone modifications[14–16].

Second, enhancers contain clusters of TF binding motifs[17], and binding of TFs to these motifs underlies their enhancing activity[18,19]. Third, enhancers are likely to loop in 3D space into proximity with their target promoters[20,21].

Like many concepts in biology, these generalizations were based on a handful of examples that were studied in depth using the tools available at the time. One paradigmatic example, then and now, is the mammalian β-globin locus control region (LCR), a non-coding region that controls the developmental timing of expression of a cluster of globin genes. First discovered as a distally located deleted region in patients with β-thalassaemia who lacked mutations impacting the coding region of the β-globin gene[22–24], the β-globin LCR was hypersensitive to DNase I[12], contained motifs corresponding to relevant TFs (for example, GATA1)[25,26], was bound by these TFs[18] and was proposed to loop in 3D space so as to regulate globin genes[27]. Of note, the pattern of evolutionary conservation of the β-globin LCR — for example, in mouse[28], rabbit[29], goat[30] and chicken[31] — critically supported its functional dissection.

Notwithstanding such exemplars, relatively few enhancers had been identified by the late 1990s, orders of magnitude fewer than the number of genes known at the time. Unlike genes, enhancers could not be identified by expressed sequence tag (EST) sequencing, and, moreover, they lacked a defined grammar that supported their assignment as being actually functional (for example, an open reading frame). Indeed, the dearth of discussion of distal regulatory elements in the initial report of the human genome illustrates the difficulty of this task at the time[32,33].

One encouraging point was that nearly all the enhancers that had been deeply characterized at the time were evolutionarily conserved[34]. Taking a 'conservation first' approach, Loots and colleagues[35] identified non-coding regions regulating several interleukin genes by comparing 1 Mb of mouse–human orthologous sequences[35]. The global application of this strategy was one of the key motivations for the sequencing of the mouse genome[34,36,37].

However, immediately upon comparing the human and mouse genomes, the field faced the opposite problem, as the number of conserved non-coding regions, each a potential regulatory element, now vastly exceeded the number of genes[38–41]. How many of these conserved non-coding regions represented bona fide enhancers, as opposed to other kinds of functional elements? A further challenge was that the Human Genome Project had revealed the number of human genes to be about the same as the number of genes in the nematode, *Caenorhabditis elegans*. If the greater complexity of mammalian development was instead encoded by enhancers, a belief that took hold at the time and persists today, they too required cataloguing and characterization. In what cell types and at what developmental time points is each enhancer active? Which genes does each enhancer regulate?

To these and other ends, in the wake of the Human Genome Project, the field immediately shifted its attention to the genome-scale characterization of the epigenome — for example, through the Encyclopedia of DNA Elements (ENCODE) Consortium and similar projects. To briefly summarize an immense amount of work, genome-wide chromatin accessibility was measured by DNase I hypersensitivity[42–45], DNA methylation by bisulfite sequencing[46–48], and genome-wide histone modifications[49–51] and TF binding[52–54] by chromatin immunoprecipitation. Each such biochemical assay was coupled to a genome-wide readout, initially microarrays and subsequently massively parallel DNA sequencing[55]. In a surprising finding, whole-transcriptome RNA sequencing revealed the transcription of active enhancers ('eRNAs')[56–59]. Altogether, over the past 15 years, such biochemical methods have been applied in order to characterize the non-coding genome in hundreds of mammalian cell types and tissues[58,60–63]. This has resulted in the cataloguing of over one million candidate *cis*-regulatory elements with enhancer-like signatures; these collectively span ~16% of the human genome[64]. It is now widely recognized that a major component of the heritability of nearly all common diseases partitions to these regions, and in particular to regions that have enhancer-like signatures in disease-relevant cell types[65].

What defines an enhancer?

In the current parlance of the field, the term 'enhancer' is often used interchangeably to refer to: first, DNA sequence elements that meet the original Banerji et al.[7] (1981) definition —that is, enhancing transcription in a reporter assay; second, DNA sequence elements that bear biochemical marks associated with enhancer activity; or third, endogenous, distally located DNA sequence elements that serve to enhance the transcription of a *cis*-located gene, in vivo and in their native genomic context. But these definitions are not equivalent. There may be sequences that activate transcription in the context of a reporter assay but do not meaningfully do so in vivo. There may also be sequences that bear enhancer-associated biochemical marks but do not actually function as enhancers in vivo. Finally, there may be in vivo enhancers that are non-canonically marked or that have contextual dependencies that are not maintained in a reporter assay.

Which definition should we use? In our view, the first two are *operational definitions*, whereas the last is a *biological definition*. An operational definition is not what an enhancer *is*, but rather follows from the practical framework that we use to distinguish biological enhancers from other sequences. Much like blind men inspecting an elephant[66], operational definitions are a means to characterize a phenomenon, but they fall short of the phenomenon itself. Here, we use the term 'enhancer' to refer to the in vivo phenomenon — that is, short regions of DNA that in their endogenous genomic, cellular and organismal context bind proteins that increase the likelihood of transcription of one or more distally located genes through a *cis*-regulatory mechanism. We acknowledge that our viewpoint is not universally shared — subsets of the community may prefer to define enhancers as sequences exhibiting enhancing activity in an in vitro reporter assay, and indeed we ourselves have slipped into this definition in the past[67]. However, particularly as new assays proliferate, the

Expressed sequence tag
(EST). In the early days of genomics, shotgun sequencing of cDNA was used as an efficient strategy for discovering genes, and subsequently to quantify their relative abundance.

Open reading frame
The portion of a gene that is translatable by a ribosome; these are relatively straightforward to annotate by sequence alone, due to the required start and stop codons.

Regulatory element
A functional non-coding DNA sequence that regulates transcription; classes of regulatory elements include enhancers, promoters, silencers and insulators (further defined in BOX 2).

elephant itself (that is, the biological enhancer) must remain the primary focus, rather than the angle at which we first bumped into it.

Indeed, the operational definitions used for enhancers have been anything but static (BOX 1). The original operational definition, from Banerji et al.[8], was relatively simple: sequences that increase the expression of a reporter gene, when sequence and reporter gene are co-located on an episome[7]. However, this definition was quickly followed by efforts to characterize the biochemical features of such sequences in their native genomic and cellular contexts[8–10]. Our present understanding is that enhancers are bound by cell type-specific TFs, are associated with regions of open chromatin and are flanked by histones carrying H3K27ac and/or H3K4me1 modifications. They interact with their cognate promoters in 3D space and can be latent, primed or active[68,69]. Although the endogenous distributions of enhancer sizes and enhancer–gene distances remain important topics for exploration, a typical enhancer is probably hundreds of base pairs in length[70,71] and acts over a few to tens of kilobases[72]. Although we have many clues, the mechanistic details of how enhancers activate the expression of their target genes have yet to be fully worked out (FIG. 1a).

Particularly in recent years, operational definitions of enhancers based on biochemical annotations have been solidified by the ENCODE Consortium, as these are available throughout the genome and across many cell types. For example, enhancer-associated biochemical features have been used for regulatory variant effect prediction in the context of both rare[73] and common[74,75]

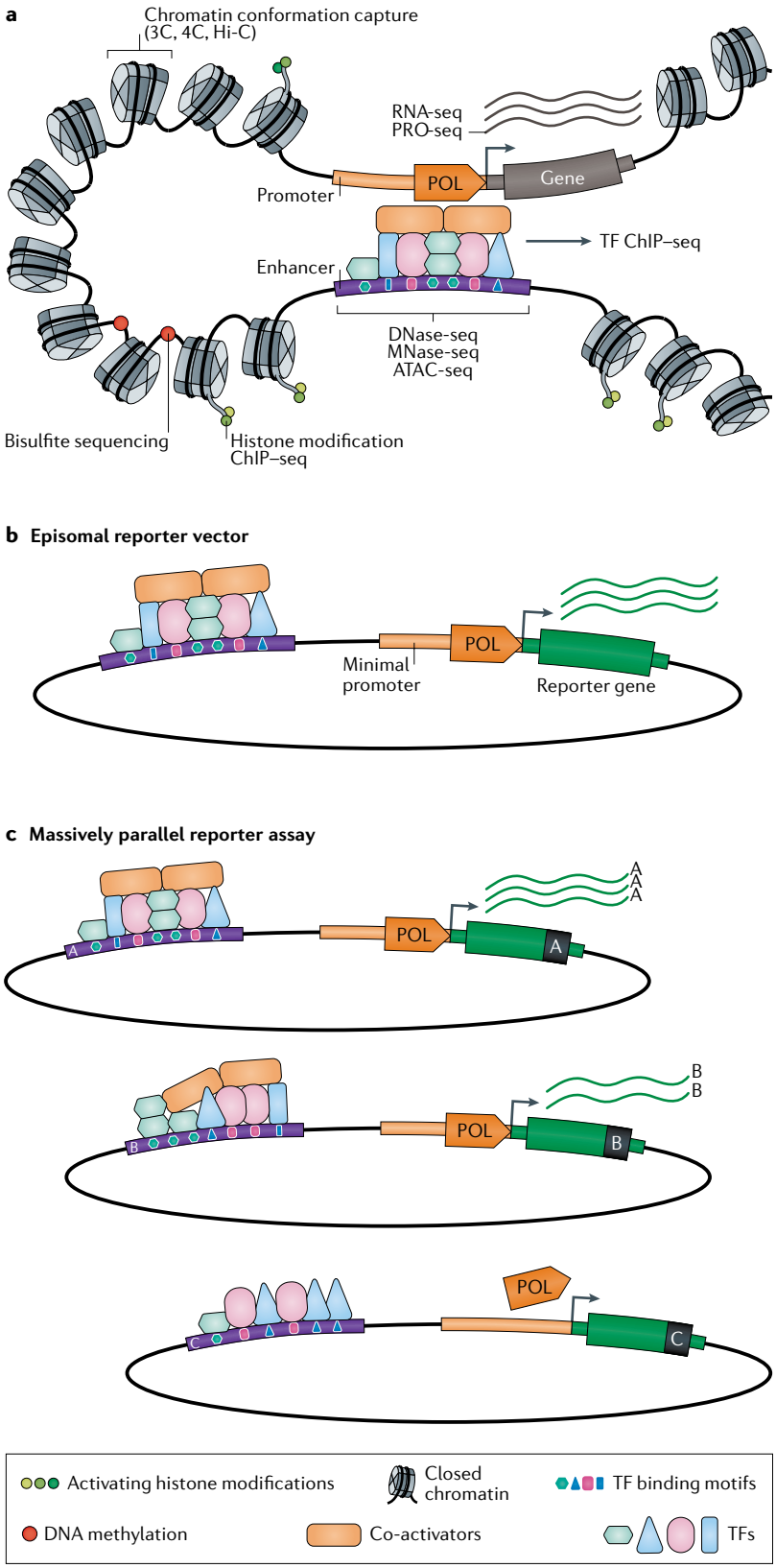


Fig. 1 | **Approaches for identifying, validating and characterizing enhancers. a** | Biochemical annotations of candidate enhancers: schematic depiction of an enhancer and a target gene, marked with the biochemical annotations used to nominate candidate enhancers and other features of non-coding DNA. Although the enhancer has been depicted in 3D proximity to its target promoter, we note that the mechanistic importance of such enhancer–promoter proximity is far from settled. We refer the reader to the Emerging approaches for biochemical annotation: 3D conformation mapping section for a discussion of open questions concerning enhancer–promoter communication and the importance of chromatin looping. **b** | Episomal reporter assay: a candidate enhancer and a reporter gene located in *cis* on an episomal vector. The candidate enhancer may increase expression of the reporter gene by recruiting transcriptional machinery. The degree of enhancer-mediated activation is measured by the abundance of reporter transcripts or the quantity of the reporter-encoded protein. **c** | Massively parallel reporter assays (MPRAs): many candidate enhancers can be interrogated simultaneously in a reporter assay if a barcode is encoded in the reporter transcript. The relative abundance of barcodes can be used to estimate the relative activities of the candidate enhancers to which they are linked. We show here just one of the many formats of MPRAs that have been developed. 3C, chromosome conformation capture; 4C, chromosome conformation capture on a chip; ATAC-seq, assay for transposase-accessible chromatin using sequencing; ChIP–seq, chromatin immunoprecipitation followed by sequencing; DNase-seq, DNase I hypersensitivity sequencing; MNase-seq, micrococcal nuclease digestion combined with sequencing; PRO-seq, precision run-on sequencing; POL, RNA polymerase; RNA-seq, RNA sequencing; TF, transcription factor. Part **a** is adapted from REF.[69], Springer Nature Limited.

Chromosome conformation capture
(3C). Methods that map the 3D positioning, looping and spatial organization of DNA within the nucleus, often relative to other segments of DNA.

CRISPR
Clustered regularly interspaced short palindromic repeats. A system that consists of the components of a bacterial immune system that have been adopted for synthetic genetic perturbation. The term is most often used in reference to the Type II Cas9 endonuclease version, which can introduce a double-stranded break into genomic DNA as directed by a synthetic guide RNA.

Topologically associating domains
(TADs). Broad regions of genomic DNA that are physically packaged together in the nucleus in 3D space, typically at a scale from hundreds of kilobases to several megabases.

disease. In human genomics, characterizations of the genome-wide sizes and distributions of enhancers rely heavily on biochemical datasets[51,76]. Many investigators are careful to qualify catalogues based on such annotations as 'predicted' or 'candidate' enhancers, but the qualification is often dropped, and such sequences simply referred to as 'enhancers'.

However, much like in vitro reporter assays, a definition based purely on biochemical annotations has clear limitations. First, biochemical annotations are based on observations made in a sequence's native genomic context but usually obtained on highly derived cell lines or on tissues that represent mixtures of many cell types. Second, although these measurements may correlate with function, they fall short of demonstrating regulation of the expression of a *cis*-encoded gene. In fact, it remains entirely possible that many biochemically identified enhancers may not be enhancing the transcription of anything[77]. Third, biochemical annotations fail to specify which genes a putative enhancer regulates, let alone the degree of activation conferred[78,79]. Fourth, many enhancer-associated biochemical features may have nothing to do with the enhancers' mechanism of action. For example, the MLL3/4 complex has been shown to serve as an essential co-activator at some enhancers, completely independent of its catalytic activity as an H3K4me1 writer[80,81]. Fifth, the coarseness of many biochemical features (for example, broad peaks) fails to resolve which specific subsequence and nucleotides underlie any enhancing function. Finally, such annotations are often used in a 'one size fits all' manner, potentially disallowing bona fide enhancers that are non-canonically marked.

We do not mean to say that operational definitions of enhancers, whether from a reporter assay or based on biochemical features, have been anything less than tremendously useful. However, we should be continually evolving towards a framework for discovering, characterizing and validating enhancers that is as close as possible to the biological phenomenon itself. To this point, new methods have recently emerged that overcome many of the key limitations of earlier technologies. These include single-cell ('sc') methods to identify cell type-specific open chromatin in complex tissues[82,83], higher-resolution chromosome conformation capture ('3C') methods to more finely map enhancer–promoter contacts[84], MPRAs to dissect or trap enhancer activity on a reporter vector at scale[85] and high-throughput CRISPR screens to directly perturb enhancers in their native genomic context and link them to their target genes[86].

The rapid maturation of these technologies should force us to re-examine how we operationally define enhancers. At the same time, given the heterogeneity of both the biochemical and functional methods that can now be applied at scale, it is important to acknowledge that this is going to be a complicated task.

What features identify an enhancer?

Enhancers are only one class of non-coding DNA regulatory element, although they are widely presumed to be the most numerically prevalent (BOX 2). For this Review, we focus on enhancers, and mammalian enhancers in particular, although many of the assays and concepts described are potentially applicable to other classes of non-coding DNA regulatory elements.

Enhancers are 'punctate' relative to broader chromatin domains (for example, chromosomes, topologically associating domains (TADs) and sub-compartments of TADs[87]), but their in vivo functionality is dependent on both the chromatin context in which they reside[88] and the *trans* milieu (for example, cell type-specific TFs)[89]. How do enhancers enhance the expression of their target genes? The classic model is that enhancers recruit cell- and condition-specific TFs and then loop in 3D space to interact with their target promoter[90]. The recruited TFs directly or indirectly (for example, via a co-activator) facilitate chromatin remodelling and recruitment of the basal transcriptional machinery at the promoter (FIG. 1a), thereby enhancing transcription[91]. However, it should be emphasized that this is not an inexorable chain of events. For example, stimulus-responsive enhancers may exhibit open chromatin and 3D interactions with their promoters before activation[92,93]. The production of a functional mRNA is a complex process, and which steps are rate-limiting varies by gene and context[94]. Mammalian promoters are typically suboptimal in one or several ways[95]. Thus, from a mechanistic perspective, enhancers might tune transcription levels by affecting any number of steps. For example, some enhancers were recently shown to regulate the release of promoter-proximal paused

Box 2 | **Other types of regulatory element**

Enhancers are only one class of non-coding DNA regulatory element. This is a brief list of other major classes, perhaps unified in that they generally correspond to open chromatin in cell types in which they are active. These are not covered in detail in this Review, but many of the assays described in this Review could also be applied to these other classes of regulatory elements. Furthermore, because all these elements share many biochemical features and/or functional characteristics, the lines between them can be blurry[59,226,227].

Promoter
An element that initiates transcription of a gene by RNA polymerase, by definition located at the 5′ end of the gene and encompassing its transcription start site (TSS). Composed of transcription factor (TF) binding sites that generally act independently of orientation, as well as core promoter elements (for example, the TATA box) that tend to be oriented relative to the TSS[228]. Most promoters contain only a few discernible core promoter elements[229].

Silencer
An element similar to an enhancer but that acts to reduce expression of a target gene. It tends to bind repressive TFs[230].

Insulator
A boundary element that restricts the ability of positive (enhancer) or negative (silencer) regulatory elements to modulate the expression of genes located on the other side of the boundary. Often bound by CTCF[231].

***Cis*-regulatory elements**
Elements that regulate a target gene by a mechanism that depends on their residing on the same chromosome or episome. Regulatory elements located on the same chromosome and within 1 Mb of their target gene[232] are often assumed to act through a *cis*-regulatory mechanism.

***Trans*-regulatory elements**
Elements whose regulation of a target gene is mediated by a *trans*-acting factor. Regulatory elements located over 1 Mb from their target gene on the same chromosome, or that are located on a different chromosome, are often assumed to act through a *trans*-regulatory mechanism.

RNA polymerase II[96], and others to act through splicing-dependent mechanisms[97]. Further heterogeneity can be introduced by the same enhancers acting via different co-regulators at different times[98].

Regardless of any such mechanistic heterogeneity, a common property is that the activity of individual enhancers is generally cell type-specific, or even condition-specific[99], and this specificity is a function of the expression levels of the TFs that are able to bind to it[89]. But even this generality is complicated by the fact that the capacity of an expressed TF to interact with an enhancer may depend on the chromatin state of the region in which the enhancer resides[100], which is in turn a function not only of a cell's present state but also of its developmental history. An enhancer's specificity may also depend on the nature of the TF — for example, whether it is a pioneer factor[101]. Finally, there are multiple models of how enhancers interact with their target promoters, including tracking, linking/chaining, short- or long-range looping, transcription factory, and hub/condensate models (reviewed in REF.[102]), more than one of which may be correct.

Well-established enhancers bear biochemical marks that are now routinely used to classify other sequences as enhancers. These include sequence-level features (TF binding site motifs and conservation); 1D biochemical annotations (accessible chromatin; H3K27ac and H3K4me1 modifications on flanking histones, for active enhancers; H3K4me1 and HK27me3, for poised enhancers; or closed chromatin that has been pre-marked by H3K4me1, for primed enhancers[103]); direct binding of TFs or secondary binding of cofactors (such as p300); and 3D biochemical annotations (nuclear spatial proximity to promoters, as measured by 3C, 3C on a chip (4C), 3C carbon copy (5C) or Hi-C) (reviewed in REF.[69]). Through projects such as ENCODE, these annotations underlie the classification of over one million candidate regulatory elements in the human genome as potential enhancers in one or more cell types[60,64].

Yet, none of these features serves as a perfect rule for identifying endogenous enhancers, as counterexamples can be found for each one. Not all distal conserved elements are detectably enhancers[39], and far more of the gene-distal non-coding genome is annotated as a regulatory element than is conserved[60,104]. TF sequence motifs alone are poorly predictive, as only a small fraction of the potential TF binding sites in the genome are typically bound in a cell type where the TF is expressed[68]. Although it is enriched, histone modification and cofactor binding is also not completely predictive of enhancer activity[54,58,72,77,105]. Furthermore, to the extent that functional activity has been measured at scale — for example, via MPRAs — its correlation with the annotations typically used to call enhancers is modest at best[106–108]. Many enhancers are spatially proximate to their target promoter in 3D[109,110], but exceptions have been described[111–114]. Genes can be affected by a single enhancer or multiple enhancers acting in concert[115]; conversely, individual enhancers can regulate multiple genes[72]. Some enhancers reside in clusters of a handful[116] to even hundreds ('super enhancers'[76,117]), whereas many are solo. At least a few enhancers reside at great distances from their target gene (for example, the ZRS enhancer, located 1 Mb from the *Shh* gene[118], and a *MYC* enhancer located 1.7 Mb downstream[119]), although most are much more proximal to their target promoters[72]. Enhancers regulating housekeeping genes may act via distinct sets of TFs and cofactors relative to the enhancers regulating developmentally specific genes[79,120]. Enhancers may have complex relationships with promoters, including feedback loops or competition with neighbouring genes[121,122].

In light of this heterogeneity, using a 'one-size-fits-all' set of annotations to catalogue enhancers seems problematic. Furthermore, as per their biological definition, enhancers are ultimately defined not by biochemical marks but by their endogenous functional activity: increasing the likelihood of transcription of one or more distally located genes through a *cis*-regulatory mechanism. It is also worth emphasizing that ruling out that a sequence is a biological enhancer may be far more difficult than proving that it is. This is simply because it would be extremely impractical to test every possible developmental time point, cell type and condition.

As we touched on above, technologies for functionally characterizing non-coding regulatory elements at scale are rapidly evolving. This creates an opportunity to rethink our operational definition of enhancers. In the next several sections, we review current and emerging technologies for the scalable characterization of enhancers and consider the evidence that each provides (TABLE 1).

Methods for scalable enhancer characterization

Current technologies and their limitations

DNA sequence. A primary sequence is modestly informative for distinguishing where enhancers lie. Evolutionary conservation can support the functional candidacy of a region[39], but not all enhancers are conserved[123–126]. Surveying a genome or candidate regulatory elements for TF binding motifs can add further support[127], but not all motifs are known or perfectly described[128]. Furthermore, the presence of a motif for an expressed TF does not mean that it is bound, and, even if it is, not all binding is functional[68]. Consequent to these limitations, automatic sequence-based enhancer annotation is helpful and worthwhile[129], but it performs modestly for predicting enhancers and the contexts in which they are active. A further limitation is that the primary sequence cannot identify the genes an enhancer regulates, beyond predictions based purely on linear proximity.

Biochemical annotations. Biochemical annotations that correlate with enhancer activity and are measurable on a genome-wide scale include assays for histone modifications or TF binding (for example, chromatin immunoprecipitation followed by sequencing (ChIP–seq) or cleavage under targets and release using nuclease (CUT&RUN)), open chromatin (for example, DNase I hypersensitivity sequencing (DNase-seq), micrococcal nuclease digestion combined with sequencing (MNase-seq) or assay for transposase-accessible chromatin using sequencing (ATAC-seq)), DNA methylation (for example, bisulfite sequencing), and the initiation and

Pioneer factor
A TF that can directly interact with compact, closed chromatin; this class of TFs are thought to initiate ('pioneer') chromatin remodelling events.

Table 1 | **Pros and cons of various strategies for identifying, validating and/or characterizing enhancers**

Type	Technologies	Single cell?	Pro	Con
Conservation	PhyloP[233], PhastCons[234]	Not applicable	Computable genome-wide; support for critical function	Not cell type-specific; not a measurement of enhancer activity; no target gene identified
Sequence motif	Databases: JASPAR[235], HOCOMOCO[236]	Not applicable	Computable genome-wide; informative as to potentially bound proteins	Limited cell-type specificity; not a measurement of enhancer activity; no target gene identified
Open chromatin	DNase-seq[43], MNase-seq[237], ATAC-seq[238]	Yes (for example, sci-ATAC-seq[82])	High-throughput biochemical annotation; associated with enhancer activity; cell type-specific	Not a measurement of enhancer activity; no target gene identified; unknown specificity
Transcription	RNA-seq, PRO-seq[239], GRO-cap[59], CoPro[240]	Yes (for example, scRNA-seq, although usually only mRNAs)	High-throughput biochemical 'eRNA' annotation; implies active RNA polymerase near enhancer	Transcription does not necessarily guarantee enhancer activity; no target gene identified
Histone marks	Enhancer-associated histone modifications on ChIP–seq	Emerging (for example, scChIC-seq[241])	High-throughput biochemical annotation; can support poised, active or silenced enhancers; cell type-specific	Not a measurement of enhancer activity; no target gene identified; unknown specificity
Protein Binding	Transcription factor ChIP–seq, CUT&RUN[242]	Emerging (for example, uliCut&Run[170])	High-throughput biochemical annotation; cell type-specific	Not a measurement of enhancer activity; no target gene identified; unknown specificity
eQTL	Many datasets available (for example, GTEx Consortium[136])	Emerging (for example, sc-eQTLGen Consortium[138])	In-genome; direct measurement from human tissues; can test all variants by all transcripts	Limited to common genetic variants; variants fall in linkage disequilibrium blocks
3D proximity	Chromatin conformation 'C's (for example, Hi-C[141], microscopy)	Yes (for example, microscopy, sci-Hi-C[163])	High-throughput biochemical annotation; cell type-specific; informs enhancer–gene links	Not a measurement of enhancer activity; unknown specificity
3D proximity + live imaging	Microscopy[172]	Yes, microscopy is inherently single-cell	Live cells, dynamic imaging of 3D proximity and transcriptional bursting across time	Limited to a small number of loci at once
3D proximity + biochemical annotation	ChIA-PET[147], HiChIP[148], DNase-Hi-C[150], PLAC-seq[149]	None yet	High-throughput biochemical annotation; cell type-specific; informs enhancer–gene links; more cost-effective than Hi-C	Not a measurement of enhancer activity; unknown specificity
Computational prediction	For example, ChromHMM[243], Segway[244]	Yes (for example, Cicero[134])	Computable genome-wide; potentially cell type-specific, can nominate enhancer–gene links	Requires experimental functional validation
Reporter plasmid activity	Luciferase, MPRAs[67,173,184], lentiMPRAs[190]	None yet	High throughput; relatively straightforward to implement; provides functional support	Episomal; removed from genomic context; no target gene identified; unknown specificity
Single-gene CRISPR screens	'Indel' scans[195], long-deletion scans[203,204], CRISPRi scans[105,208]	None yet	High throughput; in native genomic context; provides functional support; informs enhancer–gene links	Only tests candidate enhancers against one gene at a time; unknown sensitivity
Whole-transcriptome CRISPR screens	Mosaic-seq[216], multiplexed scRNA-seq[72]	Yes	High throughput; in native genomic context; provides functional support; informs enhancer–gene links; many genes at a time	Currently only implemented using epigenetic perturbation; unknown sensitivity
In vivo model organism: transgenic reporter	Episomal or transgenic delivery[219,221]	None yet	In vivo test across many developmental contexts	Low throughput; does not test enhancer in native genomic context
In vivo model organism: sequence deletion	Direct genomic sequence deletion[115]	None yet	In vivo test across many developmental contexts; potential detection of organismal phenotypes	Low throughput; not all enhancers are conserved between mouse and humans

ATAC-seq, assay for transposase-accessible chromatin using sequencing; ChIA-PET, chromatin interaction analysis with paired-end tag sequencing; scChIC-seq, single-cell chromatin immunocleavage sequencing; ChIP–seq; chromatin immunoprecipitation followed by sequencing; CoPro, coordinated precision run-on and sequencing; ChromHMM, a chromatin state annotator based on hidden Markov models; CRISPRi, CRISPR-based transcriptional interference; CUT&RUN, cleavage under targets and release using nuclease; DNase-seq, DNase I hypersensitivity sequencing; eQTL, expression quantitative trait locus; eQTLGen, eQTL Genetics Consortium; eRNA, enhancer RNA; GTEx, genotype–tissue expression program; GRO-cap, cap-enriched global nuclear run-on sequencing; HOCOMOCO, *Homo sapiens* Comprehensive Model Collection; indel, insertion or deletion; lentiMPRAs, lentiviral MPRAs; MNase-seq, micrococcal nuclease digestion combined with sequencing; MPRAs, massively parallel reporter assays; PLAC-seq, proximity ligation-assisted ChIP–seq; PRO-seq, precision run-on sequencing; RNA-seq, RNA sequencing; sc, single-cell; sci, single-cell combinatorial indexing; uli, ultra-low input.

abundance of transcription (for example, precision run-on sequencing (PRO-seq) or RNA sequencing (RNA-seq)) (FIG. 1a). Through the ENCODE Consortium and related efforts, such data have been collected in diverse cell types and tissues, to inform the cataloguing of cell type-specific enhancers. Although this effort is unquestionably useful, it remains unknown what proportion of candidate enhancers identified solely by biochemical marks are technical false positives[130–132] or products of having enhancer-like biochemical features but no meaningful impact on the expression of *cis*-encoded genes[133]. Furthermore, '1D' biochemical annotations fail to inform us which genes an enhancer regulates (biochemical annotations based on 3D 3C techniques are discussed further below). To some degree this can be overcome by correlative approaches (for example, correlating open chromatin status between promoters and enhancers across large numbers of cell types), but such links remain inferential[45,134,135].

eQTL mapping. Expression quantitative trait locus (eQTL) studies in human populations can be used to validate and characterize distally located candidate regulatory elements. In brief, genome-wide genotypes in human cohorts (measured by microarrays and imputation or by genome sequencing) are tested for correlation with the expression of genes located in *cis* (measured by bulk RNA-seq of an accessible tissue from those same individuals). Variants that are significantly associated with gene expression differences after appropriate corrections are called as eQTLs. The eQTL framework is very powerful, and, for variants residing within distally located candidate enhancers, it can provide in vivo validation of those enhancers while also linking them to their target genes[136]. On one hand, given the diversity of epigenomic contexts traversed during development, eQTL studies may represent our only hope for comprehensively observing the consequences of human enhancer disruption (as all engineered mutations will be in models such as cell lines, organoids or mice). On the other hand, the framework has clear limitations, including its reliance on naturally occurring human genetic variation (most enhancers do not harbour common variants that substantially perturb their activity), linkage disequilibrium (multiple variants in a haplotype block may equivalently explain an association) and restriction to cell types and tissues that can be practically obtained from large numbers of individuals for expression profiling (for example, peripheral blood mononuclear cells)[137,138].

Emerging approaches for biochemical annotation

3D conformation mapping. A long-hypothesized model of enhancers involves their looping in 3D space in order to access target promoters[139,140]. In recent years, successively more powerful 3C methods have yielded high-resolution 3D conformational maps of the human genome in a few cell types (FIG. 1a). With 3C methods, genomic DNA fragments are ligated to other, physically proximate genomic DNA fragments within the nucleus[141,142]. The resulting datasets have led to the identification of large-scale compartments of genome organization at various scales, including A/B compartments[141], TADs[143–146] and possibly enhancer–promoter loops[109]. 3C methods have also been paired with biochemical assays so as to enrich for potentially functional interactions — for example, such methods include chromatin interaction analysis with paired-end tag sequencing (ChIA-PET)[147], HiChIP[148], proximity ligation-assisted ChIP–seq (PLAC-seq)[149], DNase Hi-C[150] and others.

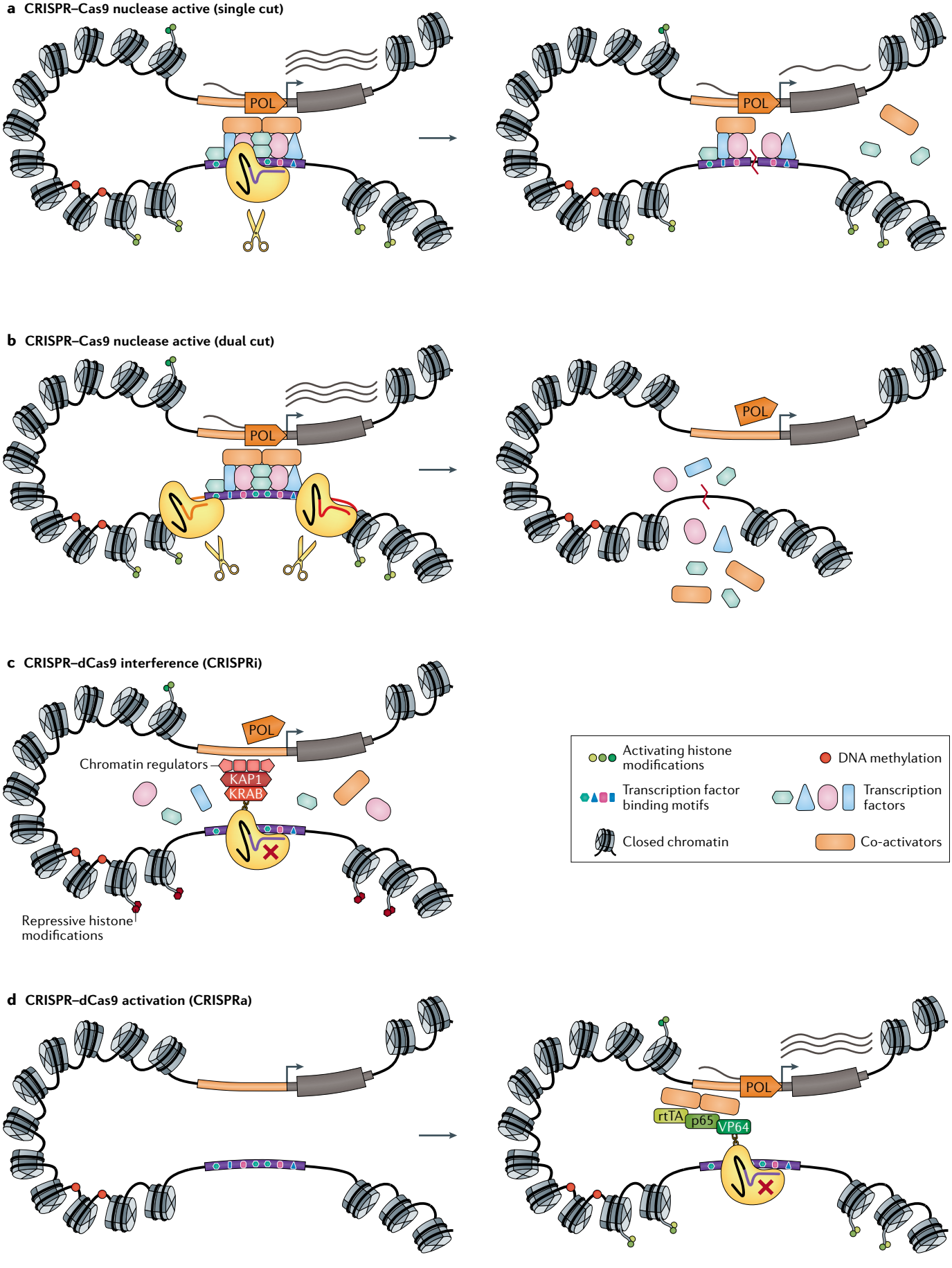
Does physical proximity strongly predict enhancer–gene links? Is it necessary and/or sufficient? In an elegant recent study that relied on live imaging, sustained proximity of an enhancer to its target was indeed required for activation[151]. Furthermore, a strong signal for distal chromatin interactions in bulk genomic assays such as Hi-C is associated with tissue-specific, presumably enhancer-dependent, expression[152]. On the other hand, proximity is sometimes maintained even when the gene or enhancer is inactive[112,153]. Other studies have found enhancer mobility, rather than proximity per se, to be a key determinant of activation[154]. Finally, the temporary disruption of 3D loops on a genome-wide scale through cohesin depletion was found to have minimal lasting effect on gene expression[155,156]. Overall, the precise mechanistic relevance of 3D proximity to enhancer-mediated gene regulation remains unclear.

Single-cell molecular profiling. Conventional or 'bulk' biochemical assays of chromatin return the mean profile of their input cells, which due to Simpson's paradox is potentially representative of none of the cells therein[157,158]. Until recently, the field has dealt with cellular heterogeneity by either ignoring it or, where possible, resorting to physical dissection or cell sorting[61,159]. However, methods for profiling chromatin state in single cells are advancing quickly and have the potential to overcome this challenge. For example, single-cell ATAC-seq has enabled the in vivo profiling of accessible chromatin at the scale of a whole organism[82,83,160,161]. Single-cell MNase-seq, ChIP–seq, Cut&Run and Hi-C methods have also been developed[162–170]. As we touched on above, microscopy — the original single-cell method — has revealed cases in which enhancer–promoter proximity either is or is not required for gene activation[151,171]. A major advantage of microscopy relative to genomic assays is the ability to study dynamic gene regulation in live cells[172]. Although currently limited to studying one or a few loci at a time, methods for multiplexing at the interface of microscopy and genomics are rapidly advancing.

Overall, single-cell methods have the potential to replace conventional bulk 1D and 3D biochemical assays. From datasets such as these, links between enhancers and promoters can be potentially nominated by their correlation across large numbers of cells, rather than large numbers of samples[134]. Single-cell methods may also enable the identification of candidate enhancers that appear to be active in extremely specific developmental contexts, or heterogeneously active within a single cell type. However, like the biochemical annotations on which they are based, any such candidate links will still lack functional validation.

Linkage disequilibrium
The population genetics phenomenon by which genetic variants are nonrandomly associated within a population. Variants are said to be in 'linkage disequilibrium' if they are found to reside on a haplotype more frequently than one would expect by completely random assortment; variants in linkage disequilibrium are nearby on a genomic locus and hence are co-inherited because they are rarely separated through meiotic recombination.

Simpson's paradox
A phenomenon in statistics in which different trends may exist in subgroups of a dataset but are undetectable when the groups are analysed as a whole.


a CRISPR–Cas9 nuclease active (single cut)
POL
POL
b CRISPR–Cas9 nuclease active (dual cut)
POL
POL
c CRISPR–dCas9 interference (CRISPRi)
POL
Chromatin regulators
KAP1
KRAB
Repressive histone modifications
Activating histone modifications
DNA methylation
Transcription factor binding motifs
Transcription factors
Closed chromatin
Co-activators
d CRISPR–dCas9 activation (CRISPRa)
POL
rtTA
p65
VP64

◀ Fig. 2 | **CRISPR-based approaches for perturbing enhancers.** The CRISPR system has been repurposed for use with four main perturbation methods that can disrupt enhancer activity **a** | Single-cut small-sequence insertion or deletion (indel). An active CRISPR nuclease such as Cas9 is directed to make a single cut that, through inaccurate repair, will usually create a small indel of <10 bp. This indel can sometimes disrupt an enhancer's function — if, for example, it overlaps a key transcription factor (TF) binding site. **b** | Dual-cut long-sequence deletions. To guarantee that a perturbation will disrupt the enhancer's functional sequence, the entire enhancer can be deleted by directing two cuts, flanking on either side. In some cells, due to inaccurate repair, deletions may occur between the two cuts. However, this is inefficient and will be only one of several possible repair outcomes that must be accounted for in an experimental design. **c** | CRISPR interference (CRISPRi)-based epigenetic repression. The nuclease domain of the CRISPR enzyme is rendered inactive ('dead', such as dCas9) but is tethered to a repressive domain (for example, KRAB) that is known to disrupt enhancer activity and expression. **d** | CRISPR activation (CRISPRa)-based epigenetic activation. A dead CRISPR enzyme is tethered to an activating domain (for example, a fusion of VP64, p65 and rtTA) that can potentially induce activation of a target gene when it is targeted to a primed enhancer. POL, RNA polymerase.

Technologies for measuring enhancer activity

Massively parallel reporter assays. An MPRA tests the functional activity of thousands of candidate regulatory sequences in a single experiment. The typical set-up of MPRAs is very similar to the original demonstration of the properties of the SV40 enhancer — that is, position-independent activity within an episomal vector[7] (FIG. 1b). Although first developed in 2009 to dissect all possible single-nucleotide variants of a promoter[173], MPRAs have mostly been used to study enhancers (reviewed in REF.[85]). Enhancer-focused MPRAs involve cloning a library of candidate enhancers into a reporter vector, wherein they have the opportunity to enhance the expression of a reporter gene via a minimal promoter (FIG. 1c). Each reporter gene transcript includes a barcode that is associated with a particular enhancer (or is the enhancer itself, in the case of STARR-seq[174]). The relative abundance of each RNA barcode, normalized to its DNA-based representation, is used to quantify the activity of its cognate candidate enhancer[175].

A clear strength of MPRAs is their ability to simultaneously test large numbers of sequences for regulatory activity via a relatively straightforward, widely accessible toolkit (that is, oligonucleotide synthesis, molecular biology, cell culture and sequencing)[175]. MPRAs have been applied to assessing biochemically annotated candidate enhancers[77,176–178], candidate enhancers harbouring variants that potentially mediate eQTLs[179–182], and even scans of the entire human genome[108,183]. A major advantage of MPRAs is that the sequences to be tested can simply be synthesized, enabling straightforward saturation mutagenesis of enhancers[67,184,185], as well as programming synthetic enhancers in order to inform modelling of their properties[186,187]. In contrast with MPRAs that rely on re-synthesis of candidate sequences, genome-wide 'shotgun MPRAs'[108,174,183,188] nicely avoid a priori assumptions about which sequences to test.

However, at least as they are usually implemented, MPRAs remain limited by several factors, including the length constraints and cost of DNA synthesis or the immense complexity of shotgun libraries, the confounding effect of the reporter's minimal promoter, and the use of episomes whose chromatin may have different properties to that of the genome[189]. Specific types of MPRA can address these concerns, at least in part — for example, by integrating MPRA reporters into the genome[88,190,191]. The fact that MPRAs test each sequence of interest entirely out of context is, on one hand, a strength, as it isolates that sequence in order to study its properties independently of that context. However, this is also a weakness, in that the properties observed out of context may be irrelevant when that native context is restored. The fact that most MPRAs only test for enhancer activity using a single promoter, or at best a handful[79], could contribute to a high false-negative rate. To put it another way, most MPRAs assume that enhancers act in a promoter-generic fashion, when that in fact may not be the case. Conventional MPRAs also fail to capture how each sequence affects and is affected by its genomic neighbourhood, as well as which promoter an 'active' enhancer endogenously affects. Users of MPRAs, including ourselves, typically fail to confirm that each 'positive' element fully meets the original Banerji definition (that is, active in both orientations and from many positions).

CRISPR screens of non-coding sequences. An exciting recent development in this area has been the emergence of pooled CRISPR-based enhancer screens for in-genome perturbation (FIG. 2). These studies springboard off CRISPR-based genome-wide screens of genes[192–194], but instead with the aim of characterizing massive numbers of enhancers in their native genomic context. In brief, such screens entail the delivery of a library of enhancer-targeting guide RNAs (gRNAs) to a pool of cells, followed by a phenotypic assay that informs as to which of those gRNAs impact the expression of a target gene or genes. To date, all such screens have used Cas9-induced perturbations, including active Cas9 for sequence disruption[195] or nuclease dead Cas9 (dCas9) tethered to an epigenetic repressor[105] or activator domain[92]. Because these genetic or epigenetic perturbations of enhancers are phenotyped by methods that directly or indirectly measure gene expression, they have the potential to functionally link enhancers to their target genes at scale, potentially filling a longstanding gap in the field.

Nuclease-active genome-editing screens. The initial CRISPR screens of regulatory elements delivered an individual gRNA per cell[195]. The gRNA–Cas9 nuclease complex directed double-stranded breaks (DSBs) at target sites, which, after repair by error-prone non-homologous end-joining (NHEJ)[196,197], resulted in 1–10-bp deletions or 1-bp insertions ('indels') in as many as 90% of cells[192,194] (FIG. 2a). These were 'single-gene' screens, in that the experiments were designed to detect expression perturbations of a specific gene. The first such screen targeted gRNAs in order to effectively tile small indels across a known cluster of enhancers of *BCL11A*[195]. The authors flow-sorted the edited cells on the basis of the *BCL11A*-dependent switch to fetal haemoglobin, sequenced guides that were enriched in cells that had or hadn't switched and, on the basis of those enrichments, successfully identified a primate-specific GATA1 motif critical for that enhancer's function. A transcription-activator-like effector nuclease (TALEN)-mediated indel scan of the same enhancer revealed the same motif, albeit via a much lower-throughput

Saturation mutagenesis
A molecular biology technique in which all possible sequence changes are generated from a parental sequence (for example, all possible amino acids in an open reading frame, or all possible single-nucleotide variants in an enhancer).

experiment[198]. Additional single-locus CRISPR screens of regulatory elements quickly followed at larger scales[199], including experiments that perturbed thousands of candidate enhancers per experiment[133,200–202].

Non-coding CRISPR screens present challenges different from those of coding CRISPR screens. In a coding screen, the indels resulting from NHEJ at a single DSB are likely to result in a frameshift and in the gene's complete loss of function. However, the rules of disrupting enhancer function are more nebulous. Although small indels are probably capable of disrupting TF binding sites within an enhancer, they might only do so if they directly overlap the binding site itself. In this respect, the ability of single guide scans to fully 'tile' a region is limited by both the distribution of protospacer-adjacent motif (PAM) sites and the non-random distribution of NHEJ-mediated mutations. Furthermore, the disruption of a single TF binding site might be insufficient to detectably disrupt the function of an enhancer. To address all these technical challenges at once, other CRISPR screens of regulatory elements have sought to program larger deletions in order to increase effect sizes and facilitate more complete tiling of regions of interest[203–205] (FIG. 2b). Such 'long-deletion' scans deliver pairs of gRNAs per cell that target closely located sites, which can result in clean deletion of the intervening sequence. However, a challenge is that the farther apart the pair of cuts induced by the gRNAs, the less often full deletion occurs — for example, ~20% of the time for a 365-bp deletion[204].

In sum, although powerful, CRISPR screens of non-coding regulatory elements are currently limited by effect size, efficiency or both. Additional challenges include that the variability of NHEJ-mediated repair outcomes plagues these screens with unprogrammed editing outcomes[204,206], and that in non-haploid cells each allele of the targeted locus can be heterogeneously edited within each cell, complicating the interpretation of results.

Nuclease-inactive epigenome-editing screens. Relying on epigenetic perturbations, rather than genetic ones, bypasses many of these limitations — for example, allowing all alleles in a given cell to be more consistently perturbed. The dCas9–KRAB repressor domain (CRISPR interference, or 'CRISPRi') was the first construct shown to synthetically silence a target enhancer by inducing ~1–2 kb of repressive marks in the vicinity of the gRNA target[207] (FIG. 2c). CRISPRi has subsequently been used in multiple single-gene screens of regulatory elements[105,208,209]. Activating domains (dCas9–VPR or dCas9–p300) have also been used to scan for poised enhancers, in an approach termed 'CRISPR activation' or 'CRISPRa'[92,208] (FIG. 2d). Additional dCas9-tethered domains have been shown to disrupt enhancer activity (for example, the histone demethylase LSD1 (REF.[210]), histone deacetylase 3 (REF.[211]) and the DNA methylator MQ[212] or DNMT3A[213–215]), and these could potentially be adapted to large-scale screens.

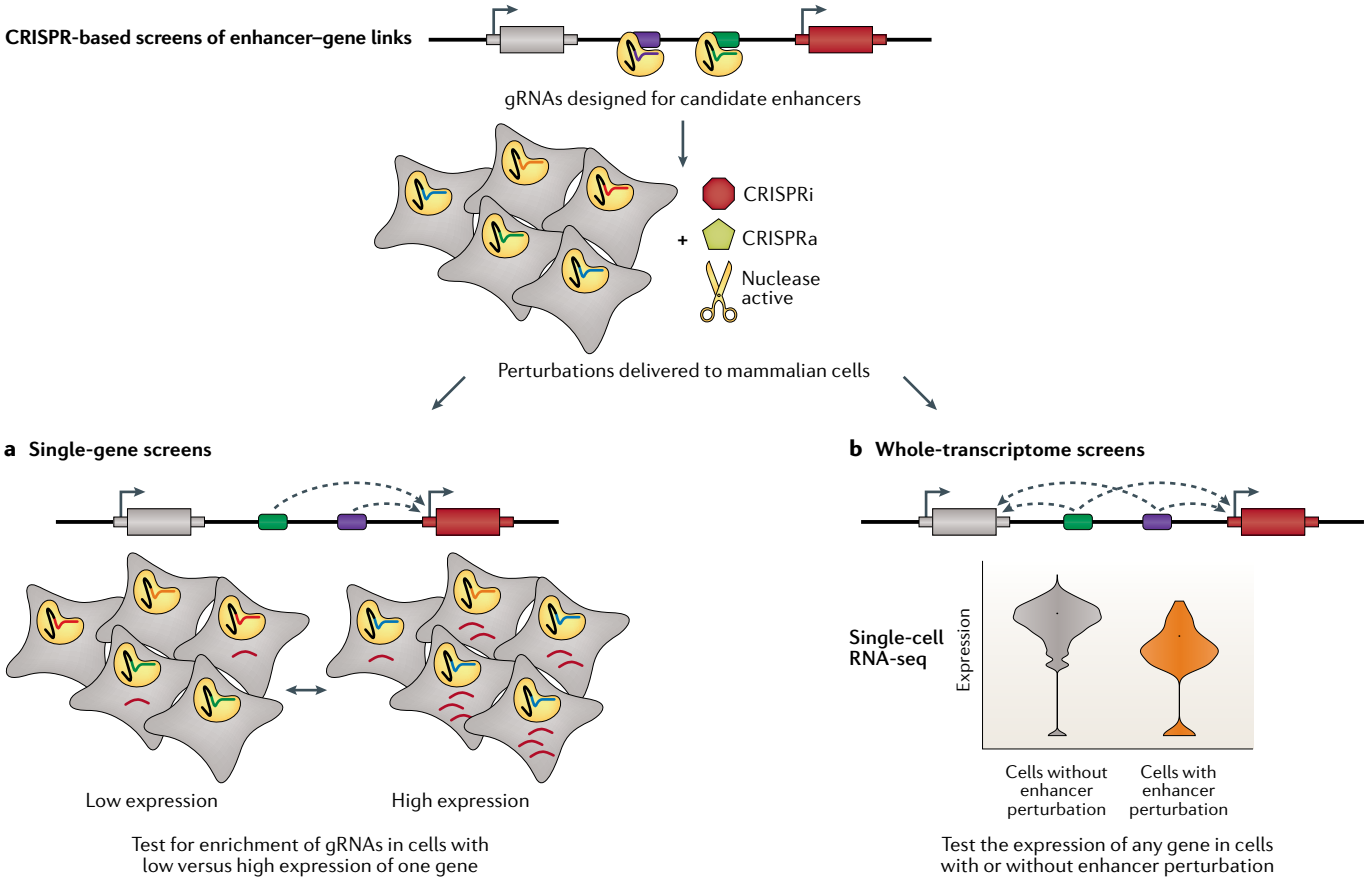
However, although nuclease-inactive epigenome scans of regulatory elements have some clear technical advantages, the synthetic nature of the perturbation leaves something to be desired. Although the epigenetic changes somewhat recapitulate how enhancers are physiologically turned on or off, the synthetic domains (for example, KRAB or VPR) used in a CRISPRi or CRISPRa system probably do not perfectly recapitulate the subtleties of enhancer regulation. This may lead to false positives (for example, through the spreading of KRAB's repressive effects or through unnatural activation by VPR) or false negatives (for example, an active enhancer that is not susceptible to CRISPRi-mediated inactivation). By contrast, wholesale deletions of candidate enhancers are unambiguously disruptive of a bounded region.

Whole-transcriptome screens. A shared limitation of single-gene screens, whether by CRISPR, CRISPRi or CRISPRa, is that the phenotyping is restricted to one or a few genes per experiment — for example, by engineering a reporter to the target gene[133,201,203,208], by labelling mRNA products with fluorescence in situ hybridization (FISH)[209] or by focusing on drug-responsive[202,204], antibody-detectable[92] or proliferation-related[105,200] genes. Each such phenotyping assay requires a specific technical set-up, which sharply limits its scalability and ease of adoption (FIG. 3a).

Towards genome-wide functional maps of enhancer–gene interactions, several groups have developed 'whole-transcriptome' screens of regulatory elements, which circumvent the need for gene-specific assays to be developed (FIG. 3b). In brief, a library of enhancer-targeting gRNAs and some form of Cas9 is still introduced to cells, but the phenotyping is performed by single-cell RNA-seq (scRNA-seq) of both mRNAs and gRNAs. The subsets of cells with versus without each gRNA are then tested for expression differences. The first such screen delivered 1 CRISPRi perturbation per cell, targeting 71 candidate enhancers across 7 genomic loci[216]. As scRNA-seq is costly and individual enhancers most likely regulate only 1 or a few genes in *cis*, we developed a related approach, wherein ~28 gRNAs were introduced per cell, enabling 5,779 candidate enhancers to be evaluated in a single experiment[72]. However, even with extensive multiplexing, such experiments are still expensive. For such screens to become routine, greater multiplexing and/or further reductions in the cost of scRNA-seq will be needed. Furthermore, such multiplex screens may be limited to epigenetic perturbation, particularly if large numbers of DSBs are toxic to cells.

Future prospects for CRISPR-based screens of non-coding sequences. Within just a few years, CRISPR and CRISPRi screens of non-coding elements have delivered clear progress in terms of validating enhancers in their native context while also linking them to their target genes. However, technical improvements are needed, and many questions remain (FIG. 3c). For example, validating each screen-based 'hit', such as by deleting it outside of a screen, remains challenging[204,206] but should probably be the standard expectation for strong claims about enhancer functionality (see Defining and cataloguing enhancers for further discussion of this point). Also, because the number of unambiguous 'positive control' enhancer–gene links remains small, the false-negative rates for these scans by and large remain unknown.

Protospacer-adjacent motif (PAM). In the original CRISPR bacterial immune system, fragments of previously encountered viral DNA are preserved in the bacterial genome; these 'remembered' sequences are processed into RNAs that guide the CRISPR nuclease to destroy newly invading viral DNA. But, to prevent the nuclease from destroying the matching 'remembered' sequence in the bacteria's own genome, a motif (the PAM) is required next to the target sequence in the viral genome. When genome editing is performed in eukaryotic cells, the presence of this sequence is still required by CRISPR nucleases.



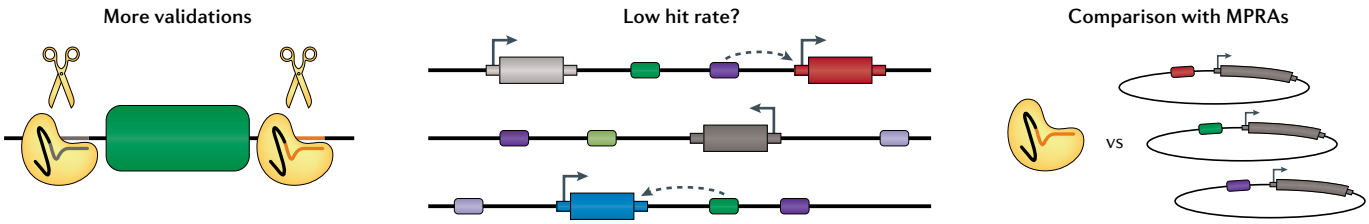


Fig. 3 | **CRISPR-based screens of enhancer–gene links.** In all such screens, guide RNA (gRNA)-based perturbations are designed for candidate enhancers and are delivered to mammalian cells as a pool. **a** | In most screens, cells are separated by the expression of a single or a few genes, and perturbations are tested for enrichment in high- or low-expression bins. **b** | In 'whole-transcriptome' screens, single-cell RNA sequencing (RNA-seq) is used to evaluate the expression of any gene against each perturbation. **c** | The future of such screens would benefit from higher standards (and better methods) to validate the screen results (for example, by deletion of individual elements), investigating why all such screens have had a low 'hit rate' thus far, and comparison of their results with massively parallel reporter assay (MPRA) readouts of activity. CRISPRa, CRISPR activation; CRISPRi, CRISPR interference.

In the vein of the latter concern over false negatives, one of the larger surprises of these studies has been that relatively high proportions of the biochemically or MPRA-supported candidate enhancers tested do not detectably influence the expression of a *cis*-encoded gene, in both CRISPR and CRISPRi screens, and even when assaying the whole transcriptome (for example, ~90% in REF.[72]). How should this be interpreted? Potential explanations include that epigenetic perturbations of enhancers have a high false-negative rate, for technical reasons; scRNA-seq fails to detect subtle changes in gene expression; shadow enhancers are buffering regulatory effects[217]; most screens to date have been in terminally differentiated, stable cell lines whose lack of dynamics masks any regulatory effects; and finally, analogous to the early estimates of the total number of human genes, there are many fewer bona fide enhancers than biochemical and MPRA-based annotations would have us believe.

On the other side of the balance sheet, putative enhancers identified by CRISPR screens may fail to show activity in MPRAs. Are such instances false positives in

Shadow enhancers
Redundant enhancers, often located far away from their target gene; enhancer redundancy is thought to enable robust buffered expression of the target gene and to provide a versatile platform for the evolution of new regulatory functions.

the CRISPR screens, or false negatives in the MPRAs? Of note, most conventional MPRAs utilize a single promoter for the reporter, which may not be sensitive to all enhancers[79]. Also, some established mechanisms of enhancer–gene interaction, such as high physical mobility[154] or weak interaction networks[218], may not translate well to an MPRA context. Finally, MPRAs will fail to recapitulate complex gene–enhancer networks[121,122]. Considerable further work will be necessary to differentiate between these and other potential explanations.

Technologies for in vivo validation

All the aforementioned methods (biochemical annotations, MPRAs, CRISPR screens and so forth) are performed in vitro on cell lines and therefore are only capable of accessing a limited number of biological contexts. As we discussed above, eQTL studies are powerful for assessing in vivo effects but are limited in critical ways. Consequently, the mouse model will remain a crucial asset for the validation and characterization of human enhancers for the foreseeable future.

First, transgenic reporter assays continue to provide valuable information regarding the tissue specificity of candidate enhancers[219]. The advantages of in vivo transgenic reporter assays include that a much broader range of developmentally and physiologically relevant contexts are 'accessed' than will ever be possible in in vitro systems and that the sequences tested have experienced the natural developmental history of these contexts, rather than being transfected or transduced into already-differentiated cells. The disadvantages of in vivo reporter assays are similar to those of MPRAs, including that elements are tested outside of their native genomic context and that the elements are not linked to their endogenous target genes.

Second, CRISPR technology has recently made it much more straightforward to delete genomic sequences in the mouse, enabling new insights into aspects such as enhancer redundancy[115] and the consequences of disrupting TADs[220]. Although observing phenotypic changes consequent to in vivo manipulation of an endogenous regulatory sequence is a powerful paradigm, a first disadvantage is that if the goal is to understand human enhancers, then such studies may be restricted to elements conserved across mammals. Furthermore, the organismal phenotypic defects caused by deleting regulatory elements can be subtle and challenging to detect[221]. Finally, similar to in vivo transgenic reporter assays, in vivo deletion of candidate enhancers will be challenging to scale beyond a handful of sequences. Despite these limitations, we envision that both murine in vivo CRISPR deletion and transgenic reporter assays of selected elements will be critical for benchmarking the validity of any emerging catalogue of functionally characterized human enhancers.

Defining and cataloguing enhancers

Ever since the Human Genome Project, a natural goal for the field of genomics has been to generate a catalogue of human enhancers. Indeed, this is one of the primary goals of the ENCODE Consortium, which has generated the vast majority of the aforementioned biochemical annotations. However, for such a catalogue to be both comprehensive and maximally useful, it should not simply comprise a list of sequences believed to be enhancers on the basis of biochemical annotations from cell lines and tissues, except perhaps in its very initial form. Rather, our goal should be to apply emerging, scalable biochemical and functional assays in order to generate a considerably more useful catalogue.

Spurred by efforts including ENCODE-4 and the Human Cell Atlas, developments that we anticipate within the next few years include the following. First, single-cell profiling, of chromatin accessibility, histone marks, TF binding and 3D conformation, will yield genome-wide catalogues of enhancer-associated biochemical marks for nearly all human cell types, from tissues obtained in vivo and from nearly all developmental stages. Second, MPRAs will be applied in order to comprehensively test candidate regulatory elements in representative cell types, quantifying the transcriptional activation potential of each element in a uniform context. Third, CRISPR screens will be applied to these same candidates in these same cell types, validating a subset of elements in their native genomic context while also revealing the targets of enhancer regulation. Finally, the number of elements tested in mouse models, either by transgenic reporters or CRISPR-mediated deletion, will continue to grow as well, albeit at a much slower rate.

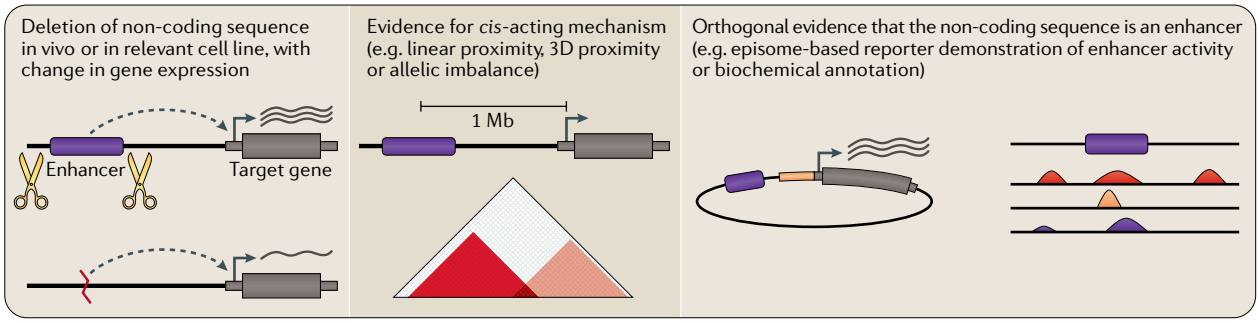
On the one hand, these developments are encouraging. They move us closer to a comprehensive catalogue of functionally supported human enhancers that is well annotated in terms of the cell types in which each element is active, the genes that each element regulates, the degree of activation each element confers and so forth. On the other hand, as compared with the current practice, in which putative enhancers are operationally identified often solely on the basis of biochemical marks, future enhancer catalogues are likely to be more nuanced. For example, it will probably more often than not be the case that specific elements are supported by some, but not all, forms of evidence. Which are we to interpret as the ground truth?

As a starting point for dealing with this anticipated heterogeneity, we propose a relatively straightforward framework for how to describe the level of support for candidate enhancers. This framework is illustrated in FIG. 4. At the very top is a new 'validated and target-linked' operational definition of enhancers, wherein a non-coding sequence has a demonstrated effect on a specific target gene's expression in its endogenous context. To be more specific, validated and target-linked enhancers would meet the following evidentiary criteria.

First, targeted deletion of the element in its native genomic context should result in altered expression of a distally located target gene. Deletion of the candidate enhancer in vivo or in a cell line should result in a measurable, reproducible change in the expression of one or more target genes. This would provide strong functional evidence that the sequence in question actually performs a regulatory function, while also linking it to at least one gene that it regulates. The deletion could be of only the one element, or possibly in combination with

ENCODE-4
The fourth generation of projects funded by the Encyclopedia of DNA Elements (ENCODE) Consortium, begun in 2017 and including a new component focused on the implementation of high-throughput functional assays.

Human Cell Atlas
An international scientific community to coordinate the generation of human single-cell datasets, with the goal of generating a reference map of every cell type in the human body.



Strong (All required)	Epigenetic genomic perturbation (but not deletion) in cell line with *cis* change in gene expression	AND	Episome-based demonstration of activity	AND	Enhancer-associated biochemical annotations
Moderate (2 of 3 required)	Genomic perturbation in cell line with *cis* change in gene expression	AND/OR	Episome-based demonstration of activity	AND/OR	Enhancer-associated biochemical annotations
Weak (One required)	Only genomic perturbation-based support	OR	Only episome-based demonstration of activity	OR	Only enhancer-associated biochemical annotations

Fig. 4 | **A tiered framework to describe the level of support for the enhancer candidacy of a non-coding sequence.** We propose 'validated and target-linked' support as the degree of evidence that we should be aiming for in cataloguing non-coding sequences as bona fide human enhancers. If the evidence falls short of that, as it currently does for nearly all candidate enhancers, we propose strong, moderate and weak tiers to describe candidate enhancers with less or conflicting evidence. The vast majority of candidate human enhancers are presently only weakly supported.

other deletions or perturbations, in order to unmask any redundancy.

Second, there should be evidence for a *cis*-acting mechanism. Perturbations of non-coding elements can have secondary effects, so there should be at least some rationale for concluding that an observed effect is mediated primarily by a *cis*-regulatory mechanism. This could simply be met through linear proximity between the candidate enhancer and its target gene (for example, <100 kb) or through other experimental data (for example, allelic imbalance or 3D proximity). Although they do not definitively demonstrate *cis* regulation, such lines of evidence at least support the possibility that the observed effects are not secondary or *trans*.

Third, there should be at least one line of orthogonal evidence that the sequence is an enhancer. Because it is plausible that a deleted sequence could influence the mRNA abundance of a *cis*-located gene through mechanisms other than serving as an enhancer, this criterion serves to add additional support. We propose that this evidence could come in the form either of the sequence episomally enhancing expression of a reporter gene on a plasmid (in accordance with the original 1981 definition[7]) or of enhancer-associated biochemical marks (in accordance with the operational definition of the ENCODE Consortium and its successors). The flexibility of this definition — that is, it requires one but not both of these lines of complementary support — allows for exceptions to the rule (for example, bona fide enhancers that do not function in reporter assays, or bona fide enhancers that bear non-canonical biochemical marks). Of course, these assays are correlated, so in many cases there will be agreement across the board.

We emphasize that we propose these as inclusionary, rather than exclusionary, criteria for defining enhancers. As we discussed above, it is very difficult to prove that a sequence is not an enhancer. Additionally, we also note that our definition may not be easily adaptable to candidate enhancers that overlap promoters or protein-coding regions.

For candidate enhancers that fall short of 'validated and target-linked' status, we propose three additional tiers (FIG. 4). 'Strongly supported' candidate enhancers should be supported by agreement of all three classes of experimental data — that is, biochemical marks, episomal reporter activity and CRISPRi/CRISPRa-based perturbation (but not necessarily deletion of the candidate enhancer, or else they would qualify as 'validated and target-linked'; see Nuclease-inactive epigenome-editing screens for related discussion). 'Moderately supported' enhancers would have support from two out of three of these, with the third being inconsistent, inconclusive or not performed. Finally, 'weakly supported' enhancers, a category that would presently apply to the vast majority of current human candidate enhancers, would be supported by only one of the three forms of evidence, with the other two being inconsistent, inconclusive or not performed.

We recognize that this scheme may be light in detail relative to the practical realities; for example, standards will be needed for how to threshold the datasets underlying each form of support, specific biochemical marks will need to be defined as enhancer-associated and so forth. However, particularly as the generation of such datasets accelerates, it seems critical that we have some framework in place for dealing with the inevitable heterogeneity in the confidence with which elements are

named as enhancers, in terms of both the kinds of assays being used and the results of those assays. In our view, the standard for declaring that an element is a biological enhancer should be better grounded in activity-based functional evidence, and the scheme in FIG. 4 is consistent with that goal. Furthermore, particularly because the functional dissection of trait-associated genetic variants from genome-wide association studies (GWAS) is likely to be a major focus of the field for the coming decade, it seems key that future efforts should prioritize the linking of enhancers to their target genes. Such links will necessarily accompany all 'validated and target-linked' and 'highly supported' enhancers according to the criteria above, as well as a subset of moderately and weakly supported enhancers.

Conclusions and future perspectives

Advances in scalable methods for the biochemical annotation and functional characterization of regulatory elements are paving the way to a comprehensive catalogue of human enhancers. In our view, such a catalogue can and should include knowledge of the cell type-specificity of each element, at least some degree of functional support for its role as a bona fide enhancer and knowledge of the element's target genes (FIG. 5). Such a catalogue could prove to be a critical resource for furthering our understanding of the human genome and its role in disease.

A first challenge to this goal is that it is already clear that the results of different types of assay will frequently disagree. How are we to explain the fact that the vast majority of biochemically nominated candidate enhancers, when perturbed by CRISPRi, do not result in detectable changes in the expression of genes located in *cis*[72]? As we touched on above, there are numerous credible technical and biological explanations for this observation, and distinguishing between them seems key to allowing the field to move forwards effectively. The broader point is that we remain largely in the dark regarding the sensitivity and specificity of most of these assays. Establishing a larger set of 'true positives' and 'true negatives' may be critical for adjudicating disagreements, which are trending towards being more prevalent than cases of agreement.

A further challenge is that although technologies are rapidly improving, it may simply not be realistic to test every candidate enhancer with every functional approach in every cell type of interest. However, as increasing numbers of elements are tested, our ability to quantitatively predict which sequences are bona fide enhancers, as well as the genes each regulates, is likely to improve as well. For example, machine-learning strategies to predict enhancer–gene links on the basis of 1D and 3D biochemical annotations are already advancing beyond the simple 'nearest gene' approach[134,209,222]. Particularly given the heterogeneous mechanisms by which enhancers might operate, establishing a community-accepted set of strongly supported enhancer–gene links, ascertained by relatively unbiased methods, seems key to calibrating the performance of such predictive tools.

We note that many of the challenges highlighted here apply not only to candidate enhancers but also to non-coding variants located within them. What standards of evidence should apply for a non-coding variant hypothesized to contribute to the association signal for

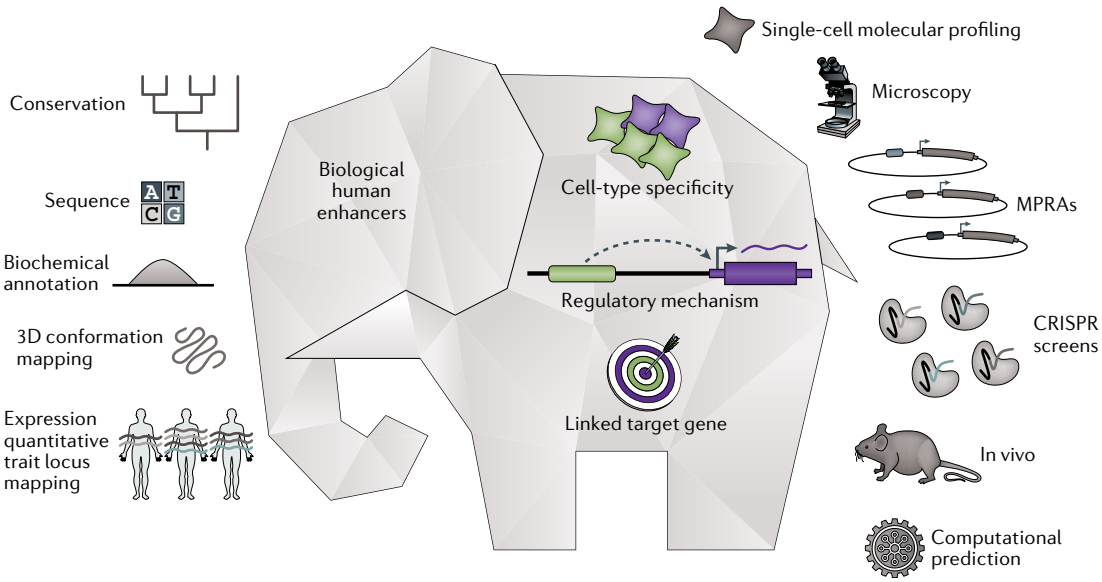


Fig. 5 | **The blind men and the elephant of human enhancer biology.** Much like blind men inspecting an elephant[66], operational definitions of enhancers are merely a means to characterize the underlying biological phenomenon, but they fall short of the phenomenon itself. As we work to develop a catalogue of bona fide biological enhancers, an updated operational definition that accommodates the heterogeneous and complementary results that are emerging from reporter assays, biochemical measurements and CRISPR screens will likely be necessary. In our view, the catalogue can and should aim to include knowledge of the cell-type specificity of each element, strong and multifaceted support for each element's role as a bona fide enhancer, and knowledge of each element's target genes. MPRAs, massively parallel reporter assays.

a common disease? Heterogeneous classes of data will be available for many variants (for example, biochemical annotations, molecular QTLs, computational predictions of variant effects, MPRAs and CRISPR perturbation), but they will not always agree. Further challenges include linkage disequilibrium, the possibility that more than one variant may contribute to a given association, the need to match the cell type in which functional characterization and/or biochemical annotation is being carried out with the disease in question, and the fact that the mechanisms by which non-coding variants exert their effects on disease risk remain unclear[223]. Although a definitive map of cell type-specific enhancers and enhancer–gene links will be critical in order to accelerate efforts to move beyond GWAS associations to causal variants and genes, it clearly will not be enough.

Additionally, for the purposes of this Review, and in line with how enhancers are broadly thought of in the field, we have focused on the modulation of gross transcript levels as an enhancer's primary activity of relevance. However, we should remain open to the possibility that many enhancer or enhancer-like sequences have more nuanced or tightly orchestrated effects, such as effects on splicing, subtle effects on the spatiotemporal unfolding of gene expression programmes during development, or other fine-grained effects. An evolving definition could also make room for surveys of enhancers' impacts on whole-cell or organismal phenotypes, although the effects on expression through which such effects were mediated would be important to know. Our overall point is that the operational definition of enhancers is likely to continue to evolve, alongside further advances in technology and biological understanding.

As we approach the 40th anniversary of their original definition[6,7], fascinating questions remain about enhancer biology. How does an enhancer pick its target gene? Is 3D chromatin structure a determinant of gene regulation, or a residual feature? How do individual enhancers coordinate within a regulatory circuit, and how widespread is redundancy within these enhancers? What constitutes the differences between the mechanisms underlying enhancer versus promoter activity? And last, what is (or are) the true precise mechanism (or mechanisms) of an enhancer's activity at a target promoter? Although it will not be enough, we anticipate that confidently identifying thousands of bona fide enhancers, ideally through some relatively unbiased method, will facilitate efforts to answer these questions, while also advancing our understanding of how this class of elements orchestrates the remarkable programme of mammalian development.

Published online 27 January 2020

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
2. Ptashne, M. Specific binding of the λ phage repressor to λ DNA. *Nature* **214**, 232–234 (1967).
3. 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).
4. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
5. Stalder, J. et al. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**, 451–460 (1980).
6. 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).
7. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981). **The first episomal observation of in vitro enhancer activity, this work coined the term 'enhancer'.**
8. Mercola, M., Wang, X., Olsen, J. & Calame, K. Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science* **221**, 663–665 (1983).
9. 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).
10. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**, 717–728 (1983).
11. Hanahan, D. Heritable formation of pancreatic β-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. *Nature* **315**, 115–122 (1985).
12. Tuan, D., Solomon, W., Li, Q. & London, I. M. The 'β-like-globin' gene domain in human erythroid cells. *Proc. Natl Acad. Sci.* **82**, 6384–6388 (1985).
13. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
14. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7**, 1395–1402 (1988).
15. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72**, 73–84 (1993).
16. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β-globin chromosomal domain. *EMBO J.* **13**, 1823–1830 (1994).
17. Serfling, E., Jasin, M. & Schaffner, W. Enhancers and eukaryotic gene transcription. *Trends Genet.* **1**, 224–230 (1985).
18. Ikuta, T. & Kan, Y. W. In vivo protein–DNA interactions at the β-globin gene locus. *Proc. Natl Acad. Sci. USA* **88**, 10188–10192 (1991).
19. Forsberg, M. & Westin, G. Enhancer activation by a single type of transcription factor shows cell type dependence. *EMBO J.* **10**, 2543–2551 (1991).
20. Ptashne, M. Gene regulation by proteins acting nearby and at a distance. *Nature* **322**, 697–701 (1986).
21. Müeller-Storm, H. P., Sogo, J. M. & Schaffner, W. An enhancer stimulates transcription in *trans* when attached to the promoter via a protein bridge. *Cell* **58**, 767–777 (1989).
22. Van der Ploeg, L. H. T. et al. γ-β-Thalassaemia studies showing that deletion of the γ- and δ-genes influences β-globin gene expression in man. *Nature* **283**, 637–642 (1980).
23. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosveld, F. G. β-Globin gene inactivation by DNA translocation in γβ-thalassaemia. *Nature* **306**, 662–666 (1983).
24. Driscoll, M. C., Dobkin, C. S. & Alter, B. P. γδβ-Thalassemia due to a de novo mutation deleting the 5′ β-globin gene activation-region hypersensitive sites. *Proc. Natl Acad. Sci.* **86**, 7470–7474 (1989).
25. Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. The β-globin dominant control region: hypersensitive site 2. *EMBO J.* **9**, 2159–2167 (1990).
26. Talbot, D., Philipsen, S., Fraser, P. & Grosveld, F. Detailed analysis of the site 3 region of the human β-globin dominant control region. *EMBO J.* **9**, 2169–2177 (1990).
27. Grosveld, F. et al. The regulation of human globin gene switching. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **339**, 183–191 (1993).
28. Moon, A. M. & Ley, T. J. Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. *Proc. Natl Acad. Sci. USA* **87**, 7693–7697 (1990).
29. Margot, J. B., Demers, G. W. & Hardison, R. C. Complete nucleotide sequence of the rabbit β-like globin gene cluster: analysis of intergenic sequences and comparison with the human β-like globin gene cluster. *J. Mol. Biol.* **205**, 15–40 (1989).
30. Li, Q., Zhou, B., Powers, P., Enver, T. & Stamatoyannopoulos, G. Primary structure of the goat β-globin locus control region. *Genomics* **9**, 488–499 (1991).
31. Reitman, M. & Felsenfeld, G. Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken β-globin locus. *Mol. Cell. Biol.* **10**, 2774–2786 (1990).
32. Dunham, I. et al. The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
33. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
34. Hardison, R. C., Oeltjen, J. & Miller, W. Long human–mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* **7**, 959–966 (1997).
35. Loots, G. G. et al. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000). **This study identifies the non-coding regions regulating several interleukin genes by comparing 1 Mb of mouse–human orthologous sequences. The global application of this strategy was one of the key motivations for sequencing of the mouse genome.**
36. Hardison, R. C. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**, 369–372 (2000).
37. Pennacchio, L. A. & Rubin, E. M. Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* **2**, 100–109 (2001).
38. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
39. Pennacchio, L. A. et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
40. Mouse Genome Sequencing Consortium et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
41. Dermitzakis, E. T. et al. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578–582 (2002).

42. Sabo, P. J. et al. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* **3**, 511–518 (2006).
43. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
44. Hesselberth, J. R. et al. Global mapping of protein–DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
45. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
46. Laurent, L. et al. Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320–331 (2010).
47. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
48. Cokus, S. J. et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
49. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
50. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
51. 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).
This study provides one of the earliest uses of genome-wide biochemical annotation datasets to annotate candidate enhancers.
52. Robertson, G. et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
53. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
54. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
55. Shendure, J. et al. DNA sequencing at 40: past, present and future. *Nature* **550**, 345–353 (2017).
56. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
57. Kim, T.-K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
58. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
This article reports an integrated analysis of cap analysis of gene expression (CAGE) datasets across hundreds of cell types and tissues performed to generate biochemical annotations of thousands of cell type-specific enhancers by their signatures of bidirectional transcription.
59. 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).
60. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
This report summarizes the wealth of data generated by the ENCODE Consortium. Hundreds of genome-wide datasets are used across hundreds of cell types and tissues to ascribe a function to the majority of the genome via biochemical annotation.
61. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
62. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1897 (2016).
63. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
64. ENCODE Project. SCREEN: search candidate regulatory elements by ENCODE. *ENCODE Project* http://screen.encodeproject.org/index/about (2019).
65. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
66. Contributors to Wikimedia projects. *Blind Men and An Elephant—Wikipedia* (Wikimedia Foundation, Inc., 2006).
67. Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
Along with Melnikov et al. (2012), this study is the first application of an MPRA to enhancer sequences.
68. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
69. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
70. Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60–63 (1998).
71. Li, L. & Wunderlich, Z. An enhancer's length and composition are shaped by its regulatory task. *Front. Genet.* **8**, 63 (2017).
72. Gasperini, M. et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390 (2019).
73. Smedley, D. et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* **99**, 595–606 (2016).
74. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
75. Gjoneska, E. et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
76. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
77. Wang, X. et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.* **9**, 5380 (2018).
78. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
79. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
STARR-seq utilizing different classes of promoters in *D. melanogaster* shows that enhancers often do not work in a promoter-generic fashion.
80. Dorighi, K. M. et al. Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* **66**, 568–576.e4 (2017).
81. Rickels, R. et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653 (2017).
82. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
83. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
84. Eagen, K. P. Principles of chromosome architecture revealed by Hi-C. *Trends Biochem. Sci.* **43**, 469–478 (2018).
85. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159–164 (2015).
86. Klein, J. C., Chen, W., Gasperini, M. & Shendure, J. Identifying novel enhancer elements with CRISPR-based screens. *ACS Chem. Biol.* **13**, 326–332 (2018).
87. Hnisz, D., Day, D. S. & Young, R. A. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188–1200 (2016).
88. Maricque, B. B., Chaudhari, H. G. & Cohen, B. A. A massively parallel reporter assay dissects the influence of chromatin structure on *cis*-regulatory activity. *Nat. Biotechnol.* **37**, 90–95 (2019).
89. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
90. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
This study shows compelling use of Hi-C contact maps to study the dynamics of enhancer–promoter interactions through cellular differentiation.
91. Thanos, D. & Maniatis, T. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100 (1995).
92. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
93. Ray, J. et al. Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proc. Natl Acad. Sci. USA* **116**, 19431–19439 (2019).
94. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192 (2009).
95. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917–922 (2006).
96. Chen, F. X. et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science* **357**, 1294–1298 (2017).
97. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
98. Murakami, S., Nagari, A. & Kraus, W. L. Dynamic assembly and activation of estrogen receptor α enhancers through coregulator switching. *Genes. Dev.* **31**, 1535–1548 (2017).
99. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).
100. Vihervaara, A. et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat. Commun.* **8**, 255 (2017).
101. Iwafuchi-Doi, M. et al. The pioneer transcription factor *foxa* maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
102. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
103. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
104. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
105. Fulco, C. P. et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
106. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
107. 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).
108. Muerdter, F. et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* **15**, 141–149 (2018).
109. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
110. Weintraub, A. S. et al. YY1 Is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588.e28 (2017).
111. Williamson, I. et al. Anterior–posterior differences in HoxD chromatin topology in limb development. *Development* **139**, 3157–3167 (2012).
112. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
113. Benabdallah, N. S. et al. Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484 (2019).
114. Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
115. Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
116. Levings, P. P. & Bungert, J. The human β-globin locus control region. *Eur. J. Biochem.* **269**, 1589–1599 (2002).
117. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
118. 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).
119. Bahr, C. et al. Author correction: A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **558**, E4 (2018).
120. Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
121. Cho, S. W. et al. Promoter of lncRNA Gene PVT1 Is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018).

122. Cinghu, S. et al. Intragenic enhancers attenuate host gene expression. *Mol. Cell* **68**, 104–117.e6 (2017).
123. Blow, M. J. et al. ChIP-seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010).
124. Schmidt, D. et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
125. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
126. Danko, C. G. et al. Dynamic evolution of regulatory element ensembles in primate CD4$^+$ T cells. *Nat. Ecol. Evol.* **2**, 537–548 (2018).
127. Kulakovskiy, I. V. et al. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* **41**, D195–D202 (2013).
128. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
129. Van Loo, P. & Marynen, P. Computational methods for the detection of *cis*-regulatory modules. *Brief. Bioinform.* **10**, 509–524 (2009).
130. 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. USA* **110**, 18602–18607 (2013).
131. Worsley Hunt, R. & Wasserman, W. W. Non-targeted transcription factors motifs are a systemic component of ChIP-seq datasets. *Genome Biol.* **15**, 412 (2014).
132. Jain, D., Baldi, S., Zabel, A., Straub, T. & Becker, P. B. Active promoters give rise to false positive 'phantom peaks' in ChIP-seq experiments. *Nucleic Acids Res.* **43**, 6959–6968 (2015).
133. Diao, Y. et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* **26**, 397–405 (2016).
134. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA Interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e8 (2018).
135. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
136. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
137. Strober, B. J. et al. Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019).
138. van der Wijst, M. G. P. et al. Single-cell RNA sequencing identifies celltype-specific *cis*-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
139. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683–689 (1988).
140. Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
141. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
142. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
143. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
144. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
145. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
146. Sexton, T. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
147. Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **107**, 30–39 (2009).
148. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
149. Fang, R. et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
150. Ma, W. et al. Fine-scale chromatin interaction maps reveal the *cis*-regulatory landscape of human lincRNA genes. *Nat. Methods* **12**, 71–78 (2015).
151. Chen, H. et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
152. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
153. Williamson, I., Lettice, L. A., Hill, R. E. & Bickmore, W. A. *Shh* and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**, 2994–3001 (2016).
154. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian *cis*-regulatory elements. *Science* **359**, 1050–1055 (2018).
155. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
Rapid cohesin depletion in cultured cells results in widespread loss of the genome's 3D organization, with minimal changes in gene expression. This has caused the field to question the importance of genome organization (that is, stable enhancer–promoter loops as they had been conceptualized).
156. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
157. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).
158. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
159. Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
160. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).
161. Cusanovich, D. A. et al. The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
162. Lai, B. et al. Publisher correction: principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **564**, E17 (2018).
163. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
164. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
165. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
166. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
167. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
168. Lee, D. S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
169. Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
170. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**, 1319–1329 (2019).
171. Benabdallah, N. S. et al. Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484.e7 (2019).
172. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
173. Patwardhan, R. P. et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
This study provides the first demonstration of an MPRA coupled to sequencing-based readout, used here to evaluate all possible single-nucleotide variants of bacteriophage and mammalian core promoters.
174. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
This is the first article to describe STARR-seq, as well as the first whole-genome shotgun MPRA. Quantitative, genome-wide maps of enhancer potential generated in two different *D. melanogaster* cell lines provide insight into the general characteristics of enhancers and enable analysis of cell-type specificity.
175. Gasperini, M., Starita, L. & Shendure, J. The power of multiplexed functional analysis of genetic variants. *Nat. Protoc.* **11**, 1782–1787 (2016).
176. Vockley, C. M. et al. Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell* **166**, 1269–1281.e19 (2016).
177. Vanhille, L. et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.* **6**, 6905 (2015).
178. Klein, J. C., Keith, A., Agarwal, V., Durham, T. & Shendure, J. Functional characterization of enhancer evolution in the primate lineage. *Genome Biol.* **19**, 99 (2018).
179. Ulirsch, J. C. et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530–1545 (2016).
180. Tewhey, R. et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **172**, 1132–1134 (2018).
181. Vockley, C. M. et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* **25**, 1206–1214 (2015).
182. Klein, J. C. et al. Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat. Commun.* **10**, 2434 (2019).
183. Liu, Y. et al. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol.* **18**, 219 (2017).
184. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271 (2012).
Along with Patwardhan et al. (2012), this study shows the first application of an MPRA to variants of enhancer sequences, in addition to coining the term 'massively parallel reporter assay'.
185. Kircher, M. et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat. Commun.* **10**, 3583 (2019).
186. 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).
187. Smith, R. P. et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* **45**, 1021–1028 (2013).
188. van Arensbergen, J. et al. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* **35**, 145–153 (2017).
189. Gilbert, N. & Allan, J. Supercoiling in DNA and chromatin. *Curr. Opin. Genet. Dev.* **25**, 15–21 (2014).
190. Inoue, F. et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* **27**, 38–52 (2017).
191. Akhtar, W. et al. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**, 914–927 (2013).
192. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
193. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
194. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
195. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
This study is the first CRISPR-based screen to identify functional non-coding sequence within an enhancer.
196. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
197. Chen, W. et al. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* **47**, 7989– 8003 (2019).
198. Vierstra, J. et al. Functional footprinting of regulatory DNA. *Nat. Methods* **12**, 927–930 (2015).
199. Wright, J. B. & Sanjana, N. E. CRISPR screens to discover functional noncoding elements. *Trends Genet.* **32**, 526–529 (2016).
200. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR–Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
201. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
202. Sanjana, N. E. et al. High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
203. Diao, Y. et al. A tiling-deletion-based genetic screen for *cis*-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629–635 (2017).

204. Gasperini, M. et al. CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192–205 (2017).
205. Aparicio-Prat, E. et al. DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genomics* **16**, 846 (2015).
206. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
207. Thakore, P. I. et al. Highly specific epigenome editing by CRISPR–Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
This study provides the first demonstration that nuclease-inactivated Cas9 tethered to the KRAB repressor and targeted to an enhancer can mediate repression of a target gene.
208. Klann, T. S. et al. CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561 (2017).
209. Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
210. Kearns, N. A. et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
211. Kwon, D. Y., Zhao, Y.-T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR–Cas9-based HDAC. *Nat. Commun.* **8**, 15315 (2017).
212. Lei, Y. et al. Targeted DNA methylation in vivo using an engineered dCas9–MQ1 fusion protein. *Nat. Commun.* **8**, 16026 (2017).
213. Vojta, A. et al. Repurposing the CRISPR–Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**, 5615–5628 (2016).
214. Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247.e17 (2016).
215. Huang, Y.-H. et al. DNA epigenome editing using CRISPR–cas suntag-directed DNMT3A. *Genome Biol.* **18**, 176 (2017).
216. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell* **66**, 285–299.e5 (2017).
This study reports the first use of single-cell RNA-seq to phenotype pools of candidate enhancer perturbations in a 'whole-transcriptome' screen.
217. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
218. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
219. 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).
220. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).
This article demonstrates that rearrangement across TAD boundaries can cause a pathogenic phenotype in humans; the study also uses CRISPR–Cas9 to effectively reconstruct human patient genome rearrangements in mouse models.
221. Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499.e15 (2018).
This report demonstrates that the knockout of ultraconserved enhancers in a mouse model had subtle but consequential effects on organismal development that were not readily detected by gross phenotyping.
222. Zeng, W., Wu, M. & Jiang, R. Prediction of enhancer–promoter interactions via natural language processing. *BMC Genomics* **19**, 84 (2018).
223. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
224. Gill, L. L., Karjalainen, K. & Zaninetta, D. A transcriptional enhancer of the mouse T cell receptor δ gene locus. *Eur. J. Immunol.* **21**, 807–810 (1991).
225. Greaves, D. R., Wilson, F. D., Lang, G. & Kioussis, D. Human CD2 3′-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell* **56**, 979–986 (1989).
226. Raab, J. R. & Kamakaka, R. T. Insulators and promoters: closer than we think. *Nat. Rev. Genet.* **11**, 439–446 (2010).
227. 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).
228. Weingarten-Gabbay, S. et al. Systematic interrogation of human promoters. *Genome Res.* **29**, 171–183 (2019).
229. Ngoc, L. V., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev.* **31**, 1289–1301 (2017).
230. Jayavelu, N. D., Jajodia, A., Mishra, A. & Hawkins, R. An atlas of silencer elements for the human and mouse genomes. Preprint at *bioRxiv* https://doi.org/10.1101/252304 (2018).
231. West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes. Dev.* **16**, 271–288 (2002).
232. Stranger, B. E. et al. Patterns of *cis* regulatory variation in diverse human populations. *PLoS Genet.* **8**, e1002639 (2012).
233. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
234. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
235. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
236. Kulakovskiy, I. V. et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).
237. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
238. 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. *Nat. Methods* **10**, 1213–1218 (2013).
239. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
240. Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet.* **50**, 1533–1541 (2018).
241. Ku, W. L. et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).
242. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
243. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
This study reports a computational framework for utilizing ChIP–seq of histone modifications to classify regions of the genome by their likely biological function (for example, strong enhancer or weak promoter).
244. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).

Acknowledgements
The authors thank S. Kim, C. Trapnell and S. Domcke, as well as other members of the Shendure Lab, for helpful discussions. J.S. is an investigator for the Howard Hughes Medical Institute.

Author contributions
M.G. and J.S. wrote the initial manuscript. M.G., J.T. and J.S. contributed to researching content for the article, discussing the content and reviewing/editing the manuscript before submission.

Competing interests
The authors declare no competing interests.

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