Housekeeping and tissue-specific cis-regulatory elements: Recipes for specificity and recipes for activity

Marta Russo, Gioacchino Natoli & Serena Ghisletti

Taylor & Francis
Taylor & Francis Group

Check for updates

POINT-OF-VIEW

Housekeeping and tissue-specific cis-regulatory elements: Recipes for specificity and recipes for activity

Marta Russo[a], Gioacchino Natoli[a], and Serena Ghisletti[b]

[a]Department of Biomedical Sciences, Humanitas University (Hunimed), Rozzano, Milan, Italy; [b]Humanitas Clinical and Research Center, Rozzano, Milan, Italy

ABSTRACT

Cell type-specific and housekeeping enhancers and promoters collectively control the transcriptional output of mammalian cells. Recent data clarify how DNA sequence features on the one hand control functional coupling of promoters with selected enhancers, and on the other impart high level of activity to a broad range of regulatory elements.

Introduction

The transcriptional output of eukaryotic cells is regulated by a complex interplay between chromatin modifications and transcription factors (TFs) acting at thousands of promoters and enhancers (heretofore referred to as *cis*-regulatory elements). The spatio-temporal regulation of gene expression is fundamental to cell differentiation and maintenance of cell identity, enabling the formation of different cell types with specialized functions. Moreover, the integration of developmental and environmental signals at *cis*-regulatory elements enables cell type-specific transcriptional responses to identical stimuli, which is another critical attribute of cell differentiation. Genomic *cis*-regulatory regions act as platforms for the recruitment of cell type-restricted and broadly expressed TFs[1,2] and include cell type-specific and housekeeping elements. Over the past few years, hundreds of thousands of *cis*-regulatory elements have been annotated in different cells of multicellular organisms. However, our functional understanding of how different combinations of TF motifs and additional DNA sequence-encoded features[3] eventually determine the distinctive functional properties of each of these elements, is still very incomplete and no accurate predictive models of enhancer or promoter activity are available. We discuss here recent significant advances in this area,

focusing on two main aspects. First, how enhancers can be functionally coupled to only a subset of core promoter elements, thus creating a layer of functional specificity; and second how a limited number of TFs are broadly used to impart high constitutive activity to both promoters and enhancers, irrespective of their housekeeping or tissue-specific activity.

Constitutively active core promoters and enhancers

In the past few years, the availability of a large panel of genome-wide transcriptomic and epigenomic data sets has contributed to increase our knowledge of the basic principles of transcriptional regulation both in lineage specification and in the control of inducible gene expression. The current paradigm assumes that cell type-specific transcriptional outputs (and thus functional properties) reflect the collective activity of thousands of active *cis*-regulatory elements that control the expression of genes that are specific to that cell type as well as the appropriate level of transcription of housekeeping genes. Stimulus-activated TFs operate within a cell type-specific landscape of accessible *cis*-regulatory elements, which is generated by the unique combination of lineage-determining TFs specifically expressed in that cell type. Active enhancers and promoters share several features, including relative nucleosome depletion, high levels of histone acetylation, the ability to

initiate transcription[4] and an overall similar sequence organization, consisting of the variable combination of motifs recognized by sequence-specific TFs and core promoter elements.[3–7]

Transcription initiates at core promoters, short sequences located in close proximity to the transcription start site (TSS) at the 5' end of genes and that can facilitate the recruitment of RNA polymerase II (Pol II), the assembly of the pre-initiation complex (PIC) and in specific cases control the accurate positioning of initiation and the direction of transcription.[8,9] Core promoter elements include the TATA box,[10,11] Initiator (Inr)[12] and the TFIIB-recognition elements (BREs).[13,14] Such classical core promoter elements, however, are not the only motifs specifically enriched in close proximity of TSS. Computational analyses of the DNA sequences in the entire complement of human gene promoters[15,16] revealed that a very few TF consensus sites are specifically over-represented within 50 bp from mapped TSS, suggesting a possible role in the control of transcription initiation. The TF motif with the strongest enrichment in the immediate vicinity (<50 bp) of the TSS is the canonical consensus DNA binding site for ETS proteins (5'CCGGAAGT3'), a metazoan-restricted family of TFs. Such a close proximity of ETS motifs to the transcription start sites suggested their direct role in the control of basic promoter properties, a hypotesis that has recently received experimental validation: a subset of ETS motifs are critical to control high constitutive transcriptional activity of promoters and enhancers, irrespective of their housekeeping or tissue-specific activity.[17]

Housekeeping vs. developmental cis-regulatory elements

Recent data point to the notion that enhancer-core promoter specificities are determined by TF motifs present in functionally coupled regulatory regions. A striking example of how specific motifs in core promoters restrict their ability to work with a selected set of enhancers comes from studies in *Drosophila*. Using STARR-seq (self-transcribing active regulatory region-sequencing), Stark and colleagues analyzed the activity of hundreds of thousands enhancer candidates towards core promoters with either housekeeping or developmental activity.[18,19] In this

high-throughput method, randomly sheared genomic DNA fragments are cloned into transfection vectors at a downstream "enhancer-like" location so that any resulting transcripts contain the sequence of the enhancer that stimulated their expression. Vector-derived cDNA sequences are then mapped to the genome, identifying enhancers that are active in a specific cell type. The authors used two different core promoters, one housekeeping -derived from the ribosomal protein gene *RpS12* and containing the TCT/DRE motif, which is commonly enriched in the promoters of housekeeping genes in *Drosophila*- and one developmental -derived from the tissue-specific *even-skipped gene (eve)* and containing tha TATA, Inr, MTE and DPE motifs. Significantly, at genome scale the two types of core promoter elements showed the selective ability to support transcription activation by only housekeeping or developmental enhancers, respectively (Figure 1). The general notion coming from these data is that for distal *cis*-regulatory elements to work, proximity to a promoter is not enough: they need to find a core promoter partner with a distinctive combination of motifs that enables functional pairing.

Enhancer-core promoter specificity within chromatin domains

Interactions of enhancers with target promoters preferentially occur within a defined physical space
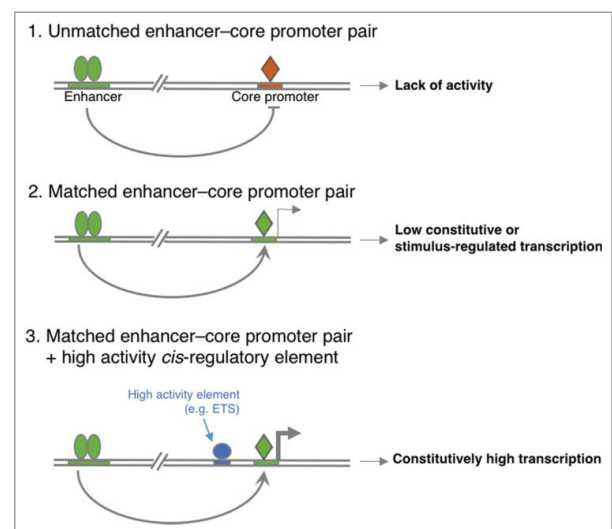

Figure 1. The schematic depicts two distinct but intertwined layers of control at *cis*-regulatory elements: the functional coupling of promoters with selected enhancers, which can be further activated in presence of high activity motifs such as ETS.

constrained by three-dimensional interactions among distant regions.[20-24] Currently, it is still unaddressed whether the chromatin architecture of housekeeping and developmental loci follows similar principles. In *Drosophila*, housekeeping enhancers tend to form multi-TSS interaction networks and to be associated with the borders of TADs, while developmental enhancers are more often bound to a few TSS and are enriched at chromatin loop anchors, regions of enriched chromatin interactions within TADs.[25]

In addition, modulating DNA accessibility within chromatin has been shown to be an important layer of specification of enhancer-core promoter contacts. In this regard, the occupancy and accessibility of nucleosomes at tissue-specific and housekeeping regulatory sequences were shown to have specific properties, at least in the mouse liver. Specifically, tissue-specific enhancers were shown to retain MNase-accessible nucleosomes significantly more than active promoters and ubiquitous enhancers. In this context, the pioneer factors FoxA1 and FoxA2 are responsible to keep nucleosmes accessible at liver-specific enhancers allowing other TFs to bind and stimulate transcription.[26]

ELF proteins control housekeeping and tissue-specific cis-regulatory elements

Data discussed above point to DNA sequence-instructed functional divergences among different *cis*-regulatory elements, with housekeeping enhancers being selectively able to activate transcription from housekeeping core promoters, and developmental or tissue-specific enhancers being instead able to activate transcription from developmental core promoters. Other data point to the only apparently conflicting notion that regulatory elements with a different range of activity across cell types (from highly tissue-specific to broadly acting elements) coopt a limited number of identical TFs to efficiently promote transcription. As mentioned above, ETS motifs are over-represented in close proximity of mammalian TSS,[15,16] indicating that selected sequence-specific TFs (among the hundreds expressed in eukaryotes cells) may act as broadly used facilitators of early steps in transcription. Interestingly, ancestors of ETS proteins appeared in primitive eukaryotes as bridges linking core promoters to components of the transcriptional machinery, thus directly enabling early steps in transcription initiation.[27]

In a well-characterized and highly differentiated metazoan cell type, namely primary macrophages,[28] a specific subset of ETS motifs recognized by ELF proteins and GABPA was found to be enriched in distinct classes of *cis*-regulatory elements whose only apparently common property was to have a constitutively high activity: housekeeping promoters and highly active tissue-specific enhancers.[17] Consistently, ChIP-seq analyses showed that the ELF subfamily of ETS proteins bound both the promoters of housekeeping genes and a specific subset of macrophage-specific enhancers characterized by very high acetylation levels and RNA Polymerase II binding.[17] Importantly, both in the context of transfected reporter vectors and at the endogenous genomic loci, point mutations in ELF binding sites were sufficient to determine a strong and in many cases almost complete loss of activity. While the mechanisms responsible for such a critical role of ETS proteins in constitutive activity of disparate regulatory elements remain to be fully elucidated, a DNA affinity purification approach coupled to mass spectrometry analysis showed an ETS-dependent recruitment of chromatin modifiers and regulators of transcriptional elongation.[17]

Concluding remarks and future directions

The data discussed in this commentary indicate the existence of two distinct layers of control that intertwine at *cis*-regulatory elements. On the one hand, specific combinations of TF motifs at enhancers and core promoters determine their ability to functionally interact, implying that proximity between an enhancer and a promoter is a required but not sufficient condition to promote transcription activation. However, a precise understanding of the rules underlying functional coupling between enhancers and core promoters is still unavailable. Moreover, although we can expect that similar principles apply to gene regulation in metazoans, definitive evidence in this direction is still lacking. If a simple code exists that controls functional coupling of enhancers and promoters, decrypting it will have an enormous impact on our understanding of transcriptional control and it will allow determining fundamental rules at the basis of gene expression control.

On the other hand, cooptation of a limited number of TFs (such as the ELFs) may represent a

transversal strategy broadly adopted across cell types to equip *cis*-regulatory elements with different functional roles and specificity with a common property: the ability to efficiently promote transcription (Figure 1). It follows that the absence of such motifs in other *cis*-regulatory region may represent a prerequisite for their tighter and dynamic regulation in response to specific micro-environmental or developmental cues. Therefore, the combination of specificity determinants and activity determinants eventually controls spectrum and level of activity of each genomic *cis*-regulatory element.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Funding

References

1. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15:272–86. doi:10.1038/nrg3682.

2. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13:613–26. doi:10.1038/nrg3207.

3. Barozzi I, Simonatto M, Bonifacio S, Yang L, Rohs R, Ghisletti S, et al. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Mol Cell. 2014;54:844–57. doi:110.1016/j.molcel.2014.04.006.

4. 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;46:1311–20. doi:10.1038/ng.3142.

5. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010;8:e1000384. doi:10.1371/journal.pbio.1000384.

6. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465:182–7. doi:10.1038/nature09033.

7. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011;474:390–4. doi:10.1038/nature10006.

8. Kadonaga JT. Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol. 2012;1:40–51. doi:10.1002/wdev.21.

9. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 1996;21:327–35. doi:10.1016/S0968-0004(96)10050-510.1016/0968-0004(96)10050-5.

10. Lifton RP, Goldberg ML, Karp RW, Hogness DS. The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb Symp Quant Biol. 1978; 42;Pt 2:1047–51.

11. Goldberg ML. "Sequence analysis of Drosophila histone genes.". PhD thesis, Stanford University, San Francisco 1979.

12. Smale ST, Baltimore D. The "initiator" as a transcription control element. Cell. 1989;57:103-

13. Deng W, Roberts SG. A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. 2005;19:2418–23.

14. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev. 1998;12:34–44.

15. FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. Comparative genomics of Drosophila and human core promoters. Genome Biol. 2006; 7:R53.

16. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, et al. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 2005;434:338–45.

17. Curina A, Termanini A, Barozzi I, Prosperini E, Simonatto M, Polletti S, et al. High constitutive activity of a broad panel of housekeeping and tissue-specific cis-regulatory elements depends on a subset of ETS proteins. Genes Dev 2017;31:399–412.

18. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 2013; 339:1074–7.

19. Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature. 2015;518:556–9.

20. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012; 485:376–80.

21. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 2014; 159:374–87.

22. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 2012; 485:381–5.

23. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 2012; 148:458–72.

24. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature 2012; 488:116–20.

25. Cubenas-Potts C, Rowley MJ, Lyu X, Li G, Lei EP, Corces VG. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res. 2017;45:1714–30.

26. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell 2016;62:79–91.

27. Schumacher MA, Lau AO, Johnson PJ. Structural basis of core promoter recognition in a primitive eukaryote. Cell 2003;115:413–24.

28. Monticelli S, Natoli G. Transcriptional determination and functional specificity of myeloid cells: making sense of diversity. Nat Rev Immunol. 2017;17:595–607.