REGULATORY ELEMENTS

Putting enhancers into context

Regulatory elements — particularly transcriptional promoters and enhancers — typically consist of multiple binding sites for transcription factors (TFs) and are often detected and interpreted based on the occurrence of consensus, high-affinity binding motifs for TFs. A new study highlights that, beyond TF binding affinity, the wider genomic context and arrangement of TF binding sites is crucial in tissue-specific enhancers.

Farley *et al.* sought to understand the regulatory logic underlying tissue-specific gene expression in the sea squirt *Ciona intestinalis*. This species has long been a model invertebrate chordate in developmental biology owing to its simple and translucent body plan. In previous work, the authors studied synthetic variants of the neural-plate-specific orthodenticle homeobox a (Otx-a) enhancer and found that it operates under the principle of 'suboptimization': the enhancer consists of low-affinity TF binding sites in a suboptimal syntax (in terms of flanking nucleotides and spacing between the sites). Optimizing either the affinities or the syntax of the sites diminished tissue specificity by making the enhancer hyper-responsive to developmental cues and caused ectopic expression of the GFP reporter gene in inappropriate tissues such as the notochord.

In the present study, the authors started by dissecting the determinants of how one of the enhancers, random synthetic Otx-a 6 (RS 6), drives gene expression in the notochord. By manipulating the constituent TF binding sites they showed that two ETS (erythroblast transformation-specific) binding sites and a ZicL (zinc-finger protein of the cerebellum 3-like) binding site were necessary and sufficient to drive gene expression in the notochord. Such combinatorial logic, involving a widespread signal (ETS activation through fibroblast growth factor signalling) and a different tissue-localized determinant (ZicL) is an emerging theme of how tissue-restricted gene expression is achieved in different organisms.

Interestingly, from the enhancer library, only 2 out of 15 enhancers with 2 ETS binding sites and a ZicL binding site were able to drive reporter expression in the notochord, indicating that the activity of the enhancer is highly dependent on the arrangement and wider context of these binding sites. Further manipulations of the enhancers confirmed key roles for binding site orientation, spacing and flanking nucleotides.

Based on the 'regulatory code' of these context-sensitive criteria, the investigators formulated a computational program for identifying putative notochord enhancers, and found that 69 enhancers were predicted in the *C. intestinalis* genome. None of these enhancers consisted of binding sites with both optimal binding affinities and optimal wider syntax, which provides further support for the pervasiveness of suboptimization in driving tissue-restricted expression, in which there is a trade-off between binding affinities and wider syntax of the binding sites. Indeed, when functionally investigating one of the predicted enhancers (upstream of the motor neuron restricted (*Mnx*) gene), optimizing both the affinities and spacing of the binding sites caused the reporter gene to lose notochord tissue specificity.

For another of the predicted enhancers (upstream of the brachyury gene), binding sites had low affinity (based on their deviation from consensus sequences); however, their optimal arrangement achieved strong and localized reporter expression in the notochord. Thus, syntax is crucial and can compensate for low-affinity TF binding sites.

An important implication of this work is that predictions of functional enhancers based on high-affinity TF binding sites — either predicted bioinformatically from consensus sequences or identified as strongly bound sites from chromatin immunoprecipitation followed by sequencing (ChIP–seq) experiments — may fail to identify many functionally relevant enhancers, especially if suboptimization is a widespread feature of enhancers. Thus, computational tools and functional assays for enhancer identification need to be sensitive to the wider context of constituent TF binding sites. Given the abundance of human-disease-associated genetic variation in non-coding regions, consideration of syntax may allow those variants causing pathological gene dysregulation to be more accurately pinpointed.

Darren J. Burgess

ORIGINAL ARTICLE Farley, E. K. *et al.* Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl Acad. Sci. USA.* http://dx.doi.org/10.1073/pnas.1605085113 (2016)
FURTHER READING Levo, M. & Segal, E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* **15**, 453–468 (2014)