Networks in Bioinformatics

Lenwood S. Heath
Department of Computer Science
Virginia Tech
Blacksburg, VA 24061-0106, USA
heath@vt.edu

Abstract

Networks of biological components are common in the life sciences. The annotation of such networks with experimental data and biological knowledge gives rise to a rich, though generally incomplete, semantics of a real biological phenomenon. One tool for investigating such networks is microarray technology, a modern window into patterns of gene expression in cells. Current bioinformatics work on gene expression data addresses mechanisms underlying successful responses to drought stress in plants, which naturally leads to problems involving biological networks. Some biological context precedes a discussion of these problems.

1. Introduction

Plants have evolved to cope with a variety of environmental stresses, both abiotic — drought, heat, cold, and salt — and biotic — pathogens and insects. The imposition of stress on a plant marshals defense responses at the cellular level. The exact defensive resources marshaled exhibit both common features and divergent features among the various stressors. In addition, defense mechanisms can operate collaboratively or independently under different circumstances. The Expresso project — a Next Generation Software system for microarray experiment management and data analysis — is actively pursuing functional genomic and bioinformatic approaches to investigate these defense mechanisms, particularly in the context of drought stress in loblolly pine, *Arabidopsis thaliana,* and potato. The project utilizes biological information from multiple sources — experimental data (especially gene expression data); sequence, protein, and other databases; and the biological literature — to build a new generation of biological models, called *multimodal networks*, that can represent multiple aspects of the current state of biological knowledge and of biological systems themselves: responses over time; response variation by subcellular compartments; uncertainty (our lack of complete knowledge of cell state); and the dynamic changes in biological information due to the boom in biological data and knowledge. This is a preliminary report on experience with networks within the Expresso project.

2. Some Biological Background

Chapter 1 of Waterman [45] gives a mathematician-friendly introduction to molecular biology. Textbooks with an intermediate or advanced treatment of the requisite biology include Cooper [12]; Brown [5]; Lewin [30]; Griffiths *et al.* [19]; and Buchanan *et al.* [6]. The Oxford Dictionary of Biochemistry and Molecular Biology [41] is a good reference for essential terminology.

All organisms consist of living subunits called *cells.* Plants are *eukaryotes* (as are animals), meaning their cells contain *subcellular compartments,* including a *nucleus. Deoxyribonucleic acid (DNA)* is the class of macromolecule that is the primary carrier of the inherited (genetic) information used to perpetuate the life of the cell. In eukaryotes, the nucleus of each cell contains DNA organized in some number of *chromosomes,* which is a characteristic of the species. each cell of a eukaryote carries the same genetic information, that is, identical copies of all the chromosomes of that individual organism. Each chromosome contains genetic information as a string over a 4-letter alphabet $\{A, C, G, T\}$. A *gene* is a substring of a chromosome that, for our purposes, is the instruction for a particular, basic function within a cell. The entirity of the genetic information within an organism is its *genome.*

Ribonucleic acid (RNA) is the class of macromolecule that carries information on a more transient basis. *Messenger RNA (mRNA)* is the form of RNA

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

that is used to *transcribe* (make a copy of) the encoding of a gene from the chromosome containing the gene. This mRNA (copy) then travels through the *nuclear membrane* to the intracellular space outside the nucleus and then to the *ribosome,* an organelle responsible for *translating* the information in the mRNA into a functional unit called a *protein,* a polypeptide determined by a sequence of *amino acids* encoded via the mRNA string. For our purposes, the amino acids are an alphabet of 20 distinct chemical components. It is the sequence of amino acids that determines the function of a protein. A protein may be an *enzyme,* a molecule that catalyzes a chemical reaction, or a structural component, such as a part of the nuclear membrance. Over time, proteins are *degraded* and their chemical constituents (the amino acids) are recycled for reuse. This occurs because a protein age or because its function is no longer needed. Hence, within a cell, new proteins must be constantly produced as needed. The transcription/translation process is the basis process for protein production. This is by no means the only common process within a cell. For example, numerous *metabolic pathways* are involved in essential cell processes such as bioenergetics and the biosynthesis of macromolecules. All such pathways participate in interesting biological networks.

Within a cell, at any particular time, only a subset of the genes are being transcribed. *Gene activation* is the process by which transcription of a gene is initiated. A gene undergoing transcription is said to be *expressed;* gene expression can be detected through the existence of the corresponding mRNA. The amount of transcription and the amount of mRNA varies from gene to gene among the genes that are expressed. The amount of mRNA in a cell is an indicator — albeit a weak indicator — of a need for the function of the corresponding protein. The exploitation of this indicator is one part of *functional genomics.*

3 Reactive Oxygen Species

A plant is subject to a variety of environmental stresses, including drought, heat, cold, and salt, that it must protect itself against. One reason these stresses are dangerous and destructive is that they promote the creation of *reactive oxygen species (ROS)* (among which are H_2O_2 and O_2^-) within the cell. ROS, in turn, are capable of dismantling essential proteins and other molecules within the cell. The cellular environment becomes more oxidized as a result of the presence of ROS. Hence it is imperative that the cell rapidly *reduce* ROS to harmless molecular species. Figure 1 illustrates some basics of the process by which the cell responds to a shift to a more oxidized (less reduced) state (a change in *redox* status). The cell always has some *antioxidant molecules* present to respond rapidly to increased ROS levels; these constitute the *metabolite defense.* Significant oxidative stress can overwhelm the metabolite defense, leading to a net increase of ROS levels in the cell. These increased levels are detected by various means, including receptors in the cell membrane, protein kinases, phosphatases, and redox sensitive transcription factors. These transcription factors travel to the nucleus and activate genes necessary to a sustained defense against ROS. The transcription and translation of these defense genes leads to an increase in proteins and other metabolites involved in antioxidant defense and in damage repair. If the stress continues for a sufficient time, the plant *acclimates* to the stress and devotes additional resources to being prepared for future stress events, even after the stress is removed.

Activation of stress resistance genes associated with three distinct functions can occur as a consequence of exposure to drought: (1) synthesis of molecules associated with specific resistance to drought stress, such as proline for osmotic adjustment to water stress, aquaporins for water movement across membranes, extensins, and proline-rich proteins for cell wall extensibility events; (2) activation of oxidative stress resistance processes, such as antioxidant-based mechanisms for sustained removal of ROS; and (3) removal or repair of damaged macromolecules, such as the action of molecular chaperones on denatured proteins, or the enzymatic removal of lipids that have undergone peroxidation.

Unless ROS are removed promptly, their action can cause protein unfolding, the inactivation of enzymes, DNA damage, mutagenesis, lipid peroxidation, and disruption of cell membrane function. A novel aldehyde reductase acts to remove the products of drought-mediated lipid peroxidation [37]. Heat shock proteins/molecular chaperones are important players in resistance to oxidative stress [20, 21, 46]. Molecular chaperones interact to protect against damage to macromolecules through the repair of denatured proteins or through targeting irreversibly damaged proteins to the ubiquitin/proteasome pathway.

In the case of adaptation to threats to cellular stability, such as oxidative stress, biologists know already that defense mechanisms are abundant and diverse. These mechanisms are organized in functional and signaling pathways that span space and time. One manifestation of cellular organization and plasticity in eukaryotes is subcellular specialization within compartments. Each subcellular compartment is specific with respect to its internal environment, the metabolic func-

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

tions that are enabled there, and the means by which it communicates with the rest of the cell. In the case of organelles such as the *mitochondrion* and the *chloroplast* that contain their own genome, an organellar-nuclear interdependence necessitates the exchange of signals to balance gene expression in the two compartments, and the coordination of responses to environmental changes. In the case of organelles such as the *peroxisome,* signals imposed external to the cell lead to organelle proliferation. These mechanisms indicate the complex, dynamic processes occurring constantly within the cell and among the structural components of its hierarchy.

Plants have evolved anti-ROS protective response mechanisms involving the production of colored pigments such as *carotenoids* and *anthocyanins* [2, 32, 43, 48], sometimes called *sunscreens for plants.* Biosynthesis of the protective pigments is stimulated upon exposure to oxidative stress. Carotenoids act to protect the genome among other anti-ROS functions [11]. Anthocyanins are synthesized via the phenylpropanoid pathway in plants; this pathway also produces *flavonoids,* another class of protectant molecules. Carotenoids are produced via the isoprenoid pathway [11].

4. Microarrays and Expresso

The desire to understand the complex networks of interactions that characterize plant response to drought stress leads to experiments that investigate patters of gene expression in cells under drought and non-drought conditions. Microarrays are a high-throughput biotechnology to access in parallel the gene expression patterns in cells under specific experimental conditions. A particular microarray experiment that my collaborators have defined but not yet performed is intended to determine which genes in Andean potato varieties are responsible for the superior drought stress resistance and vitamin content found among certain varieties. Figure 2 contains a diagram representing this experiment. The diagram indicates that many of the steps in the experimental loop are computational in nature. For our purposes, the important biological steps result in the creation — via robotic printing — of a microarray, which is a slide containing thousands of "spots" of material, each representing a particular gene; the hybridization of RNA extracted from cells under two different experimental conditions to the material on the slide; and the scanning of the hybridized microarray into two images, each representing the response of every spot to one of the two experimental conditions.

Once the two images are obtained, there is significant computation required to access the meaning contained therein. First, image processing is done in each image to reliably identify the thousands of spots to associate each spot with the corresponding gene and to correlate the intensity (brightness) of each spot with the level of gene expression of the corresponding gene. Statistical analysis of the obtained intensities must be performed to access a level of confidence in an hypothesis such as 'The gene expression of gene x under the second experimental condition is higher than under the first experimental condition." Support for such hypotheses are a first hint as to which genes are important for, say, enhanced stress resistance. Combining the results from multiple genes and multiple experiments is a more challenging problem toward that end. Expresso uses data mining techniques, especially inductive logic programming (ILP), and Bayesian networks.

Expresso is an innovative and integrated solution to microarray experiment management and data analysis that is being developed by an interdisciplinary research team at Virginia Tech. Expresso integrates all phases of microarray experiments into one system, including experiment design (selection of genes, chip layout, specification of hybridizations to be performed); image analysis; statistical analysis; data management via a unique semi-structured database; data mining via inductive logic programming (ILP) [33]; and integration of biological information from diverse sources into the database, analyses, and data mining. An important aspect of the integrated nature of Expresso is that it organically provides support for closing the experimental loop, allowing the results of the analysis of data from previous experiments to feed directly into the design of subsequent experiments. See Figure 3. The flexibility of Expresso is reflected in its support for multiple alternatives at each phase of the experiment (e.g., the statistical analysis discussed below), as opposed to the myriad stand-alone software systems that support a single alternative for a single phase. Of particular interest is the planned database of biological networks.

Numerous statistical techniques for analyzing the rich datasets generated in microarray experiments have been proposed and implemented [3, 9, 18, 24, 25, 26, 27, 28, 29, 31, 36, 42, 47, 49]. Each technique aims to address one or more aspects of the complexity of microarray datasets and none is capable of applying immediately to resolve every dilemma posed by the high-dimensional parameter space in which the datasets reside. The design of the Expresso system recognizes the value of having multiple statistical techniques available and, indeed, of applying diverse techniques to a data set. Confirmation of results from multiple analyses is analogous to confirmation via repetition of an experi-

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

ment but entails only the marginal cost of some additional computation. for microarray data.

The Expresso approach to further analysis emphasizes data mining. Data mining techniques, primarily ILP, are used to suggest high-level descriptors from expression data. Inductive logic programming (ILP) [34] provides a structured approach to finding rules that associate the level of gene expression to experimental conditions (such as levels of stress). ILP uses the language of first-order predicate logic to encode experimental conditions, gene clusters, and other properties useful for forming high-level representations. For example, `activation(expt1,gene-cluster1,0.5)` asserts that genes in `gene-cluster1` are moderately activated under the conditions of experiment `expt1`. The rules produced by ILP specify the interactions between the various predicates and their parts. This also enables domain-specific background knowledge (such as the fact that the activation of a certain group of genes are known to be inversely correlated with expression data of a different group) to be incorporated into the data mining process [4]. In addition, the induced concept descriptions are easily comprehensible — the example rule:

```
activation(E2,G,-1) :-  activation(E1,G,-0.5),
                        stresslevel(E1,S1),
                        stresslevel(E2,S2),
                        S2>S1+2.
```

expresses the mined pattern that genes in a cluster (G) go from 'moderately repressed' (E1) to 'heavily repressed' (E2) by increasing stress levels (from S1 to S4) by more than two orders of magnitude.

The networks mined by such techniques can represent temporal and causal relationships in a simple form. Expresso can use Bayesian networks [22] that propagate conditional probabilities through a graphical representation, and thus model causal relationships in a direct way. In addition, Bayesian networks can handle noise, hypothesize missing variables, and encode expert knowledge in a limited form. The complexity of learning Bayesian networks is NP-hard but various approximation algorithms such as the EM (expectation-maximization) approach lend credibility to its use in bioinformatics.

5. Experimental Results

Colleagues Ruth Grene and Boris Chevone have investigated expression patterns of genes in needles of loblolly pine seedlings that had been exposed to cycles of drought conditions over a growing season. The pine seedlings were subjected to mild severe drought stress for four *mild stress* (a level of drying in pine needles that results in little effect on growth and new flushes compared to control trees) or *severe stress* (a level of drying in pine needles that results in growth retardation with markedly fewer new flushes compared to controls) for four or three cycles, respectively. Total RNA was isolated from the samples by the method of Chang *et al.* [8], modified in Ruth Grene's laboratory and used as a source of material to probe microarrays with spots representing genes from the NSF Pine Genome Sequencing Project (Ronald Sederoff, PI, NCSU).

Many of the 60,000 mRNAs sequenced by the Pine Genome Sequencing Project have a proposed functional annotation derived from a BLAST (sequence) search of protein databases. In the first year (1999), 384 genes of known function were printed. A 2103 genes set was used on the microarrays in the second year (2001). A system of functional categories was set up to include all the genes that were printed. Using statistical and data mining algorithms incorporated in Expresso [1], genes and groups of genes involved in stress responses were identified. Signal transduction, drought acclimation, photosynthesis, and protection /repair genes are up-expressed specifically in acclimated needles. The numbers of genes whose expression was affected under mild conditions, and not under severe conditions were 38, 960, and 281 for Cycles 1, 2 and 3. Some of the categories into which these genes fell were transcription factors, drought-acclimation, oxidative stress resistance and protection and repair. At the final harvest under mild conditions in 1999, the expression of genes associated with drought acclimation, such as the dehydrins and aquaporins was increased, with either negative or undetectable change for the severe stress condition. LP-3, an established water-stress inducible gene in loblolly pine, increased under mild but not under severe condition. The same pattern was observed for glutathione-S-transferase (antioxidant function), proteases, receptor-like protein kinases (signal transduction), phosphoribulokinase, transketolase (chloroplast form), rubisco-binding proteins, protochlorophyllide reductase (photosynthesis), genes encoding protection/repair genes (HSPs) such as HSP70 (chloroplast-associated chaperone function [39]), HSP23 (LEA-like genes [14]) and HSP100 (thermotolerance [23]). These data provide a first snapshot of the status of gene expression specifically associated with acclimation during the course of a month-long exposure to cycles of drought stress in a woody species.

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

6. Two Exemplary Networks in Biology

The detailed response of plant cells to drought stress is only partly known (see Figure 1, but experimental evidence lends important clues about the process. For example, a putative osmosensor AtHK1, a histidine kinase located in the cell membrane, is thought to be one component to relay changes in osmotic potential outside the cell to intracellular signal transduction pathways [13]. Other membrane sensors are also hypothesized to be present and may respond to increasing levels of stress. These sensors trigger discrete phospholipid-based signaling pathways which are involved in early events in drought stress responses, with different pathways proposed to respond to different stress levels [7].

While the current knowledge of plant response to drought stress is incomplete, one finds complementary views represented in the literature that provide an excellent start on network models for drought stress responses. For example, Figure 4 gives a network exemplifying response to osmotic stress adapted from Munnik and Meijer [35], suggesting alternative responses dependent on the level of stress imposed and the resulting perception by a variety of osmosensors (only one of which — $ATHK_1$ — has been identified [44]). An alternative example (Figure 5) gives a related network that emphases the role of ABA in drought stress responses, adapted from Shinozaki *et al.* [40].

These networks are excellent starting points for a more detailed model of drought stress responses. They each suggest the spatial flow of events from the cell membrane to the nucleus (gene expression), as well as the temporal flow of rapid response followed by slower adaptation. The networks are incompatible, however, as the nodes are different in each (reflecting the alternate perspectives of the two research efforts) and are even at different levels of abstraction within the same network. The two networks contain nodes representing individual molecular species (PI3K, PLD, etc., in Figure 4; ABA in Figure 5), as well as gene expression, an immensely complex process involving many genes and molecular mechanisms for transcription. It should also be noted that the **kind of relationship** represented varies from node to node and arc to arc.

In developing a representation for these networks, it is essential to utilize the lessons learned from these and similar networks in the biological literature. Nodes must be of multiple types and at multiple levels of abstraction. Arcs must be capable of representing multiple types of relationships between nodes. Moreover, our networks must be able to include arcs representing many-to-many relationships; this is seen in Figure 4 through the potential of multiple osmosensors to influence multiple signal transducers. Of course, future experiments may associate particular osmosensors with particular signal transducers. Operations on networks must support refinement of networks to reduce a many-to-many relationship to, for example, numerous one-to-one relationships. As different networks representing related phenomena are typically incompatible (as discussed above), a combination of two or more networks is **partial,** in the sense that there will be only some nodes and arcs in common and the same process may be represented redundantly in the combination, though in alternate ways.

7. Multimodal Networks: General, Flexible, Extensible Models

Expresso will use networks (directed graphs or hypergraphs) as its underlying models for representing the topological and dynamic aspects of molecular transformation and transportation within the plant cell in response to stress imposition. While networks are capable of representing time, topology, and causality in natural ways, the richness of information available in cell and molecular biology requires more than just a network. As a result, the network model will be refined and extended to represent such information as hierarchical structure, and uncertainty. These extended networks are *multimodal,* as they incorporate diverse forms of information in a single framework.

Network Models. In the biological literature, models of cell metabolism or signal transduction are typically expressed visually as pathways or networks. Nodes in such a network can represent chemical reactions involving one or more chemical inputs and zero or more enzymes, producing one or more reaction products, while arcs represent the reaction products flowing from one reaction to another. Alternately, a node can represent a metabolite and an incoming arc can represent the required precursors and enzymes for the production of the metabolite. (See Polle [38] for a typical example. Also, see [16, 17].)

A first network model is a formalization of a biological pathway as a network that expresses the dependencies among the components in the pathway, much like Figures 4 and 5. The important work is to extend this initial model to carefully address some additional aspects of the plant cell and of the nature of biological knowledge.

Hierarchical Organization. Plant cells are hierarchically organized, with each cell containing organelles

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

and finer levels of structure occurring within organelles. Significant details of the hierarchical organization can be derived from the biological literature and represented in multimodal networks, utilizing hierarchical connections among network nodes at different levels of cellular organization.

Temporal Information. Many reactions within the cell occur constantly and in parallel with other reactions. Other reactions occur in response to certain internal or external conditions but still in parallel with other reactions. Finally, there are reactions that can occur only after other reactions have produced the needed precursors or an information transfer has brought the needed precursors into spatial proximity. The dependence or independence of reaction sequencing can be represented implicitly or explicitly within an augmented network model.

Uncertainty. Probabilities, reflecting uncertainty, are naturally attached to the arcs in a multimodal network to yield a probabilistic, constrained network. Such a network generalizes such concepts as reliability in communication networks [10] and Markov chains [15], and can be analyzed using extensions of stochastic techniques to be developed as part of the mathematical theory of multimodal networks.

8. Some Future Prospects

Numerous biological phenomena that should be present in models of stress response offer research challenges for computer scientists. Some of these phenomena are:

- **Compartmentalization** of genomic information and cellular processes among the nucleus, organelles, the cell wall, and the cytosol.
- **Alternative pathways** that promote redundancy and fault-tolerance in the function of a cell.
- **Metabolic pathways** involved in essential cell processes such as bioenergetics.
- **Signaling pathways and attendant downstream events** responding to environmental changes.

Research topics specific to multimodal network models include:

- Developing a mathematical theory of the space of multimodal networks and operations on that space.
- Creating a library of computational models — based on multimodal networks — for cell and molecular biology phenomena.
- Providing computational mechanisms for manipulating the library, including creation, combination, and evaluation of multimodal models.
- Developing predictive multimodal models for drought stress responses in plants.

Intriguing developments exist in the more distant future of biology and bioinformatics. These may lead to products and processes such as the following.

Biological and Micromanipulator Systems. Multimodal models can be applied in the context of a mixed biological/micromanipulator system accomplishing mechanical tasks via MEMS that are controlled by a "sea" of single-celled organisms. These organisms will be selected for survival under extreme environmental conditions and will be programmed for the cell-to-cell and cell-to-micromanipulator interactions necessary to allow the MEMS to achieve a design goal such as optimizing the drag on an airplane wing. Models for these organisms, for the micromanipulators, and for the drag on a wing can be simulated. This mixed biological/micromanipulator system will be a theoretical construct used to identify the informational, biological, and engineering challenges present in such a mixed biological/micromanipulator system.

Complex Environmental Sensors. Biological cells are capable of detecting a range of environmental conditions, including light levels, temperature, moisture levels, lack of nutrients, and chemical "attacks" and of adjusting their internal functioning in response to changes in these conditions. Conceptually, a change raises a signal within the cell that is then processed by one or more pathways. Using boolean combinations of signals available in a programming model, it is possible to have a cell detect complicated environmental conditions (e.g., dark, hot, and wet). With application of current biotechnology, it is then possible to "program" a cell that can glow at a particular frequency when those environmental conditions are detected. The multimodal modeling approach can potentially support the design of complex environmental sensors that could be used in, for example, potentially hazardous environments.

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

9. Acknowledgments

The generous support of National Science Foundation grant EIA-0103660 for the Expresso system is gratefully acknowledged. The contributions of my many collaborators in bioinformatics are also gratefully acknowledged. Among many others, I wish to thank Ananth Grama, Ruth Grene, Naren Ramakrishnan, Craig A. Struble, and Layne T. Watson.

References

[1] R. G. Alscher, B. I. Chevone, L. S. Heath, and N. Ramakrishnan. Expresso: A problem solving environment for bioinformatics: Finding answers with microarray technology. In *Proceedings of the High Performance Computing Symposium, Advanced Simulation Technologies Conference*, pages 64–69, 2001.

[2] D. Bagchi, M. Bagchi, S. Stohs, D. Das, S. Ray, C. Kuszynski, S. Joshi, and H. Pruess. Free radicals and grape seed proanthocyanidin extract: importance in human health and disease prevention. *Toxicology*, 148(2–3):187–197, 2000.

[3] P. Baldi and A. D. Long. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, 17(6):509–19, 2001.

[4] I. Bratko and S. Muggleton. Applications of Inductive Logic Programming. *Communications of the ACM*, Vol. 38(11):pp. 65–70, November 1995.

[5] T. A. Brown. *Genomes*. John Wiley and Sons, Inc., New York, 1999.

[6] B. B. Buchanan, W. Gruissem, and R. L. Jones. *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, Maryland, 2000.

[7] E. J. Calabrese, L. A. Baldwin, and C. D. Holland. Hormesis: a highly generalizable and reproducible phenomenon with important implications for risk assessment. *Risk Anal*, 19(2):261–81, 1999.

[8] S. Chang, J. Puryear, and J. Cairney. A simple and efficient method for isolating rna from pine trees. *Plant Molec. Biol. Reporter*, 11:113–116, 1993.

[9] Y. Chen, E. R. Dougherty, and M. L. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2(4):364–374, 1997.

[10] C. J. Colbourn. *The Combinatorics of Network Reliability*. Oxford University Press, New York, NY, 1987.

[11] A. Collins. Carotenoids and genomic stability. *Mutat Res.*, 475(1–2):21–28, April 2001.

[12] G. M. Cooper. *The Cell: A Molecular Approach*. Sinauer Associates, Inc., Sunderland, Massachusetts, second edition, 2000.

[13] B. Degenhardt and H. Gimmler. Cell wall adaptations to multiple environmental stresses in maize roots. *J Exp Bot*, 51(344):595–603, 2000.

[14] J. Z. Dong and D. I. Dunstan. Characterization of three heat-shock-protein genes and their developmental regulation during somatic embryogenesis in white spruce *[Picea glauca (Moench) Voss]*. *Planta*, 200(1):85–91, 1996.

[15] J. L. Doob. *Stochastic Processes*. Wiley, New York, NY, 1953.

[16] J. S. Edwards and B. O. Palsson. The escherichia coli mg1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5528–5533, May 9 2000.

[17] J. S. Edwards and B. O. Palsson. Metabolic flux balance analysis and the in silico analysis of escherichia coli k-12 gene deletions. *BMC Bioinformatics*, 1(1):1–10, July 27 2000.

[18] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. *J Comput Biol*, 7(3-4):601–20, 2000.

[19] A. J. F. Griffiths, J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart. *An Introduction to Genetic Analysis*. W. H. Freeman and Company, New York, seventh edition, 2000.

[20] N. Gustavsson, U. Harndahl, A. Emanuelsson, P. Roepstorff, and C. Sundby. Methionine sulfoxidation of the chloroplast small heat shock protein and conformational changes in the oligomer. *Protein Sci.*, 8(11):2506–2512, 1999.

[21] U. Harndahl, B. Kokke, N. Gustavsson, S. Linse, K. Berggren, F. Tjerneld, W. Boelens, and C. Sundby. The chaperone-like activity of a small heat shock protein is lost after sulfoxidation of conserved methionines in a surface-exposed amphipathic alpha-helix. *Biochim. Biophys. Acta*, 1545(1–2):227–237, 2001.

[22] D. Heckerman. Bayesian Networks for Knowledge Discovery. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 273–306. AAAI/MIT Press, 1996.

[23] S. W. Hong and E. Vierling. Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proc Natl Acad Sci U S A*, 97(8):4392–7, 2000.

[24] T. Ideker, V. Thorsson, A. F. Siegel, and L. E. Hood. Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol*, 7(6):805–17, 2000.

[25] M. K. Kerr and G. A. Churchill. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc Natl Acad Sci U S A*, 98(16):8961–5, 2001.

[26] M. K. Kerr and G. A. Churchill. Statistical design and the analysis of gene expression microarray data. *Genet Res*, 77(2):123–8, 2001.

[27] M. K. Kerr, M. Martin, and G. A. Churchill. Analysis of variance for gene expression microarray data. *J Comput Biol*, 7(6):819–37, 2000.

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 **IEEE**

[28] S. Kim, E. R. Dougherty, M. L. Bittner, Y. Chen, K. Sivakumar, P. Meltzer, and J. M. Trent. General nonlinear framework for the analysis of gene interaction via multivariate expression arrays. *J Biomed Opt*, 5(4):411–24, 2000.

[29] M. L. Lee, F. C. Kuo, G. A. Whitmore, and J. Sklar. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A*, 97(18):9834–9, 2000.

[30] B. Lewin. *Genes VII*. Oxford University Press, Oxford, 2000.

[31] A. D. Long, H. J. Mangalam, B. Y. Chan, L. Tolleri, G. W. Hatfield, and P. Baldi. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem*, 276(23):19937–44, 2001.

[32] M. Merzlyak and O. Chivkunova. Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples. *J. Photochem. Photobiol.*, 55(2–3):155–163, Apr-May 2000.

[33] S. Muggleton. Scientific knowledge discovery using inductive logic programming. *Communications of the Association for Computing Machinery*, 42(11):42–64, 1999.

[34] S. Muggleton. Scientific Knowledge Discovery using Inductive Logic Programming. *Communications of the ACM*, Vol. 41(11):pp. 56–62, November 1999.

[35] T. Munnik and H. J. G. Meijer. Osmotic stress activates distinct lipid and MAPK signalling pathways in plants. *FEBS Letters*, 498:172–178, 2001.

[36] M. A. Newton, C. M. Kendziorski, C. S. Richmond, F. R. Blattner, and K. W. Tsui. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J Comput Biol*, 8(1):37–52, 2001.

[37] A. Oberschall, M. Deak, K. Torok, L. Sass, I. Vass, I. Kovacs, A. Feher, D. Dudits, and G. V. Horvath. A novel aldose/aldehyde reductase protects transgenic plants against lipid peroxidation under chemical and drought stresses. *Plant J*, 24(4):437–46, 2000.

[38] A. Polle. Dissecting the superoxide dismutase-acorbate-glutathione-pathway in chloroplasts by metabolic modeling. computer simulations as a step towards flux analysis. *Plant Physiology*, 126:445–462, May 2001.

[39] D. V. Rial, A. K. Arakaki, and E. A. Ceccarelli. Interaction of the targeting sequence of chloroplast precursors with HSP70 molecular chaperones. *Eur J Biochem*, 267(20):6239–48, 2000.

[40] K. Shinozaki and K. Yamaguchi-Shinozaki. Molcular reponses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. *Current Opinion in Plant Biology*, 3:217–223, 2000.

[41] A. D. Smith, editor. *Oxford Dictionary of Biochemistry and Molecular Biology*. Oxford University Press, Oxford, revised edition, 2000.

[42] J. G. Thomas, J. M. Olson, S. J. Tapscott, and L. P. Zhao. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res*, 11(7):1227–36, 2001.

[43] T. Tsuda, F. Horio, and T. Osawa. The role of anthocyanins as an antioxidant under oxidative stress in rats. *Biofactors*, 13(1–4):133–139, 2000.

[44] T. Urao, B. Yakubov, R. Satoh, K. Yamaguchi-Shinozaki, M. Seki, T. Hirayama, and K. Shinozaki. A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. *Plant Cell*, 11(9):1743–54, 1999.

[45] M. S. Waterman. *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman and Hall/CRC, Boca Raton, reprint edition, 2000.

[46] N. Wehmeyer and E. Vierling. The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiol.*, 122(4):1099–1108, 2000.

[47] R. D. Wolfinger, G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol*, 8(6):625–37, 2001.

[48] K. Youdim, B. Shukitt-Hale, S. MacKinnon, W. Kalt, and J. Joseph. Polyphenolics enhance red blood cell resistance to oxidative stress: in vitro and in vivo(1). *Biochem. Biophys. Acta*, 19(1):117–122, September 2000.

[49] H. Zegzouti, B. Jones, C. Marty, J. M. Lelievre, A. Latche, J. C. Pech, and M. Bouzayen. ER5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. *Plant Mol Biol*, 35(6):847–54, 1997.

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 **IEEE**

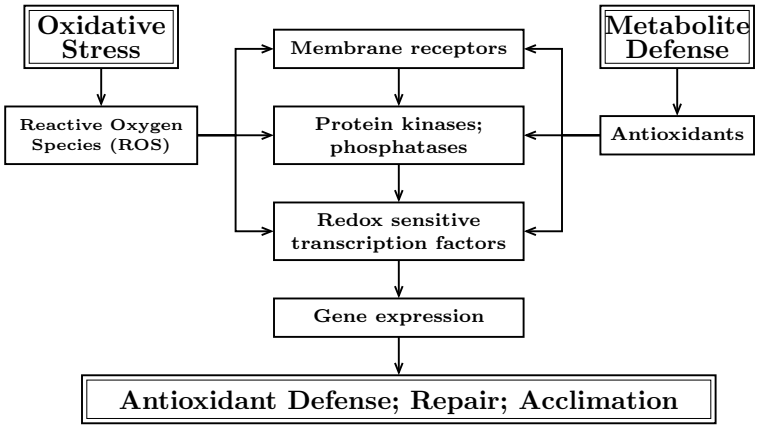

Figure 1. Cellular response to ROS.

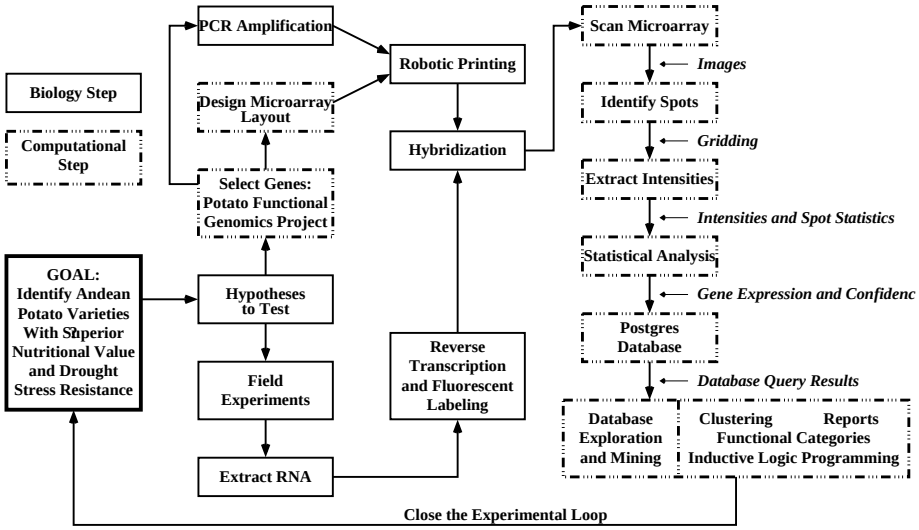

Figure 2. A representation of the flow of a microarray experiment to study nutritional value and drought stress resistance in Andean potato varieties.

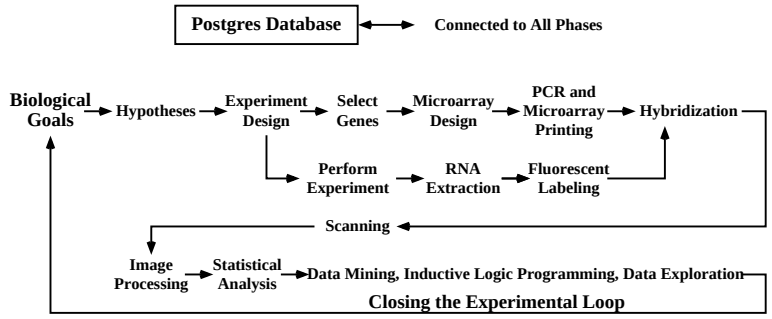

Figure 3. The flow of processing in the Expresso system.

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE

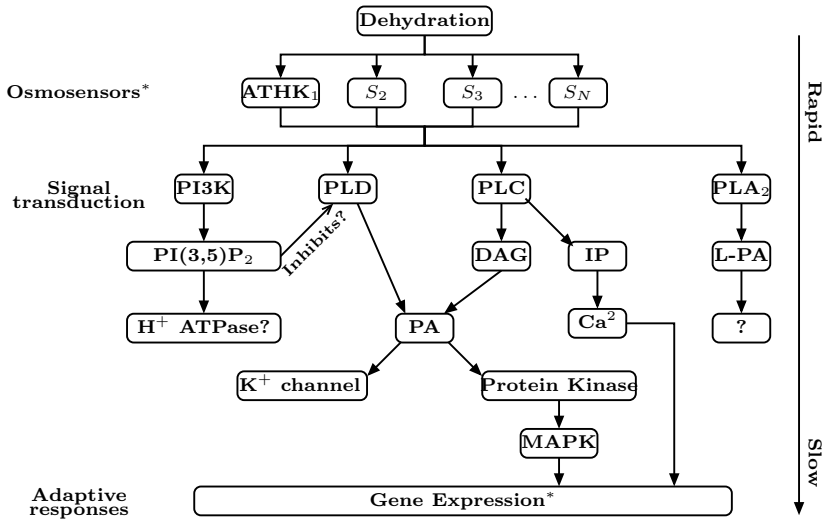

Figure 4. This network is adapted from Munnik and Meijer [35]. *Different osmosensors and signaling pathways are proposed to respond to different levels of osmotic stress.

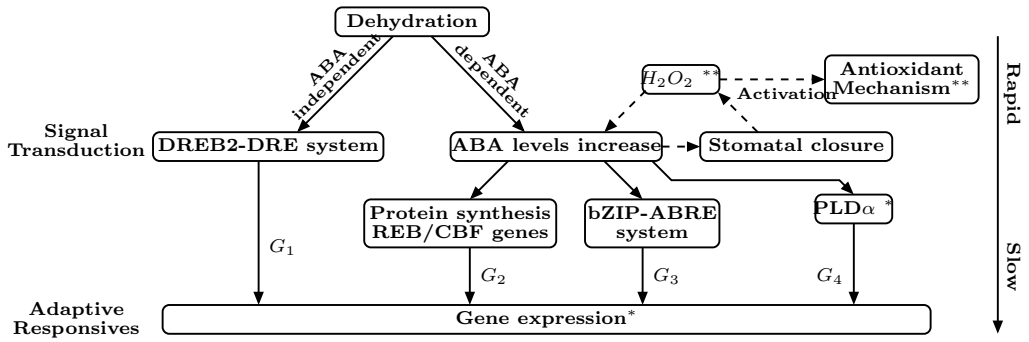

Figure 5. This network is adapted from Shinozaki and Yamaguchi-Shinozaki [40]. Four mechanisms for influencing gene expression are represented, as suggested by G_1, G_2, G_3, and G_4 on the arcs. **DREB2/DRE system (drought responsive element binding protein and drought responsive element); REB/CBF (rice endosperm binding factor and C-repeat binding factor); ABRE (ABE responsive element).

Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'02)
1087-4089/02 $17.00 © 2002 IEEE