FEBS Letters xxx (2014) xxx–xxx

journal homepage: www.FEBSLetters.org

Review

The intriguing complexities of mammalian gene regulation: How to link enhancers to regulated genes. Are we there yet?

Bence Daniel [a], Gergely Nagy [a], Laszlo Nagy [a,b,c,*]

[a] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Egyetem tér 1., Debrecen H-4010, Hungary
[b] MTA-DE "Lendület" Immunogenomics Research Group, University of Debrecen, Egyetem tér 1., Debrecen, Hungary
[c] Sanford-Burnham Medical Research Institute, 6400 Sanger Road, Orlando, FL 32827, USA

ARTICLE INFO

Article history:
Received 4 May 2014
Revised 22 May 2014
Accepted 22 May 2014
Available online xxxx

Edited by Wilhelm Just

Keywords:
Gene expression regulation
Enhancers
Genomic
Transcriptome
Cis-regulatory elements
Chromosome conformation

ABSTRACT

The information encoded in genomes supports the differentiation and function of the more than 200 unique cell types, which exist in various mammalian species. The major mechanism driving cellular differentiation and specification is differential gene expression regulation. Cis-acting enhancers and silencers appear to have key roles in regulating the expression of mammalian genes. However, these cis-acting elements are often located very far away from the regulated gene. Therefore, it is hard to find all of them and link them to the regulated gene. An intriguing and unresolved issue of the field is to identify all of the enhancers of a particular gene and link these short regulatory sequences to the genes they regulate and thus, reliably identify gene regulatory enhancer networks. Recent advances in molecular biological methods coupled with Next-Generation Sequencing (NGS) technologies have opened up new possibilities in this area of genomics. In this review we summarize the technological advances, bioinformatics challenges and the potential molecular mechanisms allowing the construction of enhancer networks operating in specific cell types and/or activated by various signals.

1. Why mammalian gene regulation is so hard to study?

Gene expression is the result of a very complex process achieved by the coordinated action of multiple layers of regulators. In prokaryotes, genes are organized into operons, and using a single promoter, the entire stretch of DNA is transcribed into RNA (Fig. 1A). These so-called "polycistronic" RNAs often encode functionally related members of an enzyme cascade regulating a particular metabolic process. Thus, in most of the cases prokaryotes utilize one promoter-proximal, restricted cis-element to initiate transcription. However, eukaryotic organisms have evolved to utilize much more complex mechanisms to regulate gene expression. In fact, one of the driving forces of eukaryotic evolution is believed to be the introduction of elaborate gene regulatory circuits. This is, in part, manifested in the concept of C value enigma, which is the observation that genome size does not correlate with organismal complexity [1]. The number of protein coding genes also does not show correlation with complexity.

Unlike prokaryotic genes, eukaryotic genes are "monocistronic" and their regulation is usually much more complex. The fact that genes can have multiple promoters with unique promoter elements makes the picture even more crowded and complicated. In addition, probably most if not all eukaryotic genes possess intergenic, as well as intragenic cis-regulatory elements (enhancers/silencers) to fine-tune their expression in a cell type and/or biological context dependent manner (Figs. 1B and 2B). To clearly understand the detailed molecular mechanisms controlling gene expression, one needs to identify the factors responsible for gene regulation and their precise action on one or multiple well-defined cis-regulatory element(s). Before the genomic era, investigations were limited methodologically and gene regulation was studied with methods restricted to the analysis of the expression of a few genes and/or biased/restricted to the immediate vicinity of a given gene. The classical way of studying gene expression used a set of so-called "promoter bashing approaches" evaluating the genomic regions in the close proximity of the transcription start site (TSS) to identify the core sequence driving the expression of the given gene [2]. These approaches were based on transient transfections and deletion and insertion mutagenesis, which are still used. The discovery of enhancers predicted the complex regulation of a gene, because these cis-regulatory elements can

* Corresponding author at: Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Egyetem tér 1., Debrecen H-4010, Hungary. Fax: +36 52 314 989.
E-mail address: nagyl@med.unideb.hu (L. Nagy).

http://dx.doi.org/10.1016/j.febslet.2014.05.041
0014-5793/

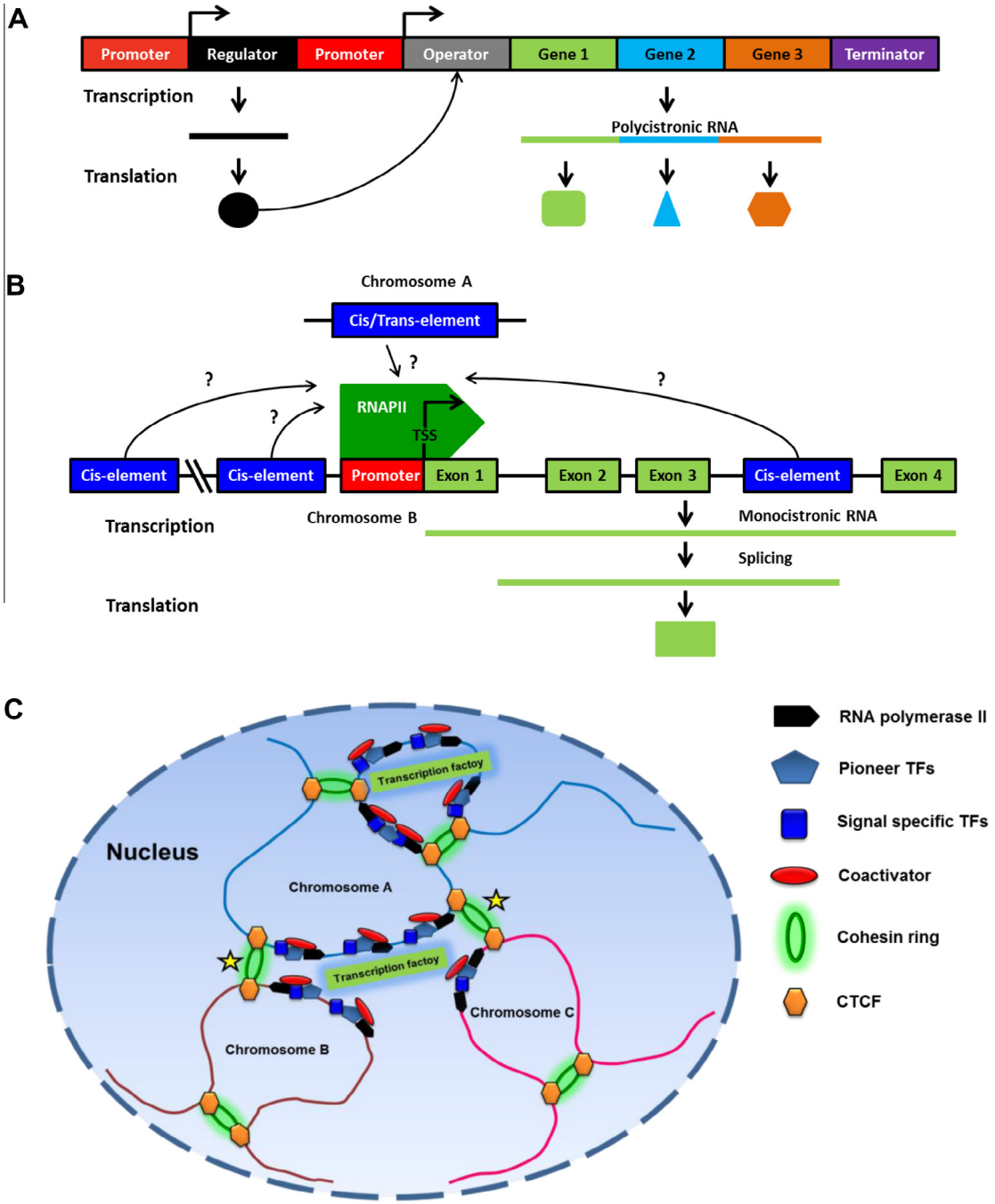

Fig. 1. Transcription regulation in prokaryotes and eukaryotes and the complexity of gene regulation in eukaryotic cells. (A) Prokaryotic gene expression is based on operons in which a gene cluster is under the control of two genomic (promoter) regions. The first promoter located at the 5′ end is responsible for the expression of the regulator protein, which in turn silences the whole operon via binding to the operator region. In the presence of an activating stimulus the regulator cannot bind to the operator region, thus the second promoter will be active and leads to the efficient expression of the enzyme coding genes producing polycistronic RNA molecules encoding more protein products. (B) Eukaryotic genes are typically regulated by cis-acting elements located in the non-coding part of the genome. These elements can be located far away from their target genes, might be even in another chromosome, thus it is challenging to pair them with their genes. RNA synthesized from a eukaryotic gene is monocistromic and undergoes the process called splicing, in which the intronic regions are excluded from the nascent transcript before translation. (C) Eukaryotic transcription might be coordinately regulated in the so-called transcription factories. These subnuclear compartments might be, in part, stabilized by the CTCF/cohesin protein complexes and permit the expression of genes in a well-coordinated manner, mechanically connecting genes regulated by the same signal, but residing on distinct chromosomes. A hypothetical scenario is depicted in which several genomic regions on distinct chromosomes are linked by CTCF/cohesin interactions and co-localize in the nucleus forming a transcription factory. Interchromosomal interactions are marked by asterisks.

be located, at least in principle, far away and either upstream 5′ or downstream 3′ of the regulated gene [3]. The inherent pitfalls of the first studies were recognized soon, but without the technical advances no one could easily go beyond the technical limitations to dissect the regulation of gene networks, not even just a particular gene. Therefore, most studies identified promoter proximal elements usually restricted to 10–20 kb upstream of the designated promoter. Using these typical "promoter bashing" technologies researchers were not able to consider and/or evaluate the contribution of multiple and/or far way enhancers or intra- or interchromosomal interactions. This represented bias and created a roadblock in understanding complex gene regulation.

An additional important aspect of gene expression regulation research is that one would need methods to assess the expression

Please cite this article in press as: Daniel, B., et al. The intriguing complexities of mammalian gene regulation: How to link enhancers to regulated genes. Are we there yet? FEBS Lett. (2014), http://dx.doi.org/10.1016/j.febslet.2014.05.041