

# Biochip platforms as functional genomics tools for drug discovery

Ivan Wick<sup>1</sup> & Gary Hardiman<sup>1,2\*</sup>

## Addresses

<sup>1</sup>Biomedical Genomics Microarray Facility (BIOGEM)

<sup>2</sup>Department of Medicine

University of California San Diego

La Jolla

CA 92093-0724

USA

Email: ghardiman@ucsd.edu

\*To whom correspondence should be addressed

**Current Opinion in Drug Discovery & Development** 2005 8(3):

© The Thomson Corporation ISSN 1367-6733

*Improvements in DNA microarray technology have resulted in the generation of data on a scale that, for the first time, permits detailed scrutiny of the human genome. These data provide the foundation for understanding not only the connections between genes and the purpose of genes in the human genome, but also the molecular basis of genetic defects. These advances have the potential to significantly improve healthcare management by improving disease diagnosis and specifically targeting molecular therapy. Herein, the current state of the technology is reviewed, the commercial platforms used by the biopharmaceutical industry are compared and contrasted, and recent efforts in cross-platform data integration are explored.*

**Keywords** Biochips, data integration, gene expression, microarrays, platform

## Introduction

Currently, the pharmaceutical industry is facing an upsurge in research and development costs, while the rate at which novel molecular entities are reaching the market is significantly slower than in the past. This has forced the industry to devise and adapt methodologies that have the potential to increase the number of new drug candidates in pipelines, within a much shorter time frame [1]. During the past decade, high-density DNA microarrays and biochips have revolutionized the field of biomedicine and helped to accelerate target validation and drug discovery efforts [2]. Microarrays are still predominantly used for gene expression analyses, but they are also finding application in genotyping and re-sequencing tools, and in comparative genomic hybridization studies. They have been utilized to address *in vitro* pharmacology and toxicology issues, and are being widely used to improve the processes of disease diagnosis, pharmacogenomics and toxicogenomics [3-6].

The power of microarray technology lies in its ability to perform a vast number of parallel gene expression profiles from a single sample. A complete view of gene expression within a sample provides a snapshot of the transcriptome in healthy and diseased states. This information is highly useful as it uncovers gene families (or more specifically pathways) that are affected, but also reveals those that are unaffected [7]. Hypotheses can also be formed regarding genes with unknown function, by comparison of their expression levels with genes of known function [8], for example, similar expression profiles would imply that genes may be co-regulated.

Microarray experimentation is a complex process, and significant time and effort are required to design biologically sound and statistically robust experiments. Once target genes are identified, additional time and expense are required to validate their selection and relevance. Drug discovery programs utilizing microarray technologies must therefore consider all available technologies before allocating precious resources. Several complementary microarray technologies for measuring gene expression have evolved. Platform evaluations are, however, impractical for the majority of researchers, as this involves considerable expenditure and often a commitment to dedicated hardware and software [9,10••].

## From cancer classification to drug discovery

The seminal publication that recognized the power of gene expression and microarray technologies, demonstrated accurate classification of the hematological malignancies acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL) [11••]. This study demonstrated that molecular signatures could be used to clearly classify patients at the clinic. Furthermore, it revealed the utility of microarrays in several areas, such as tumor classification, prediction of tumor classes and molecular diagnostics, and also in elucidating the genetic defects that lead to cancers. Subsequently, gene expression in many solid tumors was studied using DNA microarrays. In the case of breast and lung cancers, the focus switched from tumor classification [12-14], toward dissecting solid tumors in the context of patient survival [15] or defining tumors by metastasis signatures [16,17]. These cancer studies uncovered a method of gene expression class-based molecular classification of cancers. Thus, it was suggested that transcriptional profiling could aid in the discovery of new tumor classes and pathway defects, in patient stratification for treatment and in the discovery of new drug candidates.

## Microarray expression platforms

Many competing technologies have gained acceptance within the pharmaceutical industry, including full-length cDNA arrays, and pre- or *in situ*-synthesized oligonucleotide probes [17,18]. The standard experimental paradigm compares messenger RNA (mRNA) abundance in two different biological samples, on the same or replicate microarrays. Each of the respective platforms has been optimized to function with either a single- or dual-color detection system. Recent key trends have been to move away from cDNA-based microarrays toward oligonucleotide-based microarrays, and from in-house or 'home-brew' systems to higher quality, commercial platforms. The salient features of the platforms predominantly utilized by the pharmaceutical industry are presented below. Advances in laboratory automation have improved the sensitivity, specificity and reproducibility of microarray experimentation, including advances in automated hybridization, sample formulation and system preparation.

### **Affymetrix - GeneChip**

Affymetrix Inc pioneered the field of microarrays and has dominated this area for many years, applying photolithographic technologies derived from the semiconductor industry to the fabrication of high-density microarrays. GeneChip arrays have become the pharmaceutical industry standard owing to their extensive genetic content, high levels of reproducibility and minimal start-up time. GeneChip arrays consist of short single-stranded DNA segments, oligonucleotides or oligos, which are built to order by chemical synthesis [17]. A major advantage of the GeneChip technology is that the arrays are designed *in silico*, thereby eliminating the need for management of DNA clone libraries and the possibility of misidentified tubes, clones or features [19••]. The disadvantage of this platform is that it demands a dedicated scanner and utilizes short 25mer oligonucleotides, which are less sensitive than the longer 60mers utilized in other technologies. Additionally, multiple oligonucleotides are required for transcript detection.

### **Agilent Technologies – cDNA and oligonucleotide-based arrays**

Alternative platforms are emerging that are challenging the dominance of Affymetrix. One of these, a two-color system from Agilent Technologies Inc, relies on the *in situ* synthesis of probes by inkjet printing using phosphoramidite chemistry. Inkjet technology has previously been utilized by Agilent Technologies to fabricate spotted cDNA arrays from polymerase-chain reaction (PCR) amplicons. This cDNA-based platform, however, has largely been retired in favor of the superior oligonucleotide format, which utilizes 60mer probes, in contrast with the short 25mer probes employed by Affymetrix. Although short oligonucleotides should in theory provide the greatest discrimination between related sequences, they often have poor hybridization properties. The 60mers provide enhancements in sensitivity over 25mers, due in part to the larger area available for hybridization. Another advantage is that only one 60mer per gene or transcript is required [20].

### **Amersham Biosciences - CodeLink Bioarray platform**

Another platform widely used by the pharmaceutical industry is the CodeLink Bioarray platform from Amersham Biosciences Corp (now part of GE Healthcare Ltd). For this platform short 30mer oligonucleotides are synthesized *ex situ* using standard phosphoramidite chemistry. This has the advantage that the probes can be validated by mass spectrometry prior to non-contact piezoelectric deposition on a proprietary three-dimensional (3D) polyacrylamide gel matrix. Covalent attachment is achieved via covalent interactions between 5'-amine groups on the oligonucleotide probes and functional groups on the slide surface. Advantages with this platform are that the 3D nature of the slide surface promotes an aqueous biological environment and solution-phase kinetics, providing enhanced assay sensitivity. CodeLink is an open system that can be utilized with any microarray scanner. The disadvantages are that only one oligonucleotide probe, a 30mer, is used to interrogate a particular gene, which is potentially less sensitive than a 60mer [21].

### **Applied Biosystems - Expression array system**

A recent addition to the microarray field is the Expression Array System from Applied Biosystems. Standard phosphoramidite chemistry is employed to synthesize 60mer oligonucleotides, which, similarly to the CodeLink platform are validated offline by mass spectrometry, and are then deposited onto a modified nylon microarray substrate. The 3'-end of the oligonucleotide is covalently coupled to the nylon via a carbon spacer, which raises the oligonucleotide and helps avoid steric hindrance. The use of chemoluminescence detection distinguishes this system from other commercial platforms and provides increased detection sensitivity, since, unlike with fluorescence detection, an excitation step is not required, thereby minimizing background noise. The disadvantages of this platform are that it is not amenable to customization and requires a dedicated chemoluminescence reader.

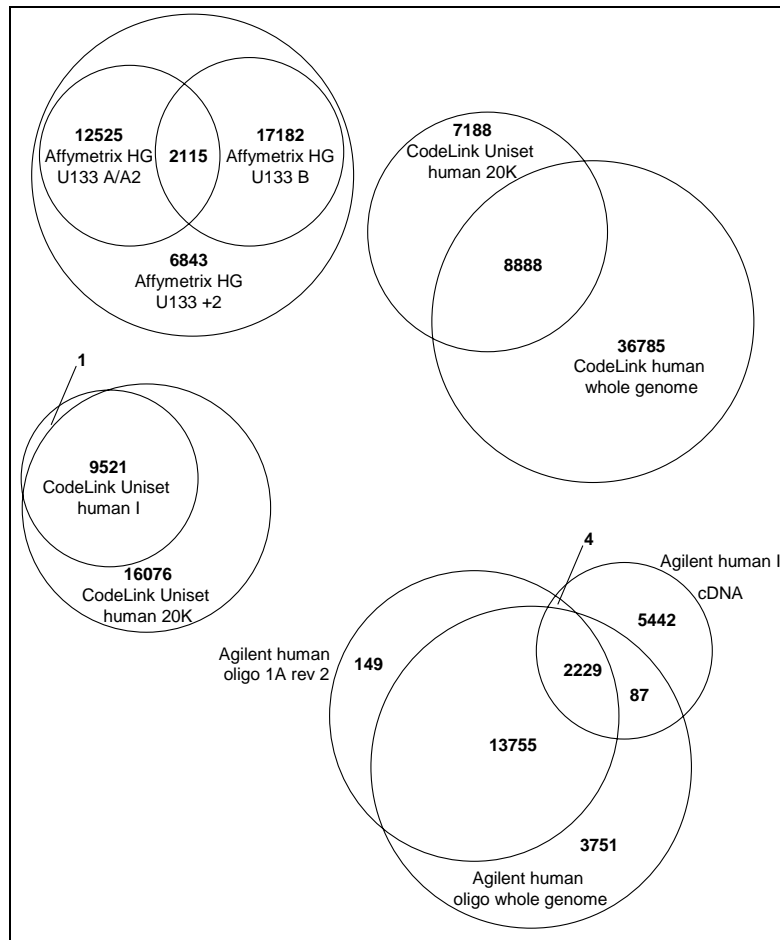
### **Illumina - Sentrix Beadchip and Array Matrix**

Illumina Inc has developed a bead-based technology for single-nucleotide polymorphism (SNP) genotyping and gene expression profiling applications on two distinct substrates, the Sentrix LD BeadChip and the Sentrix Array Matrix (which facilitate genotyping of up to 8 and 96 samples, respectively). Both are fabricated to utilize an 'array of arrays' format, which enables the simultaneous processing of multiple samples. Each array on each substrate contains thousands of tiny etched wells, into which thousands to hundreds of thousands of 3- $\mu$ m beads self-assemble in a random fashion. Gene-specific probes (50mer) concatenated with 'address or zip-code' sequences are immobilized on the bead surface. After bead assembly, each array is 'decoded' via a proprietary process, to determine which bead type containing which sequence is present in each well of the substrate. The advantages of this platform are its sensitivity and reproducibility. The oligonucleotide probes can be validated offline, and the low-density bead chip format can be utilized with any microarray scanner capable of scanning at 5- $\mu$ m resolution. The Sentrix Array Matrix provides major increases in throughput for the pharmaceutical industry, although this high-density format requires a dedicated scanner.

### **Which platform is superior?**

Each technology possesses inherent advantages and disadvantages, and currently no single platform offers overall superiority. In summary, longer probes provide greater sensitivity at the expense of reduced specificity, while 25mer probes require the use of multiple oligonucleotide probes per transcript. For example, as previously mentioned, the optimal feature of the Amersham Biosciences CodeLink system is its excellent sensitivity owing to the 3D nature of the surface. This platform also has the advantage of being open, and can be used with a range of microarray scanners. However, this platform is fabricated via deposition rather than *in situ* synthesis, so certain features may exhibit poor spot morphology and contamination artifacts. The Affymetrix system is favored by the fact that it is a mature platform with an extensive array catalog, which has been widely used by the pharmaceutical industry. There is, however, a requisite commitment to dedicated hardware. The Agilent Technologies platform is

**Figure 1. Illustration of the evolution and overlap of catalog human arrays from Affymetrix, Amersham Biosciences and Agilent Technologies.**



highly reproducible and the most sensitive of the various array platforms. Furthermore, considerable cost savings are realized with this platform as it comprises a two-color system. The two-color approach has potential disadvantages, however, since different fluorescently labeled nucleotides may be incorporated with different frequencies, altering ratios due to enzymatic parameters rather than actual transcript abundance. Additionally, multiple-experiment comparisons are not possible without replicating the reference sample (which, in the case of biopsy material, may be difficult to obtain).

The evolution and overlap of catalog human arrays from Affymetrix, Amersham Biosciences and Agilent, respectively, are highlighted in Figure 1. With improvements in the technology and in the availability of annotated human genome sequences, each of the major providers has released higher density arrays with smaller feature sizes. UniGene identifiers are the most complete and common identifiers among probes on every array system, and this information was used to determine the genes most commonly represented among all the arrays.

### Expression platforms and data variability

Several DNA chip technology formats have evolved, and carefully designed studies have been performed to evaluate

the interchangeability of data from various platforms. The ability to combine expression measurements from different technologies within a single analysis would result in considerable cost savings and reduce the need for duplicate experiments in separate laboratories. From the early comparisons that were conducted, the outlook appeared bleak for such cross-platform data integration. Discordance was observed, which was attributed to the disparity inherent in each of the respective platforms. Differences arise from the intrinsic properties of the biochips themselves, and also from the various processing and analytical steps involved in each system. This led to a number of questions: is one platform generating incorrect data? Do different biochips accurately reflect true biological expression? If differences are reported, then what is the true biological expression profile?

### ***The Affymetrix system versus cDNA microarrays***

Variability in measured gene expression levels associated with different platforms, particularly cDNA-based microarrays, has hindered integrative efforts. Kothapalli *et al* were the first to report inconsistencies in cross-platform data [22••]. Gene expression levels in peripheral blood mononuclear cells (PBMCs) from a patient with large granular lymphocyte leukemia, and also from a healthy control, were determined using Affymetrix oligonucleotide

GeneChips and UniGEM V spotted cDNA microarrays from Incyte Corp. This study highlighted the problems often encountered with cDNA microarray platforms, namely inconsistent sequence fidelity of the spotted microarrays, variability in differential expression levels, low specificity of cDNA microarray probes, discrepancies in the fold-change calculation compared with the Affymetrix system, and lack of probe specificity for different isoforms of a gene. Considerable variation exist in the cDNA libraries used to generate spotted cDNA microarrays and errors in handling bacterial clones and cross contamination have been well documented [19••].

Subsequent studies revealed similar discrepancies in cDNA-based platforms [23••,24•,25•]. Li *et al* employed Incyte cDNA arrays and Affymetrix GeneChips to analyze gene expression changes induced by *tert*-butylhydroxyquinone treatment of human neuroblastoma cells [24•]. Cross-hybridization of the cDNA probes partially contributed to discrepancies between the data generated by the two platforms. Data generated from the oligonucleotide microarrays were more reliable for interrogating changes in gene expression, when compared with data from the cDNA microarrays. Kuo *et al* conducted the first large-scale analysis of reproducibility between spotted cDNA microarrays and Affymetrix GeneChips [23••]. The cDNA microarray studies utilized 9703 cDNA probes, while the Affymetrix HU6800 arrays comprised 7245 probe sets. A comparison of mRNA measurements of 2895 sequence-matched genes in 56 cell lines from the standard National Cancer Institute panel of 60 cancer cell lines (NCI-60), revealed extremely poor correlation and discordance between the two platforms. However, an important limitation of this study, which negated the value of the data, was the lack of replicates. Additionally, data were derived from separate microarray experiments that were conducted at two different laboratories, using different materials and experimental protocols. Although identical cells lines were studied, the cells had been cultured independently and both the mRNA samples and hybridization targets were prepared separately. This revealed that, although differences existed between the technologies, the low correlations observed were likely partially due to inherent systemic variations caused by the nature of the slide chemistries, target labeling, printing and scanning instrumentation [19••].

Woo *et al* compared the variability in measured gene expression levels associated with three types of microarray platform, Affymetrix Mouse Genome Expression Set 430 GeneChips (MOE430A and MOE430B), spotted cDNA microarrays and spotted oligonucleotide microarrays [25•]. Total liver RNA from four male mice, two each from inbred strains A/J and C57BL/6J, were assayed on all three platforms. Variances associated with measurement error were comparable across all microarray platforms. The MOE430A GeneChips and cDNA arrays demonstrated higher precision across technical replicates than the MOE430B GeneChips and oligonucleotide arrays. The Affymetrix platform provided the greatest range in the magnitude of expression levels, followed by the oligonucleotide arrays. Good concordance was observed in both the estimated expression level and statistical

significance of common genes between the Affymetrix MOE430A GeneChip and the oligonucleotide arrays. Despite having high precision, the cDNA arrays demonstrated poor concordance with other platforms [25•].

In recent years, arrays containing a single long oligonucleotide probe for each gene have become popular, and are rapidly replacing the problematic cDNA-based arrays [24•,26••]. Spotted long oligonucleotide arrays are created by the deposition of long oligonucleotides, 49 to 90 bases in length. Barczak *et al* analyzed gene expression in two dissimilar RNA samples, K652 (an erythroleukemia cell line) and Stratagene Universal Reference RNA (from Stratagene Corp) [26••]. Two platforms were compared, home-spotted oligonucleotide arrays fabricated from 70mer probes, and the Affymetrix GeneChip system. Each measurement was expressed as a pair of log-transformed differential expression (M) and total signal (A) values. Comparison of the expression measurements for 7344 genes that were represented in both platforms revealed strong correlations (0.8 to 0.9) between the relative gene expression measurements. Preliminary analysis of the probe sequences used in the study suggested a high degree of overlap between the probes on both platforms [26••].

### **Comparison of commercial platforms**

The first comprehensive study of data generated from the three most widely used commercial platforms, was conducted by Tan *et al* using PANC-1 cells, which have a pancreatic ductal phenotype [27••]. The authors examined gene expression measurements generated from identical RNA preparations on all three platforms. This experimental design facilitated a direct comparison of all the array formats and eliminated experimental or biological variation that may have arisen from independent cell culture and RNA extraction. Microarrays from Affymetrix (U95Av.2 GeneChips, multiple 25mer oligonucleotide probe sets), Agilent Technologies (human I, cDNA probes) and Amersham Biosciences (CodeLink Uniset human I bioarrays, 30mer oligonucleotide probes) were hybridized with PANC-1 RNA collected from cells grown in serum-rich medium, and were compared with serum-depleted cells (24 h following transfer to serum-free media). Three biological replicates were generated for each condition, and three experimental or technical replicates were produced for the first biological replicate. GenBank identifiers (IDs) were chosen over UniGene IDs for the comparison, to eliminate variability owing to platform-dependent probes for different splice variants. In total, 2009 common genes were identified as being present on all three platforms, and using this subset, correlations in expression levels and comparisons for significant expression changes were calculated. This revealed considerable divergence across the three platforms. Unsupervised clustering and principle component analysis (PCA) suggested that the largest variation in measurements was attributable to the platforms themselves. Despite the fact that both the Affymetrix and Amersham Biosciences platforms were single-color, short oligonucleotide platforms, no significant agreement was observed between the two platforms. The best level of agreement between the target gene sets for different platforms was only 21% (between Amersham Biosciences CodeLink and Agilent cDNA using a

2-fold induction and  $p < 0.001$  criteria). Although gene sets overlapped to some extent across these platforms, the majority of genes identified as differentially expressed by each technology were uniquely identified by that technology. An encouraging observation, however, was that when the 200 highest-ranking downregulated genes were classified according to biological themes rather than individual genes, stronger concordance was observed between the Agilent Technologies and Amersham Biosciences platforms [27••]. This suggests that sufficient genes within distinct gene ontology categories were detected by each platform to arrive at a common biological theme. The low concordance across the technologies can partly be attributed to the detection of distinct types or sets of alternatively spliced transcript variants.

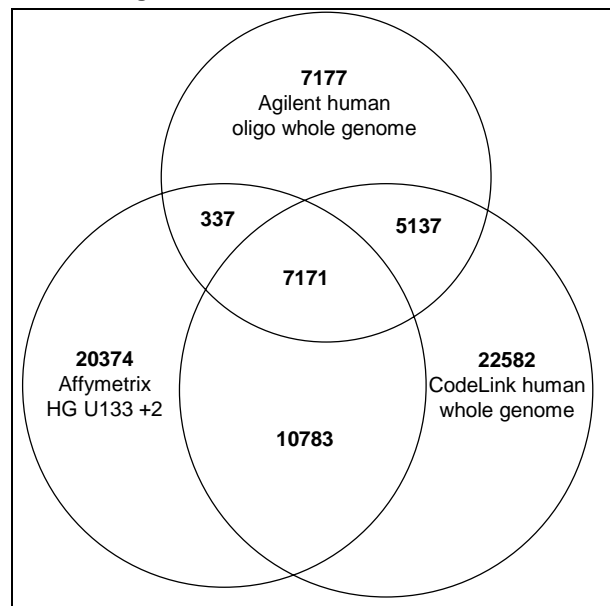
A subsequent study expanded on this research by examining six platforms: two cDNA-based (one from Agilent Technologies and one 'homemade'), three short oligonucleotide-based (one from Affymetrix, one from Amersham Biosciences and one from Mergen Ltd) and one long oligonucleotide-based platform (from Agilent Technologies) [28••]. The authors explored the hypothesis that gene expression profiles were biological rather than technological. They utilized the mouse *lacZ* model transgenic mutation test system (Muta™ Mouse), and RNA from whole lung tissue and an immortalized lung cell line (FE1) were compared on the various microarrays. Pearson product moment correlations were performed on common genes across the six platforms, based on comparisons of UniGene ID for the reporter. The resulting correlation coefficients provided a simple measure of agreement between platform pairs. The oligonucleotide arrays were highly correlated with each other, and moderately correlated with Agilent cDNA chips. Using condition tree analysis, the nature of the tissue was found to account for the measured differences in gene expression among microarray slides, regardless of the particular platform used. The expression profiles of the FE1 and whole-lung samples were observed to split on two main branches, with the Mergen spotted 30mer oligonucleotide array (MO3) and academic cDNA platforms appearing outside of these clusters. Within the two main branches, the biological replicates clustered according to platform type. Within tissue types the primary determinant of clustering was the platform, with biological replicates grouping together according to platform. The data support the hypothesis that, for broad comparisons between two sample types, conducted using different commercial and homemade platforms at different times, the primary determinants of microarray gene expression changes result from true biological differences, rather than artifacts of platform choice. Using significance analysis of microarrays (SAM), differences were observed among platforms in their ability to detect differential gene expression, with CodeLink, Affymetrix and Agilent Technologies oligonucleotide-based arrays demonstrating the best detection.

### Sequence-based probe matching

The poor overlap between the Affymetrix, Agilent Technologies and Amersham Biosciences platforms using UniGene IDs is illustrated in Figure 2. Only 7171 probes were in common among all three platforms, using this reporter. Mecham *et al* matched probes across different

platforms on the basis of sequence rather than gene identifiers [29••]. Breast cancer cell line-derived RNA was examined using Agilent Technologies cDNA and Affymetrix oligonucleotide microarray platforms to assess any advantages of this method. It was noted that, with regard to gene expression ratios and difference calls, cross-platform consistency was significantly improved by sequence-based matching. Sequence-based probe matching produced more consistent results when comparing similar biological datasets obtained by different microarray platforms. This strategy allowed more efficient transfer of classification of breast cancer samples between datasets produced by cDNA microarray and Affymetrix GeneChip platforms.

**Figure 2. Illustration of the overlap between the Affymetrix, Agilent Technologies and Amersham Biosciences CodeLink platforms using UniGene IDs.**



The lower correlation demonstrated by non-sequence-overlapping UniGene-matched probes can be explained by several factors. It may reflect splice variants or 3'/5' degradation of microarray signals along genes [30,31]. UniGene clusters assemble putative genes from cDNA clones using a variety of algorithms, although a subset of these clusters is incorrect [32]. A significant fraction of these errors have been removed in subsequent updates of UniGene and by comparison with annotated human genome sequences. The actual UniGene build utilized in some of the cross-comparative studies, however, may still contain several cases in which two cDNA clones are incorrectly listed as belonging to the same UniGene cluster. Consequently, it is possible that the cDNA feature on the spotted microarray and the Affymetrix probe may measure the expression levels of two entirely different transcripts. The low correlation observed between non-overlapping probe sets are mostly caused by this error, and this likely explains the discordance among cross-platform data. Clearly, sequence-based probe matching is required to adequately compare data, although this is not currently a trivial pursuit as the probe sequences for certain of the platforms represents proprietary information.

## Normalization, filtering and meta-analysis

After appropriate filtering, ratio and intensity data from different platforms can be compared and integrated. It is interesting to note that good agreement between platforms was obtained in the studies conducted by Barczak *et al*, after the filtering out of non-reproducible profiles using replicates from different experiments [26••]. Following image processing, data generated for the arrayed genes must be normalized before differentially expressed genes can be identified. This process is necessary to adjust for experimental variables, such as target labeling differences and variation in the detection sensitivity of fluorescent labels. Depending on the nature of the experimental design, there are different approaches for calculating normalization factors from the relative fluorescence intensities in the two scanned channels. A popular approach that might help future cross-platform studies is the use of spike controls, in which synthetic xenogenic sequences are added in increasing, but equimolar concentrations to the samples under study. The measured intensities for the added equimolar spike controls should be similar, thereby facilitating cross-platform data standardization.

Meta-analysis is a combination of techniques where the results of two or more independent studies are statistically combined to answer a particular question of interest. The underlying rationale is that the combination provides a test with more power than the individual studies themselves. Rhodes *et al* utilized meta-analysis to combine multiple datasets from different studies [33••]. Meta-analysis tests score genes by reporting a p value that expresses the probability that the observed level of differential expression could have occurred by chance. Implementation of this model revealed that four prostate cancer gene expression datasets shared significantly similar results, independent of the method and technology used. This inter-study cross-validation approach identified a panel of genes that were consistently and significantly dysregulated in prostate cancer. Analysis of these genes revealed a synchronous network of transcriptional regulation in the polyamine and purine biosynthesis pathways. This study established the first model for the evaluation, cross-validation and comparison of multiple cancer profiling studies. Choi *et al* established an alternative meta-analysis procedure based on a Bayesian approach, in which the change of gene expression in cancer was recorded as 'effect size', a standardized index measuring the magnitude of a treatment or covariate effect [34]. The effect sizes were combined to obtain an estimate of the overall mean effect. Statistical significance was determined by a permutation test extended to multiple datasets. Recently, Zhou *et al* described a second-order expression analysis approach that addressed the challenge of interpreting cross-platform data by first extracting expression patterns as meta-information from each dataset (first-order expression analysis) and then analyzing them across multiple datasets [35••]. Using yeast as a model system they were able to identify genes of similar function, without co-expression patterns. Furthermore, they elucidated the co-operativities between transcription factors for regulatory network reconstruction by overcoming a key obstacle, namely the quantification of activities of transcription factors.

## Conclusions

### Major findings

DNA microarray and oligonucleotide gene chips have emerged as powerful tools for gene expression profiling on a genomic scale, and for establishing functional relationships between large numbers of genes involved in distinct cellular processes. In addition to the detection of DNA copy-number and localization of transcription factor binding, nucleic acid arrays have been extensively utilized for the detection of gene transcription. Several DNA chip technology formats have evolved and carefully designed studies have been performed to evaluate the interchangeability of data from the various platforms. The outlook for cross-platform integration of data is currently more encouraging than the initial studies suggested. The discordances observed are attributable to the differences inherent in each of the respective platforms. The probes utilized in cDNA arrays may cause inaccurate expression measurements owing to overlap with related gene family members and the inability to discriminate between splice variants. In view of these studies, data from microarray analyses need to be interpreted cautiously, and preferably using sequence-matched probes.

Recent meta-analysis studies have been encouraging. They have provided a much-needed model for the evaluation, cross-validation and comparison of multiple cancer profiling studies.

### Future directions

As commercial manufacturers adopt standard DNA chip manufacturing practices, and arrays become clinical diagnostic tools, it is likely that many of the quality control methods currently employed in the semiconductor industry will be applied to array platforms. This will result in higher quality, higher density arrays with greater sensitivity and reproducibility, facilitating a more robust analysis of subtle changes in cellular gene expression. However, an underlying disadvantage of microarrays from a drug discovery perspective is that mRNA abundance in a cell often correlates poorly with the amount of protein synthesized, and proteins, rather than mRNA transcripts, are the major effector molecules in the cell. DNA microarrays have little utility in identifying physiologically relevant post-translational modifications of proteins, which influence protein function. Therefore there will remain a need to perform diverse assays in addition to transcriptome analysis.

## References

- of outstanding interest
  - of special interest
1. Kennedy T: **Managing the drug discovery/development interface.** *Drug Disc Today* (1997) **2**(10):436-444.
  2. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts C, Stoughton R, Burchard J, Slade D, Dai H, Bassett DE Jr *et al*: **Drug target validation and identification of secondary drug target effects using DNA microarrays.** *Nat Med* (1998) **4**(11):1293-1301.
  3. Waring JF, Ciurlionis R, Jolly RA, Heindel M, Ulrich RG: **Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity.** *Toxicol Lett* (2001) **120**(1-3):359-368.
  4. Hamadeh HK, Amin RP, Paules RS, Afshari CA: **An overview of toxicogenomics.** *Curr Issues Mol Biol* (2002) **4**(2):45-56.

5. Johnson JA: **Drug target pharmacogenomics: An overview.** *Am J Pharmacogenomics* (2001) 1(4):271-281.
6. Kruglyak L, Nickerson DA: **Variation is the spice of life.** *Nat Genet* (2001) 27(3):234-236.
7. van Someren EP, Wessels LF, Backer E, Reinders MJ: **Genetic network modeling.** *Pharmacogenomics* (2002) 3(4):507-525.
8. Vilo J, Kivinen K: **Regulatory sequence analysis: Application to the interpretation of gene expression.** *Eur Neuropsychopharmacol* (2001) 11(6):399-411.
9. Hardiman G: **Microarray platforms - comparisons and contrasts.** *Pharmacogenomics* (2004) 5(5):487-502.
10. Stafford P, Liu P: **Microarray technology comparison, statistical analysis and experimental design.** In: *Microarray Methods and Applications - Nuts and Bolts*. Hardiman G (Ed), DNA Press, Eagleville, PA, USA (2003):273-324.
  - Provides an overview of four separate microarray platforms with sample data, including an examination of differences in spot morphology. Also discusses quality control methodology, and a procedure for determining technical and biological variability, with corresponding resolution of the expression array.
11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD *et al*: **Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring.** *Science* (1999) 286(5439):531-537.
  - Seminal publication that indicates the power of gene expression and microarray technologies, and demonstrates accurate classification of the hematological malignancies AML and ALL.
12. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O *et al*: **Molecular portraits of human breast tumours.** *Nature* (2000) 406(6797):747-752.
13. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Petersen I *et al*: **Diversity of gene expression in adenocarcinoma of the lung.** *Proc Natl Acad Sci USA* (2001) 98(24):13784-13789.
14. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Meyerson M *et al*: **Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.** *Proc Natl Acad Sci USA* (2001) 98(24):13790-13795.
15. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ *et al*: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* (2002) 415(6871):530-536.
16. Beer DG, Kardias SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML *et al*: **Gene-expression profiles predict survival of patients with lung adenocarcinoma.** *Nat Med* (2002) 8(8):816-824.
17. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP: **Accessing genetic information with high-density DNA arrays.** *Science* (1996) 274(5287):610-614.
18. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* (1995) 270(5235):467-470.
19. Knight J: **When the chips are down.** *Nature* (2001) 410(6831):860-861.
  - The enormous power of DNA microarray technology is also the source of many of its problems. Errors in handling cDNA clones and cross contamination problems are discussed in this article.
20. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S *et al*: **Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.** *Nat Biotechnol* (2001) 19(4):342-347.
21. Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X *et al*: **An assessment of Motorola CodeLink™ microarray performance for gene expression profiling applications.** *Nucleic Acids Res* (2002) 30(7):e30.
22. Kothapalli R, Yoder SJ, Mane S, Loughran TP: **Microarray results: How accurate are they?** *BMC Bioinformatics* (2002) 3(1):22.
  - Provides a critical evaluation of microarray data obtained from two different commercially available systems, revealing several inconsistencies in the data obtained from the two different microarrays.
23. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS: **Analysis of matched mRNA measurements from two different microarray technologies.** *Bioinformatics* (2002) 18(3):405-412.
  - Presents a comparison of mRNA measurements for 56 cancer cell lines using Stanford type cDNA microarrays and Affymetrix oligonucleotide microarrays.
24. Li J, Pankratz M, Johnson JA: **Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays.** *Toxicol Sci* (2002) 69(2):383-390.
  - Data generated from oligonucleotide rather than cDNA microarrays were found to be more reliable for interrogating changes in gene expression.
25. Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, Churchill G: **A comparison of cDNA, oligonucleotide and Affymetrix GeneChip gene expression microarray platforms.** *J Biomol Tech* (2004) 15(4):276-284.
  - Good concordance was observed in both the estimated expression level and statistical significance of common genes between the Affymetrix MOE430A GeneChip and oligonucleotide arrays. Despite high precision, cDNA arrays demonstrated poor concordance with other platforms.
26. Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ: **Spotted long oligonucleotide arrays for human gene expression analysis.** *Genome Res* (2003) 13(7):1775-1785.
  - Provides a comparison of Affymetrix GeneChip and spotted 70mer oligonucleotide arrays, revealing strong correlations ( $r = 0.8$  to  $0.9$ ) between relative gene expression measurements.
27. Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC: **Evaluation of gene expression measurements from commercial microarray platforms.** *Nucleic Acids Res* (2003) 31(19):5676-5684.
  - Provides a comparison of gene expression measurements generated from identical RNA preparations that were obtained using three commercially available microarray platforms.
28. Yauk CL, Berndt ML, Williams A, Douglas GR: **Comprehensive comparison of six microarray technologies.** *Nucleic Acids Res* (2004) 32(15):e124.
29. •• Presents an exploration and validation of the hypothesis that gene expression profiles are biological rather than technological
30. Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z: **Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements.** *Nucleic Acids Res* (2004) 32(9):e74.
  - Demonstrates that sequence-based probe matching produces more consistent results than identifier-based probe matching.
31. Wang H, Hubbell E, Hu JS, Mei G, Cline M, Lu G, Clark T, Siani-Rose MA, Ares M, Kulp DC, Haussler D: **Gene structure-based splice variant deconvolution using a microarray platform.** *Bioinformatics* (2003) 19(Suppl 1):i315-i322.
32. Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci G, Kornacker K: **Chipping away at the chip bias: RNA degradation in microarray analysis.** *Nat Genet* (2003) 35(4):292-293.
33. Burke J, Davison D, Hide W: **d2 cluster: A validated method for clustering EST and full-length cDNA sequences.** *Genome Res* (1999) 9(11):1135-1142.
34. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM: **Meta-analysis of microarrays: Inter-study validation of gene expression profiles reveals pathway dysregulation in prostate cancer.** *Cancer Res* (2002) 62(15):4427-4433.
  - Describes the first meta-analysis model, which revealed that four prostate cancer gene expression datasets shared significantly similar results, independent of the method and technology used.
35. Choi JK, Yu U, Kim S, Yoo OJ: **Combining multiple microarray studies and modeling inter-study variation.** *Bioinformatics* (2003) 19(Suppl 1):i84-i90.
36. Zhou XJ, Kao MC, Huang H, Wong A, Nunez-Iglesias J, Primig M, Aparicio OM, Finch CE, Morgan TE, Wong WH: **Functional annotation and network reconstruction through cross-platform integration of microarray data.** *Nat Biotechnol* (2005) 23(2):238-243.
  - Describes the identification of genes of similar function, yet without co-expression patterns, using yeast as a model system.