

Chapter 17

The Genotype–Phenotype Maps of Systems Biology and Quantitative Genetics: Distinct and Complementary

Christian R. Landry and Scott A. Rifkin

Abstract The processes by which genetic variation in complex traits is generated and maintained in populations has for a long time been treated in abstract and statistical terms. As a consequence, quantitative genetics has provided limited insights into our understanding of the molecular bases of quantitative trait variation. With the developing technological and conceptual tools of systems biology, cellular and molecular processes are being described in greater detail. While we have a good description of how signaling and other molecular networks are organized in the cell, we still do not know how genetic variation affects these pathways, because systems and molecular biology usually ignore the type and extent of genetic variation found in natural populations. Here we discuss the quantitative genetics and systems biology approaches for the study of complex trait architecture and discuss why these two disciplines would synergize with each other to answer questions that neither of the two could answer alone.

1 Evolution and the Molecular Underpinnings of Phenotypic Variation

Evolution proceeds in two phases: variation is generated and then sorted into the next generation. We now have a detailed knowledge of these two levels of evolutionary change. On the one hand, advanced research in molecular biology

C.R. Landry (✉)

Institut de Biologie Intégrative et des Systèmes (IBIS), Département de biologie, PROTEO, Université Laval, QC, G1V 0A6, Canada
e-mail: christian.landry@bio.ulaval.ca

S.A. Rifkin (✉)

Ecology, Behavior, and Evolution, Division of Biology, University of California, San Diego, CA, USA
e-mail: sarifkin@ucsd.edu

has provided clear descriptions of how mutations and chromosomal changes take place in organisms and estimates of the rates at which they occur [1, 2]. Population genetics has repeatedly seized upon new technologies to dissect the evolutionary forces acting on this genetic variation, now at thousands of loci genome-wide. At the other end of the spectrum of evolutionary biology, quantitative genetics has provided us with statistical models and descriptions of how phenotypic traits evolve under natural selection and genetic drift. However, except for relatively simple cases, we know little about how mutations modify the activity and dynamics of cellular networks and how this mechanistically translates into variation in phenotypes. For instance, comparative genomics of closely related *Drosophila* species has suggested that a large fraction of amino acid differences were fixed by natural selection, but their effects on phenotype remain unknown [3]. In parallel to the advances in population genetics, detailed descriptions of many cellular networks have emerged from investigations in cell and systems biology. In several cases, we have a clear picture of how cells perceive signals and how these signals are integrated to modify the physiology and the development of the organisms. Current models of these networks explain some of their dynamic properties including robustness, thresholds, homeostasis, and bistability.

Despite this tremendous progress, understanding how natural genetic variation affects complex networks and phenotypes remains one of the most important challenges in life sciences, as it would enable us to predict phenotypes from genotypes [4]. As the molecular details of how cellular networks integrate and translate genetic and environmental cues into complex phenotypes accumulate, we should be better able to describe how genetic variation affects phenotypes in molecular terms. However, because many developmental and cellular studies are based on single genetic backgrounds in a restricted set of environmental conditions, it is often far from clear how phenotypic variation arises, including context-dependent effects (epistasis, genotype-by-environment interaction) and incomplete penetrance of different alleles. To understand the generation of variation with existing conceptual and experimental tools, we propose that there needs to be a merger of quantitative genetics and systems biology. Here we discuss quantitative genetics and systems biology approaches for understanding phenotypic trait architecture and their limitations. We point to possible ways to combine them in order to gain a better understanding of how mutations translate into phenotypic variation to ultimately fuel evolution. We mainly draw our examples from research on the budding yeast *Saccharomyces cerevisiae* because this species has been used extensively as a test bed for both quantitative genetics and systems biology.

As we will see below, genotype–phenotype maps are virtual representations of how genes and alleles of genes relate to particular phenotypes. In quantitative genetics, these representations are often based on statistical associations between genotypes (alleles found in natural populations) and phenotypes. In systems biology, these maps most often represent functional associations between genes and phenotypes and are based on the systematic perturbation of the organism (gene deletion, drug treatments). While these two approaches both aim at describing

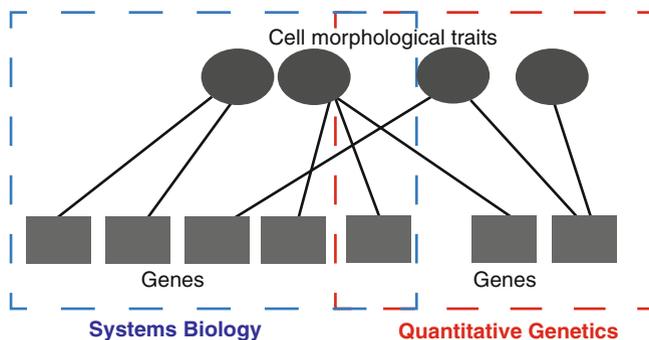


Fig. 17.1 QTL and systems biology approaches for identifying genes involved in cell shape in yeast identified two distinct groups of genes. Genetic variation that affects cell morphology among strains of yeast is not necessarily found in genes that, when deleted, affect cell morphology in the laboratory strains

gene–phenotype relationships, they often provide different pictures. An example of investigation on the genetic bases of cell morphology in the budding yeast illustrates how these two approaches can provide distinct results.

Single-celled organisms such as the budding yeast show variation in cell morphology that depends on cell-cycle stages, growth environments, and genetic backgrounds. Genes involved in determining normal cell morphology have been systematically identified using multidimensional phenotypic screening of 500 parameters on a set of 5,000 strains for which one gene was systematically deleted [5]. Half of the gene deletions of nonessential genes were found to affect one or more of the parameters describing cell morphology. Cell morphology is also known to vary among yeast strains. If one were planning on mapping genetic variation in natural populations that affects cell shape and morphology, would it be sufficient to sequence the 2,378 genes known to be involved in controlling cell morphology to find the causal genetic variation? A following study showed that this would absolutely not be the case. The same parameters were measured using exactly the same techniques in a pair of yeast strains and their F₂ progenies [6] in order to identify loci that would associate with the morphological differences between these two strains. Quantitative trait loci (QTL) for 95 of the traits could be mapped to specific loci. Only in one case did the QTL fall in the vicinity of a gene that had been found to affect cell morphology in the initial gene deletion screen (Fig. 17.1). If natural selection were to act on these traits in natural populations where the strains were sampled, it would most likely favor the fixation or elimination of alleles of loci that are not those found to affect cell morphology by gene deletion. This example is a particularly relevant one as both experiments were performed with identical platforms by the same laboratory and thus discrepancies are unlikely to come from technical differences. Why are different genes identified? Why do we need to study natural variation if we have identified the key genes in the laboratory strains?

We start with a review of the rationale of the two types of approaches and then discuss how their combination would enhance our comprehension of the molecular underpinnings of phenotypic evolution.

2 How are Quantitative Traits Transmitted Between Generations?

Quantitative genetics is a century old discipline with a rich theoretical foundation and a set of techniques that can be used for a range of purposes. Evolutionary biologists and agricultural breeders have tended to use these techniques to ask questions about the short-term phenotypic effects of selection in populations under particular mating schemes. In the last 20 years, molecular geneticists have used quantitative genetic techniques to identify loci underlying differences of interest in specifically constructed populations. Often these methods serve as foundries for generating candidate genes to feed into traditional molecular and developmental biology research projects. These complementary aspects are beginning to merge, particularly in plant and animal breeding and evolutionary genomics.

Quantitative genetics arose from an attempt to reconcile the inheritance of continuously varying traits with the particulate transmission genetics of Mendelism. In the late 1800s and early 1900s, the biometricians argued that the abundant variation in these quantitative traits could not be tied to the factors that were posited to underlay the discontinuous characters favored by the Mendelians. Continuous characters, they claimed, must have a different hereditary basis and different evolutionary properties. Because the Mendelians focused on transitions between discrete characters, they thought evolution proceeded by large steps—that mutations in the few loci underlying a trait would have large phenotypic effects. It took a series of experiments and theoretical work in the 1910s to demonstrate how particulate genes, when combined in large numbers, could generate the quantitative variation and covariation among relatives that so exercised the biometricians [7].

At its root, this disagreement was about how trait values and the distribution of these values in a population are transmitted to the next generation—they disagreed over the logic of genetics. Consider the fanciful case of an asexual organism that perfectly transmits its phenotype to its progeny. In this case the distribution of trait values in the population would only change from one generation to the next if individuals begat different numbers of offspring. For real organisms, traits are not perfectly inherited but are instead passed on with some variation. Nonetheless, offspring do often resemble their parents. Quantitative genetics asks whether the distribution of trait values (in particular the mean and variance of these values) changes in predictable ways from one generation to the next depending upon the mating system, transmission genetics, and evolutionary forces such as selection, mutation, migration, and drift. This often depended upon certain assumptions about

how trait values could be reconstructed from properties of the underlying genetic factors. This *genetic architecture* underlying a quantitative trait consists of [8,9]:

1. The number of loci involved;
2. The magnitude of the phenotypic effects of alleles at these loci or at least their average size and distribution;
3. How these effects are tempered by intra-locus (dominance) and/or inter-locus (epistatic) interactions;
4. Correlations between phenotypic effects of a locus on multiple traits (pleiotropy)

For much of its history, quantitative genetics was independent of the details of the actual loci underlying the trait. It was a theory of shifts in means and variances of phenotypic variation across generations given assumptions about the genetic architecture. By making reasonable assumptions about the genetic architecture, researchers could partition the variance in a trait into statistical components that reflected the aggregated properties of the underlying loci and had different implications for the response of the population to selection [10]. This worked remarkably well and was used to improve agricultural yields, study the effects of selection on morphological, life-history, and behavioral traits, and explain the fitness effects of inbreeding and bottlenecks [10]. There was no real way to drill down from trait variation to the molecular level, nor was it necessary for many types of predictions. The introduction first of highly variable molecular markers and then the ability to massively catalog single-nucleotide polymorphisms by sequencing made it possible to estimate the phenotypic effects of specific molecular differences between genotypes using QTL analysis [11]. These techniques would finally make it possible to work out the particulate details of inheritance of continuous traits. They have also instigated a reconceptualization of how quantitative genetic concepts should be defined.

QTL analysis and its congeners are widely used tools in medical, agricultural, and evolutionary genetics, and on a coarse level they have the same goal as systems biology—identifying important loci underlying a trait in order to predict phenotypes from genotypes. However, at a finer level of resolution the two differ in what kinds of loci they identify and what kinds of predictions they make possible. To parse these differences we will start with a concrete picture of genotype and phenotype spaces and examine how each field samples and connects them.

3 Phenotype Space

A phenotype is a description of an aspect or trait of an organism (or other biological entity such as a protein or a cell). At the most basic level, describing a trait in a collection of organisms consists of associating a phenotypic description with each organism thereby constructing a set of phenotypes. The structure of the phenotype space depends on the properties of this set, for example whether it has an ordering

or whether it makes sense to talk about distances between two different trait values. For example, height is a phenotype with a clear ordering and distance measure. It is less straightforward to think of an ordering and distance measure between different possible structures of a protein: a derived phenotype such as Gibbs free energy might serve this purpose. For both quantitative genetics and systems biology, phenotypes can often be represented as numbers on the real number line which both helps with intuition and computation. Indeed, in cases where a phenotype is discrete but ordered, biologists may posit that there is a hidden continuous phenotype which is thresholded to produce the discrete pattern and then proceed to work with this posited hidden phenotype to the extent that the data allows [12–14].

4 Genotype Space

The genotype describes the identity of the alleles of an organism at each locus. A locus can be thought of as a location on the chromosome that houses a gene while an allele is one of several variants of the gene. Alleles could differ by as little as a single base pair or as much as the whole locus (as with a knockout). The genotype is a discrete space. The number of possible alleles at each locus, the ploidy of the organism, and the rules for moving from one genotype to another determine its structure. One common simplification is to assume a haploid or diploid organism with two alternative alleles at each locus where the allele at a single locus can be changed in a single time step. For a haploid, the genotype space would then be a hypercube where genotypes are the vertices of the hypercube and the dimensionality of the hypercube depends upon the number of loci considered (Fig. 17.2). An edge of the hypercube would correspond to changing one allele for another at a

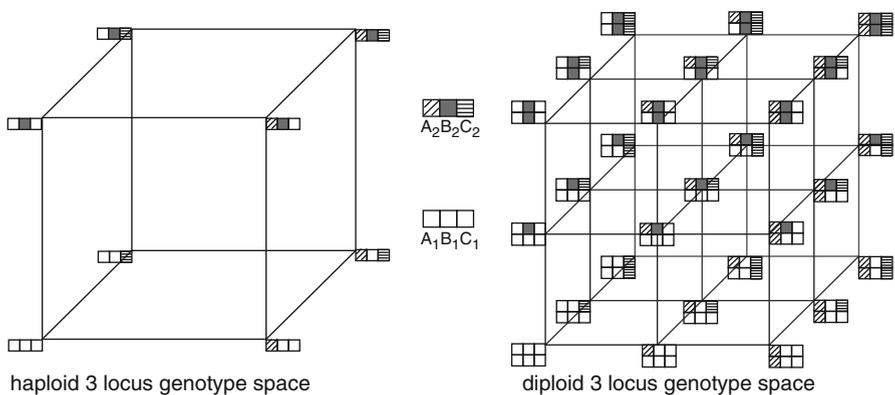


Fig. 17.2 Representation of genotype spaces. A haploid genotype space with three loci (A,B,C) each with two alleles (1,2) is on the *left*. Genotypes are marked at the vertices and edges are single allele changes. A diploid genotype space with three loci each with two alleles is on the *right*

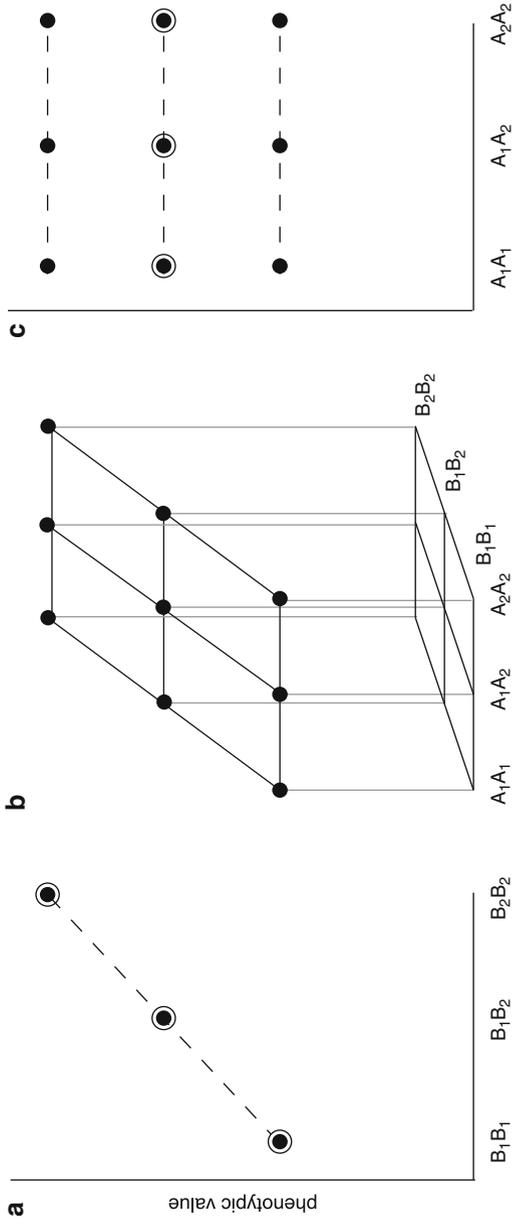
particular locus. The diploid space could also be arranged into a hypercube with a slight twist. Homozygous genotypes would populate the outer vertices of the cube, but an intermediate vertex representing the heterozygote for the corresponding locus would lie in between the homozygotes (Fig. 17.2). As in the haploid case, moving between two vertices would correspond to changing the identity of a single allele. To demonstrate the concepts of quantitative genetics, we will consider the diploid two-locus, two alleles case. More alleles, more loci, or higher ploidy are harder to visualize but the concepts extend straightforwardly.

5 Imperfect Sampling Complicates Estimating Genotype–Phenotype Maps in Finite Populations

A typical genotype–phenotype map consists of pairing each genotype with one or more phenotypes (Fig. 17.3).¹ Quantitative genetics is concerned with identifying regularities in this map. One way to approach this would be to sample a population measuring phenotypes and measuring or inferring genotypes or at least relatedness. A researcher would use this data to estimate how changing alleles changes trait values and variances of these values. In practice, however, each possible genotype can be sampled only if a small number of loci are considered. This becomes a problem if the phenotypic effect of swapping one allele for another depends upon the genotype—upon the identity of alleles at other loci. Although the *actual* genotypes in a population could be randomly sampled, the set of *possible* genotypes would not be. In this situation, estimates of the effects of swapping alleles could be biased, and in various quantitative genetic methods (usually general linear models [15]) allele frequencies and genotype-specific effects are entangled. The average effect of changing from allele A1 to allele A2 in a particular population (with its particular set of genotypes) will not be the same as the effect of changing from allele A1 to allele A2 in general, i.e. averaged across all possible genotypes.

Entangling these effects is often fine for some purposes—for example if the goal is to predict changes in the distribution of trait values in a specific population in response to selection [14–16]. But from a mechanistic perspective we would like to be able to predict how an individual trait value would change upon moving from one genotype to another—by mutation, for example. This is more akin to the approach of systems or synthetic biology where scientists investigate the phenotypic effects of specific perturbations. If we understood this map, we could then predict the phenotype distribution in a population by combining this mechanistic map with information on allele or genotype frequencies.

¹A single genotype can sometimes give rise to multiple phenotypic values depending on environmental conditions or random factors such as developmental and gene expression noise.



d

A_1A_1	A_2A_2	B_1B_1	5	5	5	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
A_1A_2	A_2A_1	B_1B_2	5	5	5	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
B_1B_2	B_2B_1	A_1A_1	6	6	6	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
B_2B_1	B_1B_2	A_1A_2	6	6	6	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
A_1A_1	A_2A_2	B_1B_1	5	5	5	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
A_1A_2	A_2A_1	B_1B_2	5	5	5	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
B_1B_2	B_2B_1	A_1A_1	6	6	6	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0
B_2B_1	B_1B_2	A_1A_2	6	6	6	$a_{A,11}$	$a_{B,11}=2$	0	0	0	0	0	0	0	0	0	0	0

$G_{11,11}=3$ $a_{A,11}$ $a_{B,11}=2$

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$d_{A,11}ed_{B \rightarrow A}$ $d_{B,11}$ $d_{B,11}ed_{A \rightarrow B}$ $add_{12,12}$ E_{AB}

6 An Idealized Diploid, Two-Locus, Two-Allele Case

To demonstrate the concepts of quantitative genetics, we will focus on the ideal case of a one-to-one or many-to-one mapping between genotype and phenotype where we do not have to sample because we know all genotype–phenotype pairs. Following the model of Wagner et al. [15], we will illustrate how a matrix of phenotypic values can be constructed from a set of more basic components. This model is but one of several mathematical representations of epistasis [16–18]. We chose to focus on it because it lends itself more readily to a mechanistic interpretation than other representations [15].

The most fundamental objective of most uses of quantitative genetics is to estimate the phenotypic effect of swapping one allele for another because this is how evolution by natural selection proceeds (see [19] for an alternative

←

Fig. 17.3 A genotype–phenotype map for a diploid, two-locus, two-allele system and its genetic architecture. Panel (b) depicts the map with the genotype space as the base and the heights of spheres above the base representing phenotypic values. Panel (a) depicts the projection of the phenotype landscape looking across the A alleles [from right to left in panel (b)]. In this example, the phenotypes collapse to a single line. The *closed circles* are the projections of the spheres from panel (b). The *open circles* are the average phenotypes at each B locus genotype. Panel (c) is similar to panel (a) except looking across the B alleles (from front to back in panel (b)). *Open and closed circles* are as described for panel (a). Panel (d) depicts the decomposition of the genotype–phenotype map into additive, dominance, and epistatic components [15]. The matrices represent these components for each genotype (first matrix on the left) and can be summed to generate the phenotype landscape. $\mathbf{G}_{11,11}$: phenotypic value for genotype $A_1A_1B_1B_1$. This is the reference genotype and components are defined as deviations from this base value. $\mathbf{a}_{A,11}$: the additive effect of substituting an A_2 allele for an A_1 allele. $\mathbf{a}_{B,11}$: the additive effect of substituting a B_2 allele for a B_1 allele. $\mathbf{d}_{A,11}$: the dominance effect of substituting an A_2 allele for an A_1 allele in the genotype A_1A_1 . Note that the right column is zero indicating that there is no dominance effect of this substitution when the starting genotype is A_1A_2 . $\mathbf{d}_{B,11}$: the dominance effect of substituting a B_2 allele for a B_1 allele in the genotype B_1B_1 . $\mathbf{ed}_{B \parallel A}$: a factor denoting the increase in the dominance deviation at locus A due to an additive change from B_1 to B_2 . This is additive-by-dominance epistasis. The total dominance deviation for locus A then becomes $\mathbf{d}_{A,11}(1 + \mathbf{ed}_{B \parallel A})$. $\mathbf{ed}_{A \parallel B}$: a factor denoting the increase in the dominance deviation at locus B due to an additive change from A_1 to A_2 . $\mathbf{edd}_{12,12}$: additional deviations in the double heterozygote including dominance-by-dominance epistasis. \mathbf{E}_{AB} : the additional additive effect of additive substitutions at the A locus due to a B_1 to B_2 change at the B locus. This is symmetric with respect to the loci and so is mathematically equivalent to the additional additive effect of additive substitutions at the B locus due to an A_1 to A_2 change at the A locus. In other words, additive-by-additive epistasis introduces the same deviation at each locus. This symmetry is due to differences in how this kind of epistasis scales the additive effects at each locus. $\mathbf{a}_{A,11}\mathbf{e}_{B \parallel A} = \mathbf{a}_{B,11}\mathbf{e}_{A \parallel B} = \mathbf{E}_{AB}$ where the $\mathbf{e}_{X \parallel Y}$ terms indicate the factor by which each additive deviation is scaled. Note that if the additive deviation at locus A ($\mathbf{a}_{A,11}$) is larger than that at locus B ($\mathbf{a}_{B,11}$), the additive-by-additive epistatic effect of locus A on locus B ($\mathbf{e}_{A \parallel B}$) is necessarily bigger than the equivalent for locus B ($\mathbf{e}_{B \parallel A}$). Moreover, $\mathbf{a}_{A,11}/\mathbf{a}_{B,11} = \mathbf{e}_{A \parallel B}/\mathbf{e}_{B \parallel A}$

conceptualization).² These are allelic effects. Unfortunately, estimating this effect is not always straightforward. Figures 17.3–17.5 depict genotype–phenotype maps for a diploid, two-locus, two-allele case with a real-valued phenotype represented by a vertical height.³ As the phenotypic landscapes become more complicated, it becomes less straightforward to determine the effect of swapping alleles at a locus because this effect becomes context dependent in several different ways. Each figure has four panels. In each figure, panels a and c show the projections of the phenotypes across variation in the A locus (panel a; looking across the panel b right to left) and across variation in the B locus (panel c; looking across panel b front to back), and the open circles in panels a and c represent the averages of the phenotypes for each genotype, averaged across the other locus. The subpanels of panel d decompose the phenotypic values into 9 orthogonal components in matrix form (see Fig. 17.3 caption).

In this two-locus, two-allele system, each genotype is accessible from any other genotype via 0, 1, or 2 substitutions at each locus. This means that we can arbitrarily assign one genotype to be a reference from which we measure deviations due to various allele swaps. We will use the genotype A1A1B1B1 as our reference. Note that the phenotypic value of the reference does not affect the phenotypic effect of substituting one allele for another.

Our goal will be to uncover regularities in how the phenotypic value changes when one allele is swapped for another—when moving along an edge of the genotype space of Fig 17.2. This involves partitioning the effect of any given allele swap into different components. There are three main categories. First are context independent effects: for example, changing from allele A1 to A2 adds 3 units to the phenotypic value. This is the additive effect. Second, the effect of changing alleles at a locus may depend upon the starting genotype at that locus. For example, if the genotype is A1A2, then changing from A1 to A2 adds an *extra* 2 units to the phenotypic value. This is a dominance effect. Third, the effect of changing alleles at a locus may depend upon the genotype at other loci. For example, changing from A1 to A2 adds an *extra* 4 units to the phenotypic value if the genotype at locus B is B1B2. This is an epistatic effect. These effects can be combined. For example, the size of the dominance effect may depend upon the genotype at locus B. This would be an epistatic effect on dominance. The total phenotypic effect of an allele swap would then be the sum of these component effects.

²It is increasingly clear that copy number differences are pervasive within populations. How duplications or deletions are handled within quantitative genetics depends upon how the genotype space is set up and conceptualized. Traditionally the edges of a genotype space (see Fig. 17.3) represent mutations between different alleles at a locus where each locus is a single copy. However, these genotype spaces could be used to represent movement between copy number variants. The “allele swapping” represented by an edge would not be point mutation or small indels but would instead be duplications or deletions. In this case the “allele” would be the copy number of the gene.

³Usually, the genetic component of a phenotype for a genotype that is predicted by a quantitative genetic model is called the *genotypic value*. In these examples we do not have any environmental effect and so the phenotypic landscape is also the landscape of genotypic values. For consistency with the systems biology section, we will talk in terms of phenotypes instead of genotypic values.

7 Additivity

Figure 17.3c demonstrates that if the B genotype is held constant, all three A locus genotypes have the same value. There is no effect of substituting A2 for A1 and the parallel lines indicate that this relationship between the A genotypes does not change depending upon the genotype at the B locus. Figure 17.3a indicates that there is an effect of changing from B1 to B2 and that it is the same effect whether going from B1B1 to B1B2 or from B1B2 to B2B2; the average heterozygote falls on the line connecting the two average homozygotes. The relationship between the B genotypes collapses to a single line—the average—in the left panel indicating that swapping between these two alleles at the A locus plays no role at all in the phenotypic variation here. This does not necessarily imply anything about the essentiality or mechanistic importance of the A locus or whether the protein from this locus physically interacts with the B locus protein or any other proteins. It does not mean that swapping between any alleles at the A locus has no effect. It only means that changing between the two A locus alleles under consideration has no phenotypic effect.

8 Dominance and Epistasis

The genotype–phenotype map of Fig. 17.4 is more complicated because it includes two deviations from additivity. The curves in Fig. 17.4a are straight but not parallel. The effect of changing from B1 to B2 does not depend upon the genotype at the B locus but *does* depend upon the genotype at the A locus. With A1A1 in the genetic background, substituting B2 for B1 increases the phenotypic value while in an A2A2 genetic background, the same substitution decreases the phenotypic value. This dependence of the additive effect of an allelic change at one locus on the genetic background is called additive-by-additive epistasis. In this example with only two loci, this is second-order epistasis. However, if more loci were considered, the allelic effects at locus B could depend upon genotypes at one, two, or more other loci, constituting ever higher orders of epistasis. In the panel on the right, the averages across the B locus (open circles) indicate that the effect of changing A1 to A2 depends on the genotype at the A locus. Moving from A1A1 to A1A2 has negligible effect while A2A2 has a lower phenotypic value than A1A2. This curvature of the lines connecting the average values of genotypes at the A locus indicates dominance, which is a deviation of the heterozygote value from the average value of the homozygotes. The relationship between the three curves corresponding to the A locus values at each of the three B locus genotypes seems rather complicated, but is simply a superposition of additive, dominance, and epistatic effects. The complicated pattern on the right can be reconstructed by adding the additive and dominance patterns at A to the additive-by-additive epistasis pattern (Fig. 17.4d).

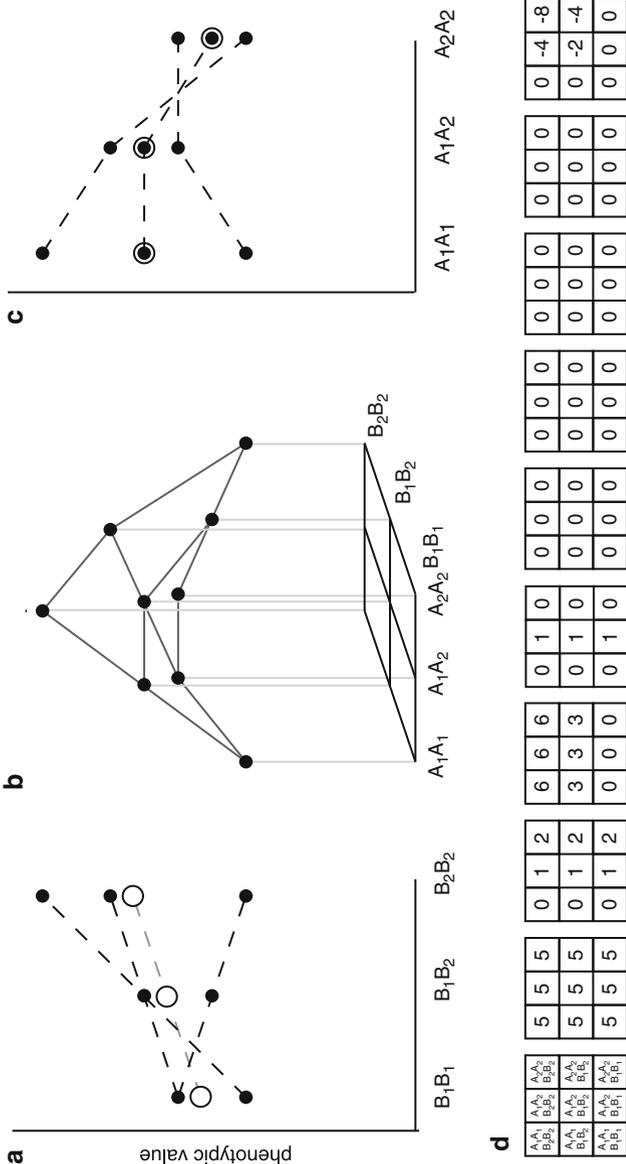


Fig. 17.4 A genotype-phenotype map that includes dominance and epistasis. The lines in panel (a) are straight indicating no dominance at the B locus. However, they are not parallel, indicating that there is additive-by-additive epistasis. The curves in panel (b) are not straight indicating dominance at the A locus. All three curves, however, have the same shape indicating that the genotype at locus B does not affect the dominance deviation at locus A

9 Higher Order Genetic Interactions

Even more complicated patterns can result when dominance relationships at one locus depend on the genotype at the other locus (dominance-by-additive epistasis). That is, when the form of an intra-locus relationship is a function of genotypes at more than one locus (Fig. 17.5). The curves on projection panels 5a and 5c show that there is no straightforward relationship between genotype and phenotype in this example. However, if a particular phenotype were measured using systems or molecular biology approaches only for heterozygotes at the A locus, the map might appear misleadingly simple. When the A locus is heterozygous, changing alleles at locus B has no effect. However, the same changes at B have strikingly different effects if the genotype is homozygous at locus A. Figure 17.5 demonstrates that studying a system in several genetic backgrounds can be crucial for truly understanding how phenotypes are generated by their underlying factors. Quantitative genetics can highlight when allele changes are likely to have an effect and when they will be masked.

10 Robustness

An allelic substitution in a homozygous dominant genotype has no phenotypic effect. This is a single locus example of robustness or canalization [20]. In general, if substituting alleles at a particular genotype has little effect, this genotype is robust to mutation or allelic substitution. One could visualize this by considering a local region (neighborhood) of the genotype space and the phenotypes associated with it. A 1-mutant neighborhood of a genotype, for example, would be the set of genotypes which differ from the focal genotype by a single mutation. This focal genotype would be robust to mutation if the phenotypes of its neighbors were similar. In this situation, the phenotype landscape would be relatively flat and unchanging *with respect to mutation*. This is akin to a parameter sensitivity analysis that is commonly used in dynamical systems modeling. If mutations have the effect of changing rate constants of reactions and other biochemical parameters, one might expect that a robust genotype would locate the biological system in a relatively insensitive region of parameter space. Although intuitive, this need not necessarily be the case: the phenotypic landscape *with respect to mutation* need not look the same as the phenotypic landscape *with respect to parameter change*. The phenotype landscape on the left of Fig. 17.6 could be generated by varying two parameters in a dynamical systems model. The mapping from a four-locus, two-allele haploid genotype space on the right to the phenotype landscape is indicated for a 1-mutant neighborhood around a focal genotype. The genotype is robust in the sense that mutants maintain the same phenotype even though the map from parameter value to phenotype is not flat [21].

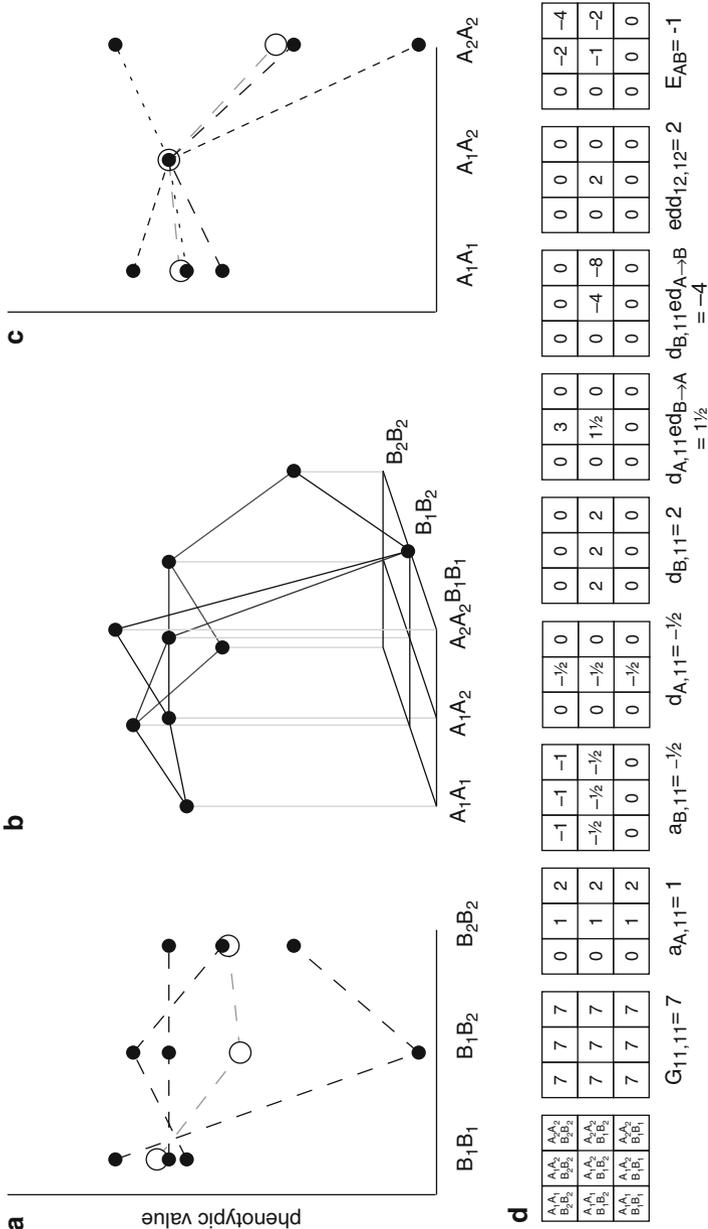


Fig. 17.5 A complicated genotype-phenotype map involving all eight deviations from the reference phenotype. Despite the complexity of the underlying genetic architecture, all phenotypes converge to a single value when the A locus is heterozygous. This demonstrates the importance of studying phenotypic phenomena in several genetic backgrounds

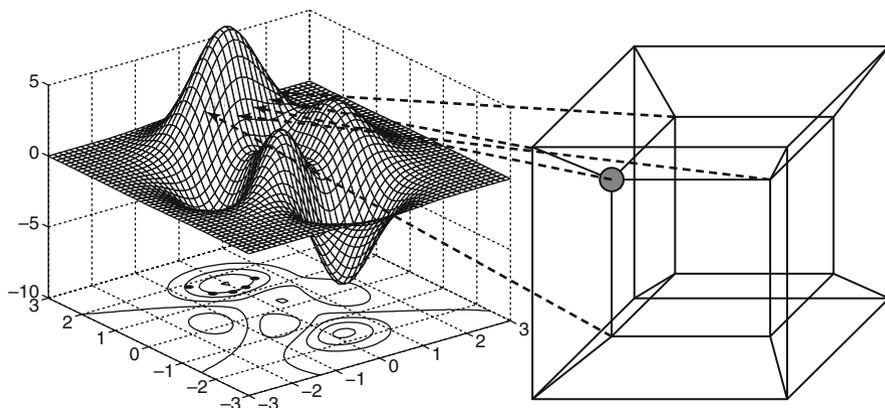


Fig. 17.6 The mapping of a local genotypic neighborhood onto phenotype space. The neighbors of the vertex with a *gray circle* around it (on the *right*) all have a similar phenotypic value (on the *left*), but the phenotypic landscape is not flat. In this case, the phenotype landscape is defined as it would be in systems biology—by varying two different parameters over some range. Distances in the phenotype space are therefore defined with respect to unit changes in the parameters. In quantitative genetics, phenotype landscapes are often defined with respect to unit changes in genotype—i.e. mutation

Empirical studies of genotype–phenotype maps in quantitative genetics mostly concentrate on the QTL mappings of traits of agricultural or ecological interests. Typically, these studies involve crosses between two genotypes that show large differences in the phenotypes of interest, analysis of recombinant genotypes (F₂ hybrids or backcrosses) and phenotypes, and tests for an association between genotypes and phenotypes. Molecular markers that co-segregate with the phenotypes of interest allow loci with significant effects on the phenotype to be identified. Their relative contributions to the trait, the level of pleiotropy of each QTL (how many traits each QTL affects) and epistasis among QTLs, can also be estimated. For the vast majority of studies, the QTLs identified are not dissected to a level where the actual causal DNA variants can be identified. There are two main reasons for this. First, most studies do not have the necessary resolution to narrow down QTL positions to specific nucleotides. This could be due to the small number of markers used and the small number of recombinant genotypes (number of recombination events) in the cross. The second reason is that estimating the variance due to additive, epistatic, and dominance effects even without identifying individual loci is often sufficient to answer fundamental questions about the evolution of quantitative traits in agriculture and in the wild.

11 What is a Genotype–Phenotype Map as Described by Systems Biology Approaches?

As seen above, quantitative genetics models of genotype–phenotype maps help predict and understand the outcome of evolution under specific selection regimes, the number of loci affecting the trait and the maintenance of genetic variation for a particular trait. Quantitative genetics is, however, largely blind to the mapping between the actual DNA sequences of the loci involved and the phenotypes at the molecular levels. Even when the actual causal DNA variants have been identified in QTL analysis, it remains difficult to draw the functional map between the sequence and the trait while including all the intermediate endophenotypes [4] (mRNA levels, protein levels, protein localization and modifications, signaling pathways activation, etc.), which is necessary for a complete understanding of the mechanisms of evolution and to eventually be able to predict phenotypes from DNA sequences alone. To overcome these limitations, many evolutionary biologists are turning to systems biology approaches where the main goals are to systematically identify all the genes involved in a trait and map the interactions among the genes and gene products involved. However, as we will see below, the two types of genotype–phenotype maps considered in the two approaches might not be completely equivalent and the best way to go might be to combine them.

12 Modular Biology

Systems biology is rich in operational definitions that help researchers formulate testable hypotheses and experiments at the molecular level. Typical approaches of experiments designed to directly link genes to phenotypes include the perturbation of a large number of genes and the measurement of the effects of these perturbations on traits of interest. Some types of experiments lead, for instance, to the annotation of genes as being essential for normal development in multicellular organisms [22] or genes that allow growth in a particular condition [23]. The genotype–phenotype map then consists in connecting a gene to a trait when perturbing that gene affects the trait (Fig. 17.7). As with QTL mapping experiments, these results allow researchers to identify the number of genes involved in each trait, their relative contributions, and their pleiotropic effects. When combinations of perturbations are considered, they allow interactions among genes to be estimated [24]. Very often investigators expect to identify a few key genes that are functionally related and that are responsible for the trait. Indeed, one of the predominant models describing how cells work posits that cellular functions—and thus phenotypes—are accomplished by groups of interacting molecules that form independent modules [25]. Accordingly, complex cellular functions cannot be reduced to particular genes but can be attributed to group of genes or proteins that interact in a particular manner. By definition, these modules are to some extent independent of each other [26], and

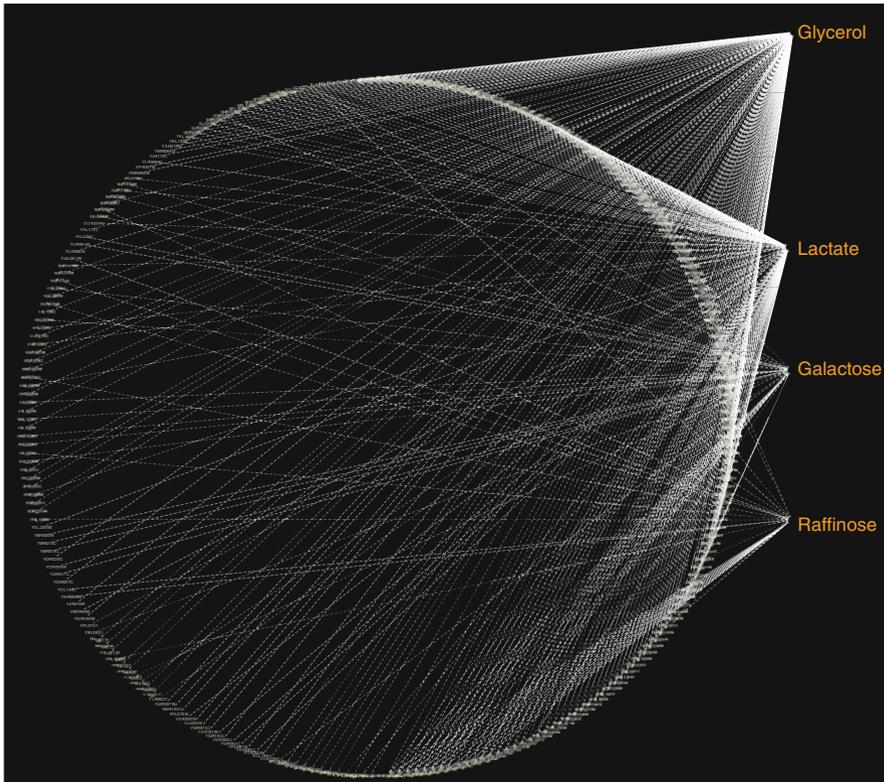


Fig. 17.7 Genotype–phenotype map of carbon source utilization in the budding yeast. Dudley et al. [23] grew a set of about 5,000 strains of budding yeast that each had a gene deleted on different carbon sources. By measuring the growth rate of the strains, they could associate hundreds of genes that are each required for normal growth on glycerol, lactate, galactose, and raffinose. These maps reveal that some genes are required to grow in several conditions (pleiotropy) and that some growth conditions require more genes than others. Only genes that were identified as being required for growth in at least one condition are represented

if we could comprehend their responses to intracellular and extracellular factors, we would understand the development of the particular trait to which this module contributes.

This modular vision of the cell is key to major advances in systems biology because it restricts the number of genes, proteins and RNAs and other molecules that need to be considered in mathematical models of complex behaviors such as cell decisions and commitment. This approach is extremely powerful. For instance, modeling, mutating, and replacing some of the key elements of these modules can modify the dynamic behavior of the cell in a predictable manner, i.e. they can make genotype–phenotype maps predictable. Clear demonstrations that we understand the function of a module include its isolation and its functional reconstitution from a

minimal set of elements. This has been shown for instance for the eukaryotic cell cycle control network whereby a minimal control system has been engineered to drive cell division events in a coordinated fashion [27] when introduced in a cell or for the assembly of a synthetic MAP kinase cascade that shows complex and predictable behavioral responses to external stimuli [28]. These experiments show that the elements necessary and responsible for these dynamic phenotypes have been identified and can be manipulated to work in a non-native context in a predictable fashion. With the development of synthetic biology approaches that enable the rational design of cell signaling circuits [29], we expect more demonstrations of this kind to support existing models of how modular structures regulate cell functions.

The success of systems biology at manipulating cellular behavior through the modification and isolation of cellular modules suggests that by identifying these modules, we are moving closer to a complete understanding of how cells and organisms work and thus of establishing functional links between genotypes and phenotypes. Accordingly, high-throughput experiments are aiming at describing cellular networks and providing descriptions and visualizations of key functional modules. In protein–protein interaction networks, these modules represent protein complexes with a well-defined function such as the proteasome, the nuclear pore complex, the RNA polymerase and many other unknown complexes or groups of proteins that interact with each other in one particular molecular pathway (Fig. 17.8) [30–32]. In the case of genetic interaction networks [24], these modules may represent genes that have coherent patterns of interactions with other genes in the genome and are thus constituted of genes with shared functions. They can also be groups of genes that show positive genetic interactions that reflect their membership to a particular molecular pathway or complex [24,33]. In models of metabolic networks, functional modules can be identified from the patterns of epistatic interaction among genes [34] or groups of genes that are highly connected among them based on network topology [35]. In the case of transcriptional networks, gene modules may represent co-regulated groups of genes and thus genes that are regulated by the same transcription factors or that are induced or repressed by the same signals upstream in the network [36]. Finally, in systematic genetic screens, modules could be groups of genes that, when individually inactivated, have similar effects on a phenotype of interest such as inability of the organism to develop a particular structure or to proliferate in a particular growth condition. Ultimately, these modular maps serve to associate genes with particular functions or phenotypes, to the extent that the function of a gene can be inferred simply from its patterns of association with other genes. This is the case, for instance, for the protein–protein interaction modules whereby the best predictor of a protein’s knockout phenotype is the knockout phenotype of the other proteins that form a protein complex with this protein [37]. With this modular organization in mind, building genotype–phenotype maps in systems biology results in connecting specific modules with traits of interests.

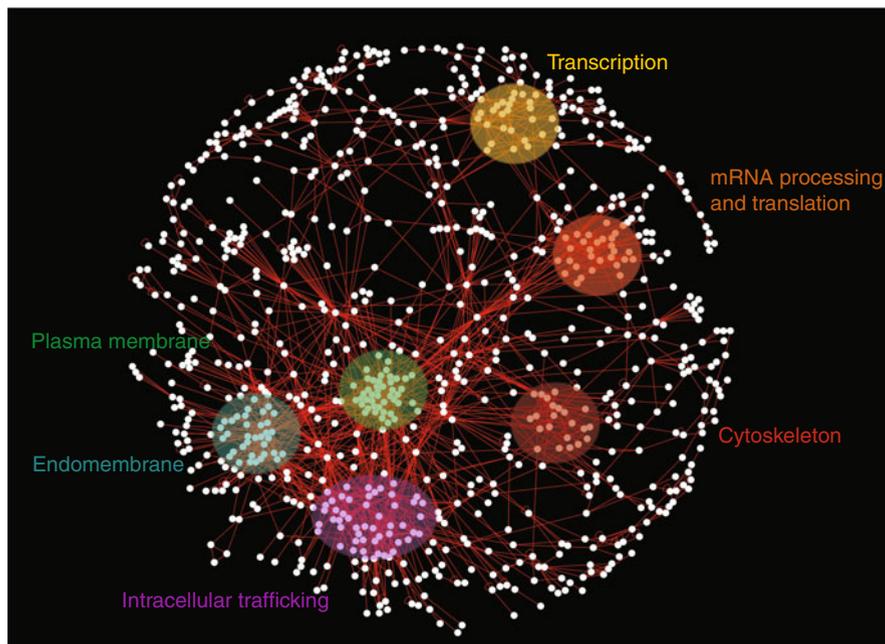


Fig. 17.8 The yeast protein interaction map as established by Tarassov et al. [30]. *White circles* represent proteins and *red arcs* pairwise interactions. These maps allow to visualize molecular modules (highly connected sets of nodes) that are involved in common molecular functions and their interconnections (figure provided by G. Diss)

13 Incomplete Congruence Between Systems Biology and Quantitative Genetics Maps

The identification of such functional cellular modules should in theory facilitate the identification of the genetic variation that underlies a trait of interest within or between species. For instance, when two individuals vary in a particular phenotype, the place to look at in the genome to find the underlying polymorphisms should be in the gene modules that have been identified as being involved in this phenotype. Similarly, one could model genetic variation in the trait of interest as slight modifications in the parameters of the reconstituted modules, such as concentrations of key elements, affinity constant, and half-lives. Intuitively, one would expect QTLs for a phenotype of interest to fall in the genes that have been shown through molecular genetics or systems biology to be involved in the trait. However, identifying the genes involved in a particular function or phenotype (necessary for the function) is quite different from identifying the mutations that may affect the trait in natural populations, as shown by the study on yeast morphology mentioned above. There are several reasons for this.

First, in quantitative genetics, the effect of an allele is a property of the allele in a particular genotypic and environmental context, but not of the locus. Mutant alleles are not necessarily interchangeable. Most gene annotations and genetic screens are derived from loss-of-function mutations and this type of alleles is likely to be rare in natural populations. Also, many effects caused by loss-of-function mutations may simply be masked by the presence of buffering mechanisms such as duplicated genes and alternative pathways [38]. Second, gain-of-function mutations are rarely studied and when they are, they are most often limited to gene overexpression, which represents only one particular case of gain of function. Others could be, for instance, amino acid substitutions that increase the catalytic activity of an enzyme or that make protein activity constitutive. These two types of genetic perturbations (complete deletion and overexpression) already confirm that different types of mutations are rarely equivalent: loss-of-function mutations by deletions and gain-of-function mutations by overexpression give strikingly different phenotypes when targeting the same genes [39]. Third, these studies are almost exclusively focused on single genetic backgrounds for each species and thus ignore complex gene-by-background interactions (epistasis), even if these have proven to be common. Even a very strict definition of a function or phenotype such as gene essentiality is highly dependent on the genetic background in which experiments are performed. For instance, the laboratory strain of *S. cerevisiae* was shown to have around 1,000 of its 5,000 genes (20%) as being essential [40]. A recent study on a closely related strain of the same species shows that 894 genes are essential in the two strains and 44 and 13 genes are essential in a strain-specific manner, and this, despite the fact that nearly 50% of the gene coding sequence are 100% identical between these two strains [41]. Genes that are reported to be essential in the laboratory background also show nucleotide polymorphism in nature and cause large phenotypic differences among individuals. Brown et al. [42] mapped the genetic basis of a complex gene expression phenotype segregating among vineyard yeast strains to a single nucleotide polymorphism. The polymorphism is a frame-shift mutation in *SSY1*, a gene encoding an amino acid transporter that is annotated as being essential in the laboratory strains. These strains have auxotrophic markers that impede the synthesis of certain amino acids, which makes their importation necessary.

14 Modular Biology, Distributed Genetic Effects

Another reason why systems biology and quantitative genetics maps may have limited overlap could be that systems biology approaches identify only the most important constituents—genes with the strongest effects on the phenotypes—and largely underestimate pleiotropic effects. Several lines of research suggest that what have been viewed as isolated, canonical molecular pathways and modules in the cells are in fact more connected than previously assumed [43]. The component with major contributions, i.e. those that can be measured and detected in typical large-scale experiments, may in fact be the components that would form the core

of the modules. There may be marginal contributions of many other genes in the genome that are missed through typical experiments and thus be excluded from current representation of the molecular networks that underlie key cellular functions. A dense pathway organization may in fact be only visible when more sensitive and direct measurements of endophenotypes are performed. For instance, this view is emerging in studies of biomolecular networks. A recent protein–protein interaction map aimed at establishing links among cellular regulators (protein kinases and phosphatases) indeed revealed that unlike what is traditionally shown in the linear representations of signaling pathways, regulatory proteins make many interactions with other regulatory proteins and do not restrict their activity to a limited number of modules [44, 45]. This model of a highly densely connected network of cellular regulators is also supported by sensitive proteomics screen that showed that inactivation of most protein kinases and phosphatases affect large parts of the cell signal transduction machinery and are not limited to canonical pathways or modules [46], despite what is suggested from the modular view of cellular systems. It is therefore possible that most QTLs are not located at the core of the modules but act in the periphery.

15 Data Integration in Evolutionary Systems Biology

Despite the disparity between the two types of genotype–phenotype maps, there is ample evidence that they are not completely orthogonal and that the two types of maps can be integrated. Indeed, there are several examples where both maps are used to better interpret how cellular networks are organized and evolve. For instance, we used data from the functional dissection of the yeast transcriptional network to show that when a gene was highly connected in the transcriptional network, it was more likely to evolve new expression levels under neutral evolution and to show genetic variation for gene expression in natural populations [47]. The integration of large-scale systems biology data with that of yeast expression QTLs also allowed to build predictive models of causal relationships between DNA variation and endophenotypes. In this case, the use of prior knowledge from systems biology enriched the types and power of the inferences that can be made [48]. A recent paper by Jelier et al. [49] offers an elegant illustration of how the combination of systems biology and population genomics can be used to predict the effect of mutations on phenotypic variation. Using the partial genomic sequences of 19 strains of yeast, the authors used phylogenetic comparisons to estimate the likelihood that mutations will have an effect on proteins functions. Using phenotypic data on the effects of gene deletions collected in systems biology investigations in laboratory strains, the authors were able to make and test predictions on the growth phenotype of the natural strains in specific conditions. Surprisingly, the approach works and shows that from comprehensive systems biology genotype–phenotype maps, we can start to build predictive models of how natural genetic variation may affect cellular phenotypes.

The systematic combination of systems biology and quantitative approaches will provide more information than these two independent fields can provide on their own. The integration of large-scale systems biology data with that of yeast expression QTLs, for instance, allowed to build predictive models of causal relationship between DNA variation and endophenotypes. In this case, the use of prior knowledge from systems biology enriched the types and power of the inferences that can be made [48]. Ultimately, a complete description of the genotype–phenotype maps of all the molecular levels between DNA sequence and organismal phenotypes such as morphology or behavior would be necessary to fully comprehend how phenotypic variation is generated. This would allow mapping the causal relationships between different levels of organizations and phenotypic variation that affects fitness in an ecological context. In principle, any molecular trait that can be quantitatively measured and that is heritable can be assessed using these approaches. Recently, these systems approaches have been applied to the genetic dissection of natural variation in molecular traits. Instead of measuring organismal traits and relating them to genotypes, systems genetics approaches have focused on quantifiable molecular phenotypes. The budding yeast, which has been the test bed for the development of most systems biology approaches, provides several examples of such approaches. Molecular phenotypic traits such as gene expression levels [50] (mRNA abundances), stochasticity in protein abundance [51] and transcription factor DNA binding intensities [52] have been genetically mapped among natural strains.

One pioneering series of studies on the combination of systems biology approaches with quantitative genetics comes from the comparison of the transcriptional landscape of two yeast strains and their segregants. In these experiments, more than 100 haploid segregants of a cross between a laboratory (BY) and a vineyard strain (RM) have been densely genotyped and expression profiled [50]. The results show that gene expression levels are highly heritable among yeast strains. Whereas a large number of transcripts (up to 75%) map to at least one QTL, 50% of all transcripts may have at least five additive QTLs and 20% at least 10 additive QTLs [53]. Furthermore, more than 57% of transcripts are influenced by a genetic interaction and a similar proportion (47%) is influenced by genotype-by-environment interaction. This confirms that even relatively simple traits such as transcript abundances may have very complex genotype–phenotype maps, with many QTLs per trait and abundant context dependent effects.

In order to completely elucidate genotype–phenotypes maps at the molecular levels, we need integrative approaches where not only mRNA abundances are considered but also many other endophenotypes such as protein levels, protein activity (post-translational regulatory states) as well as metabolite levels, signaling network activities and cell physiological states. Combining multiple levels of QTL analysis, from macroscopic to microscopic phenotypes, will then allow drawing causal relationships between DNA polymorphism, expression polymorphism, protein interaction, cell physiology and morphology and eventually organismal phenotypes. Identifying one QTL that affects traits at several of these levels of organization would reveal how a mutation affects, in a causal manner, mRNA

expression, protein abundance, protein activity, and eventually cellular or organismal phenotypes. Another advantage of this integrative approach is that we expect that significant levels of polymorphism that are not visible at the level of mRNA abundance to be visible at other levels of cellular organization. For example, several cellular responses are taking place in a time frame that is much shorter than what is needed for gene expression to be induced or repressed, such as phosphorylation cascades and neural activity. Unfortunately technologies are much more advanced in terms of gene expression profiling than they are for any other type of measurement of molecular phenotypes. However, there have been important advances in the development of tools that allow us to study cellular responses systematically. For instance, molecular tools and reporters are available to study the dynamic of signaling cascades in vivo in model organisms such as *S. cerevisiae* using protein-interaction reporters that can be integrated into the genome [54]. Protein activities can now be measured on a proteome-wide scale using large-scale phosphoproteomics [55] or TF-DNA binding sites [52].

Another key advantage of these integrative approaches is that it will allow us to compare the genotype–phenotype map of multiple levels of organization, which will clarify how genotypic information is “translated” in a cell. While we now know of a few examples where the molecular contributions of a QTL at the cellular level can be suggested from the sequence data, many quantitative genetics phenomena still remain almost completely unresolved at the molecular and cellular level. These include for instance the buffering of genetic variation at one locus by another or by cellular and developmental processes, as well as genotype-by-sex interactions, genotype-by-environment interactions and incomplete penetrance. These interactions are all cases where the effect of an allele depends on the state of the cellular networks. Quantitative genetics has very little to say on how these complex interactions could take place. The joint analysis of natural variation with the combination of the measurements of several endophenotypes and/or perturbations will be key to achieve these goals. The recent study of QTL for transcript levels and protein abundance in yeast exemplifies this rationale [56]. It was found that in many cases, heritable gene expression differences (mRNA) do not translate into differences in protein abundance. This means that a significant fraction of the genetic variation at one level maybe filtered out by the cell in subsequent steps. Inversely, some variation that affects protein abundance is not present or detectable at the transcriptional level. Either there is no transcriptional variation of that gene or it is amplified and becomes detectable only at the protein level, or it is during translation or protein degradation that heritable genetic variation is exposed. Another recent study on the divergence of gene expression levels between species pointed towards such key levels in the cell where the extent of genetic variation is modified. Different species of yeast show divergent patterns of gene expression levels. Tirosh et al. examined the role of chromatin regulators in shaping this divergence between *S. cerevisiae* and *S. paradoxus* [57, 58]. When chromatin regulators are deleted in these two species, the authors observed a systematic increase in the divergence of gene expression levels, which is consistent with a model under which chromatin regulators buffer genetic variation that acts upstream in transcriptional networks. Together, these

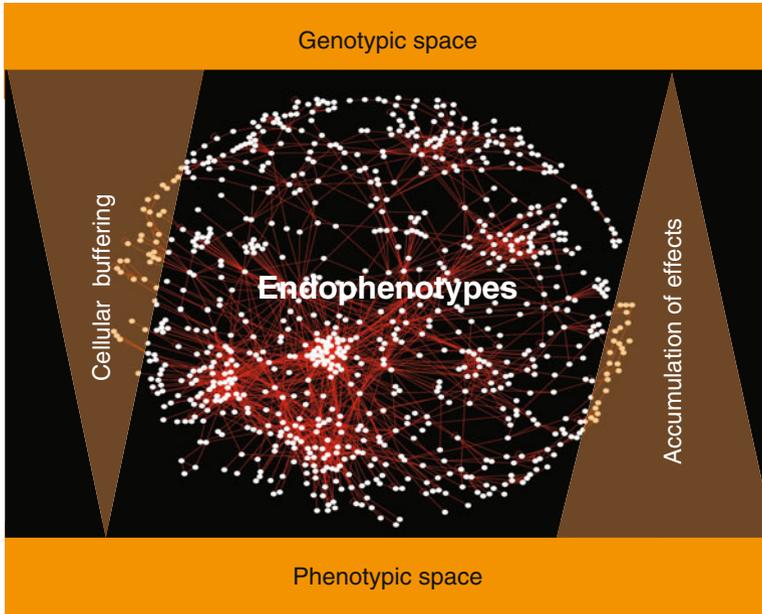


Fig. 17.9 The mapping of phenotypic traits on genotypic traits must consider the different layers of endophenotypes in order to determine how cellular networks buffer genetic variation from one level to the next (phenotypic variation decreases as we move from the genotype to organismal phenotypes) and how the effects found at one level may influence higher levels (accumulation of effects and increase phenotypic variation). Elucidating these mechanisms will allow to understand the context dependence of allelic effects in quantitative genetics

studies show that the relationship between alleles and phenotypes greatly depend on the organization of cellular networks and exemplify the power of combining systems biology approaches with natural genetic variation. The exact mechanisms by which variation at one level is buffered by other levels of cellular organization remain to be examined but we suggest that common mechanisms and rules will emerge as more investigations are performed. As we move from the genotype towards organismal phenotypes, there may be more opportunity for mutations to affect the trait—because each level depends on the previous one plus other factors—but this may be counterbalanced by cellular buffering mechanisms (Fig. 17.9). Systematic studies that combine organismal phenotypes with endophenotypes will be key to identifying these mechanisms and thus understanding how condition-dependent allelic effects take place.

16 Conclusion

Quantitative genetics has provided the evolutionary biology community with strong theoretical and analytical bases for the analysis of phenotypic traits. The next challenge will now be to be able to predict phenotypes from genotypes. This challenge requires a good understanding of how biological systems work, which is now made possible by systems biology, but also how natural genetic variation affects the component of this system and how the organization of these systems influences allelic effects. This can only be achieved by combining the two disciplines into integrative, evolutionary systems biology approaches.

Acknowledgements We thank two anonymous reviewers for their comments. CRL's and SAR's research on evolutionary systems biology is funded by the Human Frontier Science Program (RGY0073/2010)

References

1. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105 (27):9272–9277. doi:0803466105 [pii] 10.1073/pnas.0803466105
2. Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445(7123):82–85. doi:nature05388 [pii] 10.1038/nature05388
3. Sawyer SA, Parsch J, Zhang Z, Hartl DL (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc Natl Acad Sci USA* 104(16):6504–6510. doi:0701572104 [pii] 10.1073/pnas.0701572104
4. Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10(8):565–577. doi:nrg2612 [pii] 10.1038/nrg2612
5. Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL, Saka A, Fukuda T, Ishihara S, Oka S, Suzuki G, Watanabe M, Hirata A, Ohtani M, Sawai H, Frayssé N, Latgé J-P, François JM, Aebi M, Tanaka S, Muramatsu S, Araki H, Sonoike K, Nogami S, Morishita S (2005) High-dimensional and large-scale phenotyping of yeast mutants. *Proc Nat Acad Sci USA* 102(52):19015–19020
6. Nogami S, Ohya Y, Yvert G (2007) Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS Genet* 3(2):e31–e31
7. Provine WB (2001) *The origins of theoretical population genetics*. University of Chicago Press, Chicago
8. Cheverud JM (2006) Genetic architecture of quantitative variation. In: Wolf JB, Fox CW (eds) *Evolutionary genetics: concepts and case studies*. Oxford University Press, Oxford, pp 288–309
9. Fox CW, Wolf JB (2006) *Evolutionary genetics: concepts and case studies*. Oxford University Press, Oxford
10. Roff DA (1997) *Evolutionary quantitative genetics*. Chapman & Hall, New York
11. Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121(1):185–199
12. Giurumescu CA, Sternberg PW, Asthagiri AR (2009) Predicting phenotypic diversity and the underlying quantitative molecular transitions. *PLoS Comput Biol* 5(4):e1000354–e1000354

13. Rendel JM (1967) Canalisation and gene control. Logos Press, New York
14. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, New York
15. Wagner GP, Laubichler MD, Bagheri-Chaichian H (1998) Genetic measurement of theory of epistatic effects. *Genetica* 102–103(1–6):569–580
16. Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139(3):1455–1461
17. Gjuvsland AB, Plahte E, Ådnøy T, Omholt SW (2010) Allele interaction – single locus genetics meets regulatory biology. *PLoS ONE* 5(2):e9379–e9379
18. Hansen TF, Wagner GnP (2001) Modeling genetic architecture: a multilinear theory of gene interaction. *Theor Popul Biol* 59(1):61–86
19. Rice SH (2004) Evolutionary theory: mathematical and conceptual foundations. Sinauer Associates, Sunderland
20. Bagheri HC, Wagner GnP (2004) Evolution of dominance in metabolic pathways. *Genetics* 168(3):1713–1735
21. Wagner A (2008) Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet* 9(12):965–974. doi:nrg2473 [pii] 10.1038/nrg2473
22. Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, Hopkins N (2004) Identification of 315 genes essential for early zebrafish development. *Proc Natl Acad Sci USA* 101(35):12792–12797. doi:10.1073/pnas.0403929101 0403929101 [pii]
23. Dudley AM, Janse DM, Tanay A, Shamir R, Church GM (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol Syst Biol* 1:2005 0001. doi:msb4100004 [pii] 10.1038/msb4100004
24. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibzadeh S, Papp B, Pal C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C (2010) The genetic landscape of a cell. *Science* 327(5964):425–431. doi:327/5964/425 [pii] 10.1126/science.1180823
25. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402(6761 Suppl):C47–C52. doi:10.1038/35011540
26. Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8(6):450–461
27. Coudreuse D, Nurse P (2010) Driving the cell cycle with a minimal CDK control network. *Nature* 468(7327):1074–1079. doi:nature09543 [pii] 10.1038/nature09543
28. O’Shaughnessy EC, Palani S, Collins JJ, Sarkar CA (2011) Tunable signal processing in synthetic MAP kinase cascades. *Cell* 144(1):119–131. doi:S0092-8674(10)01432-7 [pii] 10.1016/j.cell.2010.12.014
29. Lim WA (2010) Designing customized cell signalling circuits. *Nat Rev Mol Cell Biol* 11(6):393–403. doi:nrm2904 [pii] 10.1038/nrm2904
30. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, Malitskaya Y, Vogel J, Bussey H, Michnick SW (2008) An in vivo map of the yeast protein interactome. *Science* 320(5882):1465–1470. doi:1153878 [pii] 10.1126/science.1153878
31. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadian V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O’Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440(7084):637–643. doi:nature04670 [pii] 10.1038/nature04670

32. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446(7137):806–810. doi:nature05649 [pii] 10.1038/nature05649
33. Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol* 7(7):R63–R63
34. Segre D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nat Genet* 37(1):77–83. doi:ng1489 [pii] 10.1038/ng1489
35. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297(5586):1551–1555. doi:10.1126/science.1073374 297/5586/1551 [pii]
36. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34(2):166–176. doi:10.1038/ng1165 ng1165 [pii]
37. Fraser HB, Plotkin JB (2007) Using protein complexes to predict phenotypic effects of gene mutation. *Genome Biol* 8(11):R252. doi:gb-2007-8-11-r252 [pii] 10.1186/gb-2007-8-11-r252
38. DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. *Nat Genet* 40(5):676–681. doi:ng.123 [pii] 10.1038/ng.123
39. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B (2006) Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 21(3):319–330. doi:S1097-2765(05)01853-8 [pii] 10.1016/j.molcel.2005.12.011
40. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429):901–906. doi:7737 [pii]
41. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK, Boone C (2010) Genotype to phenotype: a complex problem. *Science* 328(5977):469. doi:328/5977/469 [pii] 10.1126/science.1189015
42. Brown KM, Landry CR, Hartl DL, Cavalieri D (2008) Cascading transcriptional effects of a naturally occurring frameshift mutation in *Saccharomyces cerevisiae*. *Mol Ecol* 17(12):2985–2997. doi:MEC3765 [pii] 10.1111/j.1365-294X.2008.03765.x
43. Friedman A, Perrimon N (2007) Genetic screening for signal transduction in the era of network biology. *Cell* 128(2):225–231. doi:S0092-8674(07)00063-3 [pii] 10.1016/j.cell.2007.01.007
44. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, Ahn J, Dewar-Darch D, Reguly T, Tang X, Almeida R, Qin ZS, Pawson T, Gingras AC, Nesvizhskii AI, Tyers M (2010) A global protein kinase and phosphatase interaction network in yeast. *Science* 328(5981):1043–1046. doi:328/5981/1043 [pii] 10.1126/science.1176495
45. Levy ED, Landry CR, Michnick SW (2010) Cell signaling. Signaling through cooperation. *Science* 328(5981):983–984. doi:328/5981/983 [pii] 10.1126/science.1190993

46. Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, Gerrits B, Picotti P, Lam H, Vitek O, Brusniak MY, Roschitzki B, Zhang C, Shokat KM, Schlapbach R, Colman-Lerner A, Nolan GP, Nesvizhskii AI, Peter M, Loewith R, von Mering C, Aebersold R (2010) Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal* 3(153):rs4. doi:3/153/rs4 [pii] 10.1126/scisignal.2001182
47. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) genetic properties influencing the evolvability of gene expression. *Science* 317(5834):118–121
48. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 40(7):854–861. doi:ng.167 [pii] 10.1038/ng.167
49. Jelier R, Semple JI, Garcia-Verdugo R, Lehner B (2011) Predicting phenotypic variation in yeast from individual genome sequences. *Nat Genet* 43(12):1270–1274. doi:10.1038/ng.1007 ng.1007 [pii]
50. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296(5568):752–755
51. Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrman S, Francois J, Yvert G (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet* 4(4):e1000049. doi:10.1371/journal.pgen.1000049
52. Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M (2010) Genetic analysis of variation in transcription factor binding in yeast. *Nature* 464(7292):1187–1191. doi:nature08934 [pii] 10.1038/nature08934
53. Ehrenreich IM, Gerke JP, Kruglyak L (2009) Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BYxRM cross. *Cold Spring Harb Symp Quant Biol* 74:145–153. doi:sqb.2009.74.013 [pii] 10.1101/sqb.2009.74.013
54. Malleshaiah MK, Shahrezaei V, Swain PS, Michnick SW (2010) The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. *Nature* 465(7294):101–105
55. Freschi L, Courcelles M, Thibault P, Michnick SW, Landry CR (2011) Phosphorylation network rewiring by gene duplication. *Mol Syst Biol* 7:504. doi:10.1038/msb.2011.43 msb201143 [pii]
56. Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A (2011) Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol* 9(9):e1001144–e1001144
57. Tirosh I, Reikhav S, Sigal N, Assia Y, Barkai N (2010) Chromatin regulators as capacitors of inter-species variations in gene expression. *Mol Syst Biol* 6:435. doi:10.1038/msb.2010.84
58. Landry CR, Rifkin SA (2010) Chromatin regulators shape the genotype-phenotype map. *Mol Syst Biol* 6:434. doi:msb201097 [pii] 10.1038/msb.2010.97