DDT Vol. 7, No. 5 March 2002

# Current themes in microarray experimental design and analysis

**William R. Foster* and Reid M. Huber,** The Pharmaceutical Research Institute, Bristol-Myers Squibb Company, Department of Biotechnology, Rt. 141 and Henry Clay Roads, Wilmington, DE 19880, USA, e-mail: *William.Foster@bms.com

Cambridge Healthtech Institute's *Microarray Data Analysis* conference (Alexandria, VA, USA; 12–13 November 2001) provided some common sense approaches to the overwhelming task of deriving meaning from the thousands of measurements that a typical microarray experiment produces. This conference explored several approaches to achieving the promise of microarrays while avoiding costly misdirections. Topics included experimental design, use of statistics, application of data manipulations intrinsic to microarrays, practical application of microarray research and the move towards standards that enable leveraging of microarray data from multiple sources.

## Experimental design

The opening session provided the audience with connections to a tradition of well designed experiments in areas not well known to those with backgrounds in molecular biology or bioinformatics. Ed Spitznagel (Washington University, St Louis, MO, USA) gave an introduction to experimental design based on Box's *Statistics for Experimenters* [1]. The example given was that of James Lind's 1747 effort to evaluate the impact of alternative diets on scurvy. This early 'clinical trial' used a completely randomized experimental design with random assignment of the subjects to the treatment groups, the point being that these same basic principles can be applied to the design of microarray experiments. A cautionary example was presented based on the seminal study by Golub *et al.* [2] where microarrays were used to classify human acute myelocytic leukemia (AML) and acute lymphocytic leukemia (ALL) using transcriptional profiles. Analysis of covariance (ANCOVA) of the published data showed a significant contribution of the site of sample collection to observed variations in gene expression, concluding that the study design had confounded transcriptional changes with the site where the samples were prepared.

Gary Churchill (Jackson Laboratory, Bar Harbor, ME, USA) referred to the tradition of block study designs in agricultural research because similar designs can apply to microarray experiments. To prevent confounding of fluorescent dye biases with measured transcript variation in cDNA-based systems, it is common to perform fluorescent dye reversals; several circular permutation designs were presented that require the same number of hybridizations as conventional dye swap designs but can lead to improved statistical estimates: a drawback is that improved signal estimates often require more complex hybridization planning and these designs might lead to increased propagation of experimental errors, especially in cases with extreme outliers.

## The use of statistics

Properly designed microarray studies begin and end with well-characterized statistical goals and their associated experimental requirements. Several speakers pointed out that, because statistics are calculated on each measurement independently, the complexity of the statistical treatment of microarray data does not scale with the number of measurements.

The importance of replicates was emphasized; three being the minimum for meaningful analysis. However, the nature of the biological system as well as the hypothesis being tested has a substantial impact on the number of replicates required for a study. For example, one might expect to require significantly more biological replicates in a compound mechanism-of-action study in the hypothalamus than in a surrogate-marker-identification study in an immortalized cell line.

Other points of agreement included the primacy of good experimental design, maintaining statistical simplicity and the use of intensity data for statistical analysis when appropriate, rather than relying solely on gene expression ratios, which are exceedingly sensitive to noise at low expression levels. Several standard techniques were presented, including t-tests, sign tests, rank tests, analysis of variance (ANOVA) and ANCOVA.

The value of Bonferroni corrections to scale significance by the large number of measurements present in a microarray experiment was also stressed. Sorin Draghici (Wayne State, Detroit, MI, USA) and Tom Downey (Partek, St Charles, MO, USA) recommended consideration of other similarly motivated corrections, such as the less-conservative Sidak correction or the Westfall and Young step-down procedure. These corrections become less important as the measures of significance are driven by experimentally derived determinations of variation (e.g. by increased replicates).

 1359-6446/02/$ – see front matter  PII: S1359-6446(02)02184-0

### Data preprocessing and validation

Several data manipulations are specific to microarrays and are not a component of general statistical analysis. These are such things as normalization across genes, fluorescent dye bias and quality control of hybridizations. Jason Goncalves (Iobion, Toronto, Canada) presented a comparison of normalization methods, concluding that lowess normalization (which removes intensity dependent dye bias in two colour cDNA experiments) across defined microarray sub-regions performed better than commonly used alternatives. Global intensity-based normalization, which is one of the most frequently used methods in the literature, proved to be the weakest alternative.

Given the need to limit false-positive rates in microarray experiments, several organizations have focused on implementing quality control (QC) steps in the microarray analysis process. Steve Perrin (Biogen, Cambridge, MA, USA) showed a QC approach using graphs of the number of transcripts scored present in Affymetrix hybridizations versus the scaling factor of the background. This method effectively identified substandard probes and arrays. Michael Elashoff (Gene Logic, Gaithersburg, MD, USA) revealed the extensive QC insights gained by performing tens of thousands of Affymetrix (Santa Clara, CA, USA) hybridizations. Gene Logic is validating automated methods for detecting chip defects that most users perform manually, such as detecting 'crop circles' or 'haze bands'. Importantly, their methods (including principal component analysis) can identify defects such as grid misalignment or subtle alterations in RNA quality that are difficult or impossible to detect by visual inspection or by 3′:5′ intensity ratios.

An overview of the statistical methods underlying GeneChip 5.0 (Affymetrix) were presented by Tarif Awad. The move towards tunable parameters that enable user-defined trade-offs between sensitivity and specificity, as well as the use of p-values, was well received by the audience; as was the recent decision by Affymetrix to release their oligo nucleotide probe sequences to the public. There was, however, some controversy over the decision to replace perfect match measurements with estimates when mismatch signals are high, and led some to question whether this imputation could lead to incorrect estimates of gene expression.

### Applications

Interestingly, none of the presentations describing the practical use of microarrays focussed on target discovery, the classic niche for genomic technologies. Instead, applications were described in lead identification and optimization (Reid Huber; Bristol-Myers Squibb, Wilmington, DE, USA) and clinical biomarker identification (Greg Tucker-Kellogg; Millennium Pharmaceuticals, Cambridge, MA, USA). These presentations illustrated the use of various hierarchical and non-hierarchical clustering methods to identify surrogate transcript markers for a biological event. Broadening the activity read-outs of compounds in *in vitro* and *in vivo* models and identifying markers correlated with efficacious or non-efficacious outcomes in the clinic could dramatically impact the drug discovery process.

Perhaps Elashoff (Gene Logic) presented the most elegant application of transcriptional profiling and illustrated the use of gene expression databases to assess toxicity potential. The database that was presented contained the profiles of hundreds of known hepatotoxicants of rat liver. Each compound was profiled at three different doses and at three different time points, and five independent animals were profiled for each dose and time point; this design produced 45 profiles for each hepatotoxicant. Using a blinded dataset of 18 expression profiles from Pfizer, Gene Logic was able to identify the toxic class of 15 of the unknown samples; the three that did not show a significant correlation to any compound in the database all induced liver steatosis, a hepatic insult that was not present in the database during the initial queries. Similar organ-specific toxicity databases are being generated by many biotech and pharmaceutical companies and could have a dramatic impact on pharmaceutical toxicology research.

### Data standardization

The ability to warehouse and leverage biological data gathered from many disparate sources is a key concept in present-day computational biology. One need look no further than the unprecedented success of the Human Genome Project to realize the impact of such endeavors. With the formation of the Microarray Gene Expression Database (MGED) group in 1999, similar efforts are currently under way in the transcriptional profiling community. The ultimate goal of the group, as presented by Chris Stoeckert (University of Pennsylvania, Philadelphia, PA, USA), is to facilitate the adoption of standards for DNA microarray experiment annotation and data representation. Two documents have emerged from the group's activities: (1) the Minimal Information About a Microarray Experiment (MIAME) [3]; and (2) Microarray Gene Expression Mark-up Language (MAML; http://www.mged.org/). The MIAME document provides a guideline for information capture. Included is everything from experimental design and array design (format, platform, element identities), to sample descriptions (type, treatment, extract preparation, labeling) and hybridization procedures. Even image quantitation procedures and normalization methods will be captured according to the MIAME protocols.

Given the intrinsic heterogeneity in the types of biological systems that can be studied, as well as the lack of well accepted and standardized profiling platforms, standardization efforts, such

as MIAME, are bound to face considerable challenges in the future. In fact, it is possible that an as yet unknown or 'fringe' technology could displace all current oligonucleotide and cDNA-based platforms.

John Weinstein (National Cancer Institute, NIH, Bethesda, MD, USA) pointed out, in a round-table discussion, that it might be 'too soon to standardize' microarray databases and warned that the community runs the risk of establishing a standard that actually 'restricts rather than enables.' After all, there has not yet even been a standardization of the most fundamental unit of a transcriptional profiling experiment, gene annotations.

## Future directions

This conference provided a current perspective on the field of transcriptional profiling. It is clear that the field has moved well beyond the era of pure technology development, where success could be measured simply by the production of a functional microarray. We now find ourselves in the luxurious position of contemplating proper experimental design, choice of statistical methods and database construction – a sure sign of progress. Nonetheless, the challenges of data integration and compilation are readily apparent. Just as the utility of the first microbial genome sequence has increased by the sequencing of other genomes, the overall utility of any one series of microarray experiments will scale over time with increased data for comparison. Identifying enabling methods of standardization across technology platforms, processing sites and biological systems will be essential for the future success of this endeavor.

## References


1 Box, G.E.P. *et al.* (1978) *Statistics for experimenters*. (1st edn), John Wiley and Sons

2 Golub, T.R. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression profiling. *Science* 286, 531–537

3 Brazma, A. *et al.* (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat. Genet.* 29, 365–367


# Chemical genomics: discovery of disease genes and drugs

**Moulay H. Alaoui-Ismaili, Peter T. Lomedico and Satish Jindal*,** NeoGenesis Pharmaceuticals, 840 Memorial Drive, Cambridge, MA 02139, USA, *tel: +1 617 868 1500, fax: +1 617 868 1515, e-mail: satjin@neogenesis.com

Chemical genomics, a popular and efficient technology for exploiting genomic information, was the subject of a recent Cambridge Healthtech Institute's conference (*Chemical Genomics/Chemogenomics*, Boston, MA, USA; 16 November 2001). The one-day conference presented an opportunity for academic and industry leaders to discuss their visions of chemical genomics, and comprised two sessions, *Functional Analysis* and *Lead Identification–Drug Development*, under the common banner of *High-Throughput Discovery of Disease Genes and Drugs*.

How are chemical genomics, chemogenomics and chemical genetics defined? Unfortunately, these terms are used interchangeably. In practice, most people define chemical genetics as 'the use of small synthetic molecules in a genetic approach to identify key genes involved in specific biological pathways'. Chemical genomics and chemogenomics refer to the use of small synthetic molecules that are highly specific for defined protein targets, for gene function analysis and to discover new drug leads. For gene function analysis, chemical genomics/chemogenomics can be applied to multiple members of a single gene family or to a disparate collection of disease-associated proteins (e.g. unrelated genes identified by genetics or transcription profiling).

## Functional analysis

Brent R. Stockwell (Whitehead Institute, Cambridge, MA, USA) and Steven Zheng (Washington University School of Medicine, St Louis, MO, USA) described the use of exogenous target-specific ligands in a genetic approach to dissect biological processes; both used the rapamycin–TOR system (TOR being the target of rapamycin) as a model for this approach. Their studies were conducted in yeast, taking advantage of the rapid progress made in the yeast genome deletion project, which has >6000 gene knockouts. Sensitivity screens, consisting of monitoring yeast-cell growth in the presence of rapamycin, were conducted

 1359-6446/02/$ – see front matter  PII: S1359-6446(02)02185-2