© 1999 Nature America Inc. • http://genetics.nature.com

regulated during adverse drug reactions. Such information provides us with a more detailed understanding of the molecular mechanisms underlying toxic effects. Additionally, microarray technology can be used to identify gene expression profiles for new therapeutic compounds, thus providing an early screening method for potential toxic effects. An important step in using microarray analysis in toxicity screening is to construct a library of genes that are regulated positively or negatively during hepatotoxicity, as these genes may not be represented in libraries prepared from normal tissue. In order to facilitate this, we have treated rats with hepatotoxins representing diverse structural and therapeutic classes. These compounds induce a variety of hepatotoxic responses including peroxisomal proliferation, hepatic necrosis and apoptosis, cholestasis, DNA damage, protein synthesis inhibition, oxidative stress, mitochondrial damage, P-450 induction and phospholipidosis. The livers were harvested from the treated rats and mRNA was isolated. The mRNA was pooled and gene expression analysis of treated versus untreated rats was assayed using microarray analysis. The data gathered from microarray and sequence analysis allows us to identify signature sets of genes regulated during toxicity. These gene responses are then used to identify mechanisms of toxicity. Knowledge of specific toxicologic pathways can help determine potential human liabilities, and the identification of signature gene sets can be used to evaluate drug candidates for potential adverse effects. In a follow-up case study, using microarrays, we evaluated changes in gene expression in rat liver associated with hypertrophy, hyperplasia and single cell necrosis (apoptosis) following administration of a drug candidate.

Wayne, Marta

Statistical issues in functional genomics: temperature-dependent gene expression in Drosophila melanogaster

M.L. Wayne[1], L.G. Harshman[2], K.L. Simonsen[3] & A. Salmon[2]

[1]Zoology, University of Florida. Gainesville, Florida, USA
[2]Genetics, Cellular and Molecular Bioloy Section, University of Nebraska, Lincoln, Nebraska, USA
[3]Department of Statistics, Purdue University, W. Lafayette, Indiana, USA

Using the full potential of microarray data as an experimental tool will require appropriate statistical analyses. By its nature, it is quantitative data: data which exhibit continuous, rather than discrete, variation. The data may be fluorescence intensity, or ratios of fluorescence intensity at different wavelengths representing two- or four-color systems. Investigators may impose thresholds on the data to make them appear discrete, but information may be lost or is at best underutilised in the process. Further, some if not all of the genetic variation being measured is itself quantitative (*e.g.* regulatory variation). An appropriate statistical analysis which takes this into account is desirable. Further, microarray data has the additional complication that hundreds of tests are assessed simultaneously. Thus it is likely that some of the tests will appear significant purely by chance (type I error). However, traditional approaches to multiple testing, such as the Bonferroni approach, are unsatisfactory because they are too conservative, that is, too likely to reject effects of biological interest. This is particularly true if the tests are not independent, as would be the case where genes are part of a single developmental cascade or network. Two tests are proposed that address these issues, an analysis of variance (ANOVA) approach and permutation. The ANOVA allows the examination of all sources of variation contributing to continuous data on a spot by spot basis, evaluating not only the statistical significance but the relative contributions of these effects. The ANOVA includes the main effects of genotype and treatment, as well as assessing variation between independent RNA preparations and between chips. This test functions to identify all loci of potential interest, but ignores the multiple testing problem. In contrast, the permutation test focuses on multiple testing. This test reshuffles, or permutes, the identities of the spots and their fluorescence values to generate an empirical null distribution appropriate to each unique dataset. In contrast to Bonferroni corrections, however, permutation takes the correlation structure within the dataset into account, thereby generating a biologically appropriate response to non-independence. The tests will be compared using a sample microarray dataset comparing the well-studied response of *Drosophila melanogaster* to heat shock across five genetically distinct stocks. By using the heat shock system, we will be able to assess how well the tests perform in a defined context.

Weaver, Daniel C.

Simulation and predictions of gene expression behaviour through a novel mathematical model

Daniel Weaver

Genomica Corporation, R&D, 4001 Discovery Dr., Suite 130, Boulder, Colorado 80303, USA

With large sets of gene expression data, biologists will be able to construct mathematical models accurately simulating and predicting gene expression behaviour of biological systems. We present one approach to generating such mathematical models using linear coefficients (or “weights”) to calculate the regulatory interactions between any two genes. In this system, a gene’s expression level at some time t is a function of the summed regulatory inputs from the previous time step. We will show that simulations using this approach yield stable and cyclically stable gene expression levels, like those observed in biological systems. Furthermore, we will demonstrate that this system is competent to model the effects of environmental inputs and that these inputs can drive a system between steady states. Finally, we will show that this model system can predict all the regulatory interactions between a set of genes even starting with noisy simulated gene expression data.

Weinstein, John

A cDNA microarray gene expression database for cancer drug discovery

J.N. Weinstein, U. Scherf, D. Ross, M. Waltham, R. Reinhold, Y. Zhou, D.A. Scudiero, L.H. Smith, J.K. Lee, D. Shalon, D. Lashkari, M. Eisen, T.M. Myers, E.A. Sausville, D. Botstein & P.O. Brown

National Cancer Institute. Bethesda, Maryland 20892, USA
Stanford University Medical Center, Stanford, California, USA
Incyte Pharmaceuticals, Inc., Genome Systems, St. Louis, Missouri 63134, USA

Application of gene expression technology to cancer pharmacology and drug discovery is particularly challenging because most clinical tumour specimens have complex and imperfect treatment histories. In contrast, the 60 cell lines of the National Cancer Institute’s drug discovery program can be likened to 60 patients in a clinical trial in which each patient has been treated with more than 70,000 different chemical compounds one at a time and independently. This analogy provided a major reason to characterize gene expression in the 60 cell lines, despite the obvious fact that cell lines are not fully representative of clinical tumours. Accordingly, we have used 9,704-gene cDNA microarrays to assess gene expression patterns in the cell lines and found correlations of those patterns with profiles of sensitivity to a set of 1,400 potential anticancer agents. In tandem with the experimental studies, we have developed web-based analytical and data visualization tools (http://discover.nci.nih.gov and http://rana.stanford.edu) for examination of the large amounts of data derived from the gene expression patterns, drug sensitivity profiles and molecular structure descriptors of the tested compounds.

© 1999 Nature America Inc. • http://genetics.nature.com