CMLS, Cell. Mol. Life Sci.

DOI 10.1007/s00018-005-5064-9

CMLS Cellular and Molecular Life Sciences

Visions & Reflections

Genomic sources of regulatory variation in *cis* and in *trans*

P. J. Wittkopp

Department of Molecular Biology and Genetics, 227 Biotechnology Building, Cornell University, Ithaca, New York 14853 (USA), Fax: +1 607 255 6249, e-mail: pw72@cornell.edu

Received 15 February 2005; received after revision 12 April 2005; accepted 26 April 2005

Abstract. Regulatory variation results from genetic changes with both *cis* and *trans* acting effects on gene expression. Here I describe the types of genetic variants that alter *cis* and *trans* regulation and discuss differences in the potential for *cis* and *trans* changes among different classes of genes. I argue that the molecular function of the protein encoded by each gene and how the gene is wired into the genomic regulatory network may influence its propensity for *cis* and *trans* regulatory changes.

Key words. Gene expression; regulation; *cis*; *trans*; regulatory network.

Variation in gene expression is extremely common. Up to one quarter of the genes in the human, yeast and fruit fly genomes are differentially expressed among individuals of each species, and most of these expression differences are heritable (e.g. [1–6]). A complex network of molecular interactions orchestrates gene expression, providing many sources of regulatory variation [7]. Gene expression, or transcription, is controlled by *cis*-regulatory DNA sequences that contain binding sites for *trans*-acting regulatory proteins known as transcription factors. These transcription factors interact with a variety of cofactor proteins to activate and sustain transcription. Availability of these regulatory proteins is controlled by cellular systems such as signal transduction cascades that monitor the cell and elicit the appropriate transcriptional response. Genetic mutations that alter any step in this genomic regulatory network can affect gene expression and contribute to regulatory variation (for review, see [7–10]).

cis- and *trans*-regulatory changes underlie variable gene expression

Regulatory variants can be classified into two types, *cis* and *trans* acting, depending on the relationship between the genomic location of the mutation and the gene(s) whose expression it affects (fig. 1). A *cis*-regulatory variant alters expression of the associated transcript in an allele-specific manner, whereas a *trans*-regulatory change affects a diffusible molecule that regulates expression of other genes. Most often, *cis*-regulatory variants are assumed to lie either in the basal promoter region located near the transcription start site or in an enhancer located in non-coding sequences surrounding the transcribed region. But *cis*-regulatory effects can also arise from epigenetic changes that alter chromatin structure and mutations in the transcribed sequence that affect the transcription rate and/or the half-life of the transcript [9, 10]. The genetic determinants of chromatin structure, transcription rate and messenger RNA (mRNA) stability are not well understood, and it is unclear how often these types of changes contribute to *cis*-regulatory differences. *trans*-regulatory mutations alter the concentration or activity of a protein or RNA that controls expression of other genes either directly (i.e. by binding to *cis*-regulatory DNA sequences) or indirectly (i.e. by influencing the binding and/or activity of direct regulators) [9, 10]. Changes in the coding sequence of the regulator are typically considered to be the source of *trans*-regulatory effects, but *cis*-regulatory changes that modify the expression level of a regulatory molecule can also have *trans*-acting effects on gene expression. The difference between

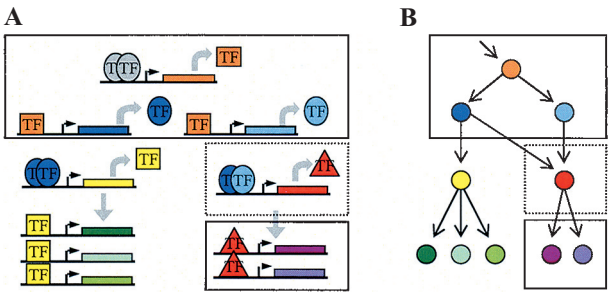

Figure 1. *cis*- and *trans*-regulatory changes result from genetic modifications in regulatory networks. (*A*) Schematic of a gene regulatory network. For each gene in this network, *cis*-regulatory DNA (line), protein coding sequence (colored box) and the transcriptional start site (arrowhead) are depicted. Colored shapes represent the proteins encoded by these genes. In this simplified diagram, all proteins are transcription factors (TF) that regulate expression of target genes by binding to their *cis*-regulatory regions. Consider the location of mutations that can affect expression of the gene in the dotted box. *trans*-regulatory mutations will be located in genes encoding any of the upstream regulators, which are included in the uppermost box. Note that these genes include both direct and indirect regulators of gene expression. Mutations within the dotted box can have *cis*-acting effects on the gene's own expression, as well as *trans*-regulatory effects on the expression of downstream genes (indicated by the lower box). (*B*) Directed graph representation of the network in (*A*). Each circle represents a gene and each arrow signifies a regulatory interaction between a protein and *cis*-regulatory DNA.

cis- and *trans*-regulatory changes is thus not the location of the mutation within the gene (i.e. coding versus non-coding), but rather whether or not expression of other genes is affected by the change.

Until recently, *cis*- and *trans*-regulatory changes had only been discerned for a handful of genes (e.g. [11–15]). New methods for measuring allele-specific gene expression now allow the rapid detection of *cis*-regulatory differences [16–19]. In addition, the combination of expression profiling with quantitative trait locus (QTL) mapping allows regions of the genome harboring variants that affect gene expression (i.e. eQTL) to be identified on a genomic scale [10, 20]. By comparing the location of eQTL to the location of the affected gene(s), each locus can be classified as *cis* or *trans* acting: eQTL located near the affected gene are assumed to indicate a *cis*-regulatory change, whereas eQTL that do not coincide with the affected gene(s) are inferred to have a *trans*-regulatory effect. Genomic studies of intraspecific regulatory variation in yeast, mice, maize and humans found that 25–35% of eQTL were consistent with *cis*-regulatory effects, with the remaining eQTL classified as *trans* acting [3, 21–23]. The striking similarity in the amount of *cis*- and *trans*-regulatory variation found in genomes from such diverse organisms suggests common mechanisms underlying regulatory variation.

Cis- and *trans*-regulatory changes result in modifications to the genomic regulatory network, and the architecture of this network is expected to shape the distribution of regulatory variants within the genome. Differences in the function of the protein encoded by each gene and how the gene is wired into the regulatory network may affect its propensity for *cis*- and *trans*-regulatory variation. Here I consider genomic sources of *cis*- and *trans*-regulatory variation and discuss how they may vary for different types of genes.

Enhancer size and structure may affect the frequency of *cis*-regulatory variants

Assuming that most *cis*-regulatory changes reside in promoter and enhancer regions, differences in the number of bases required for promoter and enhancer function among different types of genes may affect their tendency to acquire a *cis*-regulatory mutation. Basal promoter sequences contain binding sites for proteins associated with the basic polymerase machinery. Most promoters, both TATA box-containing and TATA-less, are approximately 100 bp long and subject to strong functional constraint [24], suggesting that they contribute little to differences among genes in the likelihood of *cis*-regulatory changes. Variation in the number, size and complexity of enhancers associated with different types of genes may thus be the primary cause of differences in the frequency of *cis*-regulatory mutations among genes.

Only a small fraction of the enhancers that exist in eukaryotic genomes have been characterized, yet general features of these complex *cis*-regulatory regions have already begun to emerge. Genes encoding pleiotropic regulatory proteins (e.g. transcription factors, signaling proteins) that are expressed in many different cell types and function at multiple times during the life cycle tend to have a greater number of enhancers than effector genes that are primarily expressed in one or few cell types [25]. Enhancers of regulatory genes also tend to contain a greater diversity of binding sites than enhancers regulating expression of effector genes [26]. In *Drosophila melanogaster* and *Caenorhabditis elegans*, genes with a high degree of regulatory complexity are flanked by larger regions of non-coding DNA than genes with more limited functions, possibly because of the increased number of enhancers [27]. These findings suggest that genes encoding regulatory proteins may have a higher probability of acquiring *cis*-regulatory changes than other classes of genes. Consistent with this hypothesis, eQTL identified by Brem et al. [3] associated with genes classified as transcriptional regulators in the Gene Ontology database [28] were more likely to be deemed *cis* acting than eQTL for genes in other functional classes: 41% of eQTL called *cis* for transcriptional regulators vs. 20% for genes with other functions; χ^2 test, df = 1, P = 0.01.

Regulatory network architecture determines the potential for *trans*-regulatory variants

The number of *trans*-regulatory mutations that can affect expression of a gene is related to the number of genes that control its expression. Regulatory networks describing interactions among all of the genes in a genome provide a roadmap of potential *trans*-regulatory changes for each gene [29] (fig. 1). Luscombe et al. examined the genomic regulatory network for yeast and found that genes involved in different types of biological processes are wired together differently [30]. Pathways controlling endogenous processes, such as the cell cycle and sporulation, typically involved extensive interactions among regulatory genes before activating expression of terminal effector genes (fig. 2A). In contrast, regulatory pathways controlling exogenous processes that are initiated by external cues (e.g. stress response) tended to contain fewer regulatory proteins that directly regulate expression of target genes (fig. 2B). In multicellular eukaryotes, regulatory pathways controlling development appear similar to the highly connected 'endogenous' pathways in yeast [31–33], whereas pathways controlling physiological responses to environmental stimuli appear to have the 'exogenous' pathway structure [34, 35]. Because 'endogenous' networks are more interconnected than 'exogenous' networks (fig. 2), both regulatory and terminal effector genes involved in endogenous processes may be more likely to be affected by *trans*-regulatory changes than genes controlling exogenous traits. Within each type of network, mutations in regulatory genes are expected to have a greater *trans*-regulatory effect than genetic changes in terminal genes. Regulatory genes encoding proteins with both direct and indirect effects on gene regulation have been shown to contribute to *trans*-regulatory variation [22].

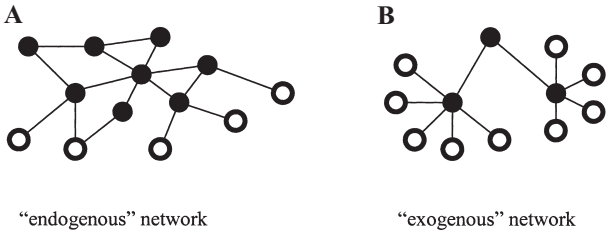

Figure 2. Regulatory networks controlling different types of biological processes have different architectures. (*A*) Example of an endogenous regulatory network consisting of many interconnected regulatory genes (filled circles) and few terminal effector genes (open circles). (*B*) Example of an exogenous regulatory network with few regulatory genes directly controlling expression of many target genes. Network structures and terminology are adapted from Luscombe et al. [30].

Molecular mechanisms reduce the proportion of mutations that alter gene expression

As described above, the mutational target size for *cis*-and *trans*-regulatory changes is expected to differ among classes of genes, but not every mutation will alter gene regulation. Within promoters and enhancers, degeneracy and flexible organization of transcription factor binding sites may reduce the number of sequence changes in *cis*-regulatory regions that impact transcription (e.g. [36, 37]). The proportion of potential *trans*-regulatory mutations that actually change *trans*-regulation may be reduced by the properties of regulatory networks that make them robust to genetic and environmental changes (e.g. feedback loops, threshold effects, redundancy) (e.g. [38–40]). Mutations in direct regulators could be more likely to cause transcriptional changes than mutations with indirect effects if each regulatory step between the genetic variant and the affected gene provides an opportunity for buffering. However, the mutational target size, which is expected to be larger for indirect than direct regulators, may offset any differences in buffering and allow many *trans*-acting eQTL with indirect effects to arise (as was observed for yeast [22]).

Genetic basis of regulatory variation: a promising future

The ideas presented here are drawn from burgeoning areas of research under active investigation. Ultimately, understanding how genetic variation is translated into variable gene expression will require the isolation of regulatory variants and knowledge about how these variants alter the genomic regulatory network. Determining the distribution of *cis*- and *trans*-regulatory changes within the genome is an important step toward this goal (e.g. [3, 19, 21–23]), but vital information is still missing. An international consortium of scientists is currently working to characterize all of the *cis*-regulatory elements in 30 Mb of human DNA [41], and a similar project may soon be initiated for the entire *D. melanogaster* genome. The resulting encyclopedia of *cis*-regulatory elements will facilitate the identification of genetic variants with *cis*-acting effects [42]. *trans*-regulatory variants will become easier to identify with the continued elucidation of regulatory networks. Genomic maps of protein-protein interactions [43–45] and transcription factor binding sites [46, 47] provide the framework for these networks, and expression profiling of mutant strains will reveal the dynamic changes that result from genetic perturbations [48, 49]. Understanding how regulatory variation is structured within genomic regulatory networks is a necessary step toward connecting genotypic and phenotypic diversity.

Acknowledgements. I thank K. Montooth and B. Payseur for helpful discussions and comments on the manuscript, and the Damon Runyon Cancer Research Foundation for support as a Damon Runyon Fellow (DRG #1795-03).

1 Cavalieri D., Townsend J. P. and Hartl D. L. (2000) Manifold anomalies in gene expression in a vineyard isolate of saccharomyces cerevisiae revealed by DNA microarray analysis. Proc. Natl. Acad. Sci. USA **97:** 12369–12374
2 Jin W., Riley R. M., Wolfinger R. D., White K. P., Passador-Gurgel G. and Gibson G. (2001) The contributions of sex, genotype and age to transcriptional variance in drosophila melanogaster. Nat. Genet. **29:** 389–395
3 Brem R. B., Yvert G., Clinton R. and Kruglyak L. (2002) Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752–755
4 Enard W., Khaitovich P., Klose J., Zollner S., Heissig F., Giavalisco P. et al. (2002) Intra- and interspecific variation in primate gene expression patterns. Science **296:** 340–343
5 Oleksiak M. F., Churchill G. A. and Crawford D. L. (2002) Variation in gene expression within and among natural populations. Nat. Genet. **32:** 261–266
6 Fay J. C., McCullough H. L., Sniegowski P. D. and Eisen M. B. (2004) Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae.* Genome Biol. **5:** R26
7 Davidson E. H. (2001) Genomic Regulatory Systems: Development and Evolution, Academic Press, San Diego
8 Carroll S. B., Grenier J. K. and Weatherbee S. D. (2001) From DNA to Diversity**:** Molecular Genetics and the Evolution of Animal Design, Blackwell Science, Oxford
9 Wray G. A., Hahn M. W., Abouheif E., Balhoff J. P., Pizer M., Rockman M. V. et al. (2003) The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. **20:** 1377–1419
10 Stamatoyannopoulos J. A. (2004) The genomics of gene expression. Genomics **84:** 449–457
11 Cavener D. R. (1992) Transgenic animal studies on the evolution of genetic regulatory circuitries. Bioessays **14:** 237–244
12 Stern D. L. (1998) A role of ultrabithorax in morphological differences between *Drosophila* species. Nature **396:** 463–466
13 Wang X. and Chamberlin H. M. (2002) Multiple regulatory changes contribute to the evolution of the *Caenorhabditis* lin-48 ovo gene. Genes Dev **16:** 2345–2349
14 Wittkopp P., Vaccaro K. and Carroll S. (2002) Evolution of yellow gene regulation and pigmentation in *Drosophila*. Curr. Biol. **12:** 1547–1556
15 Montooth K. L., Marden J. H. and Clark A. G. (2003) Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. Genetics **165:** 623–635
16 Cowles C. R., Hirschhorn J. N., Altshuler, D. and Lander E. S. (2002) Detection of regulatory variation in mouse genes. Nat. Genet. **32:** 432–437
17 Yan H., Yuan W., Velculescu V. E., Vogelstein B. and Kinzler K. W. (2002) Allelic variation in human gene expression. Science **297:** 1143
18 Pastinen T., Sladek R., Gurd S., Sammak A., Ge B., Lepage P. et al. (2004) A survey of genetic and epigenetic variation affecting human gene expression. Physiol. Genomics **16:** 194–193
19 Wittkopp P. J., Haerum B. K. and Clark A. G. (2004) Evolutionary changes in cis and trans gene regulation. Nature **430:** 85–88
20 Jansen R. C. and Nap J. P. (2001) Genetical genomics: the added value from segregation. Trends Genet. **17:** 388–391
21 Schadt E. E., Monks S. A., Drake, T. A., Lusis A. J., Che N., Colinayo V. et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature **422:** 297–302
22 Yvert G., Brem R. B., Whittle J., Akey J. M., Foss E., Smith E. N. et al. (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat. Genet. **35:** 57–64
23 Morley M., Molony C. M., Weber T. M., Devlin J. L., Ewens K. G., Spielman R. S. et al. (2004) Genetic analysis of genome-wide variation in human gene expression. Nature **430:** 743–747
24 Lee T. I. and Young R. A. (2000) Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. **34:** 77–137
25 Arnone M. I. and Davidson E. H. (1997) The hardwiring of development: organization and function of genomic regulatory systems. Development **124:** 1851–1864
26 Howard M. L. and Davidson E. H. (2004) Cis-regulatory control circuits in development. Dev. Biol. **271:** 109–118
27 Nelson C. E., Hersh B. M. and Carroll S. B. (2004) The regulatory content of intergenic DNA shapes genome architecture. Genome Biol. **5:** R25
28 Harris M. A., Clark J., Ireland A., Lomax J., Ashburner M., Foulger R. et al. (2004) The gene ontology (go) database and informatics resource. Nucleic Acids Res. **32:** D258–261
29 Rung J., Schlitt T., Brazma A., Freivalds K. and Vilo J. (2002) Building and analysing genome-wide gene disruption networks. Bioinformatics **18:** S202–210
30 Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. and Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature **431:** 308–312
31 Oliveri P. and Davidson E. H. (2004) Gene regulatory network controlling embryonic specification in the sea urchin. Curr. Opin. Genet. Dev. **14:** 351–360
32 Schroeder M. D., Pearce M., Fak J., Fan H., Unnerstall U., Emberly E. et al. (2004) Transcriptional control in the segmentation gene network of *Drosophila*. PLoS Biol. **2:** E271
33 Stathopoulos A. and Levine M. (2004) Whole-genome analysis of *Drosophila* gastrulation. Curr. Opin. Genet. Dev. **14:** 477–484
34 Hahn J. S., Hu Z., Thiele D. J. and Iyer V. R. (2004) Genome-wide analysis of the biology of stress responses through heat shock transcription factor. Mol. Cell. Biol. **24:** 5249–5256
35 Kwon J. Y., Hong M., Choi M. S., Kang S., Duke K., Kim S. et al. (2004) Ethanol-response genes and their regulation analyzed by a microarray and comparative genomic approach in the nematode *Caenorhabditis elegans.* Genomics **83:** 600–614
36 Ludwig M. Z., Bergman C., Patel N. H. and Kreitman M. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:** 564–567
37 Ludwig M. Z., Palsson A., Alekseeva E., Bergman C. M., Nathan J. and Kreitman M. (2005) Functional evolution of a cis-regulatory module. PLoS Biol. **3:** e93
38 Gibson G. (1996) Epistasis and pleiotropy as natural properties of transcriptional regulation. Theor. Popul. Biol. **49:** 58–89
39 Siegal M. L. and Bergman A. (2002) Waddington's canalization revisited: developmental stability and evolution. Proc. Natl. Acad. Sci. USA **99:** 10528–10532
40 Ingolia N. T. (2004) Topology and robustness in the *Drosophila* segment polarity network. PLoS Biol. **2:** e123
41 The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia of DNA Elements) project. Science **306:** 636–640
42 Dermitzakis E. T., Reymond A. and Antonarakis S. E. (2005) Conserved non-genic sequences – an unexpected feature of mammalian genomes. Nat. Rev. Genet. **6:** 151–157
43 Uetz P., Giot L., Cagney G., Mansfield T. A., Judson R. S., Knight J. R. et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae.* Nature **403:** 623–627
44 Giot L., Bader J. S., Brouwer C., Chaudhuri A., Kuang B., Li Y. et al. (2003) A protein interaction map of *Drosophila melanogaster.* Science **302:** 1727–1736

45 Li S., Armstrong C. M., Bertin N., Ge H., Milstein S., Boxem M. et al. (2004) A map of the interactome network of the metazoan *C. elegans*. Science **303:** 540–543
46 Lee T. I., Rinaldi N. J., Robert F., Odom D. T., Bar-Joseph Z., Gerber G. K. et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae.* Science **298:** 799–804
47 Sun L. V., Chen L., Greil F., Negre N., Li T. R., Cavalli G. et al. (2003) Protein-DNA interaction mapping using genomic tiling path microarrays in *Drosophila*. Proc. Natl. Acad. Sci. USA **100:** 9428–9433
48 Hughes T. R., Marton M. J., Jones A. R., Roberts C. J., Stoughton R., Armour C. D. et al. (2000) Functional discovery via a compendium of expression profiles. Cell **102:** 109–126
49 Jansen, R. C. (2003) Studying complex biological systems using multifactorial perturbation. Nat. Rev. Genet. **4:** 145–151