Gene 491 (2012) 284–288

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Are clustered genes in the genomes of *Arabidopsis* and *Drosophila* regulated differently?

Alexander M. Boutanaev [a,b,*], Lev G. Nemchinov [a]

[a] USDA/ARS, Plant Sciences Institute, Molecular Plant Pathology Laboratory, Beltsville MD 20705, USA
[b] Institute of Basic Biological Problems, Russian Academy of Sciences, Institute Street, 2, Pushchino, Moscow Region, 142290 Russia

ARTICLE INFO

Article history:
Accepted 24 September 2011
Available online 8 October 2011

Received by A.J. van Wijnen

Keywords:
Gene profiling
EST mapping
Transcription factors
Gene clusters
Evolution

ABSTRACT

In the eukaryotic genome, genes with similar functions tend to co-localize in close proximity. Such gene clusters together with non-clustered genes constitute a chromatin domain which is a higher order regulatory unit. On a lower level co-expressed genes are regulated by differential activity of transcription factors (TF). We compared genome-wide distributions of TF in gene clusters in the genomes of *Drosophila melanogaster* and *Arabidopsis thaliana*. This revealed a significant excess of TF genes in gene clusters of the *Arabidopsis* genome, whereas in the genome of *Drosophila* distribution of TF in gene clusters did not differ from stochastic. We speculate that these alternatives could lead to different pathways of regulation of clustered genes in two species and to evolutionary-progressive changes in architecture of regulatory networks, governing the activity of clustered genes in the animal kingdom.

1. Introduction

Distribution of genes along chromosomes in eukaryotes is far from random. One of the possible reasons is that functionally-related, non-homologous genes tend to co-localize in close proximity. Earlier studies have found such gene clusters in the human (Bortoluzzi et al., 1998; Ko et al., 1998), *Drosophila* (Boutanaev et al., 2002; Spellman and Rubin, 2002), mouse (Williams and Hurst, 2002) and, more recently, zebrafish (Ng et al., 2009) genomes. Clusters of co-expressed genes were also discovered in the *Arabidopsis* genome (Williams and Bowles, 2004) and were further investigated in Schmid et al. (2005), Zhan et al. (2006).

Apparently, there is a close relationship between chromatin domains and clusters of unrelated genes with similar function (Blanco et al., 2008; Zhan et al., 2006). The chromatin domain seems to present a regulatory unit of higher order consisting of both clustered and non-clustered genes, not all of them co-expressing (Dillon, 2006; Kalmykova et al., 2005; Sproul et al., 2005). On a lower level, fine-tuning of co-regulated gene expression in open chromatin, including clustered genes, occurs by differential activity of transcription factors (TF). Although the number of surveys in the field is growing fast, little is known about mechanisms of regulation of clustered genes, except that they are co-regulated. In particular, there is significant lack of data on the genome-wide regulation of such genes.

Understanding a higher-order organization of the eukaryotic genome and corresponding genetic regulatory networks is a real challenge. Since clusters of functionally-related genes constitute a considerable part of the global genome architecture, any new information on the regulatory mechanisms involved in their activity would be of substantial importance.

2. Results and discussion

In this work we examined the distribution of TF genes in gene clusters of two model evolutionarily distant genomes, *Drosophila melanogaster* and *Arabidopsis thaliana*. Firstly, we used genome-wide computer profiling based on the mapping of expressed sequenced tags (EST), as it was described earlier (Boutanaev et al., 2002, 2009). As DNA databases are constantly supplied both with new EST sequences and more exact versions of annotated genomes, it could be a good idea to keep upgrading computer-generated profiles.

A computer analysis of multiple EST collections publicly available through Genbank resulted in specific gene expression profiles of *Drosophila* and *Arabidopsis*. In the case of *Drosophila*, profiles were associated with larvae and pupae, testes, ovaries, embryos and head whereas, in the case of *Arabidopsis*, with roots, seedlings, ovules, silique and flowers, developing seeds and response to biotic stress. To build an expression profile associated with biotic stress responses,

Abbreviations: EST, expressed sequence tags; TF, transcription factors; GO, Gene Ontology; SA, salicylic acid; *AHA2*, H(+)-ATPASE 2 gene; *GLT1*, GLUCOSE TRANSPORTER 1 gene; *PAL2*, PHENYLALANINE AMMONIA-LYASE 2 gene; *HOS9/PFS2*, HOS9/PFS2, gene; *ERL1*, ERECTA-LIKE 1 gene; *SHP2*, SHATTERPROOF gene; *r-cup*, ryder cup gene; *Klp59D*, Klp59D gene; *gskt*, gasket gene; *cta*, concertina gene; *jing*, jing gene; *cact*, cactus gene; MADS, MADS-box protein family; C2H2, C2H2-box protein family.

* Corresponding author at: Institute of Basic Biological Problems, Russian Academy of Sciences, Institute Street, 2, Pushchino, Moscow Region, 142290 Russia. Tel.: +7 4967 732864; fax: +7 4967 330532.
E-mail address: boutanaev@mail.ru (A.M. Boutanaev).

0378-1119/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2011.09.023

heterologous EST were used as previously described (Boutanaev et al., 2009). Sequentially, these profiles allowed us to identify gene clusters as groups of neighboring co-expressed genes (Boutanaev et al., 2002). Finally, we determined TF/cluster ratio for clusters of different sizes of both species by dividing the number of TF in clusters by the number of clusters.

By definition, a cluster of functionally-related genes consists of co-regulated and specifically expressed gene members. In this regard, a criterion of specificity is important for the identification of gene clusters. Except for response to biotic stress, specificity of all other expression profiles originated from the sources of EST libraries. In the case of biotic stress, leaves or whole above-ground parts of the infected plants were used to generate EST. Unfortunately, available descriptions of the EST libraries do not allow getting more information on the specificity of gene expression. Nevertheless, it is obvious that genes in the profiles are co-regulated and are not overlapped because expression profiles were built by subtracting all profiles within the same genome from the profile in question (see Materials and methods). Consequently, all expression profiles of the two studied genomes are comparable. Although GO annotation provides more detailed insights on gene expression and function, this discussion is not within the scope of our study. It should also be emphasized that many *Drosophila* and *Arabidopsis* genes still do not have functional annotation.

Some well-annotated examples of the specific genes found as a result of *Arabidopis* expression profiling are the following: root profile-associated genes AHA2 (AT4G30190) and GLT1 (AT5G53460), expressed in roots; PAL2 (AT3G53260), expressed in root cortex (according to the GO annotation); ovule profile-associated genes HOS9/PFS2 (AT2G01500) and ERL1 (AT5G62230), expressed during embryo sac morphogenesis; SHP2 (AT2G42830), expressed during carpel development. Some examples of the specific genes found as a result of expression profiling in *Drosophila* are: r-cup (Dmel_CG10998), expressed during male meiosis; Klp59D (Dmel_CG12192), expressed during microtubule-based movement; gskt (Dmel_CG31003), expressed during male gonad development (all three are from the profile of *Drosophila* testes); cta (Dmel_CG17678), responsible for regulation of embryonic cell shape; jing (Dmel_CG9397), responsible for specification of segmental identity; cact (Dmel_CG5848), expressed during dorsal appendage formation (all are associated with the embryo profile).

Next, using a random number generator, we created a stochastic model which simulated a random distribution of TF genes in gene clusters in corresponding genomes. This model used such parameters as genome size, the number of TF in the genome, the number of clustered genes and distribution of clusters in the respective genome according to their size. According to the Wilcoxon signed-rank test, a comparison between observed and stochastic distributions (Fig. 1) showed that in the *Arabidopsis* genome there is a significant ($P<=0.008$) excess of TF genes in gene clusters, whereas in the *Drosophila* genome, distribution of TF genes in gene clusters did not differ significantly from stochastic ($P<=0.84$). For instance, in the *Arabidopsis* genome, TF/cluster ratio for clusters containing 5 genes was 0.36 for the observed distribution vs. 0.33 for the stochastic distribution, in 8-gene clusters this fraction was 0.91 and 0.54, and in 9-gene clusters — 2.00 and 0.57, respectively. In other words, the observed number of TF in 9-gene clusters exceeded stochastic by 3.5 times. Another non-parametric test, Kolmogorov–Smirnov test, gave the similar result with $P<<0.001$ ($P=6.303e-06$) for *Arabidopsis* and $P=1.0$ for *Drosophila.*

Potentially, tandem gene duplication could lead to overestimation of the TF/cluster ratio. In order to exclude influence of paralogous gene pairs on the earlier result, we calculated the number of paralogs in gene clusters of both *Drosophila* and *Arabidopsis* genomes. The result showed that, in the genome of *Drosophila,* gene clusters contained 4.7% paralogs (141 of 3010 clustered genes), whereas there were no clustered paralogous TF at all. In the *Arabidopsis* genome,

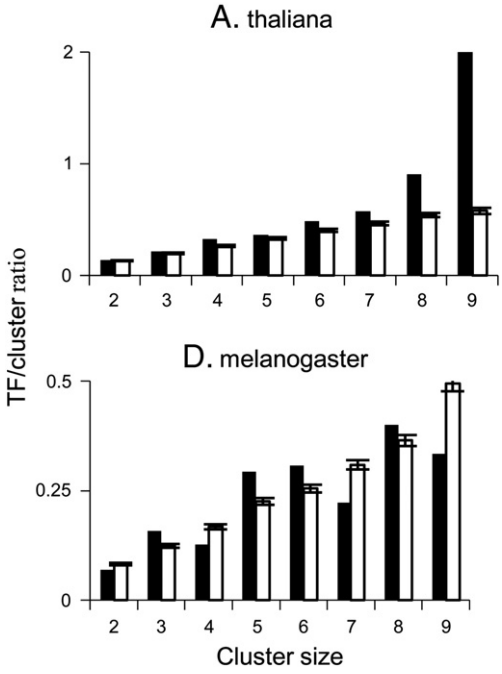

Fig. 1. Clusters of functionally-related genes in the *A. thaliana* genome are enriched by TF genes. On contrary, in the genome of *D. melanogaster* distribution of TF in gene clusters does not differ from stochastic. The height of the bars represents the fraction of TF genes per cluster calculated from all expression profiles of *A. thaliana* or *D. melanogaster*. Cluster size is the number of genes per cluster. Black bars designate data revealed by EST profiling and open-box bars designate the corresponding stochastic estimation. Raw data are presented in Supplementary Table S1.

gene clusters included 2.1% paralogs (190 of 9179 clustered genes) and only 28 clustered paralogous TF, which were located in duplets (22) and triplets (6). Moreover, 58.9% and 83.2% of non-TF paralogs were also found in duplets and triplets in *Drosophila* and *Arabidopsis*, respectively. Obviously, the effect of gene duplication events on TF/cluster ratio was minimal, if any.

For this study, out of six different gene expression profiles of *Arabidopsis*, we have chosen one associated with response to biotic stress because of an almost complete lack of any data on co-expressed non-homologous clustered genes responsive to biotic and abiotic factors. The profile contained 1590 clustered genes or 30% of all genes responsive to biotic stress. One hundred and two clusters belonging to this profile and containing 6 or more genes included at least one TF. In order to confirm non-random co-localization of TF and non-TF genes, we arbitrarily selected a single 8-gene cluster located on chromosome 5 containing 3 TF genes. The cluster did not include either TF or non-TF paralogs.

To investigate gene expression in the cluster, we took advantage of the experimental model, based on salicylic acid (SA) treatment of *Arabidopsis* plants. As SA represents a signal triggering the global defense response of the plant to pathogen attack (Durrant and Dong, 2004), the model imitates biotic stress. It might be expected that the cluster's members also could be activated in response to SA treatment because the selected cluster derived from the expression profile of putative responsive genes.

Indeed, 3 to 6 hours after treatment by 1 mM SA, we observed co-activation of the cluster's genes, including all 3 TF (Fig. 2). Variation in gene expression was considerable and exceeded corresponding levels in control, water-treated plants from 2.54 (*AT5G49540*) to 80.0 (*AT5G49480*) times. Activation of two genes (*AT5G49480* and *AT5G49520*) belonging to this cluster was also observed in the host-pathogen interaction model (Ascencio-Ibáñez et al., 2008). The authors, however, did not emphasize their possible clustering. Thus,

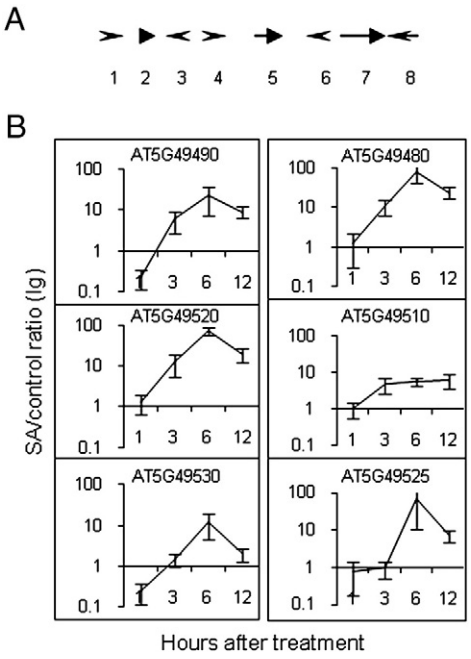

Fig. 2. Members of an eight gene cluster (A) in the *A. thaliana* genome are upregulated in response to salicylic acid (SA) treatment (B). Numbers 1 through 8 designate the *A. thaliana* genes *AT5G49480, AT5G49490, AT5G49500, AT5G49510, AT5G49520, AT5G49525, AT5G49530, AT5G49540*. Numbers 2, 5 and 7 designate the TF genes *AT5G49490*, *AT5G49520* and *AT5G49530*. Arrowheads denote direction of transcription. Ordinates in (B) represent upregulation level of six cluster members including three TF genes (left panel). In fact, all eight genes are upregulated (Supplementary Table S2). Bars represent confidence intervals (P = 0.05). Primer sequences are presented in Supplementary Table S3.

localization of TF among other co-activated genes of this cluster was not random.

Escape of TF genes beyond the borders of gene clusters in the *Drosophila* genome would be conceivable if the fraction of genes encoding TF in *Drosophila* was less than in *Arabidopsis*. In general, this might justify a reduction of TF genes in individual clusters. Alternatively, if the ratio TF/genome was similar in both species, location of TF outside gene clusters in *Drosophila* would suggest a presence of different pathways that regulate expression of clustered genes in two genomes.

In order to elucidate this, we have determined the percentage of TF genes in the genomes of *Drosophila* and *Arabidopsis* using computer analysis of the corresponding genomic files in the Genbank format. This resulted in 5.9% (823 TF out of 13831 protein coding genes) and 8.5% (2334 TF out of 27379 protein coding genes) in *Drosophila* and *Arabidopsis,* respectively (Table 1). However, considering that a broad implementation of alternative splicing in the *Drosophila* genome often leads to multiple proteins corresponding to the same gene, we counted alternative splicing in both genomes, which resulted in TF percentage of 11.4 for *Drosophila* (1582 TF out of 13831 protein coding genes) and 10.9 for *Arabidopsis* (2999 TF out of 27379 protein coding genes).

Thus, on the one hand, the TF/genome ratio was slightly less in the *Drosophila* genome comparative to the *Arabidopsis* one (1.4 times). On the other hand, taking into account alternative splicing, this ratio was very similar in both genomes. However, comparison of average values of the TF/cluster ratios in two species, calculated for clusters of all sizes, showed that in *Arabidopsis* it is 2.6 times greater than in *Drosophila* (0.63 and 0.24 respectively). Taken together, the results may suggest different regulatory mechanisms participating in expression of clustered genes in two model genomes. These differences could be both on the higher order chromatin level and on the level of regulation of individual clustered genes by interaction of TF and their corresponding cis-acting binding sites.

Table 1
Percentage of TF in the genomes of *D. melanogaster* and *A. thaliana* computed with or without alternative splicing.

	D. melanogaster	*A. thaliana*
No alternative splicing	5.9	8.5
Alternative splicing	11.4	10.9

In theory, the real genome structures such as the length of the genes and intergenic sequences can affect regulation of the clustered genes and the two studied genomes differ in this respect. However, in this study we assumed that this is more a question of functionality rather than of the established structure of the clusters. In our investigation we used a formal approach, which permits abstraction from the length of coding and non-coding sequences of clustered genes and gene spacers. The conclusions of this study were not drawn from the specific details of the genome structure and are based on the more generalized picture.

An important fact is that many TF in *Arabidopsis* and *Drosophila* belong to the different protein families. Besides, the same families can differ considerably in their size between two species. For example, lineage-specific TF constitute 45% in *Arabidopsis* and 14% in *Drosophila* due to disproportionally amplified C2H2 zinc finger TF genes (Riechmann et al., 2000). Another example of different expansion of the TF family is the MADS-box family which has four members in *Drosophila* and more than 100 members in *Arabidopsis* (Parenicova et al., 2003). Potentially, divergence of regulatory protein structures in two genomes can contribute to divergence of regulatory mechanisms in general and of regulation of clustered genes in particular. As mentioned earlier, the formal approach used in this study permits abstraction from functional details as well as it permits to define a TF gene simply as a regulatory unit (according to the GO functional annotation).

One can argue that there are homeobox (hox) gene clusters consisting of homeotic TF genes in the *Drosophila* genome (Lewis, 1978). However, the hox gene clusters in the fly genome came into existence as a result of several duplications and subsequent divergence; that is, they are not functionally-related, non-homologous genes. Perhaps this may be the exception rather than the rule. Even though clusters containing non-random TF genes could be found sufficiently frequently in the *Drosophila* genome, the frequency of such clusters, according to the results of this investigation, would be significantly lower than in *Arabidopsis*.

The development of new techniques allowed construction of genomic maps of chromatin protein binding (Ren et al., 2000; van Steensel et al., 2001) which in turn revealed contiguous regions of open and condensed interphase chromatin. Chromatin domains can be very large and include dozens of genes. Some of these domains display a nested architecture when larger domains are subdivided into smaller regions (de Wit et al., 2008). Perhaps transcriptional activation of a small chromatin region, corresponding to a gene cluster, by cooperative action of ATP-dependent chromatin remodeling complexes (Clapier and Cairns, 2009; Lorch et al., 2010) and histone modifying enzymes (Kouzarides, 2007) results in accessibility of promoters to the RNA II initiation complex. Consequent regulation of gene expression is, possibly, biased either to the activity of clustered TF genes in the *Arabidopsis* genome or to the activity of non-clustered TF genes in the genome of *Drosophila*.

The gene cluster in a chromatin domain can be considered as a subsystem with respect to the domain itself. In other words, such a cluster is part of a system which could be regulated either externally by outward TF genes or internally by TF genes belonging to the same cluster, or both. As we have shown earlier, in the *Arabidopsis* genome, gene clusters include non-random TF genes, which could be co-regulated together with gene-members of the same clusters.

It assumes that TF genes may indeed regulate gene expression of "their own" cluster and/or other clusters. There are at least two experimental works presenting data in favor of this point of view (Finnegan et al., 2004; Hirai et al., 2005).

If clustered TF genes do regulate the expression of members of the same cluster, escape of the regulator (TF gene) out of the regulated system (gene cluster) is an evolutionarily progressive event as the regulator becomes independent of the regulated system. Regulatory mechanisms of gene expression in *Drosophila* should be more elaborate than in *Arabidopsis*, since molecular and cellular organization of animals is much more complex than plants. Overall, the difference in distribution of TF genes in gene clusters of the two genomes possibly suggests evolutionarily progressive changes in the architecture of regulatory networks governing the activity of clustered genes in the animal kingdom.

Whether expression of clustered genes in *Arabidopsis* is regulated primarily by TF genes located in the same clusters or not still remains to be elucidated. Nevertheless, availability of large collections of mutant *Arabidopsis* plants offers unlimited opportunities to investigate this problem.

3. Materials and methods

3.1. EST mapping. Gene profiling and identification of clustered genes

EST mapping and building of gene expression profiles consisted of a number of steps (Boutanaev et al., 2002, 2009): i) BLAST homology search in the annotated genome using EST derived from different sources; ii) assignment of each EST to a corresponding gene based on coordinates of the homology region present in the BLAST output file; iii) building of an expression profile by assigning each gene a fraction of homologous EST; iv) a number of profiles originated from different EST sets were subtracted from the profile in question in order to obtain specific expression profile. Genes were considered to be specific if their profile values were positive. Two neighboring genes were considered to be located in the same cluster if their profile values were positive. Home-made software was used for all computational work.

3.2. Building a stochastic model

In the well-studied genomes of *Arabidopsis* and *Drosophila*, all genes can be numbered beginning with the first gene of the first chromosome and ending with the last gene of the last chromosome. Knowing the number of the specific genes derived from all expression profiles of the genome in question, it is possible to generate an equal set of random numbers in the range of 1 to the genome size (total number of genes) employing a random number generator. Two numbers are considered to represent clustered genes if they differ only by one. For instance, a row of 8, 9, 10, and 11 would represent a random 4-gene cluster. The generation of random numbers was repeated 100 times creating 100 independent sets of clusters and the latter were randomly selected from that pool to build a stochastic distribution corresponding to the observed counterpart. The next randomly selected cluster was eliminated if it overlapped with the previously selected ones. As the genomes of *Arabidopsis* and *Drosophila* are well-annotated, we used the functional Gene Ontology annotation tool to find all the TF included in the random clusters. Next, the TF/cluster ratio for each cluster size was found. The computing was cycled 200 times and average of the TF/cluster ratio for each cluster size was calculated.

3.3. Finding paralogs in gene clusters

BLAST stand alone engine was used for homology search ($e = 10-6$) among all protein sequences corresponding to every gene in the genomes of *Arabidopsis* or *Drosophila*. After computer treatment of the BLASTP output file, paralogs located in gene clusters were identified for both genomes.

3.4. Plant growth

Arabidopsis thaliana plants (ecotype Columbia-0, accession Col-0/Redei-L206440) were obtained from Lehle Seeds, Round Rock, TX, USA and grown at a density of 2–3 plants per pot to 21 days of age, in a Percival growth chamber (model E30BHOC8, Percival Scientific, Perry, IA, USA), set for a 16-h photoperiod and 24 °C. Five leaves on each of the 3 plants per time point were harvested at 1, 3, 6, and 12 h after spraying, snap frozen in liquid nitrogen, and stored at −80 °C until RNA extraction.

3.5. RNA extraction, first-strand synthesis and real-time RT-PCR

Total RNA was extracted using TRIzol reagent as described by the manufacturer (Invitrogen Corp., Carlsbad, CA, USA). Copy DNA was synthesized using SuperScript First-Strand cDNA Synthesis System according to the manufacturer's directions (Invitrogen). Real-time PCR was performed with iQ SYBR Green Supermix kit (Bio-Rad Laboratories, Inc., Hercules, CA, USA) on the MiniOpticon Real-Time PCR system (Bio-Rad) using the following parameters: 94 °C–1 min (1 cycle); 94 °C–30 s, 60 °C–30 s, 72 °C–30 s (30 cycles). Amplification was performed in several (no fewer than five) biological and two technical replicas. *Arabidopsis* actin gene *ACTIN1* (*AT2G37620*) was used as a reference in all real-time PCR experiments.

3.6. Data source

Genbank at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) was used as the data source. The *Arabidopsis* and *Drosophila* genomes were versions TAIR9 and 5.22, respectively. These versions presented 27379 and 13831 protein coding genes for *Arabidopsis* and *Drosophila* (http://www.arabidopsis.org and http://flybase.org), respectively. All information on EST libraries is freely available on the NCBI site: http://www.ncbi.nlm.nih.gov

Acknowledgments

We thank J. Hammond for critical reading of the manuscript and valuable comments and Wesley Schonborn for language editing. This work was supported by the U.S. Department of Agriculture, Agricultural Research Service and by the Institute of Basic Biological Problems, Russian Academy of Sciences.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.gene.2011.09.023.

References

Ascencio-Ibáñez, J.T., et al., 2008. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. 148 (1), 436–454.

Blanco, E., et al., 2008. Conserved chromosomal clustering of genes governed by chromatin regulators in Drosophila. Genome Biol. 9 (9), R134. doi:10.1186/gb-2008-9-9-r134.

Bortoluzzi, S., et al., 1998. A comprehensive, high-resolution genomic transcript map of human skeletal muscle. Genome Res. 8, 817–825.

Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., Nurminsky, D.I., 2002. Large clusters of co-expressed genes in the Drosophila genome. Nature 420, 666–669.

Boutanaev, A.M., Postnikova, O.A., Nemchinov, L.G., 2009. Mapping of heterologous expressed sequence tags as an alternative to microarrays for study of defense responses in plants. BMC Genomics 10, 273. doi:10.1186/1471-2164-10-273.

Clapier, C.R., Cairns, B.R., 2009. The biology of chromatin remodeling complexes. Annu. Rev. Biochem. 78, 273–304.

de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H.J., van Steensel, B., 2008. Global chromatin domain organization of the Drosophila genome. PLoS Genet. 4 (3), e1000045. doi:10.1371/journal.pgen.1000045.
Dillon, N., 2006. Gene regulation and large-scale chromatin organization in the nucleus. Chromosome Res. 14, 117–126.
Durrant, W.E., Dong, X., 2004. Systemic acquired resistance. Annu. Rev. Phytopathol. 42, 185–209.
Finnegan, E.J., Sheldon, C.C., Jardinaud, F., Peacock, W.J., Dennis, E.S., 2004. A cluster of Arabidopsis genes with a coordinate response to an environmental stimulus. Curr. Biol. 14, 911–916.
Hirai, M.Y., et al., 2005. Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. J. Biol. Chem. 280 (27), 25590–25595.
Kalmykova, A.I., Nurminsky, D.I., Ryzhov, D.V., Shevelyov, Y.Y., 2005. Regulated chromatin domain comprising cluster of co-expressed genes in Drosophila melanogaster. Nucleic Acids Res. 33, 1435–1444.
Ko, M.S.H., et al., 1998. Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the T-complex and under-representation on the X chromosome. Hum. Mol. Genet. 7, 1967–1978.
Kouzarides, T., 2007. Chromatin modifications and their function. Cell 128 (4), 693–705.
Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. Nature 276, 565–570.
Lorch, Y., Maier-Davis, B., Kornberg, R.D., 2010. Mechanism of chromatin remodeling. Proc. Natl. Acad. Sci. U. S. A. 107 (8), 3458–3462.
Ng, Y.K., Wu, W., Zhang, L., 2009. Positive correlation between gene coexpression and positional clustering in the zebrafish genome. BMC Genomics 10, 42 (Jan 22).
Parenicova, L., et al., 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15, 1538–1551.
Ren, B., et al., 2000. Genome-wide location and function of DNA binding proteins. Science 290, 2306–2309.
Riechmann, J.L., et al., 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290 (5499), 2105–2110 (Dec 15).
Schmid, M., et al., 2005. A gene expression map of Arabidopsis thaliana development. Nat. Genet. 37, 501–506.
Spellman, P.T., Rubin, G.M., 2002. Evidence for large domains of similarly expressed genes in the Drosophila genome. J. Biol. 1 (1), 5.
Sproul, D., Gilbert, N., Bickmore, W.A., 2005. The role of chromatin structure in regulating the expression of clustered genes. Nat. Rev. Genet. 6, 775–781.
van Steensel, B., Delrow, J., Henikoff, S., 2001. Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. 27, 304–308.
Williams, E.J., Bowles, D.J., 2004. Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res. 14, 1060–1067.
Williams, E.J., Hurst, L.D., 2002. Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. J. Mol. Evol. 54, 511–518.
Zhan, S., Horrocks, J., Lukens, L.N., 2006. Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosome domains. Plant J. 45, 347–357.