

# **A Current Profile of Microarray Laboratories: the 2002-2003 ABRF Microarray Research Group Survey of Laboratories Using Microarray Technologies**

**K.L. Knudtson<sup>1</sup>, C. Griffin<sup>2</sup>, D.A. Iacobas<sup>3</sup>,  
K. Johnson<sup>4</sup>, G. Khitrov<sup>5</sup>, S. Levy<sup>6</sup>, A. Massimi<sup>3</sup>,  
N. Nowak<sup>7</sup>, A. Viale<sup>8</sup>, G. Grills<sup>9</sup>, A.I. Brooks<sup>10</sup>**

***<sup>1</sup>University of Iowa, <sup>2</sup>University of California, San  
Francisco, <sup>3</sup>Albert Einstein College of Medicine,  
<sup>4</sup>Jackson Laboratories, <sup>5</sup>The Rockefeller University,  
<sup>6</sup>Vanderbilt University, <sup>7</sup>Roswell Park Cancer Institute,  
<sup>8</sup>Memorial Sloan-Kettering Cancer Institute, <sup>9</sup>Harvard  
University, <sup>10</sup>University of Rochester Medical Center***

# ABSTRACT

New microarray-based methods and instrumentation are constantly being introduced and the number of investigators using these technologies is rapidly expanding. This study is the third annual general survey conducted by the ABRF Microarray Research Group (MARG). The goals of this study are: (1) to build a current profile of microarray facilities; (2) to compare the current survey results with previous profiles to illustrate how microarray technology is evolving; and (3) to provide insight as to where this rapidly developing technology is going. This survey focused on spotted DNA microarray and Affymetrix GeneChip technologies. Information was also requested regarding the use of other microarray platforms. Data was requested from laboratories by posting instructions for participation on microarray related electronic discussion groups. The survey was aimed at gathering information from academic, pharmaceutical, and commercial laboratories that offer microarray technologies as a shared resource. Individual laboratories that have these technologies could also participate. A web based survey was used to collect information such as instrumentation, protocols, staffing, funding, costs, and throughput. The resulting analysis gives a picture of the state of the art of microarray analysis.

# INTRODUCTION

**DNA microarray technology has emerged as a powerful and prominent tool for the molecular biologist to assess changes in gene expression at 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 a profile previously generated by the MARG, and perhaps provide some insight as to where this technology is going.**

**Currently, two DNA microarray platforms dominate the field. Namely, 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. One of the additional aims of this survey, therefore, has been to address specific aspects unique to each platform.**

# METHODS

**Survey Development.** A survey was created that contained questions designed to collect information concerning instrumentation, protocols, staffing, funding, and throughput in a microarray facility. The survey consisted of 3 sections: a General Survey Section (28 questions), a Custom Array Section (32 questions), and an Affymetrix Section (23 questions). The survey was converted to a web form by the Web Design Group who also developed the back-end data collection module. The responses to the survey were collected by a third-party who removed any information that would identify 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 by posting instructions for participation on the ABRF and microarray related electronic discussion groups and listservs. 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 2003 MARG Survey. A detailed analysis of the survey will be posted on the ABRF web site ([www.abrf.org](http://www.abrf.org)) after the ABRF 2003 Meeting.

## General Section.

### •Demographics and Facility Profile

- 115 individuals responded

- 74% were from an academic setting

- 73% were from laboratories located in the United States and Canada

- 78% work in a core or service setting

- There are an average of 2.2 MA facilities per institution

- 76% offer their services on a non-profit basis

- 65% offer their services to users outside their institution

- 57% offer teaching programs to their users

These results are similar to the demographics of the respondents of the 2001 MARG Survey in which 81% of the respondents were from an academic setting located in the US and Canada. Thus, comparisons can be made between the results of the 2003 survey and the previous survey.

## **General Section (cont.)**

### **•Personnel**

- Average number of personnel per MA facility = 4.7**

- Staff members have an average of 3.4 years of experience**

- MA facility director/manager has an average of 4.5 years of experience**

- 52% of the labs plan to expand the number of personnel**

- 49% of facilities have dedicated bioinformatics staff**

**While the number of personnel ( $4 \pm 0.7$ ) per MA facility have remained the same since the MARG 2001 survey, the years of experience of both the director and staff members have increased from an average of 2.29 and 1.46 years, respectively. Interestingly, in 2001, 83% of respondents indicated that they would expand the number of personnel in their facilities. The ca. 2 year increase in average levels of experience is likely due to the 2 year time period that has elapsed since the last survey.**

## General Section (cont.)

### •Data Validation and Challenge Areas

- 80% validate their MA results using real-time PCR
- 62% of the respondents indicated that bioinformatics which includes MA data management and analysis creates the greatest 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. Thus, real-time PCR has become the method of choice to validate MA data. Bioinformatics continues to be the greatest struggle for MA facilities. One user complained that their greatest challenge was working with explaining statistics to biologists.

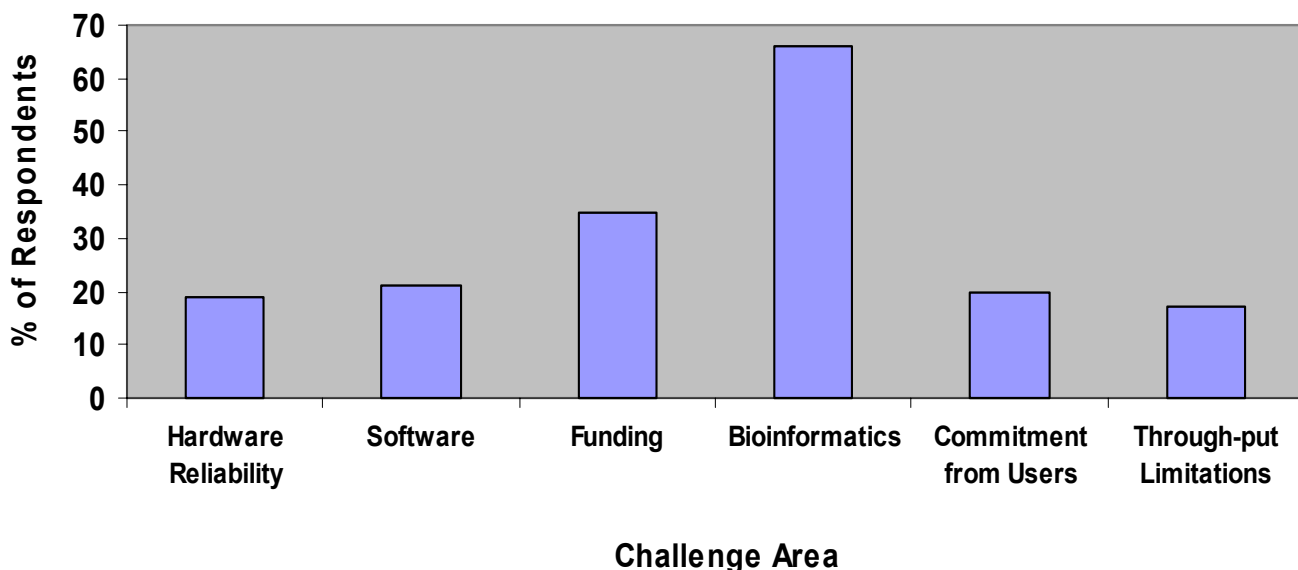
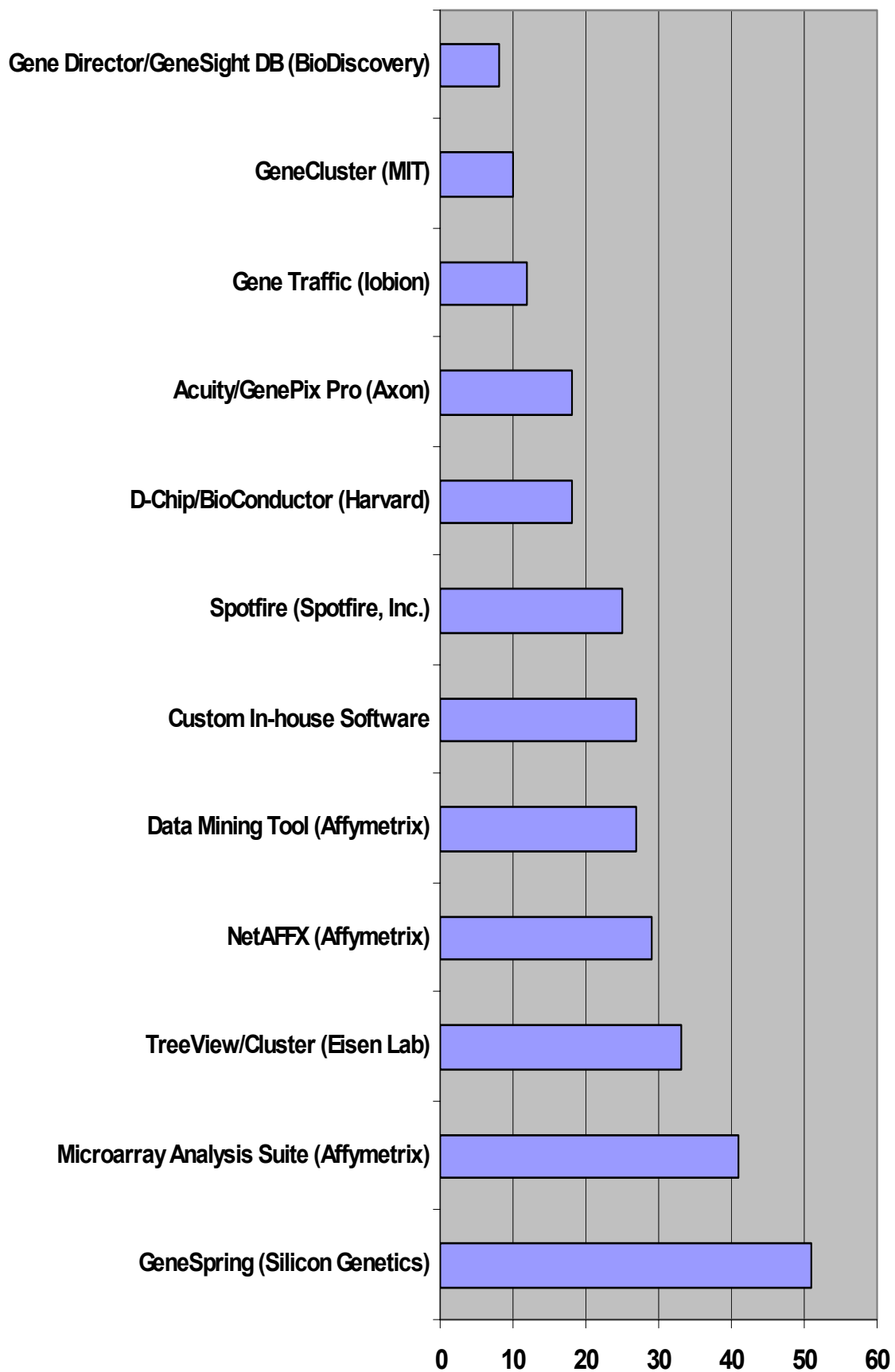


Figure 1. Challenge areas for microarray facilities.

## **General Section (cont.)**

- **Data Analysis and Storage**
  - **36% of respondents indicated their facility uses a LIMS**
  - **Most MA facilities are storing their data on servers, DVD/CD or both**
  - **63% export and/or store their data in Excel, however, more users are employing more robust storage systems such as GeneTraffic (Iobion) and MySQL.**
  - **As shown in Fig. 2, many different MA software programs are being used by MA facilities.**

**In contrast to the 2001 survey, more MA facilities appear to be adopting the use of LIMS and more robust data management software packages. The number of software packages reported to be used to analyze MA data has greatly increased since the 2001 survey. The Silicon Genetics product, GeneSpring, was the most popular choice, but only represented less than 13% of all the software packages used by MA facilities. Thus, there does not appear to be one MA software analysis package that fits everyone's needs. This notion is supported by the fact that 27 respondents indicated that they are using custom analysis software that was developed in-house.**



**Figure 2. Software used by MA facilities. Software packages used by 8 or more respondents are listed. An additional 38 software packages, not listed, were also named.**

## **Custom Array Section.**

### **•Custom MA Facility Profile**

**•71 individuals responded to the Custom Array Section**

**•74% were able to generate usable data in less than 1 year**

**•At least 90% of the facilities provide services for printing, hybridization and scanning of arrays (Fig. 3)**

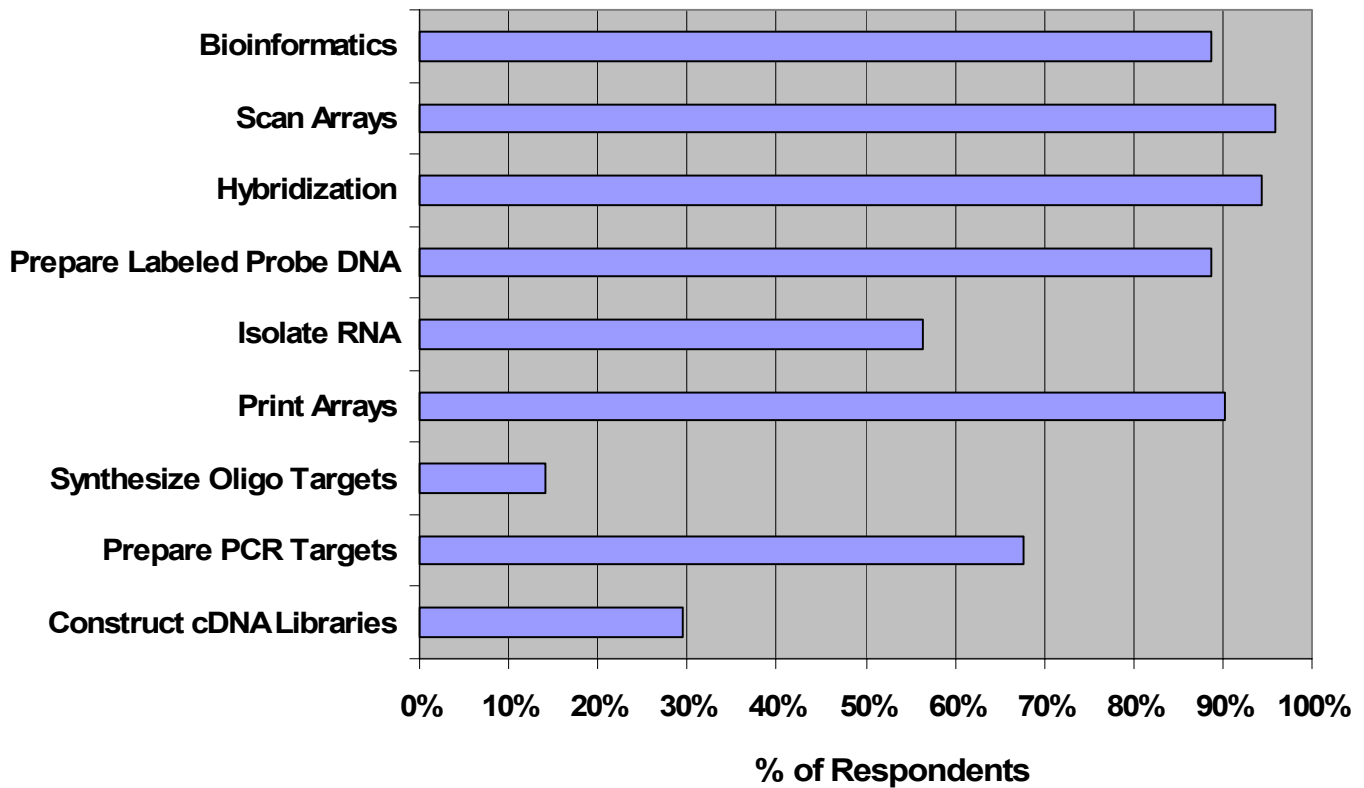
**•70% have just one arrayer, 54% have just one scanner, and 32% of respondents use a hybridization station**

**•Fig. 4 illustrates most facilities either print, label and scan less than 50 arrays or 100 to 299 arrays per month**

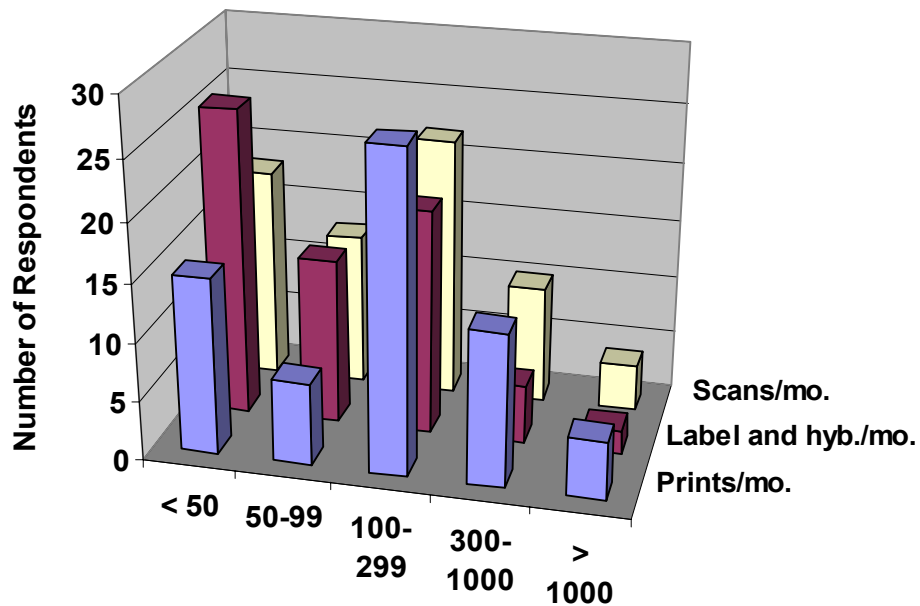
**•41% of all facilities provide arrays to 5 or less groups**

**•84% of facilities are printing arrays for use with differential gene expression studies, 6% used for SNP studies**

**Most users seemed to be at least moderately satisfied with the performance of their arrayers (89%) and scanners (91%). But, users continue to be less satisfied with the support they receive for their instruments. This represents a slight increase in performance satisfaction from the the 2001 survey. Thus, it appears that most users are becoming more confident in the performance of their instrumentation. The most common complaint was the lack of reproducibility between array print runs. Curiously, it appears that only a handful of groups at any institution are taking advantage of their custom MA facility.**



**Figure 3. Types of services performed by custom MA facilities (N=71)**



**Figure 4. Average number of arrays printed, labeled and hybridized, and scanned per month.**

## Custom Array Section (cont.).

### •Sample Printing and Preparation

- 47% of labs use an amplification protocol in their sample preparation

- 99% of facilities are using fluorescent labeling and 99% of those labs are labeling with Cy3 and Cy5

- Nearly 75% are printing human and mouse targets

- Most users are normalizing between arrays by using an average of array global intensity

- As shown in Fig. 5, over 70% of users are employing the use of a universal RNA reference or synthetic exogenous controls

- As shown in Figs. 6 & 7, 39% of users are using amino silane coated slides and 80% are using a contact pin printing format.

- As shown in Fig. 8, 3X SSC and/or 50% DMSO are the preferred printing solutions.

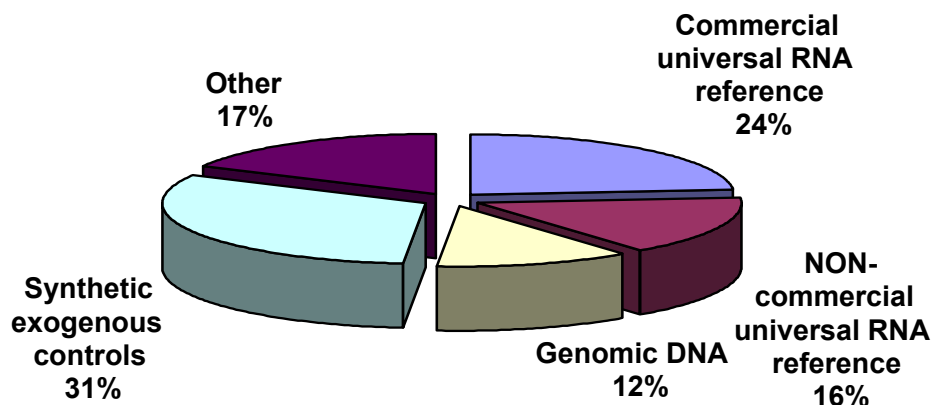


Figure 5. Sources of Normalization Markers (N=57) 26% of respondents report employing multiple sources

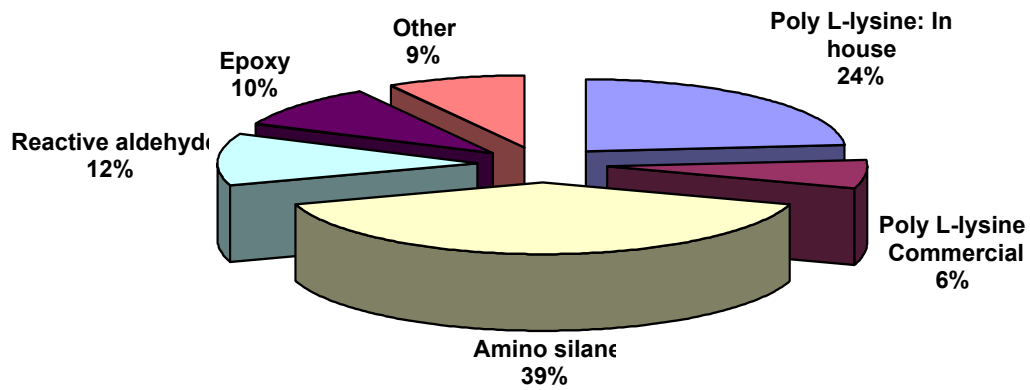


Figure 6. Material used to coat slides (N=67) 34% of respondents report employing multiple slide types

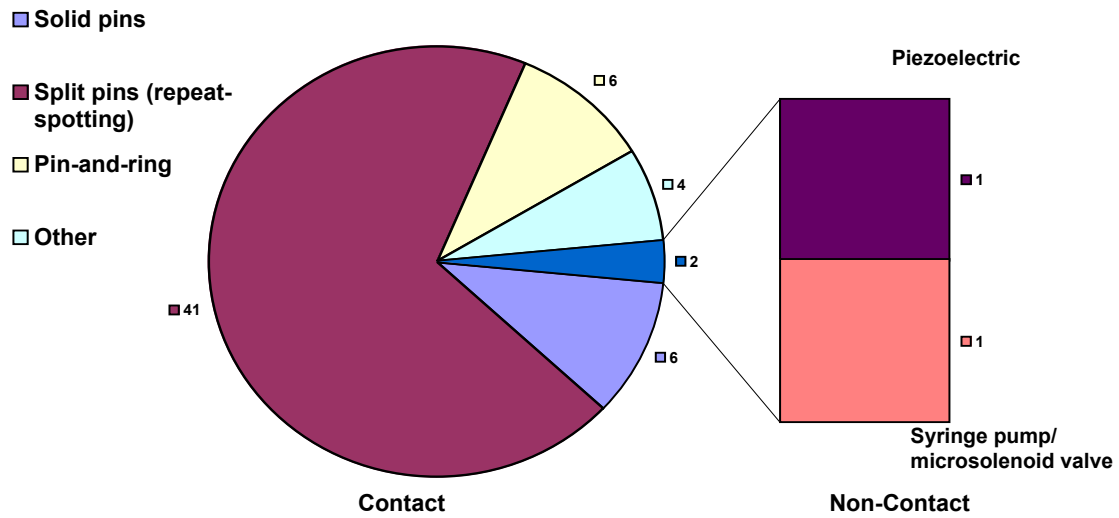


Figure 7. Pin types used by MA facilities (N=51)

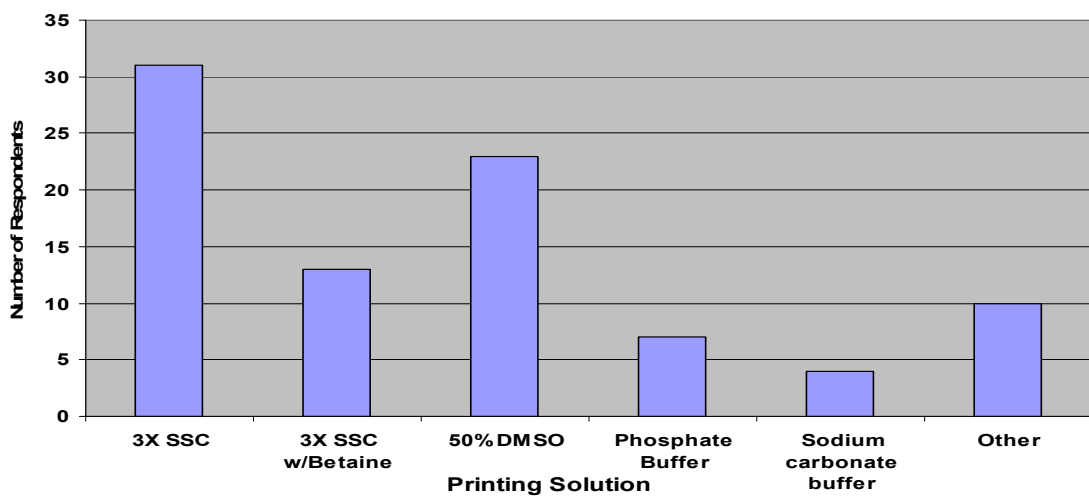


Figure 8. Types of printing solutions used by MA facilities. (N=66) 26% respondents report utilizing multiple spotting solutions

## **Affymetrix Array Section.**

### **•Affymetrix MA Facility Profile**

- 42 individuals responded to the Affymetrix Section**
- 54% were from facilities that have been operational for 2 or more years**
- 77% were from labs that were able to generate usable data in < 6 months**
- Over 95% of the facilities provide services for hybridization and scanning of arrays (Fig. 9)**
- A typical facility has one fluidics station (56%), one scanner (63%), and 2 or 3 computer work stations (62%)**
- Half the facilities routinely run test arrays (Fig. 10)**
- As shown in Fig. 11, there appears to be 3 distinct types of facilities--those that process over 100, those that process 30 to 60, and those that process less than 30 arrays/month.**
- 92% were at least moderately satisfied with the performance of the arrays and the quality of the data generated.**
- Users were the least satisfied with the array cost and documentation**

## **Affymetrix Array Section (cont).**

It appears as though most facilities were able to generate usable data very soon after setting up their Affymetrix system. As a result, most users seem to be very happy with their Affymetrix system. These findings agree with the previous 2001 survey. Despite decreases, arrays costs continue to be the primary complaint about the Affymetrix system. It is unclear from our survey if the negative sentiments to array costs are due to the actual price paid for the array or the fact that colleagues may be receiving a better price for his/her array.

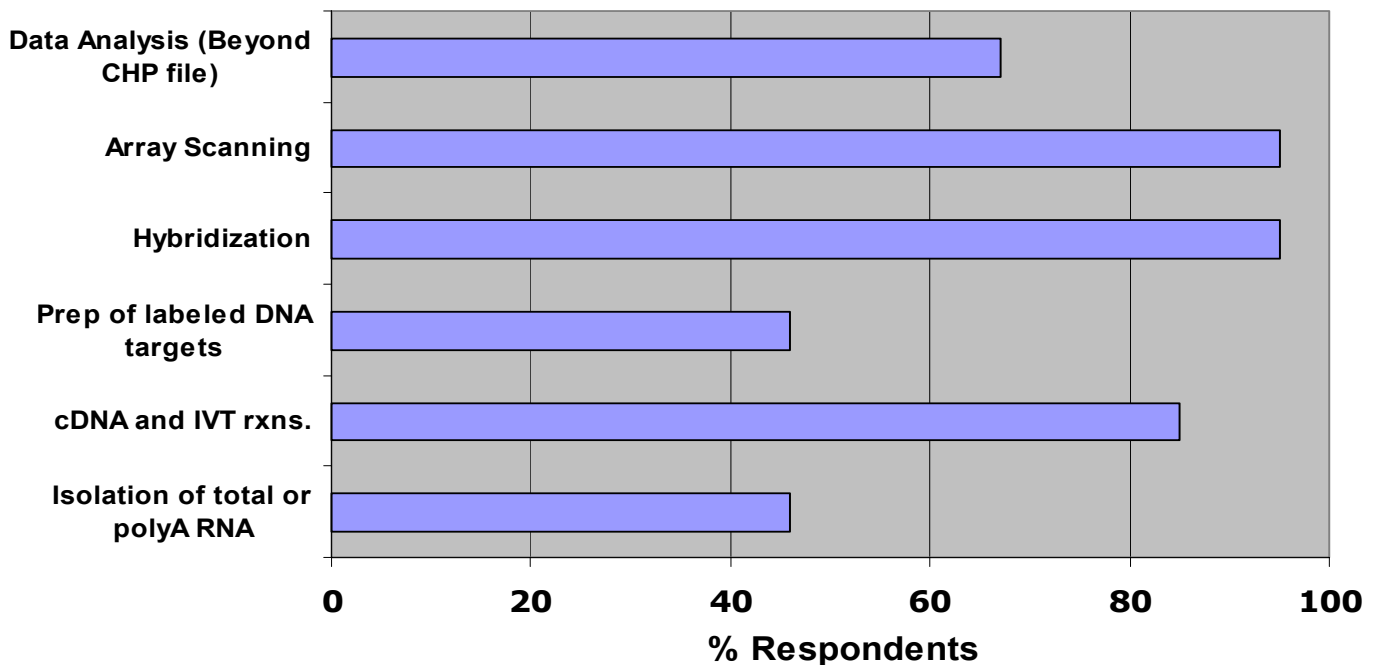


Figure 9. Services provided by Affymetrix MA facilities (N=41)

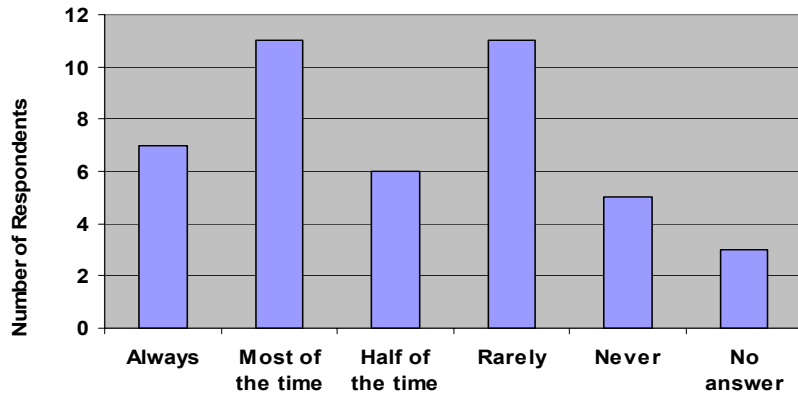


Figure 10. Frequency of use of test arrays.

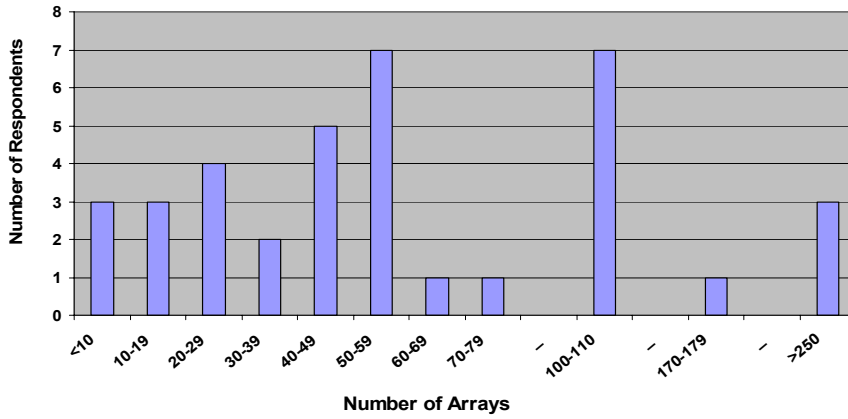


Figure 11. Number of arrays processed per month (N=36)

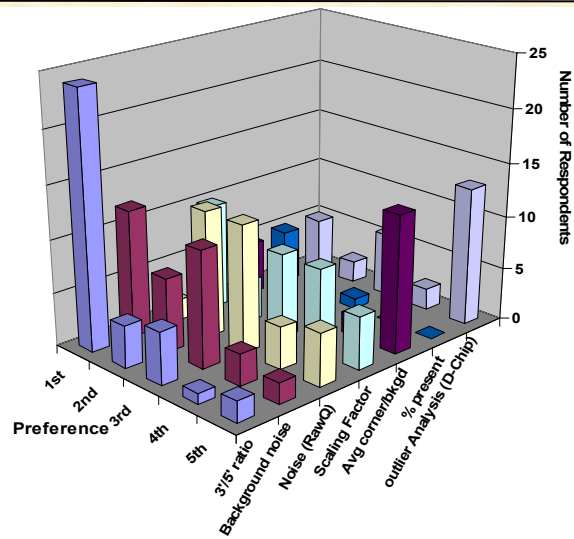


Figure 12. Methods used to evaluate the quality of Affymetrix array data.

## **Affymetrix Array Section (cont).**

The number of facilities that routinely use test array has decreased. The decrease in use of Test arrays may be due to higher user confidence in the system, a means of reducing costs, or that the cost of an expression array (which also contains the information placed on the test array) has significantly dropped.

### **•Sample preparation and analysis of arrays**

- As shown in Fig. 9, less than half of the facilities isolate RNA for their users**

- 79% use the Agilent Bioanalyzer and 94% of those users feel that the instrument is reliable in assessment of RNA/cRNA quality**

- 87% use the BioB, BioC, BioD and CreX hybridization controls, but only 32% use the controls to assess transcript levels**

- 62% do not use the *B. subtilis* spike control, however, those that use the control use it to assess sample quality**

- 43% of labs are using an amplification protocol**

- As shown in Fig. 12, the 3'/5' ratio appears to be the preferred method of evaluating whether data from the array is likely to be of good quality. Background, RawQ and %present also appear to important.**

- 87% are using global scaling to compare data between arrays.**

# FUTURE DIRECTIONS

Fig. 13 illustrates that custom array facilities will probably print more arrays containing oligonucleotides.

Respondents to both the custom and Affymetrix array sections indicate that they plan to use their arrays for SNP and genotyping studies.

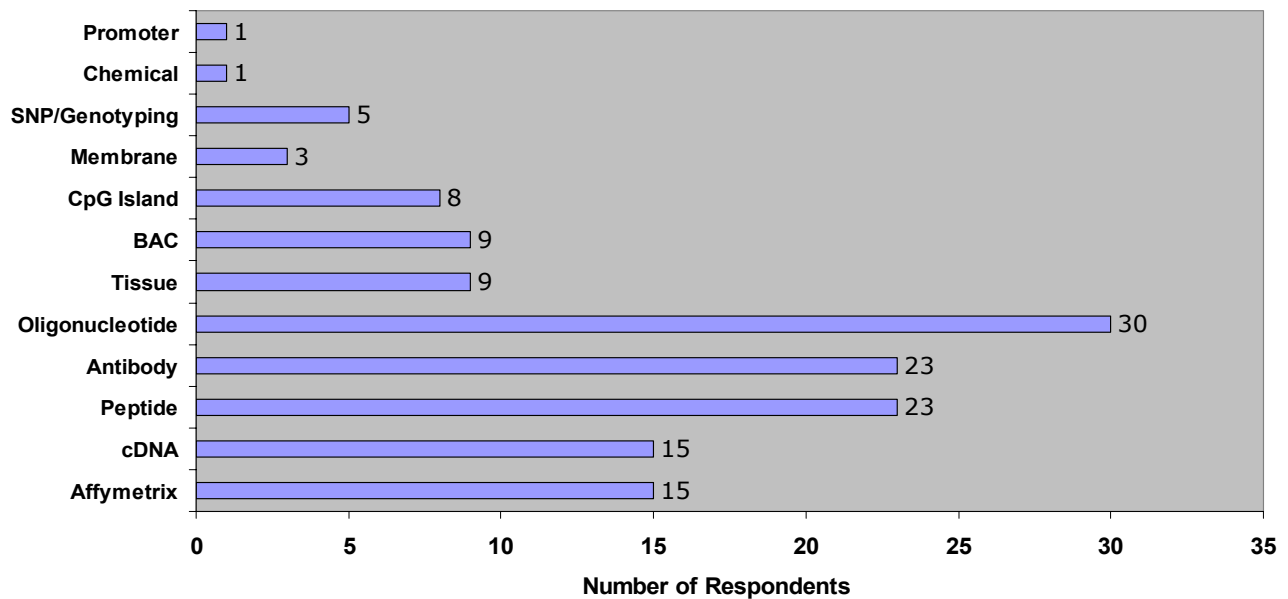


Figure 13. Types of arrays considered for future expansion