



Center for
Medical Genomics
@ Indiana University
School of Medicine

Gene expression studies using Affymetrix GeneChips®

Overview of experimental design, processing and analysis

The Center for Medical Genomics
Indiana University School of Medicine

Howard J. Edenberg, Ph.D., Director
Chancellor's Professor

Department of Biochemistry and Molecular Biology and
of Medical and Molecular Genetics

Key steps:

- Experimental design. Consider analysis before the experiment.
Information on the experimental design and samples should be submitted to Center for Medical Genomics before the samples
- Sample receipt and QC evaluation
- cDNA synthesis, cRNA synthesis
- Hybridization, scanning, data extraction
- Data Analysis:
Users should be aware that analysis is usually time-consuming, and should be comfortable setting aside the time or *discuss more extensive analytical needs with Dr. Edenberg or Dr. Jeanette McClintick (274-8450)*. There are several individuals (including faculty) associated with the CMG and/or Biostatistics who can help with analysis, but that must be arranged and budgeted. Budgeting in a grant for the efforts of such faculty is recommended if the project is complex.

• Experimental design

Obtaining the best microarray data requires careful experimental design and consideration of all possible sources of unwanted variability. Note that, unlike spotted arrays, Affymetrix microarrays do *not* use a reference sample. Each sample is hybridized to a separate GeneChip. Comparisons of single experimental samples with single control samples, even if the single samples are derived from pooled RNA, do *not* provide reasonable statistics, and is nearly useless- many of the differences are false positives that arise by chance. We have found that analysis of *at least 4 biologically independent samples for each experimental group* allows reasonable statistical analysis. To detect small changes (e.g. in brain regions) we recommend more biologically independent samples: at least 6-10 or more. [See *McClintick, J.N. and H.J. Edenberg (2006) Effects of filtering by Present call on analysis of microarray experiments. BMC Bioinformatics 7: 49* for examples of how power is affected by sample size.]

Please discuss your potential experiment with the Center at the earliest possible stage; by the time RNA is obtained, it is often too late to insure the best results. We will advise on RNA preparation and necessary Quality Controls. We do *not* generally recommend pooling of samples; biologically independent samples provide much more useful information.

A sample submission form must be filled out before submitting RNA samples to the Center for Medical Genomics. When all of the necessary information has been gathered, an experiment tracking number will be assigned.

NO RNA will be accepted before the information has been provided; then a time will be set up for bringing the samples to the CMG.

- **Information needed includes**

A written, concise description of the experiment and of EACH sample. Include species, tissue/cell type, treatment (time, drug, condition, etc), and other pertinent information. It is very important that you make clear what the key experimental comparisons will be (e.g. drug vs. control, 2x2 design). Please see Submission Instructions and Excel spreadsheet "Affy_Sample_Submission.xls".

- **Sample receipt and QC evaluation**

The investigator is responsible for preparation of **total RNA** from the samples, and for initial quality control. We can arrange to do parts of the QC at the CMG for an additional charge.

Recommended extraction:

- Animal tissue: Invitrogen Trizol (use sufficient volume and vigorous tissue disruption), followed by Qiagen RNeasy.
- Animal cell culture: Qiagen RNeasy.

The RNA should be in RNase free water; we prefer to receive at least 6 micrograms at a concentration of 0.75 $\mu\text{g}/\mu\text{l}$ or higher. *NOTE: we can work with much smaller samples, using an amplification protocol, but need to discuss this with the investigators first.*

A spectrum (220-350 nm) is necessary for QC. Often we have found contaminants that lead to grossly overestimated RNA amounts (often 8-10 fold overestimates) and poor reactions. Bring the spectra with the sample. IF QC is not good, we will consult with you about further purification. *If you have too little RNA, discuss the situation with us first.* If you don't have appropriate instrumentation, we can do that step for you, but at a minimum please check A260, A280 for every sample.

A photo of a 1% agarose gel (use 0.5 to 1 μg aliquot if possible of every RNA, and a 1 kb DNA ladder) run far enough to get good separation between the 28S and 18S ribosomal RNA (and showing the wells) is necessary. The 28S band should be brighter than the 18S band. There should be no DNA at the wells. Include a photo of your gel for ALL samples. *If there is too little RNA, please discuss it with us before bringing the sample.* We can arrange to do parts of the QC at the CMG for an additional charge, but that will require prior arrangement. If necessary, the RNA can be repurified, usually through RNeasy columns, and concentrated; arrangements need to be made with the CMG.

Contact Jun Li, email:lijun@iupui.edu or Praveena Yarru, email:lpyarru@iupui.edu, Phone: 8-9744, to arrange a suitable time to bring samples. We receive samples at the BRTC, 16th Street and Indiana Avenue (1345 West 16th St).

- **cDNA synthesis, cRNA synthesis**
- **Hybridization, scanning, data extraction**

The Center for Medical Genomics receives total RNA from investigators and carries out all of the steps of processing the RNA, hybridization to the Affymetrix GeneChips®, washing, scanning and initial analysis. We prefer to receive at least 6 microgram of total RNA for each sample; we can work with less if necessary (down to less than 0.1 microgram with amplification; that adds experimental variability, however). We follow the protocols recommended by Affymetrix in their GeneChip® Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA).

For samples of 2 microgram or more, we do single cycle labeling. This involves synthesizing cDNA using a T7 promoter-dT24 oligonucleotide as primer with the Invitrogen Life Technologies SuperScript Choice system. Following second strand cDNA synthesis and incubation with T4 DNA polymerase, the products are purified using an Affymetrix Cleanup Module. Biotinylated cRNA is made using the Affymetrix IVT kit.

When RNA is limiting, we can use the 2-cycle labeling protocol, in which the initial cRNA is made into cDNA and a second round of IVT is performed. This increases the variability between samples, so we recommend a larger number of samples per experimental condition.

The cRNA is purified using Qiagen RNeasy columns, quantitated and then fragmented by incubation at high temperature with magnesium. Fifteen µg of biotinylated cRNA is added to a total hybridization cocktail of 300 µl, and 200 µl is hybridized to a GeneChip® after adding control oligonucleotides. Hybridization is at 45°C for 17 h with constant rotation. The hybridization mixture is then removed and the GeneChips® are washed, stained with phycoerythrin-labeled Streptavidin, washed, incubated with biotinylated anti-streptavidin, and then restained with phycoerythrin-labeled Streptavidin to amplify the signals; these steps are carried out using the Affymetrix Fluidics Station. **To reduce non-random error, balanced groups of samples are handled in parallel.**

Arrays are then scanned using the dedicated scanner, controlled by Affymetrix GCOS software. Images are examined for defects. The Affymetrix® Microarray Suite version 5.2 (MAS5) algorithm analyzes the hybridization intensity data from GeneChip® expression probe arrays and calculates a set of metrics that describe probe set performance. The average intensity on each array is normalized by global scaling to a target intensity of 1000. All data including scanned images and extracted MAS5 data (expression level and detection call) are sent to the PI on a DVD.

Data Analysis

Users are responsible for their own data analysis but we can also provide analysis support or assistance. There are several individuals (including faculty) associated with the CMG who can help with analysis, but such collaborations must be arranged and budgeted. Budgeting in a grant for the efforts of such faculty is recommended if the project is complex.

The primary type of analysis starts with “absolute expression analysis.” In an absolute expression analysis, Microarray Suite examines the hybridization intensity data from each array individually to calculate a set of absolute metrics. Some of the metrics are used to determine a Detection Call for each transcript: Present (P), Absent (A), or Marginal (M), with an associated probability. We recommend not analyzing any probe sets that are not reliably detected on at least half of the arrays (for experiments of 4-6 samples per condition) in at least one of the experimental conditions. [See *McClintick, J.N. and H.J. Edenberg (2006) Effects of filtering by Present call on analysis of microarray experiments. BMC Bioinformatics 7: 49* for details about how this works and how well it eliminates false positives. Data presented there can also help determine the appropriate level of filtering for experiments of different size.] MAS5 provides a measure of the expression level (“signal”).

Most PIs will spend more time on analysis and interpretation of the experiments than on carrying out the experiment. The statistical issues for simple comparisons are not complex (we recommend t-tests on log-transformed data), but one must be aware of the problem of multiple testing corrections (we recommend False Discovery Rate per Benjamini and Hochberg). There is still the issue of finding biological meaning from the gene lists produced. ***For complex experiments, PIs are advised to consult with or collaborate with faculty of the CMG or a statistician of their choice, and consider budgeting for a fraction of their time to get assistance that will be important in turning the data into understanding (and then papers and further grants).***

It is usually important to confirm key findings with another technique, such as quantitative real-time PCR. The CMG has experience in that.

Publications

It is important to acknowledge the Center for Medical Genomics in papers.

“The microarray studies were carried out using the facilities of the Center for Medical Genomics at Indiana University School of Medicine. The Center for Medical Genomics is supported in part by the Indiana Genomics Initiative at Indiana University (INGEN®, which is supported in part by the Lilly Endowment, Inc.).”

Please send copies of papers, and information about grants you receive that use the Center for Medical Genomics facilities, so we can include them in our reports. Thanks in advance.