COMMENTARY

How to tune an enhancer

Scott Barolo[a,1]

Every cell's genome contains two main classes of functional DNA. The best understood type of DNA sequence, which was also the first to be discovered, is that which encodes RNA and protein products via the near-universal "genetic code" (1). A more mysterious but equally important class of functional DNA is *cis*-regulatory sequence, which does not have a physical product but, instead, encodes the conditions under which a particular RNA will be produced. *Cis*-regulatory DNA sequences are the primary (although not the only) determinant of gene expression: Not only the rate of RNA production but also the timing, spatial patterning, and environmental control of every gene's activity are largely controlled by these DNA sequences, which are usually in the general vicinity of the gene they regulate. These DNA segments (often called enhancers) have no enzymatic activity on their own, but act as scaffolds for large complexes of proteins and RNAs that directly control the activity of a gene's promoter, sometimes over distances of 1 million base pairs or more (2). Transcription factors (TFs), key protein regulators of gene expression, bind DNA in a sequence-specific manner, which means that the nature of the complex assembled at a given enhancer at a given time depends on its DNA sequence (*cis* information), in conjunction with the set of TFs present and active in the cell at that time (*trans* information). The chromatin state of regulatory DNA is also a very important factor, but local chromatin states are also specified, indirectly, by DNA sequence, because most chromatin-modifying enzymes are recruited to the genome by sequence-specific factors (3).

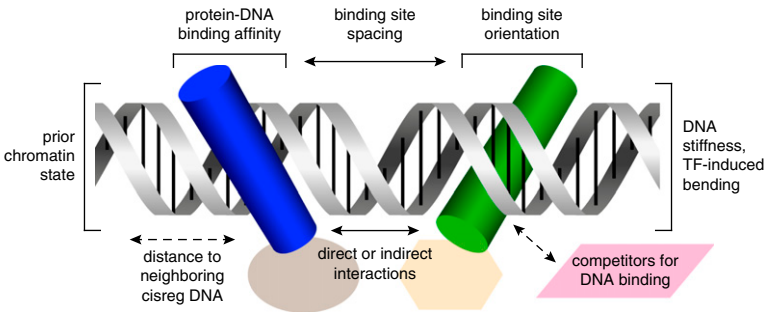

Fig. 1. Constraints and trade-offs in enhancer design. DNA sequence features modulate TF function and enhancer activity, directly or indirectly, in many ways. Not depicted are larger scale but equally sequence-driven regulatory mechanisms, such genome compartmentalization, enhancer-promoter pairing, and the combined action of multiple enhancers. A pair of TF sites is shown here for the sake of simplicity, but developmental enhancers in the wild are, as a rule, larger and much more complex.

In part, the DNA sequence of an enhancer encodes a pattern of gene expression through the combinations of sequence-specific regulators that bind to it (and the non–DNA-binding factors they recruit, in turn). However, *cis*-regulatory DNA sequences do more than merely determine which combinations of TFs will be recruited to an enhancer. The linear arrangement of binding sites (sometimes called "grammar" or "syntax") can play an important role in controlling enhancer output, especially by setting thresholds for gene activation (4–7). In addition, the affinity with which a TF binds to its DNA site, which is controlled, in part, by the particular sequence of that binding motif, is increasingly recognized as a crucial factor in both the quantitative and qualitative control of gene expression (8–16). In PNAS, Farley et al. (17) investigate how the binding affinity and the spatial arrangement of TF binding sites within enhancers interact to encode precise patterns of gene expression in developing embryos.

The authors of this study build on the results of their recently published high-throughput screen (14) in which hundreds of thousands of short DNA sequences were tested for enhancer activity in transfected *Ciona* embryos. The library of tested DNAs were all variants of Otx-a, a 69-bp neural enhancer of the *Ciona Otx* gene (18), in which the 4-bp cores of five binding sites for GATA and ETS transcription factors were preserved but all other base positions were randomized. One synthetic enhancer variant drove strong expression of a GFP reporter gene in the notochord, a tissue that does not normally express *Otx*. Sequence analysis suggested two likely explanations for the ectopic expression pattern: a new ETS motif with high predicted binding affinity or a new site predicted to bind the transcription factor ZicL. Both ZicL and an ascidian ETS ortholog are expressed in the notochord.

[a]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109

Author contributions: S.B. wrote the paper.
The author declares no conflict of interest.

See companion article on page 6508.

[1]Email: sbarolo@umich.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1606109113

Farley et al. (17) conducted gain-of-function and loss-of-function mutational analyses in the context of several enhancers, demonstrating that both ETS and ZicL sites contribute to notochord expression. Further, they find that altering the spacing between ETS and ZicL sites by as little as 1 bp, or reversing the orientation of a single binding site, strongly affects the ability of these sites to activate gene expression.

Within the wild-type Otx-a enhancer, certain ETS and GATA sites have relatively poor predicted binding affinity but optimal pairwise spacing, whereas other sites have higher affinity but suboptimal spacing (14). The discovery of a synthetic Otx-a variant with ectopic notochord activity, and comparisons against other synthetic modules with slightly different compositions, suggest that similar trade-offs can modulate gene expression in the notochord (17). This observation led Farley et al. (17) to ask an intriguing question: Can we use what we have learned about constraints on TF binding site identity, affinity, and arrangement to find new notochord enhancers in the *Ciona* genome, using sequence data alone? By first screening for genomic regions containing combinations of ETS and ZicL sites, and then focusing on those clusters with strong binding sites or optimal syntax, Farley et al. (17) identify two new enhancers of the notochord-expressed genes *Mnx* and *Brachyury*. Importantly, these tissue-restricted enhancers have either high-affinity ETS/ZicL sites or optimal spacing, but not both: Optimizing both affinity and syntax causes ectopic expression (17).

It has been suggested that computational methods of enhancer prediction, which often emphasize the highest quality TF binding motifs and/or the strongest biochemical TF binding signals, may fail to capture a large proportion of functional enhancers, namely, enhancers with binding sites of low to moderate affinity (13, 19). Taking the rules of syntax into account may provide an essential source of information for predicting enhancer outputs (4, 6). However, the importance of syntax/grammar for enhancer function is still debated (5, 6, 20); what has been needed is more experimental data. In their new report, Farley et al. (17) propose an interesting idea: Because current methods of enhancer prediction focus on high-affinity TF binding sites, perhaps they are best at identifying those enhancers that are the least subject to constraints on binding site syntax. If true, this approach would tend to paint a biased picture of the constraints that shape enhancer sequences. Now that a more complete view of the relationship between DNA sequence and enhancer function is slowly emerging (Fig. 1), future predictive methods may be more effective at computing gene expression patterns from genomic DNA sequences. After all, that is a task that cells perform every day, with astonishing reliability and precision.

1 Nirenberg M, et al. (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci USA* 53(5):1161–1168.
2 Levine M, Cattoglio C, Tjian R (2014) Looping back to leap forward: Transcription enters a new era. *Cell* 157(1):13–25.
3 Heinz S, Romanoski CE, Benner C, Glass CK (2015) The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16(3):144–154.
4 Erives A, Levine M (2004) Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc Natl Acad Sci USA* 101(11):3851–3856.
5 Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 94(5):890–898.
6 Swanson CI, Evans NC, Barolo S (2010) Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev Cell* 18(3):359–370.
7 Weingarten-Gabbay S, Segal E (2014) The grammar of transcriptional regulation. *Hum Genet* 133(6):701–711.
8 Driever W, Thoma G, Nüsslein-Volhard C (1989) Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* 340(6232):363–367.
9 Jiang J, Levine M (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* 72(5):741–752.
10 Gaudet J, Mango SE (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* 295(5556):821–825.
11 Swanson CI, Schwimmer DB, Barolo S (2011) Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr Biol* 21(14):1186–1196.
12 Parker DS, White MA, Ramos AI, Cohen BA, Barolo S (2011) The *cis*-regulatory logic of Hedgehog gradient responses: Key roles for gli binding affinity, competition, and cooperativity. *Sci Signal* 4(176):ra38.
13 Ramos AI, Barolo S (2013) Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos Trans R Soc Lond B Biol Sci* 368(1632):20130018.
14 Farley EK, et al. (2015) Suboptimization of developmental enhancers. *Science* 350(6258):325–328.
15 Crocker J, Noon EP, Stern DL (2016) The soft touch: Low-affinity transcription factor binding sites in development and evolution. *Curr Top Dev Biol* 117:455–469.
16 Lorberbaum DS, et al. (2016) An ancient yet flexible *cis*-regulatory architecture allows localized Hedgehog tuning by *patched/Ptch1*. *eLife* 5:e13550.
17 Farley EK, Olson KM, Zhang W, Rokhsar DS, Levine MS (2016) Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc Natl Acad Sci USA* 113:6508–6513.
18 Bertrand V, Hudson C, Caillol D, Popovici C, Lemaire P (2003) Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* 115(5):615–627.
19 Jaeger SA, et al. (2010) Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics* 95(4):185–195.
20 Junion G, et al. (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148(3):473–486.