

Statistical Epistasis Is a Generic Feature of Gene Regulatory Networks

Arne B. Gjuvsland,^{*,1} Ben J. Hayes,[†] Stig W. Omholt^{*} and Örjan Carlborg[‡]

^{*}Centre for Integrative Genetics (CIGENE) and Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway, [†]Animal Genetics and Genomics, Department of Primary Industries, Attwood, Victoria, Australia 3049 and [‡]Linnaeus Centre for Bioinformatics, Uppsala University, SE-751 24 Uppsala, Sweden

Manuscript received April 3, 2006

Accepted for publication September 18, 2006

ABSTRACT

Functional dependencies between genes are a defining characteristic of gene networks underlying quantitative traits. However, recent studies show that the proportion of the genetic variation that can be attributed to statistical epistasis varies from almost zero to very high. It is thus of fundamental as well as instrumental importance to better understand whether different functional dependency patterns among polymorphic genes give rise to distinct statistical interaction patterns or not. Here we address this issue by combining a quantitative genetic model approach with genotype–phenotype models capable of translating allelic variation and regulatory principles into phenotypic variation at the level of gene expression. We show that gene regulatory networks with and without feedback motifs can exhibit a wide range of possible statistical genetic architectures with regard to both type of effect explaining phenotypic variance and number of apparent loci underlying the observed phenotypic effect. Although all motifs are capable of harboring significant interactions, positive feedback gives rise to higher amounts and more types of statistical epistasis. The results also suggest that the inclusion of statistical interaction terms in genetic models will increase the chance to detect additional QTL as well as functional dependencies between genetic loci over a broad range of regulatory regimes. This article illustrates how statistical genetic methods can fruitfully be combined with nonlinear systems dynamics to elucidate biological issues beyond reach of each methodology in isolation.

MANY, if not most, biologists are prone to believe that genetic interactions are common in the genetic architecture of complex traits. It is, however, more debatable how important these interactions are in contributing to the expression of phenotypes in individuals and in determining population responses to selection, maintenance of genetic variation, and speciation processes. Studies of genetic interactions, or epistasis, are commonly based on hierarchical genetic models with additivity as the main effect and dominance and epistasis modeled, if at all, as single- and multilocus deviations from the main effects. Using these models, hybridization experiments have shown an important overall contribution of epistasis to the phenotypic differences among (DOEBLEY *et al.* 1995) and within (HARD *et al.* 1992; LAIR *et al.* 1997; CARROLL *et al.* 2001, 2003) species. The same observations have been made in studies that dissect quantitative genetic variation into contributions from individual quantitative trait loci (QTL) using epistatic genetic models (CARLBORG and HALEY 2004). PHILLIPS (1998) predicted that interaction between gene products that form molecular machines and signaling pathways would become increasingly important to genetic analysis and reinforce the concept of

epistasis. His predictions are supported by the appearance of the first genomewide mapping studies of epistatic interactions underlying gene expression in yeast (BREM and KRUGLYAK 2005), by the detection of epistatic pairs of genes, and by interpretation of these observations in terms of regulatory pathways (BREM *et al.* 2005).

Recent studies seeking to estimate how epistatic effects of individual loci contribute to phenotypic variance show that the proportion of genetic variation that can be attributed to statistical epistasis varies greatly among studies, where a very high proportion of the genetic variance is due to epistasis in some studies and virtually none in others (CARLBORG and HALEY 2004). As the studies are based on similar analytical approaches this suggests that there are biological reasons for the observed differences, and it is of importance both for understanding and for exploiting genetic variance to settle whether different functional dependency patterns between polymorphic genes (regulatory architectures) give rise to distinct statistical interaction patterns or not (we define gene A to be functionally dependent on gene B if the rate of change of expression of gene A changes when the level of gene B changes). If they do, statistical interaction patterns may reveal insights about underlying biological mechanisms. If not, it means that allelic variation within a given regulatory architecture determines the statistical interaction pattern and that very

¹Corresponding author: Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Aas, Norway. E-mail: arne.gjuvsland@cigene.no

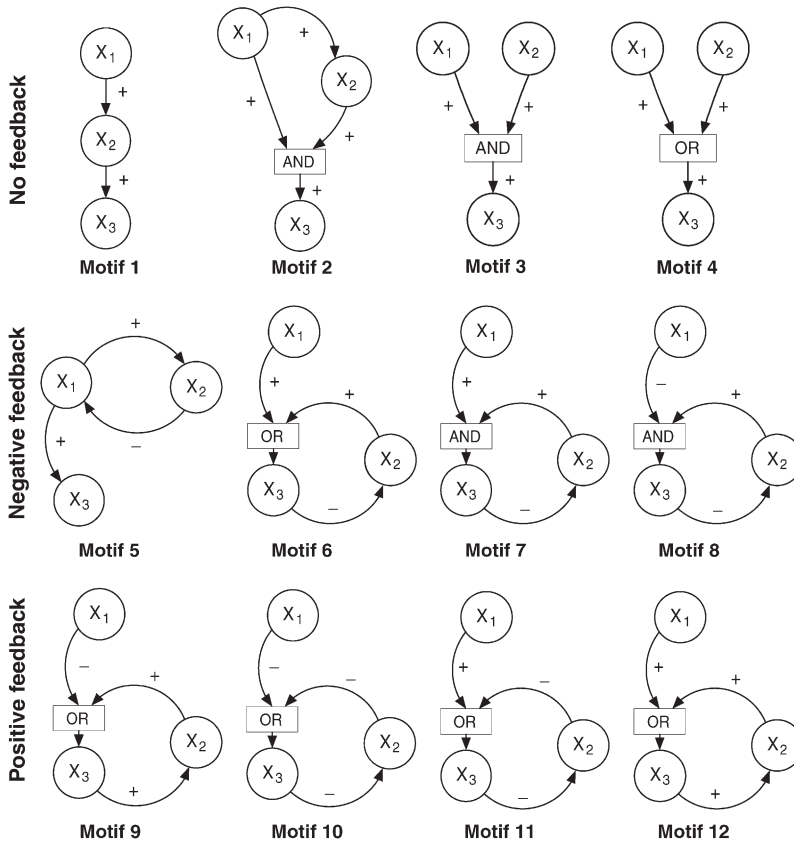


FIGURE 1.—Connectivity diagrams for the 12 network motifs in the simulation study. Each motif consists of three genes, named X_1 , X_2 , and X_3 and represented by circles. In the text they are called gene 1, gene 2, and gene 3, respectively. Genes without any arrows pointing at them are constitutively expressed, while an arrow pointing from gene i to gene j means that gene i is regulating the expression of gene j . The sign of the arrow indicates whether the type of regulation is activation (+) or inhibition (−). When a gene has two regulators the individual signals are combined with a logic block, represented by a rectangle, mapping the two signals into one by the continuous analog of the Boolean functions AND or OR.

little can be inferred about the underlying architecture from observed epistatic patterns alone.

Our work is part of an ongoing effort to understand population-level variation in terms of individual-level genotype–phenotype maps. Attempts to refine the concept of epistasis have been made [*e.g.*, “physiological epistasis” (CHEVERUD and ROUTMAN 1995) and “functional epistasis” (HANSEN and WAGNER 2001)] and studies have addressed the genetics of biological network models (WAGNER 1994; FRANK 1999; OMHOLT *et al.* 2000; YOU and YIN 2002; PECCOUD *et al.* 2004; COOPER *et al.* 2005; MOORE and WILLIAMS 2005; SEGRE *et al.* 2005; WELCH *et al.* 2005; AZEVEDO *et al.* 2006; OMHOLT 2006). In this work we study the relationship between statistical epistasis and functional dependency by doing quantitative genetic analysis of synthetic data sets obtained from genotype–phenotype models where phenotypic variation at the level of gene expression arises from allelic variation in model parameters. Using three-locus motifs of gene regulatory networks we elucidate the effects of no and one-way functional dependency in four regulatory situations in a no-feedback setting and the effects of one-way and two-way functional dependency in four regulatory situations in a negative- as well as in a positive-feedback setting. Our approach, where mathematical models generating phenotypic variability based on how genes work and interact are embedded into a statistical genetics context, illustrates how statistical methodology can be combined with nonlinear systems

dynamics to elucidate biological issues beyond reach of each of them in isolation.

METHODS

Network motifs: We made use of 12 gene regulatory models, each with a unique regulatory motif (Figure 1). Motifs 1–4 involve one-way functional dependency only (*i.e.*, regulatory actions), motifs 5–8 include two-way functional dependency in the form of negative feedback, and motifs 9–12 have two-way functional dependency in the form of positive feedback. It should be noted that although the motifs contain only three genes each, the models reflect a level of abstraction where the functional dependency does not necessarily involve direct biochemical interaction. That is, the models implicitly account for the possible presence of numerous nonpolymorphic additional genes in the networks. The models thus potentially capture a wide range of regulatory situations.

Model framework and equations: For modeling the gene regulatory motifs, we use the sigmoid formalism (MESTL *et al.* 1995; PLAHTÉ *et al.* 1998) for diploid organisms (OMHOLT *et al.* 2000). A gene regulatory network is described by a set of ordinary differential equations (ODEs),

$$\frac{d\bar{x}}{dt} = F(\bar{x}, \bar{\alpha}, \bar{\gamma}, \bar{\theta}, \bar{p}), \quad (1)$$

where the $2n$ -vector $\bar{x} = (x_{11} \ x_{12} \ x_{21} \ x_{22} \ \dots \ x_{n1} \ x_{n2})$ contains the expression levels of the two alleles for each of n genes in the gene regulatory network, while the vectors $\bar{\alpha}$, $\bar{\gamma}$, $\bar{\theta}$, and $\bar{p}$ contain allelic parameter values. Each allele, x_{ji} (the i th allele of gene j) has the parameters α_{ji} , the maximal production rate of the allele, and γ_{ji} , the relative decay rate of the expression product. In addition, for each gene x_k regulating the expression of allele x_{ji} , there is a threshold parameter θ_{kji} and a steepness parameter p_{kji} used to describe the dose-response relationship between x_k and the resulting production rate of x_{ji} . We assume for simplicity the allele products to be equally efficient as regulators and use just their sum ($y_k = x_{k1} + x_{k2}$) in the regulatory function.

We have used the Hill function (HILL 1910) in our simulations to generate a flexible dose-response relationship between regulator and production at the regulated gene,

$$H(y, \theta, p) = \frac{y^p}{\theta^p + y^p}, \quad (2)$$

where θ gives the amount of regulator needed to get 50% of maximal production rate while p determines the steepness of the response. The Hill equation describes Michaelis–Menten-like regulation for $p = 1$ and more switchlike response as p increases. If the regulatory effect is inhibitory, the regulatory function $1 - H(y, \theta, p)$ is used. Concerning our choice of the gene regulatory function, the Hill function is widely used in modeling of gene regulatory networks (BECSKEI *et al.* 2001; DE JONG 2002; ROSENFELD *et al.* 2002). There is a large body of literature supporting the presence of sigmoidal gene regulation functions, and the relationship can be due to cooperativity (VEITIA 2003), multiple transcription-factor binding sites, multiple phosphorylations (MARIANI *et al.* 2004), and spatially constrained kinetics (SAVAGEAU 1995). Thermodynamic modeling of *cis*-regulatory architecture also yields sigmoidal relationships (BINTU *et al.* 2005), and a recent empirical study of the λ -phage P_R promotor in *Escherichia coli* identifies regulatory functions closely resembling the Hill function (ROSENFELD *et al.* 2005).

Table 1 contains diploid ODE models of the diagrams in Figure 1. In all the equations $i = 1, 2$ and $y_j = x_{j1} + x_{j2}$, $j = 1, 2, 3$, and we use the notation $H_{kji}(y_k) = H(y_k, \theta_{kji}, p_{kji})$ for the dose-response relationships. In those motifs where the regulatory functions involve double inputs, we made use of the logical functions

$$\begin{aligned} \text{AND}(Z_1, Z_2) &= Z_1 Z_2, \\ \text{OR}(Z_1, Z_2) &= Z_1 + Z_2 - Z_1 Z_2. \end{aligned} \quad (3)$$

Simulations: We created 2000 homozygous parental lines for each regulatory model. Functional genetic variation between the lines was introduced by assigning allelic values for the regulatory parameters of each gene in the model (*i.e.*, production rates, regulation thresholds, regulation steepness, and relative decay rates)

through sampling from uniform distributions except for relative decay rates that were held fixed (Table 2). Then 1000 ideal F₂ populations of 960 individuals each were constructed by randomly crossing pairs of the 2000 inbred lines. The populations were ideal in the sense that the three genes in the network were at exact intermediate allele frequencies and in perfect Hardy–Weinberg and linkage equilibrium. The genotypes at each gene were recorded for each F₂ individual. The phenotype P for each individual was obtained by solving the differential equations describing the regulatory model to find the stable equilibrium level of gene 3 (y_3). To avoid artifacts arising from numerically solving the differential equations, equilibrium values < 0.01 were set to 0, and F₂ populations in which all phenotypes were 0 were discarded (this happened for 3 of 12,000 F₂ populations). For the remaining populations, the equilibrium levels were standardized to a unitary variance. To explore the statistical significance of the epistatic variance generated by the 12 interacting gene network motifs, we also generated F₂ populations (1000 populations for each motif and heritability) with two different (broad sense) heritabilities (0.2 and 0.05). This was done by adding independent normally distributed noise to the standardized equilibrium levels of the individuals in the population with phenotypes generated without noise.

Since we scaled the noise to predefined heritabilities the somewhat arbitrary absolute ranges from which parameters are sampled become less important than the relative variation between parameters. We paid particular attention to the latter type to ensure that the full spectrum of different regulatory situations could be reached for every regulatory function. The chosen values of α and γ gave steady-state levels in the range (20, 40) for a constitutively expressed gene and (0, 40) for a regulated gene. As these ranges overlap with the range (10, 30) for θ , it allowed the regulatory function to attain values close to the limits 0 and 1. This ensured a range of behaviors from all regulated genes being switched off to all being switched on for each regulatory function. For simplicity we fixed the decay rates, but since the production rates are under genetic control we should not lose any generality by this.

Genetic model and estimation of parameters and variance components: Following ZENG *et al.* (2005), and extending to three loci, the full genetic model

$$\begin{aligned} P = \mu &+ \sum_{j=1}^3 (a_j w_j + d_j v_j) \\ &+ \sum_{k=1}^2 \sum_{l=k+1}^3 (aa_{kl} w_k w_l + ad_{kl} w_k v_l + da_{kl} v_k w_l + dd_{kl} v_k v_l) \\ &+ aaa_{123} w_1 w_2 w_3 + aad_{123} w_1 w_2 v_3 + ada_{123} w_1 v_2 w_3 \\ &+ daa_{123} v_1 w_2 w_3 + add_{123} w_1 v_2 v_3 + dad_{123} v_1 w_2 v_3 \\ &+ dda_{123} v_1 v_2 w_3 + ddd_{123} v_1 v_2 v_3 \end{aligned} \quad (4)$$

TABLE 1
Ordinary differential equation (ODE) representations of the gene regulatory motifs

Motif 1	Motif 2
$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i}H_{12i}(y_1) - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i}H_{12i}(y_1) - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}H_{23i}(y_2) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{AND}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$
Motif 3	Motif 4
$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i} - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i} - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}\text{AND}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{OR}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$
Motif 5	Motif 6
$\dot{x}_{1i} = \alpha_{1i}(1 - H_{21i}(y_2)) - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i}H_{12i}(y_1) - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i}(1 - H_{32i}(y_3)) - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}H_{13i}(y_1) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{OR}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$
Motif 7	Motif 8
$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i}(1 - H_{32i}(y_3)) - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i}(1 - H_{32i}(y_3)) - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}\text{AND}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{AND}(1 - H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$
Motif 9	Motif 10
$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i}H_{32i}(y_3) - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i}(1 - H_{32i}(y_3)) - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}\text{OR}(1 - H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{OR}(1 - H_{13i}(y_1), 1 - H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$
Motif 11	Motif 12
$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$	$\dot{x}_{1i} = \alpha_{1i} - \gamma_{1i}x_{1i},$
$\dot{x}_{2i} = \alpha_{2i}(1 - H_{32i}(y_3)) - \gamma_{2i}x_{2i},$	$\dot{x}_{2i} = \alpha_{2i}H_{32i}(y_3) - \gamma_{2i}x_{2i},$
$\dot{x}_{3i} = \alpha_{3i}\text{OR}(H_{13i}(y_1), 1 - H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$	$\dot{x}_{3i} = \alpha_{3i}\text{OR}(H_{13i}(y_1), H_{23i}(y_2)) - \gamma_{3i}x_{3i}.$

was used, and using the F_2 metric we let

$$w_i = \begin{cases} 1 & \text{for } AA \\ 0 & \text{for } Aa \\ -1 & \text{for } aa \end{cases} \quad \text{and} \quad v_i = \begin{cases} -\frac{1}{2} & \text{for } AA \\ \frac{1}{2} & \text{for } Aa \\ -\frac{1}{2} & \text{for } aa, \end{cases}$$

where $\{AA, Aa, aa\}$ gives the genotype at gene i . On the basis of these equations and the simulated genotypes, we constructed a design matrix, X , containing vectors of regression variables for marginal effects and element-wise products of these variables for two-way interaction effects and three-way interaction effects such that

TABLE 2
Parameter ranges used for sampling alleles in creation of parental lines

Parameter	Range
α	[100, 200]
θ	[10, 30]
p	[1, 10]
γ	10

$$\begin{aligned}
X_{\text{marg}} &= [w_1 \ v_1 \ w_2 \ v_2 \ w_3 \ v_3], \\
X_{\text{two-way}} &= [w_1 w_2 \ w_1 w_3 \ w_2 w_3 \ w_1 v_2 \ w_1 v_3 \ w_2 v_3 \\
&\quad v_1 w_2 \ v_1 w_3 \ v_2 w_3 \ v_1 v_2 \ v_1 v_3 \ v_2 v_3], \\
X_{\text{three-way}} &= [w_1 w_2 w_3 \ w_1 w_2 v_3 \ w_1 v_2 w_3 \ v_1 w_2 w_3 \\
&\quad w_1 v_2 v_3 \ v_1 w_2 v_3 \ v_1 v_2 w_3 \ v_1 v_2 v_3], \\
X &= [X_{\text{marg}} \ X_{\text{two-way}} \ X_{\text{three-way}}]. \tag{5}
\end{aligned}$$

The dimensions of X are $n \times 27$, where n is the number of simulated individuals.

In our ideal populations there is no covariance between the columns of X . This allows us to estimate parameters in both the full genetic model (4) and any reduced model by regressing the simulated phenotypes on X and then just extracting the results from the columns associated with the particular model of interest.

Significance testing: We tested the significance of terms in various genetic models by the general linear hypothesis test (MONTGOMERY *et al.* 2001), with the test statistic

$$F_0 = \frac{[\text{SS}_{\text{Res}}(\text{RM}) - \text{SS}_{\text{Res}}(\text{FM})]/r}{\text{SS}_{\text{Res}}(\text{FM})/(n - p)}, \tag{6}$$

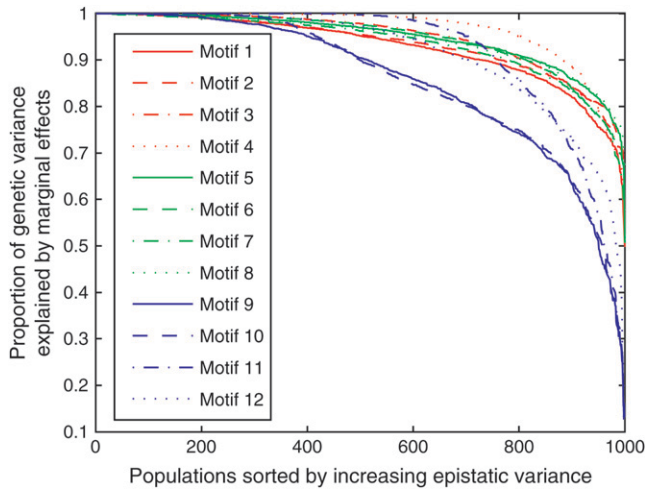


FIGURE 2.—Curves showing the distribution of the proportion of genetic variance explained by marginal (additive and dominance) effects of the three genes in motifs 1–12. The 1000 F_2 populations for each motif are sorted by an increasing amount of epistatic variance. The three different types of motif are represented by different colors, red for no feedback, green for negative feedback, and blue for positive feedback.

where the full model (FM) has p parameters, r is the number of parameters removed in the reduced model (RM), and n is the number of individuals in the F_2 population. Under the null hypothesis that none of the removed parameters are different from zero, the test statistic has the distribution $F_{1-\alpha, r, n-p}$.

RESULTS

Variance component signatures: To assess to which degree different functional dependency patterns between genes generate distinguishable statistical signatures, *e.g.*, the relative amount of variance explained by marginal, two-way interaction, and three-way interaction effects, these effects were estimated for the 12 motifs in the F_2 populations using phenotypes without noise. For all motifs, the expression level phenotype varied over the 1000 simulated populations from fully monogenic on one extreme to displaying large levels of statistical epistasis (maximum ranging from 29 to 88%; Figure 2). However, the positive-feedback motifs (motifs 9–12) generate larger amounts of statistical epistasis than the no- or negative-feedback cases (motifs 1–4 and 5–8). All positive-feedback motifs produce some data sets with >80% epistatic variance (maximum range from 81 to 88%), a level not reached by any of the other motifs (maximum range from 29 to 50%). Moreover, on average, motifs with an upstream inhibitor of the positive-feedback loop (motifs 9 and 10) generate more epistatic variance than motifs with upstream activation of the positive-feedback loop (motifs 11 and 12). The distribution of the marginal variance for motifs 1–8 is rather similar with the only exception being motif 4 that pro-

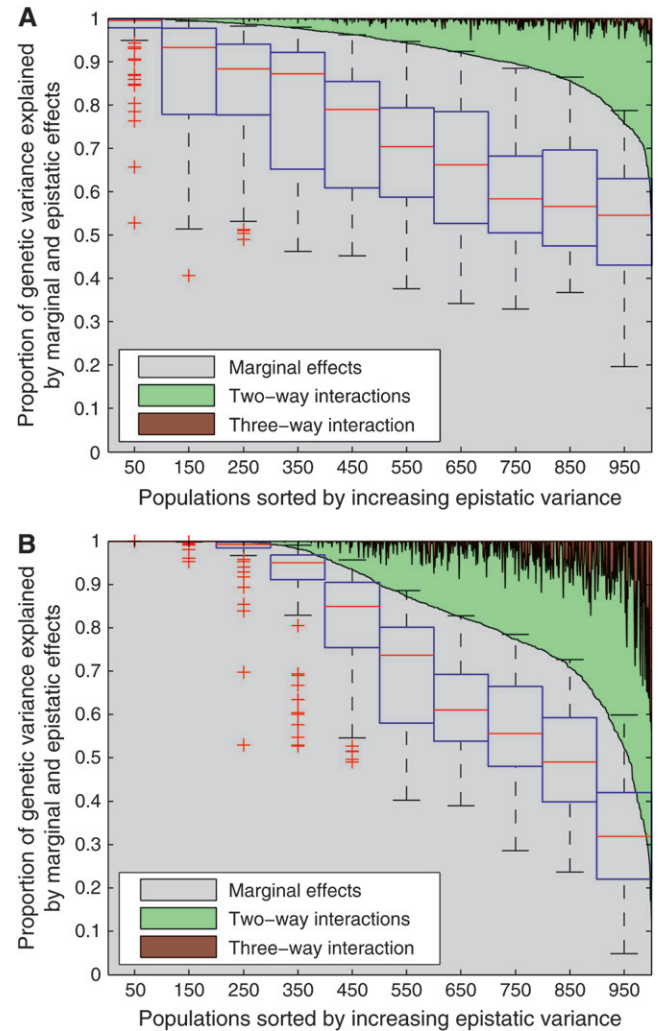


FIGURE 3.—The statistical genetic signature of the biological interactions in motifs 1 and 10 as the proportion of genetic variance explained by marginal (additive and dominance) effects, two-way interactions, and three-way interactions between the three genes of (A) motif 1 and (B) motif 10. The 1000 simulated F_2 populations are sorted by an increasing amount of epistatic variance. The box plots show the distribution, within bins of 100 populations, of the largest proportion of genetic variance explained by marginal effects of one gene.

duces data sets with particularly low levels of epistatic variance.

Figure 3 depicts in more detail the marginal-, two-way, and three-way epistatic variance distributions for a typical representative of the low epistatic variance group (motif 1, Figure 3A) and a typical representative of the high epistatic variance group (motif 10, Figure 3B). Among the 300 F_2 populations with the lowest level of epistatic variance, both motifs display an almost entirely marginal statistical genetic signature of the expression level phenotype with very low levels of epistatic variance (<2% for motif 1, <1% for motif 10). The marginal effects for motif 1 are mainly from one gene (averages of 97, 86, and 83% of the variance for the three 100-population bins) and even more so for motif 10

(averages of 100, 100, and 98%). In the 200 populations with the highest epistatic variance, the average proportion of the variance explained by the single largest gene alone decreases to ~55% for motif 1 and 40% for motif 10, and the levels of epistatic variance are considerably higher for motif 10 (25–88% of the variance) than for motif 1 (12–50%). It is also notable that a substantial portion of the explained genetic variance in motif 10 is due to three-way interactions (up to nearly 45% for some populations).

The general picture is that even though all motifs are capable of generating a flexible range of variance components, network motifs with positive feedback can generate higher amounts of two-way as well as three-way epistatic variance than motifs containing no feedback or negative feedback.

Statistical significance of two-way and three-way epistatic components: The statistical significance of the observed epistasis was explored using the simulated mapping populations with broad sense heritabilities (H^2) of 0.2 and 0.05. First, reduced models containing only marginal parameters were compared to full models containing all two-way interaction parameters. This was done for all three genes at once (19 *vs.* 7 parameters) and all three pairwise combinations of the three genes (9 *vs.* 5 parameters). Results for $H^2 = 0.2$ are summarized in Figure 4 ($H^2 = 0.05$ exhibited a similar pattern among the motifs and the results are not shown). We find that significant interactions occur much more often than expected by chance (5% significance level) for all motifs when using the full model including all three genes and their two-way interactions (27–62% of all populations). This is also true for the pairwise combinations of genes 1 and 3 and genes 2 and 3 (17–45% and 18–57% of populations, respectively). The percentage of significant interactions between gene 1 and 2 ranges from 3 to 26%, but for motifs 3 (3%) and 4 (6%) there are no more significant interactions than expected by chance (type I errors). As gene 1 and gene 2 in these two motifs are the only pairs in the whole study where either gene is functionally independent of the other, the simulated data sets show correspondence between the type of functional relationship between genes and the significance of the statistically detectable interactions. However, although this provides us with a conceptual link between functional dependency and statistical epistasis, it should

TABLE 3

Percentage of F_2 populations in which significant (5%) three-way interaction was detected

H^2	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
0.2	9.8	7.6	5.0	5.3	8.7	6.3	6.0	6.1	28.9	20.3	16.3	14.8
0.05	6.3	5.8	5.2	5.4	5.3	5.4	5.7	5.3	13.1	9.7	8.0	7.4

be noted that our analysis does not allow us to refine this link much further.

The significance of three-way interactions is generally lower than that for the two-way interactions (Table 3), but in the case of positive-feedback motifs, the level clearly exceeds what would be expected by chance.

Here we observe that the capacity to generate statistically significant two-way epistatic interactions is a generic feature of regulatory networks and that the capacity to generate statistically significant three-way interactions at the gene expression level is an additional generic feature of positive-feedback motifs.

Statistical significance of two-way interaction parameters: To get a better view of how the various two-way interaction parameters (additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance interactions) contribute to statistically significant two-way interactions in the various motifs, the significance of individual two-way interaction parameters was tested for all three pairwise combinations of the three genes (6 *vs.* 5 parameters) in the populations with $H^2 = 0.2$. The results are summarized in Figure 5, and we see that the positive-feedback motifs (especially motifs 9 and 10) frequently generate significance for all four types of interactions, while this is much less pronounced for the other motifs. Additive-by-additive interaction is the most frequently significant type of interaction for all pairs in all motifs. It is most frequent in pairs involving gene 3 and is in some cases significant in nearly half of the populations. Although significant additive-by-dominance and dominance-by-additive interactions are in general rather infrequent, they do appear more often than expected by chance, especially for motifs 9 and 10 where da_{23} and ad_{23} are significant in 20–37% of the populations. Except for motifs 3, 4, and 6, single *ad* or *da* parameters are significant in >10% of the populations. Significant

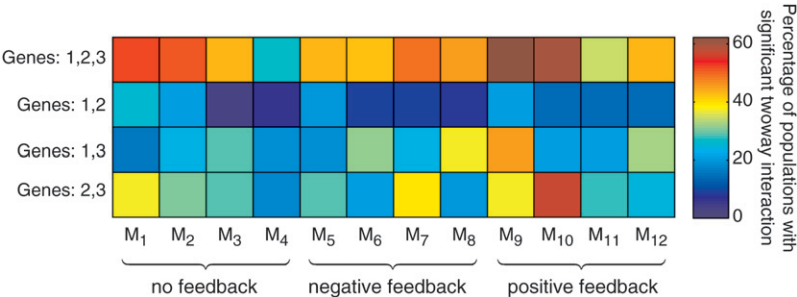


FIGURE 4.—The amount of significant two-way interactions between all pairs of genes in the 12 simulated network motifs for the broad-sense heritability $H^2 = 0.2$. The color coding indicates the percentage of the 1000 simulated F_2 populations for which a full model, including all marginal and two-way interaction parameters of the genes indicated on the y-axis, fits significantly better than a reduced model with only the marginal parameters.

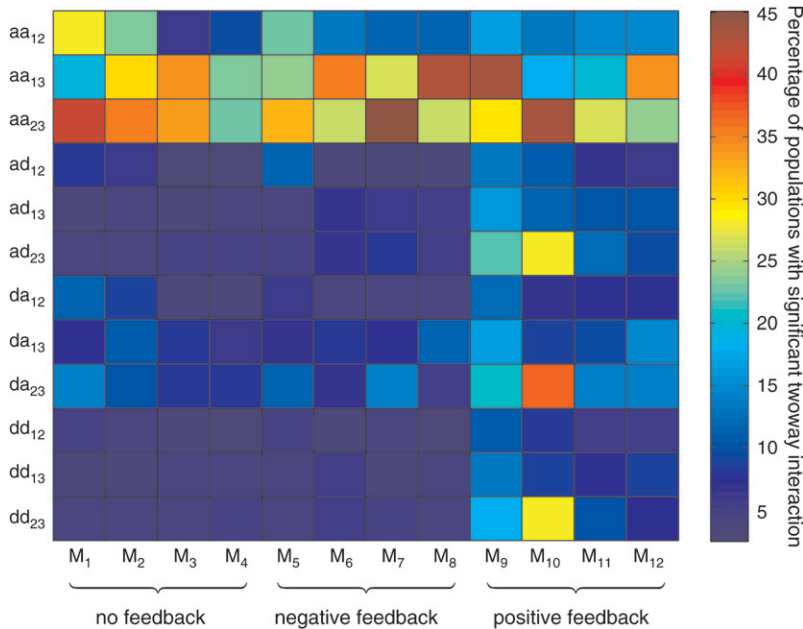


FIGURE 5.—The statistical significance of individual two-way interaction parameters for the 12 simulated network motifs. The color coding indicates the percentage of the 1000 simulated F_2 populations where the given interaction parameter is significant when a full model, containing all marginal parameters of the gene pair and the single interaction parameter indicated on the y-axis, is compared to a reduced model with only the marginal parameters.

dominance-by-dominance interactions occur more often than expected by chance only for positive-feedback motifs.

Number of genes detected: The fact that the capacity to generate statistically significant two-way and three-way interactions seems to be a generic feature of regulatory nets suggests that the search for such interactions may quite generally reveal more QTL underlying a complex trait than by solely searching for QTL independently of each other. To test this we did a three-step search for significant QTL effects in the populations with $H^2 = 0.2$. First the power to detect QTL independently was evaluated by testing for the significance of the marginal additive and dominance effects for each gene (3 *vs.* 1 parameters). Then the additional power to detect functionally dependent QTL using genetic models including epistasis was explored by testing for pairwise interaction effects in addition to the marginal effects (9 *vs.* 5 parameters) and finally we looked for trigenic interaction effects in addition to marginal and pairwise interaction effects (27 *vs.* 18 parameters). Figure 6 summarizes the distribution of the number of significant QTL after each step. We see that when

interactions are taken into account, there is for all motifs a considerable increase in the number of populations where three QTL are detected. Compared to testing for marginal effects only, an additional search for two-way interactions increased the number of cases where three QTL were detected by 22–36% more populations for motifs with no feedback or negative feedback and by 47–87% more populations for positive-feedback motifs. A further search for trigenic interactions increased the number of populations where three QTL were detected by 28–53% in the no-feedback and negative-feedback motifs and by 74–133% in the positive-feedback motifs (relative to testing for marginal effects only).

DISCUSSION

Possible shortcomings of our approach: In our network models we have focused on *cis*-regulatory mutations changing the dose-response relationship and the maximal production rate. As mutations may influence the gene expression patterns from given regulatory motifs in ways that are not accounted for here, we do not

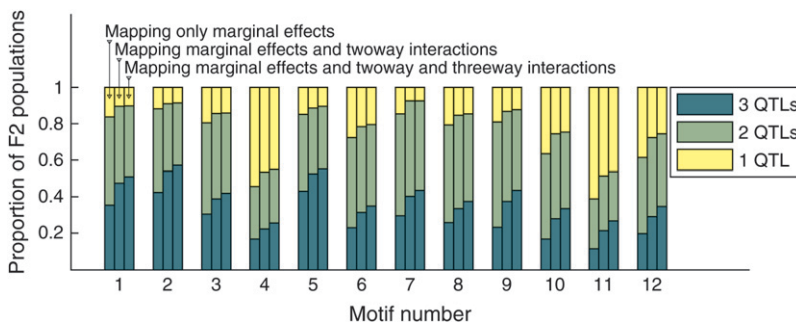


FIGURE 6.—The cumulative number of significant QTL for all 12 simulated network motifs at $H^2 = 0.2$ after testing for significant marginal effects only, when testing for significant two-way interactions in addition to the marginal effects, and when testing for significant three-way interactions in addition to marginal and two-way interaction effects.

pretend to have generated an exhaustive list of epistatic expression patterns. However, the behavior that can be generated from our models is quite extensive and we expect it to cover the majority of situations normally encountered. In addition, regulatory variation is indeed identified as an important factor in explaining complex traits (YAN *et al.* 2002). There are still few reports characterizing the mutations underlying QTL effects in terms of biological parameters. Two examples are described in a genetical genomics study in mouse (SCHADT *et al.* 2003), where allelic variation in transcript decay rates at the C5 gene and variation in the number of copies of the Alad gene, leading to higher production rates of transcript, were found to underlie two *cis*-acting eQTL. In addition, a genomewide study of regulatory variation underlying self-linkages in yeast (RONALD *et al.* 2005) identified a high proportion of *cis*-acting (promoters, transcription factor-binding sites, mRNA stability) variation.

The set of regulatory motifs in this study is not a complete collection of three-gene motifs, but we have included well-documented elements such as feed-forward loops and double input (LEE *et al.* 2002; SHEN-ORR *et al.* 2002). We also have a strong focus on feedback that is ubiquitous in biological systems (THOMAS and D'ARI 1990; CINQUIN and DEMONGEOT 2002) and contributes vital systemic features, where, *e.g.*, negative feedback is associated with homeostasis and positive feedback is a necessary prerequisite for multistationarity (PLAHE *et al.* 1995). Several regulatory motifs including feedback have been shown to be involved in the regulation of gene expression (LEE *et al.* 2002; DAVIDSON *et al.* 2003; WRAY *et al.* 2003) and it is likely that it will be an important component in the regulation of other complex traits as well.

In our simulations we include environmental variation by adding random noise to the equilibrium values of the regulatory systems. This is the standard way of doing simulations of quantitative genetic data and gives no covariance between genotype and environment. In many transcriptional regulatory systems external factors play an active role in regulating gene expression, for instance, in responses to stress conditions and utilization of nutrients. The approach used here could be expanded by including environmental variables as inputs to the regulatory functions. This would probably lead to significant genotype-by-environment interactions in much the same manner as we find genotype-by-genotype interactions in this study.

Testable predictions: Our studies confirm that traditional quantitative genetic models are, at least to some extent, able to detect functional dependencies within gene regulatory structures. This might seem like an obvious conclusion, but in our opinion it is not. Most evaluations of epistatic QTL-mapping methods have not aimed at exploring the ability of the method to detect various types of biological gene (actions and) in-

teractions, but rather at demonstrating and testing the properties of these methods for mapping of QTL whose inheritance conforms to standard quantitative genetics nomenclature (SEN and CHURCHILL 2001; CARLBORG and ANDERSSON 2002; KAO and ZENG 2002). Such simulations are thus useful for comparing mapping methods, but do not have any strong implications on the causal functional dependencies underlying the genetic interaction effects. In contrast to this, our simulations are based on the systemic features of a gene rather than its statistical effects. The genetic variance that can be detected by the statistical genetics model in our simulations thus emerges from polymorphisms describing allelic differences in properties affecting the expression of a gene in the context of a network of other genes. Our approach provides several new testable predictions concerning the ability of QTL-mapping methods to detect functional polymorphisms and dependencies in a genetical genomics context.

First, the amount of statistical epistasis generated by a biological network depends on system-level features such as the existence and sign of feedback. Regulatory structures with positive feedback are capable of generating more statistical epistasis than those with negative feedback, and these interactions are thus easier to detect in a QTL-mapping study.

Second, the amount of statistical epistasis that can be detected for a particular regulatory structure will vary widely depending on which of the regulatory parameters are affected by the genetic polymorphism. Figures 3 and 4 clearly show how the same regulatory structure can generate very different amounts of statistical epistasis: although polymorphisms are segregating at all loci, a three-gene network can statistically appear to be everything from a single major gene to a three-gene network with two- and three-way interactions. This also implies that in mapping studies where there are low levels of statistical epistasis such as in FLINT *et al.* (2004), there can still be functional relationships and network structures causally connecting the QTL.

Third, there is no clear pattern discerning one-way and two-way functional dependencies when it comes to the amount of statistical interaction. An example of this is that although all motifs with positive feedback show high amounts of epistatic variance (Figure 2), the gene pair most frequently showing significant epistasis differs between the motifs even though all motifs have the same underlying structure (Figure 4).

Fourth, the results strongly suggest that the inclusion of statistical interaction terms in the genetic model will increase the chance to detect additional QTL as well as functional dependencies between genetic loci. It thus seems worthwhile to put more effort into development of methods for mapping functional dependencies and to interpret statistical epistatic estimates in functional terms. Our simulations identify additive-by-additive as the most commonly produced interaction, and it is

therefore a strong candidate for inclusion in a reduced-interaction model. On the other hand, since the other types of interactions are less frequent, such patterns are of particular interest when it comes to biological interpretation of mapping results.

Although we in this article have limited ourselves to studying statistical epistasis patterns in a genetical genomics context, it should be noted that in addition to accounting for the possible presence of numerous other genes in the networks studied, polymorphisms in a given gene in our models can in principle influence the gene expression of another gene in the network through very complex routes involving higher-order phenotypic levels. In general, the relationship between genetic polymorphisms, regulatory dynamics, and statistical variance components can be monitored and analyzed at any phenotypic level, and there is no limit to how many systemic levels the genotype-to-phenotype models can include or how sophisticated these models can be. Fortunately, systems biology methodologies enabling us to make empirically well-founded mathematical genotype-phenotype models of more complex multilevel phenotypes are emerging very fast. This will open the way for a systematic investigation of the systemic conditions under which different types of functional dependency between polymorphic genes make detectable contributions to the genetic variance components of complex traits.

This study was supported by the National Programme for Research in Functional Genomics in Norway (FUGE) in the Research Council of Norway (grant no. NFR153302). Ö.C. is thankful to the Knut and Alice Wallenberg foundation for financial support. Visits by A.B.G. to the Linneaus Centre for Bioinformatics were supported by the Access to Research Infrastructures (ARI) program (project no. HPRI-CT-2001-00153).

LITERATURE CITED

- AZEVEDO, R. B., R. LOHAUS, S. SRINIVASAN, K. K. DANG and C. L. BURCH, 2006 Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**: 87–90.
- BECSKEI, A., B. SERAPHIN and L. SERRANO, 2001 Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**: 2528–2535.
- BINTU, L., N. E. BUCHLER, H. G. GARCIA, U. GERLAND, T. HWA *et al.*, 2005 Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15**: 125–135.
- BREM, R. B., and L. KRUGLYAK, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* **102**: 1572–1577.
- BREM, R. B., J. D. STOREY, J. WHITTLE and L. KRUGLYAK, 2005 Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**: 701–703.
- CARLBORG, O., and L. ANDERSSON, 2002 Use of randomization testing to detect multiple epistatic QTLs. *Genet. Res.* **79**: 175–184.
- CARLBORG, O., and C. S. HALEY, 2004 Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**: 618–625.
- CARROLL, S. P., H. DINGLE, T. R. FAMULA and C. W. FOX, 2001 Genetic architecture of adaptive differentiation in evolving host races of the soapberry bug, *Jadera haematoloma*. *Genetica* **112/113**: 257–272.
- CARROLL, S. P., H. DINGLE and T. R. FAMULA, 2003 Rapid appearance of epistasis during adaptive divergence following colonization. *Proc. Biol. Sci.* **270**(Suppl. 1): S80–S83.
- CHEVERUD, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455–1461.
- CINQUIN, O., and J. DEMONGEOT, 2002 Positive and negative feedback: striking a balance between necessary antagonists. *J. Theor. Biol.* **216**: 229–241.
- COOPER, M., D. W. PODLICH and O. S. SMITH, 2005 Gene-to-phenotype models and complex trait genetics. *Aust. J. Agric. Res.* **56**: 895–918.
- DAVIDSON, E. H., D. R. MCCLAY and L. HOOD, 2003 Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci. USA* **100**: 1475–1480.
- DE JONG, H., 2002 Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* **9**: 67–103.
- DOEBLEY, J., A. STEC and C. GUSTUS, 1995 *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**: 333–346.
- FLINT, J., J. C. DEFRIES and N. D. HENDERSON, 2004 Little epistasis for anxiety-related measures in the DeFries strains of laboratory mice. *Mamm. Genome* **15**: 77–82.
- FRANK, S. A., 1999 Population and quantitative genetics of regulatory networks. *J. Theor. Biol.* **197**: 281–294.
- HANSEN, T. F., and G. P. WAGNER, 2001 Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* **59**: 61–86.
- HARD, J. J., W. E. BRADSHAW and C. M. HOLZAPFEL, 1992 Epistasis and the genetic divergence of photoperiodism between populations of the pitcher-plant mosquito, *Wyeomyia smithii*. *Genetics* **131**: 389–396.
- HILL, A. V., 1910 The possible effect of the aggregation of the molecules of hemoglobin. *J. Physiol.* **40**: IV–VIII.
- KAO, C. H., and Z.-B. ZENG, 2002 Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* **160**: 1243–1261.
- LAIR, K. P., W. E. BRADSHAW and C. M. HOLZAPFEL, 1997 Evolutionary divergence of the genetic architecture underlying photoperiodism in the pitcher-plant mosquito, *Wyeomyia smithii*. *Genetics* **147**: 1873–1883.
- LEE, T. I., N. J. RINALDI, F. ROBERT, D. T. ODOM, Z. BAR-JOSEPH *et al.*, 2002 Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**: 799–804.
- MARIANI, L., M. LOHNING, A. RADBRUCH and T. HOFER, 2004 Transcriptional control networks of cell differentiation: insights from helper T lymphocytes. *Prog. Biophys. Mol. Biol.* **86**: 45–76.
- MESTL, T., E. PLAhte and S. W. OMHOLT, 1995 A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.* **176**: 291–300.
- MONTGOMERY, D. C., E. A. PECK and G. G. VINING, 2001 *Introduction to Linear Regression Analysis*. Wiley, New York.
- MOORE, J. H., and S. M. WILLIAMS, 2005 Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *BioEssays* **27**: 637–646.
- OMHOLT, S. W., 2006 From bean-bag genetics to feedback genetics: bridging the gap between regulatory biology and classical genetics, in *Biology of Dominance*, edited by R. A. VEITIA. Landes Bioscience, Georgetown, TX (<http://www.landesbioscience.com/books/id/887>).
- OMHOLT, S. W., E. PLAhte, L. OYEHAUG and K. F. XIANG, 2000 Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* **155**: 969–980.
- PECCOUD, J., K. V. VELDEN, D. PODLICH, C. WINKLER, L. ARTHUR *et al.*, 2004 The selective values of alleles in a molecular network model are context dependent. *Genetics* **166**: 1715–1725.
- PHILLIPS, P. C., 1998 The language of gene interaction. *Genetics* **149**: 1167–1171.
- PLAhte, E., T. MESTL and S. W. OMHOLT, 1995 Feedback loops, stability and multistationarity in dynamical systems. *J. Biol. Syst.* **3**: 409–413.
- PLAhte, E., T. MESTL and S. W. OMHOLT, 1998 A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.* **36**: 321–348.
- RONALD, J., R. B. BREM, J. WHITTLE and L. KRUGLYAK, 2005 Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet.* **1**: e25.
- ROSENFELD, N., M. B. ELWITZ and U. ALON, 2002 Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**: 785–793.

- ROSENFELD, N., J. W. YOUNG, U. ALON, P. S. SWAIN and M. B. ELOWITZ, 2005 Gene regulation at the single-cell level. *Science* **307**: 1962–1965.
- SAVAGEAU, M. A., 1995 Michaelis-Menten mechanism reconsidered: implications of fractal kinetics. *J. Theor. Biol.* **176**: 115–124.
- SCHADT, E. E., S. A. MONKS, T. A. DRAKE, A. J. LUSIS, N. CHE *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- SEGRE, D., A. DELUNA, G. M. CHURCH and R. KISHONY, 2005 Modular epistasis in yeast metabolism. *Nat. Genet.* **37**: 77–83.
- SEN, S., and G. A. CHURCHILL, 2001 A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.
- SHEN-ORR, S. S., R. MILO, S. MANGAN and U. ALON, 2002 Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**: 64–68.
- THOMAS, R., and R. D'ARI, 1990 *Biological Feedback*. CRC Press, Boca Raton, FL.
- VEITIA, R. A., 2003 A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol. Rev.* **78**: 149–170.
- WAGNER, A., 1994 Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc. Natl. Acad. Sci. USA* **91**: 4387–4391.
- WELCH, S. M., Z. S. DONG, J. L. ROE and S. DAS, 2005 Flowering time control: gene network modelling and the link to quantitative genetics. *Aust. J. Agric. Res.* **56**: 919–936.
- WRAY, G. A., M. W. HAHN, E. ABOUHEIF, J. P. BALHOFF, M. PIZER *et al.*, 2003 The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.
- YAN, H., W. YUAN, V. E. VELCULESCU, B. VOGELSTEIN and K. W. KINZLER, 2002 Allelic variation in human gene expression. *Science* **297**: 1143.
- YOU, L., and J. YIN, 2002 Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* **160**: 1273–1281.
- ZENG, Z.-B., T. WANG and W. ZOU, 2005 Modeling quantitative trait loci and interpretation of models. *Genetics* **169**: 1711–1725.

Communicating editor: S. MUSE