Microarray Facilities for the Vermont Genetics Network

Bruce Aronow

Shawn Levy

Timothy McClintock

accompanied by

Edward Derrick

American Association for the Advancement of Science
Research Competitiveness Service

April, 2002

Table of Contents

Introduction

This is a report of recommendations for the installation of a facility for microarray-based expression technology infrastructure for the Vermont Genetics Network, conducted by a panel convened by the Research Competitiveness Service of the American Association for the Advancement of Science (AAAS). Since 1996, the AAAS Research Competitiveness Service has provided expert peer review and guidance to institutions around the country that are engaged in research, development and innovation.

The Vermont Genetics Network (VGN) has secured funds for the development of a microarray facility as part of its efforts to build the research infrastructure that will make Vermont geneticists more competitive. It will be housed within the Nucleic Acids core facility of the Vermont Cancer Center and College of Medicine. The director of that facility will oversee the use of and training on the new microarray equipment.

The original plans of the VGN envisioned purchasing the equipment for an Agilent-based facility. However, the VGN has an opportunity through a supplement request to NIH to re-structure the microarray facility, making it more inclusive, user-friendly, and capable of accommodating a substantial variety of user-based needs. The VGN plans to contract with an already identified research faculty member to oversee the set-up and use of the facility. The existing technical staff and dedicated Director of the Nucleic Acids Facility will be in charge of the day-to-day running of and training in the facility.

VGN requested AAAS provide advice to the VGN about the equipment that would be appropriate for these new objectives. For this purpose, AAAS convened the panel of Bruce Aronow, University of Cincinnati and Cincinnati Children's Hospital; Shawn Levy, Vanderbilt University; and Tim McClintock, University of Kentucky, accompanied by Edward Derrick, Director, Research Competitiveness Service, AAAS. The panel reviewed background materials provided by VGN and conducted a site visit March 27 and 28, 2002. The agenda for the visit is included as Appendix A.

Taking all factors into account, the panel recommends investing in the Affymetrix GeneChip system, along with a hybridization station and scanner. The rationale and technical details are explained in the next section, along with a discussion of other options. There are many accompanying factors that must be considered for this system to be effective. Recommendations for roles for personnel, details for implementation, and additional considerations, follow. A set of links to web resources for microarrays is also included in Appendix B for reference.

Technical Overview and Rationale

The Vermont Genetic Network (VGN) seeks to provide mechanisms for its member faculty to perform genome-scale analysis of mRNA abundance (i.e., gene expression). There are three general methods available to perform these types of experiments: clone and count methods, differential subtraction/amplification methods involving PCR, and microarray methods. Conventional clone and count methods such as SAGE (Serial Analysis of Gene Expression) are laborious, costly, and difficult to use in large experimental designs. A novel clone and count method, Massively Parallel Signature Sequencing, is newly available and may have promise for the future. However, it is untested in an academic setting, its technical drawbacks are not fully understood, and its future depends on the success of a relatively new and small company, making it an unacceptable choice at this time. PCR-based differential amplification methods are laborious, slow, involve significant recombinant DNA expertise, and can only be used in experimental designs involving two samples. These limitations make them unsuitable for use as the basis for a core facility.

Over the past several years, DNA microarrays have had a major impact on biomedical research and emerged as a powerful tool for the parallel measurement of relative gene expression levels. The evolution of these technologies has been driven by the development of powerful analytical approaches to utilize the enormous amount of genomic data and resources being acquired through the genome sequencing projects. Efforts by the Brown lab at Stanford University established an open source modality in DNA microarrays that helped fuel the explosion of microarray technology, making it a predominant tool in genomic research. Microarrays are flexible, a variety of commercial systems and associated support systems are available, and are by far the most common general method. Microarray methods therefore are identified as the superior choice, and indeed, VGN has focused its planning on providing access to microarray methods.

DNA microarrays can be grouped into two general categories: 1) commercial arrays with defined content such as those produced by Affymetrix, Agilent, Motorola and a number of other manufacturers; and 2) microarrays produced with variable and customizable content, generally using custom spotted cDNAs or oligonucleotides.

Custom microarrays generally involve samples of DNA with known sequence being spotted and immobilized onto a substrate, most commonly a glass microscope slide. Next, RNA isolated from samples of interest is reverse transcribed into cDNA and labeled with one of two spectrally distinct fluorescent dyes. Using dyes with distinct fluorescent characteristics allows the two labeled cDNA samples to be pooled and hybridized on a single microarray. Strands of cDNA in the pooled samples hybridize to their complementary sequence immobilized on the substrate and any unhybridized cDNA is washed off. The ratio between fluorescent signal intensities of the two dyes at a particular spot is representative of the relative abundance of the corresponding mRNA in the samples of interest. While custom microarrays allow the directed study of specific groups of genes of particular interest or the production of microarrays for virtually any model system as well as a very cost-effective means for large-scale study, they routinely suffer from a lack of standardization and rigorous quality control. The difficulty in defining analysis and statistical methods for the analysis of microarray data can be partially attributed to the lack of standards for the production and use of microarrays. There have been several recent publications regarding microarray data analysis and experimental design that have helped identify and provide mechanisms to deal with the complex variables that exist in studies using microarrays, but technical challenges still exist for the production and quality control of cDNA microarrays such that the cost of core-lab based spotted arrays can be very high and the data quality can be less than optimal.

In choosing which systems to purchase for a core facility, the experiences of previous academic microarray facilities are instructive. The most important conclusion from these experiences is that while providing the ultimate in flexibility, custom production of cDNA microarrays is a laborious multi-step process that takes a long time to establish, and is very difficult to perform with uniformly high success and reproducibility. In order to avoid potential quality control and consistency problems and thereby maximize the return on the investment in terms of grants obtained and papers published, the panel recommends VGN establish a facility that focuses on the use of commercially available oligonucleotide microarrays.

The Affymetrix GeneChip system is a well-established commercial platform with several features that make it the most logical choice for the VGN. One major concern with the Affymetrix system is the limited number of species and model systems that are currently available. To serve the wide range of research interests and model systems in use at UVM, the VGN must go beyond just this system. To that end, this report includes recommendations for mechanisms by which the VGN microarray facility can support the needs of VGN member faculty who wish to test gene expression in organisms whose genes are not arrayed on the standard sets of Affymetrix GeneChips.

Recommended Technologies

Affymetrix GeneChips

The single most expedient means of ensuring most Vermont Genetics Network investigators a means of obtaining useable gene expression data is to implement an Affymetrix GeneChip system. This system represents the current leading technology for this purpose and is deployed in a large number of major biomedical research institutions. Over 100,000 Affymetrix GeneChip analyses have been performed by the biomedical research community. By adopting this standard system, VGN investigators have the opportunity to attain a high level of inter-comparability of their data with that of other investigators from around the world. The system represents essentially a turnkey method for obtaining gene expression data from a large number of biological systems. Affymetrix currently supports a wide variety of experimental model organisms by manufacturing a series of GeneChip arrays that contain probe sets complementary to the transcription products of the corresponding organism's genome. The organisms and chip types supported by the recommended Affymetrix system currently include:

- Arabidopsis Genome Array—close to whole genome representation. Important model organism for plant biology.
- Drosophila Genome Array—whole genome array. Highly advanced annotation, will integrate extremely well with Drosophila community research efforts.
- E. coli Genome Array—whole genome array. Advanced annotation.
- GenFlex™ Tag Array—provides a mechanism for using bridge linkers to analyze limited numbers of user-defined gene products. Suitable for dozens to low hundreds of genes. Has not attained a wide following.
- Human Genome U133 Set—a whole genome array set that uses A&B arrays based on late 2001 freeze of NCBI Unigene public domain whole-genome assembly. Covers approximately 20,000 well known genes, an additional 13,000 probable genes and an additional 12,000 possible genes. Highly advanced annotations.
- Human Genome U95 Set—a whole genome array set from late 1999 that uses the five A,B,C,D,&E arrays to cover approximately 50,000 Unigene entries and includes approximately 11,000 known genes on the A chip. Replaced by the U133 chip pair.
- HuSNP Probe Array—a probeset array that recognizes human genome SNPs with a moderately high density. The PCR primer set available through Affymetrix. Has not attained a wide following.
- Murine Genome U74v2 A,B & C Set. A widely used whole genome mouse array set that contains genes from a 1999 Unigene build. Contains approximately 7,000 known genes and 26,000 independent UniGene clusters. Scheduled to be replaced late in 2002 or early 2003 upon completion of the first whole mouse genomic assembly annotation.
- Rat Genome U34 A, B, & C Set. A genome array set that covers approximately 4,000 known genes and 21,000 independent rat Unigene

Clusters. This version is useful, but far from comprehensive. It will be replaced upon completion of the rat genome sequencing, assembly, and gene annotation project.

- Yeast Genome S98 Array—Complete yeast genome single chip probeset, fully annotated, well used in the community.
- Custom Array –The custom array allows for the synthesis of a large format GeneChip with up to one million 25mer oligonucleotides. This is a very expensive option, but does allow for the capability of creating an array that is capable of probing up to 25,000 genes on single chip format.
- CustomExpress™ Arrays–The express version of the custom array allows for the synthesis of a small format GeneChip with probes for up to one thousand gene products. This is relatively less expensive option that has the capability of creating an array able to probe organisms whose complete genomes are still several years off in terms of knowing a relatively full complement of genes. All Affymetrix platforms require knowledge of the exact sequences of the gene products. Unlike cDNA-based arrays, the Affymetrix platform has no tolerance of less-than-100% matches.

Custom or Commercial cDNA or Oligonucleotide Arrays

There are many reasons why VGN members may need to use cDNA arrays or make custom arrays. Some VGN members use organisms not available on standard GeneChips. One noteworthy example is the group of researchers studying bovine health and disease. To accommodate those researchers, the panel recommends VGN provide a support mechanism for the use of pre-made spotted arrays. This entails: 1) purchase of a hybridization station to provide an automated system for the hybridization of glass-slide based microarrays and 2) purchase of a multi-color microarray scanner for acquiring images of the hybridized microarrays. This will allow the VGN facility to analyze cDNA or oligonucleotide arrays purchased from sources other than Affymetrix. Users who develop their own gene sets would have two options. They could take advantage of CustomExpress™ Arrays from Affymetrix which are custom arrays produced for the Affymetrix platform. The second option is to have these clone sets amplified and printed as cDNA arrays. Spotting services for cDNA arrays are available from either larger academic centers or from commercial sources. Following spotting, these custom arrays would be processed using the recommended hybstation and multicolor scanner.

Data Analysis

Microarray data analysis is a critical requirement for the successful implementation of expression technology infrastructure. Whereas simple experiments are relatively easily analyzed using Affymetrix software, more complex experiments require software that is capable of managing complex experiment-associated variables and contains enough statistical and visualization power to accomplish complex data pattern recognition, filtering, and clustering. State-of-the-art software should also provide high flexibility for access to internet data sources and allow for rapid capture and reanalysis of microarray data from any platform and for any organism. GeneSpring from Silicon Genetics allows the accomplishment of all of those goals and is a well-supported software environment suitable for beginners and experts. The panel's recommendation is to obtain licenses for copies of the software for use by the core and key power users.

Specific Equipment and Software Recommendations

Approximate costs for major items are as follows:

Affymetrix gene chip system	$200K
Agilent Bioanalyser 2100	$22K
Axon 4000B reader	$55K
Hybridization station	$50K
Genespring (5 copies)	$15K/year

There will be other expenses associated with the effective implementation of these facilities. These include subsidizing the development of custom chips for researchers working with organisms not available on standard Affymetrix chips. These CustomExpress™ Arrays currently cost $25,000 for design and approximately $250 per array. Furthermore, VGN must make seed money available to fund pilot and feasibility grants for faculty considering microarrays in their research projects.

Other Options Beyond Our Recommendations

The general drawbacks of competing methods were described above. Most experiments that users request are sufficiently complex that microarray methods are the only viable choice at this time. The critical decision faced by the VGN group is whether to invest in commercial or custom array technology. As described above, the many disadvantages of dealing with thousands of plasmid clones to prepare cDNA for spotting custom cDNA microarrays strongly argue against this approach. Instead, the use of commercial oligonucleotide arrays is recommended, including both an Affymetrix GeneChip system and a more general fluorescence reader to read other commercially available slides. The alternative to provide equipment to spot cDNAs prepared by members of the VGN faculty (but not by the core facility staff) would require spending about $100,000 on additional equipment. Even though it avoids having the core facility deal with amplifying gene sets for spotting, this alternative appears unnecessary because the number of users at UVM needing this service appears rather small, and this spotting service can be contracted outside of UVM.

Recommended Organization Plan

Roles for Personnel

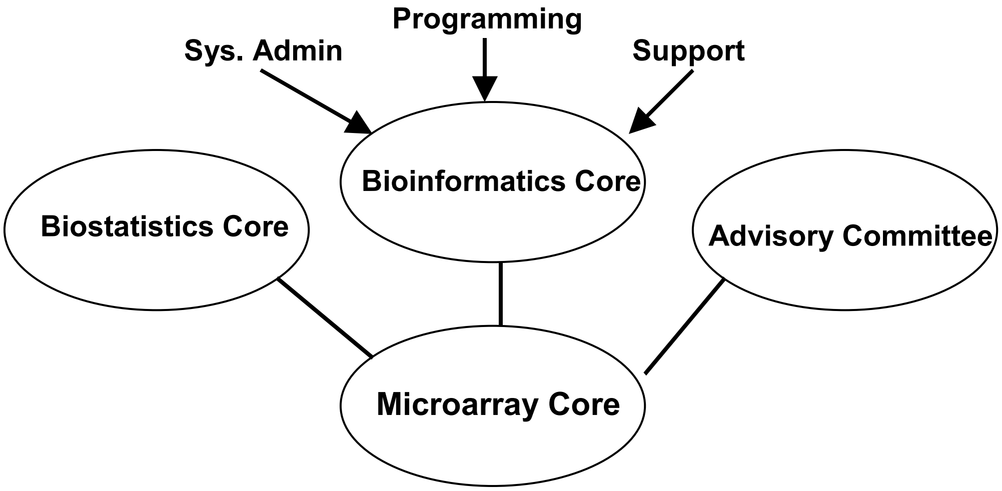

Figure 1. Relationship of the Microarray Core to other resources.

As diagramed in Figure 1, the microarray core should use the existing biostatistics resources for statistical support or consultation and the bioinformatics core for technical and programming support. The bioinformatics core should provide computer system administration, programming resources and support for the use of existing tools such as the NCBI Toolbox or other broad-base packages. The relationship of the bioinformatics core to the microarray core does not eliminate the need for a dedicated bioinformatics staff member in the microarray core. Rather, it will provide general support to the core and help remove some of the administrative distractions that the microarray core may encounter. The implementation of an advisory committee will be vital to the continued success of the microarray core. Members of the committee should be solicited from those centers that support or make heavy use of the microarray core and a quarterly meeting schedule should be established. By formalizing the meeting schedule for the advisory committee, progress and success can be closely monitored and new equipment or staffing needs can be rapidly evaluated and resources identified in a timely fashion. As with any new technology, the implementation of a microarray core facility should be considered an evolving process. Very analogous to the computer industry, the lifespan of the initial investment in equipment and technology should be considered approximately three years.

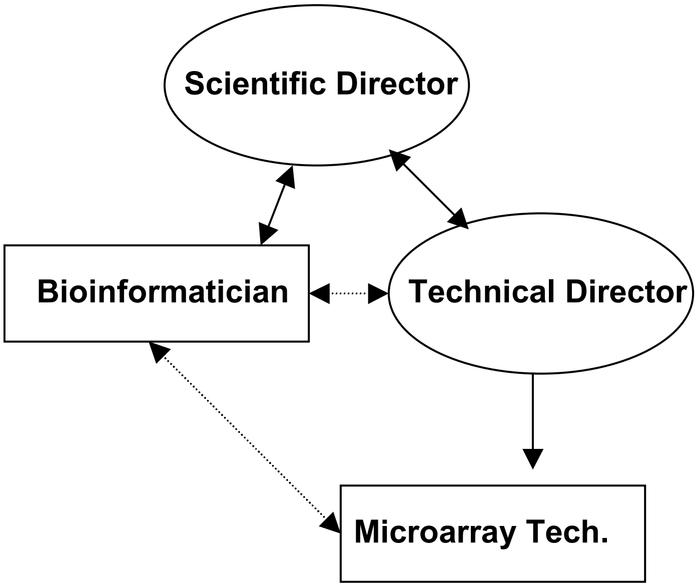

Figure 2. Staffing relationships in the Microarray Core.

Assuming a user base of 20 investigators or less, Figure 2 would represent an adequate staffing model for an Affymetrix-based microarray core facility. The scientific director would oversee all operations and provide a liaison with university administrators to ensure financial expectations are met and adequate support for the core is continued. The technical director would oversee all day-to-day operations of the microarray core and provide supervision for the microarray technician. The microarray technician would perform all of the technical procedures, provide sample handling and record keeping as well as maintain quality control and quality assurance. The bioinformatician is responsible for the design and subsequent data analysis of microarray experiments. This person would actively communicate with the microarray technician regarding file management and technical performance and maintain active relationships with the scientific and technical directors.

Comprehensive biostatistic and bioinformatic support is vital to the success of any genomic scale research program. Technologies including genome sequencing, high-throughput SNP detection, microarrays analysis of gene expression and proteomics are dependent on proper data analysis and interpretation. Most researchers are not formally trained to deal with the extremely high volume of data that is generated using the above technologies and classically trained computer scientists and statisticians generally do not have the biological background to provide the appropriate context to data analysis. Therefore, an institutional level initiative should be started to identify and recruit faculty with expertise in bioinformatics and biostatistics. These scientists would be an asset to any number of projects and future programs at the University of Vermont. In the short term, the microarray core should recruit an individual with strengths in biology as well as statistics, mathematics or computer science. Candidates should be identified from a pool of applicants with formal training in one of the mentioned disciplines and a broad exposure to the others. Personal interests and motivation should also be highly considered when evaluating candidates. A single bioinformatician should be appropriate for an initial user base of 15-20 investigators. The candidate should have significant computer experience and expertise in programming languages such as PERL, Java, PHP or C++ as well as experience with SQL and database manipulation.

This individual will be the primary interface that microarray core users interact with in the initial design of their experiments and the analysis of the resulting data sets. He will also be responsible for the organization and archival of all data files and providing software consultation and training. This individual should not be responsible for any of the operational or technical aspects of the core. To ensure efficiency and reduce distractions, software training should be given in a seminar format using a weekly to monthly format, depending on need.

The geographical location of the University of Vermont and the excellent information technology training programs in Canada may provide a good talent pool for recruiting. If recruiting at the faculty level will not be possible for the short term, the desired candidate should have significant experience in the design and interpretation of microarray data sets. If an individual currently at the University of Vermont is identified to assume the role of bioinformatician in the microarray core, he should take immediate advantage of the microarray data sets that are publicly available and become as familiar with as many different software and analysis procedures as possible. He should also take advantage of professional training offered by the recommended software and hardware vendors.

The scientific director for the microarray core should expect to spend 50% to 100% of their time in the initial implementation of the microarray core with a goal of reducing that commitment to 20% once the facility is operating. The scientific director will play a major role in evaluating the overall performance of the facility and ensuring that it is operating at the highest technological level. He should also be responsible for working with the university administration and investigators to identify and acquire external support for the microarray core.

The technical director should oversee all day-to-day operations of the microarray core and provide supervision for the microarray technician. Working closely with the scientific director, the technical director will develop and implement standard operating procedures (SOP) for the preparation of RNA, submission of sample material, labeling and hybridization of materials, data acquisition, quality control and record keeping in the microarray core. The strict implementation of these procedures will help to minimize technical variability in the core. Tim Hunter (Director of the DNA Sequencing Core) was identified as an ideal candidate for the role of technical director of the microarray core. His proven record in customer service and technical aptitude clearly show that he would be an asset to the microarray core. The DNA sequencing facility has adequate space to accommodate the microarray core and would represent the most efficient location for the implementation of microarray technologies at the University of Vermont.

Implementation Recommendations

Experiences from other microarray facilities and features of the current structure of core facilities at UVM also suggest several recommendations for implementation of the microarray facility.

1. The bioinformaticists and statisticians should immediately begin working on their skills by re-analyzing published data. This is probably best done in collaboration with a faculty member who is a potential user of the facility.
2. Set up training by Affymetrix staff. They provide several types of informative seminars and hands-on training sessions. It is not necessary to wait for the delivery of the equipment to begin training on the design of experiments.
3. Establish SOPs (Standard Operating Procedures) for the various functions of the core facility.
4. Once equipment is in place, run GeneChips and analyze the results to build expertise in technical performance and data handling. Work on achieving accuracy and reproducibility. This phase may take a few months, but it is time well spent.
5. Provide a pilot and feasibility program to allow users to generate preliminary data for extramural grant funding. This program could include small grants to allow potential users to do small GeneChip experiments, or it may involve supporting efforts of faculty using CustomExpress™ Arrays to develop reagents for organisms not supported by standard Affymetrix GeneChips.

Additional Considerations

It may be possible to expand the user base of the facility if a service is provided to isolate total RNA for users not expert in this procedure. However, most DNA microarray facilities do not provide this service, in part because most investigators should want to control this critical step themselves. It may be possible that other support facilities within the institution that would be willing to perform this service.

Plans must also be made for depreciation and replacement of the equipment. Some facilities pass this cost on to users, others do not. For planning purposes, a life span of 3 – 5 years should be expected of the DNA microarray equipment. What you purchase may very well outlive this time frame, but it is reasonable to expect that advances in technology in this still evolving field will make it advisable to upgrade by then.

Sustainability is another issue that requires planning for the future. During the start-up phase of the facility, the VGN grant or the institution should be prepared to subsidize training, supplies, and the occasional piece of standard lab equipment (e.g, desktop centrifuge, refrigerator, etc.). Incorporating support for the microarray facility into program project grants, center grants, and core grants is one way of subsidizing and holding down costs beyond the life of the VGN grant.

Appendix A: Schedule for Panel Visit

March 27

8:00 - 9:00am	Breakfast with Judy van Houten, VGN co-Director
10:00 - 11:00am	Meet with Jeffrey Bond, Director, Bioinformatics Core Rama Kocherlakota, member, Bioinformatics Core Ahmad Chaudry, Director, VGN Microarray facility
11:00 - noon	Meet with potential users of microarray facilities: Yvonne Janssen-Heininger, Pathology Tom McFadden, Animal Sciences Karen Plaut, Animal Sciences Cedric Wesley, Microbiology & Molecular Genetics
Noon - 1pm	lunch
1:00 - 2:00pm	tour the Vermont Cancer Center DNA Analysis Facility with Tim Hunter, Director
2:00 - 2:45pm	Meet with Russell Tracy, Associate Dean, College of Medicine
3:00 - 3:30pm	Meet with David Yandell, Director, Vermont Cancer Center
3:30 - 4:30pm	Meet with potential users of microarray facilities: Richard Galbraith, Medicine Corey Teuscher, Medicine Ming Hau, Medicine
4:30 - 5:30pm	Meet with potential users of microarray facilities: Russ Hovey, Animal Sciences Joanne Knapp, Animal Sciences Victor May, Anatomy and Neurobiology Felix Eckenstein, Neurobiology
6:30pm	Dinner with Judy Van Houten, Jeffrey Bond, Rama Kocherlakota, Ahmad Chaudry

March 28

10:00 - 10:30am	Meet with John Burke, Vice Provost for Research
10:30 - 11:30am	Meet with potential users of microarray facilities: Dieter Gruenert, Medicine Barbara Beatty, Pathology Ralph Budd, Medicine
11:30 - 12:30pm	Meet with potential users of microarray facilities: Uma Wesley, Microbiology and Molecular Genetics Maria Ramos, Pathology Research Associate Nick Heinz, Pathology
12:30 - 2:20pm	Lunch and panel work time
2:20 - 3:30pm	Meet with Judy van Houten and Chris Allen

Appendix B: Links to Web Resources on Microarrays

Please note that several of these sites are themselves collections of links to other sources of information.

Stanford Microarray Forum
http://cmgm.stanford.edu/cgi-bin/cgiwrap/taebshin/dcforum/dcboard.cgi

MGED main web resources (standards for microarray annotation)
http://www.mged.org

Stanford Microarray Database
http://genome-www.stanford.edu

Collection of links and resources for microarrays and instructions to join the UCSF mailing list
http://www.gene-chips.com

Public microarray resource
http://www.microarrays.org