

FUNCTIONAL GENOMICS MODULE POLICIES AND PROCEDURES

Director: Jeannie Chen, Ph.D.
Co-Director: Timothy Triche, M.D., Ph.D.
Supervisor: Sitara Waidyaratne, R.T.

ORIENTATION AND TRAINING

All new users (investigators, technicians, fellows, residents, and students) who wish to use the major equipment in the Functional Genomics Core Facility (i.e., Affymetrix GeneChip scanners, Cepheid QRT PCR devices, ABI DNA sequencer, computers, software, and personal consultation) consult directly with Sitara Waidyaratne or Betty Schaub for technical issues and with Xuan Chen for bioinformatics and statistical analysis issues. Any principal investigator who would like training for his/her personnel must contact Sitara Waidyaratne or Betty Schaub and request training and permission for access to the requested equipment. In general, direct access is limited to heavy, repeat users who wish to minimize turn-around time for their experiments. All exceptions to the general use of the Core must be approved by Ms. Waidyaratne or Ms. Schaub, one of whom will then train the individual in the use of the scanner, sequencer, or PCR device, as appropriate. In the case of bioinformatic analysis, any user can request preliminary analysis by Ms. Xuan Chen, the full-time Core computational biologist, or use of the computer and software or both.

USE OF EQUIPMENT

Most of the equipment in the Functional Genomics Module is not available for general use, due to its highly specialized nature and heavy use (e.g., 24 hours per day, 5-6 days per week). However, two desktop computers with various statistical analysis packages [GeneSpring (Silicon Genetics), Genetrix (our proprietary software developed at USC), MAS 5.0 (Affymetrix), S Plus, SAS, MatLab] are available to all users within the Core. In addition, all users may download Genetrix for free and use it at will for their data analysis (see below for details). Users who have been trained on the equipment as described above may use the equipment, with supervision, after approval by the lab supervisor, Sitara Waidyaratne, or technician, Betty Schaub.

AFTER HOURS USE OF CORE FUNCTIONAL GENOMICS FACILITY

Most equipment in the Core is used 24 hours a day and 5-6 days per week. However, when unused periods occur between projects, repeat users with suitable permission and training, as described above, may use the Core, with or without supervision, depending on their experience level, as determined by Sitara Waidyaratne or Betty Schaub.

PILOT PROJECTS

We recognize that successful grant applications generally include pilot data. Accordingly, whenever possible the Core module will assist investigators in the design and execution of pilot projects, specifically with the intent of obtaining extramural grant support. These projects generally include a limited number of microarrays, suitable quality controls (with a limited number of specimen replicates; data from technical replicates are provided by the Core module as part of its ongoing quality assurance program, which entails processing reference RNA standards at least once a month simultaneously with user samples), and assistance with data analysis, initially by consultation with the on-site computational biologist, Xuan Chen, and by faculty level consultation with Dr. Triche or Dr. Buckley when warranted. In addition, investigators are provided with a letter of support assuring ongoing assistance with a fully developed project if successfully funded.

PROBLEMS AND ANCILLARY SUPPORT SERVICES

Problems with functional genomics core services most frequently involve large scale genomic datasets generated by high-density microarrays. Two general categories occur: 1.) problems with gene expression values from one or more gene expression arrays; or 2.) problems with SNP (single nucleotide polymorphism) calls on SNP arrays.

In the first category, the Core Director or lab supervisor frequently notes “outlier” samples before data are made available to the user. If analysis suggests a chip failure, the sample will be re-run on a new array and the resultant data added to the dataset, with a notation that the “A” (amended) array (with the same core accession number, with an “A” suffix) should be substituted for the original array. If analysis suggests either degraded or insufficient RNA, the user is notified that the sample fails QC, and a repeat sample is requested. The user may then decide whether to do so or to use lesser quality data.

A related issue that frequently occurs is whether a gene expression value from microarray analysis is a valid estimate of actual expression. To determine this, primers for any gene can be designed by the lab and used for QRT PCR, multiplexed with a standard housekeeping gene, β actin. In this way, an externally validated, comparative quantitative expression value can be generated.

The second major issue is the reliability of SNP calls on high-density SNP microarrays. In general, DNA is vastly more stable than RNA, and ambiguous calls are a function of marginal sensitivity for a given set of SNPs on the microarray. These SNPs can be eliminated from the dataset, after analysis by the Core computational biologist, with the user’s concurrence. Generally, this accounts for less than 5% of the 116,000 SNPs on the current 100K SNP arrays. In rare instances, the issue is not equivocal SNP probe sets on the array, but instead is the result of inadequate DNA available for hybridization. In this case, specific SNPs can be validated and/or detected using direct DNA sequencing on the Core DNA sequencer.

PROBLEM RESOLUTION

In the event that the above described procedures fail to satisfy the user, the user may schedule a meeting with the director, who will review the data and determine whether additional studies (at the user's or Core expense) may be warranted, whether unsatisfactory results are related to sample quality, or whether there has been a lab error. In the latter event, the user will be credited for all expenses and the Core will attempt to process the user's samples again, if so desired.

SPECIFIC LAB PROCEDURES

The Functional Genomic Core offers a broad range of services. The most common services are described below, along with the responsibilities of the user.

1. Sample Preparation:

Although the Core does not routinely process tissue samples for DNA and/or RNA, in selected cases Dr. Triche's lab will accept freshly procured tissues, frozen tissue, tissue cultures, buccal smears, etc., and extract either RNA or DNA (or both, if sufficient material is submitted). This service is generally offered only to new investigators who seek to develop the necessary skills to perform whole-genome studies and who have no prior experience. The goal of this service is to teach such users or staff from their lab the basic principles of RNA or DNA extraction that will result in high quality material suitable for microarray analysis. *Because this is not a standard service, potential users will need to contact Dr. Triche prior to submitting such specimens.*

2. Gene Expression Profiling:

Transcriptome profiling is offered for any species for which Affymetrix Gene Chips are available. In addition, custom arrays can be ordered by the user directly from Affymetrix for shipment to the Core. In each case, the user is responsible for submitting a sample for analysis in the form of total RNA. The Core will then analyze the amount, purity, and relative degradation. If there is sufficient RNA of adequate quality, the Core will prepare labeled cRNA for hybridization to the requested array. If initial QC suggests potentially inadequate quality RNA, a test array will be run. If the results warrant full analysis, the same hybridization mix will then be hybridized to the requested array. If insufficient RNA is present in the initial sample or after creating labeled cRNA, the user will be notified and no further action will be taken.

3. High Density SNP analysis:

DNA polymorphism analysis is offered using Affymetrix SNP Chips, as well as direct DNA sequencing of selected SNPs, after PCR amplification when required. Any reasonable quality DNA can be used for starting material. However, when less than 100 ug is available from a specimen, pre-amplification using Qiagen DNA amplification services is indicated. The Core will assist users in procuring this service directly from Qiagen. The sample can then be sent to the user of the Core, at user discretion. Qiagen will perform an initial QC analysis, with an estimate of "usability" for SNP analysis or DNA sequencing. If less than class I

(highest quality), the user will be contacted before further processing to ensure the user wishes to proceed. Sub-optimal call rates are then the responsibility of the user. If desired by the user, the sample will be processed along with all samples passing initial QC for analysis by SNP chips.

4. Gene Expression by QRT PCR

Frequently, users will wish to determine whether the expression values for a gene or group of genes are realistic estimates of actual expression values. The Core offers a confirmatory expression determination service, using quantitative real time PCR (QRT PCR). Up to 32 genes at a time can be processed using two Cepheid devices. Primer design by the Core is available from the Core. Samples from the hybridization mix (or fresh RNA) are run in parallel for beta actin and GAPDH, as well as for the gene of interest. The cycle threshold for control genes and unknown are determined, and relative quantification is determined by plotting Ct's for known standards (beta actin and GAPDH) versus unknown. The expression level of the unknown can then be compared to the standards and compared to the results from microarray analysis.

5. SNP confirmation

Although ~95% of SNP calls are accurate when confirmed by other methods, there remains the problem of non-uniformity across a patient population or study set. Specifically, about 5% of the SNP calls on any given array may be ambiguous, and this 5% may vary across a study population (e.g., not the same 5% of SNPs). In order to confirm potential SNP associations with a co-variate of interest, such as disease susceptibility, phenotype, etc., it is sometimes useful to analyze the population for a specific SNP. In such cases, the Core will design suitable PCR primers, amplify the SNP in question from multiple individuals, and direct sequence the amplified sequence to identify the relative proportion of the target polymorphism. In addition, the Core is evaluating alternative methods based on solution hybridization for their potential utility. Finally, the Core is also actively evaluating methods better suited for regional interrogation of polymorphisms, in order to identify specific SNPs in a susceptibility locus. It is expected that one or more of these methods will be available to Core users by early 2006.

DATA ANALYSIS

- a) New users are assigned a user ID and a temporary password by the Core computational biologist. From this point on, all specimens received by the Core from this user are linked to this user ID and password, as described below.
- b) All specimens received in the Core are assigned a unique microarray core specimen number and entered into the laboratory information system (LIS), GenoWorks. All information supplied with the sample, including the user, species, amount of RNA, and any experimental ID or covariate data are added at this time. Subsequently, with each successive processing step, the sample data is updated with amount, quality, etc., as noted below. Finally, the data are made available to the user within GenoWorks as described below.

- c) The onsite computational biologist initially reviews all data generated by the Core to determine that each array and each dataset meet minimal quality criteria and to ensure no technical errors have been introduced by the lab. In addition to the standard chip brightness, percent genes present, and 3'-5' ratios (<3) for beta actin and GAPDH performed on each array, the entire dataset is subjected to comparative analysis for outliers (noise, brightness, rare genes, physical defects). If a defective array is detected, that sample is repeated. Once data quality and integrity is confirmed, the dataset is available for use by the user.
- d) All datasets are stored in the Core, and backed up in a terabyte SAN with terabyte RAID array storage. Most files are used for archival data retrieval; only those files normally used by users are actually stored for instantaneous access on the GenoWorks server. These files can be accessed by the user via Web browser (url: http://genome_core.chla.usc.edu) and downloaded, annotated, and shared with collaborators (with read only or read/write privileges, at the user's discretion).
- e) A general purpose bioinformatic tool, Genetrix, developed at USC is available free of charge to all Core users. This software can be downloaded by users via hotlink from the Core homepage, along with the appropriate gene annotation files for their samples. During download, the user is presented with options for annotation file downloads and a warning as to the estimated download time (many of these files are quite large and not necessary for a given study). The download also determines whether the annotation files are current and/or need to be updated.
- f) If the user elects to perform data analysis on-site in the Core, two computers running not only Genetrix, but GeneSpring, MatLab, R, SPlus, and SAS are available in the Core for use by the user. This generally requires the user to meet with the on-site computational biologist and undergo a brief familiarization with the desired software, dependent on the user's prior knowledge and expertise.
- g) If the user is uncomfortable performing his or her data analysis, the on-site computational biologist will work directly with the user to perform a basic data analysis. In addition, the computational biologist will teach the rudiments of genomic data analysis, since most users find it desirable to learn at least basic analytic skills to enable datamining of their own datasets.
- h) If a sophisticated level of data analysis is required, the user will be referred to either Dr. Jonathan Buckley or Dr. Paul Marjoram, USC faculty with particular bioinformatic expertise. This is generally undertaken as a collaborative endeavor and is particularly suitable for those who intend to obtain grant funding for more comprehensive studies, as described under "Pilot Projects."

INSTRUMENTS AVAILABLE

1. Affymetrix GeneChip Scanner (2) (four-color, 5 micron capable, with 48 sample autoloaders)

2. Four fluidics work stations for GeneChip hybridization (four arrays per work station)
3. Dell desktop computers (2) loaded with GCOS and MAS software (Affymetrix), for data acquisition and archiving, as well as initial QA/QC.
4. Two additional Dell desktop computers running analytic software and GenoWorks LIS (as described above).
5. Two Cepheid Smart Cycler Sybr Green 16 channel quantitative real time PCR devices, for confirmation of gene expression values.
6. Agilent 7200 16 channel capillary electrophoresis DNA sequencer for DNA sequencing SNPs and PCR products, as needed.
7. General laboratory equipment for RNA and DNA isolation, purification, amplification, sizing, quantitation, and photographic documentation.
8. A separate PCR module for pre-amplification sample preparation.
9. A separate laboratory for tissue DNA and RNA extraction, when needed.

ACKNOWLEDGMENT OF CORE GRANT SUPPORT

All researchers who use any of the Core grant support services or facilities must acknowledge this support, as required by the National Institutes of Health, on any resulting research publications. For example: "Supported in part by Core grant EY03040, National Eye Institute."