

The ABRF MARG Microarray Survey 2008: Sensing the State of Microarray Technology

C. Harrington¹, S. Hester^{2*}, H. Auer³, N. Jafari⁴, S. Potter⁵, J. Tiesman⁶, R. Jensen⁷, L. Reid⁸, A. Massimi⁹, N. Denslow¹⁰;

¹Oregon Health & Science University, Portland, OR, United States, ²US EPA, Durham, NC, United States, ³Institute for Research in Biomedicine, Barcelona, Spain, ⁴Northwestern University, Chicago, IL, United States, ⁵Children's Hospital Medical Center, Cincinnati, OH, United States, ⁶Procter & Gamble, Cincinnati, OH, United States, ⁷Virginia Bioinformatics Institute, Blacksburg, VA, United States, ⁸Formerly With Expression Analysis, Durham, NC, United States, ⁹Albert Einstein College of Medicine, New York, NY, United States, ¹⁰University of Florida, Gainesville, FL, United States.

ABSTRACT

Over the past several years, the field of microarrays has grown and evolved drastically. In its continued efforts to track this evolution, the ABRF-MARG has once again conducted a survey of international microarray facilities and individual microarray laboratories.

The goal of the survey is to profile the current state of microarrays and to gain insights into new trends in the field. The survey is composed of seven parts:

- 1- General Section
- 2- Platforms, Services & Throughput
- 3- Commercial Arrays
- 4- Bioinformatics: Data Analysis
- 5- Bioinformatics: Data Management
- 6- Next Generation Technologies
- 7- Future Directions

This is the fifth microarray survey conducted by the ABRF-MARG since 1999. The results of the survey are presented and new trends are discussed. Additionally, the survey is evaluated against past surveys to provide insights into the growth and evolution of the community of microarray researchers over time.

INTRODUCTION

DNA microarray technology has emerged as a powerful and prominent tool for the molecular biologist to assess changes in gene expression and DNA sequence on a global scale. As a result, a multi-billion dollar industry has evolved that is based on the production, use, and analysis of DNA microarrays. Moreover, many of the principles that have originally been developed for use with DNA microarrays have since been applied to the study of other macromolecules (e.g., proteins and antibodies) in a microarray setting. The aims of this survey were to construct a current profile of a microarray facility, compare the current profile with that of profiles previously generated by the MARG, and provide some insight as to where this technology and global RNA and DNA analysis is going. In the first half of this decade two DNA microarray platforms dominated the field, the slide-based technologies developed in the laboratories of Patrick Brown and Ronald Davis at Stanford University and the GeneChip technology developed by Affymetrix, Inc. As reflected in the current survey results, new array platforms and new global profiling technologies have made significant inroads in the research community and the core lab.

METHODS

Survey Development.

The survey was designed to collect information concerning instrumentation, protocols, staffing, arrays and throughput in a microarray facility. The survey consisted of 7 sections: General Section (23 questions) Platforms, Services & throughput (29 questions), Commercial Arrays, (7 questions) Bioinformatics (8 questions), Data Management (7 questions), Next Generation Technologies (9 questions), and Future Directions & MARG Projects (19 questions). The survey was hosted by SurveyMonkey.com. The responses were collected by a third-party who removed any information identifying the participant prior to making the data available to the MARG for analysis. Thus, all participants have remained anonymous.

Survey Dissemination and Participation.

The survey was announced on Nov. 1, 2007 by posting instructions for participation on the ABRF and microarray related electronic discussion groups and listservers, including the GeneArray List Serv. The survey was closed on Dec. 31, 2007. In addition to laboratories that offer microarray technologies as a shared resource, individual laboratories that have these technologies were also invited to participate. Participation was open to anyone regardless of whether they were affiliated with the ABRF. Participants had the option of completing only sections of the survey that related to their microarray operation. This survey data was analyzed to build a current profile of microarray analysis laboratories.

RESULTS AND DISCUSSION

Presented and discussed herein is a summary of the responses to the 2007 - 2008 MARG Survey. A detailed analysis of the survey will be posted on the ABRF web site (www.abrf.org) after the ABRF 2008 Meeting.

General Section:

Demographics and Facility Profile

- Number of survey respondents: 149
- Institution Type: 77% Academic/Gov, 21% Commercial/Pharma
- 71% operate as core facility (open access)
- Geographic Location: 74% US/Canada, 24% Europe, 2% Asia
- # MA Facilities/Institution: 1 (52%) >1 (42%), >4 (4%)
- Use formal Good Laboratory Practice program: 21% Yes
- Offer genomic workshops: 61% Yes
- Full Service Facility: from RNA isolation, 55%; from sample labeling to data acquisition, 79%
- Assess sample quality via Bioanalyzer: 87% Yes

These results are similar to the demographics of 2001, 2003 and 2005 MARG survey respondents in which at least two-thirds of the respondents were from an academic setting located in the US and Canada. Thus, comparisons can be made between the results of the 2008 survey and the previous surveys.

Personnel

- number of personnel per MA facility: 1=18%; >1 =82%; >3=39%
- 66% staff members have at least 4 years of experience
- 88% MA facility director/manager has >3 yrs; 54% >6yrs years
- 68% of the labs plan not to expand the number of personnel over the next year
- 45% of facilities have dedicated bioinformatics staff

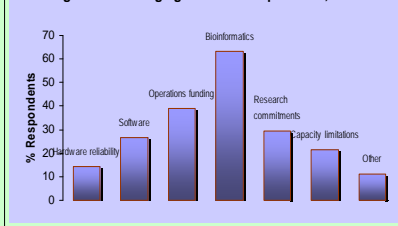
The typical number of staff per MA facility has dropped from the average number reported in 2003 of 3.8; the majority of cores now have 3 or fewer staff. The years of experience of both the director and staff members have increased from previous averages of 2.3 and 1.5 years, respectively. More than half of the facility directors now have more than 6 yrs experience with microarrays. In 2005, 52% of respondents indicated that they would expand the number of personnel in their facilities while only 32% will in 2008.

Data Validation and Challenge Areas

- 84% validate their MA results using real-time PCR
- 63% of the respondents indicated that bioinformatics (data management and analysis) is the major challenge for a MA facility (Fig. 1)

In 2001, equal numbers of labs were using Northern blot, RNase protection and real-time PCR to validate data. In the 2003, 2005, and 2008 surveys, 80%, 88%, and 84% of the respondents, respectively, validate with real-time PCR. Thus, real-time PCR continues as the method of choice to validate MA data. Bioinformatics continues to be the greatest struggle for MA facilities.

Figure 1: Challenging Areas for Respondents, N=98



Platforms, Services, and Throughput Platform

- The Affymetrix GeneChip system is the most widely used platform (58% of respondents currently use); Agilent and in-house spotting are in use in about half of facilities, and the Illumina BeadStation in approximately one third (Figure 2)
- 19% of respondents are using a robot for labeling samples, but only 12% of remaining facilities are planning to buy a labeling robot in the next year.

Services

- RNA isolation services offered by 50% of facilities; DNA isolation by 29%
- 85% of facilities provide sample labeling for expression; 64% label DNA
- 78% provide limited data analysis; comprehensive data analysis (normalization, low level statistical analysis) available in 59% of responding facilities
- 54% of facilities provide support for data publication

Throughput

- 29% process fewer than 200 arrays (RNA & DNA analysis) per year; 38% 200-1000 arrays; 18% 1000-2000 arrays, and 18% more than 2000 arrays per year (Figure 3)
- DNA analysis: 43% <1000 per yr; 35% <1000 per yr; 7%, 2000 per yr; 15% >2000 per yr
- Throughput predictions for 2008: Increase - 52%, stay the same - 38%, Decrease - 11%

Commercial Arrays

Array use

- 77% of respondents use commercial arrays for most or all of their array throughput
- Among respondents (n=90), Affymetrix arrays are preferentially used for expression analysis in 48% of facilities, Agilent arrays in 21%, Illumina arrays in 20%, and in-house spotted arrays in 16%
- Whole genome expression profiling is the most widely used array application on both Affymetrix and Illumina platforms (Table 1)

Table 1. Array type use on Affymetrix and Illumina platforms

ARRAY TYPE	AFFYMETRIX (n=63)	ILLUMINA (n=37)
EXPRESSION:		
whole genome	98% (3/TVT)	86%
whole genome: exon	60	N/A
targeted expression	N/A	32
custom expression	32	8
GENOTYPING:		
whole genome: 10+ markers	29	27
whole genome: <10 markers	44	60
targeted	8	49
custom	16	35
OTHER:		
tiling	48	N/A
resequencing	22	N/A

QA/QC

- 99% of respondents routinely evaluate RNA quality prior to expression analysis
- Measurements used in evaluation:
 - OD₂₆₀/OD₂₈₀ ratio - 77%
 - Bioanalyzer trace - 92%
 - Agarose gel pattern - 19%
- While almost all respondents use Bioanalyzer electropherograms to evaluate RNA quality, only 44% find the trace "highly predictive" of array performance; an additional 50% find that the trace "usually correlates" with performance
- Following hybridization, respondents (n=83) use the following QC steps for expression analysis arrays as indicated:
 - Image inspection - 89%
 - Review of analysis metrics - 90%
 - Data visualization - 80%

Figure 2: MA Platform Usage N=109

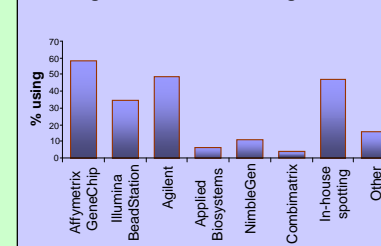
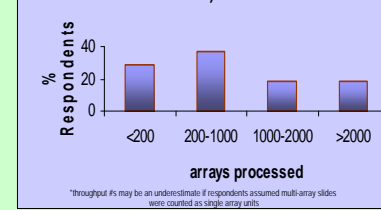


Figure 3: Total Microarray Throughput* (12 months) N=104



*throughput #s may be an underestimate if respondents assumed multi-array slides were counted as single array units

Bioinformatics

Bioinformatics continues to present challenges (Figure 1) as noted in previous MARG surveys. Data analysis is time intensive and frequently requires input from individuals with different expertise. Therefore, bioinformatics issues still represent a significant bottleneck in the array process

Infrastructure

Respondents to this section: 82/149 (73%)

- Who analyzes the data?: Individual PI (74%) Statistician (55%) Contractor (23%) Other (10%)
- 45% of respondents have 1 or more FTE's dedicated to analysis

Data Analysis

- 60% use GeneSpring, 45% use R packages, 30% use Excel for array analysis
- 50% use Ingenuity, 35% use EASE/DAVID, 33% use Pathway Assist/Pathway Studio for analysis of biologic significance

Data Management

- Only 47% of the respondents keep their data in a database, using software that most find, at best, satisfactory
- 80% of the respondents sometimes or always upload data to MIAME compliant public databases: 84% Gene Expression Omnibus; 31% Array Express. Over 50% of respondents find data publishing "not easy, but doable"
- Although most users are satisfied with software solutions provided by array manufacturer, more than 50% also rely on alternative software for array processing and analysis.

Acknowledgements: We thank ABRF for use of Survey Monkey.com. Dr. Harrington has a significant financial interest in Affymetrix, Inc. This potential conflict of interest has been reviewed and managed by OHSU. The contents do not necessarily reflect the views and policies of the EPA, nor does mention of trade names or commercial products constitute endorsement or recommendation of use.

Next Generation Technologies

HT Sequencing

- 8% have HT Sequencer (n=78). Instrument most used by respondents: Illumina-454>ABI (Figure 4)
- 24% respondents collaborate with someone who has a HT sequencer
- 46% respondents will purchase a HT sequencer in near future
- 36% process >100 samples per year
- Only 9% respondents think HT sequencing will displace microarray technology within next 5yrs; 62% think HT sequencers will partially displace MAs

Figure 4: HT Sequencer Used N=23

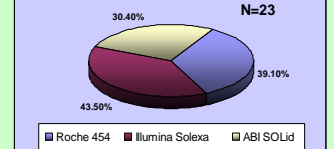
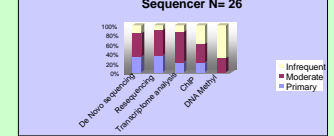


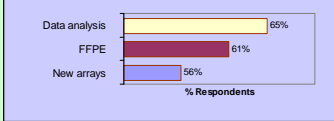
Figure 5: Applications Used on HT Sequencer N= 26



Future Directions

- 68% respondents interested in Splice Variant (SV) applications
- Of those half are currently running SV arrays
- 51 respondents run SNP arrays; 2 platforms-Illumina & Affy
- 50% interested in running arrays on clinical samples
- 38% of these have investigated what regulations must be considered to run these arrays
- 65% would like to see Data Analysis as a future MARG research project (Figure 6)

Figure 6: Suggested New MARG Projects



SUMMARY

- Affymetrix GeneChip system is most widely-used platform, followed by Agilent and in-house spotted arrays. Illumina Bead Station is close behind.
- Expression profiling is still the most frequently used array application.
- 91% of respondents expect array throughput to increase or stay the same in 2008.
- As in previous MARG surveys, bioinformatics continues to be the primary challenge for MA users and core facilities.
- Only 9% of respondents expect HT sequencing to completely replace microarrays.