DRUG DISCOVERY TODAY BIOSILICO

Artificial intelligence and robotics in high throughput post-genomics

Aroosha Laghaee, Chris Malcolm, John Hallam and Peter Ghazal

The shift of post-genomics towards a systems approach has offered an ever-increasing role for artificial intelligence (AI) and robotics. Many disciplines (e.g. engineering, robotics, computer science) bear on the problem of automating the different stages involved in post-genomic research with a view to developing quality assured high-dimensional data. We review some of the latest contributions of AI and robotics to this end and note the limitations arising from the current independent, exploratory way in which specific solutions are being presented for specific problems without regard to how these could be eventually integrated into one comprehensible integrated intelligent system.

Aroosha Laghaee*
Chris Malcolm
Institute for Perception,
Action and Behaviour (IPAB),
School of Informatics,
James Clerk Maxwell
Building, University of
Edinburgh,
Mayfield Road,
Edinburgh EH9 3JZ, UK
*e-mail: a.laghaee@
sms.ed.ac.uk
John Hallam
The Maersk Mc-Kinney Moller
Institute for Production
Technology, University of
Southern Denmark,
Campusvej 55,
DK-5230 Odense M,
Denmark
Peter Ghazal
Scottish Centre for Genomic
Technology and Informatics
(GTI), College of Medicine,
The Chancellor's Building,
The University of Edinburgh,
49 Little France Crescent,
Edinburgh EH16 4SB, UK

Biology is undergoing a knowledge revolution due to the increasing volume of data and qualitative analysis with which biological systems are being studied, and the shift from the more traditional reductionist to holistic and systems based approaches. In part this results from and facilitates an increasing role for artificial intelligence (AI) and robotics.

AI attempts to build intelligent agents (robotic or software). An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors. AI systems have the ability to achieve stated ends in the face of variation, difficulties and complexities posed by the task [1]. They are essential when a solution cannot be programmed in a conventional deterministic fashion. AI has inherited many ideas, viewpoints and techniques from other disciplines. For example, theories of reasoning and learning have emerged from philosophy. Mathematics has lent AI theories of logic, probability, decision-making and computation. Computer science provides the tools with which to make AI a reality.

Some AI methodologies and tools can provide solutions to life-science problems where conventional mathematical, statistical and brute force methods prove to be ineffective or inefficient. The AI fields of machine learning, knowledge management and multi-agent systems can make significant contributions to experimental implementation. The fields of agents, human–computer interaction, natural language processing, vision and computational creativity can provide methodological and technological support for integrating and optimizing capabilities of humans and robots. Finally, fields such as data mining, knowledge-based systems and computational theories can offer a natural link from data to knowledge (see Box 1 for a glossary of these terms).

Whole genome sequences and new high throughput technologies enable researchers to approach questions systematically and on a grand scale. They can study all the genes in a genome or the transcripts in a particular tissue or organ or tumour, or the way in which our tens of thousands of genes and proteins

1359-6446/05/$ – see front matter PII: S1359-6446(05)03581-6

BOX 1

Glossary of terms used in the text

Bayes Theorem is a means of quantifying uncertainty. Based on probability theory, the theorem defines a rule for refining a hypothesis by factoring in additional evidence and background information, and leads to a number representing the degree of probability that the hypothesis is true. Bayesian inference techniques have been a fundamental part of computerized pattern recognition (a field within the area of machine learning) techniques since the late 1950s. Some machine learning researchers create methods within the framework of Bayesian statistics.

Brute force refers to any method that does not involve a heuristic or rely on any intelligent observation but tries every possible solution to find the best solution.

Data mining is an analytical process designed to explore large amounts of data in search of consistent patterns or systematic relationships between variables and then to validate the findings by applying the detected patterns to new subsets of data. Data mining consists of three stages: (i) initial exploration, (ii) model building or pattern identification with validation/verification, and (iii) deployment. It is often considered to be a blend of statistics, AI and database research.

Knowledge-based systems are computer systems that are programmed to imitate human problem-solving by means of AI and reference to a database of knowledge on a particular subject.

Machine learning is an area of artificial intelligence (AI) concerned with the development of techniques that enable learning by analyzing data sets. Machine learning overlaps heavily with statistics but unlike traditional statistical data analysis (which is usually concerned with the estimation of population parameters by statistical inference) the emphasis in machine learning is usually on the accuracy of prediction regardless of whether or not the 'models' or techniques used to generate the prediction is interpretable or open to simple explanation.

Microarrays comprise a large number of genes deposited on to a solid support, such as glass or nylon, which are used for a multiplex reaction. Nucleic acid (usually cDNA or long oligonucleotides) is spotted in a grid arrangement. The microarrays serve as hybridization targets for labeled nucleic acids extracted from tissue or cell lysates. This is a multi-step process involving different stages entailing both human and machine operations.

Multi-agent systems (MAS) are composed of several agents that are capable of mutual interaction. The interaction can be in the form of message passing or producing changes in their common environment.

TABLE 1

Pre-genomic versus post-genomic approaches to research

Pre-genomic	Post-genomic
Reductionist (DNA or RNA or protein)	Global/holistic
Observational/phenomenological	Systems approach (DNA and RNA and protein)
Generally qualitative and non-numeric	Quantitative data and highly numeric
Hypothesis driven	Not only hypothesis driven but also data driven

work together in their interconnected networks to make up a complex or single cell organism.

In turn, the resulting knowledge has many applications in medical diagnosis and therapy. Once the gene and protein networks and functions of healthy tissue have been established, this information could be used to deduce the differences between such networks and diseased (e.g. cancerous) tissue, helping to determine which genes might cause or cure cancer or identify which treatment a specific tumour might respond to best.

High-dimensional (omic) biology

The delivery of whole genome sequences (genomics) and the investigation of how DNA and proteins work with each other and the environment (post-genomics) is generating vast amounts of data at multiple levels of complexity in biological systems research, resulting in a new field of high-dimensional biology. This emergent discipline is challenging the methodologies of biological investigation.

Table 1 provides a list of the hallmarks of pre-genomic science in comparison with the post-genomic challenge. The pre-genomic era has been deeply rooted in the methodology of reducing complex systems to their elemental forms. This has resulted in the establishment of the independent fields of genetics (study of DNA), biochemistry (study of small molecules, enzymology) and molecular biology (study of RNA and regulatory mechanisms). Many of these approaches are qualitative and not particularly numeric, with information generally provided in an unstructured format. They also provide an important phenomenological basis to biological observations. The pre-genomic approach is by necessity exclusively hypothesis-driven and involves assessing a limited number of variables.

In marked contrast, post-genomic science takes a global systems approach and integrates information from the DNA to protein levels. High throughput technologies are often used with numeric outputs. This rapidly provides vast quantities of data whose high dimensionality requires the use of relational databases. This affords an entirely new opportunity in combining hypothesis-driven research with a data-driven approach.

Figure 1 illustrates the process flow characterizing post-genomic research, which involves the formation of a question based on known knowledge, a proposed experiment to answer the question, experimental implementation, data analysis and interpretation, which in turn might be added to the knowledge base. These stages could be categorized into design, implementation and analysis, each dealing with a different kind of knowledge being manipulated. Currently, experiments are designed by researchers and implemented partly by human technicians and partly by robotic stations. The researchers then analyze and interpret the results and feed the results back to the research community. However, the rapid increase in the number of parameters and variables involved and the growing need for quality assured results during the three stages make it highly desirable to construct intelligent automated systems that are capable of learning (to improve performance over time) to aid these processes.

The data driven aspects of design and analysis make AI systems well-suited solutions for automation. The need for accuracy, repeatability and efficiency in Implementation mean that robotics can make useful contributions. Here we concentrate on the possible contributions of AI and robotics in post-genomic research by reviewing recent applications of AI systems and robotics to the automation of the design, implementation and analysis stages. We discuss the shortcomings of the current approaches and how these could be overcome. We discuss only an illustrative selection of papers and give further references for readers to pursue.

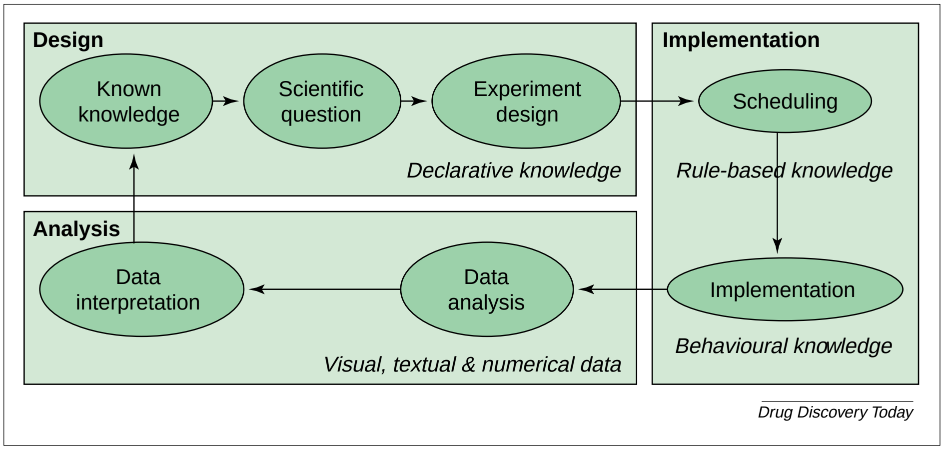

FIGURE1

An example of a full cycle of stages in post-genomic research and the different kinds of knowledge manipulation involved. In design, declarative knowledge is used (e.g. which probes to include), scheduling deals with rule-based knowledge (e.g. interdependency between processes), implementation uses behavioural knowledge (e.g. DNA-labelling protocol) and, finally, data analysis involves textual, numerical and visual data.

Automation in design

At present, experimental design is conducted predominantly by researchers. Little effort has been directed towards automating this process. The sheer volume of data available in the many hundreds of non-integrated databases on the Internet means that a lot of human time and effort is spent searching these databases, cross comparing and deciding which probes (RNA, DNA or antibody) to include in a study. Given the crucial dependency on the latest data from these large databases, AI can help to extract the information relevant for experimental design. In this section we describe recent research that tries to address this issue by applying AI systems to automate the design process.

King *et al.* [2] present a 'robot scientist' that is capable of functional genomic hypothesis generation and experimentation. This work is highly specific in that it models the relatively well-known aromatic amino acid synthesis pathway in yeast to rediscover gene function, using deletion mutants of yeast and auxotropic growth experiments. This system is capable of forming hypotheses about the function of genes in the pathway because it has a logical model of the pathway and would not be useful in gene function discovery experiments, that is, in the absence of a known pathway.

It is important for such systems to be generic otherwise it would be necessary to produce a completely independent system for each pathway that is being studied. Moreover, as more information becomes available about the pathway in question such systems will need constant modification. It is also desirable that these systems should have an interface to the available databases on the Internet so the latest results can be taken into account.

There are emerging tools that try to address some of these issues. For example Draghici *et al.* [3] describe a tool called the Onto-Design (OD) that enables the quick design of a microarray by constructing an optimal set of genes for a given set of biological processes or pathways. OD is an ontology-based tool that takes advantage of the Gene Ontology (GO) terms to construct its gene list. The input to OD is a set of GO terms based on molecular function, cellular function and biological function. OD output is the set of genes involved in each of the GO terms selected (OD can be found at http://vortex.cs.wayne.edu/projects.htm). However, this complete reliance on the GO Consortium means the exclusion of domain structure, 3D structure, evolution data and expression data. The future incorporation of such information could make GO a highly useful resource here.

Abduction (inference to the best explanation) goes from data describing something to a hypothesis that best explains or accounts for the data (a kind of theory forming or interpretive inference). This makes AI research that develops automated inference systems relevant to the problem of experimental design based on background knowledge or previous experimental results. Zupan *et al.* [4] have applied abductive inference to elucidate network constraints based on background knowledge and experimental results in a system called 'GenePath', which can infer genetic networks and propose genetic experiments that might refine the discovered network and establish relations between genes that could not be related purely based on the original experiment. Zupan *et al.* [4] illustrate the approach of GenePath and its use for the analysis of data on aggregation and sporulation of the soil amoeba *Dictyostelium discoideum* (GenePath; http://www.genepath.org/).

Yoo and Cooper [5] have modelled the expected value of experimentation (intervention and observation) to discover causal pathways in gene expression data. The system, called GEEVE, can help discover gene regulation pathways by recommending which experiments to perform, the number of measurements to include in the experimental design and by providing a Bayesian analysis that combines prior knowledge with the results of recent microarray experiments to derive posterior probabilities of gene regulation relationships. GEEVE incorporates experimenter preferences regarding which genes to study, and considers this information along with costs and the current model of causal relationships to recommend the most cost effective experiment in its search space of experiments (for more examples, see [6–10]).

All of these approaches use different AI methodologies to simplify, explain and partially automate the complex and time-consuming processes involved in the experimental design of post-genomics. AI systems that model and automate the inference mechanisms used by geneticists can contribute towards designing cost-effective experiments, which can use previous data to further our understanding. They can explain why a given experiment is suggested. Moreover, machine learning techniques enable us to formalize and partially automate the reasoning and scientific method of geneticists with the ability to explain the reasoning used by the system, thus enhancing our understanding of how these AI systems operate.

Automation in implementation

Today, there are instruments that can automate nearly every step of the large-scale sequencing process: isolating DNA, cloning or amplifying DNA, preparing enzymatic sequencing reactions, purifying DNA, and separating and detecting DNA fragments with fluorescent labels to obtain DNA sequence. To meet the growing demand for genomic sequencing and analysis biologists, engineers and scientists have been working together to develop the technologies that have enabled this revolution in the way biology is performed and analyzed [11].

Currently, there are two major platform technologies used for the analysis of gene expression: microarrays and Affymetrix chips. Affymetrix's technology involves a process known as photolithography to create a huge number of short oligonucleotides (25 mers) directly on a silicon wafer. A single chip measuring 1.28 cm × 1.28 cm can hold >400,000 probe molecules. This is a fully automated process enabling Affymetrix to manufacture 5 to 10,000 DNA chips per month.

Both of these technologies owe their success to technologies used in the computing industry. To manufacture its chips, Affymetrix uses a manufacturing technique used in the fabrication of semiconductor chips, which depends on thorough and precise computer control. The spotted microarrays depend on robot precision, which much exceeds the human arm.

Other areas in post-genomic research benefiting from robotic automation include plaque and colony picking and gridding [12,13], DNA template preparation, DNA sequencing reactions [14], gel loaders, microarray production, liquid handling machines (e.g. Genetix and Biomek), resequencing and variation detection [15], and cell culture techniques ([16]; for more comprehensive reviews, see [11,17]).

Although this automation has been a key contribution towards the shift of post-genomics to a systems approach, the laboratories that use these technologies do not, as yet, enjoy the full advantages of intelligent automation and robotics. The robots and stations designed are highly accurate and sensitive but lack the useful generalized adaptability and learning capabilities that humans exhibit and AI research addresses. Each machine or robotic arm has been designed with a highly specialized problem in mind and, therefore, cannot be applied to other similar problems. For example, a robotic arm designed to undertake the task of loading and unloading a liquid handler with microtitre plates cannot be used for the similar task of loading and unloading a scanner with slides, whereas such task generalization is trivial to a person. Once a child learns how to place a particular shaped object in its designated slot, it will be a mere case of slight adjustment to place a different shaped object in its respective slot. However, such robotic stations as liquid handling machines can only cope with the specific layout of their interiors and objects of manipulation (e.g. microtitre plates). Slight dislocations of plates or tubes could result in malfunction and potential loss of expensive biological material. This has resulted in scattered 'unintelligent' robotic stations, which potentially could be integrated usefully with other automated stations in such laboratories.

Although there has been much effort in making modular instruments that can be integrated into larger systems, there are still shortcomings in practice. First, it is common for research laboratories to use instruments from different brands. This means that even if one instrument could have a modular extension, this might not necessarily be compatible with the other branded instruments. Second, it is also common for the different instruments to be located in different rooms (e.g. spotting machines need to be in clean rooms, whereas scanners do not). It is the job of human technicians to transfer material from one workstation to the next.

We do not suggest the construction of an intelligent robot that could undertake all of the different tasks. However, a generic intelligent robot that could generally conduct the material transfer stages between the different workstations would usefully assist the integration of the different instruments.

Automation in analysis

The next challenging step is how to interpret the resulting copious, noisy and complex data so as to feed this new information back into the community. It is nearly impossible for any one person to extract subtle patterns from such huge and messy tables of numbers. Intelligent automated assistance can help make sense of this.

In particular, data mining, machine learning and pattern recognition techniques have made contributions to data analysis. Machine learning in the context of data mining denotes the application of generic model fitting or classification algorithms for predictive data mining. Machine learning overlaps heavily with statistics (because it includes some advanced statistical methods) but unlike traditional statistical data analysis, which is usually concerned with the estimation of population parameters by statistical inference, the emphasis in data mining and machine learning is usually on the accuracy of prediction. These methods usually involve the fitting of complex generic models that are not related to any reasoning or theoretical understanding of underlying causal processes. Instead, these techniques can be shown to generate accurate predictions or classification in cross-validation samples.

Pattern recognition, a field within machine learning, is typically an intermediate step in a longer process. These steps generally are acquisition of the data to be classified, pre-processing to remove noise or normalize the data in some way, computing features, classification and finally post-processing based upon the recognized class and the confidence level. Pattern recognition itself is primarily concerned with the classification step and uses classification algorithms, such as Naïve Bayes classifier, k-nearest neighbour, decision trees, Bayesian networks, Hidden Markov models and neural networks.

Inza *et al*. [18] have applied four machine learning supervised classifiers to perform sample classification of gene expression microarray data: (i) A nearest neighbour algorithm (KNN), (ii) a Naïve Bayes classifier, (iii) a decision tree, and (iv) a classification model by a set of IF-THEN rules. KNN is a method for classifying phenomena based on observable features. Naïve Bayes classifiers are based on probability models that assume that independent variables are statistically independent. Naïve Bayes classifiers can be trained efficiently in supervised learning settings and often work well in many complex high-dimensional problems.

Additionally, filter and wrapper approaches are applied to the dataset to reduce redundant features by selecting gene candidates that actively contribute to classification. It was shown [ibid] that considerably more accurate results were obtained in comparison to non-gene selection methods, with most of the genes selected by the proposed procedures appearing on the lists of relevant genes detected by previous studies.

Walker *et al*. [19] use BioMiner, a data-mining software, to discover genes associated with Alzheimer's disease from microarray data. First, several pre-processing experiments are conducted to examine the data to obtain a general view of the data and identify missing values and abnormal or interesting characteristics. Second, hierarchical clustering is used to identify the inherent class characteristics of the data. Third, a set of machine learning experiments, using inductive algorithms to identify the most informative genes with an associated threshold for classification along with a measure of interest, are performed. Finally, data visualization techniques are used to better investigate the overall patterns of identified genes. The results of this system could be used to define therapeutic strategies to prevent the loss of specific components of neuronal function in susceptible patients or to stimulate the replacement of lost cellular function in damaged neurones.

Gras *et al*. [20] suggest that some key proteomics problems (i.e. automatic protein identification, biological motif inference and multiple sequence alignment) can be seen as combinatorial optimization problems, where the set of feasible solutions is discrete or can be reduced to a discrete one, and the goal is to find the best possible solution. They apply co-operative metaheuristics (a subset of metaheuristics efficient to give good approximated solutions in a reasonable time) that enable parallel exploration of the search space by several entities with information exchange between them. In the protein identification problem, a first level of co-operation swarm intelligence (SI) is applied to the comparison of mass spectrometric data with biological sequence database, followed by a genetic programming (GP) method to discover an optimal scoring function. When applied to a set of 271 MS–MS spectra, their system was able to identify 86.7% of the spectra correctly.

Swarm intelligence is an AI technique based on the study of collective behaviour in decentralized, self-organized, systems. SI systems are typically made up of a population of simple agents interacting locally with one another and with their environment. Although there is normally no centralized control structure dictating how individual agents should behave, local interactions between such agents often lead to the emergence of global behaviour.

GP is an automated methodology inspired by biological evolution to find computer programs that best perform a user-defined task. It is a particular machine learning technique that uses an evolutionary algorithm to optimize a population of computer programs according to a fitness landscape determined by a program's ability to perform a given computational task.

McClean *et al*. [21] have successfully adopted a principled model-based approach that uses a Hidden Markov Model for clustering and characterizing gene expression sequences classified according to heterogeneous classification schemes (for more examples see [22–28]). All of these examples illustrate the assistance of AI in automating the analysis of post-genomic's high-dimensional data. Examples include gene classification, protein identification, search optimization, knowledge discovery and gene expression profiling. By helping to explain the results, AI can make it easier for researchers to interpret the data and help define possible therapeutic strategies, as well as designing further experiments. It is also important that these systems think like geneticists, so that a geneticist can understand the explanations of the system about the results and can trust these explanations.

Concluding remarks

As illustrated in Figure 1, post-genomic research can be divided into three stages of design, implementation and analysis. Four different types of knowledge manipulation are involved: (i) designers deal with declarative knowledge such as the appropriate probes to put on glass slides, (ii) task planners deal with rule based knowledge such as the dependencies between the resources available and how they should be allocated and sequenced, (iii) implementation incorporates behavioural knowledge such as what actions achieve a certain task (e.g. protocol used to label DNA), and (iv) data analysis requires the interpretation of a wide spectrum of visual, numerical and textual data.

This has meant that researchers in different disciplines (e.g. data miners, engineers, roboticists, and so on) have had the chance to pick out those problems with the most interest to them and try to provide a solution (be it a robotic arm that can help with the liquid handling or an AI system that is capable of data analysis). However, in our view the wide range of solutions and systems produced so far suffer from several limitations.

The vision behind much of this automation work in software and hardware has been eventually to produce a collaboration between automated systems and people that could undertake all the different stages of a research program from experimental design through scheduling and task specification to data analysis. Of course, the formation of interesting questions to ask and the data interpretation cannot be fully automated yet because we are still far from being able to automate intuition and creative imagination.

To generate such a system it is important to take into consideration all the different aspects of the system from the initial stage of design. Otherwise, if each part is made independently, integration problems will inevitably arise when the whole system is put together. The larger and more complex the system, the more difficult these problems will be. If subsystems have been based on different assumptions then integration might be impossible.

Post-genomic research and AI-driven bioinformatics are both in their infancy, therefore, the solutions thus far generated have been the result of exploratory and independent problem solving amongst the different disciplines, with little thought about eventual integration. It is important to ensure that independent development does not preclude bridging the notoriously awkward gap between the lower-level control algorithms (e.g. robot arm control) and the higher-level cognition (e.g. experimental design).

Moreover, automation on its own is not sufficient; such a system should be able to improve its performance through experience. Learning is a fundamental part of intelligence, giving the advantage of adaptability. With adaptive systems we can move towards generalized solutions of wider scope and considerably reduce the work involved in tailoring a specific application. Adaptability is also essential because post-genomic research takes place in highly dynamic environments.

Currently, none of the automated stations and robotic arms available to the laboratories has learning capabilities built into them. This has resulted in a lot of expensive and unintelligent machinery scattered around the laboratories capable of undertaking only highly specific tasks and with a lot of expense incurred every time more functionality is needed (because a whole new piece of equipment has to be purchased).

It is important not to preclude the future possibility of building intelligent generic robots that could undertake a wide range of tasks and learn more as the need arises. For example, a generic and adaptive robot technician that knows how to load and unload could learn how to do this for a variety of objects and platforms available in a given laboratory once purchased. This would greatly reduce the amount of time and expense in purchasing all the different necessary kits and learning how to use and set them up.

Having delved into what possible solutions there might be to automate the numerous and complex problems in post-genomic research, it is now time to move on to a more comprehensible and integrated approach to designing these autonomous systems. It is necessary to consider the various component forms of automation from the point of view of their ultimate integration, otherwise we run the serious risk of devising components which will not fit together.

References

1 Newell, A. and Simon, H.A. (1976) Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19, 113–126

2 King, R.D. *et al.* (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252

3 Draghici, S. *et al.* (2003) Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto Translate. *Nucleic Acids Res.* 31, 3775–3781

4 Zupan, B. *et al.* (2003) GenePath: a system for inference of genetic networks and proposal of genetic experiments. *Artif. Intell. Med.* 29, 107–130

5 Yoo, C. and Cooper, G.F. (2004) An evaluation of a system that recommends microarray experiments to perform to discover gene-regulation pathways. *Artif. Intell. Med.* 31, 169–182

6 Karp, P.D. *et al.* (1999) Integrated pathway/genome database and their role in drug discovery. *Trends Biotechnol.* 17, 275–281

7 Kerr, M.K. and Churchil, G.A. (2001) Experimental design for gene expression microarrays. *Biostatistics* 2, 183–201

8 Karp, R.M. *et al.* (1999) Algorithms for choosing differential gene expression experiments. *3rd Annual Conf. Res. Comput. Biol.*, pp. 208–217

9 Ideker, T.E. *et al.* (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac. Symp. Biocomput.* 2000, 305–316

10 Tong, S. and Koller, D. (2001) Active learning for structure in Bayesian networks. *International Joint Conference on Artificial Intelligence*, Seattle, WA, USA, pp. 863–869

11 Meldrum, D. (2000) Automation for Genomics, Part One: Preparation for sequencing. *Genome Res.* 10, 1081–1092

12 Uber, D.C. *et al.* (1991) Application of robotic and image processing to automated colony picking and arraying. *Biotechniques* 11, 642–647

13 Panussis, D.A. *et al.* (1996) Automated plaque picking and arraying on a robotic system equipped with a CCD camera and a sampling device using intramedic tubing. *Lab. Robot. Autom.* 8, 195–203

14 Marziali, A. *et al.* (1999) An automated sample preparation system for large scale DNA sequencing. *Genome Res.* 9, 457–462

15 Warrington, J.A. *et al.* (2002) New developments in high throughput resequencing and variation detection using high density microarrays. *Hum. Mutat.* 19, 402–409

16 Vogt, G. (2002) Multi-axis robots bring automation to life sciences. *Ind. Robots* 29, 49–52

17 Meldrum, D. (2000) Automation for Genomics, Part Two: Sequencers, Microarrays, and Future Trends. *Genome Res.* 10, 1288–1303

18 Inza, I. (2004) Filter versus wrapper gene selection approaches in DNA microarray domains. *Artif. Intell. Med.* 31, 91–103

19 Walker, P.R. (2004) Data mining of gene expression changes in Alzheimer brain. *Artif. Intell. Med.* 31, 137–154

20 Gras, R. *et al.* (2003) Cooperative Metaheuristics for Exploring Proteomic Data. *Artif. Intell. Rev.* 20, 95–120

21 McClean, S. *et al.* (2003) Cenceptual Clustering of Heterogeneous Gene Expression Sequences. *Artif. Intell. Rev.* 20, 53–73

22 Futschik, M.E. *et al.* (2003) Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue.

Artif. Intell. Med. 28, 165–189

23 Maran, U. and Sild, S. (2003) QSAR modelling of genotoxicity on non-congeneric sets of organic compound. *Artif. Intell. Rev.* 20, 13–38

24 Karzynski, M. *et al.* (2003) Using a genetic algorithm and a perceptron for feature selection and supervised class learning in DNA microarray data. *Artif. Intell. Rev.* 20, 39–51

25 Domachowske, J.B. (2004) Microarrays and gene expression profiling in microbiology and infectious disease: A clinical perspective. *Clin. Microbiol. Newsl.* 26, 155–159

26 Carbonaro, M. (2003) Proteomics: present and future in food quality evaluation. *Trends Food Sci. Technol.* 15, 209–216

27 Peterson, C. and Ringner, M. (2003) Analysing tumor gene expression profiles. *Artif. Intell. Med.* 28, 59–74

28 Stetter, M. *et al.* (2003) Large-scale computational modelling of genetic regulatory networks. *Artif. Intell. Rev.* 20, 75–93