

cations to oncology research we have developed the AmpliOnc™ I array, which contains 3 spots for each of over 50 target loci (mostly oncogenes) that have been shown to be potentially amplified in tumour tissues. In hybridizations of normal female (green) against normal male (red) total human DNA, the spots containing the androgen receptor (X chromosome) revealed a green/red ratio significantly higher than the ratios of all other loci (>99% confidence level). This suggests sufficient sensitivity for single copy-change detection. In tests with defined tumour cell lines, the expected ratios were obtained for multiple amplifications with more than 95% confidence on each chip. Multiple tumour cell lines and primary tumours have been analysed with the AmpliOnc™ chip. In all cases known oncogene amplifications were confirmed, and in some instances additional and so far unknown amplifications were discovered.

Munson, Peter

P-SCAN: quantification and statistical analysis package for cDNA microarrays

Peter J. Munson & Vinay V. Prabhu

Mathematical and Statistical Computing Laboratory, Center for Information Technology, NIH, Bethesda, Maryland 20892-5626, USA

We have developed an algorithm and software package, peak quantification with statistical comparative analysis (P-SCAN), for quantification and analysis of a wide variety of commercially available and special-purpose cDNA microarrays. The program is oriented toward experiments using radiolabelled probes (P32 or P33), nylon filter-based arrays and image data acquired from phosphorimager devices, but has also been adapted to experiments using fluorescent probes, glass-based arrays and laser scanners. The program correctly reads the Molecular Dynamics, Fuji and TIFF data storage formats. The peak quantification strategy employs a hybrid of automated and interactive methods for determination of the placement of each spot in the array, and permits several choices of algorithms for quantifying individual spots. The strategy is designed to minimize operator time and effort in processing each array while maintaining data quality. Statistical comparison of results of two hybridizations is shown in a scatterplot of normalized log-intensities together with lists of genes yielding extreme ratios, after accounting for spots falling below the background level. Output files are generated for use in subsequent analysis by such statistical packages as JMP™. These files merge information from several sources, including spot address, spot location, clone identifier, gene name if available, spot intensity, normalized intensity, intensity ratios and log-ratios. Simultaneous visualization of the original filter layout, the log-intensity scatterplot and the table of spot intensities is thereby facilitated. One may choose genes by name and highlight the corresponding spots in the scatterplot and filter layout, or choose spots by relative intensity or by array position and highlight them in the other two representations. This approach has proven invaluable in recognizing experimental and image artefacts while still uncovering real expression changes. We have applied this approach to more than 400 sets of experimental data drawn from almost 24 laboratories at the NIH, using Genome Systems, Clontech and Research Genetics commercial arrays on nylon membranes, as well as special-purpose arrays on glass. Our experience has demonstrated the value of performing independent replicate hybridization experiments to distinguish reproducible expression level changes from experimental variability and artefact. The P-SCAN program runs on many computer platforms (Macintosh™, Windows™ or Unix machines) and source code is freely available on the web (<http://abs.cit.nih.gov>). P-SCAN is written using the commercial package MATLAB™ available from The Math Works. A self-contained Java version of P-SCAN is under development.

Nadon, Robert

Statistical informatics: software tools for statistical inference about expression arrays

Robert Nadon¹, Peide Shi², Edward Susko³ & Peter Ramm¹

¹Imaging Research Inc. and Brock University, St. Catharines, Ontario, L2S 3A1, Canada

²Imaging Research Inc., St. Catharines, Ontario, L2S 3A1, Canada

³Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada

The characteristics of systematic and random measurement errors associated with expression arrays present unique challenges for statistical inference. Existing methods for estimating both types of error (e.g., use of reference genes) can be imprecise and dependent upon potentially invalid assumptions. We report novel statistical modelling procedures which estimate measurement error based upon replicate variability and properties inherent to the array. The procedures, which can be used with both micro- and macro-arrays, include outlier detection, confidence intervals, statistical tests of differences between conditions and statistical power analysis for determining the number of replicates needed to detect between-condition differences of specified magnitude.

Nagalla, Srinivasa R.

Gene expression profiles for prognostic stratification: generation of potential neuroblastoma specific microarrays by combining suppressive subtractive hybridisation and microarray analysis

S.R. Nagalla¹, Q. Tian¹, Z. Wang¹, P. Pattee¹ & R.C. Seeger²

¹Department of Pediatrics, Oregon Health Sciences University of, Portland, Oregon 97201, USA

²Children's Hospital, Los Angeles, California 90027, USA

Neuroblastomas is the most common extracranial solid tumour in children. The clinical behaviour of this tumour ranges from spontaneous regression to non-responsiveness to therapy. The molecular events that regulate the unique spontaneous regression, induced differentiation or heterogeneity in outcome within the same risk group are poorly understood. To develop a molecular basis for stratifying patients into various risk groups, we examined gene expression patterns that correlate with: (1) stage of the disease and (2) cellular differentiation of the tumour. We performed suppressive subtractive hybridisation (SSH) and microarray screening simultaneously, to compare and to combine the markers identified by both methods to develop a neuroblastoma specific array. The subtracted pool of 900 cDNA's were used to prepare a custom array. Unsubtracted and subtracted probes were used to screen neuroblastoma custom arrays and general microarrays representing known genes and EST's (10,000 cDNA's). A pool of 60 cDNAs (40% known and 60% novel genes) that were consistently differentially expressed were identified. The differential expression of 20 cDNAs was further evaluated using the Taqman sequence detection system based on real time RT-PCR quantitation. In preliminary studies, the expression of selected target genes correlated with the stage of disease and differentiation. The expression of known gene products such as cyclin A1 and p21 was further examined using immunohistochemistry, and correlated with mRNA levels. A large number of samples from retrospective clinical trials are currently being analysed for prognostic value of the identified targets using Taqman sequence detection and quantitation. Combining SSH and microarray screening will facilitate toward developing potentially disease specific and organ specific custom arrays that can be used for routine clinical stratification of patients.