EJSO 2001; **27**: 504–508
doi:10.1053/ejso.2001.1116, available online at http://www.idealibrary.com on IDEAL®

EJSO
European Journal of Surgical Oncology

EDUCATIONAL SECTION

DNA microarray technology in cancer research

D. A. Rew

Royal South Hants Cancer Centre, Southampton University Hospitals, Southampton, SO14 0YG, UK

Microarray technology transforms the study of functional genetics. The entire genomic activity of cells and tissues can be analysed and compared on single slides, or gene chips. In cancer research, this will allow the better understanding of the regulation of activity of cells and tumours in various states. It will also allow the classification of individual tumours by their gene expression patterns, which may also describe and predict therapeutic resistance and sensitivity patterns. This short article provides a short introduction to the technology and its applications.

Key words: gene chip; microarray technology; cancer research.

INTRODUCTION

DNA chip technology is a remarkable advance for molecular biology and for cancer research. DNA microarrays and macroarrays, or Gene Chips, allow for the global assay of all genetic activity in the cell or sample in the same time frame. Thus, for example, the expression of all the known oncogenes, or all of the genes for drug resistance and metabolism in a cell can be detected and measured simultaneously.[1–3,W1]

A gene chip is founded upon a silicon or glass slide or nylon solid-state base.[4] The glass slide is a simple and convenient substrate whose adhesiveness for DNA is enhanced with a coating of polylysine or a silane. Upon the base is etched an orderly framework of genetic samples using a variety of physicochemical processes. Macroarrays comprize sample spot sizes of 300 microns diameter or more. Microarrays comprize sample spot sizes of less than 200 microns in diameter. Using such arrays, many thousands or tens of thousands of discrete spots can be accommodated on one standard slide.[W2] Each spot has known coordinates and defined characteristics, and is labelled with one of a selection of gene probes of known DNA base sequence. These can either be large probes of DNA complementary (cDNA) to genes of interest and up to 5000 nucleotide bases long, or short synthetic sequences of DNA (oligonucleotides) up to 25 bases long. The large probes are usually used for RNA expression analysis, while the short probes are used for both RNA and DNA sequence anaysis. The key structural aspects of probe attachment, densities and *in situ* synthesis on the array substrate are reported by Southern *et al.*[5]

The process which is the key to the utility of gene chips is the accurate binding, or hybridization, of strands of DNA with their precise complementary copies in experimental conditions where one sequence is also bound onto a solid state substrate. The chemistry in itself is not new, as hybridization techniques have been well established for several decades. It is the breadth and scale of the simultaneous assays which transforms the technology from a single gene to a whole system analytical tool. Gene chips now enable the simultaneous detection and semiquantitative assay of thousands of genes from the human genome or from any other research organism with great sensitivity.

PRODUCTION OF MICROARRAY CHIPS

Chips are manufactured by a variety of 'spotting' and photolithographic techniques. Spotting is the physical placement of molecules on the chip under precise

Corresponding author: David A. Rew TD MA MB MChir FRCS, Consultant Surgeon and Honorary Senior Lecturer, Royal South Hants Cancer Centre, Southampton University Hospitals, Southampton, SO14 0YG, UK.

0748–7983/01/050504+05 $35.00/0

positional control. Photolithography allows the *in-situ* high density spatial synthesis of oligonucleotides on the slide substrate, such that up to 400 000 distinct sequences can be produced on one slide.[2] This is akin to semiconducter chip manufacture, whereby layers of nucleotides (or amino acids in the case of peptide chips) are built up using a series of chemical reactions, mask and washings. Commercial chips can be purchased to pre-defined profiles.[6,7] The enthusiastic and well-resourced laboratory can manufacture chips to its own specifications.[3,4]

DETECTION OF GENES ON THE CHIP

Binding of target sequences of DNA or RNA can be detected by a variety of methods, including radiolabels (such as 32P and 33P linked to dCTP) and red and green fluorescent dyes linked to the nucleotides (such as Cye-3 or Cye-5 to dUTP). Fluorescence offers considerable advantages over radioisotopes. Fluorescence measurements are convenient, immediate, quantitative and very flexible. There is a wide range of available dyes with precisely known excitation and emission spectra. Laser-based instruments allow precise, rapid and reproducible data collection and analysis. The computer-driven laser scanning microscope measures and records the fluorescence intensity at each and every spot (and hence gene marker) on the chip.

Fluorescence also allows the gene expression in two different samples to be compared directly. If each sample is labelled with a dye of different emission wavelength, and the samples are mixed and applied to the chip, then the ratio of the intensity of fluorescence on each spot is the ratio of expression of each gene in the two samples.[3] Thus, for example, the expression of each and every gene in a normal cell and a derivative cancer cell can be compared directly.

Genomic maps

The informative use of data generated from arrays presents new challenges. Conventional genetic and chromosomal maps describe the position of genes on chromosomes. This format does not describe the simultaneous, quantitative or relative expression of large numbers of genes. For this, functional maps are more helpful. For example, Brown and Botstein[3] describe a tabular model for comparing large numbers of whole genome observations on the yeast cell in different experimental conditions in which many thousands of measurements can be conveniently displayed. This model uses colour coding of the relative intensity of expression of each gene. It clusters groups of genes in hierarchies according to their similarities and differences in expression in the various experiments, as for example, in relation to cell cycle regulation. Such tabulations allow unforeseen correlations to be made between the expression of genes in various cell states and molecular processes. The classification of multiple genes by function, suggested by Lander as akin to a biological periodic table, may help this functional interpretation.

THE UTILITY OF GENE CHIPS

cDNA and oligonucleotide arrays yield 'static' information on the cells and tissues in which particular genes are expressed, and 'dynamic' information on the expression of genes relative to others in time and place. Genes appear to be switched on only where and when their protein product makes a contribution to the activity of the cell. Gene expression is a highly sophisticated and interactive process of multiple, rapid and dynamic changes in time and space.

Microarray studies of gene expression describe the gene by gene totality of genetic activity taking place in the cell at the time of study. The technology thus finds many new applications. It allows for the study of the normal functioning of cells in any number of physiological or morphological states. This is achieved through the analysis of those genes which are switched on and hence producing mRNA, the precursor to the protein products which determine cell form and function. It allows for the study of disease states, such as the total expression of genes in a cancer or in response to an environmental insult, and comparison with the expression in normal or unchallenged cells. It allows for the detection of a wide range of specific mutations in a small number of critical genes in individual patients and hereditary diseases. It allows for the rapid determination of infectious organisms or disease states against a standard profile of genes. It allows for the genotypic characterization of individuals. It provides for a host of new studies across the spectrum of human, animal, plant and pathogenic research, drug discovery, toxicology and forensics.

THE APPLICATIONS

Expression profiling

The structure and function of cells and tissues is determined at all times by the selective, differential and collective expression of many genes, translated into RNA and transcribed into proteins. Microarrays will allow for the simultaneous measurement of all RNA, and hence the active genes, in any one state of activity and differentiation. Such a pattern of gene activity is an expression profile.[8] There are many possible applications. For example, they allow the direct comparison of genetic activity in diploid and aneuploid cancer cells with their normal counterparts in any tissue.[9]

Microarray assays in cancer studies

Microarray technology allows for the detailed genomic comparison of each and every tumour class and even renders conceivable the economic description of each individual tumour and metastasis in genomic variance terms to complement conventional histopathology reports.[10] The masses of microarray and genomic data generated by public and commercial research lend themselves to storage and interrogation on the Internet. The US National Cancer Institute (NCI) has developed the Cancer Genome Anatomy Project to collate and publish complete libraries of complementary DNA sequences for the major cancers.[W3]

Mutation detection

Detailed study of many genes associated with human diseases reveals large numbers of specific mutations in each gene. Thus, for example, over 400 specific mutations have been found in the BRCA1 gene in different individuals.[11] Mutations in a known target gene DNA can be detected by creating arrays of oligonucleotides of up to 25 A, C, T and G nucleotides and with large numbers of single nucleotide variations, thus allowing rapid and precise identification of mutations in individuals, families and populations. This process could in theory be scaled up to seek out mutations or variants simultaneously in large fragments of the genome.

Gene copy number analysis

Many tumours contain abnormal numbers of genes through processes such as chromosomal and gene deletion, amplification and aneuploidy. Quantifying the number of gene copies per cell can be helpful in understanding a disease process. Pollack *et al.*[12] describe an elegant application of microarray technology to study ErbB2 expression in breast tumour cell lines using two colour fluorescence. The DNA fragments from the test sample and labelled with (red) fluorescent dye, and mixed with DNA fragments from normal, control cells labelled with a different (green) dye. The binding ratio (i.e. copy number) at each spot on the microchip is accurately reported by the fluorescence ratio measured by confocal (high definition) laser scanning microscopy. cDNAs thus allow simultaneous study of copy number changes across the entire spectrum of genes in any one sample, including clinical syndromes such as Turner's (XO).

Cancer therapeutics

Human cancer cell lines form the basis of many therapeutic studies. Ross *et al.*[13] have recently described the pattern of expression of some 8000 genes in each of the 60 tumour cell lines used in the NCI screen for anticancer drugs. Each tumour type and its tissue of origin expressed distinctive gene subsets ('molecular signatures') which helped correlate genotype with known function. Thus, for example, melanoma cell lines prominently expressed genes associated with melanin metabolism. These patterns may help classify the lineages of complex clinical samples such as poorly differentiated tumours and metastases of unknown primary origin.

Perou *et al.*[14,W4] report an extension of this study wherein 65 surgical breast specimens from 42 patients, including primary and metastatic samples, and from patients before and after 18 weeks of doxorubicin treatment, were analysed and compared for the expression of these 8000 genes. The paired tumour samples were reported as being more similar to each other than to any other tumour. Tumours could be subclassified according to differences in their gene expression patterns. The data is given in detail on Wx.

Drug sensitivity

Microarrays offer a number of new approaches to drug development and therapeutic efficacy. Many genes in any one cell affect its reponse to therapeutic agents, in relation to drug exclusion, resistance, metabolism and DNA repair, for example. The simultaneous study of the expression of these genes may thus allow for the characterization of each cell, tissue and tumour in terms of the mechanism of drug action, its predicted drug sensitivity and side-effect profile to a wide spectrum of reagents. For example, Scherf *et al.*[15] have analysed a subset of 1400 genes from the study reported by Ross *et al.*[13] in respect of the mechanisms of action of 118 selected from 70 000 well-characterized anticancer drugs for patterns of correlation between gene expression and drug action. This study demonstrates that it is possible to integrate and interrogate massive data sets on drug action and microarray generated genomic information, although no clinically applicable findings have yet resulted.

The expression of particular gene patterns in particular tissues, as for example cathepsin K expression in osteoclasts, or inflammation associated genes in rheumatoid tissue,[16] may identify possible treatment strategies. The testing of drugs in model organisms such as the yeast and fruit fly will prove more informative now that patterns of gene response in challenged cells can be accurately characterised in relation to human homologues.

SUMMARY

Microarray technology creates new problems and challenges in experimental design and control, and in data handling and interpretation. The design, acquisition and interpretation of the massive amounts of data from gene chips is addressed by the science of bioinformatics. Computers and mathematical algorithms are essential to

the interpretation of sequence data. The assignation of previously unknown genes to functions and to one or more functional groups can be improved using self-learning computer algorithms.[17]

Clinicopathological studies in cancer research, in cell and molecular biology have tended to address one or a few variables or biomarkers at a time. We may expect the principle of arrays to be more widely applied, as to huge panels of proteins, antibodies and cells. Given the economy of conservation of genes and their functions in evolution, the use of gene chips to study model organisms such as bacteria, yeast, nematode worm, fruit fly and mouse is likely to allow rapid progress to the definition of gene functions across the entire human genome in normal and disease states (Brown and Botstein 99).

The place of the technology in the armamentarium of molecular and genetic techniques is not yet fully defined. The assessment of its likely contribution to clinical medicine and oncology is as yet purely speculative. However, the remarkable concurrence of technologies which make microarray systems possible can be expected to invigorate cancer research for decades to come.

TABLE OF TERMS AND DEFINITIONS

Gene chip: A glass slide or nylon substrate on which has been a pattern of timy spots, each containing a genomic DNA or RNA fragment of known characteristics for binding to a test sample.

Microarray and macroarray: The pattern and scale of spots on the gene chip.

Spotting: The process of synthesizing gene chips by accurate placement of tiny samples (spots) of DNA or RNA on gene chips in patterns with known coordinates, which can be identified by measuring technologies.

Confocal Laser Scanning Microscopy/Scanning Arrayer: The technology for rapid and accurate analysis of gene chips. A fine, computer controlled laser beam is moved from spot to spot on the slide. The fluorescence resulting from the binding of fluorochromatic dye tagged test samples to the hybridization spots allows rapid quantitative recording of findings.

Hybridization: The process of precise binding between complementary sequences of DNA and RNA through specifc pairing of the Adenosine-Thymine or Uracil and Cytosine-Guanine nucleotide bases.

Photolithography: The process of synthesizing complex patterns of many thousands of different oligonucleotides of up to 25 bases long, layer by layer on a gene chip, using chemical masks akin to the manufacture of silicon chips in computer semiconductor technology.

Bioinformatics: The science of documentation and analysis of huge data sets resulting from modern molecular biology techniques, including genome sequencing and gene chips.

Oligonuceotide: A short synthetic length of defined sequence of nucleotides.

dUTP: deoxyuridine triphosphate.

dCTP: deoxycytidine triphosphate.

Key Learning Points

1. Microarray technology allows the simultaneous study of the expression of many thousands of genes or their RNA products.
2. These patterns can be revealed using fluorescence chemistry and displayed in tables known as expression profiles.
3. Through economy in nature, genes are only switched on when their products are needed by cells. Thus, an expression profile give an accurate picture of the pattern of gene activity in the cell at the time of study.
4. DNA or RNA expression profiles can be generated from a wide range of biological samples, including human tumours, and in a wide range of experimental physiological and developmental states.
5. Tissues and tumours can be classified according to their gene expression profiles. Preliminary studies indicate a close correlation between the observed characteristics of tissues and the expression of genes of known association.
6. The treatment sensitivity of tumours may be predictable from the expression profiles of genes associated with drug resistance and metabolism.
7. The mass of data generated by microarray technology mandates new systems of data analysis and interpretation: the science of bioinformatics.
8. Microarray technology promises to be economic for widespread clinical and oncological research applications.

REFERENCES

1. Collins FS. Microarrays and macroconsequences. *Nature Genetics* (Suppl) 1999; **21**: 2.
2. Lander ES. Array of hope. *Nature Genetics* (Suppl) 1999; **21**: 3–4.
3. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nature Genetics* (Suppl) 1999; **21**: 33–7.

4. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nature Genetics* (Suppl) 1999; **21**: 15–24.
5. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. *Nature Genetics* (Suppl) 1999; **21**: 5–9.
6. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density oligo-nucleotide arrays. *Nature Genetics* (Suppl) 1999; **21**: 20–4.
7. Bowtell DDL. Options for obtaining expression data by microarray. *Nature Genetics* (Suppl) 1999; **21**: 25–32.
8. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nature Genetics* (Suppl) 1999; **21**: 10–4.
9. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genetics* 2000; **25**: 333–7.
10. Cole KA, Krizman DB, Emmert-Buck MR. Review: The genetics of cancer – a 3D model. *Nature Genetics* (Suppl) 1999; **21**: 38–41.
11. Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics* (Suppl) 1999; **21**: 42–7.
12. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genetics* 2000; **23**: 41–6.
13. Ross DT, Scherf U, Eisen MB *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics* 2000; **24**: 227–35.
14. Perou CM, Serlie T, Eisen MB *et al.* Molecular portraits of human breast tumours. *Nature* 2000; **406**: 747–52.
15. Scherf U, Ross DT, Waltham M *et al.* A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* 2000; **24**: 236–44.
16. Debouck C, Goodfellow PM. DNA microarrays in drug discovery and development. *Nature genetics* (Suppl) 1999; **21**: 48–50.
17. Gaasterlaand T, Bakiranov S. Making the most of microarray data. *Nature Genetics* 2000; **24**: 204–6.

Web references (current, July 2000)

W1. www.Gene-Chips.com Leming S. DNA Microarray: monitoring the genome on a chip. An extensive non-commercial resource of references on microarrays.
W2. www.affymetrix.com Affymetrix was the first company to commercialize microarray technology.
W3. www.ncbi.nlm.gov/ncicgap The US National Cancer Institute's Cancer Genome Anatomy project.
W4. http://genome-www.stanford.edu/molecularportraits Extended data sets to the paper by Perou *et al.*