

Minireview

Gene expression neighborhoods

Brian Oliver, Michael Parisi and David Clark

Address: Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

Correspondence: Brian Oliver. E-mail: oliver@helix.nih.gov

Published: 1 July 2002

Journal of Biology 2002, 1:4

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/1/1/4>

© 2002 BioMed Central Ltd ISSN 1475-4924

Abstract

The finding that neighboring eukaryotic genes are often expressed in similar patterns suggests the involvement of chromatin domains in the control of genes within a genomic neighborhood.

Reductionist approaches have been a tremendous boon to understanding the regulation of transcription, one of the vital steps defined by the central dogma of molecular biology. Gene-by-gene analysis has clearly shown that control regions within the DNA sequence bind protein transcription factors that up- or down-regulate the activity of promoters. But now that patterns of gene expression can be studied across the entire genome, new findings suggest that, as well as being controlled individually, genes may also be subject to regulation according to their location within the genome.

It has been clear for some time that genomic location has some impact on gene expression. For example, in various species when transgenes are removed from their local environment and reinserted elsewhere in the genome the transgenes tend to work more-or-less normally but almost always show some alteration in expression due to insertion site - and sometimes the effect on expression is dramatic. That even subtle differences in gene expression can have consequences in some circumstances is also well known, and is illustrated by the dramatic effects of minute concentration differences in the gradients of pattern-determining morphogens during development [1], and in the dosage compensation mechanisms that have evolved to ensure that X-linked genes are expressed at similar levels in male and female animals [2].

In this issue, Spellman and Rubin [3] describe a transcriptional profiling study that reveals a surprising correlation between the organization of genes along *Drosophila* chromosomes and their expression levels. Specifically, neighborhoods composed of an average of 15 contiguous genes show markedly similar relative expression levels. Although the average neighborhood contains 15 genes, there is a very wide range. These neighborhoods are not obviously composed of genes with related functions that might be expected to exhibit co-regulation, as is the case for the rRNA, histone, Hox, and globin gene clusters.

Two other recent papers also suggest that genes with similar expression levels are non-randomly distributed, in this case within the human genome [4,5]. In humans, it has been suggested recently that expression neighborhoods serve to regulate housekeeping functions [5]. In *Drosophila* this is less likely, however, because Spellman and Rubin [3] demonstrate that embryos and adults differ dramatically in the organization of their neighborhoods of similarly expressed genes (although one could argue about whether the vermiform *Drosophila* larvae and adults might be expected to show two different housekeeping gene sets). The compelling and intriguing *Drosophila* data are rather mysterious and warrant closer examination: what could underlie the observed similarity of gene expression within neighborhoods?

Perhaps the simplest explanation is that co-regulation within an expression neighborhood may be due to incidental interactions between promoters and transcriptional enhancers (Figure 1a). In this model, transcription of one or more genes in a genomic cluster is regulated by the usual suspects (transcription factors) binding at the appropriate sites and activating nearby genes as well as the target gene - and the resulting inappropriate expression of genes other than the target is tolerated because it has little biological effect. If this is the case then, if sites that bind strong transcriptional activators, such as the yeast protein GAL4, were seeded in the *Drosophila* genome they should create new neighborhoods. Transcription factors have a limited range of effect [6], so if strong activators are responsible one might expect to see a steep fall-off in the effects of a given factor with distance from its core binding site (Figure 1a). But the data presented by Spellman and Rubin [3] suggest that in fact the pattern of gene expression within a neighborhood is essentially a 'square wave' (as shown in Figure 1b).

Spellman and Rubin [3] therefore favor a structural chromatin domain model (Figure 1b), involving the opening of the chromatin of an entire neighborhood as a result of activation of a target gene within the neighborhood. The creation of a domain of open chromatin structure would, it is argued [3], increase the availability of the promoters and enhancers of all the genes in the neighborhood to the transcriptional machinery, leading to correlated increases in expression. Such a domain could be delimited by boundary elements or insulators, accounting for the square wave profile (Figure 1b). A problem with this model is that increased chromatin accessibility is just as likely to facilitate the binding of repressors as activators, with the result that some genes would be up-regulated and some down-regulated. This is not consistent with neighborhoods of co-regulation. But if increased accessibility primarily affects basal (that is, non-activated) expression, there could be a general increase in transcription of all the genes in the neighborhood. Indeed, modification of the chromatin of the male X chromosome in *Drosophila* results in global up-regulation of gene expression [2], as does depleting histones from yeast [7]. And if neighborhoods influence all genes within them - and not just those that evolved so as to be regulated within a particular neighborhood - then inserted transgenes that land in a neighborhood should come under neighborhood control, and chromosome deletions and inversions should alter the extent of particular neighborhoods.

Spellman and Rubin [3] tested a short list of known chromosomal structures to look for correlations with expression neighborhoods. The cytology of *Drosophila* chromosomes and chromosome puffs has long suggested

that the chromosome is divided into loop domains with differing degrees of compaction. Indeed, heterochromatin and euchromatin were recognized long before we knew that

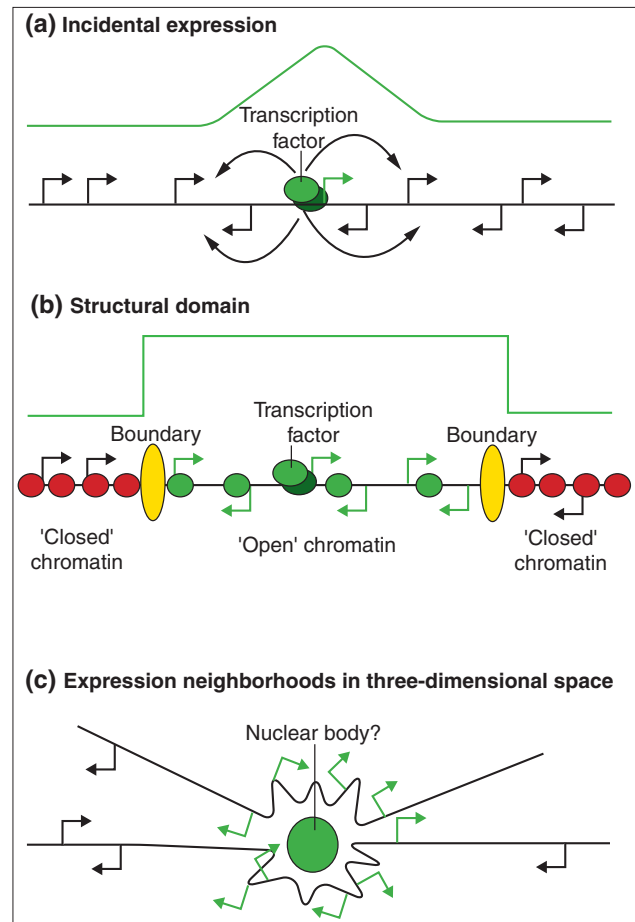


Figure 1
Models to account for gene expression neighborhoods. Several models (or combinations of models) could account for the observed phenomenon of gene expression neighborhoods. **(a)** Incidental regulation. A transcription factor (green oval) binds at a target gene (green arrow) and incidentally up-regulates neighboring genes. In this model, the level of expression of neighboring genes is determined by proximity to the target gene and is expected to decrease with distance from the target gene (the green line at the top of each panel indicates the gene expression profile across the neighborhood). **(b)** A structural domain model. A discrete 'open' chromatin domain is created as a result of activation of a target gene within the domain. Flanking boundary or insulator elements (yellow ovals) define the neighborhood and the limits of the open chromatin domain. (Note the 'square wave' expression profile.) **(c)** Expression neighborhoods in three-dimensional space. In this model, activation of a target gene results in its recruitment to a specific nuclear location. This would necessarily involve the co-recruitment of neighboring genes. The particular subnuclear location exposes the neighborhood to increased concentrations of components of the transcriptional machinery (the image shows two segments of chromatin with two neighborhoods in the vicinity of a (green) nuclear body).

chromosomes were the carriers of genetic information. Molecular biologists know that chromatin has various accessibility states and binds to a nuclear matrix at defined locations. Which of these is the structural basis of a neighborhood? The short and surprising answer appears to be 'none of the above'. Although the stunning block-like organization of neighborhoods along a chromosome [3] indicates that there must be *cis*-acting structures, no known structures correlate with the blocks. But it is increasingly clear that the nucleus is a highly organized three-dimensional space (Figure 1c). Sub-nuclear structures of various types, such as insulator bodies and the PML macromolecular bodies found in mammalian nuclei, may be distinct from structural elements such as loop-domain boundaries and matrix-attachment regions [8,9]. The hunt for the structural basis of expression neighborhoods will be an exciting one.

What do expression neighborhoods mean for the organism? One possibility, favored by Spellman and Rubin [3], is that they mean nothing. They suggest that although expression domains reveal some sort of structural feature, only one or a few genes in the neighborhood are *bona fide* targets. The bottom line for any would-be gene-expression profiler is that the 'interesting' genes identified in a microarray experiment are accompanied by a large amount of chaff. Spellman and Rubin suggest that the inappropriate expression of gene neighbors does no harm, an idea that is supported by the lack of dominant phenotypes when single genes are mutated. But it is also true that deletions removing greater than 1% of the *Drosophila* genome (around 140 genes) have severe dominant deleterious effects on the organism [10]. Such deletions are likely to remove whole neighborhoods.

It seems to us that expression neighborhoods should greatly favor the evolution of genes that benefit by being within that neighborhood. For example, a *de novo* function that is encoded in a gene is of no consequence if it is never expressed in a tissue that it could influence. As pointed out by Spellman and Rubin [3], the sequencing of related *Drosophila* species will allow us to determine whether neighborhood structures are maintained intact through evolutionary time. If the neighborhoods identified by Spellman and Rubin are less often broken by inversions than other non-neighborhood regions of the genome (assuming that there are indeed any non-structured regions), then neighborhoods are likely to be functionally significant. Expression neighborhoods could help create, capture and maintain gene function within a framework of expression defined by that neighborhood, providing evolution with additional tools with which to work. From this fascinating starting point we can expect further insights into the significance of gene-expression neighborhoods

and the mechanisms that generate them as more genomes are sequenced and more expression patterns studied over coming months.

References

1. Tabata T: **Genetics of morphogen gradients.** *Nat Rev Genet* 2001, **2**:620-630.
2. Pannuti A, Lucchesi JC: **Recycling to remodel: evolution of dosage-compensation complexes.** *Curr Opin Genet Dev* 2000, **10**:644-650.
3. Spellman PT, Rubin GM: **Evidence for large domains of similarly expressed genes in the *Drosophila* genome.** *Journal of Biology* 2002, **1**:5.
4. Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus MC, van Asperen R, Boon K, Voute PA, et al.: **The human transcriptome map: clustering of highly expressed genes in chromosomal domains.** *Science* 2001, **291**:1289-1292.
5. Lercher MJ, Urrutia AO, Hurst LD: **Clustering of housekeeping genes provides a unified model of gene order in the human genome.** *Nat Genet* 2002, **31**:180-183.
6. Dorsett D: **Distant liaisons: long-range enhancer-promoter interactions in *Drosophila*.** *Curr Opin Genet Dev* 1999, **9**:505-514.
7. Wu J, Grunstein M: **25 years after the nucleosome model: chromatin modifications.** *Trends Biochem Sci* 2000, **25**:619-623.
8. Carmo-Fonseca M: **The contribution of nuclear compartmentalization to gene regulation.** *Cell* 2002, **108**:513-521.
9. West AG, Gaszner M, Felsenfeld G: **Insulators: many functions, many mechanisms.** *Genes Dev* 2002, **16**:271-288.
10. Lindsley DL, Sandler L, Baker BS, Carpenter AT, Denell RE, Hall JC, Jacobs PA, Miklos GL, Davis BK, Gethmann RC, et al.: **Segmental aneuploidy and the genetic gross structure of the *Drosophila* genome.** *Genetics* 1972, **71**:157-184.