Gene networks: how to put the function in genomics

Paul Brazhnik, Alberto de la Fuente and Pedro Mendes

An increasingly popular model of regulation is to represent networks of genes as if they directly affect each other. Although such gene networks are phenomenological because they do not explicitly represent the proteins and metabolites that mediate cell interactions, they are a logical way of describing phenomena observed with transcription profiling, such as those that occur with popular microarray technology. The ability to create gene networks from experimental data and use them to reason about their dynamics and design principles will increase our understanding of cellular function. We propose that gene networks are also a good way to describe function unequivocally, and that they could be used for genome functional annotation. Here, we review some of the concepts and methods associated with gene networks, with emphasis on their construction based on experimental data.

Published online: 02 September 2002

Paul Brazhnik
Alberto de la Fuente
Pedro Mendes*
Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA.
*e-mail: mendes@vt.edu

Experimental biology is finally catching up with theoretical biology in analyzing life at a systems level [1,2]. With the development of high-throughput genomics and, more recently, functional genomics, data on thousands of cellular species are being gathered by the gigabyte. This is a significant shift from the traditional molecular biology approach of exploring the complex networks of interacting cellular components by focusing on single molecules and reactions. Technologies to measure genome-wide differential gene expression at the level of mRNA [3–5] (see also http://www.gene-chips.com) are extremely popular and costs are coming down. Methods to profile proteins [6–8] and metabolites [9–11] are equally important, but are not yet as widespread as methods to profile gene expression. These three levels (proteins, metabolites and gene expression) will be needed for accurate descriptions of cellular biochemistry at a system, but important results can be achieved with data from transcriptomics (differential mRNA measurements). The reconstruction of gene networks from experimental (microarray) data to study their dynamics is timely and important. As mentioned by Loomis and Sternberg [12], the challenge is to link the genes and their products into functional pathways, circuits and networks.

A very successful model in biochemistry is a depiction of relationships between molecules as networks of interactions. These biochemical networks can be constructed at several levels and can represent different types of interaction. Often, 'pathways' rather then networks are referred to when one is interested in a particular series of interactions. Yet, it is important to remember that such pathways never exist in isolation and that they are part of larger networks. Thus, in systems biology the concept of the pathway might have relatively little relevance. Several biochemical networks have traditionally been considered: (1) metabolic networks represent the chemical transformations between metabolites; (2) protein networks represent protein–protein interactions, such as formation of complexes and protein modification by signaling enzymes (also known as signaling networks); and (3) gene networks represent relationships that can be established between genes, when observing how the expression level of each one affects the expression level of the others. Each of these types of network is a simplification of the complete cellular system, which we refer to as the global biochemical network to emphasize that it explicitly includes all three types of molecule (metabolites, proteins and mRNA). Adoption of those simplifications for description of specific phenomena depends largely on which cellular components were observed experimentally. Thus, when exclusively monitoring gene expression to study some phenomenon, one is limited to constructing a gene network to explain the data.

Figure 1 represents a model of a global biochemical network in which the three levels are shown explicitly as planes. In any global biochemical network, genes do not interact directly with other genes (neither do the corresponding mRNAs); instead, gene induction or repression occurs through the action of specific proteins, which are, in turn, products of certain genes. Gene expression can also be affected directly by metabolites, or through protein–metabolite complexes. However, it is often useful to abstract these actions of proteins and metabolites, and represent genes acting on other genes in a gene network (also called genetic regulatory, transcription or expression networks). This simplification of going from the global biochemical network to a gene network is akin to a projection of all interactions to the 'gene space' (Fig. 1).

Traditional molecular biology (molecular genetics in particular) might have propagated the idea that genes dictate all that goes on inside a cell. This materialized in the central 'dogma' of molecular biology, which emphasizes that proteins, and consequently metabolites, are only synthesized when genes are activated. This 'dogma' failed to acknowledge that gene expression is also influenced by the levels of protein and metabolite. Systems in which there is no feedback from proteins or metabolites to genes are called 'dictatorial' [13], but are currently only used

 0167-7799/02/$ – see front matter PII: S0167-7799(02)02053-X

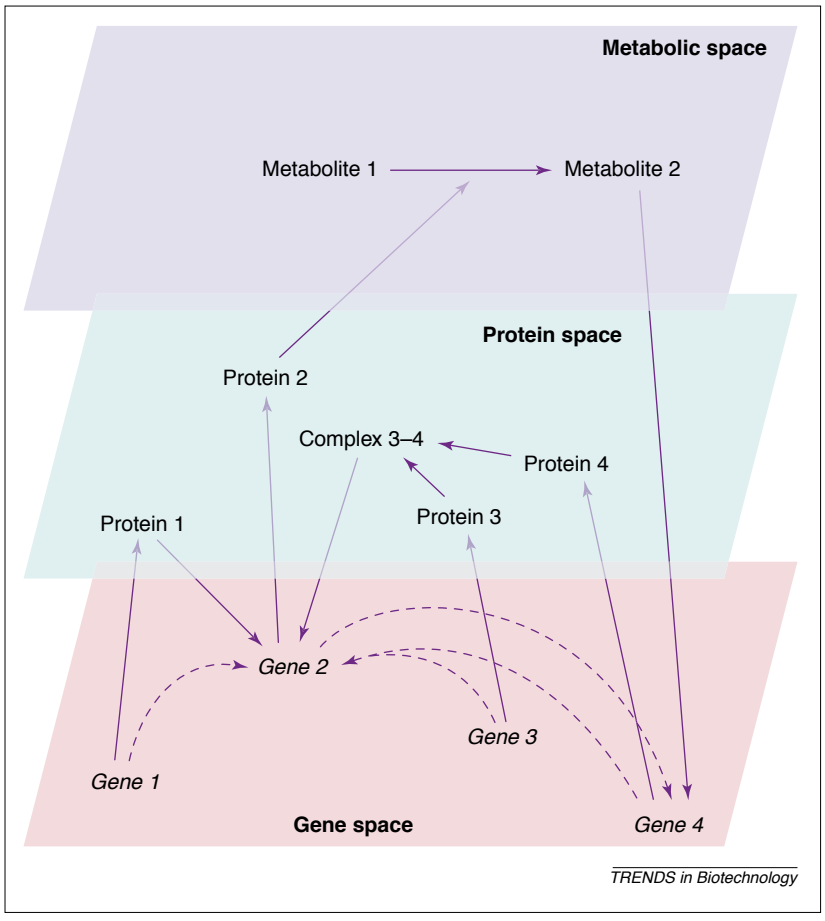

Fig. 1. An example of a biochemical network. Molecular constituents (nodes of the network) are organized in three levels (spaces): mRNAs, proteins, and metabolites. Solid arrows indicate interactions, the signs of which (activation or repression) are not specified in this diagram. Three different mechanisms of gene–gene interactions are shown: regulation of gene 2 by the protein product of the gene 1; regulation of the gene 2 by the complex 3–4 formed by the products of gene 3 and gene 4; and regulation of gene 4 by the metabolite 2, which in turn is produced by protein 2. Projections of these interactions into the 'gene space', indicated by dashed lines, constitute a corresponding gene network.

conceptually. It is now well established that regulation is distributed over all levels, and accordingly such systems are referred to as 'democratic'[13]. A recent study quantified how the control of glycolytic flux in three species of parasitic protist was partitioned between gene expression and metabolism. It was concluded that the flux is rarely regulated by gene expression alone; in a specific case, it was regulated 30% by gene expression and 70% by metabolism [14]. Although this indicates that future studies need to make more effort to monitor all three levels of regulation, it is still useful to study gene networks alone.

Why gene networks?

Increasingly, gene networks are being used as models to represent phenomena at the level of gene expression, and research on their construction from experimental data are rife. The gene network model has several applications and advantages over other approaches:

Gene networks provide a large-scale, coarse-grained view of the physiological state of an organism at the mRNA level. The mRNA phenotype can be a very important representation of cell function, offering much more precise description than can be achieved with words [15], even when these words are part of a controlled vocabulary, such as the Gene Ontology™ [16]. For instance, if the gene of a certain protein kinase is linked to genes involved in synthesis of a flagellum, one could conclude that it has a role on the chemotaxis signal transduction pathway. In this sense, not only are gene networks (and especially their graphical representations) capable of describing a large number of interactions in a concise way, but also they might represent the dynamic properties underlying those interactions at a systems level. Cells exhibit complex interacting behavior that is usually not predictable from the properties of individual system components alone. Gene networks provide such a system view at the level of gene activities. We propose that gene networks should be used for describing functions, and thus become a sophisticated means for annotation of genomics and functional genomics data.

The detailed molecular mechanisms of how the products of one gene affect expression of another gene are often unknown but the effect itself can be easily observed in gene-expression experiments. It is therefore appropriate and timely to use genome-wide gene-expression data to identify gene networks, an important step towards uncovering the complete biochemical networks of cells. Research focused on developing methods for this identification of gene networks from microarray data are now an important part of bioinformatics.

Knowledge about gene networks might provide valuable clues and lead to new ideas for treating complex diseases. It will aid pharmaceutical research in prioritizing targets, tailoring drug therapy to the individual needs of each patient [17], and can form the basis for rational gene therapy.

Cellular responses and actions are often a result of coordinated activity of a group of genes. Gene networks might allow genes to be ranked according to their importance in controlling and regulating cellular events. There is a growing indication that most single-gene mutations do not have marked phenotypes (most genes in genomes are not of 'known function'). Most phenotypes are the result of a collective response of a group of genes. Gene networks help rationalize how these complex traits arise and which groups of genes are responsible for them.

Recent estimates on the number of genes in the human genome [18] suggest that it is only about twice that of the *Caenorhabditis elegans* worm [19] (but see also [20]). There are several hypotheses to explain this relative 'simplicity' of the human genome. First, the mean number of proteins encoded by human genes could be larger than the number encoded by genes in other genomes [21]. Second, the proportion of regulatory genes (encoding signaling proteins, transcription factors, etc.) in the human genome might be higher than in other genomes. Third, the human gene network could have a higher mean number of connections per gene than do other genomes (which implies that the encoded proteins contain more binding sites). Both the second and third hypotheses could be tested by determining and comparing gene networks of various organisms. Gene networks are then also well suited for comparative genomics.

Some studies [22,23] indicate that the topology of gene networks might be largely responsible for the

robustness shown by living organisms. A particular gene network topology might have been selected during evolution to permit the type of system robustness against drastic perturbations at the genetic level that is currently observed in many species (e.g. 40% of the genes of *Saccharomyces cerevisiae* can be removed without causing noticeable phenotypes). Comparative genomics at the level of gene networks would provide a suitable test of this hypothesis.

Representations of gene networks

Gene networks are models that display causal relationships between gene activities, usually at the mRNA level, and are commonly represented by directed graphs (Fig. 2). The nodes of the graph are genes and the directed edges are causal relationships between genes. A widely adopted norm is to use arrow tips on edges to represent positive interactions, where an increase in activity of the originating gene causes an increase in the target gene, and bars on edges to represent negative interactions, where an increase in activity of the originating gene causes a decrease in activity of the target gene. Gene networks can also be represented through matrices (Eqn 1).

$$R = \begin{array}{c} \begin{matrix} G1 & G2 & G3 & G4 \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ -0.4 & -0.92 & 0.44 & 0.55 \\ 0 & 0 & -1 & 0 \\ 0 & -0.14 & 0 & -0.92 \end{pmatrix} \end{array} \begin{matrix} G1 \\ G2 \\ G3 \\ G4 \end{matrix} \qquad \text{[Eqn 1]}$$

The matrix in Eqn 1 was obtained by applying our method of regulatory strengths [24] to simulated data. Each column and row of the matrix represents one gene, and the matrix elements represent causal relationships. These matrices can be qualitative, in which positive interactions are represented with the number 1, negative interactions with the number –1 and 0 for the case of no interaction between genes. Matrices are also well suited for quantitative representations, in which case its elements take real values representing the strength and sign of the interaction. Graph representations can also express quantitative values, which are expressed with real numbers next to the edges to indicate the strength of the interaction.

An interaction between two genes is said to be direct if it does not run through any other genes in the network. For example, in Fig. 2, gene 1 directly affects gene 2. Gene 1 also affects gene 4, but only in an indirect way because the effect has to run through gene 2. Non-additive interactions are those that require the simultaneous action of two or more genes (i.e. when each of them alone has no effect and only together do they become a cause). An example is the effect of genes 3 and 4 on gene 2 in Fig. 1. Non-additive interactions are not easily captured in any of the two representations above. The graph representation needs to be generalized to hyper-graphs, in which edges can connect more than two nodes: edges originate from all the cause genes and end up in the effect gene (Fig. 1). The matrix representation would equally have to be extended to a higher-order object – 3D or even higher dimensionality matrices. Given that 3D matrices are already very cumbersome to draw and require large amounts of computer memory, this representation is not the most appropriate when it is important to include non-additive effects.

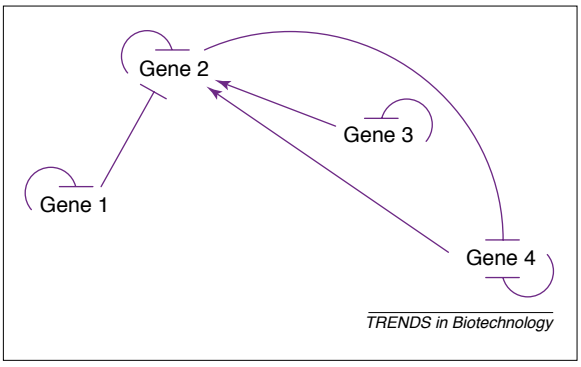

Fig. 2. Graph representation of the gene network corresponding to the biochemical network in Fig. 1. Lines show direct effects, with arrows standing for activation, and bars for inhibition. The edges implicitly include the effects of the proteome and metabolome as shown in Fig. 1. Most genes in gene networks will have a negative effect on their own concentration because the degradation rate of their mRNA is proportional to its concentration.

Inferring gene networks from experimental data

Research in gene networks has been geared towards two major goals: first, to understand the dynamics and design principles of gene regulation; and second, to reverse engineer gene networks from experimental measurements. Activities started with the pioneering work of Kauffman [25] and then Thomas [26] on random Boolean (gene) networks. More recently, the assumption that the topology of gene networks is random has been called into question, as more convincing arguments indicate that gene networks follow a 'small-world' topology with a power law distribution for node connectivities [27,28]. Research into the dynamics and structure of gene networks is still active and fruitful. However, in the past five years or so, the majority of research in gene networks has focused on methodologies for reconstructing gene networks from experimental observations, perhaps owing to the abundance of microarray data.

Many interactions between genes have been discovered through traditional molecular biology approaches. Gene networks can be obtained by combining knowledge about these interactions. The GeNet database [29] (http://www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm) is a convenient electronic repository for such information. Ideker *et al.* [30] constructed a gene network of 348 genes of *S. cerevisiae* based on information of 2709 protein–protein interactions [31,32] combined with 317 known protein–DNA interactions collected from the databases TRANSFAC [33] (http://transfac.gbf.de/TRANSFAC) and SCPD [34] (http://cgsigma.cshl.org/jian/). In a similar way, a network of

ten genes was proposed for the control of flower morphogenesis in *Arabidopsis thaliana* [35].

Experimental data of mRNA levels obtained with the use of high-throughput technologies are now abundant. These are snapshots of the molecular state of cell populations at the transcript level and are rich in information about gene networks. It seems logical that these data are the best to uncover gene networks, and indeed this strategy is presently the most widely adopted, with several methods available for this purpose. This process of establishing cause–effect relationships between genes on the basis of observed expression levels is referred to as 'reverse engineering' and is a traditional inverse problem. Several approaches have been proposed for inferring gene networks from experimental data. These have already been reviewed in some depth [36,37], but it is useful to repeat here their main characteristics.

A popular method used for gene-expression data analysis, sometimes called 'guilt by association', assumes that genes with similar expression patterns are functionally related to each other. These associations are usually explored with the use of clustering algorithms [38] and principal component analysis [39]. Although in widespread use, this method is not really appropriate to uncover gene networks. It might work when the underlying networks are modular (i.e. with small number of connections). However, it would provide ambiguous results when applied to heavily connected networks, and is therefore not generally useful for this purpose. Other methods are based on more sophisticated statistical analysis [40], including Bayesian belief networks [41–43] and pair-wise correlation methods [44]. Often such methods are aimed, not necessarily at providing a detailed reconstruction of the network, but rather, at the extraction of maximum of information from a set of noisy measurements. In general, measuring correlations is not sufficient to infer causality between genes.

Several methods exist that rely on the simplification of considering genes to be either expressed at a fixed rate, or not at all [25,45–47]. These methods also consider time to be a discrete process, and have rules that govern whether genes are on or off at a given time step, based on the values taken by the genes at the previous time step. Boolean approaches suffer from their inability to capture intermediate levels of gene expression, and can easily generate spurious results owing to their discrete nature [48].

More challenging, but potentially more accurate representations of gene networks use continuous functions, in which expression levels are allowed to take any positive value. These approaches are mathematically implemented by difference or differential equations, either linear or nonlinear. In linear additive models, the expression level [49,50] of each gene transcript depends linearly on the expression level of other genes. Each interaction is characterized by one parameter that is positive for activation, negative for inhibition, or zero for no interaction. More realistic, but also more difficult approaches use nonlinear kinetics to represent the rates of transcription, such as neural network-like sigmoidal functions [51] or empirical rate laws similar to those of enzyme kinetics [2]. In both cases, nonlinear optimization methods are used to fit the model equations to the observed data. The use of nonlinear kinetics is handicapped by requiring larger amounts of data than do linear or Boolean methods, but have the advantage of much greater predictive power.

A graph theoretical approach has been proposed to analyze gene-expression data obtained from null mutants [52]. This method is promising because it uses the most abundant type of data currently available. Unfortunately, it would not possible to distinguish between different gene networks of the same class, so the most parsimonious network must be adopted [52]. However, evidence from molecular biology suggests that the underlying gene networks will not necessarily be parsimonious. In addition, this approach is only applicable to acyclic graphs, a feature common with Bayesian belief networks [41,42]. Nevertheless, there is plenty of evidence that it is common for gene networks to have circular dependencies (gene A affects gene B but gene B also affects gene A) that originate from feedback loops (Becskei and Serrano highlight the importance of feedback loops in this context [53]).

In a recent approach, the gene network determining sea urchin development was proposed [54]. The approach was to knock out single genes and measure the response of the whole network. The authors considered that the perturbation data in itself was not sufficient to distinguish between direct and indirect effects (in contrast to what was proposed by Wagner [52]). However, using previous knowledge about *cis*-regulatory elements in this genome [55], a network specified by direct effects was proposed. Genes were taken to be directly affected if they responded in an experiment in which a certain transcription factor was perturbed and they also contained the specific target sites of that transcription factor in their *cis*-regulatory elements. Similar approaches correlate gene-expression data with the DNA sequence at gene promoters, allowing the discovery of new transcription regulatory elements [40,56].

We have proposed, in a separate publication [24], pa method based on systematic perturbation of gene transcription rates and microarray measurements to infer the underlying gene networks. The method itself is based on developments from metabolic control analysis, particularly co-response analysis [57,58] and regulatory strengths [59]. Briefly, the method is capable of identifying and quantifying direct interactions between genes, requiring several experiments equal to the number of genes considered in the analysis (c.f. with most other methods described here that require more experiments).

A conclusion that arises from the descriptions in this section is that many more experiments are needed to infer gene networks with high accuracy.

Furthermore, methods that are based on specific experimental designs [24,52] are expected to perform better than are those that disregard how the data were obtained (e.g. most applications of clustering).

Connectivity of gene networks

Gene networks only describe dynamics of gene activity. However, the interactions on the proteome and metabolome levels are implicitly present in gene networks, because the dynamics of gene activity depends on them. Presence or absence of interactions in a gene network is determined by the kinetic properties of each step along the path of interaction that passes through the proteome and metabolome. If a certain step along that path is kinetically saturated, the interaction will not be revealed in the gene network. Many of the interactions will reveal themselves only in certain physiological states and so all gene networks are phenomenological models.

Connectivity of gene networks might have important functional consequences. How dense are gene networks? Several authors [25,26,51,52,60] argue that gene networks are sparsely connected. However, there are plenty of arguments that indicate the opposite. First, because genes are connected by metabolic networks, if a certain metabolite affects the rate of transcription of a certain gene, then other genes encoding enzymes that have some level of control over that metabolite will also appear to interact with the gene in question. Second, transcription of all genes depends on metabolic energy, which implies that all genes are affected, to some extent, by genes that encode energy-metabolism enzymes. Similarly, all genes can be said to depend on the genes encoding for nucleotide metabolism enzymes. Third, when (and if) RNA polymerases exist in low concentrations, interactions between any two genes could develop by competition between their polymerase-binding sites for the few polymerase molecules. In this circumstance, increased expression of one gene would result in decreased expression of all others. Similar situations can exist for genes that encode transporters of proteins and metabolites that modulate transcription. These are ongoing arguments, and further experiments are required before any conclusion be made about the magnitude and impact that these numerous interactions might have on the density of connections in a gene network.

Many interactions in gene networks might arise from non-intuitive phenomena. The exception is when the protein product of one gene affects the expression of another, in which case it is obvious that the two genes are related. Although the process of gene expression depends on transcription factors, the interactions between genes in a gene network will in many cases not depend on transcription factors at all.

Conclusion

Gene networks are collections of gene–gene regulatory relations in a genome (or a subset thereof). In contrast to metabolic networks, the focus is not on mechanisms, but simply on the existence and perhaps magnitude of interaction between two genes. Gene networks are phenomenological models of how changing activity of genes affects the activity of other genes. Gene networks are useful to rationalize phenomena in terms of how external perturbations propagate through the expression of genes. We propose that gene networks could be used to annotate functions in genomics because these networks describe in unambiguous ways what processes each gene is involved in. Taking into account the progress in gene-expression profiling, elucidating gene networks is an appropriate and timely step on the way to uncovering the complete biochemical network of a cell type. Starting from a high-level description of gene regulation in cells provided by the gene network, one could systematically add details of the mechanism of physical interaction and expand the network to include proteins and metabolites explicitly.

Acknowledgements
We thank the National Science Foundation (grant BES-0120306) and the Commonwealth of Virginia for financial support.

References

1 Kitano, H. (2000) Perspectives on systems biology. *New Gener. Comput.* 18, 199–216
2 Mendes, P. (2001) Modeling large scale biological systems from functional genomic data: parameter estimation. In *Foundations of Systems Biology* (Kitano, H., ed.), pp. 163–186, MIT Press
3 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
4 Lockhart, D.J. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675–1680
5 Noordewier, M.O. and Warren, P.V. (2001) Gene expression microarrays and the integration of biological knowledge. *Trends Biotechnol.* 19, 412–415
6 Gygi, S.P. and Aebersold, R. (2000) Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* 4, 489–494
7 Moseley, M.A. (2001) Current trends in differential expression proteomics: isotopically coded tags. *Trends Biotechnol.* 19, S10–S16
8 Zhou, H. *et al.* (2001) Solution and chip arrays in protein profiling. *Trends Biotechnol.* 19, S34–S39
9 Oliver, S.G. *et al.* (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 373–378
10 Fiehn, O. *et al.* (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161
11 Raamsdonk, L.M. *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19, 45–50
12 Loomis, W.F. and Sternberg, P.W. (1995) Genetic networks. *Science* 269, 649
13 Westerhoff, H.V. and van Workum, M. (1990) Control of DNA structure and gene expression. *Biomed. Biochim. Acta* 49, 839–853
14 ter Kuile, B.H. and Westerhoff, H.V. (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* 500, 169–171
15 Kell, D.B. and King, R.D. (2000) On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning. *Trends Biotechnol.* 18, 93–98
16 Consortium, T.G.O. (2001) Creating the gene ontology resource: design and implementation. *Genome Res.* 11, 1425–1433
17 Somogyi, R. and Greller, L.D. (2001) The dynamics of molecular networks: applications to therapeutic discovery. *Drug Discov. Today* 6, 1267–1277
18 Ewing, B. and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* 25, 232–234
19 Kim, S.K. *et al.* (2001) A gene expression map for *Caenorhabditis elegans*. *Science* 293, 2087–2092
20 Das, M. *et al.* (2001) Assessment of the total number of human transcription units. *Genomics* 77, 71–78
21 Roberts, G.C. and Smith, C.W. (2002) Alternative splicing: combinatorial output from the genome. *Curr. Opin. Chem. Biol.* 6, 375–383
22 von Dassow, G. *et al.* (2000) The segment polarity network is a robust developmental module. *Nature* 406, 188–692

23 Little, J.W. *et al.* (1999) Robustness of a gene regulatory circuit. *EMBO J.* 18, 4299–4307
24 de la Fuente, A. *et al.* (2002) Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genet.* 18, 395–398
25 Kauffman, S. (1969) Homeostasis and differentiation in random genetic control networks. *Nature* 224, 177–178
26 Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor. Biol.* 42, 563–585
27 Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
28 Wagner, A. and Fell, D.A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. Ser. B* 268, 1803–1810
29 Serov, V.N. *et al.* (1998) Graphical interface to the genetic network database GeNet. *Bioinformatics* 14, 546–547
30 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
31 Uetz, P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627
32 Schwikowski, B. *et al.* (2000) A network of protein-protein interactions in yeast. *Nat. Biotechnol.* 18, 1257–1261
33 Wingender, E. *et al.* (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* 28, 316–319
34 Zhu, J. and Zhang, M.Q. (1999) SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* 15, 607–611
35 Mendoza, L. *et al.* (1999) Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics* 15, 593–606
36 D'Haeseleer, P. *et al.* (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16, 707–726
37 de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* 9, 67–103
38 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
39 Hilsenbeck, S.G. *et al.* (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl. Cancer Inst.* 91, 453–459
40 Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285
41 Friedman, N. *et al.* (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620
42 Pe'er, D. *et al.* (2001) Inferring subnetworks from perturbed expression profiles. *Bioinformatics* 17, S215–S224
43 Imoto, S. *et al.* (2002) Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. *Pac. Symp. Biocomput.* 7, 175–186
44 Arkin, A. *et al.* (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648
45 Liang, S. *et al.* (1998) REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* 3, 18–29
46 Akutsu, T. *et al.* (1998) A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. *Genome Inform.* 9, 151–160
47 Ideker, T.E. *et al.* (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac. Symp. Biocomput.* 5, 305–316
48 Hatzimanikatis, V. and Lee, K.H. (1999) Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab. Eng.* 1, 275–281
49 D'Haeseleer, P. *et al.* (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.* 4, 41–52
50 Holter, N.S. *et al.* (2001) Dynamic modeling of gene expression data. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1693–1698
51 Wahde, M. and Hertz, J. (2000) Coarse-grained reverse engineering of genetic regulatory networks. *Biosystems* 55, 129–136
52 Wagner, A. (2001) How to reconstruct a large genetic network from *n* gene perturbations in fewer than n^2 easy steps. *Bioinformatics* 17, 1183–1197
53 Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* 405, 590–593
54 Davidson, E.H. *et al.* (2002) A genomic regulatory network for development. *Science* 295, 1669–1678
55 Yuh, C.H. *et al.* (1998) Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896–1902
56 Bussemaker, H.J. *et al.* (2001) Regulatory element detection using correlation with expression. *Nat. Genet.* 27, 167–171
57 Hofmeyr, J.H.S. *et al.* (1993) Taking enzyme kinetics out of control – putting control into regulation. *Eur. J. Biochem.* 212, 833–837
58 Hofmeyr, J.H. and Cornish-Bowden, A. (1996) Co-response analysis: a new experimental strategy for metabolic control analysis. *J. Theor. Biol.* 182, 371–380
59 Kahn, D. and Westerhoff, H.V. (1993) The regulatory strength: how to be precise about regulation and homeostasis. *Acta Biotheor.* 41, 85–96
60 Thieffry, D. and Thomas, R. (1998) Qualitative analysis of gene networks. *Pac. Symp. Biocomput.* 3, 77–88

Genetically tailored grapevines for the wine industry

Melané A. Vivier and Isak S. Pretorius

Grapevine biotechnology is one of the most promising developments in the global wine industry, which is increasingly faced with conflicting demands from markets, consumers and environmentalists. In the grapevine industries, this technology and its supporting disciplines entail the establishment of stress tolerant and disease resistant varieties of *Vitis vinifera*, with increased productivity, efficiency, sustainability and environmental friendliness, especially regarding improved pest and disease control, water use efficiency and grape quality. The implementation and successful commercialisation of genetically improved grapevine varieties will only be realized if an array of hurdles, both scientific and otherwise, can be overcome.

Published online: 06 September 2002

Economically, the grapevine constitutes the most important fruit species globally and has been linked to agricultural and religious activities in the earliest writings and chronicles. This ancient species has evolved from a bushy, sun-loving plant to a trailing climber. The grapevine has been domesticated with ease, giving rise to ~8 million hectares of intensely pruned and manicured grapevines that are typical of vineyards across the world [1].

When considering the necessity and possible impact of plant biotechnology on the wine industry, it is imperative to consider the long-term objectives of the wine industry. From the production and resources perspectives (which are the only ones discussed here), several key issues should be considered. In the end, the overriding question will always be whether a vineyard and its derived products are economically viable.

 0167-7799/02/$ – see front matter PII: S0167-7799(02)02058-9