

Experimental Procedures

Affymetrix GeneChip analysis of gene expression will utilize the ----- . Each gene or EST is represented on these chips by at least 11 x 25-mer oligonucleotides. The studies will be carried out by submitting RNA samples to the Keck Microarray Resource (<http://info.med.yale.edu/wmkeck/affymetrix/>) which is a full service, cost recovery unit that carries out sample processing, hybridization, and data analyses. Total RNA will be isolated using TRIzol reagent and will be further purified using an RNeasy kit (Qiagen, Valencia, CA). The quality of total RNA will be evaluated by A_{260}/A_{280} ratio, which should be at least 1.9, and by electrophoresis on the Agilent Bioanalyzer. For each experimental condition 10 μ g of very high quality total RNA will be provided to the facility who will then carry out the cRNA labeling, hybridization, and data analysis as described below.

Preparation of labeled cRNA for hybridization onto Affymetrix GeneChips will follow the recommended Affymetrix protocol. Double stranded cDNA and biotin-labeled cRNA will be synthesized from 1 to 5 μ g of total RNA by using a one cycle target labeling kit (Affymetrix). Biotin-labeled cRNA will be purified using the GeneChip Cleanup Module prior to fragmenting to a size of 35-200 bases by incubating at 94⁰ C for 35 minutes in fragmentation buffer (40 mM Tris-acetate, pH 8.1, 100 mM potassium acetate, 30 mM magnesium acetate). Integrity of cDNA, cRNA and fragmented cRNA will all be assessed by running aliquots on the Bioanalyzer prior to proceeding further.

We will take advantage of the relatively inexpensive Test-3 Gene Array to assess the quality of the biotin-labeled cRNA prior to using the considerably more expensive U133 arrays. Array hybridization buffer (100 mM MES, 1M [Na⁺], 20 mM EDTA, 0.01 % Tween 20) will be used to prehybridize the Test-3 Gene array for 10-15 minutes at 45⁰C. The prehybridized solution will be removed and replaced with 80 μ l of hybridization mixture containing hybridization buffer, fragmented cRNA (0.05 μ g/ μ l) and herring sperm DNA (0.1 mg/ml; Promega). Also included in the hybridization buffer are acetylated BSA (0.5 mg/ml, Invitrogen) and four control bacterial and phage cRNA (1.5 pM BioB, 5 pM Bio C, 25 pM BioD, and 100 pM Cre) samples to serve as internal controls for hybridization efficiency. The arrays will be hybridized for 16 h at 45⁰ in a rotisserie oven. After hybridization, arrays will be washed using the Affymetrix fluidics station and stained with streptavidin-phycoerythrin (10 μ g/ml, Molecular Probes) according to the Affymetrix technical manual. Washed arrays will be scanned on a Affymetrix GeneChip Scanner 3000.

Scanned output files will be visually inspected for hybridization artifacts and then analyzed by using Affymetrix GeneChip Operating Software (GCOS). Arrays will be scaled to an average intensity of 500 and analyzed independently. The quality of the data will be evaluated by checking the following quality control parameters:

a. Presence of spiked control cRNAs: Bio-B, C, D and CRE serve as positive controls for hybridization and are spiked at different concentrations as follows: BioB: 1.5 pm, BioC: 5 pM, BioD: 25 pM, CRE: 100 pM. The data will be considered acceptable if the hybridization intensities for these controls are present in increasing amounts, B being the least and CRE the highest.

b. Background values: The background value provides a measure of the signal intensity resulting from auto-fluorescence of the array surface and nonspecific binding of target or stain molecules. The background values for all the arrays in one experiment should be within $\pm 25\%$ of the overall average background; otherwise, this will introduce additional error. Non-specific binding causes low signal to noise ratio, which may result in genes for transcripts present at very low levels in the sample being incorrectly called as absent. Thus, high background noise creates an overall loss of sensitivity in the experiment.

c. Q value. Noise (Q value) results from small variations in the digitized signal observed by the scanner as it samples the probe array's surface. It is measured by examining pixel-to-pixel variations in signal intensities. The noise value for all the arrays in one experiment should be very similar to each other.

d. Scaling factor provides a measure of the overall brightness of the array. The brightness (image intensities) varies from one array to another. Non-biological factors (amount and quality of the cRNA, amount of stain or other experimental variation) can contribute to the overall variability in hybridization intensities. In order to reliably compare data from multiple arrays, it is essential that the overall intensity of the arrays be brought to the same level. Scaling is a mathematical technique used by the Affymetrix GCOS software to minimize differences in overall signal intensities between two arrays, allowing for more reliable detection of biologically

relevant changes. GCOS calculates the average intensity of an array by averaging the intensity values of every probe set on the array. The average intensity of any array is then multiplied by a factor (Scaling factor) to bring it to an arbitrary Target Intensity value (usually 500) set by the user. Thus scaling allows a number of experiments to become normalized to one Target Intensity, allowing comparison between any two experiments. In a particular set of experiments, Scaling factor values for all the arrays should be within two-fold of each other.

If the quality control parameters described above for the Test-3 array data are within the acceptable range, the samples will be hybridized to the standard array. The standard array images will be analyzed and quality parameters will be evaluated as described for the Test 3 arrays above. The comparison analysis will be carried out as per the guidelines provided by Affymetrix. Genes initially will be considered up or down regulated if the relative expression is changed 2-fold or more above the baseline control. The Keck Biostatistics Resource will then assist in carrying out rigorous statistical analyses as described at: <http://info.med.yale.edu/wmkeck/biostats/overview.htm#gene> to identify differentially regulated genes and ESTs. In addition to GCOs, other statistical packages, e.g. R and Dchip will be used for data normalization, intensity estimation, and other statistical analyses. Finally, expression levels of highly differentially regulated genes and ESTs will be independently verified by real time quantitative RT-PCR using the ABI 7900 Sequence detection system in the Keck Microarray Resource.