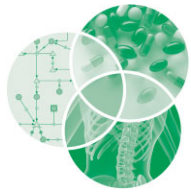
Advanced Review



Evolution of transcriptional regulatory networks in yeast populations

Duccio Cavalieri*

Saccharomyces cerevisiae is the most thoroughly studied eukaryote at the cellular, molecular, and genetic level. Recent boost in whole-genome sequencing, array-based allelic variation mapping, and genome-wide transcriptional profiling have unprecedentedly advanced knowledge on cell biology and evolution of this organism. It is now possible to investigate how evolution shapes the functional architecture of yeast genomes and how this architecture relates to the evolution of the regulatory networks controlling the expression of genes that make up an organism. A survey of the information on genetic and whole-genome expression variations in yeast populations shows that a significant score of gene expression variation is dependent on genotype-by-environment interaction. In some cases, large *trans* effects are the result of mutations in the promoters of key master regulator genes. Yet *trans*-variation in environmental sensor proteins appears to explain the majority of the expression patterns differentiating strains in natural populations. The challenge is now to use this information to model how individual genetic polymorphisms interact in a condition-dependent fashion to produce phenotypic change. In this study, we show how fruitful application of systems biology to the progress of science and medicine requires the use of evolution as a lens to reconstruct the hierarchical structure of regulation of biological systems. The lessons learned in yeast can be of paramount importance in advancing the application of genomics and systems biology to emerging fields including personalized medicine. *WIREs Syst Biol Med* 2010 2 324–335

MODELING YEAST TRANSCRIPTIONAL REGULATORY NETWORKS

The goal of systems biology is to obtain a superior level of knowledge of how function arises in dynamic interactions, representing the dependencies between genes and molecules in cellular pathways as a mathematical model.[1] Bottom-up systems biology relies on the availability of *a priori* knowledge on biological pathways. A dynamic mathematical model identifies co-regulated modules by mapping biological knowledge of reaction kinetics on metabolic maps.[2,3] Modeling approaches allow to make predictions on the outcome of the system starting from a set of initial conditions that ultimately need to be proved with specifically tailored experiments. Fitness measurements on the yeast deletion collection integrated with mathematical modeling of flux balance using approaches such as metabolic control analysis (MCA) have successfully been used to identify high flux genes in metabolic networks.[4]

Although systems biology does not necessarily require analysis of large 'omics' datasets, the wealth of public genomic data represents an extraordinary resource for deciphering the core signaling networks of the cell. The so-called top-down systems biology approaches uses whole-genome data collected from a system exposed to environmental or genetic perturbations. Performing the analyses in a chemostat allows the scientist to selectively change a set of conditions, leaving others untouched, so to analyze one variable at a time. This approach aims at reconstructing modules associated to the response without any preconceptions.

*Correspondence to: duccio.cavalieri@unifi.it

Department of Pharmacology, University of Florence, Florence, Italy

DOI: 10.1002/wsbm.68

Transcriptional regulatory networks are conceptually central to the integration of large datasets with the aim of generating robust biological hypotheses.

In order to experimentally investigate the structure governing transcriptional regulation, it is necessary to integrate measurements on gene expression, binding of transcription factors (TFs) to promoters, variation in promoter sequence, and information on protein–protein interaction (PPI) networks. This integration requires generation of 'multi-omics' datasets measuring at genomic level information on the transcriptome, proteome, and interactome. The ability to capture the regulatory layer of the biological system can hence be used to predict the behavior of the system in a condition-dependent manner. Elements of the transcriptional regulatory networks are classically the transcription regulatory proteins (TFs) and the sequences they bind upstream from the controlled genes (promoters).

The majority of the 'multi-omics' datasets have been so far produced in cells of the laboratory strain (S288c) of the baker's yeast *Saccharomyces cerevisiae*, the most thoroughly studied eukaryote at the cellular, molecular, and genetic levels. Yeast provides public access to different types of high-throughput experimental data,[5,6] including the complete PPI network, a resource that has proved instrumental in identifying co-regulated modules.[7,8] This well controlled genetic system allowed to study the transcriptional response following exposure to environmental (expression studies comparing treated versus untreated) perturbations. An idea of the amount of transcriptional information available can be obtained by visiting the *Saccharomyces* genome database (SGD) (http://www.yeastgenome.org/), mining the compendium of 2000 genome-wide gene expression experiments developed by Tanay and colleagues.[9] Using the term *cerevisiae* as query in gene expression omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) or in Arrayexpress (http://www.ebi.ac.uk/microarray-as/ae/) returns more than 650 experiments involving the baker's yeast for a total of more than 10,000 microarray hybridizations. S288c and its derivatives are the only system where TF-binding sites have been predicted computationally and tested experimentally on a whole-genome scale.[10,11] Regulatory networks have been reconstructed in yeast using transcription data identifying motif combinations regulating expression patterns during the cell cycle, sporulation, and various stress responses.[12] A number of *in silico* methods have been developed to link regulatory motifs with the TFs that bind them, eventually clustering genes by using TFs as seeds to group together genes whose expression patterns correlate with that of a particular TF.[13] Bi-clustering approaches integrating genome-wide location of putative TFs binding sites and expression data have proved useful for discovering networks of regulatory modules,[14–17] while co-regulated modules can be used to infer novel pathways.

These studies rapidly evolved from methods for extraction of biologically meaningful knowledge from microarray data to construction of models. The availability of large data collections in a well controlled genetic system allowed scientists to rapidly expose models to experimental proof and made *S. cerevisiae* the cradle of systems biology.

The ability to infer regulation of metabolic fluxes from transcriptomics data in yeast is made possible by the observation that metabolic control is strictly linked to transcriptional control. Modeling the energy constraints on the evolution of gene expression[18] demonstrated that even when considered separately, mRNA and protein synthesis can increase on average by no more than 10% without causing significant energy costs. The constraint between mRNA expression and protein levels increases under limiting conditions.

The contribution of transcriptional regulation to flux balance has been modeled for several fundamental processes in yeast cells of the reference laboratory strain S288c.

The interplay between signal transduction through protein phosphorylation along the MAPK signal transduction cascade and gene expression regulation have been described in a model on the role of MAP kinase Hog1 in regulating signal transduction in the yeast high osmolarity glycerol (HOG) pathway.[19]

Workman and collaborators[20] used an interaction model, based on protein–DNA interactions, for the integrative analysis of transcriptional data and CHIP-based genome-wide analysis of the binding of 30 TFs involved in DNA repair. This analysis reconstructed the causality between promoter binding of a TF and change in transcription for genes of the DNA-damage repair pathways.

Application of modeling to the binding of specific TFs involved in regulation of yeast cell cycle and stress response confirmed known transcriptional regulations and predicted novel TF–gene interactions.[21] Kim and O'Shea[22] also faced the challenge of developing quantitative, predictive models of gene regulation, using the PHO5 promoter to model the effects of differential affinity and accessibility of TF-binding sites (TFBS). The authors quantified the relationship between TF input and gene expression output, termed the gene-regulation function (GRF), capturing

variable interactions among TFs, nucleosomes, and the promoter in a way that faithfully reproduced the observed quantitative changes in the GRF in increasing the diversity of gene expression profiles.

A very recent work addressed the decision-making processes deciding the temporal scale of utilization of potential routes within a cellular network.[23] The authors applied a framework of 'activity motifs' to the analysis of yeast cell metabolism. The yeast cell uses different timing activity motifs to optimize transcription upon metabolic switches: forward activation to produce metabolic compounds efficiently, backward shutoff to rapidly stop production of a detrimental product and synchronized activation for co-production of metabolites required for the same reaction. The measurements of protein abundance and RNA levels over time revealed that mRNA timing motifs also occur at the protein level. The authors discovered that timing motifs significantly overlap with binding activity motifs, where genes in a linear chain have ordered binding affinity to a TF. This order-generating process seems to regulate adaptation to an ever changing environment. The results suggest that measuring gene expression is a relatively precise indication of changes in TF-binding and is a good measurement of transcriptional regulation.

Overall the results from this plethora of studies have generated highly connected motif-association maps showing global views of transcription networks, suggesting that a small number of master regulators are responsible for adaptive expression patterns in diverse conditions. Yet a large number of expression changes cannot be attributed to direct binding of a TF to a promoter and chromatin structure. In addition, cooperation between many TFs seems to play an important role.[24–26]

The results obtained on laboratory strains of *S. cerevisiae* served as testing ground for systems biology[27] including its potential application for the reconstruction of conserved reactions in the mammalian cell.[28–30] Yet before being possibly generalized to mammals, it is important to evaluate the evolutionary robustness of many of the concepts on the regulation of transcriptional networks.

Information on which elements of the network are more prone to show sequence or expression variation between strains of the same species is crucial to assess plasticity and evolvability of the system.

It is likely that the knowledge achieved in the yeast model about master regulators of transcriptional and functional plasticity will be exportable to the mammalian system, and will prove important to model transcriptional regulatory networks.

THE NATURE OF TRANSCRIPTIONAL REGULATORY VARIATION

Building on the results obtained on the laboratory strain, yeast researchers have recently investigated the relative contributions of transcriptional variation to generation of phenotypes, and diversity among isolates of the species *S. cerevisiae*.

According to the Darwinian theory of evolution, the accumulation of genetic mutations leads to generation of diversity.[31] Natural selection acts on the genotype of organisms indirectly through their phenotype and explains the adaptive features arising in a population as a result of natural processes.[32] Consequently to understand the phenotypic effects of mutations, these must be observed in the context of the environmental factors acting as selective forces on a given species or cell type.[33]

Regulatory evolution involves genomic variation which directly or indirectly influences gene expression. Divergence, temporal regulation, and different topology of gene expression are central to the evolutionary process.[34,35] When discussing the causes of strain-specific variation in regulatory circuits supporting differences in gene expression, it is important to dissect the effects of non-coding variation, causing differences in regulatory sequences, promoters, or chromatin remodeling sequences (*cis*), compared to the effects of genetic mutations in coding sequences (*trans*).[36,37]

Promoters are ideal candidates for generation of *cis*-diversity; a single change in TF-binding site's affinity can result in alteration of multiple functional modes and consequently differences in transcription programs.[38]

Alternatively, a genetic polymorphism, resulting in changes in the properties of the encoded protein, can alter expression in another gene. These so-called *trans*-regulatory changes can modify the activity or expression of factors interacting with *cis*-regulatory sequence motifs of other genes.[39,40]

The effects of coding variation on gene expression can derive from differences in two classes of *trans*-acting master regulators, TFs, and sensor proteins responding to environmental stimuli or intracellular stimuli. Mutations in environmental sensors or members of feedback loops can possibly act indirectly on transcription by mediating the concerted action of several TFs by integrating signaling from metabolites or end products of biochemical reactions in positive or negative feedback loops.

We will now discuss how *cis*- and *trans*-acting sequence variation explain differences in gene expression between species and within individuals of the same species.

GENOME-WIDE SEQUENCE AND TRANSCRIPTIONAL VARIATION AMONG SPECIES OF THE *SACCHAROMYCES SENSU STRICTU*

One of the seminal papers investigating the bases of variation between species was published by Wittkopp et al.[40] The authors investigated the relative contribution of *cis*- versus *trans*-regulatory mechanisms to differential expression of 29 genes in interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans* and found that 28 genes showed evidence of *cis*-acting changes and 16 of these were affected by *trans*-acting changes.[40]

In a recent review, Fay and Witkopp[41] surveyed gene expression studies dealing with the extent of variability both within and between a wide range of species, concluding that published studies provide considerable evidence for adaptive changes in gene expression. However, they were unable to ascertain the relative importance of regulatory changes versus changes in protein function.[41]

In order to fully address the relative importance of *cis* versus *trans*-variation in shaping the transcriptional landscape, we need information on sequence as well as transcriptional variation in as many conditions as possible. In this regard, yeast might be able to provide a large enough amount of information.

S. cerevisiae was the first eukaryote to have its genome completely sequenced, revealing about 6200 genes located on 16 chromosomes. Fundamental information on the cross species organization of regulatory variation was obtained from the sequence of four very close relatives to *S. cerevisiae: Saccharomyces paradoxus* (5–10 M years), *Saccharomyces mikatae (*10–15 M years), *Saccharomyces kudriavzevii* (15–20 M years), and *Saccharomyces bayanus* (20 M years).[42,43] Interpreting the information on sequence variation among these species in the context of gene ontology and regulatory structure[44] revealed that, although duplicated genes rarely diverge with respect to biochemical function, they do with respect to regulatory control. According to this vision, gene duplication may drive the modularization of functional networks through functional specialization of gene expression control. The same datasets were also analyzed regarding selection on transcriptional regulatory motifs. Raijma et al.[45] used sequence information to develop a probabilistic model of the evolution of promoter regions in yeast. The model expresses explicitly the selection forces acting on TFBS in the context of a dynamic evolutionary process, and reported weak selection on most TFBS. Finding few constraints on promoters suggests that transcriptional regulation in yeast is highly dynamic. Much of the functionality of transcriptional networks is encoded in ways other than strong TFBSs, and binding sites are under continuous remodeling as a result of high levels of redundancy. The same conclusion regarding the volatility of the *cis*-regulatory structure can be derived from the work of several authors[46–48] who reported substantial divergence of TFBS across related yeast species and strains.

Recent research from the Barkai group[49–51] investigated the impact of allelic variation on gene expression measurements on a genomic scale on three of the sequenced yeast species of *S. sensu strictu*. Barkai and collaborators studied systematically the effect of changes in *cis*-regulatory elements on the divergence of gene expression comparing gene expression profiles for species where gain and loss of TFBS could be predicted. The results indicate that the divergence rate of gene expression is not correlated with that of its associated coding sequence, and that promoter divergence is not correlated with divergence in gene expression both in yeast and in mammalian expression datasets.[49] A possible reason for the lack of observed change in gene expression could be that the transcription analysis was performed under neutral conditions for expression from that binding site. To rule out this possibility, the authors performed transcriptional analysis on cells of the three species following exposure to mating pheromone, studying specifically the Ste12 TF controlling the response. The authors discovered that divergence of Ste12-binding sequence motifs could account for 30% of the transcriptional differences observed between the species, containing 50% of the STE regulated genes. The variation in expression that could not be accounted for by promoter divergence could be because of compensatory effects on neighboring binding sites, appearance of binding sites for other TFs that would buffer the effect of lack of a binding site by allowing binding of other TFs, a general possibility in case of transcription driven by coordinated function of many TFs. The authors reported that presence of TATA box was correlated with expression divergence. TATA box is present in only 24% of the genes of the genome, mainly stress response genes. Essential housekeeping genes, on the contrary, are TATA-less. This feature could increase robustness of transcriptional regulation of essential genes, preserving important phenotypes. TATA seems to be important for generating heritable transcriptional variation, maintaining evolvability of regulatory networks in response to environmental variation. In agreement with other studies, the authors found that chromatin structure could play a role in buffering the effects of promoter divergence.[52]

TATA-containing genes are more dependent on chromatin structure regulation and have a peculiar pattern of DNA bendability. Differences in promoter nucleosome occupancy, corresponding to species-specific patterns of nucleosome positioning, could account for some of the differences observed in the Ste12 response. The impact of chromatin structure on modulating the effects of sequence variation seems a general theme with much more profound effects than those so far reported.[51] The authors observed that proteins involved in mating and fertility are among the most rapidly evolving genes. Based on the overall results, it is reasonable to think that *trans* effects caused by divergence in these proteins could have as profound if not more profound effects on variation in gene expression than *cis*-promoter variation.[50]

A tremendous leap forward in understanding the forces shaping the evolution of the regulatory network was very recently obtained in studies of interspecific hybrids. Tirosh et al.[53] compared the allele-specific expression of two yeast species, *S. cerevisiae* and *S. paradoxus* and their hybrid. Expression divergence between species was generally a result of changes in *cis*. In agreement with evolutionary theory, *cis* (but not *trans*) effects were correlated with sequence divergence at promoters and regulatory elements, whereas *trans* (but not *cis*) effects were enriched with genes whose expression was altered upon deletion of transcription or chromatin regulators, reflected a differential response to the environment and explained the tendency of certain genes to diverge rapidly. A small portion of the *trans* effects (<1%) were enriched within contiguous chromosomal regions that display correlated expression divergence, possibly indicating epigenetic effects.

The authors also discovered that parental genomes accumulate a substantial fraction of compensating *cis–trans* effects, a signature of purifying selection, that are released in the hybrid and account for approximately 20% of over-dominance expression patterns. *Trans* effects discovered are primarily attributable to mutations affecting environmental sensing and transduction of sensory signals and not to mutations in TFs and account for the majority of the hybrid-specific expression. The interaction between *cis* and *trans* effects can potentially enhance expression divergence if they act in the same direction or compensate divergence if they act in opposite directions. Under complete neutrality, they should be equally favored. If natural selection acts to eliminate differences, compensating interactions would be favored. Finally, if expression changes are beneficial, then enhancing interactions would be preferred. The authors found a small but significant enrichment of compensating interactions, suggesting a role of purifying selection in the buffering of gene expression divergence. Consistent with this hypothesis, both *cis* and *trans* effects are underrepresented among essential genes.

Overall cross species variation suggests that control of transcriptional programs is less deterministic than expected and comprised of plastic networks that are continuously changing during evolution. Interspecific hybrid formation might scramble these networks and provide tremendous potential to boost evolution of transcriptional networks. As a consequence hybrid formation might be advantageous under selective environmental conditions.

EXPLORING THE GENETIC DETERMINANTS OF VARIATION IN GENE EXPRESSION IN *S. CEREVISIAE* POPULATIONS

Experiments on strains of different species are informative yet limited as different species are not interfertile and it is unusual to obtain progeny from interspecific hybrids. This limits the potential of testing genetic concepts and using differences in recombination frequencies of genetic markers and linkage studies to map quantitative traits loci (QTLs) to phenotypes, including expression phenotypes.

In the past 10 years, the application of whole-genome transcriptional analysis to different interfertile yeast isolates of the *S. cerevisiae* species has shown genome-wide rates of genetic variation in gene expression among yeast strains of different origins[54–57] and in heterozygosis within one strain.[58,59] The application of high-throughput genotyping in combination with gene expression analysis to strains showing differences in gene expression enabled discovery of the heritable determinants of variation in both single genes[54,56,59,60] and QTLs.[60–64] Most expression differences in progeny from crosses of strains of the same species[54,56] mapped to *trans*-acting loci, the most noteworthy being GPA1, a G-protein alpha subunit encoding for a pheromone receptor, AMN1, a negative regulator of the mitotic exit network in yeast and the TF HAP1, a regulator of respiratory metabolism.

Brown et al.[59] recently cloned a master regulator of gene expression variation responsive to different environmental conditions. The segregating effects on transcription of this allele were first described in the progeny of a vineyard isolate from Montalcino region by Cavalieri et al.[58] and confirmed by Landry et al.[6] The extent of variation caused by this single mutation was in puzzling parity with the variation observed among isolates by Townsend et al.[55] Interestingly,

the transcriptional effects are not caused by variation in a *cis*-regulatory region or in a TF, but by an insertion in a homonucleotide repeat within the coding sequence of the gene SSY1, which encodes a sensor of external amino acids (AAs), Ssy1p. The insertion causes a STOP codon and a truncated protein that makes the haploid derivatives of M28 'blind' to AAs in the culture environment. As a consequence, the cell represses transporters of AAs and the autarchic strain, deprived of the possibility to import AAs, and activates genes for AA biosynthesis (methionine, leucine, lysine, histidine pathways) through the action of the respective transcriptional regulators.

Interestingly, the genes responsible for the repressive feedback loops are not activated, indicating production of basal levels of AA biosynthetic enzymes without overproduction of the end product of the pathway. An example is the overexpression of the genes of the methionine pathway in the wine and lab strains carrying a truncated Ssy1. This overexpression of all the genes leading to methionine production (Figure 1) is supported by the overexpression of MET28, the TF activating gene expression of this pathway, without activation of genes leading to production of *S*-adenosyl methionine (SAM1 and SAM2), and without changes in the transcript levels of the transcriptional repressor MET30.

The cascading effects of the mutation affect as many as 400 genes in the wine yeast genetic background and approximately 200 in the laboratory strain genetic background (depending on the p-value), implying significant epistatic interactions alleviating or aggravating the effects of the mutation in different genetic backgrounds. The effects of this mutation on gene expression are totally environmentally dependent as they can be observed only in media with AAs but disappear in conditions where no AAs are supplemented in the media. It is also noteworthy that characterization of this mutation was made through the elegance of genetics and the stringent logical reasoning of a systems biology approach using pathway analysis of transcriptomic data united with profound knowledge of yeast metabolism and physiology.[65,66]

The results from Brown et al.,[59] describing the centrality of genetic variation in enzymes regulating environmental sensing, AA and carbon sources metabolism on the complexity of yeast regulatory networks, are in agreement with the observations of Zhu et al.[67] integrating allelic variation, whole-genome expression, TFBS, and PPI on a yeast population segregating from a laboratory strain (S288c) and a wine strain (RM11).[54] The authors discovered that a network integrating TFBS and PPI data was the most predictive. This Bayesian network was extremely useful in predicting causal regulators responsible for hotspots of gene expression variation. The method discovers Leu2 and ILV6, AA biosynthetic genes, linked to Leu3 and GCN4, TFs regulating AA metabolism, as central players in these networks. The result again suggests that integration of signals from specific AA regulators, with MSN2

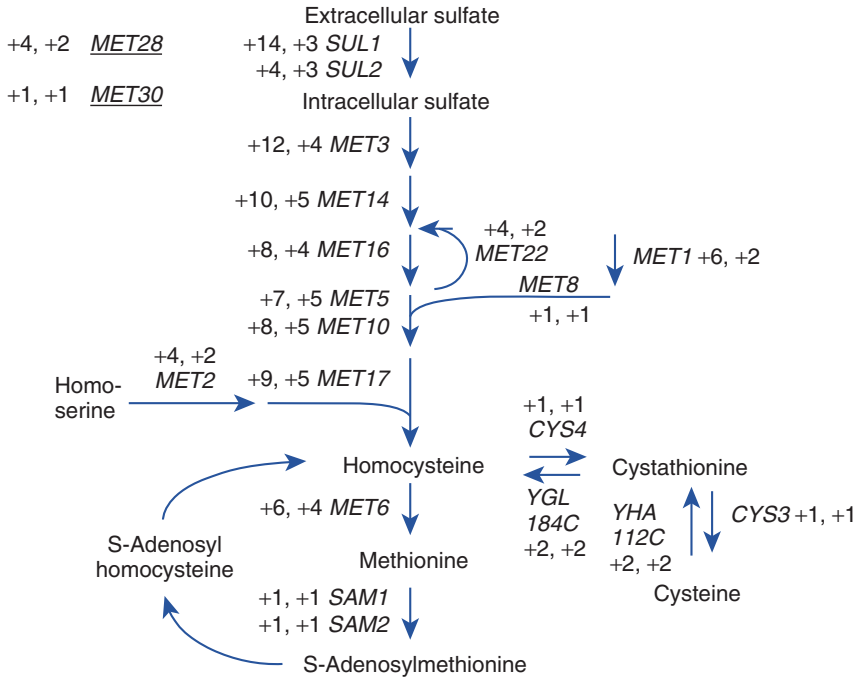


FIGURE 1 | Regulation of differential gene expression of the methionine regulatory network caused by the mutated Ssy1p described in Ref 61. The numbers next to genes express an approximation of the relative expression of mutated versus wild-type allele in the M28 (first number) and S288c (second number) genetic backgrounds over at least four replicates. Transcription factors are underlined.

and SKN7 TFs, associated to GCN4, responding to stress and starvation. One of the sub-networks was regulated by PHM7, a gene with unknown function, regulated by phosphate levels with a *cis*-polymorphism affecting differential gene expression in S288c and RM strains.[60] In addition to the *cis*-polymorphism in PHM7, Brem et al.[60] described also another *cis*-polymorphism in a sensor, GPA1 as important to regulate gene expression, and estimated *cis*-acting locus pairs to potentially influence 57% of the yeast transcripts, yet only 225 were actually identified in the segregants from the cross between RM and BY strains, varying often at low levels.

An important role of *cis*-variation was thus reported by Ronald et al.[62] that measured allele-specific expression in the By-RM hybrid for 77 genes showing strong self linkage concluding that in 52–78% of these genes local regulatory variation was acting in *cis*. These genes showed increased polymorphism in the promoter region, some in predicted TFBS. It is hard to assess whether the *cis* effect observed is as a result differences in TF binding or other effects such as chromatin structure and nucleosome positioning.

When this review was ready for submission, a new study was just published which re-affirmed the predominance in *trans*-variation in differentiating yeast strains, but reported the first example in which multiple, interacting loci encoding three TFs: IME1, RME1, and RSF1, explain natural variation in the efficiency of sporulation, and in the transcriptional program in yeast that initiates the sexual phase of the life cycle.[64] The authors discovered that the difference in sporulation efficiency observed between oak tree and vineyard strains is due to allelic variation in three loci demonstrating that selection has shaped quantitative variation in yeast sporulation between strains, again acting on protein sequence of three TFs rather than on their promoters.

The results of studies on variation between strains of the same species or in heterozygosity in one strain of the same species indicate that constrains on functional variation in *trans* seem to play a greater role than previously expected. It is noteworthy that significant *trans* variations are mapped to sensors and at a lesser extent to TFs. The wealth of genome-wide information on nucleotide variation between strains of the same species recently made available,[68,69] if associated with transcriptional and fitness data under variable environmental conditions, will soon make it possible to obtain conclusive evidence on the relative role of *cis–trans* transcriptional regulation in *S. cerevisiae* diversification, a task so far not possible in any other group of organisms.

GENETIC VARIATION FOR TRANSCRIPTIONAL PLASTICITY

The genetic determinants of transcriptional responses to environmental perturbations were first investigated using a combination of fitness analyses on the yeast barcode deletion collection[70] and transcriptional studies on selected mutants.[71]

Studies of experimental evolution of laboratory populations of yeast coupled to the ability to systematically find mutations in evolved organisms provides an opportunity to observe the evolutionary dynamics of adaptation and assess the transcriptional consequences of these mutations.[72,73] Yet, studying artificial evolution though important is not necessarily ecologically relevant.

A number of studies have demonstrated that natural strains exhibit important changes in gene expression in response to environmental perturbations, including conditions simulating natural environments (for a review, see Ref 6). Landry et al.[6,74] first demonstrated the existence of significant variation in genotype-by-environment interactions modulating the transcriptional response to environmental perturbations (transcriptional plasticity) in yeast populations. Genetic variation in transcriptional plasticity is evidenced by significant genotype-by-environment interaction. Landry and colleagues measured gene expression in a set of six independent isolates grown in four different environments chosen to mimic the stress that yeast may encounter in the wild, the harshest environment being nitrogen starvation. The authors then estimated and modeled the way yeast cells respond to environmental variation with functions and reaction norms, relating the average level of expression of a genotype across a range of environments. The level of expression of each gene was standardized to the expression in the reference laboratory strain (BY4743). For a global assessment of the variation, a self-organizing map (SOM) algorithm was used to group genes into a number of discrete reaction norms. Interestingly, an important fraction of genes showed significant genotype-by-environment interaction, meaning that there is genetic variation in transcriptional plasticity, variation which acts as a background for natural selection. When plotting the mean genetic variance in expression level among the genes increased across the environmental gradient, Landry and collaborators observed that greater variation can be observed in 'harsh' environments compared to non selective conditions, in agreement with Schmalhausen's law, which predicts that organisms will be more sensitive to variation at the boundary of their tolerance in any dimension.[33]

When testing for enrichment for specific cellular functions, Landry et al.[59,74] discovered that genetic variation for environmental plasticity was enriched in metabolic pathways such as ribosome function, purine metabolism, translation factors, beta-alanine metabolism, and arginine and proline metabolism, consistent with the nitrogen starvation selection applied, requiring mobilization and recycling from proteins and AAs. The take home message is that genetic variation in the level of gene expression among natural isolates of *S. cerevisiae* depends on the environment in which it is measured and that studies performed in single, rich environment may not necessarily extrapolate to other conditions. In selective conditions networks associated to the regulation of AA and nitrogen metabolism seem to be highly interconnected and pleiotropic, in agreement with the previously discussed results from Brown et al.[59] and Zhu et al.[67]

Smith and Kruglyak[75] also investigated gene–environment interaction. The authors employed thousands of genes as quantitative traits to discover the genetic and molecular bases of variation in gene expression between a laboratory strain (BY) and a vineyard strain (RM) using different carbon sources as environmental variables, glucose, and ethanol. The authors observed that most transcripts vary by strain and condition, with 2996, 3448, and 2037 transcripts showing significant strain, condition, and strain–condition interaction effects, respectively. Gene expression profiling of over 100 meiotic segregants from the cross between these two strains in both growth conditions identified 1555 linkages for 1382 transcripts with significant gene–environment interaction. The authors observed that linkages corresponding to polymorphisms in *cis*-regulatory elements tend to show the same effect or the same direction of effect in ethanol or/and glucose. Linkages corresponding to polymorphisms influencing *trans*-acting factors are more condition-dependent, and often show opposite effects in the two conditions. These genes can be either TFs or enzymes and the result is in perfect agreement with the findings of Landry et al.[6] A polymorphism in IRA2W locus explained the condition-dependent expression of many growth-related transcripts. This result showed that most variation is explained by polymorphisms causing different IRA2W-mediated repression of Ras/PKA signaling.

We can conclude that both studies indicate a prevalent role of signal transduction modules acting in *trans* in regulating the phenotypic (transcriptional) plasticity of the environmental response.

FROM NATURAL VARIATION TO EXPERIMENTAL EVOLUTION OF REGULATORY NETWORKS

We have so far discussed results based primarily on analysis of natural existing variation. Experimental evolution is also important, although beyond the scope of this review.[72,76] Nevertheless, measuring the effects of spontaneous mutations without the confounding effect of natural selection recently allowed Landry and collaborators[77] to identify the properties shaping neutral evolution of gene expression.

The effects of naturally occurring mutations on gene expression without natural selection demonstrate that gene expression sensitivity to mutations is enhanced by increasing *trans*-mutational target size and the presence of a TATA box.[77,78] This finding is in agreement with studies on variation between strains of different species, assigning a central role in shaping transcriptional variation to TATA boxes.[51]

Altogether the results of Landry et al.[77] and Tirosh and Barkai[51] conclude that there is a robust lack of correlation between coding sequence and expression divergence, essential genes diverge less in expression, and neutral drift with purifying selection is the driving force in gene expression divergence in yeast.

The observation that genes with greater sensitivity to mutations are also more sensitive to systematic environmental perturbations and stochastic noise indicates that *cis*-acting sequences, and in particular TATA boxes, are a reservoir of variation which has the potential to contribute to gene expression evolvability.

CONCLUSION

Since 1969,[79] regulatory divergence has been proposed as a driving force for genetic variation, and several studies indicate that protein divergence is insufficient to account for the extensive phenotypic differences observed between species, again stressing the relevance of gene regulation.[41] In our review, we critically reassess this statement showing that intraspecific adaptive transcriptional variation could result from few segregating polymorphisms with manifold pleiotropic effects rather than from numerous mutations affecting the regulation of individual genes. The idea that most *trans* effects reflect differences in sensory signals is supported now by an increasing number of studies in which variations in expression of multiple genes does not map to transcription regulators but to environmental sensors or signal transduction genes.

The pivotal role of sensory *trans* divergence in indirectly shaping variation in the transcriptional

landscape results from being upstream of the immediate regulators coordinately influencing several TFs and consequently a large number of genes. This means that a broadening of the elements included in models of gene expression variation might be necessary.

Cis-acting variation seems to play a fundamental role in fine-tuning *trans* effects and in explaining interspecific expression variation. Studies on gene expression variation in interspecific hybrids indicate variation in promoter regions as key to species-specific expression differences. A possible explanation is that mutations in environmental sensors could initially drive specialization and occupation of specific environmental niches. Divergent evolution could have then isolated species-specific regulatory circuits through specialization of promoter regions. The balance between buffering and enhancing effects on expression divergence resulting from the environmental dependent interactions between *cis* and *trans*-acting mutations is the force that shapes population genetic variation in gene expression. In our opinion, the interpretation of the effects of transcriptional variation in the context of the evolutionary forces present in the environment is essential to the assessment of the proportion of expression divergence attributable to natural selection. In future, gene expression measurements should be associated to fitness measurements in a much wider number of environmental conditions.

Although the adaptive significance of a large extent of this variation remains to be determined, the organization of the variation into specific pathways and biological functions should help identify its molecular basis and ecological significance.

The promise of the investigation of natural variation in gene expression is based on the idea that nothing makes sense in biology unless interpreted through the lens of evolution and nothing can be meaningfully modeled in systems biology except from an environmental and evolutionary perspective.

The knowledge obtained in the yeast model on how individual genetic polymorphisms interact in a condition-dependent fashion to produce phenotypic change could advance the application of genomics and systems biology to emerging fields including personalized medicine.

Jacob Monod [quotation from François Jacob *DNA Interactive DVD*, 2003 (http://www.dnai.org/)] paradigmatically described how the evolution of scientific common sense from the 1950s to today has reversed the belief that 'Cows had cow molecules and goats had goat molecules and snakes had snake molecules' as more molecules of various organisms are resulting very similar. Two-hundred years from Darwin's birthday, it is becoming increasingly clear that not only molecules but also regulatory programs might share more common themes than expected. Yet, understanding which networks are conserved and which show fundamental differences is increasingly important. The incorporation of information on robustness of a regulatory circuit and on its plasticity in systems biology models will be crucial to understanding to what extent information obtained for fundamental processes in a model can be exported to evolutionarily distant organisms. The possibility of extrapolating meaningful information obtained in yeast to model processes in the mammalian cell is dependent on the extent of conservation of the process among strains of the same species and of evolutionarily similar species.

REFERENCES

1. Bruggeman FJ, Westerhoff HV. The nature of systems biology. *Trends Microbiol* 2007, 15:45–50.
2. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 2005, 102:2685–2689.
3. Oliver SG. From genomes to systems: the path with yeast. *Philos Trans R Soc Lond B Biol Sci* 2006, 361:477–482.
4. Delneri D, Hoyle DC, Gkargkas K, Cross EJ, Rash B, et al. Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat Genet* 2008, 40:113–117.
5. Liti G, Louis EJ. Yeast evolution and comparative genomics. *Annu Rev Microbiol* 2005, 59:135–153.
6. Landry CR, Townsend JP, Hartl DL, Cavalieri D. Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol* 2006, 15:575–591.
7. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001, 292:929–934.
8. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 2002, 18(Suppl 1):S233–S240.

9. Tanay A, Steinfeld I, Kupiec M, Shamir R. Integrative analysis of genome-wide experiments in the context of a large high-throughput data compendium. *Mol Syst Biol* 2005, 1:2005.0002.
10. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002, 298:799–804.
11. Bar-Joseph Z, Gerber GK, Lee TI, et al. Computational discovery of gene modules and regulatory networks. *Nat Biotechnol* 2003, 21:1337–1342.
12. Pilpel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 2001, 29:153–159.
13. Zhu Z, Pilpel Y, Church GM. Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm. *J Mol Biol* 2002, 318:71–81.
14. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 2003, 34:166–176.
15. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999, 22:281–285.
16. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, et al. Revealing modular organization in the yeast transcriptional network. *Nat Genet* 2002, 31:370–377.
17. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004, 36:1090–1098.
18. Wagner A. Energy constraints on the evolution of gene expression. *Mol Biol Evol* 2005, 22:1365–1374.
19. Klipp E, Nordlander B, Kruger R, Gennemark P, Hohmann S. Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* 2005, 23:975–982.
20. Workman CT, Mak HC, McCuine S, Tagne JB, et al. A systems approach to mapping DNA damage response pathways. *Science* 2006, 312:1054–1059.
21. Li H, Zhan M. Unraveling transcriptional regulatory programs by integrative analysis of microarray and transcription factor binding data. *Bioinformatics* 2008, 24:1874–1880.
22. Kim HD, O'Shea EK. A quantitative model of transcription factor-activated gene expression. *Nat Struct Mol Biol* 2008, 15:1192–1198.
23. Chechik G, Oh E, Rando O, Weissman J, Regev A, et al. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol* 2008, 26:1251–1259.
24. Pokholok DK, Harbison CT, Levine S, Cole M, Hann NM, et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 2005, 122:517–527.
25. Rando OJ, Ahmad K. Rules and regulation in the primary structure of chromatin. *Curr Opin Cell Biol* 2007, 19:250–256.
26. Rando OJ. Chromatin structure in the genomics era. *Trends Genet* 2007, 23:67–73.
27. Mustacchi R, Hohmann S, Nielsen J. Yeast systems biology to unravel the network of life. *Yeast* 2006, 23:227–238.
28. Hartwell LH. Yeast and cancer. *Biosci Rep* 2004, 24:523–544.
29. Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: understanding cancer using microarrays. *Nat Genet* 2005, 37(Suppl):S38–S45.
30. Petranovic D, Nielsen J. Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol* 2008, 26:584–590.
31. Darwin C. *On the Origin of Species by Means of Natural Selection*. London: John Murray; 1859.
32. Ayala FJ. Darwin's greatest discovery: design without designer. *Proc Natl Acad Sci USA* 2007, 104(Suppl 1):8567–8573.
33. Levins R. *Toward a Population Biology, Still* Cambridge: Cambridge University Press; 2004.
34. Carroll SB, Grenier JK, Weatherbee SD. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. New York: Blackwell Science; 2001.
35. Davidson EH. *Genomic Regulatory Systems: Development and Evolution*. San Diego, CA: Academic Press; 2001.
36. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science* 1975, 188:107–116.
37. Wray GA. Transcriptional regulation and the evolution of development. *Int J Dev Biol* 2003, 47:675–684.
38. Tanay A, Gat-Viks I, Shamir R. A global view of the selection forces in the evolution of yeast *cis*-regulation. *Genome Res* 2004, 14:829–834.
39. Wray GA. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet* 2007, 8:206–216.
40. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 2004, 430:85–88.
41. Fay JC, Wittkopp PJ. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 2008, 100:191–199.
42. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, et al. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* 2003, 301:71–76.
43. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 2003, 423:241–254.

44. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* 2007, 449:54–61.
45. Raijman D, Shamir R, Tanay A. Evolution and selection in yeast promoters: analyzing the combined effect of diverse transcription factor binding sites. *PLoS Comput Biol* 2008, 4:e7.
46. Doniger SW, Fay JC. Frequent gain and loss of functional transcription factor binding sites. *PLoS Comput Biol* 2007, 3:e99.
47. Borneman AR, Gianoulis TA, Zhang ZD, et al. Divergence of transcription factor binding sites across related yeast species. *Science* 2007, 317:815–819.
48. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, et al. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* 2008, 4:e1000183.
49. Tirosh I, Weinberger A, Bezalel D, Kaganovich M, Barkai N. On the relation between promoter divergence and gene expression evolution. *Mol Syst Biol* 2008, 4:159.
50. Tirosh I, Barkai N. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet* 2008, 24:109–113.
51. Tirosh I, Barkai N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* 2008, 18:1084–1091.
52. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell* 2007, 128:707–719.
53. Tirosh I, Reikhav S, Levy AA, Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 2009, 324:659–662.
54. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002, 296:752–755.
55. Townsend JP, Cavalieri D, Hartl DL. Population genetic variation in genome-wide gene expression. *Mol Biol Evol* 2003, 20:955–963.
56. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, et al. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 2003, 35:57–64.
57. Fay JC, McCullough HL, Sniegowski PD, Eisen MB. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol* 2004, 5:R26.
58. Cavalieri D, Townsend JP, Hartl DL. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc Natl Acad Sci USA* 2000, 97:12369–12374.
59. Brown KM, Landry CR, Hartl DL, Cavalieri D. Cascading transcriptional effects of a naturally occurring frameshift mutation in *Saccharomyces cerevisiae*. *Mol Ecol* 2008, 17:2985–2997.
60. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci USA* 2005, 102:1572–1577.
61. Storey JD, Akey JM, Kruglyak L. Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol* 2005, 3:e267.
62. Ronald J, Brem RB, Whittle J, Kruglyak L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet* 2005, 1:e25.
63. Ronald J, Akey JM. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. *PLoS ONE* 2007, 2:e678.
64. Gerke J, Lorenz K, Cohen B. Genetic interactions between transcription factors cause natural variation in yeast. *Science* 2009, 323:498–501.
65. Townsend JP. Sleuthing the difference a nucleotide can make. *Mol Ecol* 2008, 17:2793–2795.
66. Clark TA, Townsend JP. Quantifying variation in gene expression. *Mol Ecol* 2007, 16:2613–2616.
67. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 2008, 40:854–861.
68. Liti G, Carter DM, Moses AM, Warringer J, Parts L, et al. Population genomics of domestic and wild yeasts. *Nature* 2009.
69. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 2009, 458:342–345.
70. Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. *Nat Protoc* 2007, 2:2958–2974.
71. Harrison R, Papp B, Pal C, Oliver SG, Delneri D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci USA* 2007, 104:2307–2312.
72. Gresham D, Desai MM, Tucker CM, Jeng HT, Pai DA, et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 2008, 4:e1000303.
73. Airoldi EM, Huttenhower C, Gresham D, Lu C, Caudy AA, et al. Predicting cellular growth from gene expression signatures. *PLoS Comput Biol* 2009, 5:e1000257.
74. Landry CR, Oh J, Hartl DL, Cavalieri D. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene* 2006, 366:343–351.
75. Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression. *PLoS Biol* 2008, 6:e83.
76. Dunham MJ. Synthetic ecology: a model system for cooperation. *Proc Natl Acad Sci USA* 2007, 104:1741–1742.

77. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. *Science* 2007, 317:118–121.
78. Lynch M, Sung W, Morris K, Coffey N, Landry CR, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 2008, 105:9272–9277.
79. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science* 1969, 165:349–357.

