Activating genes at long distances

Lan T M, Ariel O Galindo-Albarrán, Jaime A Castro-Mondragon, Charlotte Andrieu-Soler, Alejandra Medina-Rivera, Charbel Souaid , Guillaume Charbonnier, Aurélien Griffon, Laurent Vanhille, Tharshana Stephen, Jaafar Alomairi, David Martin, Magali Torres, Nicolas Fernandez, Eric Soler, Jacques van Helden, Denis Puthier & Salvatore Spicuglia*

* Corresponding author: salvatore.spicuglia@inserm.fr

A very intriguing question since early days of molecular biology is how do cells know what function to perform? For instance, how does a cell know if it should become a neuron or a lymphocyte? The answer is given by the genes, sequences of DNA containing the information to produce proteins or RNA molecules, and the way they are regulated. Humans have around 18,000 genes, but not all of them are active in all the cells, for example, the active genes in a neuron are not the same than in a lymphocyte. Gene regulation is a complex process that we fairly understand, it's highly dynamic within cells and tissues, and changes through time (development).

In the classical model of eukaryotic gene regulation, genes are activated through the interaction of regulatory proteins and two kinds of regulatory sequences. These sequences are classified by their biochemical properties and relative distance to the regulated gene in: i) the proximal (500 nucleotides away), known as promoters, and ii) the distal ones (thousands of nucleotides away) called enhancers. Enhancers interact physically with promoters to activate genes (E-P interaction).

For a long time, promoters and enhancers were considered as distinct classes of sequences, however in Dao, L.T. M *et al.*, 2017[1] , we challenged such classification by demonstrating that hundreds of promoters in human cells indeed activate distal genes by a promoter-promoter (P-P) interaction.

In this study, we have developed a method to quantify the enhancer activity for any sequence of the human genome (asking if a given sequence may activate genes at long distances). We focused on ~18,000 promoters (one promoter per gene), in two different cell types and by using this method, we identified ~400 candidate promoters with potential enhancer activity.

Furthermore, using a genomic engineering approach, we observed a reduction in the expression of target genes after mutating many of our candidate promoters, indicating that this special set of promoters may activate genes at long distances (as enhancers do), therefore we called *e-promoters*. Many of the regulated genes by these e-promoters are associated to a rapid activation, for example, to stress-response activated by viruses or heat-shock.

Our results and other recent studies[2] have challenged the classical view of gene regulation, instead we propose a unified model of gene regulation, where enhancers and promoters are seen as a single class of regulatory elements, whose differences rely on the regulatory proteins they interact with and not anymore by the distance to the regulated genes.

1. Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., … Spicuglia, S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nature Genetics, 10. https://doi.org/10.1038/ng.3884

2. Rennie, S., Dalby, M., Lloret-Llinares, M., Bakoulis, S., Dalager Vaagensø, C., Heick Jensen, T., & Andersson, R. (2018). Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities. Nucleic Acids Research, 46(11), 5455–5469. https://doi.org/10.1093/nar/gky244