Shaping the nebulous enhancer in the era of high-throughput assays and genome editing

Edwin Yu-Kiu Ho[1*], Qin Cao[2*], Mengting Gu[3], Ricky Wai-Lun Chan[2], Qiong Wu[2,4], Mark Gerstein[3,5,6] and Kevin Y. Yip[1,2,7,8,9]

[1]Department of Biomedical Engineering, [2]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong
[3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA
[4]School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
[5]Program in Computational Biology and Bioinformatics, [6]Department of Computer Science, Yale University, New Haven, Connecticut, USA
[7]Hong Kong Bioinformatics Centre, [8]CUHK-BGI Innovation Institute of Trans-omics, [9]Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong

Abstract

Since the first discovery of transcriptional enhancers in 1981, their textbook definition has remained largely unchanged in the past 37 years. With the emergence of high-throughput assays and genome editing, which are switching the paradigm from bottom-up discovery and testing of individual enhancers to top-down profiling of enhancer activities genome-wide, it has become increasingly evidenced that this classical definition has left substantial gray areas in different aspects. Here we survey a representative set of recent research articles and report the definitions of enhancers they have adopted. The results reveal that a wide spectrum of definitions is used usually without the definition stated explicitly, which could lead to difficulties in data interpretation and downstream analyses. Based on these findings, we discuss the practical implications and suggestions for future studies.

Keywords

Enhancers, cis-regulatory elements, transcriptional regulation,

---

[*] Equal contributions

distance-independence, high-throughput reporter assays

Introduction

The concept of transcriptional enhancers originated from three papers published in 1981 that studied gene expression in Simian Virus 40 (SV40) [1–3]. These studies identified a 72bp repeat sequence around 150-200bp upstream of the early genes of SV40 that was necessary for their transcription. In one of the studies, it was further discovered that this "enhancer" element (term coined in Banerji et al. (1981)) could greatly enhance reporter gene expression at various distances from the gene no matter it was placed upstream or downstream of it [1]. The enhancer was also found to be *cis*-acting, in that covalent linkage to the recombinant was required for the stimulation of expression. Based on these initial findings, enhancers have since been commonly defined as *cis*-regulatory elements that can enhance the expression of target genes in a distance- and orientation-independent manner [4–7].

Research in enhancers has been accelerated in recent years by the emergence of new technologies for identifying and characterizing enhancers [8–14], quantifying enhancer activities in particular cell and tissue types [15–18], determining enhancer targets [19,20,29,21–28], and evaluating the functional consequence of genomic [30–37] or epigenomic [31,38–44] perturbation of enhancers in normal conditions and diseases. These new studies have greatly advanced current understandings of enhancers. Yet at the same time they also created questions about the classical definition of enhancers in various aspects, from the functional roles of enhancers and relationships between them and their targets to the necessary and sufficient defining features (Figure 1, to be discussed in detail below). The once clearly defined concept of "enhancers" now appears to be fairly ambiguous.

In this era of high-throughput assays and genome editing, it will soon become possible to experimentally test the enhancer function of every region in the genome in an unbiased manner without the need for a candidate set of putative enhancers defined *a priori* [45–47]. Accordingly, the focus of enhancer studies will be switched from identifying the genomic locations of enhancers and testing the functional roles of a small subset of them, which has been an active area of research in the past few years, to interpreting and mining the high-throughput testing data. Depending on how enhancers are defined, the resulting sets of enhancers identified from such data can be drastically different.

In the followings we discuss issues related to enhancer definitions that are causing substantial differences among the enhancers reported in different studies, with our arguments supported by a survey of a representative set of 54 research articles about experimental and computational identification of enhancers and their targets (Table S1). The survey results show that in these studies the same term “enhancer” can actually refer to very different concepts, and sometimes a precise definition is not explicitly provided. We also illustrate the issues by comparing several major sets of human enhancers.

Based on these findings, we discuss the practical implications of the ambiguity and possible efforts for alleviating the situation, which would be fundamental for interpreting the results of single studies and comparing across multiple studies.

For the more general topics of enhancer properties and their experimental and computational identification methods, readers are referred to separate reviews [6,48–56].

Promoters initiate, enhancers enhance?

Classically, promoters and enhancers are considered separate classes of regulatory elements. This distinction has been substantiated by epigenomic studies, which identified different characteristic histone modifications for promoters and enhancers [57,58]. In this classical view, promoters provide the platform for transcription factors, co-factors and RNA polymerase to initiate transcription at a nearby transcription start site (TSS), whereas enhancers enhance transcription by coming in contact with promoters through looping, although the detailed mechanisms have not been completely mapped out. This dichotomy is also seen in our survey, in which 46% of the research articles explicitly exclude annotated promoters from the definition of enhancers.

However, this classical view has been challenged by a number of recent studies in several ways [4,58–62] (Figure 1a).

First, enhancers have been reported to produce short non-coding RNAs in one or both directions [4,63–66]. These enhancer RNAs (eRNAs) have even been considered hallmarks of active enhancers and used to quantify enhancer activities [18]. In our survey, 13% of the articles made use of data produced by protocols that could detect eRNAs, and it is becoming a popular feature for identifying enhancers.

Second, some promoters have been shown to enhance the expression of reporter genes when inserted near them [67,68]. In addition, using CRISPR-Cas9 to study the effects of a large number of mutations, disruption of some promoters were also found to affect expression of other genes that can be far away [46,69,70].

These findings blur the line between promoters and enhancers. Is it then possible to distinguish them by the transcripts they produce? In terms of transcript types, just like protein-coding genes, many functional non-coding genes also have their own promoters [71,72]; on the other hand, though not as commonly, enhancers can also transcribe messenger RNAs by acting as alternative promoters [30,73]. In terms of transcript properties, in contrast to mRNAs transcribed from promoters, eRNAs are generally shorter, unspliced and non-polyadenylated. However, exceptions do exist, such as some polyadenylated eRNAs with lengths that are even longer than typical mRNAs [74]. There is thus not a simple demarcation of the types of transcripts that can be produced at promoters and enhancers, and their difference lies more on the tendency of producing one type of transcripts or another.

In terms of directionality and stability, eRNAs can be unidirectional or bidirectional, which weakly correspond to the more stable and less stable transcripts, respectively [74]. In the same way, promoters have also been proposed to produce transcripts in both directions, which usually lead to upstream anti-sense RNAs (uaRNAs), or promoter upstream transcripts (PROMPTs) that are less stable, and a more stable transcript going the other direction. Both the length and stability of transcripts from the two directions depend on the presence of nearby U1 snRNP recognition sites and poly(A) signals (PASs) [75–77]. Directionality and stability, therefore, do not provide a simple black-and-white rule for classifying promoters and enhancers either.

Given all these similarities, whether a functional sequence element should be called a promoter or an enhancer could be just a matter of degree [78]. This view is best illustrated by the H3K4me3-to-H3K4me1 ratio, which is commonly used to distinguish between promoters and enhancers [79,80]. Just as this ratio can take on any non-negative value, “promoter” and “enhancer” could simply be the names given to the two extremes of a continuous spectrum of sequence elements having transcriptional initiating and enhancing abilities. It remains not clear though whether a sequence element can act as both promoter and enhancer simultaneously in the same cell type or even in the same single cell.

Distance-, direction- and orientation-independent?

Another core component of the classical definition of enhancers is that its distance, direction and orientation with respect to the target gene do not affect its capability of enhancing the gene's expression. Specifically, direction concerns whether the enhancer is upstream or downstream of the target gene, whereas orientation concerns the DNA strand that contains the enhancer sequence (Figure 1b).

There are many examples of the distance-independent nature of enhancers, but it should be noted that a lot of these studies involve reporter constructs in plasmids, which limit the maximum distance between the enhancer and the reporter gene to a few kilobases [81]. In native genomic contexts, indeed there are cases in which the distance between an enhancer and its target gene can be as far as 1Mbp [82–84], but the overall distribution of enhancer-target distance is skewed toward the low end with higher interaction frequency for shorter distance. Studies based on ChIA-PET, Hi-C and computational inference of enhancer-target interactions have estimated the median distance between an enhancer and the genes it regulates to be around 60-125kbp depending on the cell type [85,86]. For example, in IMR90 cells, around 25% of enhancer-target pairs are within 50kbp, while around 57% are 100kbp or more apart [86]. This distance distribution is strongly governed by the three-dimensional structure of chromosomes, where a large portion of enhancer-target interactions happen within topologically associating domains (TADs) or chromatin contact domains (CCDs) that are on the scale of hundreds of kilobase pairs to several megabase pairs in size [26,86–89]. All these data suggest that in genomes, there is a strong tendency for an enhancer to regulate genes that are close to it, and there are physical constraints due to the chromosome architecture that make it difficult for an enhancer to regulate genes very far away.

In artificial constructs, an enhancer can indeed enhance expression of a gene at various distances from it, yet the level of expression induction could still vary. Based on a variety of examples from the first discovery of enhancers to the latest research, with a displacement of hundreds of base pairs to several kilobase pairs, the level of induction can change for many folds [3,90–95].

Interestingly, in our survey we also found very different distance thresholds were used in different studies. At one extreme, some studies defined the potential target gene of an enhancer as the closest one or no more than 100kb apart. At another extreme, some studies allow potential target genes of an enhancer to be as far as

5Mb or even imposing no distance limits.

In terms of direction, most studies do agree that an enhancer should function no matter it is upstream or downstream of its target genes. In our survey, only two studies required a particular direction of the enhancers relative to the genes they regulate, but that is due to the design of the experimental protocol (of using the transcription of the enhancer itself in the transcript as an indicator of enhancer activity) rather than a believe that the enhancers would function better in that direction.

As for orientation, some studies have defined the stringent requirement that a DNA sequence can be considered an enhancer only if it can enhance expression no matter it is inverted or not [96–98]. However, many examples have shown that inverting an enhancer can in fact substantially alter the expression of the target gene [98–104]. Nonetheless, this issue seems to be not commonly considered in our surveyed articles, with none of them specifying the orientation explicitly.

Cis-acting?

The terms "*cis*" and "*trans*" are among biological concepts with the most erratic definitions. In the context of enhancers, they could mean whether the enhancer and target gene are on the same molecule (e.g., chromosome) [4,105–109], whether they are close to each other based on a certain genomic distance threshold [110], or whether the function of enhancer depends on its DNA sequence alone ("*cis*-regulatory element") but not its products ("*trans*-acting factors") [111–114] (Figure 1c).

The first definition, that an enhancer is *cis*-acting if it works only for target genes on the same molecule, originated from the SV40 enhancer. It increased transcription of the target gene by hundreds of folds when the enhancer was inserted together with the target gene in the same plasmid, but had no effect when they were inserted into separate molecules [3,49].

On the other hand, a strong evidence that enhancers can also regulate genes on another molecule, and thus work "in *trans*" by this first definition, comes from studies in Drosophila. It was found that enhancers can regulate genes on the other homologous chromosome by a phenomenon known as transvection [115–119]. Recent works suggest that transvection is fairly common in Drosophila [120], and it

appears to be a general property of all Drosophila enhancers [109].

Inter-chromosomal enhancer-target interactions can happen in species other than Drosophila, and, although relatively rare, between non-homologous chromosomes [121,122]. The general mechanisms of inter-molecular enhancer-target interactions remain unclear, but some evidence suggests that the formation of protein bridges may play a role [123].

In our survey, among the relevant articles, 95% of research articles required potential targets of an enhancer to reside on the same chromosome or DNA molecule as the enhancer. These studies thus assumed that enhancers act in cis according to this first definition. However, many of these studies likely made this decision mainly for the technical purpose of reducing false positives, at the expense of a (likely small) number of missed targets on other chromosomes.

The second definition of "*cis*" and "*trans*" based on genomic proximity is commonly used in the discussion of effects of single nucleotide variants (SNVs) and expression quantitative trait loci (eQTL) at a certain distance from the genes they affect [124–126]. As discussed above, some recent studies have estimated the distribution of genomic distance between an enhancer and its target gene in specific cell types. In order to determine whether enhancers act in *cis* then, what it takes is drawing a distance threshold and computing the fraction of such interactions involving enhancer-target pairs below the threshold.

Different values of this threshold have been used in different studies, such as 500kbp [18], 1Mbp [85,125,127] and 2Mbp [124,128], and these values appear to be arbitrarily picked. As mentioned above, in our survey we observed a variety of values used in our collection of articles. In order to find a suitable threshold, one could look at Hi-C contact maps and picked a threshold such that most DNA contacts happen between sites within this distance threshold [85,86,127,129,130]. However, this approach cannot solve the conceptual problem of finding a suitable distance threshold for "*cis*", since it presumes that most enhancers act in *cis*, and would thus lead to a circular argument if this threshold is then used to study whether enhancers are *cis*-acting in general. In addition, this approach still needs to decide on a threshold for defining "most DNA contacts", which could end up with another arbitrary choice such as 99%.

An extreme case of this second definition is that an enhancer is assumed to regulate

its closest gene. While simple and reasonably accurate given its simplicity, this approach could also lead to many missing targets. Indeed, in our survey only two articles have used this approach to determine enhancer targets.

The third definition is based on mechanism rather than genomic location. A *trans*-regulatory element produces diffusible factors (such as transcription factors, non-coding RNAs and signaling molecules) that regulate transcription of a gene by binding its *cis*-regulatory elements.

Traditionally, enhancers are described to interact with target promoters by forming chromatin loops [5], and thus they regulate target gene expression without necessarily involving any diffusible products. On the other hand, it has been discovered that the expression of eRNAs correlate positively with the expression of target genes [64,131]. Whether eRNAs play an active role in the transcriptional regulation process has become a question under debate, with three non-exclusive models [132]:

- The *trans* model, that eRNAs are free to move within the nucleus and play an active role in transcriptional regulation of promoters that could be far away. An example supporting this model is the Evf-2 noncoding RNA transcribed from an enhancer of Dlx-5/6, which complexes with the Dlx-2 protein and enhances its activity [81].
- The *cis-trans* model, that eRNAs are involved in mediating the interaction between the enhancer and target promoter [133].
- The *cis* model, that eRNAs are simply by-products when the transcription pre-initiation complex is brought close to the enhancer, and play no roles in regulating the target genes

In another classification [4], the *trans* and *cis-trans* models are grouped into a single class in which the eRNA has an active function, while the *cis* model is retained as a class with no functional roles for the eRNA. There is also a third class in which it is the transcription of eRNA, but not the eRNA itself, that is important for the regulatory process.

The lack of knowledge about eRNA functions is also reflected in our survey, with only seven articles trying to discuss this topic.

Before the functions of eRNAs in transcriptional regulation (if any) become clear, no

conclusions can be made as to whether enhancers are largely *cis*-acting based on this third definition of "*cis*" and "*trans*".

In general, few studies have explicitly considered all the above three definitions of "*cis*" and "*trans*" when studying enhancers, but a number of reviews on enhancers have already explicitly pointed out that enhancers can act both in *cis* and in *trans*, based on certain definitions of these terms [49,134,135]. A study also no longer mentions "*cis*-acting" in the definition of enhancers, but describes them in a general way as "regulatory elements that increase the transcriptional output of target genes" [136].

Necessary or sufficient?

The classical definition of enhancers does not explicitly specify whether a sequence element should be necessary or sufficient for enhancing the expression of a gene in order to be called an enhancer. It also does not specify whether every nucleotide in the enhancer sequence is necessary (Figure 1d).

Considering an enhancer as a whole, one obvious reason that necessity should not be considered in defining enhancers is the possibility for a gene to be regulated by multiple enhancers. It has been estimated, based on 3C-based methods and computational predictions, that each gene is regulated by on average 2-5 enhancers in each cell/tissue type, depending on the exact cell/tissue type and whether only genes with at least one regulating enhancer are considered [85,86]. For some genes, when the main enhancer is absent, other "shadow enhancers" can take up the role [137]. Recent studies indicate that shadow enhancers may be quite prevalent, at least in the Drosophila genome [138]. To complicate things further, enhancer redundancy may change with the context. For instance, an enhancer having redundancy provided by another enhancer under normal conditions could become essential under more stressful conditions [139,140]. In general, the fact that an enhancer is not necessary for a target genes' expression due to the presence of other enhancers does not naturally affect its status as an enhancer.

If enhancers are to be defined based on their sufficiency instead, one would consider a sequence element an enhancer if there is a detectable amount of induction of the target gene's expression when the enhancer becomes present/active. This is the principle behind reporter assays, that the expression of the target gene with only a weak promoter is compared to the expression when the enhancer is present. There

is no standard as to the amount of induction required for the tested sequence to be considered an active enhancer, but in general defining whole enhancers by their sufficiency is commonly accepted.

It becomes trickier if we consider individual nucleotides within an enhancer. When a sequence is considered an enhancer, does it mean every nucleotide is necessary for the enhancer function, or some nucleotides can be dispensable as long as the whole sequence is sufficient for enhancing target gene expression?

Since an enhancer sufficient for enhancing the expression of a gene may contain regions not necessary for its function, various studies have attempted to define minimal enhancers by progressively deleting DNA sequences from both ends of an enhancer until its enhancing ability is largely abolished [97,98,141–147]. These studies implicitly define minimal enhancers as a sequence element that is sufficient as a whole, and every nucleotide is necessary.

Interestingly, some authors went further and defined an enhancer by the necessity of the nucleotides alone. For example, Smyth et al. (2008) identified a 20bp region that is necessary but not sufficient for receptor cone photoreceptor-specific expression, and called it an enhancer.

As a recurrent theme of this review, while sufficiency and necessity are concepts of Boolean logic, in reality the expression level of a gene is not binary. This is well exemplified by an early study of an enhancer of the type II collagen gene [148]. A sequence of 1.5kb was identified to enhance the expression of this gene. Taking the minimal enhancer approach, a 100bp sub-sequence was found to be able to drive the collagen gene expression at a similar level. However, when 6 additional nucleotides were deleted from the 5'end of this 100bp sequence, expression dropped to 68%. When 15 nucleotides were deleted from the 5'end instead, expression dropped to 10%. Similarly, when 11 nucleotides were deleted from the 3'end, expression dropped to 78%. While the 100bp sequence was clearly sufficient for the enhancer activity, whether the 94, 85 and 89bp truncated sequences should be considered sufficient is again a matter of choosing an expression threshold of the collagen gene. This type of gradual reduction of enhancer activity when trimming off more and more nucleotides can also be found in other studies [12,149].

One may be tempted to take a simple approach to consider an enhancer sufficient only if it can drive the full expression of the target gene. The full expression level is

usually defined based on a certain reference situation, such as the original genomic context. However, there could be silencers and insulators in the original sequence, the deletion of which can lead to an expression level of the target gene even higher than the natural context, thereby contradicting the definition of “full” expression [99,142]. Similarly, when multiple enhancers can regulate the same gene, it is unclear whether full expression should be defined based on the enhancers that actively regulate the gene in the natural context, or when all of them are made active artificially. Furthermore, the exact sequence of an enhancer may also affect the degree of target gene induction. For example, when the natural spacing between GATA1 and ETS1 binding sites in the Otx-a enhancer is increased, the enhancer activity can be increased as well [150]. These and other factors make it difficult to define the full expression of a gene.

In our survey, we found that most studies defined enhancer lengths based on practical considerations, such as size of enzyme-digested DNA fragments or length of ChIP-seq signal peaks, rather than functional necessity. We did observe a growing trend of large-scale random perturbations of enhancers, which provide information about the effects of individual nucleotides and help define the necessary regions of enhancers as discussed below.

By effect, features or mechanism?

A systematic way to test enhancer function is to perturb a potential enhancer sequence and observe the resulting change of target gene expression. For example, saturation mutagenesis coupled with massively parallel reporter assay has enabled studying the effects of mutated sequence elements on reporter genes [54]. With the invention of genome editing methods based on CRISPR-Cas9, it is now also possible to perturb specific sequences genetically [30–37] or epigenetically [31,38–44]. In these experiments, if the expression level of a gene after the perturbation of a sequence element is reduced, the sequence element is a potential regulator of the gene. Some of these elements have been found to lack typical chromatin and epigenetic features of enhancers [46,70]. Consequently, they were named unmarked regulatory elements (UREs) in some studies [46].

If we take the perturbed sequence as reference, having the unperturbed UREs can increase the target gene expression, and thus they can be called enhancers. One may further require that these UREs to be able to drive reporter gene expression in reporter assays before they can be called enhancers.

Yet the conundrum is that UREs do not have typical features of enhancers. Is it necessary for a sequence element to contain some features indicative of enhancers in order to be qualified as an enhancer? To answer this question, first there should be a common set of features used for defining enhancers. Many different feature sets have previously been used, such as H3K4me1, H3K27ac, P300 binding and eRNA [18,149,151–153]. Extensive comparisons of these features have revealed that different features have different strengths and weaknesses [154,155]. While the debate of the best enhancer-defining feature set continues, it is clear that no single feature can identify enhancers without false positives and false negatives. In our survey, 98% of the relevant articles used some features to define enhancers, but the features used differed substantially from study to study. There are also features such as H3K64ac and H3K122ac recently proposed to define a new class of enhancers [156]. Whether UREs can be considered enhancers could therefore evolve over time as more features of enhancers are discovered.

The fact that disruption of a sequence element reduces target gene expression could also be due to other factors, such as the change of the local chromatin structure. Reduction of target gene expression alone is therefore insufficient for concluding that the sequence element plays enhancer role functionally. In the literature, enhancers are usually described to function by binding transcription factors and co-activators, forming a loop to get in touch with the target promoter [6,130]. Should a sequence element be called an enhancer only if it satisfies these criteria [81]? With an incomplete list of proteins that bind enhancers and the lack of DNA three-dimensional contact information, currently it is difficult to define enhancers based on these mechanistic criteria. As a result, most “enhancers” reported in the literature were based on either their outcomes or features instead of their proven mechanisms (Figure 1e). In our survey, among the articles that tried to study the mechanisms of enhancers, the majority determined physical contacts of promoters and enhancers, which should be considered relevant but not a direct proof of the mechanistic functions of the enhancers.

Illustration of the issues by comparing several major enhancer sets

To further illustrate the issues discussed above, we compared several major sets of human enhancers to see whether they have differences related to enhancer definitions:

1. Roadmap+ENCODE [157]: Enhancers inferred by the Roadmap Epigenomics

Consortium for 127 human cell and tissue samples (including some cell lines from the Encyclopedia of DNA Elements, ENCODE) based on histone modifications and other chromatin features

2. FANTOM5 [18]: Enhancers defined by CAGE (Capped Analysis of Gene Expression) signals that exhibit eRNA-like patterns from around 1,800 human cell and tissue samples
3. VISTA [158]: Human enhancers identified from various methods tested *in vivo* using transgenic mouse assays
4. TCGA [159]: A subset of FANTOM5 enhancers with expression signals in around 9,000 human tumors based on RNA sequencing (RNA-seq)

Before comparing the genomic locations of the enhancers in these data sets, we already observed various significant differences among them, including the average length (from 276bp to 2,043bp), total number (from less than 1,000 to more than 9.3 million) and saturation (Figures 2a and 2b) of the enhancers. In particular, the FANTOM5 enhancers have saturated after including around 150 samples, but the Roadmap+ENCODE ones are not yet saturated after including all 127 samples. If the two sets of enhancers are directly compared, the differences observed could be due to either the incompleteness of the Roadmap+ENCODE set or intrinsic differences of the enhancers from the two sources.

Therefore, we compared the four data sets in three different ways with three corresponding goals. In the first comparison, we divided the whole genome into 200bp bins and considered each of them as an enhancer bin if it overlapped with an enhancer by at least 100bp, which allowed us to easily determine the intersections of the four sets. From the results (Figure 2c), the different sets of enhancer bins intersected poorly, with only 73 of them commonly shared by all four sets. This is due to a combination of reasons, including the different cell and tissue types involved in defining the different sets, their different types of samples, their ways of defining enhancers (by features or by reporter activities), as well as other technical differences such as how the length of each enhancer was determined (based on feature signals or trimmed to the same length). We also acknowledge that there are sub-structures among the data sets, in that the TCGA enhancers are a subset of the FANTOM5 enhancers, while many VISTA enhancers were originally selected based on chromatin features similar to the features used by Roadmap+ENCODE. Therefore, the four data sets tend to form two clusters that are less similar from each other.

Another way to interpret the results is by the ratio of overlap. For example, around

88% of FANTOM5 enhancer bins were also in Roadmap+ENCODE, while only around 1% of Roadmap+ENCODE enhancer bins were also in FANTOM5. Given that Roadmap+ENCODE contained almost 100 times more enhancers than FANTOM5, it is surprising to see that 12% of FANTOM5 enhancer bins were still not covered by Roadmap+ENCODE. One potential reason could be that FANTOM5 contained a lot more samples, and thus some cell- and tissue-type specific enhancers could not be covered by Roadmap+ENCODE. However, even when we included only 127 random samples from FANTOM5, there were still 12% of enhancer bins not covered by Roadmap+ENCODE (Figure 2d), showing that these two sets were different also in other additional ways. Since FANTOM5 and VISTA also contained mouse enhancers, we repeated the above analysis and found that the two resulting sets of enhancer bins also intersected poorly (Figure 2e).

As mentioned above, a potential issue of the above comparisons was the different lengths of enhancers from the different sets, such that longer enhancers would occupy more bins, and it could happen that only some of them intersected with the enhancer bins from another data set, even if both sets contained a similar enhancer. Therefore, in the second setting, we compared the enhancers directly rather than enhancer bins, and varied the required fraction of overlapping bases for two enhancers to be considered the same. An exception was made here for the Roadmap+ENCODE enhancers, which we still used 200bp bins since its enhancers from different samples do not align, unlike the FANTOM5 enhancers. We performed such comparisons for every pair of data sets. From the results (Figure 3), only in two cases was one set of enhancers completely covered by another set, namely VISTA enhancers covered by Roadmap+ENCODE (Figure 3d) and TCGA enhancers covered by FANTOM5 (Figure 3e). Both cases were due to technical rather than biological reasons as explained above. In all other cases, the overlap ratios were not very high. For example, only around 80% of the FANTOM5 enhancers could be covered by Roadmap+ENCODE even when two enhancers would be considered overlapping when they had only minimal overlap (Figure 3b, required fraction of overlapping bases close to 0%), suggesting that differences in enhancer lengths could not explain the 12% FANTOM5 enhancer bins not covered by Roadmap+ENCODE in the first comparison setting.

Finally, to deal with the issue that the different enhancer sets included different biological samples, we compared the enhancers in four ENCODE cell lines included in both the FANTOM5 and Roadmap+ENCODE data sets. Using the same comparison strategy as in the second setting above, but considering full-length

Roadmap+ENCODE enhancers rather than 200bp bins, we again observed that the overlap ratios were low (Figure 4), showing that discrepancies in biological samples also cannot explain the differences among the enhancer data sets.

Overall, this comparison of the four enhancer sets reveal that they are very different and there are many factors behind their differences. Some of these factors touch on the issues about enhancer definitions discussed above, such as whether enhancers are defined by features (CAGE/eRNAs or histone marks and DNA accessibility) or by effects (reporter activities), whether distance and orientation to target gene are involved (when designing the reporter construct), and how the exact span of the enhancer region is determined (arbitrarily or based on patterns of feature signals).

Practical implications and possible actions

Considering all the aspects of enhancer definition discussed above, and the fact that in our surveyed articles enhancers have been defined in very different ways, we argue that instead of trying to give a single universal definition of enhancers, it is more useful to define enhancers in a case-by-case manner according to the practical purpose. Here we illustrate this idea by discussing several key contemporary problems related to enhancers.

Benchmarking the performance of enhancer identification methods

The power of benchmarking computational and experimental methods based on common data sets, standardized procedures and multi-faceted evaluations has been well-demonstrated by various contests covering topics from low-level tasks such as sequence assembly and transcript quantification to high-level tasks such as prediction of structures, functions and interactions [160–165]. While many methods have been proposed for identifying enhancers, there has not been an analogous large-scale open contest for benchmarking these methods [166].

There are two major ways for setting up such a contest, namely pre-creating a “gold-standard” set of positive and negative enhancers and asking the competing methods to distinguish between them, or post-validating the candidate enhancers identified by the contesters experimentally. In both cases, the experimental details involved play a major role in defining sequences that should be considered an enhancer. First, reporter assays test sequences that are sufficient for driving reporter gene expression, while perturbation experiments test sequences that are necessary

for maintaining target gene expression, leading to different objectives. Second, the length requirements of the sequences to be tested, whether they will be automatically extended to a minimum length, as well as the distance, direction and orientation to the target gene can all affect the prediction strategy. Third, since enhancer activity is affected by the chromatin environment and the availability of relevant transcription factors, if the experimental method depends on these factors, it is necessary to provide some relevant information, for example by means of supplementary ATAC-seq and RNA-seq data of the testing system.

Enhancer annotation

There are various sets of comprehensive annotations for genes, with a reasonably high level of consistency between them, at least for protein-coding genes [167]. Correspondingly, by defining promoters as a (albeit arbitrary) small region around TSSs, they are also fairly well defined. In contrast, enhancer annotations are far from having the same level of convergence, as demonstrated in the comparisons presented in the introduction.

Going forward toward a more unified set of enhancers, there are various lessons that can be learned from gene annotations, which naturally land on definitions of enhancers. First, gene annotations aim at including the full set of genes and transcripts. A gene is included as long as it is expressed in at least one cell/tissue type or condition. In the same way, an enhancer annotation set should include all sequence elements that have enhancing capability in any context, regardless of the exact definition. Second, just as genes are annotated by the different sub-elements of it (exons, introns, etc.) and at different levels (elements related to transcription and elements related to translation), enhancers can also be annotated by different sub-elements based on their functions (such as protein binding locations), features (such as location of H3K27ac peak) and outcomes (such as parts that can drive the target gene up to certain percentage of a reference expression level). Third, some gene annotations provide confidence levels based on experimental evidence, manual verification and computational inference [167]. Enhancer annotations can as well be associated with confidence levels based on such information. This three-level hierarchy of information allows users of the enhancer annotations to pick the most suitable enhancer definition and correspondingly the boundaries of each enhancer based on their need.

Identifying target genes of enhancers

Since the target genes of an enhancer can be upstream or downstream of it and at various distances, determining enhancer targets is a non-trivial task and many methods have been proposed [19,20,29,21–28]. Having a clear definition of enhancers helps clarify the scope of this task and guide the design of the methods.

Most existing methods define a maximum distance between an enhancer and its potential target genes, due to physical, computational and statistical reasons. If enhancers act in *cis* alone, setting this maximum distance to the typical size of a TAD should be sufficient to cover a large portion of enhancer targets. On the other hand, if enhancers can act in *trans* by having functional roles of eRNAs over a long genomic distance, this distance threshold should be abandoned and it could be useful to incorporate new types of information into the target identification methods, such as protein-eRNA interactions [168].

As discussed above, the boundaries of an enhancer are hard to define, depending on the sufficiency/necessity requirement and level of target gene expression that it needs to drive. Enhancer identification methods based on activity signals could lead to enhancers that are slightly shifted in different samples, making it difficult to define a unified set of enhancers. In addition, whether a large genomic region with broad enhancer feature signals should be considered a single enhancer unit (such as a super enhancer [169,170]) or multiple separate enhancers also affects the target identification methods. A general way to deal with these issues is to divide up such broad regions into small enhancer units and allow them to exert joint effects on common targets [29].

Estimating functional effects of non-coding genetic variants

An important application of enhancer annotation and enhancer target gene sets is to estimate changes of gene expression levels due to non-coding genetic variants that hit enhancer regions [171,172]. Existing methods mostly look for simple overlaps between the variants and annotated enhancers, or specific protein binding motifs or ChIP-seq signals within them. Having a detailed sufficiency/necessity map of individual nucleotides for enhancer function could greatly improve the precision of these methods.

A straightforward approach to predicting genes affected by a non-coding variant is to associate the variant with genes nearby, such as the closest genes, genes within a

certain genomic distance, or genes within the same linkage disequilibrium block [173]. Although enhancers may not regulate genes within the immediate vicinity, this simple approach is still fairly popular [174]. Advancements in enhancer target identification methods, taking into account the operational mechanisms of enhancers, will help predict the affected genes and accordingly better prioritize genes for validation and functional experiments.

Acknowledgments

We thank Nelson Tang for helpful discussions. KYY is partially supported by the HKSAR Research Grants Council General Research Fund 14145916, Collaborative Research Fund C4017-14G and C4054-16G, and Theme-based Research Scheme T12-401/13-R, T12-402/13N and T12C-714/14-R.

Key Points

- The traditional definition of transcriptional enhancers needs to be revisited due to various recent new findings.
- In the literature, many different definitions of enhancers are used, making it difficult to compare the results in different studies.
- It is important for the research community to define clear reporting guidelines regarding the definition of enhancers adopted.

Biographical Note

Edwin Yu-Kiu Ho was a master student in Department of Biomedical Engineering at The Chinese University of Hong Kong.

Qin Cao is a postdoctoral associate in School of Biomedical Sciences and Department of Computer Science and Engineering at The Chinese University of Hong Kong. She studies methods for identifying enhancer targets.

Mengting Gu is a PhD student in Department of Molecular Biophysics and Biochemistry at Yale University. She studies computational methods for identifying enhancers.

Ricky Wai-Lun Chan is a PhD student in Department of Computer Science and Engineering at The Chinese University of Hong Kong. He studies sequence patterns of

enhancers.

Qiong Wu is a PhD student in School of Biomedical Sciences and Department of Computer Science and Engineering at The Chinese University of Hong Kong. She studies cancer epigenomics.

Mark Gerstein is an Albert L Williams Professor of Biomedical Informatics at Yale University. He studies various topics in biomedical informatics, including transcriptional regulation, biological networks and genetic variations.

Kevin Y. Yip is an associate professor in Department of Computer Science and Engineering at The Chinese University of Hong Kong. He studies transcriptional regulation and human diseases.

References

1. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981;27(2 Pt 1):299-308.
2. Benoist C, Chambon P. In vivo sequence requirements of the SV40 early promoter region. *Nature*. 1981;290(5804):304-310. doi:10.1038/290304a0
3. Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res*. 1981;9(22):6047-6068.
4. Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet*. 2016;17(4):207-223. doi:10.1038/nrg.2016.4
5. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nat Rev Genet*. 2013;14(4):288-295. doi:10.1038/nrg3458
6. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15(4):272-286. doi:10.1038/nrg3682
7. Vernimmen D, Bickmore WA. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet*. 2015;31(12):696-708. doi:10.1016/j.tig.2015.10.004
8. Yip KY, Cheng C, Bhardwaj N, et al. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol*. 2012;13(9):R48. doi:10.1186/gb-2012-13-9-r48

9. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339(6123):1074-1077. doi:10.1126/science.1232542
10. Patwardhan RP, Hiatt JB, Witten DM, et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol*. 2012;30(3):265-270. doi:10.1038/nbt.2136
11. Melnikov A, Murugan A, Zhang X, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30(3):271-277. doi:10.1038/nbt.2137
12. Ernst J, Melnikov A, Zhang X, et al. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol*. 2016;34(11):1180-1190. doi:10.1038/nbt.3678
13. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res*. 2014;24(10):1595-1602. doi:10.1101/gr.173518.114
14. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247
15. Inoue F, Kircher M, Martin B, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27(1):38-52. doi:10.1101/gr.212092.116
16. Sheffield NC, Thurman RE, Song L, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. 2013;23(5):777-788. doi:10.1101/gr.152140.112
17. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011;470(7333):279-283. doi:10.1038/nature09692
18. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455-461. doi:10.1038/nature12787
19. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol*. 2011;30(1):90-98. doi:10.1038/nbt.2057
20. Beagrie RA, Scialdone A, Schueler M, et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*. 2017;543(7646):519-524. doi:10.1038/nature21411
21. Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific Genome Architecture

Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167(5):1369-1384.e19. doi:10.1016/j.cell.2016.09.037

22. Mumbach MR, Rubin AJ, Flynn RA, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13(11):919-922. doi:10.1038/nmeth.3999
23. Davies JOJ, Telenius JM, McGowan SJ, et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods*. 2016;13(1):74-80. doi:10.1038/nmeth.3664
24. Hughes JR, Roberts N, McGowan S, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*. 2014;46(2):205-212. doi:10.1038/ng.2871
25. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 2015;47(6):598-606. doi:10.1038/ng.3286
26. Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665-1680. doi:10.1016/j.cell.2014.11.021
27. Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009;462(7269):58-64. doi:10.1038/nature08497
28. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi:10.1126/science.1181369
29. Cao Q, Anyansi C, Hu X, et al. Reconstruction of enhancer – target networks in 935 samples of human primary cells , tissues and cell lines. *Nat Publ Gr*. 2017;49(10):1428-1436. doi:10.1038/ng.3950
30. Kowalczyk MS, Hughes JR, Garrick D, et al. Intragenic enhancers act as alternative promoters. *Mol Cell*. 2012;45(4):447-458. doi:10.1016/j.molcel.2011.12.021
31. Zhang X, Choi PS, Francis JM, et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet*. 2016;48(2):176-182. doi:10.1038/ng.3470
32. Canver MC, Smith EC, Sher F, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527(7577):192-197. doi:10.1038/nature15521
33. Hnisz D, Schuijers J, Lin CY, et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell*. 2015;58(2):362-370. doi:10.1016/j.molcel.2015.02.014

34. Li Y, Rivera CM, Ishii H, et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One*. 2014;9(12):e114485. doi:10.1371/journal.pone.0114485
35. Zhou HY, Katsman Y, Dhaliwal NK, et al. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev*. 2014;28(24):2699-2711. doi:10.1101/gad.248526.114
36. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014;157(2):369-381. doi:10.1016/j.cell.2014.02.019
37. Canver MC, Bauer DE, Dass A, et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J Biol Chem*. 2014;289(31):21312-21324. doi:10.1074/jbc.M114.564625
38. Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell*. 2017;66(2):285-299.e5. doi:10.1016/j.molcel.2017.03.007
39. Kearns NA, Pham H, Tabak B, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*. 2015;12(5):401-403. doi:10.1038/nmeth.3325
40. Thakore PI, D'Ippolito AM, Song L, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*. 2015;12(12):1143-1149. doi:10.1038/nmeth.3630
41. Hilton IB, D'Ippolito AM, Vockley CM, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*. 2015;33(5):510-517. doi:10.1038/nbt.3199
42. Gao X, Tsang JCH, Gaba F, Wu D, Lu L, Liu P. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res*. 2014;42(20):e155. doi:10.1093/nar/gku836
43. Perez-Pinera P, Kocak DD, Vockley CM, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013;10(10):973-976. doi:10.1038/nmeth.2600
44. Qi LS, Larson MH, Gilbert LA, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013;152(5):1173-1183. doi:10.1016/j.cell.2013.02.022
45. Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol*. 2017;18(1):219. doi:10.1186/s13059-017-1345-5

46. Rajagopal N, Srinivasan S, Kooshesh K, et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol*. 2016;34(2):167-174. doi:10.1038/nbt.3468
47. Korkmaz G, Lopes R, Ugalde AP, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol*. 2016;34(2):192-198. doi:10.1038/nbt.3450
48. Plank JL, Dean A. Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell*. 2014;55(1):5-14. doi:10.1016/j.molcel.2014.06.015
49. Schaffner W. Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem*. 2015;396(4):311-327. doi:10.1515/hsz-2014-0303
50. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015;16(3):144-154. doi:10.1038/nrm3949
51. Whitaker JW, Nguyen TT, Zhu Y, Wildberg A, Wang W. Computational schemes for the prediction and annotation of enhancers from epigenomic assays. *Methods*. 2015;72:86-94. doi:10.1016/j.ymeth.2014.10.008
52. Cao Q, Yip KY. A Survey on Computational Methods for Enhancer and Enhancer Target Predictions. In: Wong K-C, ed. *Computational Biology and Bioinformatics: Gene Regulation*. CRC Press; 2015.
53. Su J, Teichmann SA, Down TA. Assessing computational methods of cis-regulatory module prediction. *PLoS Comput Biol*. 2010;6(12):e1001020. doi:10.1371/journal.pcbi.1001020
54. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. *Genomics*. 2015;106(3):159-164. doi:10.1016/j.ygeno.2015.06.005
55. Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011;12(4):283-293. doi:10.1038/nrg2957
56. Babbitt CC, Markstein M, Gray JM. Recent advances in functional assays of transcriptional enhancers. *Genomics*. 2015;106(3):137-139. doi:10.1016/j.ygeno.2015.06.002
57. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*. 2011;12(1):7-18. doi:10.1038/nrg2905
58. Andersson R, Sandelin A, Danko CG. A unified architecture of transcriptional regulatory elements. *Trends Genet*. 2015;31(8):426-433. doi:10.1016/j.tig.2015.05.007
59. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian

promoters and enhancers. *Nat Genet*. 2014;46(12):1311-1320. doi:10.1038/ng.3142

60. Kim T-K, Shiekhattar R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*. 2015;162(5):948-959. doi:10.1016/j.cell.2015.08.008
61. Gray JM, Kim T-K, West AE, Nord AS, Markenscoff-Papadimitriou E, Lomvardas S. Genomic Views of Transcriptional Enhancers: Essential Determinants of Cellular Identity and Activity-Dependent Responses in the CNS. *J Neurosci*. 2015;35(41):13819-13826. doi:10.1523/JNEUROSCI.2622-15.2015
62. Nguyen TA, Jones RD, Snavely AR, et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res*. 2016;26(8):1023-1033. doi:10.1101/gr.204834.116
63. De Santa F, Barozzi I, Mietton F, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*. 2010;8(5):e1000384. doi:10.1371/journal.pbio.1000384
64. Kim T-K, Hemberg M, Gray JM, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010;465(7295):182-187. doi:10.1038/nature09033
65. Hah N, Danko CG, Core L, et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*. 2011;145(4):622-634. doi:10.1016/j.cell.2011.03.042
66. Wang D, Garcia-Bassets I, Benner C, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*. 2011;474(7351):390-394. doi:10.1038/nature10006
67. Serfling E, Lübbe A, Dorsch-Häsler K, Schaffner W. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO J*. 1985;4(13B):3851-3859.
68. Li G, Ruan X, Auerbach RK, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 2012;148(1-2):84-98. doi:10.1016/j.cell.2011.12.014
69. Dao LTM, Galindo-Albarrán AO, Castro-Mondragon JA, et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet*. 2017;49(7):1073-1081. doi:10.1038/ng.3884
70. Diao Y, Fang R, Li B, et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods*. 2017;14(6):629-635. doi:10.1038/nmeth.4264
71. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest ARR, Kawaji H, et al. A promoter-level mammalian expression atlas. *Nature*.

2014;507(7493):462-470. doi:10.1038/nature13182

72. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775-1789. doi:10.1101/gr.132159.111
73. Skvortsova Y V, Kondratieva SA, Zinovyeva M V, Nikolaev LG, Azhikina TL, Gainetdinov I V. Intragenic Locus in Human PIWIL2 Gene Shares Promoter and Enhancer Functions. *PLoS One*. 2016;11(6):e0156454. doi:10.1371/journal.pone.0156454
74. Natoli G, Andrau J-C. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet*. 2012;46:1-19. doi:10.1146/annurev-genet-110711-155459
75. Wu X, Sharp PA. Divergent transcription: a driving force for new gene origination? *Cell*. 2013;155(5):990-996. doi:10.1016/j.cell.2013.10.048
76. Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*. 2013;499(7458):360-363. doi:10.1038/nature12349
77. Kaida D, Berg MG, Younis I, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*. 2010;468(7324):664-668. doi:10.1038/nature09479
78. Andersson R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays*. 2015;37(3):314-323. doi:10.1002/bies.201400162
79. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-112. doi:10.1038/nature07829
80. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311-318. doi:10.1038/ng1966
81. Sahlén P, Abdullayev I, Ramsköld D, et al. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol*. 2015;16:156. doi:10.1186/s13059-015-0727-9
82. Gordon CT, Attanasio C, Bhatia S, et al. Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. *Hum Mutat*. 2014;35(8):1011-1020. doi:10.1002/humu.22606
83. Sagai T, Masuya H, Tamura M, et al. Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). *Mamm Genome*. 2004;15(1):23-34. doi:10.1007/s00335-033-2317-5

84. Lettice LA, Heaney SJH, Purdie LA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*. 2003;12(14):1725-1735.
85. He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A*. 2014;111(21):E2191-9. doi:10.1073/pnas.1320308111
86. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013;503(7475):290-294. doi:10.1038/nature12644
87. Tang Z, Luo OJ, Li X, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*. 2015;163(7):1611-1627. doi:10.1016/j.cell.2015.11.024
88. Heidari N, Phanstiel DH, He C, et al. Genome-wide map of regulatory interactions in the human genome. *Genome Res*. 2014;24(12):1905-1917. doi:10.1101/gr.176586.114
89. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376-380. doi:10.1038/nature11082
90. Roepcke S, Stahlberg S, Klein H, et al. A tandem sequence motif acts as a distance-dependent enhancer in a set of genes involved in translation by binding the proteins NonO and SFPQ. *BMC Genomics*. 2011;12:624. doi:10.1186/1471-2164-12-624
91. Králová J, Jansa P, Forejt J. A novel downstream regulatory element of the mouse H-2Kb class I major histocompatibility gene. *EMBO J*. 1992;11(12):4591-4600.
92. Tam KT, Chan PK, Zhang W, et al. Identification of a novel distal regulatory element of the human Neuroglobin gene by the chromosome conformation capture approach. *Nucleic Acids Res*. 2017;45(1):115-126. doi:10.1093/nar/gkw820
93. Collins C, Azmi P, Berru M, Zhu X, Shulman MJ. A weakened transcriptional enhancer yields variegated gene expression. *PLoS One*. 2006;1:e33. doi:10.1371/journal.pone.0000033
94. Mocikat R, Harloff C, Kütemeier G. The effect of the rat immunoglobulin heavy-chain 3' enhancer is position dependent. *Gene*. 1993;136(1-2):349-353.
95. Bruce WB, Bandyopadhyay R, Gurley WB. An enhancer-like element present in the promoter of a T-DNA gene from the Ti plasmid of Agrobacterium tumefaciens. *Proc Natl Acad Sci U S A*. 1988;85(12):4310-4314.
96. Guo Z, Shao L, Du Q, Park KS, Geller DA. Identification of a classic

cytokine-induced enhancer upstream in the human iNOS promoter. *FASEB J*. 2007;21(2):535-542. doi:10.1096/fj.06-6739com

97. Madan A, Curtin PT. A 24-base-pair sequence 3' to the human erythropoietin gene contains a hypoxia-responsive transcriptional enhancer. *Proc Natl Acad Sci U S A*. 1993;90(9):3928-3932.
98. Harris LC, Remack JS, Brent TP. Identification of a 59 bp enhancer located at the first exon/intron boundary of the human O6-methylguanine DNA methyltransferase gene. *Nucleic Acids Res*. 1994;22(22):4614-4619.
99. Lorch Y, Lue NF, Kornberg RD. Interchangeable RNA polymerase I and II enhancers. *Proc Natl Acad Sci U S A*. 1990;87(21):8202-8206.
100. Walke W, Xiao G, Goldman D. Identification and characterization of a 47 base pair activity-dependent enhancer of the rat nicotinic acetylcholine receptor delta-subunit promoter. *J Neurosci*. 1996;16(11):3641-3651.
101. Bachl J, Olsson C, Chitkara N, Wabl M. The Ig mutator is dependent on the presence, position, and orientation of the large intron enhancer. *Proc Natl Acad Sci U S A*. 1998;95(5):2396-2399.
102. Smyth VA, Di Lorenzo D, Kennedy BN. A novel, evolutionarily conserved enhancer of cone photoreceptor-specific expression. *J Biol Chem*. 2008;283(16):10881-10891. doi:10.1074/jbc.M710454200
103. Ward RD, Davis SW, Cho M, et al. Comparative genomics reveals functional transcriptional control sequences in the Prop1 gene. *Mamm Genome*. 2007;18(6-7):521-537. doi:10.1007/s00335-007-9008-6
104. Bonello GB, Pham M-H, Begum K, Sigala J, Sataranatarajan K, Mummidi S. An evolutionarily conserved TNF-alpha-responsive enhancer in the far upstream region of human CCL2 locus influences its gene expression. *J Immunol*. 2011;186(12):7025-7038. doi:10.4049/jimmunol.0900643
105. Miele A, Dekker J. Long-range chromosomal interactions and gene regulation. *Mol Biosyst*. 2008;4(11):1046-1057. doi:10.1039/b803580f
106. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. *Trends Genet*. 2010;26(4):188-197. doi:10.1016/j.tig.2010.01.007
107. Bateman JR, Johnson JE, Locke MN. Comparing enhancer action in cis and in trans. *Genetics*. 2012;191(4):1143-1155. doi:10.1534/genetics.112.140954
108. Koch L. Molecular evolution: Regulatory runaways. *Nat Rev Genet*. 2016;17(1):6. doi:10.1038/nrg.2015.14
109. Blick AJ, Mayer-Hirshfeld I, Malibiran BR, et al. The Capacity to Act in Trans Varies Among Drosophila Enhancers. *Genetics*. 2016;203(1):203-218. doi:10.1534/genetics.115.185645

110. Harper AR, Nayee S, Topol EJ. Protective alleles and modifier variants in human health and disease. *Nat Rev Genet*. 2015;16(12):689-701. doi:10.1038/nrg4017
111. González-Barrios M, Fierro-González JC, Krpelanova E, et al. Cis- and trans-regulatory mechanisms of gene expression in the ASJ sensory neuron of Caenorhabditis elegans. *Genetics*. 2015;200(1):123-134. doi:10.1534/genetics.115.176172
112. Borensztein P, Germain S, Fuchs S, Philippe J, Corvol P, Pinet F. cis-regulatory elements and trans-acting factors directing basal and cAMP-stimulated human renin gene expression in chorionic cells. *Circ Res*. 1994;74(5):764-773.
113. Chalifour LE, Wirak DO, Hansen U, Wassarman PM, DePamphilis ML. cis- and trans-acting sequences required for expression of simian virus 40 genes in mouse oocytes. *Genes Dev*. 1987;1(10):1096-1106.
114. Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD, eds. *Basic Neurochemistry -- Molecular, Cellular and Medical Aspects*. 6th ed. American Society for Neurochemistry; 1999.
115. Duncan IW. Transvection effects in Drosophila. *Annu Rev Genet*. 2002;36:521-556. doi:10.1146/annurev.genet.36.060402.100441
116. Morris JR, Chen JL, Geyer PK, Wu CT. Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis. *Proc Natl Acad Sci U S A*. 1998;95(18):10740-10745.
117. Geyer PK, Green MM, Corces VG. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *EMBO J*. 1990;9(7):2247-2256.
118. Müller HP, Schaffner W. Transcriptional enhancers can act in trans. *Trends Genet*. 1990;6(9):300-304.
119. Lewis EB. The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster. *Am Nat*. 1954;88(841):225-239.
120. Mellert DJ, Truman JW. Transvection is common throughout the Drosophila genome. *Genetics*. 2012;191(4):1129-1141. doi:10.1534/genetics.112.140475
121. Patel B, Kang Y, Cui K, et al. Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. *Leukemia*. 2014;28(2):349-361. doi:10.1038/leu.2013.158
122. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. *Nature*. 2005;435(7042):637-645. doi:10.1038/nature03574
123. Mahmoudi T, Katsani KR, Verrijzer CP. GAGA can mediate enhancer function in

trans by linking two separate DNA molecules. *EMBO J*. 2002;21(7):1775-1781. doi:10.1093/emboj/21.7.1775

124. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. *PLoS Genet*. 2014;10(7):e1004461. doi:10.1371/journal.pgen.1004461
125. Kirsten H, Al-Hasani H, Holdt L, et al. Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci†. *Hum Mol Genet*. 2015;24(16):4746-4763. doi:10.1093/hmg/ddv194
126. Thompson DA, Cubillos FA. Natural gene expression variation studies in yeast. *Yeast*. 2017;34(1):3-17. doi:10.1002/yea.3210
127. Roy S, Siahpirani AF, Chasman D, et al. A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Res*. 2015;43(18):8694-8712. doi:10.1093/nar/gkv865
128. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet*. 2016;48(5):488-496. doi:10.1038/ng.3539
129. Lan X, Witt H, Katsumura K, et al. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res*. 2012;40(16):7690-7704. doi:10.1093/nar/gks501
130. Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. *Brief Bioinform*. 2016;17(6):980-995. doi:10.1093/bib/bbv097
131. Ørom UA, Derrien T, Beringer M, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 2010;143(1):46-58. doi:10.1016/j.cell.2010.09.001
132. Rothschild G, Basu U. Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends Genet*. 2017;33(2):143-154. doi:10.1016/j.tig.2016.12.002
133. Wang X, Arai S, Song X, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*. 2008;454(7200):126-130. doi:10.1038/nature06992
134. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin*. 2012;5(1):1. doi:10.1186/1756-8935-5-1
135. Kulaeva OI, Nizovtseva E V, Polikanov YS, Ulianov S V, Studitsky VM. Distant activation of transcription: mechanisms of enhancer action. *Mol Cell Biol*. 2012;32(24):4892-4897. doi:10.1128/MCB.01127-12

136. Krivega I, Dean A. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev*. 2012;22(2):79-85. doi:10.1016/j.gde.2011.11.001
137. Hong J-W, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science*. 2008;321(5894):1314. doi:10.1126/science.1160631
138. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51. doi:10.1016/j.cub.2015.11.034
139. Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol*. 2010;20(17):1562-1567. doi:10.1016/j.cub.2010.07.043
140. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010;466(7305):490-493. doi:10.1038/nature09158
141. Ma L, Merenmies J, Parada LF. Molecular characterization of the TrkA/NGF receptor minimal enhancer reveals regulation by multiple cis elements to drive embryonic neuron expression. *Development*. 2000;127(17):3777-3788.
142. Lindsley A, Snider P, Zhou H, et al. Identification and characterization of a novel Schwann and outflow tract endocardial cushion lineage-restricted periostin enhancer. *Dev Biol*. 2007;307(2):340-355. doi:10.1016/j.ydbio.2007.04.041
143. Wang C-C, Tsai M-F, Dai T-H, et al. Synergistic activation of the tumor suppressor, HLJ1, by the transcription factors YY1 and activator protein 1. *Cancer Res*. 2007;67(10):4816-4826. doi:10.1158/0008-5472.CAN-07-0504
144. Munshi N V, McAnally J, Bezprozvannaya S, et al. Cx30.2 enhancer analysis identifies Gata4 as a novel regulator of atrioventricular delay. *Development*. 2009;136(15):2665-2674. doi:10.1242/dev.038562
145. Jumlongras D, Lachke SA, O'Connell DJ, et al. An evolutionarily conserved enhancer regulates Bmp4 expression in developing incisor and limb bud. *PLoS One*. 2012;7(6):e38568. doi:10.1371/journal.pone.0038568
146. Barembaum M, Bronner ME. Identification and dissection of a key enhancer mediating cranial neural crest specific expression of transcription factor, Ets-1. *Dev Biol*. 2013;382(2):567-575. doi:10.1016/j.ydbio.2013.08.009
147. Garstang MG, Osborne PW, Ferrier DEK. TCF/Lef regulates the Gsx ParaHox gene in central nervous system development in chordates. *BMC Evol Biol*. 2016;16:57. doi:10.1186/s12862-016-0614-3
148. Krebsbach PH, Nakata K, Bernier SM, et al. Identification of a minimum enhancer sequence for the type II collagen gene reveals several core sequence

motifs in common with the link protein gene. *J Biol Chem*. 1996;271(8):4298-4303.

149. Antes TJ, Goodart SA, Huynh C, Sullivan M, Young SG, Levy-Wilson B. Identification and characterization of a 315-base pair enhancer, located more than 55 kilobases 5' of the apolipoprotein B gene, that confers expression in the intestine. *J Biol Chem*. 2000;275(34):26637-26648. doi:10.1074/jbc.M003025200
150. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. Suboptimization of developmental enhancers. *Science*. 2015;350(6258):325-328. doi:10.1126/science.aac6948
151. Rajagopal N, Xie W, Li Y, et al. RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput Biol*. 2013;9(3):e1002968. doi:10.1371/journal.pcbi.1002968
152. Koenecke N, Johnston J, Gaertner B, Natarajan M, Zeitlinger J. Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol*. 2016;17(1):196. doi:10.1186/s13059-016-1057-2
153. Corradin O, Saiakhova A, Akhtar-Zaidi B, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*. 2014;24(1):1-13. doi:10.1101/gr.164079.113
154. Dogan N, Wu W, Morrissey CS, et al. Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin*. 2015;8:16. doi:10.1186/s13072-015-0009-5
155. Kleftogiannis D, Kalnis P, Bajic VB. Progress and challenges in bioinformatics approaches for enhancer identification. *Brief Bioinform*. 2016;17(6):967-979. doi:10.1093/bib/bbv101
156. Pradeepa MM, Grimes GR, Kumar Y, et al. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet*. 2016;48(6):681-686. doi:10.1038/ng.3550
157. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248
158. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res*. 2007;35(Database issue):D88-92. doi:10.1093/nar/gkl822
159. Chen H, Li C, Peng X, et al. A Pan-Cancer Analysis of Enhancer Expression in

Nearly 9000 Patient Samples. *Cell*. 2018;173(2):386-399.e12. doi:10.1016/j.cell.2018.03.027

160. Steijger T, Abril JF, Engström PG, et al. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods*. 2013;10(12):1177-1184. doi:10.1038/nmeth.2714
161. Bradnam KR, Fass JN, Alexandrov A, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience*. 2013;2(1):10. doi:10.1186/2047-217X-2-10
162. Radivojac P, Clark WT, Oron TR, et al. A large-scale evaluation of computational protein function prediction. *Nat Methods*. 2013;10(3):221-227. doi:10.1038/nmeth.2340
163. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins*. 2016;84 Suppl 1:4-14. doi:10.1002/prot.25064
164. Stolovitzky G, Monroe D, Califano A. Reverse engineering biological networks. Opportunities and challenges in computational methods for pathway inference. Proceedings of the workshop entitled Dialogue on Reverse Engineering Assessment and Methods (DREAM). September 7-8, 2006. Bronx, New York. *Ann N Y Acad Sci*. 2007;1115:xi-xiv, 1-285.
165. Janin J, Henrick K, Moult J, et al. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins*. 2003;52(1):2-9. doi:10.1002/prot.10381
166. Li Y, Shi W, Wasserman WW. Genome-wide prediction of cis-regulatory regions using supervised deep learning methods. *BMC Bioinformatics*. 2018;19(1):202. doi:10.1186/s12859-018-2187-1
167. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-1774. doi:10.1101/gr.135350.111
168. Hu X, Wu Y, Lu ZJ, Yip KY. Analysis of sequencing data for probing RNA secondary structures and protein-RNA binding in studying posttranscriptional regulations. *Brief Bioinform*. 2016;17(6):1032-1043. doi:10.1093/bib/bbv106
169. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934-947. doi:10.1016/j.cell.2013.09.053
170. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307-319. doi:10.1016/j.cell.2013.03.035
171. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*. 2014;15(10):480.

doi:10.1186/s13059-014-0480-5
172. Khurana E, Fu Y, Colonna V, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science*. 2013;342(6154):1235587. doi:10.1126/science.1235587
173. Petersen A, Alvarez C, DeClaire S, Tintle NL. Assessing methods for assigning SNPs to genes in gene-based tests of association using common variants. *PLoS One*. 2013;8(5):e62161. doi:10.1371/journal.pone.0062161
174. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495-501. doi:10.1038/nbt.1630

Figure Legends

Figure 1. Different aspects that affect the definition of transcriptional enhancers. (a) In the classical view, a promoter initiates transcription while an enhancer enhances the expression level. However, it has been shown that enhancers can also initiate transcription, and promoters also exhibit enhancing activities. (b) The expression level of the target gene could be affected by the distance and direction of the enhancer with respect to the gene, and its orientation. (c) Three different definitions of *cis* and *trans*, respectively based on whether the enhancer and target gene are on the same molecule, whether the enhancer is close to the target gene, and whether the enhancer itself but not its diffusible products is needed for the enhancer function. (d) A sequence element is usually defined as an enhancer if it is sufficient to drive target gene expression, even if it is not necessary due to alternative enhancers. On the other hand, necessity of individual bases can be defined based on the loss of target gene expression upon removing the bases. (e) Enhancers are commonly defined based on its features such as H3K4me1 and H3K27ac and/or by its observable effects on target gene expression, although ideally the mechanism should be involved in the definition.

Figure 2. Comparisons of the four enhancer sets based on fixed-sized bins. The whole genome was divided into consecutive, non-overlapping 200bp bins. For each enhancer set, a bin was defined as an enhancer bin if it overlapped with an enhancer from this set by at least 100bp. (a,b) Saturation plots of the FANTOM5 (a) and Roadmap+ENCODE (b) human enhancer bins, where the y-axis represents the fraction of all enhancer bins that can be covered by a random subset of the samples. For each subset size, 100 random subsets of samples were drawn to form a distribution. (c) A Venn diagram of the four sets of human enhancer bins. (d) The fraction of FANTOM5 human enhancer bins covered by Roadmap+ENCODE based on all FANTOM5 samples or subsets of 127 random samples, with the distribution of the latter formed by 100 random subsets. (e) The intersection of FANTOM5 and VISTA mouse enhancer bins.

Figure 3. Pair-wise comparisons of the four enhancer sets at various fractions of overlapping bases for two enhancers to be considered the same. Each panel involves a pair of enhancer data sets. Each curve in a panel shows the fraction of enhancers from a particular set (the “subject enhancers”) that are also contained in the other enhancer set, with the fraction of overlapping bases computed using the enhancer length of one of the two sets chosen (the “length normalizer”) as the denominator.

For example, when the subject enhancers are from FANTOM5 and the length normalizer is the Roadmap+ENCODE enhancers, a point on the curve with an x-coordinate of *x* indicates the ratio of FANTOM5 enhancers that are also contained in Roadmap+ENCODE, where a FANTOM5 enhancer is considered to be contained in Roadmap+ENCODE if there is an enhancer from Roadmap+ENCODE of length *l* that share *c* common bases with the FANTOM5 enhancer with $c/l \geq x$.

Figure 4. Pair-wise comparisons of the FANTOM5 and Roadmap+ENCODE enhancer sets at various fractions of overlapping bases for two enhancers to be considered the same, based on only enhancers from the same cell lines. Each panel involves only the enhancers from a particular cell line contained in both the FANTOM5 and Roadmap+ENCODE data sets. The interpretation of the curves is the same as in Figure 3.

!"#

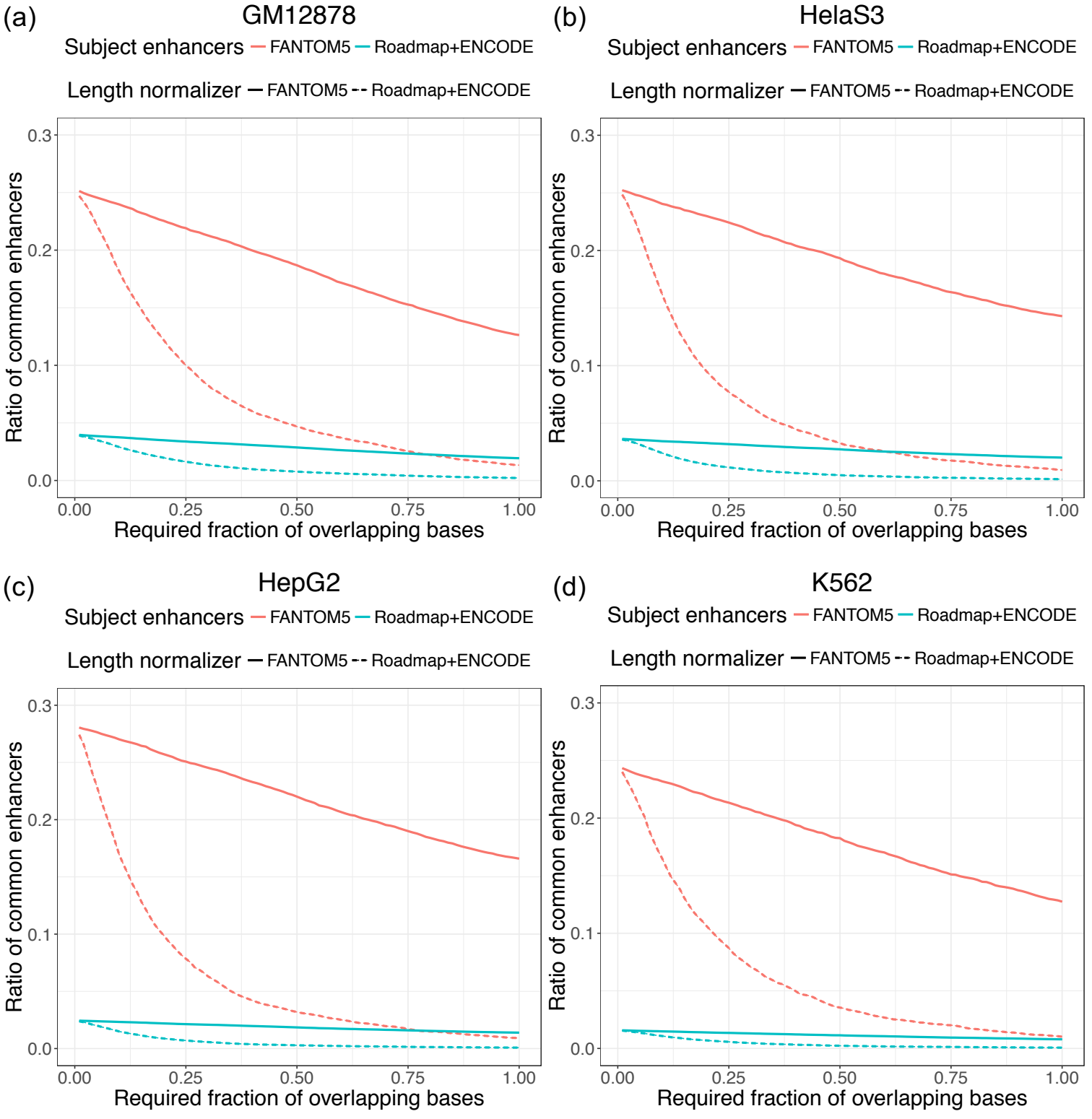
!$#

!%#

!&#

!’#

!(#


(a)
GM12878
Subject enhancers FANTOM5 Roadmap+ENCODE
Length normalizer FANTOM5 Roadmap+ENCODE
Ratio of common enhancers
0.0
0.1
0.2
0.3
0.00
0.25
0.50
0.75
1.00
Required fraction of overlapping bases
(b)
HelaS3
Subject enhancers FANTOM5 Roadmap+ENCODE
Length normalizer FANTOM5 Roadmap+ENCODE
Ratio of common enhancers
Required fraction of overlapping bases
(c)
HepG2
Subject enhancers FANTOM5 Roadmap+ENCODE
Length normalizer FANTOM5 Roadmap+ENCODE
Ratio of common enhancers
Required fraction of overlapping bases
(d)
K562
Subject enhancers FANTOM5 Roadmap+ENCODE
Length normalizer FANTOM5 Roadmap+ENCODE
Ratio of common enhancers
Required fraction of overlapping bases