Review

From peas to "chips" – the new millennium of molecular biology: a primer for the surgeon

Iain Brown*, Steven D Heys and Andrew C Schofield

Address: Section of Surgical Oncology, Department of Surgery, University Medical School Buildings, University of Aberdeen, Aberdeen, UK, AB25 2ZD

Email: Iain Brown* - i.brown@abdn.ac.uk; Steven D Heys - s.d.heys@abdn.ac.uk; Andrew C Schofield - a.schofield@abdn.ac.uk

* Corresponding author

Introduction

The mechanisms underlying the basis of heredity and the beginning of understanding of the genetic basis of life began to be unravelled some 160 years ago. These fundamental concepts, which have paved the way for the current explosion in our understanding of the genetic basis of cellular function, were established from the study of pea plants by an Augustinian monk, Gregor Mendel. The next major development in genetics was 100 years later when Watson and Crick discovered the structure of DNA. Following on from their seminal work there has been an exponential growth in knowledge regarding the structure and function of DNA and its functional unit, the gene [1].

The study of DNA itself is just a broad overview of the human genome (genomics). When trying to understand more complex genetic-based traits and diseases, such as cancer, this is inadequate because it does not allow a thorough understanding of the complex inter-related processes occurring within the cell. In order to take this further, the functions of the individual genes, the messenger RNA resulting from the gene and the subsequent protein, which is produced, need to be examined. Furthermore, there are complex interactions between the cellular environment and the genes, which can affect genetic and cellular function.

The measurement of gene expression can, therefore, provide information on regulatory mechanisms, biochemical pathways, cellular control mechanisms and potential targets for intervention and therapy in a variety of disease states. One technique, which allows this to be studied, is DNA microarray technology, which is now used to moni-

tor the expression of thousands of genes simultaneously. This paper outlines briefly the applications, limitations and the possible future of microarray techniques in oncological research.

Gene expression

The gene sequences, which are contained in DNA, are transcribed into messenger RNA (mRNA). These mRNAs encode all the information required to synthesize proteins that are the cellular effector molecules and are hence coded for by DNA. Quantifying mRNA sequences presents difficulties, not least of which is that there may be extremely small amounts present within the cell. Furthermore, the mRNA molecule itself is very quickly degraded. Robust and sensitive techniques have been developed to allow an assessment of mRNA. The Reverse Transcription-Polymerase Chain Reaction (RT-PCR) was, until recently, the gold standard of mRNA expression analysis, which allowed the *de novo* synthesis of mRNA to be assessed using the DNA as the template on which mRNA could be formed. Thus having formed mRNA, the technique of real-time RT-PCR was then able to provide quantitative data on mRNA synthesis.

RT-PCR has many limitations, one of which is that it relies on specific primer sequences, termed probes. Furthermore, it is often only used to study one or several RNA messages, at best, at one time. The novel development of the DNA microarray technique in 1995 has altered the concepts and assessments of mRNA expression [2]. This method, which allows the analysis of thousands of genes, simultaneously, in one single experiment is a
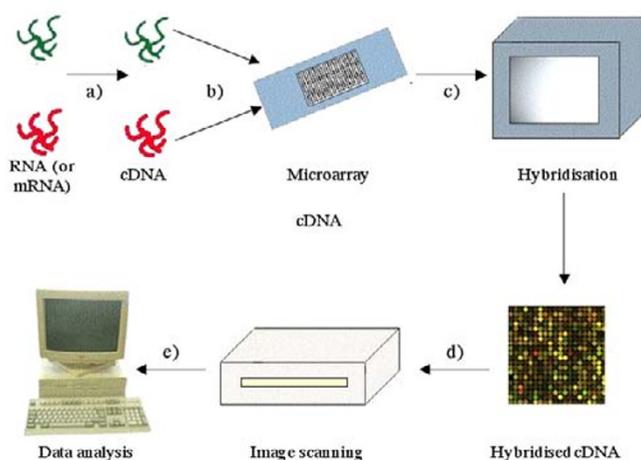
Figure 1
Simplified protocol for microarray analysis. a) RNA or mRNA from sample of interest is converted into cDNA. b) cDNA is applied to the microarray slide. c) Target cDNA and probes are hybridised under specific conditions. d) The completed chip is scanned and converted into raw data. e) Data is analysed by computer software. Different samples are indicated in red and green.

phenomenal development in molecular biological research methodology.

DNA microarray technology

A microarray consists of either complementary DNA (cDNA) arranged in a particular order onto glass slides or nylon membranes, and oligonucleotide arrays that comprise short DNA sequences (oligonucleotides) synthesized directly onto the slide. This slide is also termed a "chip".

The cDNA sequences, or oligonucleotides, correspond to genes, which may be previously identified or unidentified ones. RNA from biological samples, for example, blood, normal tissues or tumour samples, is used to create complementary cDNA. This is used to "probe" the arrays to determine if a specific gene is present. However in microarray terminology, the "probe" is actually the physically bound oligonucleotide or cDNA sequence [3].

The level of expression of each of the "probes" is determined by a specific detection method. Briefly, the bound target sequences are labelled with a (usually fluorescent or light producing) dye, or chemical, which can be detected visually. Once the targets have bound to the probes on the array, everything else is removed by washing. Scanning equipment is then used to produce a digital image of the signals produced, and these images are used for analysis.

Computer software packages are available which are used to determine the levels of expression of a particular mRNA, based on the strength of signal produced. This data can then be compared with that from different chips or samples. Statistical analysis is then used to determine the significance of any changes in gene expression taking place (Figure 1).

Applications of microarrays to oncological research

One of the initial applications of microarray analysis was in tumour classification and identification of tumour markers (oncogenomics). Microarray analysis initially revealed two previously unknown and distinct types of diffuse large B-cell lymphomas. One of these types had a better prognosis than the others in terms of survival [4]. Subsequently, the use of oligonucleotide arrays in ovarian cancers and tumour cell lines allowed a comparison with cells from the normal tissue of origin to be made [5]. It proved to be possible to identify groups of tumours on the basis of their genetic profile. Moreover, tumours were readily identifiable when compared with normal tissues. A subset of candidate genetic markers for the malignant process were identified for further study, for example, the *HE4* gene (a proposed ovarian tumour marker) and *CD24* gene (codes for a protein involved in breast cancer cell motility).

Microarray analysis has also been applied to the study of drug sensitivity and also in the identification of novel therapeutic agents (pharmacogenomics). Discovering, a target for a drug, identifying a compound suitable for that target and long-term clinical studies can all be furthered by using microarray analysis. For example, in the case of response to a known drug, cDNA microarrays were utilized to monitor the expression profiles (a particular pattern of genes being expressed) of breast cancer cell lines, which were either sensitive or resistant to doxorubicin. In this study, a distinct set of genes which were altered during treatment with doxorubicin and another subset of genes, were also identified, which were constitutively expressed in cells that were resistant to doxorubicin treatment [6]. Therefore, this opens up the possibility of *ex vivo* testing (before the commencement of any therapy) of a tumour biopsy to allow the identification of the appropriate chemotherapeutic agent for an individual patient.

The identification of interactions between therapeutic agents and genes (drug-gene interactions) has also utilized microarray technology. One such example was the examination of the effectiveness of 118 possible agents, which had anti-cancer activity, against 60 different cancer cell lines [7]. In particular, 78% of cell lines with low expression of the dihydropyrimidine dehydrogenase gene (DPYD) were more sensitive to 5-fluorouracil (5-FU). As

5-FU is commonly used for, and is one of the most effective agents used in the treatment of colon cancer the results of the study suggested that DPYD may have potential clinical use in patients with colon cancer.

Microarrays can also be used to elucidate complex biochemical pathways that occur *in vivo*. For example, oligonucleotide microarray technology was utilized to determine changes in gene expression between pre-adipocytes and adipocytes *in vitro* and *in vivo* [8]. A number of previously uncharacterised gene regulatory elements in the pathway *in vitro* were demonstrated. Furthermore, there was also a difference in gene expression between the *in vitro* and *in vivo* pathways. This may be of fundamental importance in understanding adipogenesis, which had previously only been understood at a more elementary level previously.

Analysis of mutations and polymorphisms is still a crucial part of understanding the mechanisms of disease, in particular malignant disease. Polymorphisms are actually just differences in DNA sequence at a particular location but occurring more frequently than can be attributed to their arising because of a mutation alone. Polymorphisms may have no effect on cellular function and may therefore be of no clinical consequence. However, sometimes they may be associated with disease and may be useful for tracing the progression of a disease-causing gene through families. Analysis of possible mutations may involve either determining the presence of previously characterized mutations, or alternatively, searching for all possible mutations in a sequence of DNA. One of the initial uses of screening for mutations using microarrays was the identification of all 37 known mutations in the cystic fibrosis (CF) gene, and in addition, this allowed the documentation of all the possible nucleotide substitutions [9].

Further studies have detailed the feasibility and accuracy of large-scale identification, mapping and genotyping of polymorphisms using microarray techniques [10]. Mutation analysis of the p53 gene, the most frequently mutated gene in human cancer, has also been improved by use of microarrays [11]. This study demonstrated an increased sensitivity and a more accurate detection of known mutations using microarray mutation analysis.

Application of microarrays to the clinical setting
As already discussed, microarrays have been used already to identify novel classes of B cell lymphomas [4]. Several studies have shown that microarray analysis can allow the identification of novel subtypes of breast tumours (two new subgroups of luminal epithelial/oestrogen receptor positive tumours), and predict subsequent clinical outcome. This may allow, therefore, targeting of therapy such

as adjuvant chemotherapy to be given to those patients who have the worst prognosis and are in most need of such treatment [12–14].

Microarrays have also been used to classify and predict prognosis for other cancers such as oesophageal, endometrial, and renal carcinoma. For example, the sensitivity of oesophageal tumours to chemotherapy could be given a response score based on the expression levels of a set of genes identified by microarray analysis. When applied to six unknown test samples, the response score correctly placed all tumours into the correct response groups [15]. Hepatocellular carcinoma could be categorized as either solid or pseudo-glandular types [16], and a revised classification of renal carcinoma was suggested by Higgins *et al* [17]. In particular, they identified distinct molecular expression profiles between usual renal cell carcinomas with granular cytoplasm, when compared with those with clear cytoplasm – which had been previously classed together as "conventional" carcinomas. One recent study has shown that microarrays could be used to identify particular expression profiles in patients with diffuse large B-cell lymphoma, which could predict their long term survival with 100% accuracy [18].

An extremely important area where clinicians are faced with diagnostic difficulties are with those patients who present metastatic disease from an unknown primary site. Palliative chemotherapy may prolong life but appropriate therapy is dependant on identification of the tissue of origin of the tumour. Arrays may prove to be useful tools in identifying the primary site of such metastases [19]. A recent study revealed that microarray analysis of the metastatic tissue resulted in correct identification of the primary site of origin in 81% of patients studied.

Limitations of microarray technology
One major disadvantage that occurred in the early years of microarray research was their financial cost. This resulted in the restriction of their use to well-funded larger laboratories, with these costs being beyond the funding scope of most academic research laboratories. In recent years, however, the costs of this technology has decreased as a result of advances in manufacturing technology and commercial competition to develop and make this technology available to as large a market as possible. Indeed, many research laboratories can manufacture their own arrays for relatively little cost. Spotted cDNA arrays, and arrays produced "in house" also have problems of standardization and consistency, with there being possible variations between experiments. Rigorous quality control is required to ensure that genetic changes, which are identified, are not simply due to defects and variations between arrays. Affymetrix™ oligonucleotide gene chips, however, have a multitude of internal controls within the arrays to account

for any variation between arrays (normalization) as well as controls for correct hybridisation of targets.

One of the main drawbacks of DNA microarray technology is that the levels of mRNA expression do not necessarily represent the levels of proteins in the cell. It is well recognized that proteins can also be altered by a variety of processes, which occur following transcription of DNA to mRNA or after translation of mRNA into its protein. This fact means that any interesting changes suggested by an array experiment must be further verified by RT-PCR or northern blot analysis (to ensure the altered expression does actually exist) and western analysis to determine changes in the protein levels.

It has become generally accepted that when using array analysis that a "real" change exists when there is an apparent 2-fold, or higher, change in gene expression. This means that smaller degrees of change, which may be just as important, will more often than not remain unrecognised, unless they were specifically looked for initially.

The complexity of microarray analysis means that tissue sample collection becomes a crucial factor in the data produced. As microarrays have been used with increasing frequency in recent years the amount of diversity in gene expression between samples, even from the same tissue in the same individual, has become clear. Precise sampling (including factors such as the time of day and month taken) and the ability to sample homogeneous tissue samples (by using techniques such as laser capture microdissection) [20], are crucial steps in obtaining an accurate analysis. At the same time, the amount of tissue required can also be a problem as a relatively large amount of RNA is required. However, more recent techniques to amplify RNA have been developed to allow extraction from minute tissue samples [21,22].

Figure 2
Future application of microarrays – Individualized Therapy. A. Currently, patients are likely to be given the treatment based on the best available drug at the time. B. Pre-treatment testing may allow the treatment to be tailored to a particular individual based on their gene expression profile
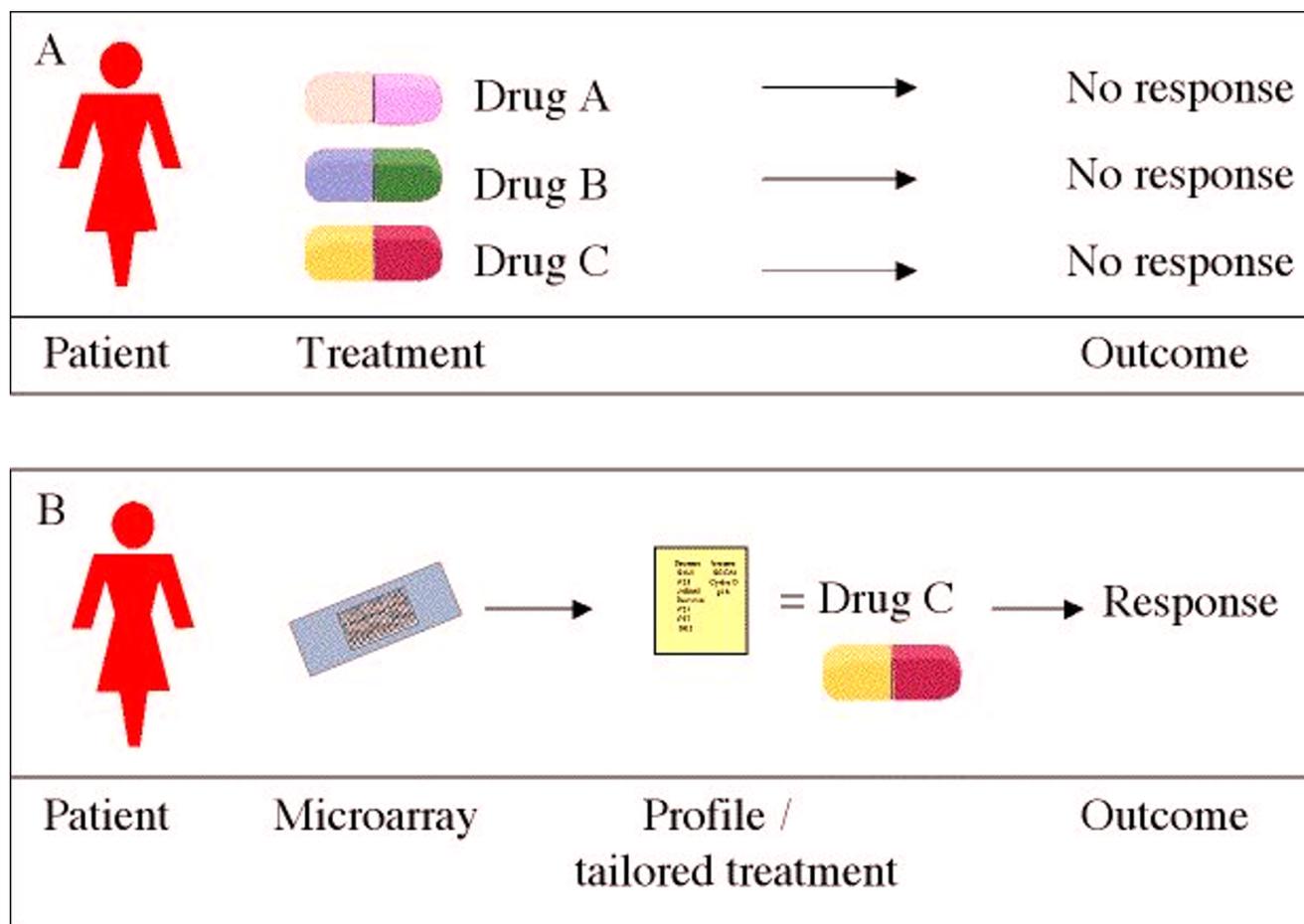
As arrays have become more and more complex, the data analysis of the results produced has also become more complicated such that it can take considerable time to analyse using powerful computers to produce the required data. Data acquisition and processing via a variety of statistical methods can identify unique patterns or profiles of gene expression. Once a set of genes of interest has been identified the ever-expanding public and commercial databases then have to be interrogated (data mining) to determine any proposed functional effects of these changes in expression. Complicated statistical algorithms and artificial intelligence networks are now used to "trawl" through the vast amounts of data, which may be produced from a single study [18,23].

The Future

As microarray technology becomes more advanced, arrays will be able to offer an increased ability to unravel complex disease processes and determine new targets for therapeutic interventions. Affymetrix™ now has made available a microarray chip, which contains the sequences of over 11,000 polymorphic sites in the genome [24]. In the past, researchers concentrated on one, or a small group of single nucleotide polymorphisms (SNPs), at any one time, because they were limited by the need to design primer sets for each SNP. Now, one array can give the information for one individual for thousands of SNPs, in one experiment.

Further developments in nanotechnology have enabled the production of the SmaSeq™, a single molecule array, which should allow the complete sequencing of an entire individual in one reaction and on one array chip [25]. The sequence would then be compared to a reference sequence and alterations recorded. This type of technology opens up the door to the possibility of individualized therapy where diagnosis and therapy of a disease will be specifically tailored to the individual patient (Figure 2).

Conclusions

Microarrays have revolutionised genetic and medical research over the last 10 years. Despite the initial limitations of variability and cost, microarrays are now more comprehensive and accurate and are easily available to most research laboratories. It is now possible to analyse thousands of genes at the same time in one experiment, and differences, or similarities, between individuals can be determined quickly and easily. As they are used more and more in clinical applications, microarrays will be adapted for diagnostic procedures and used to determine specific treatment regimens, tailored for individual patients.

References

1.   Watson JD and Crick FHC: **Molecular structure of nucleic acids.** *Nature* 1953, **171:**737-738.
2.   Schena M, Shalon D, Davis RW and Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270:**467-470.
3.   Phimister B: **Going global.** *Nature Genet* 1999, **21:**1.
4.   Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403:**503-511.
5.   Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA and Hampton GM: **Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer.** *Proc Natl Acad Sci U S A* 2001, **98:**1176-1181.
6.   Kudoh K, Ramanna M, Ravatn R, Elkahloun AG, Bittner ML, Meltzer PS, Trent JM, Dalton WS and Chin KV: **Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray.** *Cancer Res* 2000, **60:**4161-4166.
7.   Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO and Weinstein JN: **A gene expression database for the molecular pharmacology of cancer.** *Nature Genet* 2000, **24:**236-244.
8.   Soukas A, Socci ND, Saatkamp BD, Novelli S and Friedman JM: **Distinct transcriptional profiles of adipogenesis in vivo and in vitro.** *J Biol Chem* 2001, **276:**34167-34174.
9.   Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM and Miyada CG: **Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays.** *Hum Mutat* 1996, **7:**244-255.
10. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ and Lander ES: **Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.** *Science* 1998, **280:**1077-1082.
11. Wen WH, Bernstein L, Lescallett J, Beazer-Barclay Y, Sullivan-Halley J, White M and Press MF: **Comparison of TP53 mutations identified by oligonucleotide microarray and conventional DNA sequence analysis.** *Cancer Res* 2000, **60:**2716-2722.
12. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415:**530-536.
13. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P and Borresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci U S A* 2001, **98:**10869-10874.
14. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406:**747-752.
15. Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T and Nakamura Y: **Prediction of sensitivity of esophageal tumours to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles.** *Cancer Res* 2001, **61:**6474-6479.
16. Lee D, Choi SW, Kim M, Park JH, Kim M, Kim J and Lee IB: **Discovery of differentially expressed genes related to histological subtype of hepatocellular carcinoma.** *Biotechnol Prog* 2003, **19:**1011-1015.
17. Higgins JP, Shinghal R, Gill H, Reese JH, Terris M, Cohen RJ, Fero M, Pollack JR, van de Rijn M and Brooks JD: **Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray.** *Am J Pathol* 2003, **162:**925-932.
18. O'Neill MC and Song L: **Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect.** *BMC Bioinformatics* 2003, **4:**13.

19. Buckhaults P, Zhang Z, Chen YC, Wang TL, St Croix B, Saha S, Bardelli A, Morin PJ, Polyak K, Hruban RH, Velculescu VE and Shih IeM: **Identifying tumour origin using a gene expression-based classification map.** *Cancer Res* 2003, **63:**4144-4149.
20. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA and Liotta LA: **Laser capture microdissection.** *Science* 1996, **274:**998-1001.
21. Aoyagi K, Tatsuta T, Nishigaki M, Akimoto S, Tanabe C, Omoto Y, Hayashi S, Sakamoto H, Sakamoto M, Yoshida T, Terada M and Sasaki H: **A faithful method for PCR-mediated global mRNA amplification and its integration into microarray analysis on laser-captured cells.** *Biochem Biophys Res Commun* 2003, **300:**915-920.
22. Xiang CC, Chen M, Ma L, Phan QN, Inman JM, Kozhich OA and Brownstein MJ: **A new strategy to amplify degraded RNA from small tissue samples for microarray studies.** *Nucleic Acids Res* 2003, **31:**E53.
23. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS: **Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks.** *Nat Med* 2001, **7:**673-679.
24. **Affymetrix** [http://www.affymetrix.com/products/arrays/index.affx]. [Last accessed on October 10, 2003]
25. **Solexa** [http://www.solexa.co.uk/technology/overview.htm]. [Last accessed on October 10, 2003]