Variable gene expression in eukaryotes: a network perspective

Patricia J. Wittkopp

Department of Ecology and Evolutionary Biology, Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

e-mail: wittkopp@umich.edu

Summary

Changes in gene expression underlie phenotypic plasticity, variation within species, and phenotypic divergence between species. These expression differences arise from modulation of regulatory networks. To understand the source of expression differences, networks of interactions among genes and gene products that orchestrate gene expression must be considered. Here I review the basic structure of eukaryotic regulatory networks and discuss selected case studies that provide insight into how these networks are altered to create expression differences within and between species.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1567/DC1

Key words: gene regulation, transcription, *cis*-regulation, *trans*-regulation.

Introduction

Comparative physiologists strive to understand how and why cellular and organismal functions differ among environments, among individuals, and among species. These issues can be addressed by comparing complex organismal phenotypes, comparing enzyme activity and flux through biochemical pathways, or comparing the molecular mechanisms that produce these higher-order phenotypes. Recently, many comparative studies have focused on differences in gene expression as a source of organismal diversity. Although the importance of gene regulation for controlling biological processes has been recognized for over 30 years (Britten and Davidson, 1969; King and Wilson, 1975), it is only in the last decade that the tools necessary to study changes in gene expression on a large-scale have become available (Schena et al., 1995). These tools, particularly DNA microarrays, have opened new areas of exploration for comparative physiologists studying both model and non-model systems (reviewed in Cossins et al., 2006).

With the rapid accumulation of studies analyzing transcript levels, it is important to remember that transcription is only one step (albeit a critical one) in converting genotypes into phenotypes; changes in transcript levels do not always affect phenotypes and *vice versa*. The limitations of transcript analysis have been discussed in detail (Feder and Walser, 2005). Despite these limitations, comparative studies of gene expression have provided insight into the molecular mechanisms of ecological responses and phenotypic evolution.

These comparative studies can be divided into three groups: comparisons within genotypes, within species and between species.

(1) Differences in gene expression can be created by environmental cues without any genetic differences. Comparative studies of individuals exposed to different environments reveal changes in gene expression associated with physiological responses to external stimuli. The utility of this approach is illustrated by a recent study of the eurythermic goby fish (*Gillichthys mirabilis*) reared under multiple temperature regimes mimicking wild conditions (Buckley et al., 2006). Approximately 2% of the genes surveyed showed a change in gene expression between treatments, with the specific genes showing altered expression varying among tissues. Individual expression changes observed in this study (and in similar studies) may be associated with positive physiological adjustments that help an organism cope with its surroundings, may reflect a stress response to adverse conditions, or may be consequences of pleiotropy that have no direct role in adjusting to different temperatures. Additional experiments are required to determine which expression changes fall into which class.

(2) Under standardized environmental conditions, genetic polymorphisms can cause expression differences within a species. Studies comparing genetically distinct samples of yeast, fruit flies and fish reared under similar laboratory conditions found that up to 25% of genes vary in their expression level between individuals of the same species (Brem

et al., 2002; Jin et al., 2001; Oleksiak et al., 2002). These expression differences include both neutral polymorphisms and variation that contributes to phenotypic variation. For example, polymorphic gene expression explains variation in cardiac metabolism of the teleost fish *Fundulus heteroclitus*, demonstrating the adaptive potential of regulatory variation (Oleksiak et al., 2005). Putative adaptive changes in gene expression have also been observed in experimental populations of microorganisms (e.g. Cooper et al., 2003; Ferea et al., 1999; Riehle et al., 2003). Although the number of case studies demonstrating phenotypic consequences for regulatory changes is growing, the proportion of regulatory variation that is advantageous, neutral and deleterious remains subject to debate (e.g. Fay and Wittkopp, in press; Gilad et al., 2006; Lemos et al., 2005; Ranz and Machado, 2006).

(3) Genetic divergence between species also creates differences in gene expression. Regulatory evolution has been shown to contribute to divergent traits such as body armor in stickleback fishes (Colosimo et al., 2005), ear shape in maize (Hubbard et al., 2002) and pigmentation in fruit flies (Wittkopp et al., 2003). Up to 25% of genes differentially expressed between closely related *Drosophila* species show patterns of expression variation consistent with lineage-specific selection (Rifkin et al., 2003). Meta-analysis of comparative genomic expression data concludes that regulatory evolution is characterized by strong stabilizing selection with directional selection on some genes (Lemos et al., 2005).

All changes in gene expression result from modifications to regulatory networks. To understand the genetic and molecular mechanisms responsible for expression differences, we must examine the structure of regulatory networks and investigate how changes in these networks alter gene expression. Elucidating the architecture of regulatory networks will reveal connections among genes and is expected to uncover properties of regulatory networks that make certain types of changes more or less likely to occur (Wittkopp, 2005). Here, I review basic structural features of eukaryotic regulatory networks and use selected case studies to examine how networks vary between environments, between genotypes, and between species.

The structure of regulatory networks

In isolation, a gene is an inert object – nothing more than a string of nucleotide bases. Only when placed in a cellular environment and interacting with the products of other genes can a gene become active. The first step in this activation is for the gene to be transcribed, or 'expressed'. Comparative studies of gene expression on a genomic scale typically use standing transcript levels as a measure of gene expression. Although post-transcriptional regulation is expected to be equally important, techniques needed to study this process in a high-throughput manner remain limited. Biochemical interactions among proteins, RNA and DNA that control transcript levels are summarized below. These interactions form a complex network that coordinates gene expression and integrates the genome (Babu et al., 2004).

Regulatory interactions

Molecular interactions between genes and gene products form the connections that make up a regulatory network (Fig. 1A). Sequence-specific interactions between transcription factor proteins and *cis*-regulatory DNA sequences provide the basic network structure (Blais and Dynlacht, 2005). Binding sites for 106 transcription factors have been mapped genome-wide in the baker's yeast *Saccharomyces cerevisiae* (Lee et al., 2002). The number of target genes for a given transcription factor ranged from 0 to 181 in this experiment, with an average of 38 putative *cis*-regulatory targets per transcription factor. Two-thirds of the transcription factors surveyed had less than 40 targets each. Although full genomic surveys have not yet been completed in metazoans, smaller scale studies have been conducted in *Drosophila melanogaster* (Moorman et al., 2006) and *Caenorhabditis elegans* (Deplancke et al., 2006). These studies suggest that the number of regulatory factors per gene may be higher in multicellular eukaryotes than in yeast.

Protein–protein interactions affecting co-regulators, signaling pathways, and other control systems also influence transcriptional regulation and must be incorporated into transcriptional regulatory networks. Genomic surveys of protein–protein interactions in yeast, flies and worms (Giot et al., 2003; Li et al., 2004; Uetz et al., 2000) reveal a scale-free network structure (i.e. few genes with many connections and many genes with few connections) similar to that seen for transcription factor targets (Albert, 2005). In animals, an additional layer of regulatory control is provided by small, noncoding, microRNA molecules that affect transcript levels by altering chromatin state or modulating mRNA stability (Ke et al., 2003; Stark et al., 2005).

Local motifs

Despite the many possible arrangements of regulatory factors and their target genes, five common motifs (Fig. 1B) have emerged from analyses of transcriptional regulatory networks in yeast (Lee et al., 2002): (1) Feed-forward loop, which involves three genes. Gene A regulates gene B, and together they regulate gene C. This motif is over-represented in transcriptional networks (Milo et al., 2002), and has properties well-suited to transcriptional regulation (Mangan and Alon, 2003). (2) Single input module, which features a single transcription factor that activates expression of a group of target genes. This type of motif is often associated with genes that respond to exogenous signals (Luscombe et al., 2004). (3) Multiple input module, which describes cases where the same group of regulatory factors controls expression of a battery of target genes. Genes that regulate embryonic development in *D. melanogaster* display this type of regulation (Erives and Levine, 2004). (4) Autoregulatory and feedback (multi-component) loops; these describe cases in which a gene product regulates expression of the gene encoding it, either directly (autoregulation) or through interaction with other genes (feedback loops). These motifs provide stability to patterns of gene expression (Becskei and Serrano, 2000). (5) Regulatory chain: cascades of regulatory interactions in which

gene A regulates gene B, which regulates gene C, which regulates gene D, and so on. This motif contributes to the hierarchical structure of regulatory networks. All five of these regulatory motifs are also found in the regulatory networks of metazoans (e.g. Davidson et al., 2003; Levine and Davidson, 2005; Stathopoulos and Levine, 2005).

Modules: trait-specific pathways

Groups of genes regulating the same phenotype can be placed together into a pathway (Fig. 1C). Historically, pathways have been defined genetically by ordering the action of mutations that disrupt the same phenotype. When biochemical interactions responsible for these genetic effects are identified, the pathways can be integrated with transcriptional regulatory networks.

In multicellular animals, the two best-understood regulatory networks – at both the genetic and biochemical levels – control mesoderm development in sea urchins and embryonic patterning in *D. melanogaster* embryos (Levine and Davidson, 2005). These pathways share features thought to be representative of developmental regulatory systems in general (Davidson et al., 2003; Stathopoulos and Levine, 2005). For example, developmental pathways have a hierarchical structure with genes controlling the earliest regulatory events at the top and genes controlling the final differentiation at the bottom. Different functional classes of proteins act at different levels in these hierarchies, with genes encoding transcription factors and signaling molecules near the top and genes encoding enzymes and structural proteins at the bottom. Often, regulatory interactions that initiate a developmental program are followed
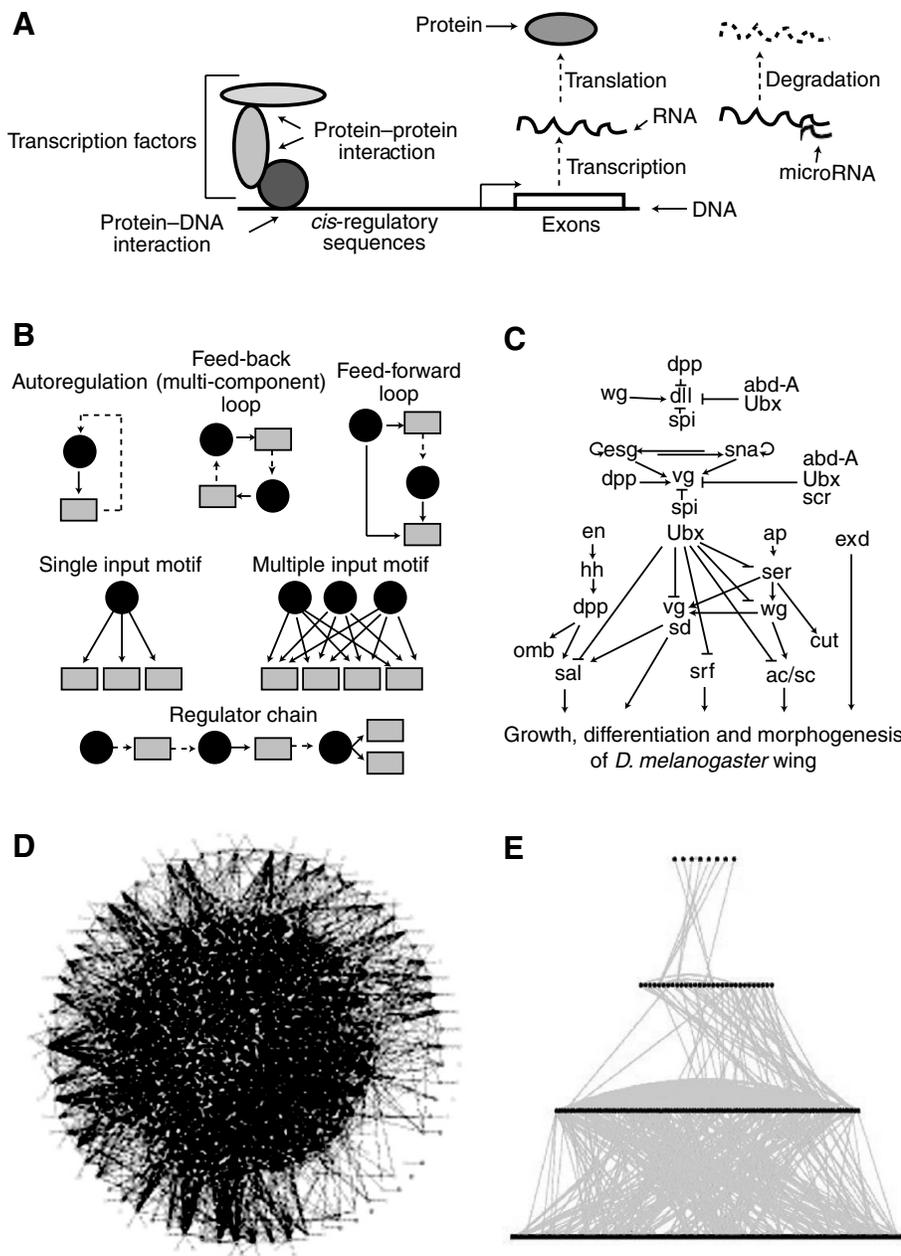

Fig. 1. Structure of eukaryotic regulatory networks. (A) The basic building blocks of regulatory networks are interactions between transcription factor proteins and *cis*-regulatory DNA sequences, protein–protein interactions affecting availability or activity of transcription factors, and hybridization of microRNAs to their targets. (B) Interactions among these factors form local regulatory motifs such as feed-forward loops, single input modules, multiple input modules, autoregulation, feed-back loops and regulator chains. In the diagrams, circles represent proteins, squares represent genes, and solid arrows represent regulatory interactions. Broken lines indicate protein production from a gene. (C) Regulatory pathways controlling biological processes (especially development) have hierarchical structure. Pathways controlling wing development in *D. melanogaster* at mid-embryonic (top), late embryonic (middle) and late larval (bottom) stages are reproduced from Abouheif and Wray (Abouheif and Wray, 2002). Reprinted with permission from AAAS. Layers of regulatory proteins interact before the ultimate activation of effector genes that give differentiated cells their characteristic properties. (D) Genes functioning in multiple pathways form an integrated genomic network. The transcriptional regulatory network for *S. cerevisiae* is shown here, modified from Babu et al. (Babu et al., 2004), with permission from Elsevier. (E) The genomic regulatory network for *S. cerevisiae* has a hierarchical structure similar to individual pathways. Image is modified from Yu and Gerstein (Yu and Gerstein, 2006, *Proc. Natl. Acad. Sci. USA* **103**, 14724-14731), with permission. Copyright (2006) National Academy of Sciences, USA.

by multi-gene feedback loops that maintain differentiated states. Both positive and negative regulators operate in these pathways and contribute to the robustness of developmental pathways. The wing development pathway of *D. melanogaster*, depicted in Fig. 1C, illustrates all of these features.

Genomic networks

Regulatory factors that function in multiple pathways link trait-specific pathways together to form a complex genomic regulatory network (Fig. 1D). These common regulators can create pleiotropy within the network. However, independent control of gene expression in different pathways can minimize pleiotropic effects and generate modularity. The modularity of regulatory networks is a critical property that facilitates evolutionary change (Carroll et al., 2001).

Like developmental pathways, genomic networks also have properties that appear to be shared among eukaryotes. For example, all known regulatory networks share a scale-free distribution with a small number of highly connected genes (i.e. 'hubs') and many genes with few connections (Albert, 2005). Genomic networks also have a hierarchical structure similar to individual developmental pathways (Yu and Gerstein, 2006) (Fig. 1E). Highly connected nodes occur at the top and middle of the hierarchy with minimally connected 'terminal nodes', which do not directly impact regulation of other genes, at the bottom. The similarity of genetic architecture among species may result from shared ancestry, selection for an optimal design, or (most likely) both. Simulation studies have shown that the structure of regulatory networks confers a robustness and stability in the face of genetic and environmental perturbations known as 'canalization' (Hornstein and Shomron, 2006; Siegal and Bergman, 2002).

Regulatory variation in a network context

Comparisons of gene expression alone, whether measured by microarrays or *in situ* hybridization, reveal little about the underlying molecular mechanisms that cause divergent gene expression. Integrating this comparative work with knowledge of regulatory networks based on genetic analysis and biochemistry will provide a more complete understanding of developmental, physiological and evolutionary processes. Here I use selected case studies to illustrate how comparative studies of gene expression can be used to examine the structure of regulatory variation within regulatory networks.

Environmental effects on gene regulation

A recent study investigating the effects of alcohol exposure on gene expression in the fruit fly *D. melanogaster* illustrates the power of using microarrays to understand physiological changes (Morozova et al., 2006). 3% of transcripts were found to have significant expression differences between flies that were and were not exposed to ethanol. Genes involved in olfaction, signal transduction, metabolism, transcription regulation, circadian rhythm and pigmentation changed expression more often than expected by chance. Some of these categories, such as olfaction

and metabolism, fit prior expectations for the types of physiological changes induced by ethanol. Other classes, such as pigmentation genes, may represent secondary pleiotropic consequences of the network structure or genes whose functions are incompletely characterized.

To determine the fraction of genes with altered expression that functionally mediate the response to ethanol, mutant strains for 20 of the affected genes were tested for ethanol tolerance (Morozova et al., 2006). 15 of the mutants had an ethanol sensitivity that differed from that of a control, non-mutant strain. These data suggest that the majority of genes whose expression is affected by ethanol exposure contribute to ethanol tolerance. Up to 25% of the genes with altered expression, however, appear to be side effects resulting from pleiotropic connections in the underlying regulatory network. Pathways controlling expression of genes induced by external stimuli (e.g. ethanol) have been shown to contain a small number of transcription factors directly regulating expression of a collection of functionally related genes (Luscombe et al., 2004). This structure may contribute to the high specificity of expression changes induced in response to ethanol.

Compared to networks regulated by exogenous cues, networks controlling development and basic cellular processes tend to be controlled by more regulators, with extensive interactions and feedback loops among the regulators (Luscombe et al., 2004). This structure provides a variety of mechanisms for altering the output of regulatory systems. The developmental basis of polyphenism in ants illustrates this point (Abouheif and Wray, 2002). Exposure to juvenile hormone causes genetically identical individuals to develop into any one of three castes (i.e. reproductive, worker, soldier). Of these, only the reproductive caste has wings. Using the regulatory networks controlling wing development in *D. melanogaster* as a guide, patterns of gene expression for wing developmental genes were compared between winged and wingless castes. The point at which wing development is disrupted was found to differ between the two wingless castes as well as between the two developing wing discs within one caste. In one case, expression was disrupted only in a gene located at the bottom of the network, whereas in the other case, the developmental pathway was blocked at a much higher step in the pathway (Fig. 2). Expression changes affecting wing development do not interfere with other functions of these pleiotropic proteins because of the modularity in regulatory networks.

Variable expression within species

One way to understand how regulatory networks vary is to identify the genetic changes affecting gene expression, determine which genes they affect, and examine their placement in the network. Quantitative trait locus mapping can be used to identify regions of the genome responsible for expression differences (eQTL) between strains or individuals of the same species (Brem et al., 2002). To determine the molecular mechanisms by which genetic variants affect gene expression, the location an eQTL is compared to the location of the affected gene. If the eQTL is located near the affected
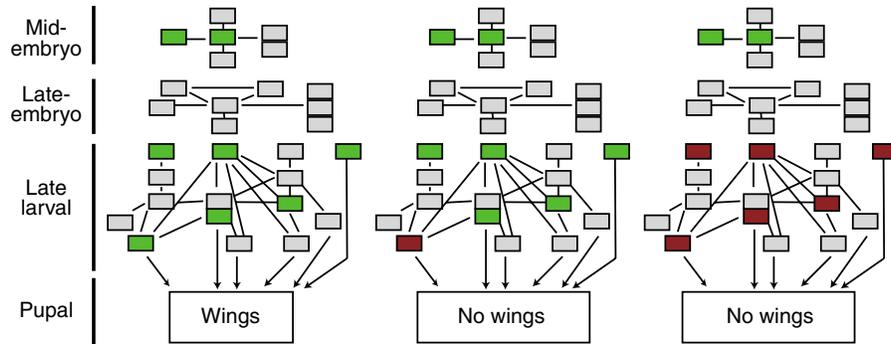
Fig. 2. The wing development pathway is disrupted at multiple points to create wingless ants. Activity (as measured by gene expression) of genes involved in wing development is shown for ant imaginal wing discs that will and will not produce wings in the adult animals. In these networks, boxes represent genes shown in Fig. 1C and lines indicate regulatory interactions among genes. Green indicates similar expression to *D. melanogaster*, red indicates divergent or absent expression, and grey indicates genes that were not surveyed. This figure is adapted from Abouheif and Wray (Abouheif and Wray, 2002).

gene, the change is inferred to act by modifying *cis*-regulation. If the eQTL is located elsewhere, the genetic change is inferred to act by modifying *trans*-regulation. Although discriminating between *cis*- and *trans*-acting changes does not identify the precise change within the regulatory network, it does separate changes at the gene itself from those residing in other factors.

The majority of eQTL identified in studies ranging from yeast to humans had *trans*-acting effects on gene expression (e.g. Brem et al., 2002; Cheung et al., 2005; Monks et al., 2004; Morley et al., 2004; Schadt et al., 2003). Yvert et al. characterized these *trans*-acting eQTL in yeast and found that eQTL for different genes often map to the same genomic region (Yvert et al., 2003). Assuming the causative sites responsible for the coincident eQTL are all located within the same gene, regulatory variation appears to be concentrated at highly connected hubs in the network. Indeed, only 100–200 distinct genes are estimated to account for all of the 1716 *trans*-acting eQTL identified in this study. The authors hypothesized that transcription factors with many target genes were responsible for these widespread effects; however, this class of proteins was found not to be over-represented near clustered eQTL. This is perhaps not surprising because a variety of biochemical classes can serve as hubs in regulatory networks. For example, a mutation in a receptor-associated G protein was shown to be responsible for one of the eQTL 'hotspots'.

The high frequency of *trans*-acting eQTL with widespread effects could be due to a biased mutational process, or to a higher fitness for changes in hubs relative to other types of genes. Studies of mutation accumulation (MA) lines in the nematode *Caenorhabditis elegans* suggest that the mutational process itself may produce many variants with pleiotropic, *trans*-acting effects on gene expression (Denver et al., 2005). MA lines were created using single hermaphrodites to propagate independent lines derived from the same starting genotype for many generations. This procedure maintains all but the most severe mutations. After 280 generations, expression was compared among four MA lines. 9% of the genes were found to have evolved in expression differences in at least one line.

By contrast, only 2% of genes were found to have expression differences among distantly related *C. elegans* strains isolated from the wild, demonstrating that natural selection eliminates many regulatory mutations. In the MA lines, but not in natural isolates, co-expressed genes were over-represented among genes with altered expression. Expression differences for groups of co-expressed genes were also observed in a similar mutation accumulation study of *D. melanogaster* (Rifkin et al., 2005). These findings suggest that only a few regulatory mutations with effects on multiple downstream genes are responsible for the extensive expression divergence observed. Some groups of coregulated genes changed expression in multiple mutation accumulation lines, suggesting that the portions of the network controlling expression of these genes are particularly susceptible to regulatory mutations. Because selection is minimal in mutation accumulation lines, the structure of the regulatory network is expected to control the distribution of regulatory mutations within the genome.

Expression divergence between species

Regulatory networks evolve by changing which genes interact as well as by changing how these genes interact (Babu et al., 2004). For example, *cis*-regulatory sequences may switch binding sites from one transcription factor to another, mutations in protein–protein interaction domains or within the DNA binding regions may abolish a connection in the network, or evolution of microRNA sequences may generate new target genes. Gene duplications and deletions can also add and subtract entire regulatory modules. Comparative studies of regulatory networks controlling development have identified divergent steps as well as sets of highly conserved regulatory interactions (reviewed in Davidson and Erwin, 2006). Differences in network structure between species may cause expression divergence, but may also reflect silent changes characteristic of developmental system drift (True and Haag, 2001). The presence of similar regulatory motifs and scale-free properties in regulatory networks from diverse eukaryotes suggests that regulatory connections are rearranged in a manner

Table 1. *Functional tests for cis- and trans-acting regulatory changes underlying interspecific expression differences*

Species A	Species B	Gene	Function[1]	cis[2]	trans[3]	Reference
Drosophila species (fruit flies)						
D. melanogaster	*D. simulans*	*Ultrabithorax*	R	Y	Y	(Stern, 1998)
D. melanogaster	*D. sechellia*	*shaven-baby*	R	Y	Y	(Sucena and Stern, 2000)
D. melanogaster	*D. subobscura, D. virilis*	*yellow*	E	Y	Y	(Wittkopp et al., 2002)
D. melanogaster	*D. biarmipes*	*yellow*	E	Y	Y	(Gompel et al., 2005)
D. melanogaster	*D. elegans, D. gunungcola, D. tristis*	*yellow*	E	Y	Y	(Prud'homme et al., 2006)
D. melanogaster	*D. simulans*	*Alcohol dehydrogenase*	E	Y	Y	(Dickinson et al., 1984)
	D. mulleri	*Alcohol dehydrogenase*	E	Y		(Fischer and Maniatis, 1986)
	D. affinidisjuncta	*Alcohol dehydrogenase*	E	Y	Y	(Brennan and Dickinson, 1988)
	D. hawaiiensis, D. affinidisjuncta	*Alcohol dehydrogenase*	E	Y	Y	(Brennan et al., 1988)
	D. grimshawi, D. affinidisjuncta, D. hawaiiensis	*Alcohol dehydrogenase*	E	Y	Y	(Wu et al., 1990)
D. melanogaster	*D. pseudoobscura*	*glucose dehydrogenase*	E	Y	Y	(Schiff et al., 1992)
D. melanogaster	*D. pseudoobscura*	*esterase 5*	E	Y		(Brady and Richmond, 1990)
D. melanogaster	*D. pseudoobscura, D. virilis*	*urate oxidase*	E		Y	(Wallrath and Friedman, 1991)
D. melanogaster	*Ceratitis capitata* (medfly)	*serum polypeptides*	E	Y	Y	(Christophides et al., 2000)
D. melanogaster	*D. quadrilineata*	*Scute*	R	Y	n.d.	(Marcellini and Simpson, 2006)
Vertebrates						
Mus musculus (lab mouse)	*Gallus gallus* (chicken)	*Hoxc8*	R	Y	Y	(Belting et al., 1998)
Mus musculus (lab mouse)	*Danio rerio* (zebrafish), *Fugu rubripes* (pufferfish)	*Hoxc9*	R	Y	n.d.	(Anand et al., 2003)
Mus musculus (lab mouse)	*Gallus gallus* (chicken)	*Lysozyme*	E	Y	n.d.	(Bonifer et al., 1996)
Mus musculus (lab mouse)	*Homo sapiens* (human)	*HLA_DRA*	E	Y	Y	(Giacomini et al., 1991)
Homo sapiens (human)[4]	*Galago crassicaudatis* (primate)	*gamma globin*	E	Y	n.d.	(TomHon et al., 1997)
Other						
Strongylocentrotus purpuratus (sea urchin)	*Lytechinus variegatus* (sea urchin)	*Endo16*	E	Y	Y	(Romano and Wray, 2003)
Caenorhabditis elegans (nematode)	*Caenorhabditis briggsae* (nematode)	*lin-48*	R	Y	Y	(Wang and Chamberlin, 2002)

All studies listed meet the following criteria: (i) an expression difference between two species was demonstrated and (ii) transgenic or genetic tests were used to compare activity of species-specific alleles in a common genetic background.

[1]Functional classification of gene product: R, a regulatory protein (e.g. transcription factor, signaling molecule), typically found near top of regulatory pathway; E, an effector protein (e.g. enzyme, adhesion, cell surface), typically found near bottom of regulatory pathway.

[2]Y indicates that *cis*-regulatory differences were inferred based on divergent activities of species-specific alleles in a common genetic background.

[3]Y indicates that *trans*-regulatory differences were inferred because *cis*-regulatory differences were insufficient to account for all aspects of expression divergence. n.d., data were insufficient to rule out *trans*-regulatory differences.

[4]Functional comparison performed in mice.

that has minimal effect on the overall network architecture. Mutations affecting the kinetics of individual regulatory interactions are expected to impact network output (i.e. gene expression) without altering its structure.

Interspecific comparisons of gene expression can be used to identify regulatory changes contributing to phenotypic divergence. Expression differences that correlate with phenotypic diversity have been observed for genes encoding transcription factors at the top of hierarchical pathways (e.g. Abzhanov et al., 2004; Averof and Patel, 1997; Gompel and Carroll, 2003; Sucena et al., 2003) as well as genes encoding enzymes at the bottom of pathways (e.g. Dickinson et al., 1984; Wittkopp et al., 2002). For all divergent phenotypes analyzed to date, only a subset of genes (often only one gene) in the developmental pathway is compared between species. Consequently, it remains unknown whether regulatory changes tend to cluster at the top, bottom or middle of a pathway.

Distinguishing between *cis*- and *trans*-acting variants is the first step for locating regulatory changes within a network. For a given gene, a *cis*-regulatory change indicates that the variant is associated with the gene surveyed. A *trans*-regulatory change indicates that the primary difference is located in a gene functioning upstream in the pathway. Transgenic and genetic studies comparing interspecific expression differences have shown that both *cis*- and *trans*-regulatory differences are common, regardless of where the gene fits within the regulatory pathway (Table 1). Consistent with these data, allele-specific analysis of 29 genes with expression differences between two *Drosophila* species found that 97% of genes with expression differences were affected by *cis*-regulatory divergence and approximately half showed evidence of *trans*-regulatory changes (Wittkopp et al., 2004).

Relating expression differences to underlying regulatory networks

The recent explosion of genomic resources and high throughput techniques is accelerating the elucidation of regulatory networks in model systems. Comparative studies of gene expression are an important tool for interpreting the function of network components and for understanding their stability and susceptibility to change. To understand how changes in regulatory networks contribute to phenotypic diversity within and between species, the following issues should be addressed.

(1) Many changes in gene expression are often induced by an environmental change. Some of the changes may be directly involved in the physiological adjustment while others may be secondary consequences of the regulatory network. Comprehensive descriptions of regulatory connections will help disentangle these types of changes by revealing connections between functional modules. However, genetic mapping and functional tests will ultimately be needed to identify the subset of genes for which expression changes impact the phenotype.

(2) Regulatory changes may be most stable when located in particular parts of a pathway. To test this hypothesis, complete pathways controlling divergent traits should be surveyed to locate all independent regulatory changes within the network. Locating regulatory variants will also make it possible to determine whether the connectivity of a gene within the network influences its propensity for change.

(3) The distribution of new regulatory mutations within a network appears to differ from the distribution of regulatory variants in the wild (Denver et al., 2005). Network architecture is expected to influence how regulatory variation arises, while the pleiotropic side effects of individual regulatory mutations are expected to influence which changes survive the test of time. To fully appreciate the impact of network architecture on evolutionary trajectories, properties that promote particular changes within regulatory networks must be identified.

(4) Some functional classes of genes may be more susceptible than others to regulatory mutations affecting their expression. Analyzing the distribution of regulatory variants among genes with different gene ontology designations will test this hypothesis. Such an analysis may also identify specific biological functions with a propensity for regulatory changes (e.g. sperm expressed genes in *C. elegans*) (Denver et al., 2005). However, any analyses using current gene ontology designations should be interpreted cautiously. At present, for most genes, gene ontology assignments of functional classes and biological processes are predicted solely based on sequence similarity and are awaiting genetic and/or biochemical verification.

As discussed in this review, existing case studies provide some insight into these issues. However, we have a long way to go toward understanding how regulatory variation is distributed within genomic regulatory networks and how network structure influences patterns of variable gene expression. A combination of genetic and biochemical dissection of regulatory networks in model systems, computational analyses of network properties, and comparative studies of gene expression among non-model species will be needed to resolve these issues. Given the recent growth in these research areas, a comprehensive understanding of regulatory variation in the context of regulatory networks may soon be achieved.

References

Abouheif, E. and Wray, G. A. (2002). Evolution of the gene network underlying wing polyphenism in ants. *Science* **297**, 249-252.

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. and Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462-1465.

Albert, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947-4957.

Anand, S., Wang, W. C., Powell, D. R., Bolanowski, S. A., Zhang, J.,

Ledje, C., Pawashe, A. B., Amemiya, C. T. and Shashikant, C. S. (2003). Divergence of *Hoxc8* early enhancer parallels diverged axial morphologies between mammals and fishes. *Proc. Natl. Acad. Sci. USA* **100**, 15666-15669.

Averof, M. and Patel, N. H. (1997). Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* **388**, 682-686.

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. and Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14**, 283-291.

Becskei, A. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593.

Belting, H. G., Shashikant, C. S. and Ruddle, F. H. (1998). Modification of expression and *cis*-regulation of *Hoxc8* in the evolution of diverged axial morphology. *Proc. Natl. Acad. Sci. USA* **95**, 2355-2360.

Blais, A. and Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes Dev.* **19**, 1499-1511.

Bonifer, C., Huber, M. C., Faust, N. and Sippel, A. E. (1996). Regulation of the chicken lysozyme locus in transgenic mice. *Crit. Rev. Eukaryot. Gene Expr.* **6**, 285-297.

Brady, J. P. and Richmond, R. C. (1990). Molecular analysis of evolutionary changes in the expression of *Drosophila* esterases. *Proc. Natl. Acad. Sci. USA* **87**, 8217-8221.

Brem, R. B., Yvert, G., Clinton, R. and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752-755.

Brennan, M. D. and Dickinson, W. J. (1988). Complex developmental regulation of the *Drosophila affinidisjuncta alcohol dehydrogenase* gene in *Drosophila melanogaster*. *Dev. Biol.* **125**, 64-74.

Brennan, M. D., Wu, C. Y. and Berry, A. J. (1988). Tissue-specific regulatory differences for the *alcohol dehydrogenase* genes of Hawaiian *Drosophila* are conserved in *Drosophila melanogaster* transformants. *Proc. Natl. Acad. Sci. USA* **85**, 6866-6869.

Britten, R. J. and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* **165**, 349-357.

Buckley, B. A., Gracey, A. Y. and Somero, G. N. (2006). The cellular response to heat stress in the goby *Gillichthys mirabilis*: a cDNA microarray and protein-level analysis. *J. Exp. Biol.* **209**, 2660-2677.

Carroll, S. B., Grenier, J. K. and Weatherbee, S. D. (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden: Blackwell Scientific.

Cheung, V. G., Spielman, R. S., Ewens, K. G., Weber, T. M., Morley, M. and Burdick, J. T. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365-1369.

Christophides, G. K., Livadaras, I., Savakis, C. and Komitopoulou, K. (2000). Two medfly promoters that have originated by recent gene duplication drive distinct sex, tissue and temporal expression patterns. *Genetics* **156**, 173-182.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Jr, Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D. and Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307**, 1928-1933.

Cooper, T. F., Rozen, D. E. and Lenski, R. E. (2003). Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **100**, 1072-1077.

Cossins, A., Fraser, J., Hughes, M. and Gracey, A. (2006). Post-genomic approaches to understanding the mechanisms of environmentally induced phenotypic plasticity. *J. Exp. Biol.* **209**, 2328-2336.

Davidson, E. H. and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800.

Davidson, E. H., McClay, D. R. and Hood, L. (2003). Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci. USA* **100**, 1475-1480.

Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M. and Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544-548.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A. M., Grove, C. A., Martinez, N. J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J. S., Hope, I. A. et al. (2006). A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**, 1193-1205.

Dickinson, W. J., Rowan, R. G. and Brennan, M. D. (1984). Regulatory gene evolution: adaptive differences in expression of *alcohol dehydrogenase* in *Drosophila melanogaster* and *Drosophila simulans*. *Heredity* **52**, 215-225.

Erives, A. and Levine, M. (2004). Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc. Natl. Acad. Sci. USA* **101**, 3851-3856.

Fay, J. C. and Wittkopp, P. J. (in press). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*.

Feder, M. E. and Walser, J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**, 901-910.

Ferea, T. L., Botstein, D., Brown, P. O. and Rosenzweig, R. F. (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. USA* **96**, 9721-9726.

Fischer, J. A. and Maniatis, T. (1986). Regulatory elements involved in *Drosophila Adh* gene expression are conserved in divergent species and separate elements mediate expression in different tissues. *EMBO J.* **5**, 1275-1289.

Giacomini, P., Ciucci, A., Nicotra, M. R., Nastruzzi, C., Feriotto, G., Appella, E., Gambari, R., Pozzi, L. and Natali, P. G. (1991). Tissue-specific expression of the *HLA-DRA* gene in transgenic mice. *Immunogenetics* **34**, 385-391.

Gilad, Y., Oshlack, A. and Rifkin, S. A. (2006). Natural selection on gene expression. *Trends Genet.* **22**, 456-461.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727-1736.

Gompel, N. and Carroll, S. B. (2003). Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* **424**, 931-935.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005). Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481-487.

Hornstein, E. and Shomron, N. (2006). Canalization of development by microRNAs. *Nat. Genet.* **38**, S20-S24.

Hubbard, L., McSteen, P., Doebley, J. and Hake, S. (2002). Expression patterns and mutant phenotype of *teosinte branched1* correlate with growth suppression in maize and teosinte. *Genetics* **162**, 1927-1935.

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.

Ke, X. S., Liu, C. M., Liu, D. P. and Liang, C. C. (2003). MicroRNAs: key participants in gene regulatory networks. *Curr. Opin. Chem. Biol.* **7**, 516-523.

King, M.-C. and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-116.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I. et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799-804.

Lemos, B., Meiklejohn, C. D., Caceres, M. and Hartl, D. L. (2005). Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126-137.

Levine, M. and Davidson, E. H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. USA* **102**, 4936-4942.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T. et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540-543.

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.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**, 11980-11985.

Marcellini, S. and Simpson, P. (2006). Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *PLoS Biol.* **4**, e386.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.

Monks, S. A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J. W., Sachs, A. and Schadt, E. E. (2004). Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094-1105.

Moorman, C., Sun, L. V., Wang, J., de Wit, E., Talhout, W., Ward, L. D., Greil, F., Lu, X. J., White, K. P., Bussemaker, H. J. et al. (2006). Hotspots

of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **103**, 12027-12032.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. and Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743-747.

Morozova, T. V., Anholt, R. R. and Mackay, T. F. (2006). Transcriptional response to alcohol exposure in *Drosophila melanogaster*. *Genome Biol.* **7**, R95.

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.

Oleksiak, M. F., Roach, J. L. and Crawford, D. L. (2005). Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat. Genet.* **37**, 67-72.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., True, J. R. and Carroll, S. B. (2006). Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050-1053.

Ranz, J. M. and Machado, C. A. (2006). Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**, 29-37.

Riehle, M. M., Bennett, A. F., Lenski, R. E. and Long, A. D. (2003). Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. *Physiol. Genomics* **14**, 47-58.

Rifkin, S. A., Kim, J. and White, K. P. (2003). Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138-144.

Rifkin, S. A., Houle, D., Kim, J. and White, K. P. (2005). A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220-223.

Romano, L. A. and Wray, G. A. (2003). Conservation of *Endo16* expression in sea urchins despite evolutionary divergence in both *cis* and *trans*-acting components of transcriptional regulation. *Development* **130**, 4187-4199.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G. et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Schiff, N. M., Feng, Y., Quine, J. A., Krasney, P. A. and Cavener, D. R. (1992). Evolution of the expression of the *Gld* gene in the reproductive tract of *Drosophila*. *Mol. Biol. Evol.* **9**, 1029-1049.

Siegal, M. L. and Bergman, A. (2002). Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci. USA* **99**, 10528-10532.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B. and Cohen, S. M. (2005). Animal microRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell* **123**, 1133-1146.

Stathopoulos, A. and Levine, M. (2005). Genomic regulatory networks and animal development. *Dev. Cell* **9**, 449-462.

Stern, D. L. (1998). A role of Ultrabithorax in morphological differences between *Drosophila* species. *Nature* **396**, 463-466.

Sucena, E. and Stern, D. L. (2000). Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo/shaven-baby*. *Proc. Natl. Acad. Sci. USA* **97**, 4530-4534.

Sucena, E., Delon, I., Jones, I., Payre, F. and Stern, D. L. (2003). Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424**, 935-938.

TomHon, C., Zhu, W., Millinoff, D., Hayasaka, K., Slightom, J. L., Goodman, M. and Gumucio, D. L. (1997). Evolution of a fetal expression pattern via *cis* changes near the gamma globin gene. *J. Biol. Chem.* **272**, 14062-14066.

True, J. R. and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* **3**, 109-119.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627.

Wallrath, L. L. and Friedman, T. B. (1991). Species differences in the temporal pattern of *Drosophila urate oxidase* gene expression are attributed to *trans*-acting regulatory changes. *Proc. Natl. Acad. Sci. USA* **88**, 5489-5493.

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.

Wittkopp, P. J. (2005). Genomic sources of regulatory variation in *cis* and in *trans*. *Cell. Mol. Life Sci.* **62**, 1779-1783.

Wittkopp, P. J., Vaccaro, K. and Carroll, S. B. (2002). Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547-1556.

Wittkopp, P. J., Carroll, S. B. and Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet.* **19**, 495-504.

Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85-88.

Wu, C. Y., Mote, J., Jr and Brennan, M. D. (1990). Tissue-specific expression phenotypes of Hawaiian *Drosophila Adh* genes in *Drosophila melanogaster* transformants. *Genetics* **125**, 599-610.

Yu, H. and Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. USA* **103**, 14724-14731.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R. and Kruglyak, L. (2003). *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**, 57-64.