



**Australian Government**  
**Department of Health and Ageing**



Australia and New Zealand Horizon Scanning Network

**ANZHSN**

AN INITIATIVE OF THE NATIONAL, STATE AND  
TERRITORY GOVERNMENTS OF AUSTRALIA  
AND THE GOVERNMENT OF NEW ZEALAND

# **Horizon Scanning Technology Emerging Technology Bulletin**

## **DNA Microarrays**

### **August 2007**



© Commonwealth of Australia 2007

Online ISBN 1-74186-406-2  
Publications Approval Number: P3-2494

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation. Apart from any use as permitted under the Copyright Act 1968, all other rights are reserved. Requests and inquiries concerning reproduction and rights should be addressed to Commonwealth Copyright Administration, Attorney General's Department, Robert Garran Offices, National Circuit, Canberra ACT 2600 or posted at <http://www.ag.gov.au/cca>

Electronic copies can be obtained from <http://www.horizonscanning.gov.au>

Enquiries about the content of the report should be directed to:

HealthPACT Secretariat  
Department of Health and Ageing  
MDP 106  
GPO Box 9848  
Canberra ACT 2606  
AUSTRALIA

**DISCLAIMER:** This Bulletin is based on information available at the time of research and cannot be expected to cover any developments arising from subsequent improvements to health technologies. This bulletin is based on a limited literature search and is not a definitive statement on the safety, effectiveness or cost-effectiveness of the health technology covered.

The Commonwealth does not guarantee the accuracy, currency or completeness of the information in this Bulletin. This Bulletin is not intended to be used as medical advice and it is not intended to be used to diagnose, treat, cure or prevent any disease, nor should it be used for therapeutic purposes or as a substitute for a health professional's advice. The Commonwealth does not accept any liability for any injury, loss or damage incurred by use of or reliance on the information.

The production of this *Emerging Technology Bulletin* was overseen by the Health Policy Advisory Committee on Technology (HealthPACT), a sub-committee of the Medical Services Advisory Committee (MSAC). HealthPACT comprises representatives from health departments in all states and territories, the Australia and New Zealand governments; MSAC and ASERNIP-S. The Australian Health Ministers' Advisory Council (AHMAC) supports HealthPACT through funding.

This *Emerging Technology Bulletin* was prepared by Ms Linda Mundy, A/Prof Annette Braunack-Mayer, and Professor Janet Hiller from the National Horizon Scanning Unit, Adelaide Health Technology Assessment (AHTA), Discipline of Public Health, Mail Drop 511, University of Adelaide, Adelaide, South Australia, 5005.

## Table of Contents

---

Executive Summary .....	1
Introduction .....	4
Background .....	5
<i>Microarray production</i> .....	5
<i>How DNA microarrays work</i> .....	8
<i>Reliability issues associated with DNA microarrays</i> .....	12
<i>Data analysis of microarray results</i> .....	13
Comparators .....	15
DNA microarrays and breast cancer .....	18
<i>DNA microarray breast cancer prognostic tests</i> .....	22
DNA microarrays and drug metabolism .....	32
Other applications of DNA microarrays .....	39
<i>Human papillomavirus (HPV)</i> .....	39
<i>Familial hypercholesteraemia</i> .....	40
<i>Blood typing</i> .....	41
<i>Burkitt's Lymphoma</i> .....	42
<i>Single nucleotide polymorphisms</i> .....	43
<i>Miscellaneous</i> .....	44
Potential Cost Impact .....	45
Ethical Considerations.....	46
Sources of Information.....	49
<i>Internet sites of interest</i> .....	50
<i>Clinical trials</i> .....	52
Conclusions .....	55
Glossary of Terms .....	58
Appendix A: Levels of Evidence .....	60
Appendix B: An example of a DNA microarray.....	62
Appendix C: Microarray companies .....	63
Appendix D: Australian Research.....	66
Appendix E: HTA Internet Sites .....	67
References .....	70

## Tables

---

Table 1	Comparative costs of microarrays.....	6
Table 2	Breast cancer patients using 70-gene panel .....	20
Table 3	MammaPrint® .....	25
Table 4	Breast cancer patients using 76-gene panel .....	28
Table 5	76-gene panel .....	29
Table 6	Effects of CYP enzyme genes on drug metabolism.....	33
Table 7	Genotyping of CYP2D6 and CYP2C19 genes .....	37
Table 8	HPV genotyping with DNA microarrays.....	40
Table 9	Search terms utilised .....	49
Table 10	Literature sources utilised in assessment .....	50

## Figures

---

Figure 1	Example of a DNA microarray .....	6
Figure 2	Microarray production techniques .....	7
Figure 3	Schematic of a DNA microarray hybridisation .....	9
Figure 4	Two-dimensional gene space of DLBCL vs FL. ....	15
Figure 5	Central dogma of molecular biology .....	16
Figure 6	Incidence of breast cancer, Australia, 1988-2002 .....	18
Figure 7	Patient criteria for 70- or 76-gene microarray .....	23
Figure 8	Expression of HPV SNPs.....	39
Figure 9	Example of DNA microarray.....	62

## Executive Summary

---

The human genome consists of approximately  $2.8 \times 10^9$  nucleotide base pairs which may encode up to 25,000 genes. However not all of these genes are “turned” on, with only a subset being expressed at any one time. Not only can gene expression be turned on and off in response to stimuli, the amount or volume of gene expression can either be up or down regulated. The degree of gene expression may give previous unknown insights into the course of disease, and may guide treatment strategies.

DNA microarrays are a relatively new technology, which were first described in 1995. Small fragments of DNA, or oligonucleotides, are attached to a glass substrate by a variety of manufacturing processes. DNA microarrays may hold 100,000s of these elements. Complementary DNA from test and reference samples are labelled with a visualisation tag such as a fluorophore and hybridised to the microarray. The intensity of the signal of the bound DNA is an indication of a given gene’s activity. Active genes will give a more intense, brighter signal than less active genes. Due to the large number of target elements able to be bound to a DNA microarray, the rapid, simultaneous characterisation of thousands of genes is possible. DNA microarrays may be useful tools in categorising diseases such as cancer by determining the presence or absence of particular genes, which may provide important biological, diagnostic and prognostic information.

This *Emerging Technology Bulletin* is aimed at providing a *non-systematic* overview of the rate of progress and development of DNA microarrays in Australia. It is not a definitive or comprehensive assessment of the safety and effectiveness associated with DNA microarrays. Although a great deal of research is conducted in Australia using DNA microarrays, no clinically orientated papers were identified for inclusion in this *Bulletin* that were published in Australia.

One of the significant areas in which the use of DNA microarrays has created great interest is in the detection of prognostic markers in breast cancer. Lymph node status still remains the best *prognostic* marker for survival, with 50 and 25 per cent of lymph node positive and negative women experiencing recurrence of disease, respectively. It has been estimated that up to 85 per cent of node-negative women may be undergoing toxic chemotherapy needlessly. It is hoped that with the use of DNA microarrays, women with a good prognosis, that is, those likely to be at low-risk of developing recurrence, can be easily identified and therefore avoid unnecessary adjuvant chemotherapy.

There are two breast cancer predictive DNA microarrays in current use. The MammaPrint<sup>®</sup> is the only FDA approved assay and utilises a 70-gene expression panel. The other assay, the 76-gene “Rotterdam” signature panel, is not a commercially available product. The MammaPrint<sup>®</sup> assay can be accessed by Australian women, at full cost to the patient, for A\$3,600. Predictive DNA microarray studies indicate that both the 70-and 76-gene panels are capable of stratifying breast cancer patients into groups that either have a low-risk (good prognosis) or high-risk (poor prognosis) for the

development of distant metastases, which will impact on their overall survival rate. Reported hazard ratios indicate that lymph node-negative women in the high-risk category are 2-5 times more likely to experience disease recurrence, even when adjusted for clinical prognostic factors. In addition, from this stratification it can be seen that a larger proportion of women considered to be at low risk of recurrence survive at 10-years than those at high risk. Only the studies conducted with the 76-gene panel reported on the sensitivity and specificity of using DNA microarrays, compared to the gold standard of bi-directional DNA sequencing, to predict recurrence of breast cancer. Reported sensitivities were good, ranging from 90-97 per cent for the risk of developing distant metastases within 5-years. However, reported specificities were poor, ranging from 31-48 per cent. Although the arrays successfully identifies those patients at high-risk of recurrence, poor specificity indicates that a number of low-risk individuals are incorrectly categorised as high-risk and will receive chemotherapy needlessly.

All of the predictive breast cancer gene expression studies included for assessment in this *Bulletin* were retrospective. To date no prospective, randomised clinical trials, assigning patients to chemotherapy regimes based on the results of gene expression assays, have been published. There are no data available which describe the impact of the 70-gene panel on patient outcomes (avoidance of toxic chemotherapy, disease-free survival and overall survival) by the identification of high (those who will benefit from adjuvant chemotherapy) and low risk women (those unlikely to benefit from adjuvant chemotherapy). To address these concerns, the MINDACT trial, is currently underway and actively recruiting node-negative women. This study is a prospective, randomised controlled trial, however results are not expected until at least 2011 due to the need to assess 5-year outcome measures.

Another area of great interest is the use of DNA microarrays to predict drug metabolism. Cytochrome P450 enzymes catalyse the oxidation of over 80 drugs and are encoded by a group of over 100 CYP genes. These genes are highly polymorphic, which may result in either an increase or decrease in enzyme activity. This in turn may affect the rate of drug metabolism and individuals may be classified as either an ultra-fast, extensive, intermediate or poor metaboliser of drugs. In 2004, the FDA approved the only DNA microarray (Affymetrix AmpliChip™ cytochrome P450 test) to investigate single nucleotide polymorphisms in two genes of particular interest: the CYP2D6 and CYP2C19 genes. Patients are assigned as either poor, intermediate, efficient or ultra-rapid metabolisers of drugs according to the combination of CYP alleles expressed.

AmpliChip™ appears to be an accurate means of genotyping patients into poor, intermediate, efficient and ultra-rapid metabolisers of drugs. High sensitivity (>99%) and specificity (>99%) values when the AmpliChip™ is compared to the gold standard indicate that the test can successfully identify individuals who are poor metabolisers. However, there is a lack of clinical data concerning whether or not genotyping patients has an impact on patient outcomes. No studies have been published that report on improvements in outcomes of patients on selective serotonin reuptake inhibitors (SSRIs),

prescribed for non-psychotic depression, who have been genotyped versus those of patients on SSRIs who have not been genotyped.

DNA microarrays are a powerful research tool for the identification of association signals, as demonstrated by the large study conducted by the Wellcome Trust Consortium. This study has identified single nucleotide polymorphisms as being associated with some of the major public health diseases including coronary artery disease and diabetes. There is no indication of the attributable risk associated with these genes ie the proportion of all instances of these diseases that is association with the identified single nucleotide polymorphisms. There are many other commercially available DNA microarrays (with European CE-marking *not* FDA approval) that are available for both human testing as well as pure research purposes.

“In-house” microarrays may be custom made and produced for \$200, however an initial outlay of approximately \$350,000 is required for basic equipment including a robotic spotter and laser scanner. Most Affymetrix microarrays require that they be processed using equipment purchased from Affymetrix. Costs for a complete microarray processing system are approximately A\$335,000. Commercial gene expression arrays cost between A\$270 -550 depending on the complexity of the target. The cost per array increases as the number of oligonucleotides per array increases.

In summary, although the possibilities for using DNA microarrays are vast there is a paucity of information regarding the use of gene expression data to give meaningful clinical outcome data. DNA microarrays have dramatically altered the ability to assess differences in gene expression on a mass, high-throughput scale. However, a great deal of work still needs to be done to translate knowledge gained from gene expression into meaningful patient outcomes.

## Introduction

---

As this *Emerging Technology Bulletin* is intended for the use of health planners and policy makers, it will focus on the current state of development, its present use and the potential future application of DNA Microarrays to medical devices and health.

The National Horizon Scanning Unit, Discipline of Public Health, University of Adelaide, on behalf of the Medical Services Advisory Committee (MSAC), has prepared an *Emerging Technology Bulletin* to provide advice to the Health Policy Advisory Committee on Technology (Health PACT) on the state of play of the introduction and use of DNA microarrays in Australia and New Zealand.

The field of DNA microarrays is rapidly developing and diffusing across many medical specialties. It is beyond the scope of this *Emerging Technology Bulletin* to present either a definitive picture of all areas of DNA microarrays under development, or a comprehensive assessment of the safety, effectiveness and other considerations associated with its use. This *Emerging Technology Bulletin* is aimed at providing a *non-systematic overview* of the rate of progress or development of DNA microarrays, its applications and its potential clinical impact. It is primarily a briefing document, as opposed to a critical evidence-based preliminary assessment, as would be the case with an Horizon Scanning Report.

There is currently great interest in the field of DNA microarrays and the implications this technology will have for the future.

DNA microarrays, or gene-chips are a relatively new technology, first described by Schena et al (1995) in the seminal *Science* paper *Quantitative monitoring of gene expression patterns with a complementary DNA microarray* (Schena et al 1995). Microarrays were developed at Stanford University to study plant gene expression, where researchers were isolating large numbers of plant proteins and factors transcribed by genes and seeking a more rapid method to characterise their functions. The concept of utilising microscopic DNA arrays (microarrays) attached to a glass substrate was proposed, building on knowledge gathered in the 1980s from DNA hybridisation experiments conducted on glass. However, this simple concept created a complex technological problem which required input from the disciplines of biology, chemistry, physics, engineering, mathematics and computer science for the idea to become a reality. Despite massive innovation in the field of microarrays, the basic principles have remained the same (Schena 2003). To put the use of DNA microarrays and their potential for clinical use in context, it should be remembered that the human genome consists of approximately  $2.8 \times 10^9$  nucleotide base pairs which may encode up to 25,000 genes. However not all of these genes are “turned” on, with only a subset being expressed at any one time. Not only can gene expression be turned on and off in response stimuli, the amount or volume of gene expression can either be up or down regulated (Berrar et al 2003).

### *Microarray production*

There are two approaches to microarray preparation, known as *delivery* and *in situ synthesis*. The delivery approach prepares the target molecules (DNA of varying length) ‘off-line’ utilising techniques such as cloning and polymerase chain reaction (PCR), with the final target products delivered onto the glass microarray surface by contact printing, the most advanced example of this being ink jetting technology (Figure 2c) (Schena 2003). The advantages of the delivery approach is that it is relatively easy to manufacture a microarray, does not require complex, expensive machinery and is therefore a method used in many research laboratories. The diversity of microarrays produced using this method is huge, limited only by the types of molecules (cDNA<sup>1</sup> clones, PCR products and synthesised oligonucleotides between 5-120 bases in length) that can be produced off-line. Benefits of the delivery method include the capability of producing microarrays with a low to medium density of target elements in the range of 10-10,000 elements per cm<sup>2</sup> which are relatively cheap to produce, however the cost of high-complexity microarrays, with 10,000-500,000 elements, increases in a linear fashion (Table 1). In addition, sample tracking may be difficult using this technique (Schena 2003).

---

<sup>1</sup> cDNA = complementary DNA, see glossary for all definitions.

**Table 1 Comparative costs of microarrays produced by delivery or synthesis methods**

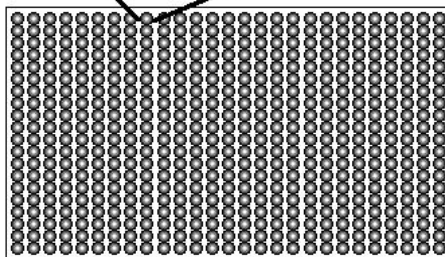
	Low complexity (100-10,000 elements)	High complexity (10,000-500,000 elements)
<b>Delivery production</b>	\$50-\$5,000 per chip	\$50-\$10,000 per chip
<b>Synthesis production</b>	\$1000-\$50,000 per chip	\$100-\$10,000 per chip

Prices quoted in \$US in 2003 (Schema 2003)

The first cDNA microarrays produced by the researchers at Stanford are an example of the delivery approach (Schema et al 1998). cDNA microarrays involve microscopic spots of DNA derived from different genes spotted in a grid-like array on glass. The molecules of cDNA originate from individual clones produced by bacterial plasmids. The cDNA molecules are amplified using PCR to a typical length of 200-500 base pairs, purified and then spotted systematically onto a glass slide using either mechanical microspotting or ink jetting (Haviv & Campbell 2002; Schema et al 1998). The DNA spots may range in diameter from 50-150  $\mu\text{m}$ , with 150-200  $\mu\text{m}$  spacing between them. As a result these microarrays may contain hundreds or thousands of genes, however at the current time they are unlikely to contain the entire human genome (Figure 1) (Haviv & Campbell 2002).

```

TCCTTTCCGG AACGGTTGGC GTCTGGGCAC GCGGTTSTGG GGCATGACAT
GCCGCCCCAG GAACAACCCC GACACGGCTT TAAGCCTCTC AAATCGCTGT
AGACATCATC TTTACGTGCT TGCTTTGCCC TCCACCATTT AGGGCTGTTC
CGCGGACGAC TCGCCATTCA ACCTCAGTCC TTCCGGTTGA GCGAGTGGGT
CGCGGCCAAG GTGCGAATGG GTGCGCGGCA AAGTGTTCGG CTGGCTGTAT
TATATGCTTC CTATAGCGAG ACTAACGACC CACACTTTCA CACAAGGATT
TCCCGCTAAT GGTACCTCG CGTCAGGACC TTGACGCAAG CCGGCTTCG
GTTGGCCCA AGCTTGCTAG GACTACTTAT CTTGAGCTCA TTTAACATCC
CGCGGCTCT CCGGAGCGG TCGTCGGAA GAAGTCAAAC CCGAACGGC
GTTGACAAAAG CGTGGAGACA TCGATACCTC TGTGTAGCG GCCACAAATC
    
```

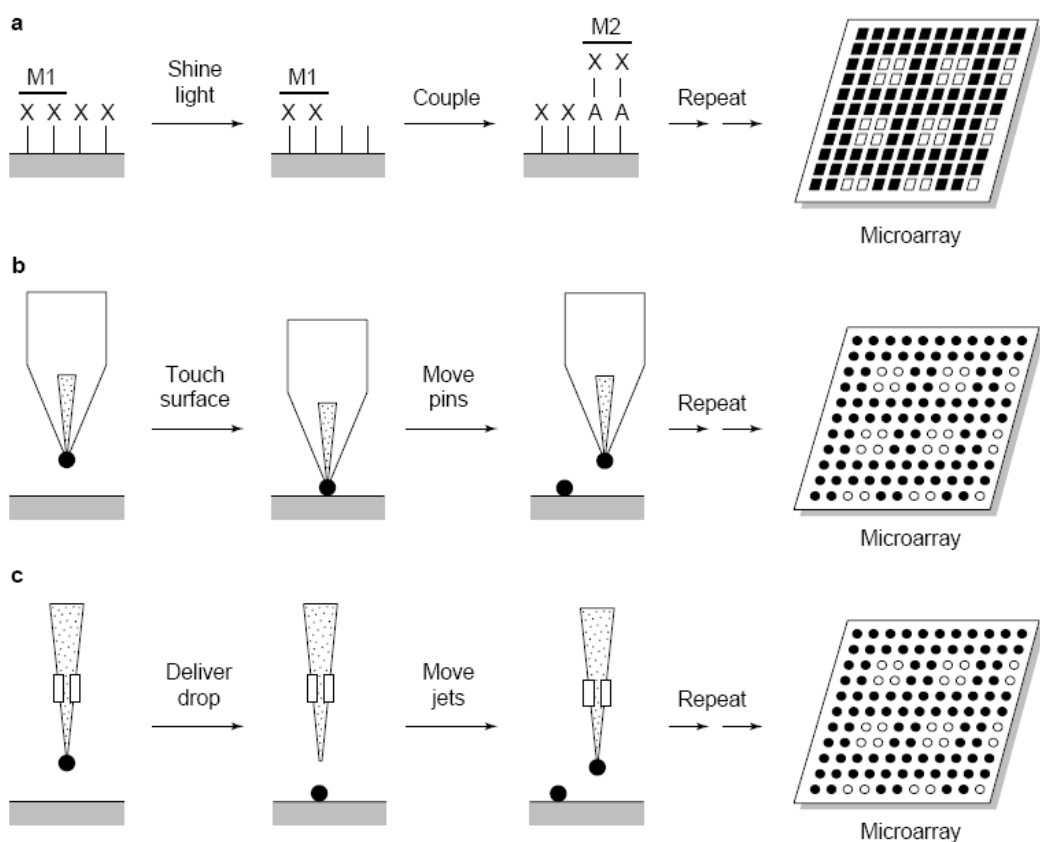


**Figure 1 Example of a DNA microarray or chip, with a single gene sequence highlighted (Campbell 2003)**

Mechanical microspotting, a *contact* printing technology, is a widely used technique for manufacturing microarrays using the *delivery* technique (Figure 2b). Using this method the target DNA molecule must be synthesised, purified and stored prior to microarray manufacture using conventional PCR techniques. Glass slides are overlaid with a positively charged coating. The prefabricated cDNA or oligonucleotides are loaded into a spotting pin by capillary action and a small volume is transferred by physical contact with the solid glass surface. After the first spotting cycle the pin is washed and a second sample is loaded and deposited at an adjacent address on the slide.

Microarrays manufactured in this manner may contain 10,000 groups of cDNA in an approximate area of 3.6 cm<sup>2</sup>. Each cDNA enables expression monitoring of its equivalent human gene and a set of four microarrays would allow expression monitoring of approximately 40,000 genes. Although photolithographic methods are capable of producing denser microarrays, microspotted microarrays are cheaper to produce and are likely to be the choice of basic research laboratories (Skena 2003; Skena et al 1998).

Another form of microarray manufacturing, which utilises the *delivery* method in a *non-contact* manner, exploits the same principles of ink jet systems that are used in printing (Figure 2c). As with the microspotting technique, target elements must be pre-synthesised, however this method can be used to spot any biomolecule of interest including cDNAs, genomic DNA, antibodies and small molecules such as proteins. The prefabricated biomolecules are loaded into a flexible capillary fitted with a piezoelectric transducer. Electric pulses from the transducer create transient pressure waves inside the capillary tube, which results in the expulsion of a small amount of liquid, containing the target molecule. After the first jetting step, the jet is washed and a second sample is loaded and deposited at an adjacent address on the glass slide. As ink jetting uses drop-on-demand delivery and does not require surface contact it is capable of high throughput production. Although ink jetting is not considered as robust a technique as microspotting or photolithography, it can be used to produce microarrays of single cDNAs at a density of 10,000 spots per cm<sup>2</sup> (Skena 2003; Skena et al 1998).



**Figure 2** Microarray production techniques (Skena et al 1998)  
a) Photolithography, b) Mechanical microspotting, c) Ink jetting

The other type of microarray preparation technique is the synthesis approach, which creates target elements *in situ*, that is directly on the microarray glass surface. The target elements are constructed in a step-wise fashion nucleotide-by-nucleotide until the desired oligonucleotide length is reached, usually up to 25 bases. The short oligonucleotide sequences are usually produced using photolithography (Figure 2a) (Schna et al 1998). The United States company, Affymetrix, pioneered the development of oligonucleotide microarrays and dominate the field, with many companies using Affymetrix gene chips as a platform for their own products<sup>2</sup>. The advantages of *in situ* microarray synthesis compared to cDNA arrays is that after the initial expense of purchasing the photolithographic technology, the production of oligonucleotide microarrays is relatively cheap. *In situ* synthesis is capable of producing low, medium or high density microarrays ranging from 100-500,000 elements per cm<sup>2</sup> and sample tracking is accurate as there is no need for sample handling. The small size of the individual oligonucleotides results in reliable target identification which impacts favourably on the cost-effectiveness of the technology. However, the coupling efficiency of oligonucleotides is reduced for longer target sequences, and is therefore lower in microarrays produced via the synthesis method (95-99%) when compared to those prepared with the delivery method (>99%) (Schna 2003).

Photolithography (Figure 2a) utilises *non-contact* technology from the microelectronics semiconductor industry, combining it with synthetic-DNA chemistry. A glass wafer is modified with a light sensitive chemical compound or photo-protecting groups (X). A lithographic mask (M1) is used to either block or transmit light onto specific locations of the wafer surface, ensuring that DNA synthesis occurs only in defined positions. The surface is then flooded with a solution containing either adenine (A), thymine (T), cytosine (C), or guanine (G), and coupling occurs only in those regions on the glass that have been “deprotected” through light activation. When the DNA base attaches to the glass surface it does so at the 3' end and is referred to as a coupled nucleotide. The coupled nucleotide is also bound to a photo-protecting group (A-X) at the 5' end, enabling the cycle of light activation (deprotection) and nucleotide coupling to be repeated. The process is repeated until the oligonucleotides are approximately 25 bases in length (Affymetrix 2007a; Schna 2003; Schna et al 1998). The advantage of this process is the sheer number of oligonucleotides able to be synthesised, with a gene chip capable of containing as many as 400,000 groups of oligonucleotides or features in an area approximately 1.6 cm<sup>2</sup>, with each feature containing in the region of 10<sup>6</sup> oligonucleotides of a given sequence. The main disadvantage of the photolithographic approach is the need for photomasks, which are expensive and time-consuming to design and construct (Schna et al 1998).

#### *How DNA microarrays work*

Once the DNA microarray has been prepared, messenger RNA (mRNA) is isolated from the test sample of interest and converted into single stranded

---

<sup>2</sup> The disadvantage of using the Affymetrix GeneChips is that the arrays can only be processed on Affymetrix instrumentation. Murphy, D. (2002). 'Gene expression studies using microarrays: principles, problems, and prospects', *Adv Physiol Educ*, 26 (1-4), 256-270.

cDNA (Figure 3). Both the test sample cDNA (DNA probe) and a reference sample (ie normal tissue) are tagged with different fluorophores which will emit light at different wavelengths, so that the extent to which each sample hybridises to the microarray may be measured independently. Using Figure 3 as an example, the sample of interest may be tagged red (eg breast cancer sample) and the reference sample tagged green. The two sample populations are mixed and hybridised to the microarray (Campbell 2003; Haviv & Campbell 2002; Ntzani & Ioannidis 2003). Hybridisation should always be carried out in the dark to minimise photobleaching of the fluorescent dyes. To ensure complete hybridisation, binding of the samples to the microarray should be carried out for 14-18 hours, as the hybridisation process is limited by diffusion (Blalock 2003). Unbound cDNA is washed away and a laser scanner measures the intensity of the red/green fluorescent signal of the bound cDNA. Active genes will have a greater representation in the DNA probe resulting in a more intense, brighter hybridisation signal. If the gene is *not expressed* in the test sample, the target spot will appear green and if the gene is *only expressed* in the test sample the spot will appear red. However, if the level of transcription is the same in the test and reference sample, the spot will bind both red and green fluorophores in equal proportion and the signal will appear yellow. An absence of gene expression in either the test or reference sample will result in a black spot (Campbell 2003; Haviv & Campbell 2002; Ntzani & Ioannidis 2003).

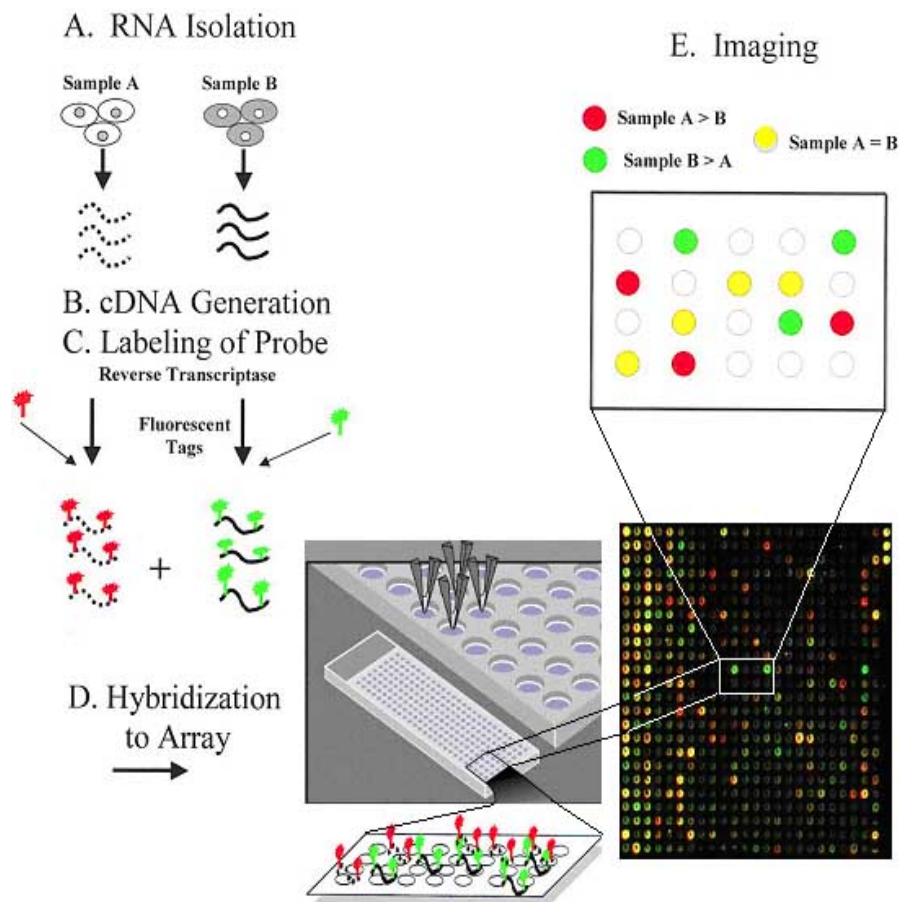


Figure 3 Schematic of a DNA microarray hybridisation (Krutovskii & Neale 2001)

For an example of an actual DNA microarray see Appendix B.

Using image analysis software, the intensity data are translated to give numerical estimates reflecting the relative level of a particular gene transcript (ie the abundance of mRNA) in both the test and normal samples. These numbers may correspond to absolute levels, or a ratio of the level of gene expression in the test sample against that of the reference standard. An expression ratio of one indicates balanced expression of test and reference samples, a ratio of 2.0 means that there is twice as much mRNA from the target sample compared to the reference, and a ratio of 0.5 indicates that the level of mRNA in the reference sample is twice that of the test sample (Berrar et al 2003).

It is important to choose an appropriate reference standard with which to compare the test sample. For low complexity experiments such as those looking for a treatment response in an *in vitro* cell culture assay, an appropriate reference would be a matched, untreated sample. In complex studies such as those using primary tumour biopsies, the reference is usually a pooled RNA sample from a mix of tumour cell lines. Pooled reference standards are also commercially available (Haviv & Campbell 2002).

Due to the large number of target elements able to be placed on a microarray slide, the simultaneous, rapid characterisation of thousands of genes is possible. As such microarrays may be useful tools in categorising diseases such as cancer by determining the presence or absence of particular genes, which may provide important biological, diagnostic and prognostic information (Haviv & Campbell 2002; Ntzani & Ioannidis 2003).

The types of studies that may be carried out using DNA microarrays include:

- Differential gene expression studies: observing a single gene profile, searching for genes which exhibit different expression levels under different conditions (eg tissue type or developmental stage). Studies of this kind include normal-versus-diseased state investigations.
- Gene co-regulation studies: instead of analysing expression variation of a single gene as above, gene co-regulation studies compares one gene profile to another. The aim is to identify expression levels of genes that correlate (either positively or negatively) with the expression of other genes. That is if the expression level of one gene increases as that of another increases, and vice versa for negative correlating genes.
- Gene function identification studies: the comparison of a novel gene's expression profile under varying conditions, compared to profiles of genes with known function to infer function of the novel gene.
- Time-course studies: requires transcript samples taken at different time points, hybridised on the array – one array per time point. This may elicit temporal changes in gene expression and may be especially useful in the study of the cell cycle.

- Dose-response studies: to uncover changes in gene expression profiles in response to either a sample or patient being exposed to differing doses of a chemical or drug. Drug-dose response studies may include an element of time-course studies.
- Identification of pathways and gene regulatory networks: pathway identification investigates the routes and processes by which genes and their products (proteins) function in cells and tissues. This process involves monitoring any changes in gene expression in response to making changes in a specified pathway.
- Predictive toxicology studies: of particular use in the pharmaceutical industry for the identification of toxic substances. Relies on accessible reference databases containing the results of large microarray screening experiments. The influence of a new compound on the expression of several genes of interest is compared to the expression profiles of known toxins in the database, then using an inference-by analogy to predict the toxicity of the new compound.
- Clinical diagnosis: to document expression patterns that are characteristic of a particular disease. May also be utilised to uncover subtypes of known diseases.
- Sequence variation studies: used in the detection of single nucleotide polymorphisms (SNPs) which may result in phenotypic changes ie disease (Berrar et al 2003).

DNA microarrays are not currently used for *diagnostic* purposes in Australia.

### **Summary**

DNA microarrays are often referred to as gene chips.

DNA microarrays may be manufactured using either the *delivery* or *in situ* technique.

The *delivery* technique requires prefabricated target molecules but can spot large amounts of the target onto the slides including 5-120 bases of DNA, proteins and antibodies.

The *in situ synthesis* technique generates small oligonucleotides (up to 25 bases) directly onto the glass slide.

Samples are hybridised to the target molecules and the intensity of attached fluorescent probes is a measure of gene activity.

Using DNA microarrays the simultaneous, rapid, characterisation of thousand of genes is possible.

Using the same principles as DNA microarrays a great deal of current research is being conducted on protein, antibody and antigen arrays. These arrays are a more powerful method than conventional protein chemistry assays as they are capable of high-throughput, parallel analysis of cellular proteins. In addition, protein arrays utilise fluorescence unlike many conventional protein chemistry assays which use radioactivity. These arrays may be used to study protein-protein, antibody-antigen, protein-drug and protein-DNA interactions, which

are important elements in areas of research including cell signalling, cellular differentiation, molecular pathology, drug discovery and diagnostics. A great deal of work utilising protein microarrays is being conducted in patients with autoimmune diseases, where individuals mount an immune response to their own cellular proteins (Schna 2003). For the purposes of this *Emerging Technology Bulletin*, protein microarrays will *not* be considered further.

### *Reliability issues associated with DNA microarrays*

#### Noise

Prior to 2001, manufacturers would identify which genes the probes would target in a microarray but not the actual nucleotide sequence of the probes. Experiments showed that when probing for the same gene using different microarray platforms (Affymetrix, Agilent and Amersham arrays) reproducibility was poor, with a high level of discordant data. Among all three brands of microarrays, only 30 per cent agreement could be reported with an increase to 52 per cent if only two brands were considered. After the sequence composition of the probes was attained it was found that different probes were responding to the same portion of a particular gene. Although large companies such as Affymetrix now release sequence data for microarray probes, many other companies still do not provide these details (Draghici et al 2006; Tan et al 2003; Marshall 2004).

One of the major issues associated with DNA microarrays is the difficulties experienced in deciphering information gained from high throughput gene expression experiments due to the noisy nature of the data. Expression noise is inherent in measurements of gene expression utilising DNA microarrays. Noise can be defined as changes in the measured transcript values between different experiments, caused by both biological variations (ie *real* differences between different cell types and tissues) and experimental noise. Expression noise may obscure useful patterns of gene expression, resulting in a false negative (Febbo & Kantoff 2006).

#### Accuracy

Accuracy, or the degree of conformity of the measured quantity to its true value, is an important consideration when using microarrays to quantify gene expression. The lack of understanding of RNA-DNA hybridisation dynamics may be the cause of microarray inaccuracies. Lack of specific binding and splicing variants may account for a great deal of variation observed in microarray experiments.

Three types of signals may be produced by a microarray probe:

- a specific signal produced by a specific probe binding to the target;
- a cross-hybridisation signal produced by transcripts which doesn't have 100 per cent sequence compatibility but can still significantly bind the probe; and
- a non-specific, background signal that is present in the absence of any sequence homology.

Studies have demonstrated that the binding affinity for non-specific sequences, as demonstrated in the last two scenarios, can at times be greater than the binding affinity of 100 per cent complementary RNA-DNA. In addition, before the translation of RNA to protein, the transcript must be processed,

removing segments or sections, which is known as splicing. However, alternative splicing of the same transcript may take place, producing alternative gene products. Different probes may respond to the same portion of a gene, as a large number splice variants exist (Draghici et al 2006).

### Sensitivity

The sensitivity threshold of microarray measurements defines the concentration range in which accurate microarray measurements can be made. Experiments measuring the number of transcripts using RT-PCR compared to cDNA and Affymetrix arrays gave consistent results down to the level of two copies per cell. Meaningful results below this threshold could not be obtained using microarrays. Other experiments found that although 83 per cent of gene products could be reliably quantified by RT-PCR, less than 55 per cent could be detected using an Affymetrix GeneChip. It was thought that sensitivity could be increased by increasing the size of the probe as signal strength increases with length. However, an increase in probe length produces only a small increase in sensitivity at the same time as decreasing the specificity due to an increase in the mismatching of base pairs. The detection limits of current microarrays ranges between 1-10 copies of mRNA per cell, which may not be low enough to detect relevant changes on low abundance genes such as transcription factors. Innovations such as labelling probes with quantum dots may overcome these sensitivity issues in the future (Draghici et al 2006).

### Reproducibility

The reproducibility characteristics of a particular microarray platform are relatively easy to assess, however it should be remembered that the reproducibility of a microarray differs greatly from its accuracy. Reproducibility can be measured by the ability of a given probe to bind repeatedly to the same number of labelled transcripts in repeated measurements of the same sample. As discussed previously, a poorly designed probe may cross-hybridise with several transcripts. Repeated measures with the same probe would generate highly reproducible results. Therefore reproducibility is a necessary but not sufficient quality for reliable DNA microarray measurements (Draghici et al 2006).

### Sample preparation

Sample preparation is of crucial importance when conducting DNA microarray experiments or assays. The purity of the RNA sample used is a critical factor for optimum hybridisation as cellular protein, lipid and carbohydrate can facilitate non-specific binding of labelled cDNAs to the surface of the matrix. Another experimental limitation of cDNA microarrays is the large amount of RNA required to produce an adequate gene expression signal over noise, especially in samples with low abundance transcripts. Fluorescence detection of hybridised samples requires  $\geq 10\mu\text{g}$  of total RNA, which is equivalent to RNA extracted from  $10^6$  cells (Murphy 2002).

### *Data analysis of microarray results*

Prior to microarrays, functional gene analysis was performed on a single gene basis. Methods for assaying gene expression levels by DNA microarray experiments produce large volumes of data, ranging from 20,000 to  $10^6$  data points at any one time. It should be noted that the description which follows

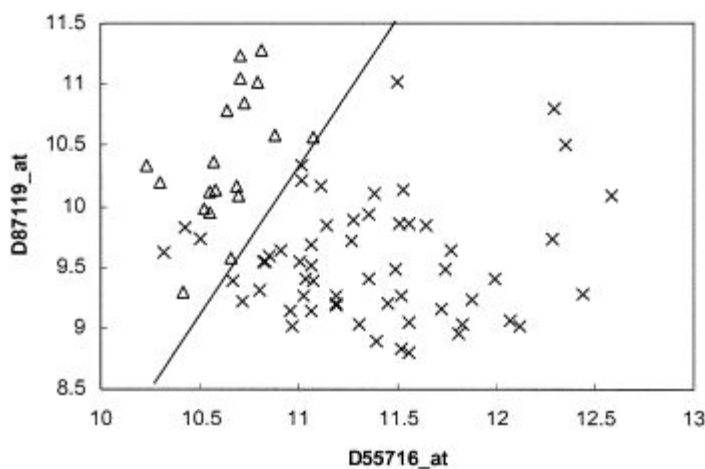
provides a brief overview of the basics of DNA microarray data analysis and entire books have been published dedicated to describing comprehensive methods of analysis (Amaratunga & Cabrera 2004; Berrar et al 2003; Causton et al 2003; Schena 2003).

Prior to any meaningful analysis of DNA microarray data, pre-processing and normalisation of the data must take place. Pre-processing takes into account variations in spot intensities due to factors inherent in the DNA microarray manufacturing process such as different print tips, differences in PCR reactions used to generate oligonucleotides, integration efficiencies of the fluorescent dyes and hybridisation specific effects. Measurements need to be normalised to “even out” variations observed in duplicates or triplicates of the same samples. If normalisation is not carried out differences in measurements may be attributed to the differential expression of genes rather than differences in hybridisation (Steinhoff & Vingron 2006).

Two methods of data analysis are currently used: *descriptive analysis*, which summarises the data and describes changes that exist in a group of samples; and *predictive analysis*, which uses inductive methods to predict the course of disease (Eschrich & Yeatman 2004).

Clustering is a descriptive analysis method for the grouping of similar data according to a measure of similarity. Clustering provides a method for summarising data both visually and graphically, and in microarray analysis patient samples or specific genes may be clustered. Clustered genes may shed light on interesting groupings of genes, whereas patient samples are clustered based on the similarity of gene expression levels within the sample. Clusters that do not correspond to existing groups may indicate a molecular origin for the difference, however information may be difficult to attain if the number of irrelevant genes is large compared to the number of informative genes. To analyse differences in genes between two groups, which may be helpful in finding new therapeutic targets or prognostic indicators, statistical methods can be applied to microarray data. Important genes may be identified by looking for a greater than 2-fold increase in gene expression between groups. However, this method is viewed as inexact as even a small change in gene expression may result in large biological differences. Therefore statistical tests such as the t-test for measuring the differences between two group means and ANOVA for more than two groups, are more likely to be used to identify significant gene expression differences between groups. It is important to remember when determining the significance level at 0.05, many genes may be identified by chance given the very large number of tests that are undertaken and associations that could be identified. For example, using the Affymetrix HG-U133A gene chip, which allows testing of 22,000 genes, as many as 1,100 genes may be identified by chance as significant. Utilising corrective methods for multiple statistical testing, such as the Bonferroni correction, may result in an over-adjustment with very few genes identified as statistically significant. An alternative technique known as SAM (Significance Analysis of Microarrays), uses the random permutation of data groups to estimate the distribution of a t-test empirically. Genes are considered to be significant if the test statistic is better than that expected by chance (from the permutations). Therefore SAM gives the rate at which genes may be called significant by chance (Eschrich & Yeatman 2004).

Predictive analysis relies on the generation of predictive models and may be referred to as “learning from examples”. Two stages are involved in “learning” from microarray data: gene selection and classifier construction. Selection of a gene of interest may be accomplished using the descriptive methods described above. A classifier is a mathematical function of gene expression values that gives a prediction value as an output. For example, data generated from DNA microarrays of two genes of interest (D55716\_at and D87119\_at) from two different types of lymphoma (diffuse large B-cell or follicular lymphoma) (Figure 4). The line is the classifier that best separates the two groups of samples. This method may be used to identify gene signatures that may predict patient outcome and disease progression, and forms the basis of products such as MammaPrint<sup>®</sup> and the 76-gene expression panel discussed later in this *Bulletin* used in the prognosis of breast cancer patients (Eschrich & Yeatman 2004).



**Figure 4** Two-dimensional gene space of diffuse large B-cell vs follicular lymphoma. Two classes are represented as the x and Δ symbols and the line is the classifier. (Eschrich & Yeatman 2004)

## Comparators

The central dogma of molecular biology dictates that DNA (genes) can either undergo replication (DNA→DNA) or transcription (DNA→mRNA) and then translation (mRNA→protein) (Figure 5). Gene expression is a common term used to describe the transcription of information encoded within DNA sequences into mRNA, and the subsequent translation of the mRNA information into proteins that regulate and control cell function. Analysis of gene expression was previously conducted on a single gene-by-gene basis, by low-throughput techniques that either utilised nucleic acid probes (in situ hybridisation, northern blotting, Southern blotting, RNase protection assays or real-time PCR) or protein probes (immunocytochemistry or western blotting). It should be noted that the expression of many genes is regulated *after* transcription by enzymes, therefore the measurement of mRNA concentrations *may not* reflect the true level of gene expression (Murphy 2002).

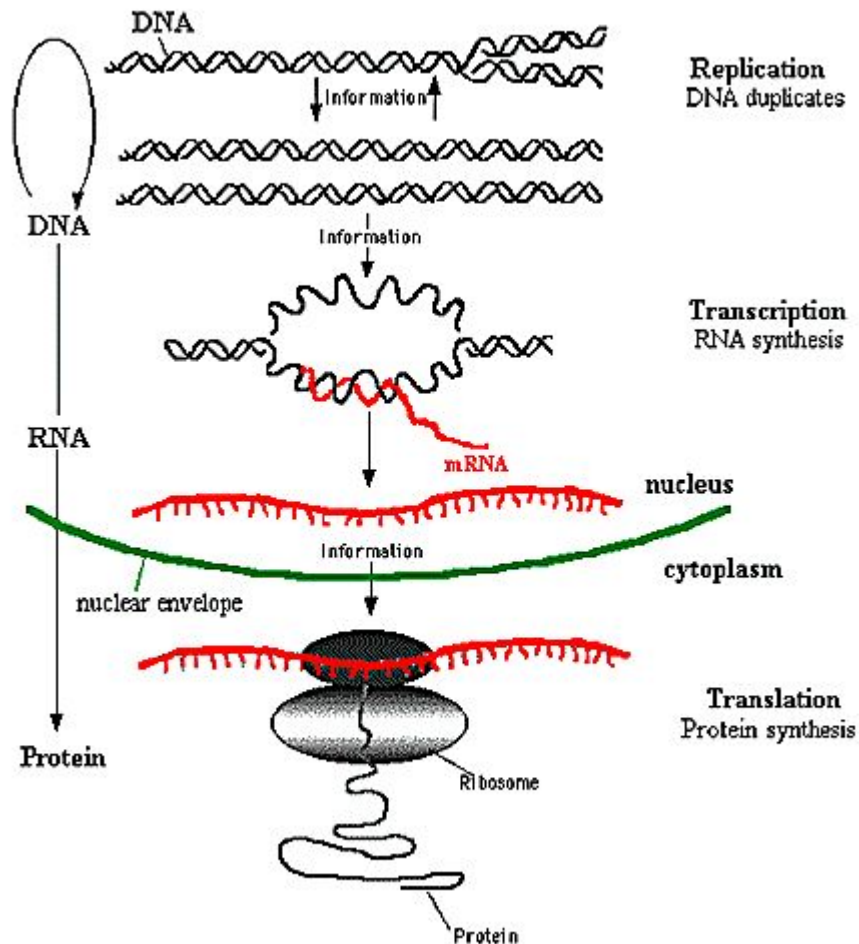


Figure 5 The central dogma of molecular biology (Access Excellence @ The National Health Museum)

The Southern blot was first reported in 1975 as a method to detect the presence of a particular DNA sequence in a sample of DNA. Southern blotting is a robust technique that was widely used in research and diagnostic laboratories. Using electrophoresis, DNA that has undergone digestion with restriction enzymes, is separated on an agarose gel and transferred to a filter membrane. A radioactive single-strand DNA or RNA probe is created which will hybridise to the DNA region of interest. After washing excess probe away, the pattern of hybridisation is visualised on X-ray film. Southern blotting is useful for detecting gene rearrangements or deletions (Watson et al 1992).

Northern blots, developed not long after Southern blots, utilise the same principle as a Southern but quantitatively measure mRNA levels instead of DNA. RNA samples are first separated by size via electrophoresis in an agarose gel under denaturing conditions. As RNA is easily degraded, RNase free reagents must be used at all times. The RNA is then transferred to a membrane, cross-linked and hybridised with either a DNA or RNA radio-labelled probe that is complementary to the target sequence. After washing off excess probe, the pattern of hybridisation is again visualised on X-ray film and the strength of the visualised band correlates to the level of gene activity or expression. Northern blots require relatively large amounts of RNA and cannot

be used to detect low level gene expression from either single cells or small numbers of cells (Watson et al 1992; Wikipedia 2007c).

For cases where the amount of starting material is limited, some degree of amplification is required using real-time PCR<sup>3</sup>. In conventional PCR, DNA undergoes rounds or cycles of replication amplifying regions of DNA with the aid of specific primers. During the amplification process there are three distinct phases: the exponential phase where product rapidly accumulates due to the availability of fresh reagents favouring a doubling of the amplicon; the linear phase where reagents are consumed and the reaction begins to slow down with product no longer doubling in each cycle of the reaction; and finally the plateau phase where the reaction slows and eventually stops with reagents fully depleted. The plateau phase is where conventional PCR takes its measurement, known as the end-point. Quantification of PCR product during the plateau phase is not as accurate as measurements taken in the exponential phase. Real-time PCR quantifies the DNA after each round of replication using either a fluorescent dye, which inserts into the double stranded DNA, or by using modified DNA oligonucleotide probes which fluoresce when hybridised with complementary DNA. To quantify gene expression via the quantification of low abundance mRNA, real-time PCR may be combined with reverse transcription PCR. The amount of RNA or DNA (depending on the real-time PCR technique used) present is calculated against a standard curve which is generated by serial dilutions of a known amount of RNA or DNA (Applied Biosystems 2007; Wikipedia 2007d).

Whether a gene is active or not can be detected by the presence or absence amount of translated protein. The most commonly used method to detect protein is using western blots. A tissue homogenate or extract is separated by electrophoresis on a polyacrylamide gel. After transfer to a nitrocellulose membrane, the separated proteins are probed with an antibody specific to the protein of interest, which is conjugated to a reporter for imaging or quantification (Watson et al 1992).

---

<sup>3</sup> Real time PCR is often designated as RT-PCR however this may lead to confusion with reverse transcriptase PCR

## DNA microarrays and breast cancer

Breast cancer is the most frequently diagnosed cancer among women in Australia with 12,027 new cases occurring in 2002. The incidence of breast cancer in females rose from 100.4 cases per 100,000 population in 1991 to 117 cases per 100,000 population in 2002. This is an increase of 1.4 per cent per annum over the eleven year period from 1991-2002, however a significant contributor to this increase has been a greater female population having early detection through mammographic screening (Figure 6) (AIHW 2006).

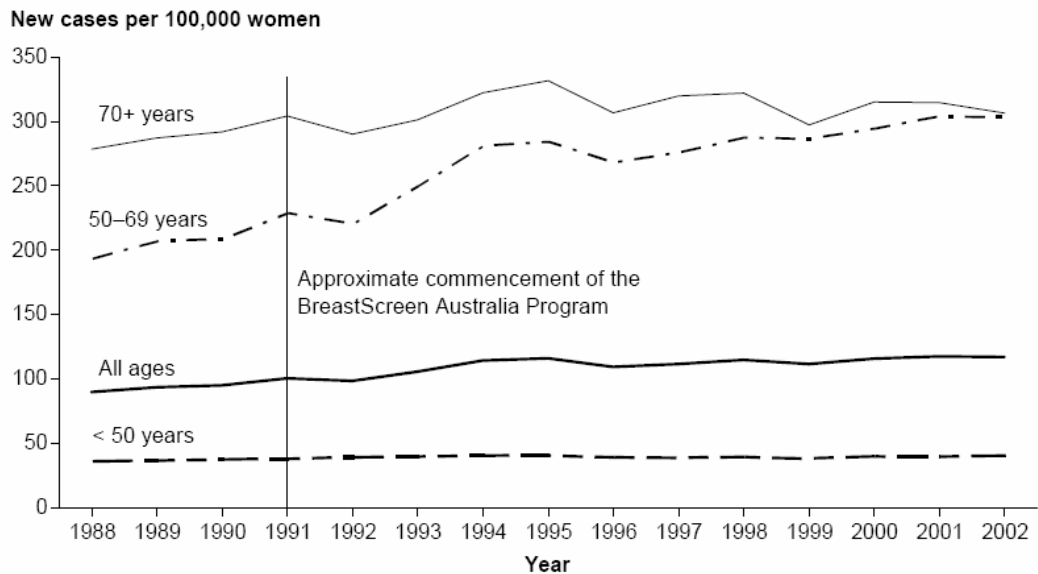


Figure 6 Incidence of breast cancer in women, Australia, 1988-2002 (AIHW 2006)

Australian age standardised incidence of breast cancer per 100,000 is much higher than the average for some developed countries of the world, but is similar to rates described by countries such as Canada, New Zealand and the United Kingdom, and lower than that of the United States (AIHW 2006). However, Australia has the 5<sup>th</sup> highest breast cancer incidence rate in the world with one in 11 Australian women developing breast cancer before reaching 75 years of age.

In addition, breast cancer is the most common cause of cancer-related death in Australian women, with 2,641 deaths attributed to breast cancer in 2004. However international comparisons of developed countries show that Australia has low mortality rates indicating good survival rates through early detection and treatment strategies (AIHW 2006).

In New Zealand, the number of new female breast cancer registrations was 2,364 and the number of registered female breast cancer deaths was 625, for the year 2002. The age standardised incidence and mortality for breast cancer is 86.0 and 19.8 per 100,000 respectively (New Zealand Health Information Service 2006).

Diagnostic tests aimed at identifying women at genetic risk of breast cancer will only benefit a small proportion of the population. Less than 10 per cent of women who develop breast cancer have an underlying, inheritable genetic susceptibility to the disease. A further 15-20 per cent of women with breast

cancer have a family history of the disease with no readily identifiable genetic pattern of inheritance (Hinestrosa et al 2007). Although mutations in genes such as BRCA1, BRCA2 and TP53 increase the lifetime risk of developing breast cancer from 25-80 per cent (NHMRC 1999a), the frequency of these mutations in the general population are low at 1:1,000, 1:1,000 and 1:10,000, respectively (NBCC 2006).

Breast cancer mortality rates are falling, due in part to the large range of treatment options now available including surgery, radiotherapy, hormonal therapy, chemotherapy and directed antibody therapy. However, one of the most significant problems facing clinicians treating women with breast cancer is the heterogeneity of the tumour itself. Treatment options are often based on algorithms which evaluate the histopathological characteristics of the tumour. Lymph node status, tumour size and grade, patient age and lymphovascular invasion are the main factors taken into account when predicting the risk of metastasis. Lymph node status remains the best prognostic marker for survival, with approximately 50 per cent of node positive breast cancer patients *not* developing recurrence (with or without adjuvant chemotherapy), whereas recurrence may occur in 25 per cent of node negative (Modlich et al 2006).

Previous research has been focussed on the prevention and detection of breast cancer, however, current research is concentrating on the identification of genes or biomarkers that may improve the quality of life of patients already diagnosed with breast cancer. In addition to determining which treatments may be most effective for individual patients or which treatments should be avoided due to potential toxic side effects, the identification of biomarkers may be able to predict tumour behaviour and the prognosis or response of patients to treatment (Hinestrosa et al 2007). Many clinical studies have identified alterations in single genes or biomarkers, such as oestrogen receptor status (ER) and the epidermal growth factor receptor, ERBB2 (also known as HER2) as important for the type of chemotherapy or hormonal treatment offered to women with breast cancer and their eventual treatment outcome (Hinestrosa et al 2007; van 't Veer et al 2002).

In both lymph node-negative and lymph node positive patients, adjuvant poly-chemotherapy may significantly improve the period of time patients are disease free and their overall survival from breast cancer. However, for patients who are node-negative, who have a better prognosis when compared to those who are node-positive, treatment with adjuvant chemotherapy results in only a small improvement in survival rates. In randomised controlled trials of women younger than 50 years of age, when poly-chemotherapy was compared to no chemotherapy, 10-year disease-free survival only increased from 58 to 68 per cent. Improvements in disease-free and overall survival decreased with increasing age. In addition, other studies have indicated that node-negative women treated with only tamoxifen after surgery have an average 10-year recurrence rate of approximately 15 per cent indicating that 85 per cent of these women may have been subjected to toxic chemotherapy unnecessarily (Blue Cross and Blue Shield Association 2005). DNA microarray technology may facilitate the classification of all breast cancers into prognostic categories depending on the expression of a panel of genes or biomarkers, predicting which patients would as well as patients who would not

benefit from adjuvant chemotherapy (Hinestrosa et al 2007; van 't Veer et al 2002).

DNA microarray analysis for the prognosis of breast cancer was first described by van 't Veer et al in 2002. Primary breast tumours from 98 women were examined. The study group included 'sporadic' cases who were lymph node negative and <55 years when diagnosed (n=78) of whom 34 developed distant metastases within five years and 44 were disease-free after five years. In addition, there were 20 cases with a genetic origin: 18 BRCA1 and two BRCA2 patients. RNA was isolated and used to derive complementary RNA (cRNA). Equal amounts of cRNA from the sporadic cases were pooled to create a reference cRNA pool. Two hybridisations for each tumour were carried out on microarrays containing approximately 25,000 human genes. Fluorescent intensities were scanned, quantified, normalised and corrected to yield the transcript abundance of a gene as an intensity ratio with respect to that of a signal of the reference pool. Using clustering algorithms, a gene expression signature strongly predictive of a short interval to distant metastases (ie a 'poor prognosis' signature) in lymph node negative women was identified. The 'poor prognosis' signature consisted of a set of 70 genes with diverse functions including regulating cell cycle, metastasis, invasion and angiogenesis. When the gene expression profile was used to classify lymph node negative patients for eligibility for adjuvant chemotherapy, it performed as well as other clinical parameters (St Gallen and NIH consensus clinical parameters<sup>4</sup>) (Table 2). Using the St Gallen and NIH criteria, 82 and 92 per cent of lymph node negative women respectively, would have been candidates for adjuvant therapy. However, even without adjuvant therapy, 70-80 per cent of these women *would not go on to develop* distant metastases, and would therefore *not benefit* from the treatment and may even suffer severe side effects from the treatment. Using the 'poor prognosis' gene expression profile, 43/78 (55%) of women would be candidates for adjuvant therapy, significantly reducing the number of patients who would otherwise receive unnecessary treatment (van 't Veer et al 2002).

**Table 2 Breast cancer patients eligible for adjuvant therapy using 70-gene panel (van 't Veer et al 2002)**

	Patient group		
	Total (n=78)	Metastatic disease at 5-years (n=34)	Disease free at 5-years (n=44)
<b>St Gallen</b>	64/78 (82%)	33/34 (97%)	31/44 (70%)
<b>NIH</b>	72/78 (92%)	32/34 (94%)	40/44 (91%)
<b>Poor prognosis profile</b>	43/78 (55%)	31/34 (91%)	12/44 (27%)

This seminal 2002 paper formed the basis for the commercial product MammaPrint<sup>®</sup>, developed by Agendia, which in February 2007 became the first FDA approved multi-gene microarray which determines the likelihood of recurrence within 5-10 years for women with lymph node negative breast cancer (FDA 2007).

<sup>4</sup> Based on histological and clinical characteristics

Another prognostic test, Oncotype DX™ developed by Genomic Health Inc, uses a 21-gene panel to quantify the likelihood of recurrence in women with early stage breast cancer. Oncotype DX™ has been commercially available since 2004 and received regulatory clearance from the United States Clinical Laboratory Improvement Amendments<sup>5</sup>. It should be noted that Oncotype DX™ utilises RT-PCR, *not DNA microarray technology*, to screen for the 21 genes involved in breast cancer progression (5 control genes and 16 breast cancer genes including those associated with proliferation, HER2, invasion and oestrogen) (Hinestrosa et al 2007). Oncotype DX™ has the advantage that it can utilise RNA extracted from formalin fixed, paraffin embedded tissue unlike DNA microarrays which require RNA extraction from fresh tissue, which may have implications for routine clinical use (Kaklamani 2006). The Oncotype DX™ test may be accessed by Australian patients, with samples being sent to Genomic Health Inc (USA) for processing, which takes approximately 10-14 days from receipt of sample. The cost of the Oncotype DX™ assay is approximately A\$4,000<sup>6</sup> (personal communication Genomic Health Inc).

The original validation trial of the 21-gene panel was conducted retrospectively on a cohort of 668 lymph node negative women (level III-3 prognostic evidence). The levels of expression of the 16 cancer-related and five reference genes were used in a defined algorithm to calculate a recurrence score and to determine a risk group (low, intermediate, or high) for each patient. The RT-PCR assay categorised 51, 22 and 27 per cent of woman as having a low, intermediate, or high-risk, respectively, of breast cancer recurrence. The 10-year distant metastases recurrence rates in the low, intermediate, and high-risk groups were 6.8% [95%CI 4.0, 9.6], 14.3% [95%CI 8.3, 20.3], and 30.5% [95%CI 23.6, 37.4], respectively. A multivariate analysis indicated that women in the high-risk category were more likely to experience recurrence than those in the low risk group (HR =2.81, 95%CI [1.70, 4.64],  $p < 0.001$ ) (Abdullah-Sayani et al 2006; Paik et al 2004).

A cost-utility analysis of this validation cohort was conducted by Hornberger et al (2005). A Markov model was used to estimate overall survival, costs and cost-effectiveness of patients classified as either high- or low-risk according to their recurrence score obtained from the 21-gene assay, compared to the status quo of the recurrence rate of women classified according to the National Comprehensive Cancer Network (NCCN) guidelines. Administration of adjuvant chemotherapy in *high-risk patients* based on NCCN guidelines was estimated to increase overall survival by 1.53 years and increase QALYs by 0.65 years. Adjuvant chemotherapy increases overall costs for breast cancer by US\$13 878, resulting in a cost utility equal to US\$21,428. In *low-risk patients*, adjuvant chemotherapy was calculated to increase overall survival by only 0.37 years compared with use of chemotherapy in high-risk patients. However, when adjusting for the effect of chemotherapy on the quality of life, quality-adjusted survival with chemotherapy is lower in this group than without

---

<sup>5</sup> Diagnostic laboratories may develop validated “in-house” tests and offer them as a service under the condition that the test meets CLIA standards Modlich, O., Prisack, H. B. & Bojar, H. (2006). 'Breast cancer expression profiling: the impact of microarray testing on clinical decision making', *Expert Opin Pharmacother*, 7 (15), 2069-2078..

<sup>6</sup> Cost per test quoted as US\$3,467= A\$3,884 on 17<sup>th</sup> July 2007

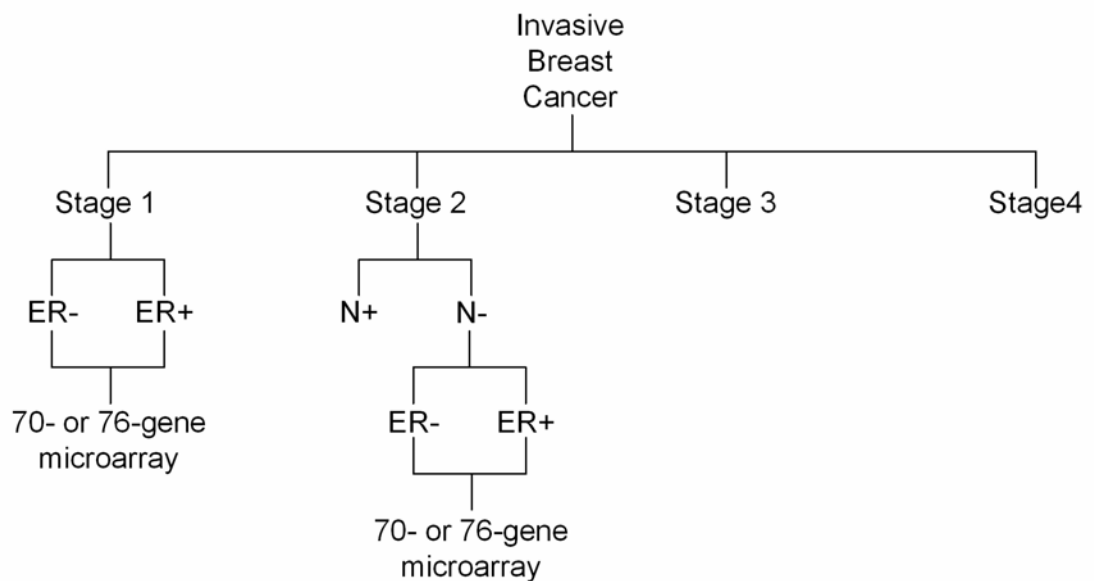
chemotherapy. When reclassifying a hypothetical cohort of women using their recurrence scores not the NCCN criteria, two patients would be expected to be reclassified from low to intermediate/ high-risk and 45 patients would be reclassified from high- to low-risk. Assuming that patients at intermediate/high-risk receive chemotherapy and patients at low-risk do not receive chemotherapy, testing with the 21-gene panel will increase QALYs in this cohort by 8.6 years. Although the assay will cost US\$345,000; the cost of adjuvant chemotherapy will decrease by 46%, from US\$1.63 million to US\$876 000. Therefore, overall costs were projected to decline by 5%, from US\$4.32 million to US\$4.12 million, for a net savings of US\$202,828. The authors concluded that, if applied appropriately, the assay may increase quality-adjusted survival and save costs (Hornberger et al 2005).

For the purposes of this *Emerging Technology Bulletin*, the Oncotype DX™ prognostic test will not be considered further, however the effectiveness of this technology is discussed in several papers which may be of interest (Blue Cross and Blue Shield Association 2005; Cobleigh et al 2005; Habel et al 2006; Kaklamani 2006; Mina et al 2007). It is of interest as there has been an economic assessment of its potential impact that may serve as a model for the analysis of other microarray tests.

#### *DNA microarray breast cancer prognostic tests*

The purpose of gene expression analysis is to determine which genes are actively transcribed into messenger RNA (mRNA) and therefore translated into proteins in tissue samples of interest, in this case breast cancer. Differences in the level of gene expression may correlate with characteristics of breast cancer and aid in the treatment and prognosis of patients. There are currently two main breast cancer gene expression microarrays described in the literature: the 70-gene microarray MammaPrint®, and a 76-gene microarray (the “Rotterdam signature”) developed by several groups in the Netherlands. Both assays are designed for use in lymph node-negative women according to the schematic in Figure 7. Both assays aim to identify those women at *low risk* of breast cancer recurrence for whom post-surgical adjuvant chemotherapy could be safely avoided (Blue Cross and Blue Shield Association 2005).

A United Kingdom review of gene expression profiling of breast cancer patients outlined the potential savings to the health system if patients were successfully stratified to an appropriate treatment regime. An outpatient course of six cycles of standard chemotherapy with adriamycin/cyclophosphamide costs approximately £2,400 (A\$5,600) (Cleator & Ashworth 2004). In Australia, there were 12,027 new cases of breast cancer reported in 2002 (AIHW 2006). If 50 per cent of diagnosed breast cancer cases receive adjuvant chemotherapy, this would equate to 6,000 women in Australia being administered chemotherapy in each year. If chemotherapy use was decreased by 70 per cent this would represent a saving of A\$23 million (Cleator & Ashworth 2004). However the current cost of a gene expression assay is high (MammaPrint® = A\$3,600). If all newly diagnosed women were tested with this assay, this would equate to an outlay of A\$43 million. This analysis does not take into account potential benefits for health related quality of life in more appropriate triage for chemotherapy.



**Figure 7** Patient criteria for the use of the 70- or 76-gene microarray (ER- = Oestrogen negative, ER+ = oestrogen positive, N+ = lymph node positive, N- = lymph node negative)

### *MammaPrint*<sup>®</sup>

As described previously, MammaPrint<sup>®</sup> is based on the initial work by van 't Veer et al (2002), and is an inkjet-synthesised oligonucleotide microarray based on 70 genes associated with metastasis of breast cancer. MammaPrint<sup>®</sup> analysis cannot be conducted on formalin-fixed, paraffin embedded tissue samples as this process degrades the RNA. Messenger RNA is isolated from fresh frozen tissue and treated with DNase to remove contaminating DNA. After transcription, cRNA is produced and labelled with a fluorescent dye. The labelled cRNA, together with the labelled cRNA from a reference sample, is then hybridised to a DNA microarray containing 25,000 oligonucleotides (corresponding to small sequences of the 70 genes) and scanned with a confocal laser scanner. Each MammaPrint<sup>®</sup> microarray contains three identical sets of fragments of the 70 genes to be analysed, giving three independent measurements of the 70-gene profile. In addition, the arrays contain several hundred normalisation genes and negative control genes, which are DNA sequences that can not bind human cRNA (Agendia; van de Vijver et al 2002). Agendia market the MammaPrint<sup>®</sup> assay, which utilises commercially manufactured Agilent gene chip microarrays produced by Agilent Technologies Inc, USA. The cost per MammaPrint<sup>®</sup> test is approximately A\$2,500<sup>7</sup> (Abdullah-Sayani et al 2006). MammaPrint<sup>®</sup> is currently available in Australia via MedVet Science Pty Ltd (South Australia), with the test itself being performed in the Netherlands. The MammaPrint<sup>®</sup> test is not funded by the Medicare Benefits Schedule, therefore the \$3,600 cost (ex-GST) of the test

<sup>7</sup> Cost per test quoted as €1,600 = A\$2,521 or US\$2,050 = A\$2,345 on 17<sup>th</sup> July 2007  
Abdullah-Sayani, A., Bueno-de-Mesquita, J. M. & van de Vijver, M. J. (2006). 'Technology Insight: tuning into the genetic orchestra using microarrays--limitations of DNA microarrays in clinical practice', *Nat Clin Pract Oncol*, 3 (9), 501-516.

must be fully met by the patient (personal communication MedVet). Samples obtained in Australia are sent to the USA and the turn around time from Agendia receiving the sample to receiving the results is 10 days (Agendia).

Two studies, which used the 70-gene panel for the prediction of distant metastases and survival in breast cancer, were included in this *Bulletin* (Table 3) (Buyse et al 2006; van de Vijver et al 2002). Several smaller studies have been published but as patient numbers were low these studies have not been included. Both of the included studies were retrospective (level III-3 prognostic evidence) and neither study used conventional DNA sequencing or PCR to compare their results. To date no prospective, randomised clinical trials, assigning patients to chemotherapy regimes based on the results of gene expression assays, have been published. However, the MINDACT trial is currently underway and recruiting (see Sources of information). A similar randomised controlled trial is being conducted using the Oncotype DX™ assay. Results from both of these trials will not be known for some time due to the expected lengthy follow-up period (5- and 10-year survival rates).

The study by Buyse et al (2006) only assayed women who were lymph node negative, whereas the study by van de Vijver et al (2002) assayed both lymph node negative and positive women, with results presented for the combined patient group as well as stratified according to lymph node status. Results were expressed as either a poor prognosis (patients were at high-risk of developing distant metastases), or as a good prognosis (patients at low-risk of developing distant metastases).

When looking at only *lymph node negative women*, the hazard ratios indicate that women in the high-risk category were either 2.32 or 5.5 times (Buyse et al and van de Vijver et al, respectively) *more likely* to experience recurrence than those in the low-risk group. Although the hazard ratios estimated by the two studies appear somewhat different, their confidence intervals did overlap (1.35, 4.0 vs. 2.5, 12.2), indicating that these results may not be as different as they appear. The confidence interval around the hazard ratio is likely to be more precise in Buyse et al compared to van de Vijver et al due to the larger sample size employed in the study (n=307 vs n=151). The van de Vijver et al study may also be limited as it included 61 lymph node negative patients who were part of the original study group used for the development of the 70-gene assay. These samples were included in the study to avoid selection bias, as the test set included a disproportionate number of poor prognosis patients (Blue Cross and Blue Shield Association 2005).

Although similarly disparate results were reported by both studies for the risk of overall survival (2.8 and 8.6) van de Vijver et al reported on the overall risk for *all* patients (lymph node positive and negative), whereas Buyse et al reported on only lymph node negative patients.

The study by Buyse et al (2006) compared the ability of the gene expression assay and conventional clinical criteria to stratify lymph node negative breast cancer patients into high- and low-risk for the development of distant metastases. Compared to members of the low risk gene expression group, members of the high-risk group were at a substantially greater instantaneous risk for reporting distant metastases (HR = 2.32, 95% CI [1.35, 4.0]). The

hazard ratio decreased only marginally after adjusting for clinical criteria (adjusted HR ranged from 2.13-2.15), suggesting that differences in the time to distant metastases between the two gene expression groups were largely independent of these criteria. This was also true when overall survival and disease free survival were adjusted for clinical criteria. It was interesting to note that the hazard ratio reported (data *not* presented in table<sup>8</sup>) for risk classification for oestrogen receptor status compared favourably with the gene signature for time to distant metastases (2.18 vs 2.32), overall survival (2.36 vs 2.79) and disease-free survival (1.49 vs 1.5), whereas other clinical criteria did not perform as well.

**Table 3 MammaPrint® as a predictor of distant metastases in breast cancer**

Study	Level of Prognostic Evidence	Study Design	Population	Outcomes																														
van de Vijver et al 2002, Netherlands	III-3	Retrospective cohort	295 consecutive patients with stage I or II breast cancer, less than 53 years old.  151 lymph node negative, 10 of whom received adjuvant therapy and 144 lymph node positive, 120 of whom received adjuvant therapy.  Follow-up: median 6.7 years (range 0.05-18.3 years).	<p><b>Free of distant metastases (% , ± SE)</b></p> <table border="1"> <thead> <tr> <th></th> <th>5 year</th> <th>10 year</th> </tr> </thead> <tbody> <tr> <td><b>All patients</b></td> <td></td> <td></td> </tr> <tr> <td>Low risk (n=115)</td> <td>94.7 ± 2.1</td> <td>85.2 ± 4.3</td> </tr> <tr> <td>High risk (n=180)</td> <td>60.5 ± 3.8</td> <td>50.6 ± 4.5</td> </tr> </tbody> </table> <p><u>High vs low risk: distant metastases in 5 years</u> Over entire follow-up period HR = 5.1      95%CI [2.9, 9.0], <i>p</i> &lt; 0.001 During first 5 years HR = 8.8      95%CI [3.8, 20.0], <i>p</i> &lt; 0.001 After 5 years HR = 1.8      95%CI [0.69, 4.5], <i>p</i> = 0.24</p> <p><b>Lymph node -ve patients</b></p> <table border="1"> <thead> <tr> <th></th> <th>5 year</th> <th>10 year</th> </tr> </thead> <tbody> <tr> <td>Low risk (n=60)</td> <td>93.4 ± 3.2</td> <td>86.8 ± 4.8</td> </tr> <tr> <td>High risk (n=91)</td> <td>56.2 ± 5.5</td> <td>44.1 ± 6.3</td> </tr> </tbody> </table> <p><u>High vs low risk: distant metastases in 5 years</u> HR = 5.5      95%CI [2.5, 12.2], <i>p</i> &lt; 0.001</p> <p><b>Lymph node +ve patients</b></p> <table border="1"> <thead> <tr> <th></th> <th>5 year</th> <th>10 year</th> </tr> </thead> <tbody> <tr> <td>Low risk (n=55)</td> <td>95.2 ± 2.6</td> <td>82.7 ± 7.8</td> </tr> <tr> <td>High risk (n=89)</td> <td>66.3 ± 5.2</td> <td>56.7 ± 6.4</td> </tr> </tbody> </table> <p><u>High vs low risk: distant metastases in 5 years</u> HR = 4.5      95%CI [2.0, 10.2], <i>p</i> &lt; 0.001</p>		5 year	10 year	<b>All patients</b>			Low risk (n=115)	94.7 ± 2.1	85.2 ± 4.3	High risk (n=180)	60.5 ± 3.8	50.6 ± 4.5		5 year	10 year	Low risk (n=60)	93.4 ± 3.2	86.8 ± 4.8	High risk (n=91)	56.2 ± 5.5	44.1 ± 6.3		5 year	10 year	Low risk (n=55)	95.2 ± 2.6	82.7 ± 7.8	High risk (n=89)	66.3 ± 5.2	56.7 ± 6.4
	5 year	10 year																																
<b>All patients</b>																																		
Low risk (n=115)	94.7 ± 2.1	85.2 ± 4.3																																
High risk (n=180)	60.5 ± 3.8	50.6 ± 4.5																																
	5 year	10 year																																
Low risk (n=60)	93.4 ± 3.2	86.8 ± 4.8																																
High risk (n=91)	56.2 ± 5.5	44.1 ± 6.3																																
	5 year	10 year																																
Low risk (n=55)	95.2 ± 2.6	82.7 ± 7.8																																
High risk (n=89)	66.3 ± 5.2	56.7 ± 6.4																																

<sup>8</sup> The complete result set was not presented in the table due to the overwhelming amount of data

				<b>Overall survival (% , ± SE)</b> 5 year                      10 year <b>All patients</b> Low risk (n=115)      94.7 ± 1.5      94.5 ± 2.6 High risk (n=180)      74.1 ± 3.3      54.6 ± 4.4  <u>High vs low risk: overall survival</u> HR = 8.6                      95%CI [4.0, 19.0], <i>p</i> <0.001  <b>Lymph node -ve patients</b> Low risk (n=60)      96.7 ± 2.3      96.7 ± 2.3 High risk (n=91)      71.5 ± 4.8      49.6 ± 6.1  <b>Lymph node +ve patients</b> Low risk (n=55)      98.2 ± 1.8      92.0 ± 4.8 High risk (n=89)      76.5 ± 4.6      59.5 ± 6.3
Buyse et al 2006 TRANSBIG consortium, Multi-centre trial	III-3	Retrospective cohort	307 patients with lymph node negative, breast cancer, tumour size <5cm and less than 61 years old who had not received adjuvant chemotherapy Median follow-up 13.6 years.  Samples analysed using MammaPrint® and conventional histopathology methods: oestrogen receptor status (ER) and histologic grade using Elston and Ellis method.	<b>Clinical low risk (n=80)</b> <u>MammaPrint®</u> <u>Proportion alive at 10 years</u> Low risk (n=52)                      88%, 95%CI [74, 95] High risk (n=28)                      69%, 95%CI [45, 84]  <b>Clinical high risk (n=222)</b> <u>MammaPrint®</u> <u>Proportion alive at 10 years</u> Low risk (n=59)                      89%, 95%CI [77, 95] High risk (n=163)                      69%, 95%CI [61, 76]  <b>MammaPrint®</b> <u>High vs low risk: distant metastases in 5 years</u> HR = 2.32                      95%CI [1.35, 4.0], <i>p</i> = 0.002  <u>High vs low risk: overall survival</u> HR = 2.79                      95%CI [1.6, 4.87], <i>p</i> <0.001  <u>High vs low risk: disease free survival</u> HR = 1.5                      95%CI [1.04, 2.16], <i>p</i> =0.032

SE = standard error

Glas et al (2006) have reported on an adaptation of the MammaPrint® assay to enable high throughput processing of a high volume of samples on a routine basis (Glas et al 2006). MammaPrint® microarrays contain 25,000 oligonucleotides, 60 bases in length, of the 70 prognostic genes. The MammaPrint® assay allows only one sample per chip and turn around time from point of sampling to a result is lengthy at 10 days. Agilent Technologies Inc has developed a mini-array, with eight identical regions or sub-arrays, containing only 1,900, 60-base, oligonucleotides of the 70 genes of interest. This allows for eight individual hybridisations to be carried out and should shorten processing time leading to the more rapid production of results. In addition, the high sample volume throughput should result in reduced costs.

The original 78 breast cancer samples on which the MammaPrint<sup>®</sup> assay is based were retrieved and run on the new mini-array. The results correlated well with the original published data (Pearson correlation =0.92,  $p<0.0001$ ). The overall accuracy of the new assay was determined by calculating the odds ratio for developing distant metastases within 5-years. The odds ratio for the mini-array (OR=13, 95%CI [3.9, 44.0]) compared favourably with that calculated from the original data (OR= 15, 95%CI [2.1, 19.0]). To further validate the mini-array, the 151<sup>9</sup> lymph node negative women from the van de Vijver 2002 study were assayed. The data generated with the mini-array correlated well with the data from the 2002 study (Pearson correlation =0.88,  $p<0.0001$ ). The hazard ratio for distant metastases (high vs low-risk) was 5.6 (95%CI [2.4, 7.3],  $p<0.0001$ ), which again compared favourably to the original published data of HR =5.5 (95%CI [2.5, 12.2],  $p<0.001$ ). The authors conclude that that the mini-array is as robust as the MammaPrint<sup>®</sup> assay and may be a useful clinical tool (Glas et al 2006).

#### *76-gene panel or The “Rotterdam Signature”*

The 76-gene prognostic microarray was developed in the Netherlands and has only three genes in common with the MammaPrint<sup>®</sup> 70-gene panel. The 76-gene assay was developed using a commercially manufactured microarray, the Affymetrix Human U133a gene chip, which contains 22,000 oligonucleotides. The patient population used to validate the 76-gene assay differed markedly than that used to validate the 70-gene panel. The women in the validation study (n=115) by Wang et al (2005) were all lymph node negative and from a wide age range (26 – 83 years), whereas patients in the study by van't Veer et al (2002) were all aged less than 53 years. Another important difference is that the 76-gene panel considered oestrogen receptor positive patients separately from oestrogen negative patients, based on the assumption that disease progression may differ in these subgroups. Of the 70 genes used in the van't Veer et al study, 48 are present on the Affymetrix H133a array, while only 38 of the 76 genes used in Wang's panel are present on the MammaPrint<sup>®</sup> array. Although the two arrays only share three genes in common (cyclin E2, origin recognition complex and TNF super-family protein), recent studies have indicated that the included genes on each array share 21 biological pathways (Blue Cross and Blue Shield Association 2005; Wang et al 2005). Using the St Gallen and NIH criteria, 92 per cent of lymph node negative women would have been candidates for adjuvant therapy (Table 4). Using the 'poor prognosis' 76-gene expression profile, only 112/171 (65%) of women would be candidates for adjuvant therapy, significantly reducing the number of patients who would receive unnecessary treatment (Wang et al 2005).

---

<sup>9</sup> Of the 151 patients only 145 RNA samples were available

**Table 4 Breast cancer patients eligible for adjuvant therapy using 76-gene panel (Wang et al 2005)**

	Patient group		
	Total	Metastatic disease at 5-years	Disease free at 5-years
St Gallen	156/170 (92%)	52/55 (95%)	104/115 (90%)
NIH	153/169 (92%)	52/55 (95%)	101/114 (89%)
Poor prognosis profile with 76-gene panel	112/171 (65%)	52/56 (93%)	60/115 (52%)

Three retrospective studies (level III-3 prognostic evidence), which used the 76-gene panel for the prediction of distant metastases and survival in breast cancer, were included in this *Bulletin* (Table 5) (Desmedt et al 2007; Foekens et al 2006; Wang et al 2005). As with the MammaPrint<sup>®</sup> assay, no prospective, randomised clinical trials have been published. All three studies assayed lymph node negative women. Results were expressed as either a poor prognosis (patients were at high risk of developing distant metastases), or as a good prognosis (patients at low risk of developing distant metastases).

When comparing the high-risk and low-risk groups for the risk of developing distant metastases, the three studies reported consistent hazard ratios of 7.41, 5.78 and 5.67, indicating that women in the high-risk category were more than five times *more likely* to experience recurrence than those in the low risk group. The hazard ratio decreased only marginally after adjusting for clinical criteria in two studies suggesting that differences in the time to distant metastases between the two gene expression groups were largely independent of these criteria. However, in the study by Foekens et al (2006) the adjusted hazard ratio rose markedly from 7.41 to 11.36, but the wide confidence intervals reported for both of these ratios may contribute to this discrepancy.

Overall survival at 10-years was reported by two studies, with hazard ratios of 4.93 and 2.87. Only the study by Desmedt et al (2007) adjusted for clinical criteria and found that as with recurrence, survival was independent of clinical criteria (adjusted HR=2.55 95%CI [1.07,6.1]).

All three studies reported on the sensitivity and specificity of using the 76-gene panel to determine the likelihood of developing distant metastases at 5-years. Sensitivity ranged from 90 to 97 per cent, and specificity ranged from 34 to 48 per cent. A high sensitivity value indicates that the arrays are capable of correctly identifying those at high-risk. As a consequence the remaining population is accurately defined as being at low-risk of distant metastases and therefore may not benefit from treatment. Although the arrays successfully identify those patients at high-risk of recurrence, the poor specificity indicates that a number of low-risk individuals are incorrectly categorised as high-risk and may receive treatment needlessly. This figure, however, is less that would occur using the St Gallen or NIH criteria.

**Table 5 76-gene panel as a predictor of distant metastases in breast cancer**

Study	Level of Prognostic Evidence	Study Design	Population	Outcomes																		
Wang et al 2005, Netherlands	III-3	Retrospective cohort	171 lymph node negative patients with breast cancer who had not received adjuvant chemotherapy Median age at surgery 52 years (range 26-83 years). Median follow-up for patients who survived was 8.4 years (range 1.7-14.3 years).	<p>Validation of 76-gene panel in 171 patients</p> <p><b>Distant metastases in 5 years</b> AUC = 0.694 Sensitivity = 52/56 (93%) Specificity = 55/115 (48%)</p> <p>Patients with relapse score &gt; threshold of assay have OR 11.9 95%CI [4.04, 35.1], <math>p &lt; 0.0001</math> To develop distant metastases within 5 years</p> <p><b>Distant metastases High vs low risk</b> HR = 5.67 95%CI [2.59, 12.4], <math>p &lt; 0.0001</math></p> <p><u>Adjusted for clinical prognostic factors</u> HR = 5.55 95%CI [2.46, 12.5], <math>p &lt; 0.0001</math></p> <p><u>Pre-menopausal patients (n=84)</u> HR = 9.60 95%CI [2.28, 40.5]</p> <p><u>Post-menopausal patients (n=87)</u> HR = 4.04 95%CI [1.57, 10.4]</p> <p><u>Tumour size 10-20mm (n=79)</u> HR = 14.1 95%CI [3.34, 59.2]</p> <p><b>Proportion free of distant metastases</b></p> <table border="0"> <tr> <td></td> <td>5 year</td> <td>6.7 year</td> </tr> <tr> <td>Low risk</td> <td>93%</td> <td>88%</td> </tr> <tr> <td>High risk</td> <td>53%</td> <td>49%</td> </tr> </table> <p><b>Overall survival proportion</b></p> <table border="0"> <tr> <td></td> <td>5 year</td> <td>6.7 year</td> </tr> <tr> <td>Low risk</td> <td>97%</td> <td>95%</td> </tr> <tr> <td>High risk</td> <td>70%</td> <td>63%</td> </tr> </table>		5 year	6.7 year	Low risk	93%	88%	High risk	53%	49%		5 year	6.7 year	Low risk	97%	95%	High risk	70%	63%
	5 year	6.7 year																				
Low risk	93%	88%																				
High risk	53%	49%																				
	5 year	6.7 year																				
Low risk	97%	95%																				
High risk	70%	63%																				

<p>Foekens et al 2006 Multi-centre trial</p>	<p>III-3</p>	<p>Retrospective cohort</p>	<p>180 lymph node negative patients with breast cancer who had not received adjuvant chemotherapy</p> <p>Median age at surgery 56 years (range 30-82 years).</p> <p>Median follow-up for patients who survived was 8.3 years (range 4 -11.4 years).</p>	<p><b>Distant metastases in 5 years</b></p> <p>Sensitivity = 27/30 (90%) Specificity = 69/147 (47%) PPV = 38% 95%CI [29, 47] NPV = 94% 95%CI [86, 97] Patients with relapse score &gt; threshold of assay have OR 8.0 95%CI [2.3, 27.4], <math>p = 0.0004</math> To develop distant metastases within 5 years</p> <p><u>High vs low risk of distant metastases</u> HR = 7.41 95%CI [2.63, 20.9], <math>p = 0.0002</math></p> <p><u>Adjusted for clinical prognostic factors</u> HR = 11.36 95%CI [2.67, 48.4], <math>p = 0.001</math></p> <p><u>High vs low risk: overall survival 10 years</u> HR = 4.93 95%CI [1.47, 16.6], <math>p = 0.0043</math></p> <p><u>High vs low risk: disease free survival 10 years</u> HR = 6.5 95%CI [2.31, 18.3], <math>p &lt; 0.0001</math></p> <p><b>Distant metastases-free survival % [95%CI]</b></p> <table border="1"> <tr> <td></td> <td>5 year</td> <td>10 year</td> </tr> <tr> <td>Low risk</td> <td>96 [89,99]</td> <td>94 [83, 98]</td> </tr> <tr> <td>High risk</td> <td>74 [64, 81]</td> <td>65 [53, 74]</td> </tr> </table>		5 year	10 year	Low risk	96 [89,99]	94 [83, 98]	High risk	74 [64, 81]	65 [53, 74]									
	5 year	10 year																				
Low risk	96 [89,99]	94 [83, 98]																				
High risk	74 [64, 81]	65 [53, 74]																				
<p>Desmedt et al 2007 TRANSBIG consortium, Multi-centre trial</p>	<p>III-3</p>	<p>Retrospective cohort</p>	<p>198 lymph node negative patients with breast cancer, tumour size &lt;5cm.</p> <p>Median age 47 years All patients &lt;61 years.</p> <p>Median follow-up 13.6 years.</p> <p>Samples analysed using 76-gene panel and conventional histopathology methods: oestrogen receptor status (ER) and histologic grade using Elston and Ellis method.</p>	<p><b>Clinical low risk (n=46)</b></p> <p><u>76-gene</u> Low risk (n=16) High risk (n=30)</p> <p><b>Clinical high risk (n=152)</b></p> <p><u>76-gene</u> Low risk (n=39) High risk (n=113)</p> <p><b>Distant metastases-free survival % [95%CI]</b></p> <table border="1"> <tr> <td></td> <td>5 year</td> <td>10 year</td> </tr> <tr> <td>Low risk</td> <td>98 [88,100]</td> <td>94 [83, 98]</td> </tr> <tr> <td>High risk</td> <td>76 [68, 82]</td> <td>73 [65, 79]</td> </tr> </table> <p>Sensitivity 97% 93% Specificity 34% 31%</p> <p><b>Overall survival % [95%CI]</b></p> <table border="1"> <tr> <td></td> <td>5 year</td> <td>10 year</td> </tr> <tr> <td>Low risk</td> <td>98 [88,100]</td> <td>87 [73, 94]</td> </tr> <tr> <td>High risk</td> <td>84 [77, 89]</td> <td>72 [63, 78]</td> </tr> </table> <p><b>76-gene panel</b></p> <p><b>Distant metastases in 10 years</b></p> <p><u>High vs low risk</u> HR = 5.78 95%CI [1.78, 18.8], <math>p = 0.001</math></p> <p><u>Adjusted for clinical prognostic factors</u> HR = 5.11 95%CI [1.57, 16.67]</p> <p><b>Overall survival 10 years</b></p> <p><u>High vs low risk</u> HR = 2.87 95%CI [1.21, 6.82], <math>p = 0.0126</math></p> <p><u>Adjusted for clinical prognostic factors</u> HR = 2.55 95%CI [1.07, 6.10]</p>		5 year	10 year	Low risk	98 [88,100]	94 [83, 98]	High risk	76 [68, 82]	73 [65, 79]		5 year	10 year	Low risk	98 [88,100]	87 [73, 94]	High risk	84 [77, 89]	72 [63, 78]
	5 year	10 year																				
Low risk	98 [88,100]	94 [83, 98]																				
High risk	76 [68, 82]	73 [65, 79]																				
	5 year	10 year																				
Low risk	98 [88,100]	87 [73, 94]																				
High risk	84 [77, 89]	72 [63, 78]																				

AUC = area under the curve, OR = odds ratio

In summary, studies indicate that both the 70-and 76-gene panels are capable of stratifying breast cancer patients into groups that either have a poor or good prognosis for the development of distant metastases, which will impact on their overall survival rate. From this stratification it can be seen that a higher proportion of women considered to be at low risk of recurrence survive at 10-years than those at high risk. However, there are no data available which describe the impact of the 70-gene panel on patient outcomes (avoidance of toxic chemotherapy, disease-free survival and overall survival) by the identification of high (those who will benefit from adjuvant chemotherapy) and low risk women (those unlikely to benefit from adjuvant chemotherapy). DNA microarrays used to predict recurrence of breast cancer have good sensitivity but poor specificity. A high sensitivity value indicates that the arrays can correctly identify those at high-risk. As a consequence the remaining population is accurately defined as being at low-risk of distant metastases and may not benefit from treatment. Although the arrays successfully identifies those patients at high-risk of recurrence, poor specificity indicates that a number of low-risk individuals are incorrectly categorised as high-risk and will receive treatment needlessly.

To date no prospective, randomised clinical trials, assigning patients to chemotherapy regimes based on the results of gene expression assays, have been published.

## DNA microarrays and drug metabolism

---

Depression represents a serious health burden for Australia, accounting for eight per cent of all years lived with disability. The indirect and direct costs of treating and living with depression were estimated to be in excess of A\$3 billion in 1999 (Hickie 2004). The majority of depression episodes are diagnosed and treated by general practitioners, who prescribed 85 per cent of all anti-depressants subsidised by the Australian Pharmaceutical Benefits Scheme (PBS). From 1975 to 1990, the prescription of antidepressants remained stable at approximately 10 defined daily doses (DDD) per 1000 people per day. The annual growth rate during this period was an average of 1.1 per cent. However, in 1990, the first selective serotonin reuptake inhibitor, fluoxetine, was marketed in Australia, followed by moclobemide in 1992, and sertraline and paroxetine in 1994. An inhibitor of serotonin and noradrenalin reuptake, venlafaxine, was released onto the market in 1994. From 1990 to 2002, the use of antidepressants in Australia increased by 352 per cent to 51.5 DDDs per 1000 people per day. By 2002, SSRIs represented 64.5 per cent of all antidepressants sold. It should be noted that these figures are based on sales to pharmacies and therefore do not necessarily represent patient prescriptions (Mant et al 2004).

Neurons send information to, or signal, other neurons by releasing neurotransmitters, of which serotonin is one, into the synapse. Neurotransmitters bind to receptors on the surface of the post-synaptic neuron, stimulating it to relay the signal. Approximately 10 per cent of the neurotransmitter is lost during this process, with the remaining 90 per cent released from the receptors and taken up again by the pre-synaptic neuron via monoamine transporters, a process termed reuptake. Depression has been linked to a lack of stimulation of the recipient neuron at the synapse. To stimulate the recipient neuron, SSRIs *inhibit* the reuptake of serotonin, so that the serotonin stays in the synaptic gap longer. The recipient neuron may now be stimulated several times by the serotonin instead of just once (Wikipedia 2007e).

Although SSRIs are considered more tolerable to a greater proportion of patients than other antidepressants, side effects of taking SSRIs include nausea, diarrhoea and headaches, which may necessitate the patient discontinuing use. In addition, they are associated with high rates of non-responsiveness, with reported response rates of only 47 per cent after 14 weeks of treatment (Matcher et al 2006).

The cytochrome P450 (CYP450) enzymes are members of an isoenzyme super-family that catalyse the oxidation of more than 80 drugs currently in clinical use including antidepressants, anti-psychotics, anti-arrhythmics, analgesics,  $\beta$ -blockers, proton pump inhibitors and some anti-cancer drugs. There are approximately 150 isoforms of the CYP enzyme, encoded for by over 100 genes (Matcher et al 2006). Several enzymes in the CYP2 and CYP3 families are considered to be important in drug metabolism, with two genes of particular interest: CYP2D6 and CYP2C19 (Roche Diagnostics 2006b). Some of the CYP450 genes are highly polymorphic, which are phenotypically expressed by either an increase or a decrease in enzyme activity. This in turn

may lead to altered rates of drug metabolism, resulting in either toxicity from an excess of drug, or at the other end of the metabolism spectrum, a lack of drug efficacy (Table 6). There are racial differences associated with the polymorphic frequency. Approximately 5-10 per cent of Caucasians of European descent are poor (PM) and up to seven per cent ultra-rapid (UM) metabolisers of CYP2D6. The prevalence of PMs and UMs in the Asian population is low (<1%) but the frequency of intermediate metabolisers is high (Matcher et al 2006).

**Table 6** Effects of genetic polymorphisms of CYP enzyme genes on drug metabolism (Matcher et al 2006)

	<b>Genotype</b>	<b>Expected effect on drug metabolism</b>
<b>Ultra-rapid metaboliser (UM)</b>	More than 2 copies of active enzyme gene alleles	Usual doses of drug may not lead to therapeutic drug concentrations, possible no-response
<b>Extensive metaboliser (EM)</b>	2 copies of active enzyme gene alleles	Usual doses lead to expected drug concentrations and response
<b>Intermediate metaboliser (IM)</b>	Homozygous for 2 reduced activity enzyme gene alleles or heterozygous for an inactive allele and a reduced activity allele	Drug effects between those of extensive and poor metabolisers
<b>Poor metaboliser (PM)</b>	Homozygous or compound heterozygote for deficiency alleles	Usual doses may lead to higher than expected drug concentrations and possible adverse effects

A number of P450 enzymes are involved in the metabolism of SSRIs. The CYP2D6, CYP2C19 and CYP2C9 enzymes are involved in the metabolism of all SSRIs. However it should be noted that enzymes other than those encoded by the CYP gene may be involved in drug metabolism and more than one CYP enzyme may be involved in the metabolism of any given SSRI. In addition, some SSRIs are potent inhibitors of CYP enzymes, in particular CYP2D6, which has the effect of raising serum concentrations of the SSRI (Matcher et al 2006).

In December 2004, the FDA approved the Roche AmpliChip™ cytochrome P450 test which utilises the Affymetrix GeneChip 3000Dx microarray platform. The test is designed to investigate the single nucleotide polymorphisms (SNPs) associated with two of the drug metabolising enzyme CYP genes, CYP2D6 and CYP2C19 (Palylyk-Colwell 2006). The AmpliChip™ test analyses 29 polymorphisms of the CYP2D6 gene and two polymorphisms of the CYP2C19 gene to determine a patient's genotype and predicted drug metabolising phenotype (poor, intermediate, extensive, or ultra rapid) (Roche Diagnostics 2006a).

The AmpliChip™ CYP450 test contains over 15,000 different oligonucleotide probes synthesised *in situ* using the photolithographic method. A blood sample is taken from the patient, DNA is isolated and a restriction enzyme digest is performed. Two separate PCR amplifications of the DNA fragments are then performed. Reaction A uses primers specific for the promoter and coding regions of the CYP2D6 gene and a CYP2D6 gene-duplication specific product. Reaction B uses primers to generate products of the three CYP2C19 genes as well as a CYP2D6 gene-deletion specific product. The products of

the two PCR amplifications are pooled and then cleaved to form smaller fragments between 50-200 nucleotides in length, which are labelled with biotin. The labelled fragments are hybridised to the AmpliChip™ CYP450 microarray. Visualisation of hybridised probes is achieved with a streptavidin-conjugated fluorescent dye and the emitted light is proportional to the amount of bound DNA (FDA 2005; FDA 2004). Patients are assigned as either poor, intermediate, efficient and ultra-rapid metabolisers of drugs according to the combination of CYP alleles expressed. The gold standard for the comparison of AmpliChip™ would be regarded as bi-directional DNA sequencing, however methods such as allele-specific PCR and restriction fragment length polymorphism PCR would also be considered acceptable as a reference standard (Matcher et al 2006).

Although the CYP genes are capable of metabolising many different drugs, the AmpliChip™ is being promoted primarily to psychiatrists for the determination of a patient's ability to metabolise certain drugs, in particular SSRIs (Palylyk-Colwell 2006).

The cost of one AmpliChip™ test is approximately US\$500. The AmpliChip™ test must be processed on Affymetrix GeneChip 3000Dx instrumentation which costs approximately US\$200,000. To date, no cost-effectiveness studies on the use of the AmpliChip™ for drug metabolism genotyping have been published (2005a; de Leon 2006).

At least three other microarray products for the detection of single nucleotide polymorphisms have been developed, DrugMEt™, CodeLink™ and PHARMAchip. DrugMEt™ is manufactured by JuriLab Ltd and can detect 27 SNPs involved in drug metabolism. DrugMEt™ received the European CE mark in May 2006. The CodeLink™ P450 SNP Bioarray is manufactured by General Electric Health Care and is capable of genotyping 68 SNPs (Amersham Biosciences 2003; Hendolin 2005). The PHARMAchip is manufactured by Progenika and allows screening for 90 SNPs of genes involved in a drugs response, including the CYP450 enzymes, drug transport, neurotransmitter receptor genes (Progenika Biopharma 2007). These assays do not have FDA approval.

<u>DrugMEt™</u>		<u>CodeLink™</u>	
Targeted gene	Number of SNPs	Targeted gene	Number of SNPs
CYP2B6	2	CYP1A1	3
CYP2B9	2	CYP1A2	2
CYP2C19	2	CYP1B1	16
CYP2D6	12	CYP2C19	5
CYP3A5	1	CYP2D6	29
TPMT	3	CYP2E1	11
NAT2	4	CYP3A4	2
MDR1	1		

Four studies were included for assessment in this *Bulletin* which described the use of the AmpliChip™ CYP450 test for the genotyping of the drug metabolism CYP genes, thus categorising patients as either poor, intermediate, efficient or ultra-rapid metabolisers of drugs according to the combination of CYP alleles expressed (Table 7).

Two studies were submitted to the FDA in the product approval application, one for the CYP2D6 component of the test and one for the CYP2C19 component. These two submissions were assessed in this *Bulletin* as if they were one study (FDA 2005; FDA 2004). All studies reported on the accuracy of the AmpliChip™ CYP450 assay in detecting genetic variations in patients when compared to standard methods of genetic testing including bi-directional DNA sequencing and allele specific PCR (level III-1 diagnostic evidence). In addition, one paper included a case-control component (level III-3 diagnostic evidence), which reported on the poor metaboliser phenotype and its association with adverse drug reactions. This study only genotyped polymorphisms of the CYP2D6 gene (de Leon et al 2005). One study reported on the use of the GeneChip, a predecessor of the AmpliChip™ which only genotyped polymorphisms of the CYP2D6 gene (Chou et al 2003).

No studies have been published that report on improvements in patient outcomes of patients on SSRI therapy who have been genotyped versus those of patients on SSRI therapy who have not been genotyped.

Chou et al (2003) reported a high concordance between the AmpliChip™ and AS-PCR methods when five major alleles were genotyped (CYP2D6 \*3, \*4, \*6 and \*9) in 236 healthy individuals. The two studies submitted by Roche Diagnostics to the FDA for product approval also reported high sensitivity, specificity, positive predictive and negative predictive values (>98%) when the AmpliChip™ was compared to both bi-directional sequencing and AS-PCR for genotyping both CYP2D6 and CYP2C19 polymorphisms. From these studies it appears that the AmpliChip™ has the ability to accurately identify genotypes and the assay is reproducible. However, the report prepared by the Agency for Healthcare Research and Quality states that one problem associated with the AmpliChip™, despite the large number of CYP2D6 variants tested for, is its failure to capture a considerable number of rare variants associated with deficient CYP enzyme activity (Matcher et al 2006).

In the study by Chou et al (2003) individuals were categorised into metabolising groups according to their genotype (poor, intermediate, efficient and ultra metabolisers) and administered a dose of dextromethorphan (DXT), which is metabolised to dextrorphan (DRP). Those individuals classified as poor metabolisers had a low mean urinary recovery of DRP (0.93%) indicating that the rate of DXT metabolism to form DRP was slow. However, there was no significant difference between the mean urinary recovery of DRP from those patients classified as either intermediate (18.6%), efficient (26.6%) and ultra (16.7%) metabolisers when corrected for multiple comparisons. These results were obtained with an assay that genotyped only CYP2D6. Additional genotyping for the CYP2C19 gene may have been able to further clarify the metaboliser groups giving a more accurate classification.

The poor sensitivity values reported by de Leon et al (2005) do not reflect on the ability of the AmpliChip™ to test patients for the CYP2D6 genotype,

instead indicates how efficient the test is in identifying those individuals with adverse drug reactions (to risperidone) who discontinued use of the drug *due* to an adverse drug reaction. The accuracy (the proportion of all tests giving the correct result) of the AmpliChip™ test was high (>94%). When a univariate analysis of the data from this study was performed, comparing poor metabolisers to non-poor metabolisers, an odds ratio of 3.4 and 3.1 was obtained for the patients stabilised on risperidone, for those with and without an adverse drug reaction, respectively. A similar odds ratio (3.0) was obtained in the group of patients who discontinued risperidone treatment for reasons other than an ADR, however those patients who discontinued risperidone treatment due to an adverse drug reaction had a much higher odds ratio of 6.0. Regardless of treatment group, this analysis demonstrates that individuals are at least three times more likely to have an ADR if they have the poor metaboliser CYP2D6 genotype.

In summary, the AmpliChip™ appears to be an accurate means of genotyping patients into poor, intermediate, efficient and ultra metabolisers of drugs. High sensitivity and specificity values when the AmpliChip™ is compared to the gold standards of bi-directional DNA sequencing and allele-specific PCR indicate that the test can successfully identify individuals who are poor metabolisers. However, there is a lack of clinical data concerning whether or not genotyping patients has an impact on patient outcomes.

The Roche AmpliChip™ CYP450 test can only be analysed on the Affymetrix GeneChip® system 3000Dx, which limits the implementation of the test. Tests such as AS-PCR and DNA sequencing, although complex and requiring skilled staff, are standard techniques widely available in pathology laboratories.

Many authors expressed the view that although personalised medicine (the adjustment of a patient's treatment regime based on the results of genotyping *not* just pharmacogenetic testing) was likely to gain increased recognition within the next five years, that further research into this area is required (de Leon 2006; Jain 2005).

**Table 7 Genotyping for polymorphisms of CYP2D6 and CYP2C19 genes using AmpliChip™**

Study	Level of Diagnostic Evidence	Study Design	Population	Outcomes
Chou et al 2003 United States	III-1	Cross classification of patients for CYP2D6 genotype on GeneChip AmpliChip™ predecessor and AS-PCR.  CYP2D6 alleles *3, *4, *6, *7, *8, *9, *10 and *11 tested by GeneChip  CYP2D6 alleles *3, *4, *6, *7, *9, *17 and *41 tested by AS-PCR  CYP2D6 alleles *3, *4, *6, *7 and *9 tested by both AS-PCR and GeneChip	236 healthy individuals, prescription and non-prescription drug free for 14 days. DNA isolated then subjects given a 60mg dose of dextro-methorphan. to determine metabolic rate of drug  Ethnicity: 87.3% Caucasian, 5.5% African-American, 1.7% Hispanic, 1.2% Asian, 2.5% multi-racial	<b>CYP2D6</b> Sensitivity 100%, 95%CI [96.7, 100] Specificity Not reported PPV 100% NPV 100%  <b>Genotype classification</b> 234/236 (99%) could be classified  <u>Ultra-rapid metabolisers (UM)</u> At least 3 functional copies of CYP2D6 gene ie *1/*1xn, *1/*2xn, *2/*1xn or *2/*2xn 5/234 (2.1%) Mean urinary recovery DRP = 16.7%  <u>Poor metaboliser (PM)</u> Carrying any 2 of the following CYP2D6 alleles: *3, *4, *4xn, *5, *6, *7, *8, or *11 10/234 (4.3%) Mean urinary recovery DRP = 0.93%  <u>Intermediate metaboliser (IM)</u> Carrying 1 CYP2D6 allele with diminished activity ie *9, *10, *17 or *41 and 1 non-functional allele : ie *3, *4, *4xn, *5, *6, *7, *8 or *11 20/234 (8.5%) Mean urinary recovery DRP = 18.6%  <u>Efficient metaboliser (EM)</u> 199/234 (85.0%) Mean urinary recovery DRP = 26.6%
(FDA 2005) for CYP2C19  (FDA 2004) for CYP2D6  Both reported in (Matcher et al 2006)	III-1	Cross classification of patients on AmpliChip™, bi-directional sequencing and PCR-RFLP	<b>CYP2D6 analysis:</b> 246 patients AmpliChip™ and bi-directional sequencing. 403 patients AmpliChip™ and PCR-RFLP.  <b>CYP2C19 analysis:</b> 103 patients AmpliChip™ and bi-directional sequencing. 798 patients AmpliChip™ and PCR-RFLP.  No information on ethnicity of patients was provided.	<b>CYP2D6 vs sequencing</b> Sensitivity 99.3%, 95%CI [97.5, 99.8] Specificity 99.0%, 95%CI [94.7, 99.8] <b>vs PCR methods</b> Sensitivity 99.2%, 95%CI [98.0, 99.6] Specificity 99.5%, 95%CI [97.4, 99.9]  PPV 99.22% NPV 99.02% Within & between laboratory precision 99.7% Reproducibility 99.6%, 95%CI [98.9, 99.8]  <b>CYP2C19 vs sequencing</b> Sensitivity 98.9%, 95%CI [94.2, 99.8] Specificity 100%, 95%CI [97.6, 100] <b>vs PCR methods</b> Sensitivity 100%, 95%CI [97.5, 100] Specificity 100%, 95%CI [99.4, 100]  PPV 100% NPV 99.6% Within & between laboratory precision 99.6% Reproducibility 99.6%, 95%CI [98.9, 99.9]

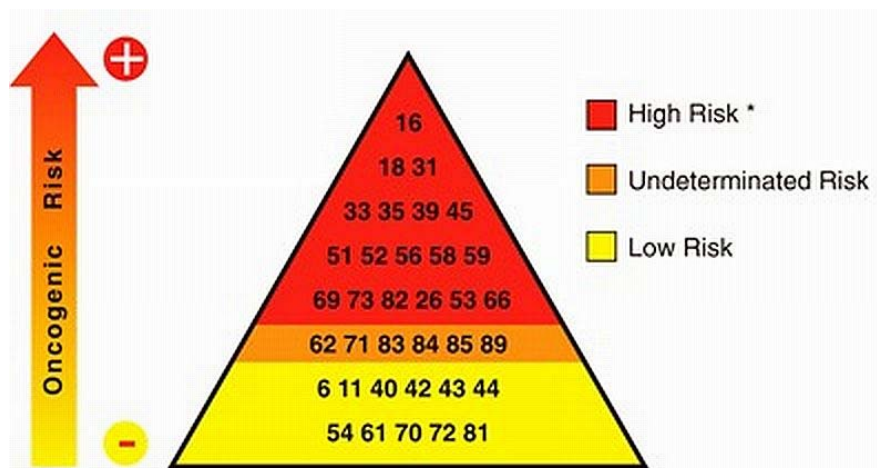
De Leon et al (2005) United States and de Leon et al (2006)	III-1	Cross-classification of patients on AmpliChip™ and AS-PCR and	325 patients stabilised on risperidone therapy. 73/325 (22%) had moderate to marked ADR 252/325 (78%) no ADR	<b>Genotyping for CYP2D6</b> <u>Patients taking risperidone (n=325)</u> Sensitivity 16% Specificity 77% Accuracy 94% Likelihood ratio Positive 2.7 Negative 0.89
	III-3	Case-control study: patients taking risperidone vs those not taking risperidone	212 patients who had been discontinued from risperidone treatment. 81/212 (38%) discontinued due to ADR, 131/212 (62%) discontinued for other reason. All patients genotyped for CYP2D6 polymorphism	<u>Patients NOT taking risperidone (n=212)</u> Sensitivity 9% Specificity 63% Accuracy 97% Likelihood ratio Positive 3.0 Negative 0.94  After correction for clinical variables <u>Stabilised on risperidone</u> Poor metabolisers 27/325 (8.3%) Non-poor metabolisers 298/325 (91.7%)  No ADR Poor metabolisers 15/27 (55.6%) OR = 3.1, 95%CI [1.4, 7.0]  Moderate ADR Poor metabolisers 12/27 (44.4%) OR = 3.4, 95%CI [1.5, 8.0]  <u>Discontinued risperidone</u> Poor metabolisers 11/212 (5.2%) Non-poor metabolisers 201/212 (94.8%)  Not due to ADR Poor metabolisers 4/11 (36.4%) OR = 3.0, 95%CI [0.9, 10.6]  Due to ADR Poor metabolisers 7/11 (63.6%) OR = 6.0, 95%CI [1.4, 25.4]

AS-PCR = allele specific PCR, PCR-RFLP = restriction fragment length polymorphism PCR, ADRs = adverse drug reactions, DRP = dextropropranolol

## Other applications of DNA microarrays

### *Human papillomavirus (HPV)*

A Spanish firm has developed a DNA microarray for the detection of 35 HPV genotypes, which according to the assay information sheet, categorises the patient into low, indeterminate or high oncogenic risk (Figure 8) by the pattern of gene expression (Genomica 2007a). DNA can be obtained from liquid cytology, cervical swabs and paraffin embedded tissue samples.



**Figure 8** Expression of HPV SNPs associated with an oncogenic increased risk (Genomica 2007a)

Several studies have reported on the detection of HPV using the EasyChip<sup>®</sup> HPV blot (39 HPV genotypes), however, a search for this product and the manufacturing company yielded no results. This may indicate that the product is now manufactured by another company. One study reported the HPV status using the EasyChip<sup>®</sup> of 2,118 Chinese women who had surgery for Stage I-IV cervical cancer (Lai et al 2007). HPV DNA sequences were detected in 96 per cent of these women. Only samples with a multiple genotype (13.8% of total) underwent type-specific PCR. The HPV assay was not compared to any other genotyping method therefore no data could be reported on the sensitivity and specificity of the assay. Lin et al (2006) reported that the overall sensitivity of the EasyChip<sup>®</sup> was 1-50 copies of the HPV genome equivalent, and that intra- and inter-batch reproducibility was 98 and 97 per cent, respectively. This was achieved by cloning all of the 39 HPV genotypes into plasmids, with the genotypes confirmed by DNA sequencing. A human DNA sample was spiked with each of the 39 genotype plasmids in serial dilutions. PCR was performed in triplicate and the amplified products hybridised on the Easy Chip<sup>®</sup> (Lin et al 2007).

A 2006 study by Huang et al compared the EasyChip<sup>®</sup> to the established Hybrid Capture II (HC-II) test for the detection of HPV genotypes in women with normal (n=146) and abnormal (n=208) cytology. The HC-II test is a rapid liquid hybridisation assay which can detect 18 HPV genotypes (13 high risk: HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 & 68, and 5 low risk: 6, 11, 42, 43 & 44), which may be used for biopsy

specimens but is primarily used for cervical scrapes. Compared with PCR-ELISA, the sensitivity and specificity of the HC-II test is 92 and 95 per cent, respectively (Venturoli et al 2002). The concordance rate between the two methods was high (81%) (Table 8). There were 15/354 (4.2%) samples found to be negative for HPV with HC-II but positive with the EasyChip<sup>®</sup> due to genotypes *not* included in the HC-II assay. Samples were confirmed to be positive by direct DNA sequencing. In addition, there were 18/354 (5.1%) of samples found to be positive with HC-II but negative with the EasyChip<sup>®</sup>. These samples were confirmed to be negative when sequenced (Huang et al 2006). Similar results were reported in an earlier study by the same group (Huang et al 2004).

Many other studies have reported on HPV genotype detection using “in-house” microarrays (Choi et al 2005; Klaassen et al 2004; Oh et al 2004; Prinsen et al 2007).

**Table 8 HPV genotyping with DNA microarrays**

Study	Level of Diagnostic Evidence	Study Design	Population	Outcomes															
Huang et al (2006) Taiwan	III-2	Cross classification study of patients on HC-II test and EasyChip <sup>®</sup> microarray. Discrepancies were evaluated by either direct sequencing, type specific PCR or HPV blot	354 cervical swabs 143 normal cytology 208 abnormal cytology of which: 105 ASCUS 94 LSIL 12 HSIL	<p><b>HPV positive</b></p> <table border="1"> <thead> <tr> <th></th> <th>HC-II n(%)</th> <th>EasyChip<sup>®</sup> n(%)</th> </tr> </thead> <tbody> <tr> <td>Normal</td> <td>18 (12.6)</td> <td>26 (18.2)</td> </tr> <tr> <td>ASCUS</td> <td>38 (36.2)</td> <td>48 (45.7)</td> </tr> <tr> <td>LSIL</td> <td>54 (57.4)</td> <td>54 (57.4)</td> </tr> <tr> <td>HSIL</td> <td>10 (83.3)</td> <td>9 (75.0)</td> </tr> </tbody> </table> <p>Overall detection rates            HC-II 120/354 (33.9%)            EasyChip<sup>®</sup> 137/354 (38.7%)</p> <p>Concordance of 2 methods 286/354 (80.8%)            Agreement of 2 methods <math>\kappa = 0.68</math></p>		HC-II n(%)	EasyChip <sup>®</sup> n(%)	Normal	18 (12.6)	26 (18.2)	ASCUS	38 (36.2)	48 (45.7)	LSIL	54 (57.4)	54 (57.4)	HSIL	10 (83.3)	9 (75.0)
	HC-II n(%)	EasyChip <sup>®</sup> n(%)																	
Normal	18 (12.6)	26 (18.2)																	
ASCUS	38 (36.2)	48 (45.7)																	
LSIL	54 (57.4)	54 (57.4)																	
HSIL	10 (83.3)	9 (75.0)																	
Huang et al (2004) Taiwan	III-2	Cross classification study of patients on direct sequencing and EasyChip <sup>®</sup> microarray.	149 paraffin embedded samples from patients with stage IB or IIA	<p><b>HPV positive</b>            149/149 (100%)            101/149 (67.8%) were single mutations            48/149 (32.2%) were multiple mutations</p> <p>HPV-16 detected in 98.7% samples            HPV-18 detected in 22.8% samples</p> <p>Concordance of 2 methods 120/149 (80.5%)            Agreement of 2 methods <math>\kappa = 0.63</math> (<math>p &lt; 0.0001</math>)</p> <p>Infection with HPV-18 or HPV-16 &amp; HPV-18 had a significant effect on overall survival            RR= 2.33, 95% CI [1.17, 4.64] <math>p = 0.016</math></p>															

HC-II, hybrid capture II assay, ASCUS = atypical squamous cells of undetermined significance, LSIL = low grade squamous intraepithelial lesions, HSIL = high grade squamous intraepithelial lesions, RR= relative risk

### Familial hypercholesterolemia

A European firm, Progenika Biopharma has developed the LIPOchip<sup>®</sup>, a DNA microarray for the diagnosis of familial hypercholesterolemia (FH). The LIPOchip<sup>®</sup> received CE-marking in June 2004 for routine clinical use (Progenika Biopharma 2007). Conventional clinical diagnosis of FH is

difficult, and the high number (>840) of reported mutations in the low-density lipoprotein receptor (LDLR) gene makes conventional genetic testing expensive (eg sequencing) and time consuming. Familial defective apolipoprotein B-100 (FDB) may also cause hypercholesterolemia and results from a missense mutation in the apolipoprotein B (APOB) gene. FDB can only be distinguished from FH by genetic diagnosis (Tejedor et al 2005). The LIPOchip<sup>®</sup> is able to detect 230 single nucleotide polymorphisms in the LDLR gene and four SNPs in the apolipoprotein B (APOB) gene. Mutations not currently identified by the LIPOchip<sup>®</sup> are continuously being added to the microarray (Progenika Biopharma 2007).

Data reported by Progenika Biopharma found that in a group of 800 patients, the LIPOchip<sup>®</sup> was able to detect at least one mutation in 65 per cent of patients who were clinically diagnosed, and a mutation in 30 per cent of patients with a probable/possible diagnosis of FH (Progenika Biopharma 2007).

A Spanish study that was used as the basis for the LIPOchip<sup>®</sup> product was conducted on 407 non-genotyped patients of whom 252 (62%) had a definite clinical diagnosis of FH and 155 (38%) had a probable diagnosis (level IV diagnostic evidence). A microarray was constructed which was capable of detecting 117 SNPs in the LDLR gene and one SNP in the APOB gene. Using this microarray, 1,180 previously DNA sequenced samples were processed yielding a sensitivity and specificity of 99.9 and 99.7 per cent, respectively. The microarray identified at least one mutation in 51.2 per cent of patients with a definite clinical diagnosis and in 37.4 per cent of probable FH cases, providing a definite diagnosis in these patients. When DNA sequencing was conducted on the remaining 123 patients with a definite clinical diagnosis, 28 mutations not previously detected in the Spanish population were identified in 43 patients. The remaining patients were expected to have a mutation in other loci associated with FH (Tejedor et al 2005). It would be expected that an increase in the number of SNPs added to the LIPOchip<sup>®</sup> would increase the efficiency of the microarray in detecting patients with FH.

### *Blood typing*

The same company, Progenika Biopharma, has a number of products in development. In particular, the BLOODchip<sup>®</sup>, is a blood group genotyping DNA microarray for the detection of 116 SNPs in the ABO, RhD, RhCE, Kell, Kidd, Duffy, MNS, Diego, Dombrock and Colton blood group genes. It is hoped that this product will assist in the prevention of alloimmunisation reactions that currently occur in approximately three per cent of individuals transfused with blood typed by conventional antibody based agglutination assays (Progenika Biopharma 2007). Transfused blood is currently tested for minor blood group antigens due to time and cost constraints.

Alloimmunisation against one or more minor antigens may complicate future transfusions. A number of studies have reported on the feasibility of blood typing using DNA microarrays (Beiboer et al 2005; Hashmi et al 2005; Montpetit et al 2006). An application for CE-marking for the BLOODchip<sup>®</sup> is in process and the product is expected to be marketed in September 2007. A large cohort of 3,000 blood donor samples is currently being tested as part of the product validation process (Nogues 2007; Progenika Biopharma 2007).

Progenika Biopharma is also developing a product called the IBDchip<sup>®</sup> for the identification of SNPs associated with a susceptibility to inflammatory bowel disease.

### *Burkitt's Lymphoma*

Burkitt's lymphoma (BL) and diffuse large-B-cell lymphoma (DLBCL) are both rare, aggressive B-cell lymphomas, which, if left untreated, are rapidly fatal. BL occurs in approximately 30-50 per cent of all lymphomas in children and 1-2 per cent of adults. DLBCL is more than 20 times as common as Burkitt's lymphoma (Dave et al 2006; Harris & Horning 2006; Hummel et al 2006). In Australia in 2001, a total of 3,900 individuals were diagnosed with an immuno-proliferative lymphoma, representing an incidence of 20.7 per 100,000 population. The mortality rate for the same period was 8.1 per 100,000 population (AIHW 2004). Treatment and management of the two diseases differ markedly. BL responds well to chemotherapy with methotrexate and cytarabine, with cure rates of approximately 90 and 70 per cent in children and adults, respectively. The treatment for DLBCL consists of cyclophosphamide, doxorubicin, vincristine, prednisone and the monoclonal anti-B-cell antibody, rituximab. This treatment regime will not cure or benefit BL patients. Approximately 30 per cent of patients with DLBCL do not respond to treatment and relapse. Differentiation and correct diagnosis of the two diseases is of utmost importance. The distinction between the two diseases by morphology, immunophenotype and genetic abnormalities is not reliably reproducible as they share many features in common. In *almost all* BL cases the locus of the *myc* oncogene is translocated to one of three immunoglobulin loci. However, *myc* translocations may also occur in DLBCL (Dave et al 2006; Harris & Horning 2006; Hummel et al 2006).

Two studies constructed a DNA microarray with a "Burkitt's lymphoma signature" and compared the results to conventional diagnosis of samples with morphology and pathology (Dave et al 2006; Hummel et al 2006). The retrospective study by Dave et al (2006) found that of 71 cases originally classified as BL, only 52 (73%) had the BL genetic signature (level III-2 diagnostic evidence). When these samples were re-classified by an expert panel of haematologists, only 45/71 (63.4%) of the original sample group were classified as having BL, and 44/45 (98%) of these samples had the BL gene signature. However, eight specimens that were pathologically diagnosed as DLBCL had the BL gene signature. This suggests that these patients would be difficult to diagnose with conventional methods and would benefit from an alternative treatment regime. The retrospective study by Hummel et al (2006) analysed 220 samples classified by conventional pathology as mature aggressive B-cell lymphomas. The BL gene signature was identified in 44/220 (20%) of samples and 11 of these were classified morphologically as DLBCL. Five of these samples did not have the *myc* translocation, and conversely of the 155 lymphomas classified as DLBCL by the microarray, 33 (21%) had a *myc* translocation, confirming that this characteristic alone can not be used as a diagnostic measure.

Although these studies are reporting early work in this field, they demonstrate the usefulness of DNA microarrays in identifying gene signatures in rare diseases which are difficult to diagnose with conventional means.

### *Single nucleotide polymorphisms (SNPs)*

SNPs occur when a single nucleotide is replaced by another eg in the small DNA sequence AAGGTTA is changed to ATGGTTA, with the second adenine base changed to a thymine. SNPs occur in the genome frequently (approximately once every 1,200 bases), however the majority are likely to occur outside the coding regions of DNA. Many genetic disorders such as cystic fibrosis, are caused by several different mutations, however disease such as diabetes and coronary artery disease may be caused by a combination of genetic and environmental factors. SNPs associated with these diseases may identify an individual's genetic predisposition, or the potential to develop the disease at some time in their life. DNA polymorphisms associated with disease may be identified by comparing differences in DNA sequences in a large group of individuals with the disease to those without, creating an association pattern. Once an SNP pattern is established for a given disease, DNA microarrays may be used to test an individual's gene expression to determine susceptibility to disease (NCBI 2004). The International HapMap Project is an international (Japan, the United Kingdom, Canada, China, Nigeria, and the United States) project that aims to identify genetic similarities and differences in humans. A database has been constructed that catalogues differences in genes which may in turn affect health, and individual responses to medications and environmental factors. Recent published data from the HapMap Project reported that more than one million SNPs were found in 269 DAN samples from four different populations. Identification of all SNPs would only be possible by complete DNA sequencing, which is impractical and expensive (2005b).<sup>10</sup>

However, identification of particular SNPs is possible with DNA microarrays, such as those described in the drug metabolism section of this *Bulletin*. Large scale screening of SNPs may become possible in the future utilising products such as the Affymetrix GeneChip<sup>®</sup> Human Mapping 500K array, which is currently for research not clinical use. This array is capable of genotyping 250,000 SNPs, which were identified from the >1.2 million SNPs held on the SNP repository database. Results from research with this chip may find their way into clinical practice in the future (Affymetrix 2007b).

The recent Nature paper published by the Wellcome Trust Case Control Consortium, described the use of the Affymetrix GeneChip<sup>®</sup> Human Mapping 500K array to genotype individuals in a large British case-control study (2007). Seven common, but complex diseases of major public health importance were targeted: bipolar disorder, coronary heart disease, Crohn's disease, hypertension, rheumatoid arthritis and type I and type II diabetes<sup>11</sup>. A

---

<sup>10</sup> James Watson's DNA was the first individual genome to be sequenced. Sequencing took 2 months to complete and cost US\$1 million (<http://www.nature.com/news/2007/070528/full/070528-10.html>)

<sup>11</sup> Additional studies undertaken by this group include a study for tuberculosis using samples from The Gambia, and association studies for breast cancer, multiple sclerosis, ankylosing spondylitis and autoimmune thyroid disease.

total of 17,000 samples were genotyped: 3,000 control individuals chosen from the 1958 British Birth Cohort and from blood donors, and 2,000 cases for each disease. This study identified 24 independent *association* signals ( $p < 5 \times 10^{-7}$ ): one in both bipolar disorder and coronary artery disease, nine in Crohn's disease, three in rheumatoid arthritis, seven and three in type I and type II diabetes, respectively. A large number of other signals were identified as likely to yield additional susceptibility loci. Despite these major findings, these results require further validation and research and are only a first step in a long process before meaningful clinical applications can be found.

### *Miscellaneous*

DNA microarrays have proved useful in the genotyping of bacterial strains. Clinicians in Germany are using microarrays to screen clinical isolates of *Staphylococcus aureus* for resistance genes and virulence factors once the strain has been cultured by conventional microbiology methods. This technique has been used in cases of unusual clinical presentation such as suspected toxic shock syndrome to detect genes known to be responsible for the condition and thus rule in or rule out a diagnosis. In addition, microarrays have been used in the study of the epidemiology of multiple resistant strains of *S. aureus* (MRSA). The information gained from typing the characteristics of strains can be used in outbreak investigations and monitoring the movement of specific strains in and out of communities such as hospitals (personal communication, Institute for Medical Microbiology and Hygiene, Desden, Germany) (Monecke et al 2007). This approach may be proved useful in the serotyping of *E. coli* (Ballmer et al 2007; Anjum et al 2007), the differentiation of Chlamydia strains (Sachse et al 2006) and a range of other bacteria carrying resistance genes (Perreten et al 2005) with not only clinical applications but in the agriculture and veterinary fields as well.

Genomica have two other products which have received CE-marking: Clinical Arrays<sup>®</sup> PneumoVir for the diagnosis of respiratory viruses and Clinical Arrays<sup>®</sup> MetaBone for the identification of SNPs associated with loss of bone mineral density and fracture risk. The PneumoVir test uses DNA extracted from nasopharyngeal washes to detect 17 types and subtypes of respiratory virus including influenza A, B and C; parainfluenza 1, 2, 3, 4a and 4b; and respiratory syncytial virus A and B (Genomica 2007c). The MetaBone test uses DNA extracted from a blood sample and detects SNPs of the collagen 1A1, vitamin D receptor, oestrogen receptor and calcitonin receptor genes (Genomica 2007b).

## Potential Cost Impact

---

Costs for “in-house” spotted DNA microarrays are estimated to be A\$200 per DNA chip. DNA chips can be produced by a robotic spotter and read by a laser scanner, which costs approximately \$280,000 and A\$80,000 respectively (personal communication University of Adelaide). The Affymetrix Dx system is a complete instrumentation system which comprises an array scanner, a FS450 Fluidics station capable of processing four arrays per wash/stain run (capable of multiple runs per day with an average processing time of 1-1.5 hours), a hybridisation oven, computer work station and operating software. The cost for the complete system is approximately A\$335,000. The price of an Affymetrix gene expression array is A\$270 -550 depending on the complexity of the target. A custom made array can cost between A\$30,000 for 90 arrays targeting 1,700 genes (\$365 per array with approximately 38,000 oligonucleotides) and up to >A\$200,000 for 180 arrays containing millions of oligonucleotides. The cost per array increases as the number of oligonucleotides per array increases (personal communication July 2007, Millenium Science Pty Ltd).

## Ethical Considerations

---

### *Ethical issues raised by genetic testing with DNA microarrays*

This bulletin has shown that the possibilities for using DNA microarrays are vast. There is little doubt that DNA microarrays will revolutionise the ability to assess changes in gene expression on a mass, high-throughput scale. The power to obtain a vast amount of genetic information on an individual in a short period of time is at the heart of any discussion of the ethical issues surrounding genetic testing using DNA arrays (Grody 2003).

At this stage, there is little evidence available about the clinical applications of DNA microarrays for prevention, diagnosis and treatment. Despite this lack of evidence, ethical analysis of DNA microarray technology has focused mainly on the potential future clinical uses for the technology, rather than the current technical state of knowledge. Ethicists have argued that the potential ethical issues are so significant that they need to be addressed *before* the technology becomes available routinely, in order to ensure that appropriate regulatory and clinical frameworks are in place. The Ethical, Legal and Social Implications (ELSI) program within the US Human Genome Project was created with this issue in mind (Human Genome Project 2004).

This section very briefly outlines the ethical issues that may arise as clinical applications of DNA microarrays become available. It does not offer an ethical analysis of these applications, nor does it consider strategies to address the ethical issues as they arise.

### *Informed consent*

Both clinical and research practice in healthcare is underpinned by an ethical requirement for the informed consent of patients and research participants. Informed consent has three components: a voluntary decision, understanding of what the decision entails (including the risks and consequences), and lack of coercion. The first two of these components may be at risk if DNA microarrays come to be used in diagnosis and screening.

### Voluntariness

Until now, decisions with patients about diagnostic and screening tests for genetic disorders have been taken on a test by test basis, one disorder at a time. The introduction of DNA microarray technologies will create an environment in which screening for thousands of genetic variations is possible on a single occasion. In addition, because specimens obtained for other purposes and at other times could be used for analysis, strict privacy regulations would need to be in place to ensure that patient samples are available for testing without consent years after samples were taken. These factors are likely to make it impossible to seek the informed consent of patients for each and every test each and every time, leading some scholars to suggest that blanket consent for all diagnostic and/or screening tests should be sought (Elias & Annas 1994; Shuster 2007). Regardless of what decisions are made, the advent of DNA microarrays will complicate consent procedures for genetic testing.

Population wide data bases such as those available in Western Australia and the United Kingdom, now make data linkage possible. This data may be used in quality assurance of the health care system activities, from which patients will ultimately benefit. Use of this data without the patients' consent runs the risk of harm from disclosure of information. Patients should expect quality assurance activities to be ethically sound and benefits to justify any risks and burdens (Wade 2007).

Testing for genetic disorders, even at the moment, is complicated by the fact that genetic information is both private and personal and familial and non-individual. For example, if a patient learns that she is BRCA1 positive and, therefore, has an increased chance of developing breast cancer in her lifetime, this has consequences for other female members of her family and their potential risk of developing breast cancer, whether they wish to know that risk or not (Braunack-Mayer 2006). Thus, genetic testing on one individual who voluntarily consents always raises the possibility that there are other family members who have not given consent and who would not do so. The use of DNA microarrays in genetic screening and testing has the potential to make this situation even more complex, since it may allow the possibility of predictive testing for a large range of mutations and polymorphisms, some of which would have significant implications for the lives of family members who did not necessarily wish to learn information about their genetic risk.

#### Understanding

In the current climate of genetic testing, the provision of adequate genetic counselling is a considerable challenge for health care workers. There is concern that the results of tests using DNA microarrays, particularly screening tests, will be difficult to explain to patients, in large part because of the sheer volume and complexity of the information to be conveyed. The information is likely to be inconclusive, perhaps ambiguous and generally anxiety provoking. Geneticists are concerned that general physicians will not be "up to the task" of interpreting and providing this information, leading to fears that specialised genetic counselling services will be swamped by the demand. Based on current knowledge, the implications of outcomes from DNA microarrays for the health and well-being of patients can only be speculated on (Grody 2003).

#### *Insurance and employment discrimination*

Genetic testing, particularly for late-onset disorders, has raised concerns about the implications of test results for patients' access to health and life insurance and employment possibilities. Although there has been little evidence to date that insurance companies have discriminated against applicants on genetic grounds, concern has been expressed that this would change if DNA microarrays made large scale screening for common disorders feasible (Grody 2003). DNA microarrays are also likely to be able to identify subtle differences in metabolism within populations (for example, individuals who may be susceptible to occupational or environmental exposure to carcinogens), suggesting the possibility that DNA microarrays might be used to exclude certain employees from certain occupations, or to deny employment completely.

### *Prenatal genetic screening*

Prenatal genetic screening has always been complicated by the fact that one of the endpoints of the screening process is abortion. Although a discussion of the ethical arguments for and against abortion is beyond the scope of this *Emerging Technology Bulletin*, it is worth noting here that DNA microarrays will complicate even further the difficult decisions that need to be made by parents faced with the results of prenatal genetic screening tests. Some scholars have expressed concern about DNA microarray technology because they can envisage its use as a screening tool to ‘genetically audition’ fetuses. Shuster has suggested that: “*Parental genetic screening allows parents to take on the role of gene police, and to erect a roadblock at which they search and examine their children-to-be before birth. Microarray genetic screening technologies give prospective parents and their physicians a means to extract information that they believe can predict the health of a child. Anomalies that might have remained hidden during a lifetime will now be revealed and assessed* (Shuster 2007).”

### *Privacy and confidentiality*

The difficulties that attach to maintaining privacy and confidentiality of genetic information have already been alluded to above. Concerns about the confidentiality of genetic information are likely to increase as DNA microarrays make it easier to carry out large-scale analysis of population samples.

These will be issues for populations as well as for individuals. DNA microarrays have the potential to provide information about mutations that are specific to particular racial and ethnic groups. Guaranteeing the genetic privacy of an ethnic group is likely to be a particularly difficult task. Awareness of these issues has led some authors to suggest that ‘group consent’ mechanisms need to be used, in addition to individual consent mechanisms (2004). There are still questions, however, about the mechanisms available for small groups, or individuals within groups, to opt out.

## Sources of Information

The medical literature (Table 10) was searched utilising the search terms outlined in Table 9 to identify relevant studies and reviews, until June 2006. In addition, major international health assessment databases were searched.

Using the search terms outlined in Table 9 a total of 11,692 references were downloaded. When duplicates were excluded 8,213 references remained. To refine the search further, only those references with “microarray” in the title or abstract were retained, leaving a total of 3,038. These references were examined and the majority were potentially relevant. References only pertaining to neoplasms were searched for leaving a total of 2,796 references. These references were further sorted using “clinical” as a key word (449 references). The majority of references appeared to be applicable to be included in an *Emerging Technology Bulletin*, however most abstracts concluded with the proviso that *this technology may be of clinical use in the near future*. It was felt that the volume of literature was too large to be included in an *Emerging Technology Bulletin*, and that synthesis of studies would be difficult due to the large number of genes considered, even if the *Bulletin* was restricted to neoplasms, specifically breast and prostate cancer. The Evaluators contacted all known companies involved in the manufacture of DNA microarrays to ascertain which products, if any, had *current* clinical applications, either prognostic or diagnostic and which products may have clinical applications in a three-year time horizon.

**Table 9 Search terms utilised**

<b>Search terms</b>
<b>MeSH</b> Gene Expression Profiling, Oligonucleotide Array Sequence Analysis, Diagnosis, Prognosis, Neoplasms
<b>Text words</b> DNA AND micro-array*, DNA AND micro AND array*, molecular AND profil*, chip AND gene AND expression, DNA AND array*, DNA AND chip*, DNA AND microarray*
<b>Limits</b> English, Humans

**Table 10 Literature sources utilised in assessment**

Source	Location
<i>Electronic databases</i>	
AustHealth	University library
Australian Medical Index	University library
Australian Public Affairs Information Service (APAIS) - Health	University library
Cinahl	University library
Cochrane Library – including, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, the Cochrane Central Register of Controlled Trials (CENTRAL), the Health Technology Assessment Database, the NHS Economic Evaluation Database	University library
Current Contents	University library
Embase	Personal subscription
Pre-Medline and Medline	University library
ProceedingsFirst	University library
PsycInfo	University library
Web of Science – Science Citation Index Expanded	University library
<i>Internet</i>	
Australian Clinical Trials Registry	<a href="http://www.actr.org.au/default.aspx">http://www.actr.org.au/default.aspx</a>
Current Controlled Trials metaRegister	<a href="http://controlled-trials.com/">http://controlled-trials.com/</a>
Health Technology Assessment international	<a href="http://www.htai.org">http://www.htai.org</a>
International Network for Agencies for Health Technology Assessment	<a href="http://www.inahta.org/">http://www.inahta.org/</a>
Medicines and Healthcare products Regulatory Agency (UK).	<a href="http://www.medical-devices.gov.uk/">http://www.medical-devices.gov.uk/</a>
National Library of Medicine Health Services/Technology Assessment Text	<a href="http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat">http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat</a>
National Library of Medicine Locator Plus database	<a href="http://locatorplus.gov">http://locatorplus.gov</a>
New York Academy of Medicine Grey Literature Report	<a href="http://www.nyam.org/library/grey.shtml">http://www.nyam.org/library/grey.shtml</a>
Trip database	<a href="http://www.tripdatabase.com">http://www.tripdatabase.com</a>
U.K. National Research Register	<a href="http://www.update-software.com/National/">http://www.update-software.com/National/</a>
US Food and Drug Administration, Center for Devices and Radiological Health.	<a href="http://www.fda.gov/cdrh/databases.html">http://www.fda.gov/cdrh/databases.html</a>
Websites of Specialty Organisations	Dependent on technology topic area

*Internet sites of interest*

Bibliography on Microarray Data Analysis, Wentian Li of Feinstein Institute for Medical Research

<http://www.nslj-genetics.org/microarray/>

DNA Microarray, Wikipedia

[http://en.wikipedia.org/wiki/DNA\\_microarray](http://en.wikipedia.org/wiki/DNA_microarray)

Open directory project: search “Microarray”

[http://dmoz.org/Science/Biology/Biochemistry\\_and\\_Molecular\\_Biology/Gene\\_Expression/Products\\_and\\_Services/](http://dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Gene_Expression/Products_and_Services/)

Biometric Research Branch: an integrated package for the visualization and statistical analysis of DNA microarray gene expression data.

<http://linus.nci.nih.gov/BRB-ArrayTools.html>

Stanford Microarray Database

<http://smd-www.stanford.edu//>

National Office of Public Health Genomics, CDC, ACCE Model System for Collecting, Analyzing and Disseminating Information on Genetic Tests

<http://www.cdc.gov/genomics/gtesting/ACCE/fbr.htm>

National Human Genome Research Institute: Microarray Project

<http://research.nhgri.nih.gov/microarray/main.html>

MicroArray Quality Control (MAQC) Project (FDA)

<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm>

### *Clinical trials*

Several clinical trials, utilising DNA microarrays, are currently registered on either the Australian Clinical Trials Registry or the Current Controlled Trials metaRegister.

**1. Clinical trial: Microarray Evaluation of Normal Breast Tissue to Identify Genetic Markers of Risk for Estrogen Receptor Positive Breast Cancer**

Sydney Cancer Centre, Department Medical Oncology, Royal Prince Alfred Hospital, Sydney

[https://actr.org.au/registry/trial\\_review.aspx?ID=739](https://actr.org.au/registry/trial_review.aspx?ID=739)

Study began: February 2005, Currently recruiting

For Invasive Breast Cancer. A translational study involving analysis of malignant and benign breast tissue taken from women undergoing definitive surgery for invasive breast cancer.

Primary outcome:

To identify genetic markers of risk of breast cancer that distinguishes between risk of Oestrogen Receptor positive and Oestrogen Receptor negative disease.

Key secondary outcome/s: To validate the predictive value of the identified markers in a second group of patients.

ACTR Number: ACTRN012605000480684

**2. MINDACT (Microarray In Node-Negative Disease May Avoid Chemotherapy): A Prospective, Randomized Study Comparing the 70-Gene Signature With the Common Clinical-Pathological Criteria in Selecting Patients for Adjuvant Chemotherapy in Node-Negative Breast Cancer**

Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam

<http://clinicaltrials.gov/show/NCT00433589>

Study began: December 2006, Currently recruiting

Primary Outcome Measures:

Distant metastasis-free survival at 5 years, disease-free survival (DFS).

Secondary Outcome Measures:

Proportion of patients treated with chemotherapy based on clinical prognosis compared to 70-gene signature prognosis, overall survival at 5 years, DFS at 5 years, safety (early and late).

Primary objectives:

Compare a molecular profiling approach (70-gene signature) vs usual clinical assessment in assigning adequate risk categories (and the need to receive adjuvant chemotherapy or not) in women with node-negative breast cancer.

Compare the efficacy and long-term toxicities of docetaxel and capecitabine vs standard anthracycline-based chemotherapy regimens in these patients.

Determine the best endocrine treatment strategy (i.e., letrozole for 7 years vs sequential tamoxifen for 2 years followed by letrozole for 5 years) in these patients.

ClinicalTrials.gov identifier NCT00433589

**3. Trial for Microarray Analysis of Colon Cancer Outcome-A (MACCO-A)**

H. Lee Moffitt Cancer Center and Research Institute, Hoffmann-La Roche, Florida, United States

<http://controlled-trials.com/mrct/trial/124619/microarray>

Study began: October 2003, Currently recruiting

Primary objectives:

To develop a tumor tissue classifier that will predict response to XELOX + bevacizumab or XELIRI + bevacizumab

Secondary Outcome Measures:

To evaluate the overall survival of colorectal cancer (CRC) patients treated with XELOX + bevacizumab (Arm A) or XELIRI + bevacizumab (Arm B)

To assess the toxicity associated with Arms A and B

Colorectal cancer is the third largest cause of cancer mortality in the United States. The treatment of metastatic colorectal cancer is undergoing rapid improvement. Currently, there are two major chemotherapy regimens, which can both be combined with anti-angiogenesis treatment. These regimens are 5-FU + irinotecan and 5-FU + oxaliplatin. Each therapy has roughly similar rates of response, but it is unclear which specific therapy would benefit which patients. The advent of genome wide expression analysis provides a tool to analyse these differences. In the microarray analysis of colon cancer outcome trial, sponsored by the National Institutes of Health (NIH) and Moffitt Cancer Center, patients with newly diagnosed metastatic colon cancer are biopsied and samples are preserved in RNA later. Patients are then randomized to either one of two state of the art regimens: capecitabine + irinotecan + avastin (bevacizumab) or capecitabine + oxaliplatin + avastin. Response to chemotherapy, time to progression, and overall survival are end points of this trial. Once accrual of patients has been met, the investigators will compare genome wide expression patterns for each group.

ClinicalTrials.gov Identifier: NCT00127036

**4. Microarray Analysis in breast cancer to Taylor Adjuvant Drugs Or Regimens, a randomized phase III study**

Dr S.C. Linn Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Departments of Molecular Biology and Medical Oncology, Amsterdam, Netherlands.

<http://www.controlled-trials.com/ISRCTN61893718>

Study began: October 2004, Ongoing

Primary Outcome Measure:

To define gene expression profiles that can predict a disease-free survival advantage for either dose dense therapy, docetaxel, and/or 6 versus 4 courses of chemotherapy.

Secondary Outcome Measures:

Is TAC better than AC dd regarding disease free survival and overall survival?

Are 6 courses better than 4 regarding disease free survival and overall survival?

Objectives of optional side studies:

To determine whether the proteomic profile of patients, with primary breast cancer, relates to patient demographic characteristics, tumour stage, tumour biologic characteristics or tumour genetic (micro-array) profile.

To identify a proteomic pattern that positively or negatively predicts relapse according to the genetic profile of the primary tumour (micro-array analysis) in each treatment arm.

To identify a proteomic pattern in follow-up serum samples that can predict for relapse.

5. The implementation of microarrays in cancer diagnosis (microarray prognostics in breast cancer). Dr S.C. Linn Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Departments of Molecular Biology and Medical Oncology, Amsterdam, Netherlands.  
<http://www.controlled-trials.com/ISRCTN71917916>  
Study began: January 2004, completed December 2006

Primary Outcome Measure:

The amount of successfully performed diagnostic microarray tests as a proportion of the total number of accrued patients.

Secondary Outcome Measures:

To assess the proportion of a 'high' versus a 'low' risk profile in lymph node negative breast cancer patients.

To assess the concordance between the 70-gene microarray risk profile and the metastasis risk as assessed with current Dutch guidelines based on clinicopathological factors (such as age, pT, tumor grade, hormonal receptor-status).

6. Roche Diagnostics are currently testing the AmpliChip Leukaemia Test in large, international multi-centre clinical trials to classify and correlate outcomes in leukaemia patients (AHRQ 2006).

## Conclusions

---

This *Emerging Technology Bulletin* is aimed at providing a *non-systematic* overview of the rate of progress and development of DNA microarrays in Australia. It is neither a definitive nor a comprehensive assessment of the safety and effectiveness associated with DNA microarrays.

The human genome consists of approximately  $2.8 \times 10^9$  nucleotide base pairs which may encode up to 25,000 genes. However not all of these genes are “turned” on, with only a subset being expressed at any one time. Not only can gene expression be turned on and off in response to stimuli, the amount or volume of gene expression can either be up or down regulated.

DNA microarrays are a relatively new technology, first described in 1995. Small fragments of DNA, or oligonucleotides, are attached to a glass substrate by a variety of manufacturing processes. DNA microarrays may hold 100,000s of these elements. Complementary DNA from test and reference samples are labelled with a visualisation tag such as a fluorophore. Samples are hybridised or bound to the DNA microarray, and the intensity of the signal of the bound DNA is an indication of a given gene’s activity. Due to the large number of target elements able to be bound to a DNA microarray, the rapid, simultaneous characterisation of thousands of genes is possible. DNA microarrays may be useful tools in categorising diseases such as cancer by determining the presence or absence of particular genes, which may provide important biological, diagnostic and prognostic information.

“In-house” microarrays may be produced for \$200, however an initial outlay of approximately \$350,000 is required for basic equipment including a robotic spotter and laser scanner. Most products purchased from Affymetrix, the market leader in supplying commercial microarrays, require that they be processed using equipment also purchased from Affymetrix. Costs for a complete microarray processing system are approximately A\$335,000. Affymetrix gene expression arrays cost between A\$270 -550 depending on the complexity of the target. A custom made array can cost between A\$30,000 for 90 arrays targeting 1,700 genes and up to >A\$200,000 for 180 arrays containing millions of oligonucleotides. The cost per array increases as the number of oligonucleotides per array increases.

One of the areas in which the use of DNA microarrays has created great interest is in the detection of prognostic markers in breast cancer. Although early detection and treatment strategies have resulted in increased survival times, breast cancer is still the most common cause of cancer-related death in Australia among women. Current genetic tests, which aim to detect women at high-risk of breast cancer (eg BRCA1 and BRCA2 carriers), are applicable to only a small proportion of the population.

Lymph node status remains the best *prognostic* marker for survival, with 50 and 25 per cent of lymph node positive and negative women experiencing recurrence of disease, respectively. Node-negative women treated only with tamoxifen and *not* adjuvant chemotherapy have a 10-year recurrence rate of 15 per cent, indicating that 85 per cent of these women may be subjected to

unnecessary toxic chemotherapy. DNA microarrays have been developed to identify women with a poor prognosis, that is, those likely to be at high-risk of developing recurrence and as a consequence have reduced survival. This may lead to better targeted therapy.

There are three main breast cancer predictive assays in current use. Two utilise DNA microarrays: MammaPrint<sup>®</sup>, the only FDA approved predictive breast cancer assay, and the 76-gene “Rotterdam” signature panel. The Oncotype<sup>™</sup> Dx is in current use in the United States but utilises RT-PCR and therefore was not considered in detail in this *Emerging Technology Bulletin*. Both the MammaPrint<sup>®</sup> and the Oncotype<sup>™</sup> Dx assays can be accessed by Australian women at full cost to the patient (A\$3,600 and A\$4,000, respectively). All studies included for assessment for both of these assays were retrospective and to date no prospective, randomised clinical trials, assigning patients to chemotherapy regimes based on the results of gene expression assays, have been published. To address these concerns, the MINDACT trial, a prospective randomised trial, is currently underway and actively recruiting node-negative women. Results from this trial will not be available until at least five years from initial recruitment.

Predictive DNA microarray studies indicate that both the 70- and 76-gene panels are capable of stratifying breast cancer patients into groups that either have a low-risk (good prognosis) or high-risk (poor prognosis) for the development of distant metastases, which will impact on their overall survival rate. Reported hazard ratios indicate that lymph node-negative women in the high-risk category are 2-5 times more likely to experience disease recurrence, even when adjusted for clinical prognostic factors. In addition, from this stratification it can be seen that a larger proportion of women considered to be at low risk of recurrence survive at 10-years than those at high risk. However, there are no data available which describe the impact of the 70-gene panel on patient outcomes (avoidance of toxic chemotherapy, disease-free survival and overall survival) by the identification of high (those who will benefit from adjuvant chemotherapy) and low risk women (those unlikely to benefit from adjuvant chemotherapy). Only the studies conducted with the 76-gene panel reported on the sensitivity and specificity of using DNA microarrays to predict recurrence of breast cancer. Reported sensitivities were good, ranging from 90-97 per cent for the risk of developing distant metastases within five years. However, reported specificities were poor, ranging from 31 to 48 per cent. A high sensitivity value indicates that the arrays can correctly identify those at high-risk. As a consequence the remaining population is accurately defined as being at low-risk of distant metastases and may not benefit from treatment. Although the arrays successfully identify those patients at high-risk of recurrence, poor specificity indicates that a number of low-risk individuals are incorrectly categorised as high-risk and will receive treatment needlessly.

The other area that has currently generated great interest is the potential of DNA microarrays to predict drug metabolism. Cytochrome P450 enzymes, encoded by a group of over 100 CYP genes, catalyse the oxidation of over 80 drugs. Some of these genes are highly polymorphic and may be phenotypically expressed as resulting in either an increase or decrease in enzyme activity.

This in turn may affect the rate of drug metabolism and individuals may be classified as either an ultra-fast, extensive, intermediate or poor metaboliser of drugs. In 2004, the FDA approved the only DNA microarray (Affymetrix AmpliChip™ cytochrome P450 test) to investigate single nucleotide polymorphisms in two genes of particular interest: the CYP2D6 and CYP2C19 genes.

Three studies were included for assessment in this *Bulletin* examining the ability of the AmpliChip™ to categorise the ability of patients to metabolise drugs according to their genotype. Patients are assigned as either poor, intermediate, efficient and ultra-rapid metabolisers of drugs according to the combination of CYP alleles expressed. However, no studies have been published that report on improvements in patient outcomes of patients on selective serotonin reuptake inhibitors (SSRIs) who have been genotyped versus those of patients on SSRIs who have not been genotyped. AmpliChip™ appears to be an accurate means of genotyping patients into poor, intermediate, efficient and ultra-rapid metabolisers of drugs. High sensitivity (>99%) and specificity (>99%) values when the AmpliChip™ is compared to the gold standards of bi-directional DNA sequencing and allele-specific PCR indicate that the test can successfully identify individuals who are poor metabolisers. However, there is a lack of clinical data concerning whether or not genotyping patients has an impact on patient outcomes.

DNA microarrays are a powerful research tool for the identification of association signals, as demonstrated by the large study conducted by the Wellcome Trust Consortium, where single nucleotide polymorphisms were identified as being associated with some of the major public health diseases including coronary artery disease and diabetes. Other commercially available DNA microarrays (with European CE-marking *not* FDA approval) are available that can be used for human papillomavirus, blood typing, and respiratory disease. In addition, many laboratories utilise “in-house” microarrays, which can be custom made to address specific research questions.

In summary, the possibilities for using DNA microarrays are vast. There is little doubt that DNA microarrays will revolutionise the ability to assess changes in gene expression on a mass, high-throughput scale. How the knowledge gained from these gene expression assays is applied to give clinically meaningful outcomes remains to be seen. The evidence that DNA does more than promise a revolution in prevention, diagnosis and treatment, has not yet been produced.

## Glossary of Terms

---

**Allele:** Alternative form of a gene. One of the different forms of a gene that can exist at a single locus.

**Base pairs:** two nitrogenous (purine or pyrimidine) bases (adenine and thymine or guanine and cytosine) held together by weak hydrogen bonds. The number of base pairs is often used as a measure of length of a DNA segment, eg 500 bp.

**Biomarkers:** any measurable cellular, subcellular or humoural factor that demonstrates the presence of malignancy or malignant potential, or predicts tumour behaviour, prognosis or response to treatment (Hinestrosa et al 2007).

**Codon:** three nucleotide bases in length (RNA) or three base pairs of nucleotides (DNA) which code for a single amino acid. A sequence of three RNA or DNA nucleotides that specifies (codes for) either an amino acid or the termination of translation.

**Compound heterozygote:** The presence of *two different* mutant alleles at a particular gene locus, one on each chromosome of a pair.

**DNA:** deoxyribonucleic acid

**cDNA:** complementary DNA. DNA synthesised from a mature mRNA template in a reaction catalysed by the enzyme reverse transcriptase.

**DNase:** an enzyme that degrades DNA to nucleotides.

**Electrophoresis:** a technique for separating the components of a mixture of charged molecules (proteins, DNAs, or RNAs) in an electric field within a gel or other support.

**Genetic polymorphism:** The occurrence together in the same population of more than one allele or genetic marker at the same locus with the least frequent allele or marker occurring more frequently than can be accounted for by mutation alone.

**Heterozygote:** Possessing two different forms of a particular gene, one inherited from each parent.

**Homozygote:** Possessing two identical forms of a particular gene, one inherited from each parent.

**Hybridisation:** to anneal nucleic acid strands from different sources ie DNA to DNA, cRNA to DNA.

**Microarray:** sets of miniaturised chemical reaction areas that may also be used to test DNA fragments, antibodies, or proteins.

**Missense mutation:** a mutation that changes a codon for one amino acid into a codon for a different amino acid.

**Nucleotide:** subunit that polymerizes into nucleic acids (DNA or RNA), the structural units of DNA or RNA (A, T, G, C).

**Oligonucleotide:** a molecule usually composed of 25 or fewer nucleotides; used as a DNA synthesis primer.

**PCR:** polymerase chain reaction: a method for amplifying specific DNA segments which exploits certain features of DNA replication. Replication

requires a primer and specificity is determined by the sequence and size of the primer. The method amplifies specific DNA segments by cycles of template denaturation; primer addition; primer annealing and replication using thermostable DNA polymerase.

**Plasmid:** autonomously replicating extra chromosomal DNA molecule. An autonomous self-replicating genetic particle usually of circular double-stranded DNA.

**Polymorphism:** The occurrence in a population (or among populations) of several phenotypic forms associated with alleles of one gene or homologs of one chromosome.

**Primer:** A short sequence (of RNA or DNA) from which DNA replication can initiate.

**Restriction digest:** restriction endonuclease (enzyme that recognises a particular short DNA sequence) which cleaves DNA sample into smaller fragments.

**RNA:** Ribonucleic acid

**mRNA:** messenger RNA is a molecule of RNA encoding a chemical "blueprint" for a protein product. mRNA is transcribed from a DNA template, and carries coding information to the ribosomes, site of protein synthesis.

**cRNA:** Synthetic RNA produced by transcription from a specific DNA single stranded template.

**RT-PCR:** reverse transcriptase polymerase chain reaction: a variation of the PCR technique in which cDNA is made from RNA via reverse transcription. The cDNA is then amplified using standard PCR protocols.

**SNP:** single nucleotide polymorphism. DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered.

**Transcription:** the synthesis of RNA using a DNA template.

**Translation:** the process of protein synthesis whereby the primary structure of the protein is determined by the nucleotide sequence in mRNA.

(Lefers 2004; Wikipedia 2007b)

## Appendix A: Levels of Evidence

Designation of levels of evidence according to type of research question

Level	Intervention §	Diagnosis **	Prognosis	Aetiology †††	Screening
I *	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among consecutive patients with a defined clinical presentation ††	A prospective cohort study ***	A prospective cohort study	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among non-consecutive patients with a defined clinical presentation††	All or none §§§	All or none §§§	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparative study with concurrent controls: Non-randomised, experimental trial † Cohort study Case-control study Interrupted time series with a control group	A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	Analysis of prognostic factors amongst untreated control patients in a randomised controlled trial	A retrospective cohort study	A comparative study with concurrent controls: Non-randomised, experimental trial Cohort study Case-control study
III-3	A comparative study without concurrent controls: Historical control study Two or more single arm study † Interrupted time series without a parallel control group	Diagnostic case-control study ††	A retrospective cohort study	A case-control study	A comparative study without concurrent controls: Historical control study Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard) ††	Case series, or cohort study of patients at different stages of disease	A cross-sectional study	Case series

## **Tablenotes**

\* A systematic review will only be assigned a level of evidence as high as the studies it contains, excepting where those studies are of level II evidence.

§ Definitions of these study designs are provided on pages 7-8 *How to use the evidence: assessment and application of scientific evidence* (NHMRC 2000b).

† This also includes controlled before-and-after (pre-test/post-test) studies, as well as indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C).

‡ Comparing single arm studies ie. case series from two studies.

\*\* The dimensions of evidence apply only to studies of diagnostic accuracy. To assess the effectiveness of a diagnostic test there also needs to be a consideration of the impact of the test on patient management and health outcomes. See *MSAC (2004) Guidelines for the assessment of diagnostic technologies*. Available at: [www.msac.gov.au](http://www.msac.gov.au).

§§ The validity of the reference standard should be determined in the context of the disease under review. Criteria for determining the validity of the reference standard should be pre-specified. This can include the choice of the reference standard(s) and its timing in relation to the index test. The validity of the reference standard can be determined through quality appraisal of the study. See Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 2003, 3: 25.

†† Well-designed population based case-control studies (eg population based screening studies where test accuracy is assessed on all cases, with a random sample of controls) do capture a population with a representative spectrum of disease and thus fulfil the requirements for a valid assembly of patients. These types of studies should be considered as Level II evidence. However, in some cases the population assembled is not representative of the use of the test in practice. In diagnostic case-control studies a selected sample of patients already known to have the disease are compared with a separate group of normal/healthy people known to be free of the disease. In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are excluded, which can lead to exaggeration of both sensitivity and specificity. This is called spectrum bias because the spectrum of study participants will not be representative of patients seen in practice.

‡‡ Studies of diagnostic yield provide the yield of diseased patients, as determined by an index test, without confirmation of accuracy by a reference standard. These may be the only alternative when there is no reliable reference standard.

\*\*\* At study inception the cohort is either non-diseased or all at the same stage of the disease.

§§§ All or none of the people with the risk factor(s) experience the outcome. For example, no smallpox develops in the absence of the specific virus; and clear proof of the causal link has come from the disappearance of small pox after large-scale vaccination.

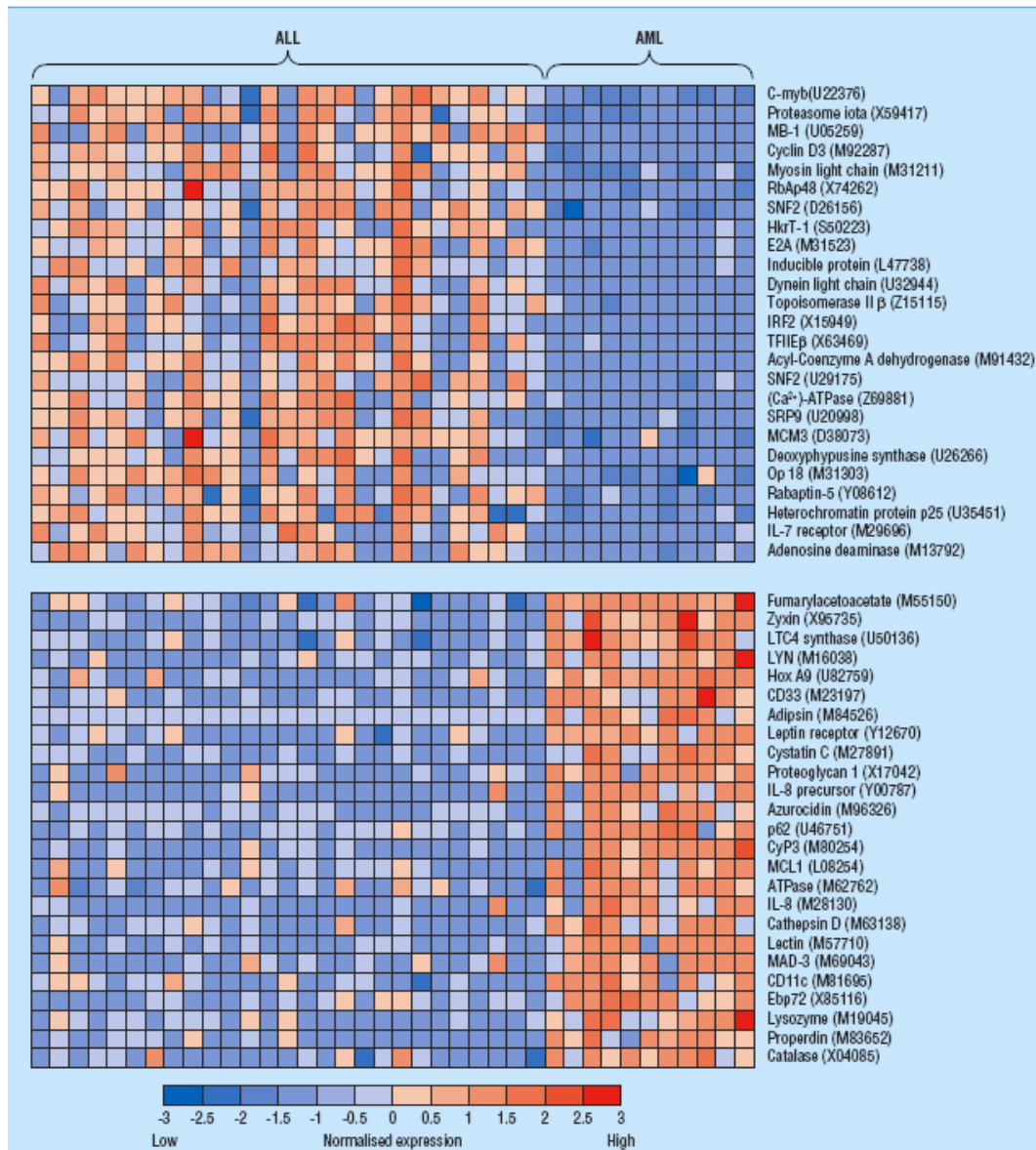
††† If it is possible and/or ethical to determine a causal relationship using experimental evidence, then the 'Intervention' hierarchy of evidence should be utilised. If it is only possible and/or ethical to determine a causal relationship using observational evidence (ie. cannot allocate groups to a potential harmful exposure, such as nuclear radiation), then the 'Aetiology' hierarchy of evidence should be utilised.

**Note 1:** Assessment of comparative harms/safety should occur according to the hierarchy presented for each of the research questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms are rare and cannot feasibly be captured within randomised controlled trials; physical harms and psychological harms may need to be addressed by different study designs; harms from diagnostic testing include the likelihood of false positive and false negative results; harms from screening include the likelihood of false alarm and false reassurance results.

**Note 2:** When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding research question eg. level II intervention evidence; level IV diagnostic evidence; level III-2 prognostic evidence etc.

Hierarchies adapted and modified from: Bandolier editorial 1999; Lijmer et al 1999; NHMRC 1999b; Phillips et al 2001

## Appendix B: An example of a DNA microarray



**Figure 9** Genes distinguishing acute lymphoid leukaemia (ALL) from acute myeloid leukaemia (AML). The figure shows the 50 genes with the greatest distinction between ALL and AML, the top panel showing genes more highly expressed in ALL and the bottom panel showing genes more highly expressed in AML. Each row corresponds to a gene, with the columns corresponding to expression levels of these genes in different patient samples. Expression levels greater than the mean are red, and those below the mean are blue. The scale indicates standard deviations above or below the mean. Adapted from (Golub et al 1999) (Aitman 2001)

## Appendix C: Microarray companies

---

The list of companies provided below is by no means exhaustive.

**Affymetrix:** Santa Clara, CA, USA <http://www.affymetrix.com/>  
Product names: see <http://www.affymetrix.com/products/index.html>

**Agendia B.V.:** Amsterdam, Netherlands <http://www.agendia.com/>  
Performs gene expression analysis (microarray) based diagnostics, founded in 2003 by the Netherlands Cancer institute.  
Product name: FDA approved MammaPrint<sup>®</sup> breast cancer prognostic test.

**Agilent Technologies :** Palo Alto, CA, USA <http://www.chem.agilent.com/>

**Amersham Pharmacia Biotech,** Piscataway, NJ, USA  
<http://www.apbiotech.com/>  
Product names: see <http://www.apbiotech.com/application/microarray/>

**Applied Maths :** Belgium <http://www.applied-maths.com/>  
Product name: GeneMaths. see <http://www.applied-maths.com/ge/ge.htm>

**BioDiscovery :** Los Angeles, CA, USA <http://www.biodiscovery.com/>  
Product names: <http://www.biodiscovery.com/products/products.html>  
Product name: ImaGene

**BioSieve:** San Jose, CA, USA <http://www.biosieve.com/>  
Product name: ExpressionSieve. see <http://www.biosieve.com/product.html>

**Clondiag :** Jena, Germany <http://www.clondiag.com/>  
Product name: IconoClust Image Software

**CombiMatrix Corp:** CA, USA <http://www.combimatrix.com/>  
Makes DNA microarrays using electrochemistry on arrays of microelectrodes. The arrays are built on semi-conductor chips. The CombiMatrix method controls the building of DNA on its semiconductor chips by the software-controlled turning on and off of electrodes. The process is intended to facilitate manufacture of custom DNA arrays, where users want the ability to do some experiments, change content, do some more experiments, change the content etc.

**Corning:** USA <http://www.corning.com/cmt/>  
Product name: CMT

**CytoGenomics :** St. Louis, MO, USA <http://www.silicocyte.com/>  
Product name: SilioCyte

**Eppendorf Biochip Systems:** Hamburg, Germany <http://www.eppendorf-biochip.com/>  
DNA microarrays for gene expression analysis, detection of infectious agents, GMOs in food and feed, and miRNA analysis.

**GeneData AG,** Basel, Switzerland <http://www.genedata.com/>  
Product name: GD Expressionist. see <http://www.genedata.com/products/expressionist/>

**Gene Logic :** Gaithersburg, MD, USA <http://www.genelogic.com>

**Gene Network Sciences ,** Ithaca, NY, USA <http://www.gnsbiotech.com/>

Product name: BioMine. see <http://www.gnsbiotech.com/biomine.shtml>

**GeneSifter** , Seattle, WA, USA <http://www.genesifter.net/>

Product name: GeneSifter

**Genetix**: Hampshire, United Kingdom <http://www.genetix.com/>

Product name: QArraymini, QArray2, QArraymax

**Genomic Solutions** , Ann Arbor, MI, USA

<http://www.genomicsolutions.com/>

Product name: GeneTAC

**Genotypic Technology** : Bangalore, India <http://www.genotypictech.com>

**Illumina** , San Diego, CA, USA <http://www.illumina.com/>

Product name: BeadArray

**Imaging Research Inc.**: St. Catharines, Ontario, Canada

<http://imaging.brocku.ca/>

Product names: ArrayVision, ArrayStat

**Incyte** , Palo Alto, CA, USA <http://www.incyte.com/>

Product name: LifeArray. see

<http://www.incyte.com/reagents/lifearray/index.shtml>

Product name: LifeExpression. see

<http://www.incyte.com/expression/index.shtml>

**LION bioscience AG** : Heidelberg, Germany <http://www.lionbioscience.com/>

**MDS Analytical Technologies**: PA, USA

<http://www.moleculardevices.com/home.html>

**Nanogen**: California, USA <http://www.nanogen.com/>

Product name: NanoChip<sup>®</sup> Electronic Microarray

**NuGEN Technologies**: San Carlos, CA, USA <http://www.nugeninc.com/>

**Oxford Gene Technology** : Oxford, UK <http://www.ogt.co.uk/>

**Packard BioChip**: a wholly-owned subsidiary of Packard BioScience

Company, Perkin-Elmer, USA <http://www.packardbiochip.com/>

Product name: ScanArray

**Partek**, St.Charles, MO, USA <http://www.partek.com/>

**Premier Biosoft International** : Palo Alto, CA, USA

<http://www.premierbiosoft.com/>

Product name: ArrayDesigner

**Progenika**: Vizcaya, Spain

[http://www.progenika.com/index.php?option=com\\_content&task=view&id=133&Itemid=167](http://www.progenika.com/index.php?option=com_content&task=view&id=133&Itemid=167)

Product names: LIPOchip<sup>®</sup>, PHARMAchip<sup>®</sup>, BLOODchip<sup>®</sup>, IBDchip<sup>®</sup>

**ResGen**, part of Invitrogen. Huntsville, AL, USA <http://www.resgen.com/>

Product name: VastArray

**Rosetta Inpharmatics:** Kirkland, WA, a wholly-owned subsidiary of Merck, USA <http://www.rii.com/>  
Product name: Resolver

**Scanalytics:** Fairfax, VA, USA <http://www.scanalytics.com/>

**Sequenom Industrial Genomics:** San Diego, CA <http://www.sequenom.com/>  
product names: MassArray

**Silicon Genetics,** Redwood City, CA <http://www.sigenetics.com/>  
Product name: GeneSpring

**SpotFire,** Somerville, MA, USA <http://www.spotfire.com>  
Product name: DecisionSite for functional genomics

**SuperArray Bioscience** , Frederick, MD, USA <http://www.superarray.com/>

**Tecan** : Maennedorf, Switzeland <http://www.tecan.com/>

**TeleChem International, Inc.:** Sunnyvale, CA, USA <http://www.arrayit.com/>  
Product name: ArrayIt

(Li 2007; Wikipedia 2007a)

## Appendix D: Australian Research

---

Australasian Microarray and Associated Technologies Association  
<http://www.amata.org.au/>

The University of Adelaide Microarray Facility  
<http://www.microarray.adelaide.edu.au/>

The University of Queensland, SRC Microarray Facility  
<http://microarray.imb.uq.edu.au/index.html>

Australian Genome Research Facility Ltd  
<http://www.agrf.org.au/Default.aspx?tabid=32>

CSIRO, Mathematical and Information Sciences, Image Analysis Group  
<http://experimental.act.cmis.csiro.au/Spot/index.php>

## Appendix E: HTA Internet Sites

---

### AUSTRALIA

- Centre for Clinical Effectiveness, Monash University  
<http://www.med.monash.edu.au/healthservices/cce/evidence/>
- Health Economics Unit, Monash University  
<http://chpe.buseco.monash.edu.au>

### AUSTRIA

- Institute of Technology Assessment / HTA unit  
<http://www.oecaw.ac.at/ita/welcome.htm>

### CANADA

- Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé (AETMIS) <http://www.aetmis.gouv.qc.ca/en/>
- Alberta Heritage Foundation for Medical Research (AHFMR)  
<http://www.ahfmr.ab.ca/publications.html>
- Canadian Coordinating Office for Health Technology Assessment (CCHOTA) <http://www.cadth.ca/index.php/en/>
- Canadian Health Economics Research Association (CHERA/ACRES) – Cabot database <http://www.mycabot.ca>
- Centre for Health Economics and Policy Analysis (CHEPA), McMaster University <http://www.chepa.org>
- Centre for Health Services and Policy Research (CHSPR), University of British Columbia <http://www.chspr.ubc.ca>
- Health Utilities Index (HUI)  
<http://www.fhs.mcmaster.ca/hug/index.htm>
- Institute for Clinical and Evaluative Studies (ICES)  
<http://www.ices.on.ca>

### DENMARK

- Danish Institute for Health Technology Assessment (DIHTA)  
[http://www.dihta.dk/publikationer/index\\_uk.asp](http://www.dihta.dk/publikationer/index_uk.asp)
- Danish Institute for Health Services Research (DSI)  
<http://www.dsi.dk/engelsk.html>

## **FINLAND**

- FINOHTA <http://www.stakes.fi/finohta/e/>

## **FRANCE**

- L'Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES)  
<http://www.anaes.fr/>

## **GERMANY**

- German Institute for Medical Documentation and Information (DIMDI)  
/ HTA <http://www.dimdi.de/dynamic/en/>

## **THE NETHERLANDS**

- Health Council of the Netherlands Gezondheidsraad  
<http://www.gr.nl/adviezen.php>

## **NEW ZEALAND**

- New Zealand Health Technology Assessment (NZHTA)  
<http://nzhta.chmeds.ac.nz/>

## **NORWAY**

- Norwegian Centre for Health Technology Assessment (SMM)  
<http://www.kunnskapssenteret.no/>

## **SPAIN**

- Agencia de Evaluación de Tecnologías Sanitarias, Instituto de Salud  
"Carlos III"/Health Technology Assessment Agency (AETS)  
<http://www.isciii.es/aets/>
- Catalan Agency for Health Technology Assessment (CAHTA)  
<http://www.aatrm.net/html/en/dir394/index.html>

## **SWEDEN**

- Swedish Council on Technology Assessment in Health Care (SBU)  
<http://www.sbu.se/www/index.asp>
- Center for Medical Health Technology Assessment  
<http://www.cmt.liu.se/>

## **SWITZERLAND**

- Swiss Network on Health Technology Assessment (SNHTA)  
<http://www.snhta.ch/>

## UNITED KINGDOM

- NHS Quality Improvement Scotland  
[http://www.nhshealthquality.org/nhsqis/qis\\_display\\_home.jsp?pContentID=43&p\\_applic=CCC&pElementID=140&pMenuID=140&p\\_service=Content.show&](http://www.nhshealthquality.org/nhsqis/qis_display_home.jsp?pContentID=43&p_applic=CCC&pElementID=140&pMenuID=140&p_service=Content.show&)
- National Health Service Health Technology Assessment (UK) / National Coordinating Centre for Health Technology Assessment (NCCHTA)  
<http://www.hta.nhsweb.nhs.uk/>
- University of York NHS Centre for Reviews and Dissemination (NHS CRD) <http://www.york.ac.uk/inst/crd/>
- National Institute for Clinical Excellence (NICE)  
<http://www.nice.org.uk/>

## UNITED STATES

- Agency for Healthcare Research and Quality (AHRQ)  
<http://www.ahrq.gov/clinic/techix.htm>
- Harvard School of Public Health – Cost-Utility Analysis Registry  
<http://www.tufts-nemc.org/cearegistry/index.html>
- U.S. Blue Cross/ Blue Shield Association Technology Evaluation Center (TEC) <http://www.bcbs.com/tec/index.html>

## References

---

- (2004). 'Integrating ethics and science in the International HapMap Project', *Nat Rev Genet*, 5 (6), 467-475.
- (2005a). 'AmpliChip CYP450 test', *Med Lett Drugs Ther*, 47 (1215-1216), 71-72.
- (2005b). 'A haplotype map of the human genome', *Nature*, 437 (7063), 1299-1320.
- (2007). 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, 447 (7145), 661-678.
- Abdullah-Sayani, A., Bueno-de-Mesquita, J. M. & van de Vijver, M. J. (2006). 'Technology Insight: tuning into the genetic orchestra using microarrays--limitations of DNA microarrays in clinical practice', *Nat Clin Pract Oncol*, 3 (9), 501-516.
- Access Excellence @ The National Health Museum (2007). *The Central Dogma of Molecular Biology* [Internet]. The National Health Museum. Available from: <http://www.accessexcellence.org/RC/VL/GG/central.html> [Accessed 16th July 2007].
- Affymetrix (2007a). *Array Manufacturing* [Internet]. Affymetrix Inc. Available from: <http://www.affymetrix.com/technology/manufacturing/index.affx> [Accessed 11th July 2007].
- Affymetrix (2007b). *Whole genome analysis* [Internet]. Affymetrix. Available from: [http://www.affymetrix.com/products/application/whole\\_genome.affx](http://www.affymetrix.com/products/application/whole_genome.affx) [Accessed 26th July 2007].
- Agendia *About MammaPrint* [Internet]. Agendia. Available from: <http://www.agendia.com/en/Professional/About-MammaPrint/About-MammaPrint> [Accessed 3rd July 2007].
- AHRQ (2006). *Genetic Tests for Cancer*, Agency for Healthcare Research and Quality, Maryland, United States, <http://www.ahrq.gov/clinic/ta/gentests/gentests.pdf>.
- AIHW (2004). *Cancer in Australia 2001*, Australian Institute of Health and Welfare, Canberra.
- AIHW (2006). *BreastScreen Australia monitoring report 2002-2003*, Australian Institute of Health and Welfare, Canberra, <http://www.aihw.gov.au/publications/can/bsamr02-03/bsamr02-03.pdf>.
- Aitman, T. J. (2001). 'DNA microarrays in medical practice', *BMJ*, 323 (7313), 611-615.
- Amaratunga, D. & Cabrera, J. (2004). *Exploration and analysis of DNA microarray and protein array data*. John Wiley & Sons Ltd, New Jersey, USA.
- Amersham Biosciences (2003). *CodeLink P450 SNP bioarrays* [Internet]. Amersham Biosciences. Available from: [http://www.apczech.cz/pdf/DF\\_Human\\_P450\\_SNP.pdf](http://www.apczech.cz/pdf/DF_Human_P450_SNP.pdf) [Accessed 23rd July 2007].

- Anjum, M. F., Mafura, M. et al (2007). 'Pathotyping Escherichia coli using miniaturised DNA-microarrays', *Appl Environ Microbiol*.
- Applied Biosystems (2007). *Real-time PCR vs traditional PCR* [Internet]. Applied Biosystems. Available from:  
[http://docs.appliedbiosystems.com/cms/groups/mcb\\_marketing/documents/generaldocuments/cms\\_042484.pdf](http://docs.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_042484.pdf) [Accessed 16th July 2007].
- Ballmer, K., Korczak, B. M. et al (2007). 'Fast DNA serotyping of Escherichia coli by use of an oligonucleotide microarray', *J Clin Microbiol*, 45 (2), 370-379.
- Bandolier editorial (1999). *Diagnostic testing emerging from the gloom?* [Internet]. Bandolier. Available from:  
<http://www.jr2.ox.ac.uk/bandolier/band70/b70-5.html> [Accessed 2004].
- Beiboer, S. H., Wieringa-Jelsma, T. et al (2005). 'Rapid genotyping of blood group antigens by multiplex polymerase chain reaction and DNA microarray hybridization', *Transfusion*, 45 (5), 667-679.
- Berrar, D. P., Dubitzky, W. & Granzow, M. (2003). *A practical approach to DNA microarray data analysis*. Kluwer Academic Publishers, Boston, USA.
- Blalock, E. M. (2003). *A beginner's guide to microarrays*. Kluwer Academic Publishers, Boston, USA.
- Blue Cross and Blue Shield Association (2005). *Gene expression profiling for managing breast cancer treatment*, Kaiser Permanente, Chicago, Illinois,  
[http://www.bcbs.com/betterknowledge/tec/vols/20/20\\_03.html](http://www.bcbs.com/betterknowledge/tec/vols/20/20_03.html).
- Braunack-Mayer, A. (2006). 'Ethics and health technology assessment: handmaiden and/or critic?' *International Journal of Technology Assessment in Health Care*, 22, 307-312.
- Buyse, M., Loi, S. et al (2006). 'Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer', *J Natl Cancer Inst*, 98 (17), 1183-1192.
- Campbell, M. A. (2003). *Introduction to DNA microarrays* [Internet]. Department of Biology, Davidson College. Available from:  
<http://www.bio.davidson.edu/people/macampbell/strategies/chipsintro.html> [Accessed 11th July 2007].
- Causton, H. C., Quackenbush, J. & Brazma, A. (2003). *Microarray gene expression data analysis: A beginner's guide*. Blackwell Science Ltd, Oxford, United Kingdom.
- Choi, Y. D., Jung, W. W. et al (2005). 'Detection of HPV genotypes in cervical lesions by the HPV DNA Chip and sequencing', *Gynecol Oncol*, 98 (3), 369-375.
- Chou, W. H., Yan, F. X. et al (2003). 'Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships', *Clin Chem*, 49 (4), 542-551.
- Cleator, S. & Ashworth, A. (2004). 'Molecular profiling of breast cancer: clinical implications', *Br J Cancer*, 90 (6), 1120-1124.

- Cobleigh, M. A., Tabesh, B. et al (2005). 'Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes', *Clin Cancer Res*, 11 (24 Pt 1), 8623-8631.
- Dave, S. S., Fu, K. et al (2006). 'Molecular diagnosis of Burkitt's lymphoma', *N Engl J Med*, 354 (23), 2431-2442.
- de Leon, J. (2006). 'AmpliChip CYP450 test: personalized medicine has arrived in psychiatry', *Expert Rev Mol Diagn*, 6 (3), 277-286.
- de Leon, J., Susce, M. T. et al (2005). 'The CYP2D6 poor metabolizer phenotype may be associated with risperidone adverse drug reactions and discontinuation', *J Clin Psychiatry*, 66 (1), 15-27.
- Desmedt, C., Piette, F. et al (2007). 'Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series', *Clin Cancer Res*, 13 (11), 3207-3214.
- Draghici, S., Khatri, P. et al (2006). 'Reliability and reproducibility issues in DNA microarray measurements', *Trends Genet*, 22 (2), 101-109.
- Elias, S. & Annas, G. J. (1994). 'Generic consent for genetic screening', *N Engl J Med*, 330 (22), 1611-1613.
- Eschrich, S. & Yeatman, T. J. (2004). 'DNA microarrays and data analysis: an overview', *Surgery*, 136 (3), 500-503.
- FDA (2004). *510 (k) substantial equivalence determination decision summary for Roche AmpliChip CYP450 microarray for identifying CYP2D6 genotype* [Internet]. United States Food and Drug Administration. Available from: <http://www.fda.gov/cdrh/reviews/k043576.pdf> [Accessed 20th July 2007].
- FDA (2005). *510 (k) substantial equivalence determination decision summary for Roche AmpliChip CYP450 microarray for identifying CYP2C19 genotype* [Internet]. United States Food and Drug Administration. Available from: <http://www.fda.gov/cdrh/reviews/k043576.pdf> [Accessed 20th July 2007].
- FDA (2007). *FDA Clears Breast Cancer Specific Molecular Prognostic Test* [Internet]. United States Food and Drug Administration. Available from: <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01555.html> [Accessed 3rd July 2007].
- Febbo, P. G. & Kantoff, P. W. (2006). 'Noise and bias in microarray analysis of tumor specimens', *J Clin Oncol*, 24 (23), 3719-3721.
- Foekens, J. A., Atkins, D. et al (2006). 'Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer', *J Clin Oncol*, 24 (11), 1665-1671.
- Genomica (2007a). *Clinical arrays HPV genotyping papillomavirus* [Internet]. Genomica. Available from: [http://genomica.artempus.net/media/docs/HPV%20Flyer\(%20english\)%20.pdf](http://genomica.artempus.net/media/docs/HPV%20Flyer(%20english)%20.pdf) [Accessed 24th July 2007].
- Genomica (2007b). *Clinical arrays MetaBone: Bone metabolic disorders* [Internet]. Genomica. Available from:

[http://genomica.artempus.net/media/docs/Tritico%20completo%20Clinical%20arrays%20MetaBone%20Ingles%20\(3\).pdf](http://genomica.artempus.net/media/docs/Tritico%20completo%20Clinical%20arrays%20MetaBone%20Ingles%20(3).pdf) [Accessed 24th July 2007].

Genomica (2007c). *Clinical arrays PneumoVir* [Internet]. Genomica. Available from:

<http://genomica.artempus.net/media/docs/GENOMICA%20TRIPTICO%20PNEUMO%20VIR.pdf> [Accessed 24th July 2007].

Glas, A. M., Floore, A. et al (2006). 'Converting a breast cancer microarray signature into a high-throughput diagnostic test', *BMC Genomics*, 7, 278.

Golub, T. R., Slonim, D. K. et al (1999). 'Molecular classification of cancer: class discovery and class prediction by gene expression monitoring', *Science*, 286 (5439), 531-537.

Grody, W. W. (2003). 'Ethical issues raised by genetic testing with oligonucleotide microarrays', *Mol Biotechnol*, 23 (2), 127-138.

Habel, L. A., Shak, S. et al (2006). 'A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients', *Breast Cancer Res*, 8 (3), R25.

Harris, N. L. & Horning, S. J. (2006). 'Burkitt's lymphoma--the message from microarrays', *N Engl J Med*, 354 (23), 2495-2498.

Hashmi, G., Shariff, T. et al (2005). 'A flexible array format for large-scale, rapid blood group DNA typing', *Transfusion*, 45 (5), 680-688.

Haviv, I. & Campbell, I. G. (2002). 'DNA microarrays for assessing ovarian cancer gene expression', *Mol Cell Endocrinol*, 191 (1), 121-126.

Hendolin, P. (2005). *The DrugMEt microarray genotyping test* [Internet]. JuriLab Ltd. Available from:

[http://www.uudenmaanosaamiskeskus.fi/fmnt/kuopio/Luennot/Hendolin-041011\\_MST.pdf](http://www.uudenmaanosaamiskeskus.fi/fmnt/kuopio/Luennot/Hendolin-041011_MST.pdf) [Accessed 23rd July 2007].

Hickie, I. B. (2004). 'Reducing the burden of depression: are we making progress in Australia?' *Med J Aust*, 181 (7 Suppl), S4-5.

Hinestroza, M. C., Dickersin, K. et al (2007). 'Shaping the future of biomarker research in breast cancer to ensure clinical relevance', *Nat Rev Cancer*, 7 (4), 309-315.

Hornberger, J., Cosler, L. E. & Lyman, G. H. (2005). 'Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer', *Am J Manag Care*, 11 (5), 313-324.

Huang, H. J., Huang, S. L. et al (2004). 'Human papillomavirus genotyping by a polymerase chain reaction-based genechip method in cervical carcinoma treated with neoadjuvant chemotherapy plus radical surgery', *Int J Gynecol Cancer*, 14 (4), 639-649.

Huang, S. L., Chao, A. et al (2006). 'Comparison between the Hybrid Capture II Test and an SPF1/GP6+ PCR-based assay for detection of human papillomavirus DNA in cervical swab samples', *J Clin Microbiol*, 44 (5), 1733-1739.

- Human Genome Project (2004). *Ethical, Legal, and Social Issues Research* [Internet]. United States Department of Energy Office of Science. Available from: [http://www.ornl.gov/sci/techresources/Human\\_Genome/research/elsi.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/research/elsi.shtml) [Accessed 2nd August 2007].
- Hummel, M., Bentink, S. et al (2006). 'A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling', *N Engl J Med*, 354 (23), 2419-2430.
- Jain, K. K. (2005). 'Applications of AmpliChip CYP450', *Mol Diagn*, 9 (3), 119-127.
- Kaklamani, V. (2006). 'A genetic signature can predict prognosis and response to therapy in breast cancer: Oncotype DX', *Expert Rev Mol Diagn*, 6 (6), 803-809.
- Klaassen, C. H., Prinsen, C. F. et al (2004). 'DNA microarray format for detection and subtyping of human papillomavirus', *J Clin Microbiol*, 42 (5), 2152-2160.
- Krutovskii, K. V. & Neale, D. B. (2001). *Forest Genomics for Conserving Adaptive Genetic Diversity* [Internet]. Food and Agriculture Organization of the United Nations. Available from: <http://www.fao.org/DOCREP/003/X6884E/x6884e03.htm#TopOfPage> [Accessed 10th July 2007].
- Lai, C. H., Huang, H. J. et al (2007). 'Human papillomavirus genotype in cervical cancer: a population-based study', *Int J Cancer*, 120 (9), 1999-2006.
- Lefers, M. (2004). *Life Science Glossary* [Internet]. Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Illinois, USA. Available from: <http://www.biochem.northwestern.edu/holmgren/Glossary/Definitions.html> [Accessed 17th July 2007].
- Li, W. (2007). *Bibliography on Microarray Data Analysis* [Internet]. Feinstein Institute for Medical Research. Available from: <http://www.nslj-genetics.org/microarray/company.html> [Accessed 2nd July 2007].
- Lijmer, J. G., Mol, B. W. et al (1999). 'Empirical evidence of design-related bias in studies of diagnostic tests.' *Journal of the American Medical Association*, 282 (11), 1061 - 1066.
- Lin, C. Y., Chen, H. C. et al (2007). 'Quality assurance of genotyping array for detection and typing of human papillomavirus', *J Virol Methods*, 140 (1-2), 1-9.
- Mant, A., Rendle, V. A. et al (2004). 'Making new choices about antidepressants in Australia: the long view 1975-2002', *Med J Aust*, 181 (7 Suppl), S21-24.
- Marshall, E. (2004). 'Getting the noise out of gene arrays', *Science*, 306 (5696), 630-631.
- Matcher, D. B., Thakur, M. E. et al (2006). *Testing for Cytochrome P450 polymorphisms in adults with non-psychotic depression treated with selective serotonin reuptake inhibitors (SSRIs)*, Agency for Healthcare Research and Quality (AHRQ), Rockville, Maryland,

- Mina, L., Soule, S. E. et al (2007). 'Predicting response to primary chemotherapy: gene expression profiling of paraffin-embedded core biopsy tissue', *Breast Cancer Res Treat*, 103 (2), 197-208.
- Modlich, O., Prisack, H. B. & Bojar, H. (2006). 'Breast cancer expression profiling: the impact of microarray testing on clinical decision making', *Expert Opin Pharmacother*, 7 (15), 2069-2078.
- Monecke, S., Berger-Bachi, B. et al (2007). 'Comparative genomics and DNA array-based genotyping of pandemic *Staphylococcus aureus* strains encoding Panton-Valentine leukocidin', *Clin Microbiol Infect*, 13 (3), 236-249.
- Montpetit, A., Phillips, M. S. et al (2006). 'High-throughput molecular profiling of blood donors for minor red blood cell and platelet antigens', *Transfusion*, 46 (5), 841-848.
- Murphy, D. (2002). 'Gene expression studies using microarrays: principles, problems, and prospects', *Adv Physiol Educ*, 26 (1-4), 256-270.
- NBCC (2006). *Advice about familial aspects of breast and epithelial ovarian cancer: a guide for health professionals*, National Breast Cancer Centre, Sydney, [http://www.nbcc.org.au/bestpractice/resources/BOG182\\_adviceaboutfamiliala.pdf](http://www.nbcc.org.au/bestpractice/resources/BOG182_adviceaboutfamiliala.pdf)
- NCBI (2004). *SNPs: variations on a theme* [Internet]. National Center for Biotechnology Information. Available from: <http://www.ncbi.nlm.nih.gov/About/primer/snps.html> [Accessed 26th July 2007].
- New Zealand Health Information Service (2006). *Cancer: New Registrations and Deaths 2002*, New Zealand Ministry of Health, Wellington, New Zealand, <http://www.nzhis.govt.nz/publications/cancer02.pdf>.
- NHMRC (1999a). *Familial aspects of cancer: A guide to clinical practice*, National Health and Medical Research Council, Commonwealth of Australia, Canberra, ACT,
- NHMRC (1999b). *National Statement on Ethical Conduct in Research Involving Humans*, NHMRC, Canberra,
- Nogues, N. (2007). 'Microarray technology and its application to blood grouping', Conference Paper: International Society of Blood Transfusion, Madrid, Spain. <http://www.parthen-impact.com/eventure/publicAbstractView.do?id=41657>.
- Ntzani, E. E. & Ioannidis, J. P. (2003). 'Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment', *Lancet*, 362 (9394), 1439-1444.
- Oh, T. J., Kim, C. J. et al (2004). 'Development and clinical evaluation of a highly sensitive DNA microarray for detection and genotyping of human papillomaviruses', *J Clin Microbiol*, 42 (7), 3272-3280.
- Paik, S., Shak, S. et al (2004). 'A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer', *N Engl J Med*, 351 (27), 2817-2826.
- Palylyk-Colwell, E. (2006). *CYP450 genotyping for determining drug metabolizer status* [Internet]. Canadian Coordinating Office for Health Technology Assessment. Available from:

[http://www.cadth.ca/media/pdf/375\\_armplichip\\_cetap\\_e.pdf](http://www.cadth.ca/media/pdf/375_armplichip_cetap_e.pdf) [Accessed 20th July 2007].

Perreten, V., Vorlet-Fawer, L. et al (2005). 'Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria', *J Clin Microbiol*, 43 (5), 2291-2302.

Phillips, B., Ball, C. et al (2001). *Levels of Evidence and Grades of Recommendations* [Internet]. Centre for Evidence-Based Medicine, Oxford, UK. Available from: Available from: [http://www.cebm.net/levels\\_of\\_evidence.asp](http://www.cebm.net/levels_of_evidence.asp) [Accessed 28th January 2004].

Prinsen, C. F., Fles, R. et al (2007). 'Baseline human papillomavirus status of women with abnormal smears in cervical screening: a 5-year follow-up study in The Netherlands', *Bjog*, 114 (8), 951-957.

Progenika Biopharma, S. A. (2007). *Progenika: Pioneering personalized medicine* [Internet]. Progenika Biopharma, S.A. Available from: <http://www.progenika.com/index.php> [Accessed 24th July 2007].

Roche Diagnostics (2006a). *AmpliChip CYP450 Test* [Internet]. F.Hoffmann-La Roche Ltd. Available from: [http://www.roche.com/prod\\_diag\\_amplichip.htm](http://www.roche.com/prod_diag_amplichip.htm) [Accessed 20th July 2007].

Roche Diagnostics (2006b). *The CYP450 gene family and drug metabolism* [Internet]. F.Hoffmann-La Roche Ltd. Available from: [http://www.roche.com/final\\_cyp\\_gene\\_family.pdf](http://www.roche.com/final_cyp_gene_family.pdf) [Accessed 20th July 2007].

Sachse, K., Hotzel, H. et al (2006). 'The use of DNA microarray technology for detection and genetic characterisation of chlamydiae', *Dev Biol (Basel)*, 126, 203-210; discussion 326-207.

Schena, M. (2003). *Microarray analysis*. John Wiley & Sons Inc, New Jersey.

Schena, M., Heller, R. A. et al (1998). 'Microarrays: biotechnology's discovery platform for functional genomics', *Trends Biotechnol*, 16 (7), 301-306.

Schena, M., Shalon, D. et al (1995). 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science*, 270 (5235), 467-470.

Shuster, E. (2007). 'Microarray genetic screening: a prenatal roadblock for life?' *Lancet*, 369 (9560), 526-529.

Steinhoff, C. & Vingron, M. (2006). 'Normalization and quantification of differential expression in gene expression microarrays', *Brief Bioinform*, 7 (2), 166-177.

Tan, P. K., Downey, T. J. et al (2003). 'Evaluation of gene expression measurements from commercial microarray platforms', *Nucleic Acids Res*, 31 (19), 5676-5684.

Tejedor, D., Castillo, S. et al (2005). 'Reliable low-density DNA array based on allele-specific probes for detection of 118 mutations causing familial hypercholesterolemia', *Clin Chem*, 51 (7), 1137-1144.

van 't Veer, L. J., Dai, H. et al (2002). 'Gene expression profiling predicts clinical outcome of breast cancer', *Nature*, 415 (6871), 530-536.

- van de Vijver, M. J., He, Y. D. et al (2002). 'A gene-expression signature as a predictor of survival in breast cancer', *N Engl J Med*, 347 (25), 1999-2009.
- Venturoli, S., Cricca, M. et al (2002). 'Human papillomavirus DNA testing by PCR-ELISA and hybrid capture II from a single cytological specimen: concordance and correlation with cytological results', *J Clin Virol*, 25 (2), 177-185.
- Wade, D. (2007). 'Ethics of collecting and using healthcare data', *BMJ*, 334 (7608), 1330-1331.
- Wang, Y., Klijn, J. G. et al (2005). 'Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer', *Lancet*, 365 (9460), 671-679.
- Watson, J. D., Gilman, M. et al (1992). *Recombinant DNA*. Second Scientific American Books, New York, USA.
- Wikipedia (2007a). *DNA microarray* [Internet]. Wikipedia. Available from: [http://en.wikipedia.org/wiki/DNA\\_microarray](http://en.wikipedia.org/wiki/DNA_microarray) [Accessed 2nd July 2007].
- Wikipedia (2007b). *Home page* [Internet]. Wikipedia. Available from: [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page) [Accessed 16th July 2007].
- Wikipedia (2007c). *Northern blot* [Internet]. Wikipedia. Available from: [http://en.wikipedia.org/wiki/Northern\\_blot](http://en.wikipedia.org/wiki/Northern_blot) [Accessed 16th July 2007].
- Wikipedia (2007d). *Real time PCR* [Internet]. Wikipedia. Available from: [http://en.wikipedia.org/wiki/Real\\_time\\_pcr](http://en.wikipedia.org/wiki/Real_time_pcr) [Accessed 16th July 2007].
- Wikipedia (2007e). *Selective serotonin reuptake inhibitor* [Internet]. Wikipedia. Available from: <http://en.wikipedia.org/wiki/SSRI> [Accessed 20th July 2007].