2008 MARG Microarray Survey

Sensing the state of microarray technology

Microarrays:

15 years of rapid development & expanding use

- Developed in early 90's
  - Pat Brown & colleagues: spotted arrays
  - Steve Fodor & Affymetrix: oligo arrays
- In-house spotted vs Affymetrix GeneChip arrays (1995-2000)
- Over last 6-7 years growth in other types and vendors of arrays: Agilent, Illumina, ABI, NimbleGen, …
- Applications: expression, DNA variation analysis, genome regulation

Goals of 2008 MARG Survey

- Describe current state of microarray use and MA facility operations
- Identify trends in use of new arrays, new applications, and new platforms
- Compare 2008 picture to results from previous MARG surveys
- Assess interest in new, high throughput technologies for genomics research

Survey development & data collection

Survey Design

- Collect information on instrumentation, protocols, staffing, arrays and throughput in a microarray facilities.
- Questions developed by members of MARG and hosted by SurveyMonkey.com.
- Responses collected by third-party who removed information identifying the participant prior to making data available to the MARG for analysis.

Survey Dissemination and Participation.

- Survey 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. Survey closed on Dec. 31, 2007.
- In addition to laboratories that offer microarray technologies as a shared resource, individual laboratories with these technologies were also invited to participate.
- Participation was open to anyone regardless of affiliation with ABRF.
- Participants had option of completing only sections of the survey that related to their microarray operation.

Survey analysis

- Data was analyzed by the MARG to build a current profile of microarray analysis laboratories.

Respondent demographics

	2008	2005	2003	2001	2000
#respondents	149	213	115	72	47
Location	74%*	67%*	74%*	73%*	NA
Inst type	77%^	75%^	73%^	55%^	30%^
Core facility	71%	76%	78%	ND	ND

USA/Canada
^academic

Conclusion: Can compare 2008 results to previous surveys

Survey sections

- General Section
- Platforms, Services and Throughput
- Commercial Arrays: types, user satisfaction, assay methods and QA/QC
- Bioinformatics: Data Analysis
- Bioinformatics: Data Management
- Next Generation Technologies
- Future directions

Facility Profile

	2008	2005	2003
Full service	**55/79%**	**76%**	
Use GLP	**21%**	**34%**	
Open access	**66%**	**64%**	**65%**
Education prgm	**61%**	**54%**	**57%**

In 2008, 74% of MA facilities had been in operation for over 4 years; only 6% less than 2 yrs.

Personnel

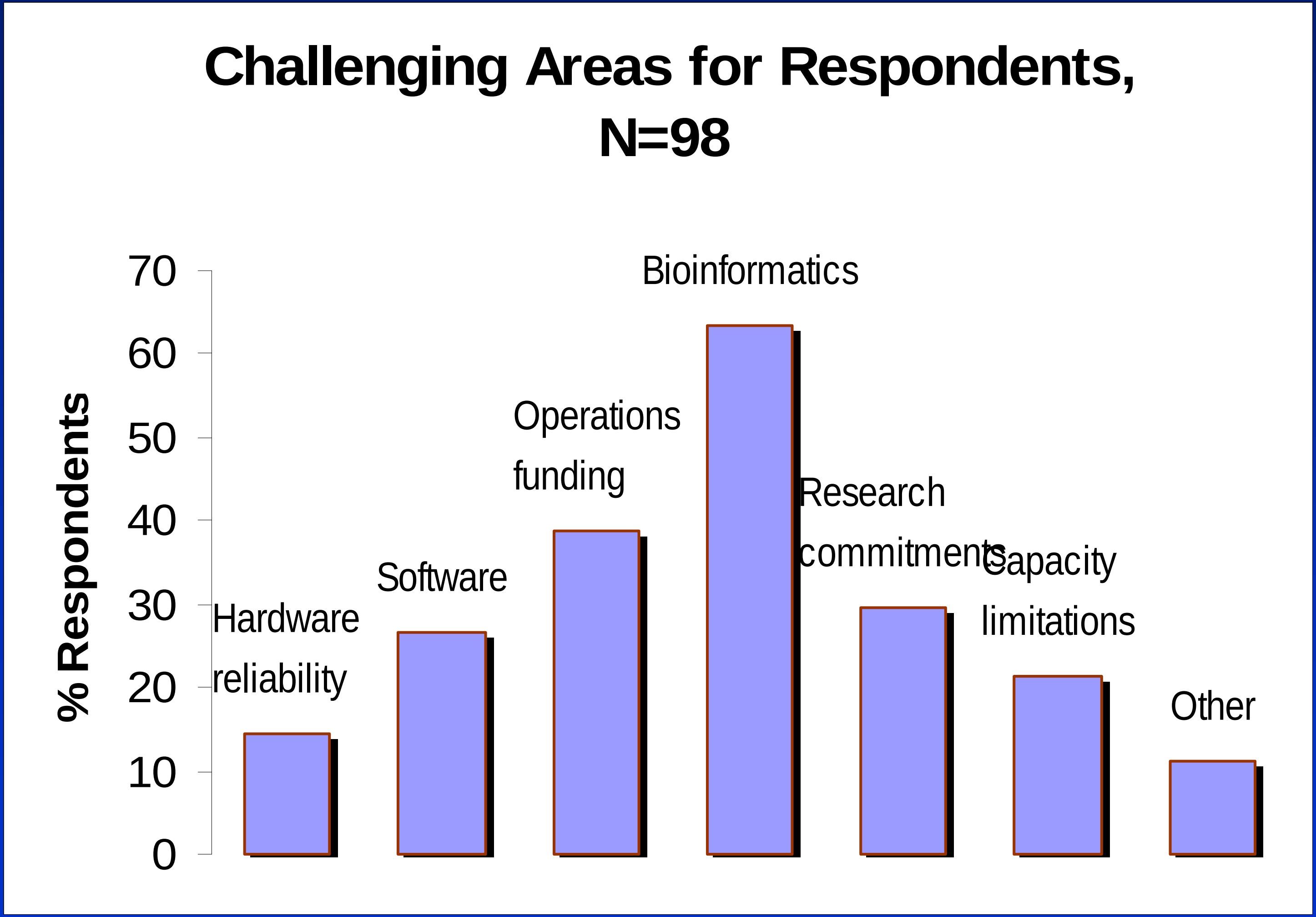
STAFF:	2001	2005	2008
#'s	N/A	<3.8>	<3.2>
Experience (yrs)	<1.4>	<3.5>	<4.2>
DIRECTOR:			
Experience (yrs)	<2.3>	<4.6>	<6.3>
Plan expansion	N/A	52%	32%

Note: 2008 #'s are estimates

Core Lab Services, N=96

- RNA isolation services offered by 50% of facilities; DNA isolation by 29%
- 85% of facilities provide sample labeling through data acquisition for expression assays; 64% label DNA for non-expression assays
- 78% provide limited data analysis; comprehensive data analysis available in 59% of facilities
- 54% of facilities provide support for data publication

Challenges

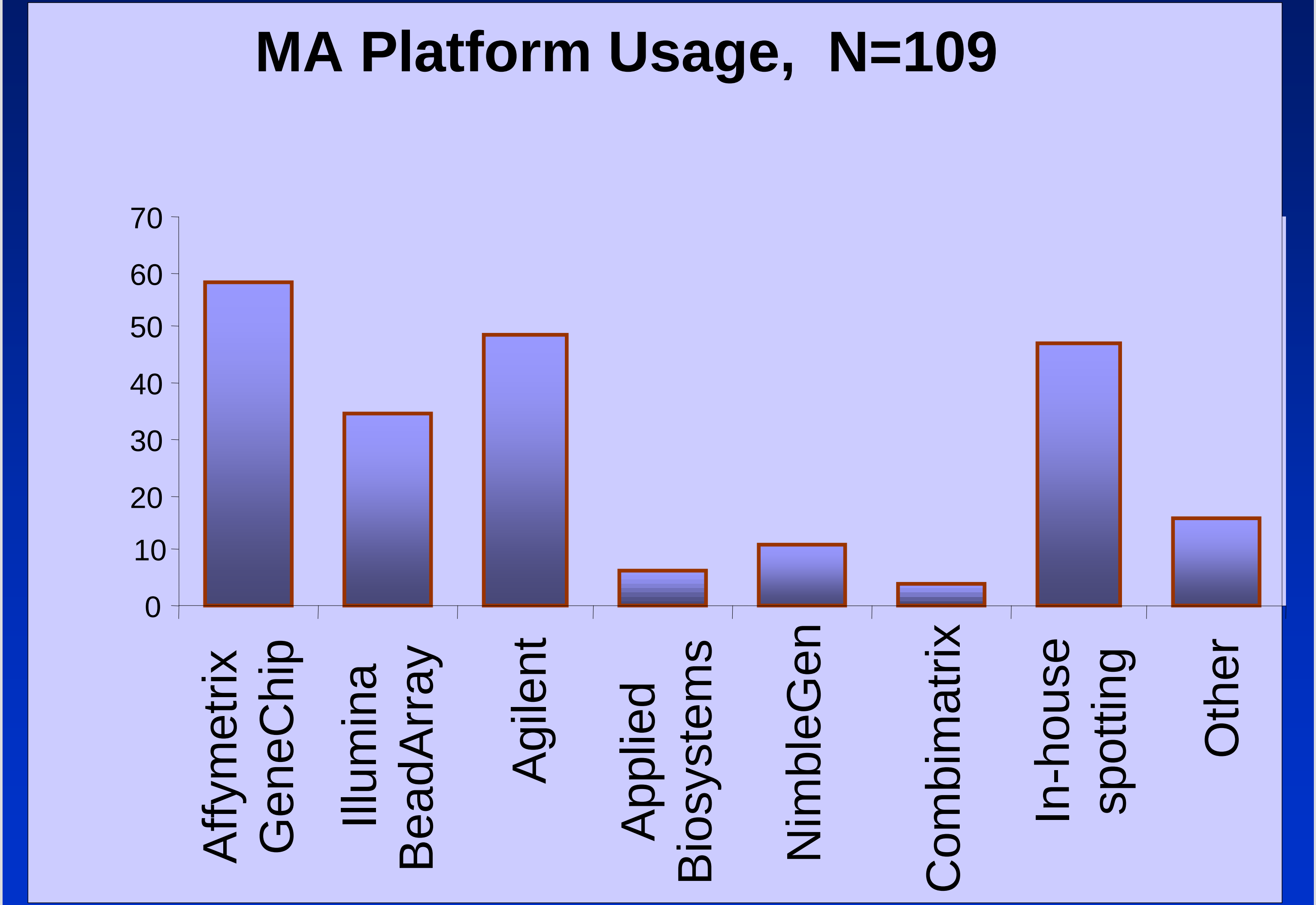

Challenging Areas for Respondents, N=98
% Respondents
70
60
50
40
30
20
10
0
Hardware reliability
Software
Operations funding
Bioinformatics
Research commitments
Capacity limitations
Other

Summary: Facility Profile 2008

- Most MA facilities are relatively mature operations: have used MA technology for more than 4 yrs; staff and directors very experienced; not planning to add staff.
- Most core labs are full service from sample labeling through data acquisition
- Bioinformatics is a challenging area for majority of MA facilities

Platforms & Array Throughput

Platform Usage



Automation

- Use robot for high throughput labeling

  19% YES
- Planning to buy robot in 2008

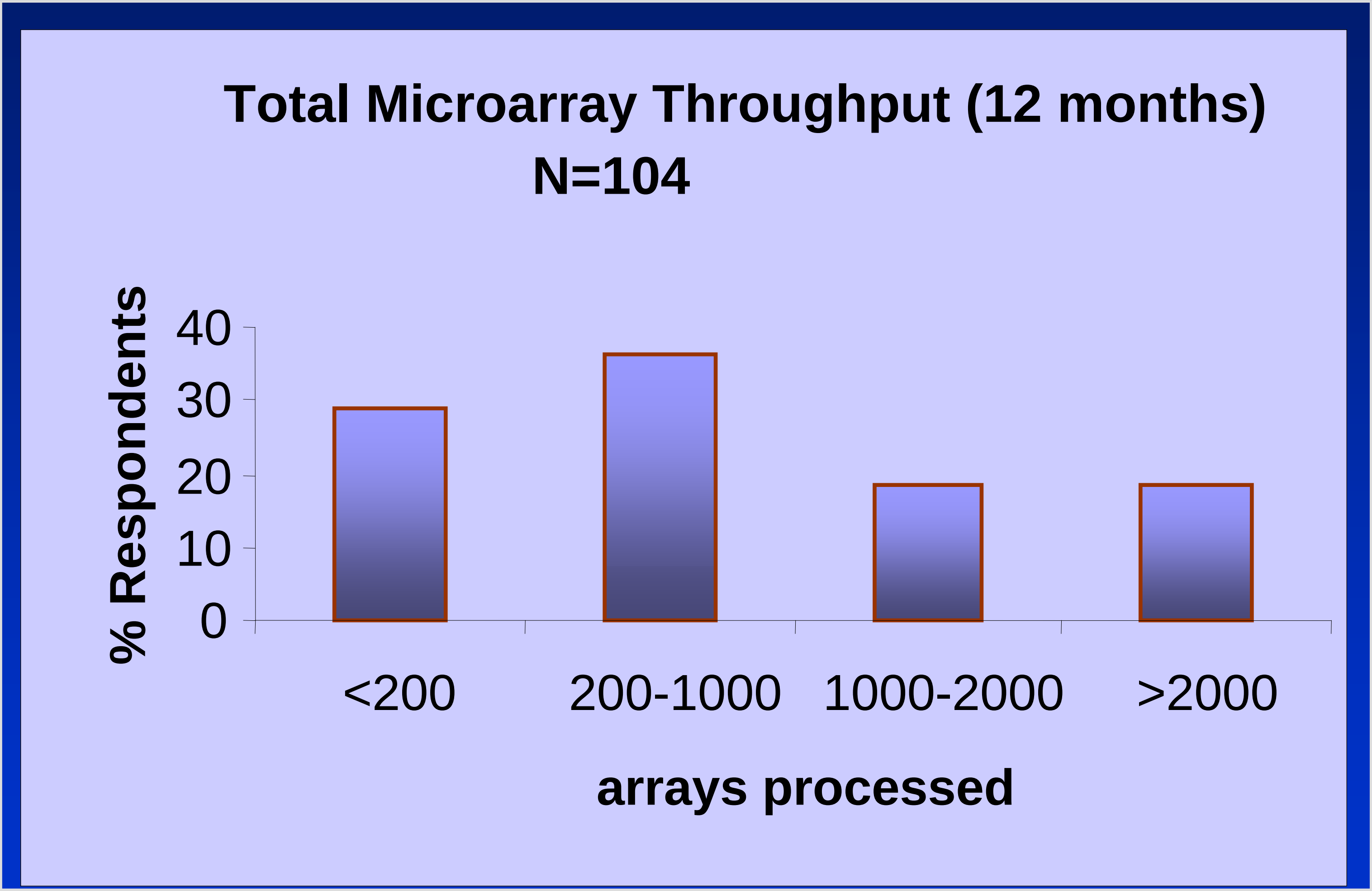
  12% YES

Array assay outsourcing

N=103

➢ 500 arrays/yr	2%
➢ 100 arrays/yr	6%
➢ 50 arrays/yr	3%
< 50 arrays/yr	12%
NONE	78%

Array Throughput



Throughput #'s may be underestimate if respondents assumed multi-array slides were counted as single array units

Array Projection 2008

Increase	52%
Stay the same	38%
Decrease	11%

Commercial Arrays

Commercial array usage

% arrays	% respondents
< 10%	9%
10-25	4
25-50	9
50-75	8
>75	69

Array types used: Affymetrix/Illumina

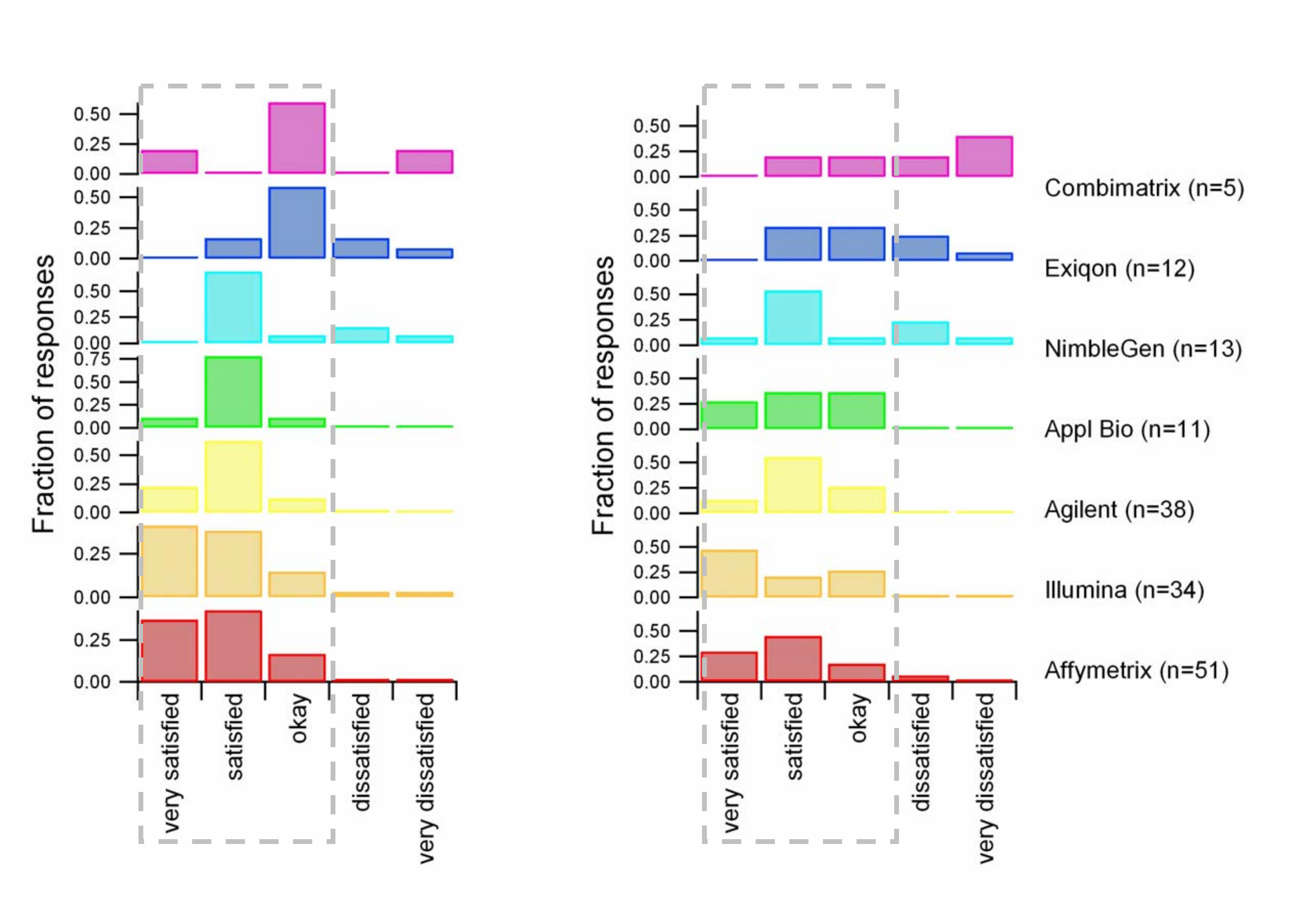
ARRAY TYPE	**AFFYMETRIX** (n=63)	**ILLUMINA** (n=37)
EXPRESSION:		
whole genome	98% (3'IVT)	86%
whole genome: exon	60	N/A
targeted expression	N/A	32
custom expression	32	8
GENOTYPING:		
whole genome: 10^6 + markers	29	27
whole genome: $<10^6$ markers	44	60
targeted	8	49
custom	16	35
OTHER:		
tiling	48	N/A
resequencing	22	N/A

Array preference: expression analysis

	Respondents
Affymetrix	48%
Illumina	20
Agilent	21
ABI	2
NimbleGen	1
Exiqon	1
In-house spotted	17
Other	4

Commercial arrays

Product satisfaction   Service satisfaction



Summary: Commercial arrays

- 77% of respondents use commercial arrays for most or all of their array throughput
- Whole genome expression profiling is most widely used array application on both Affymetrix and Illumina platforms
- Among respondents (n=90), Affymetrix arrays preferentially used for expression analysis in 48% of facilities, Agilent arrays 21%, Illumina arrays 20%, in-house spotted arrays 16%
- Most respondents very satisfied or satisfied with array products & service from Affymetrix, Illumina, Agilent, ABI, and NimbleGen

Expression Assay: Labeling Methods

Affymetrix/Illumina users

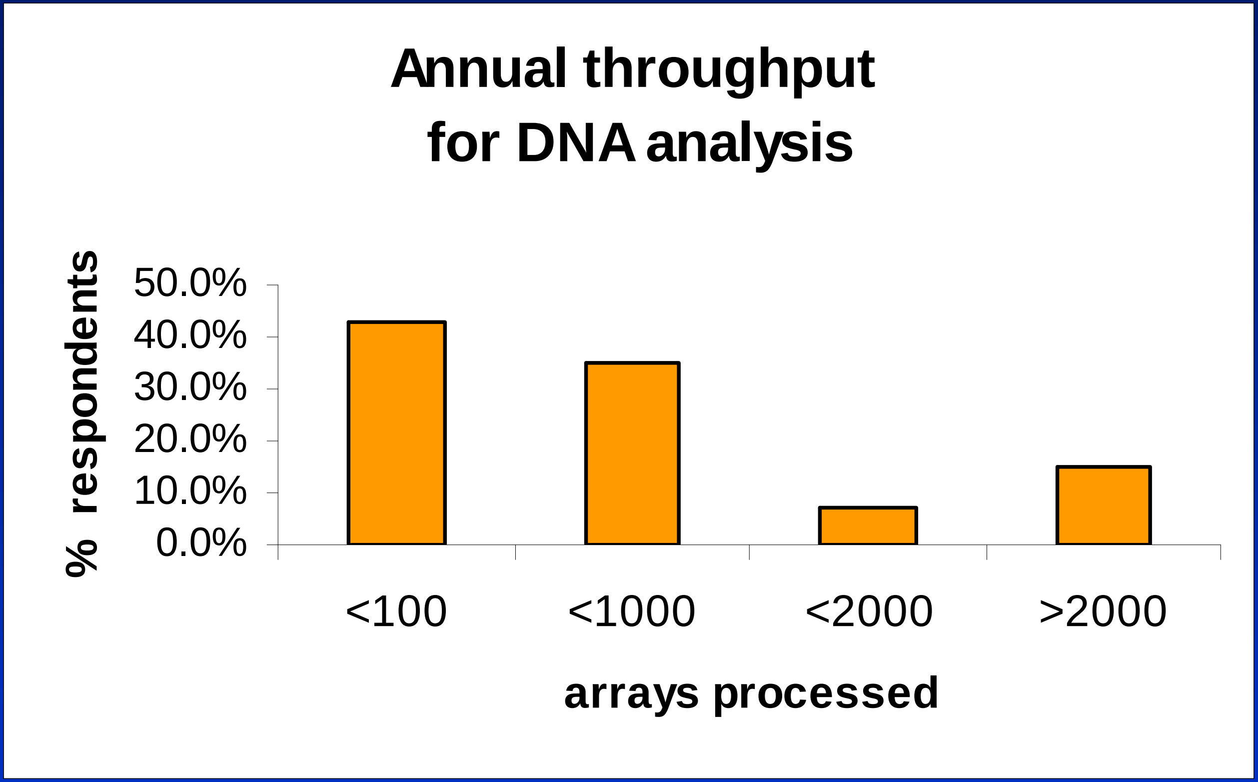
- Affymetrix one-cycle, two-cycle, and whole transcript, and Illumina Total Prep most widely-used labeling methods
- NuGEN Ovation used by 27% of respondents; Ambion one round by 21%; NuGEN WT Ovation by 17%
- Most frequently preferred:

1-5 ug n=64	Affy one-cycle/Ambion 1 round
0.01-1 ug	Illumina Total Prep/Affy two cycle/NuGEN Ovation
1–10 ng n=23	NuGEN Ovation or WT Ovation/Affy two cycle
<1 ng n= 14	NuGEN pico Ovation/Epicentre 2 round/Ambion 2 round

Expression assay: QA/QC

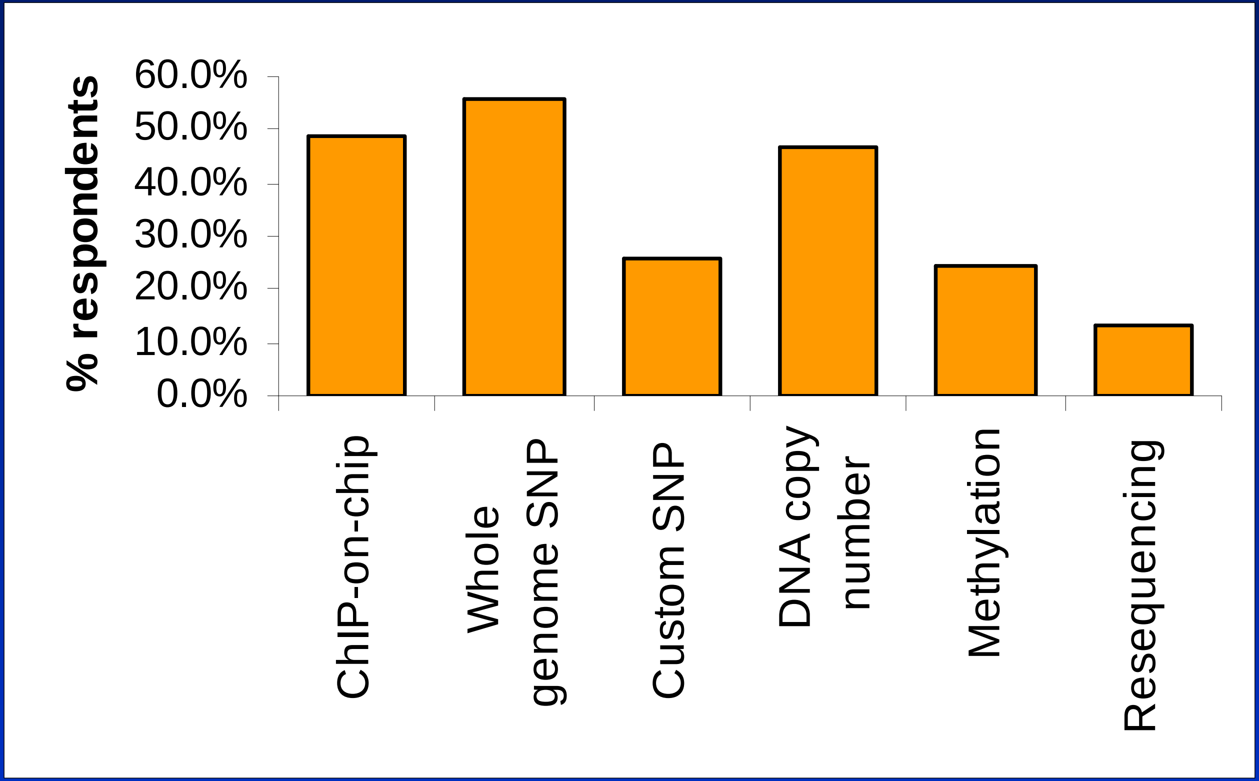
- 99% evaluate RNA quality prior to labeling
- 79% use OD260/OD280 ratio
- 92% use Bioanalyzer
- 19% use agarose gel
- Less than half (44%) find Bioanalyzer trace highly predictive of assay performance; 50% find it usually correlates

DNA analysis



68 of survey takers answered this question, suggesting about 60% of microarray facilities are currently running DNA applications

DNA analysis applications



ChIP-on-chip, whole genome SNP, and copy number analysis most widely used DNA assays

DNA analysis: QA/QC

- 73% evaluate DNA quality prior to labeling
- 90% use OD260/OD280 ratio
- 59% use agarose gel pattern
- 10% other (Bioanalyzer; OD260/OD230 ratio)

Bioinformatics:

- Data Analysis
- Data Management

BI Infrastructure

- Respondents to this section: 82/149
- Who analyzes the data?:
  - Individual PI - 74%
  - Statistician - 55%
  - Contractor - 23%
  - Other - 10%
- 45% of respondents have 1 or more FTEs dedicated to analysis

Tools for array data QA

N=67

Visual inspection of chip	83.6%
Pairwise scatterplots	70.1
Correlation coefficients	59.7
Within group similarity metrics	44.8
Principle component analysis	37.3
SimpleAffy (R Bioconductor)	19.4
Other (array metrics, R tools, external controls)	14.9

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

→ **Wide range of analysis tools being used with GeneSpring and Ingenuity having widest use.**

Data management: storage

- 48% store microarray data in database (n=75)
- No widely-used, highly rated DB system.
- Multiple respondents identified BASE, Array Track, custom systems, or Rosetta Resolver as satisfactory or better. (n=34)

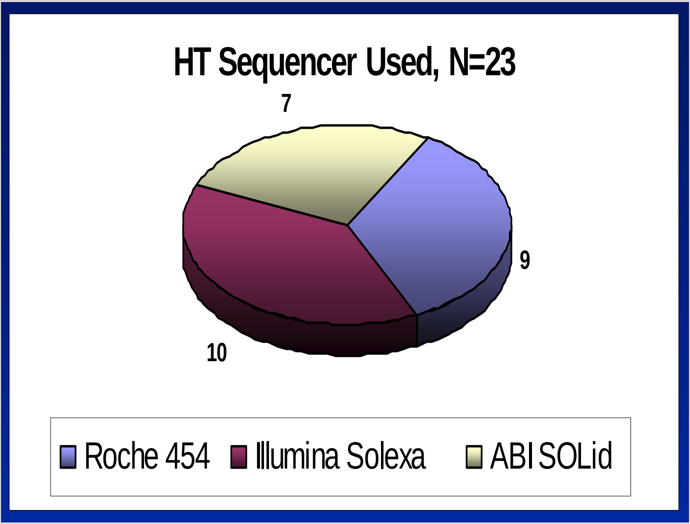
Data management: submission

- 13% always submit to MIAME-compliant public DB; 29%, sometimes; 39%, only when requested by publisher; 21%, Never
- 84% to GEO, 31% to Array Express
- Few find submission very easy; most rate it as relatively easy or not easy: GEO easier than Array Express

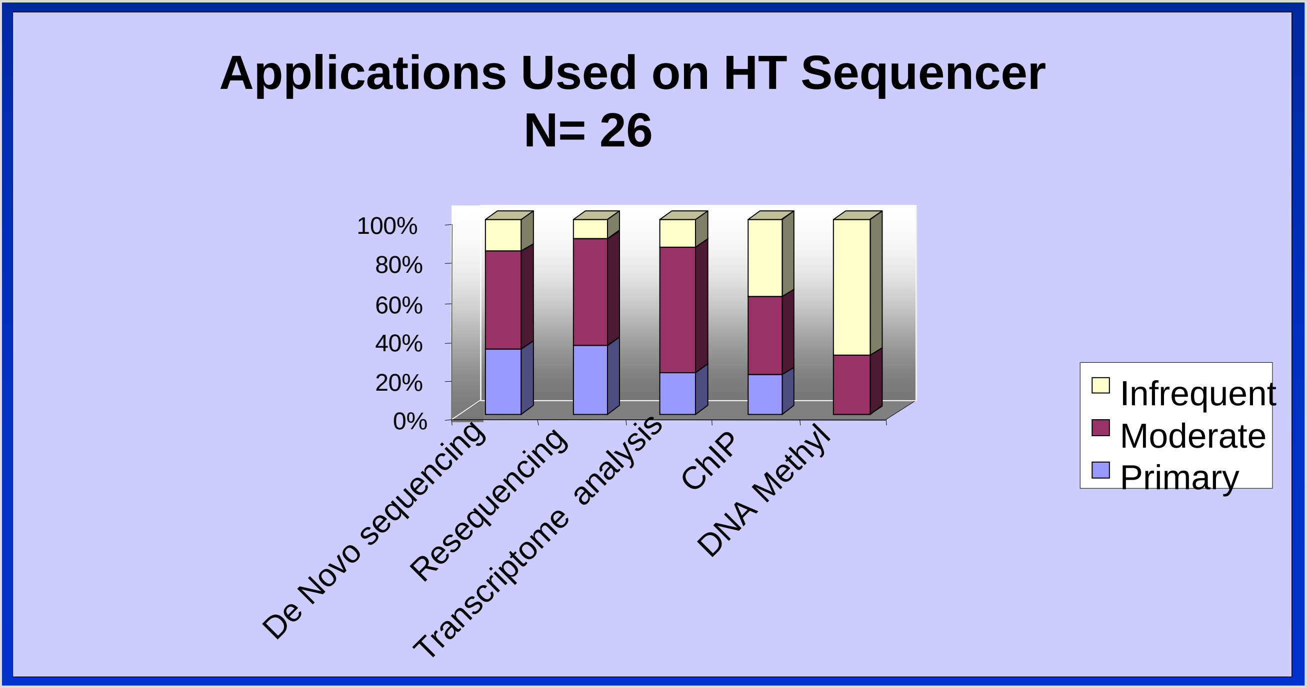
Next Generation Technologies: High Throughput DNA Sequencing

How are these new technologies impacting genomics research and microarrays?

HT DNA Sequencing



28% of respondents have high throughput (HT) DNA sequencer



Will HT sequencing displace DNA microarrays?

- In next 5years:
  - 9% – Yes, completely
  - 62% - Yes, partially
  - 17% - No
  - 12% - Not sure

Future Directions

- Splice variant arrays: 46/68 interested; 22/68 have run
- Array of arrays: 5/64 have run
- 3D surfaces: 10/57 interested

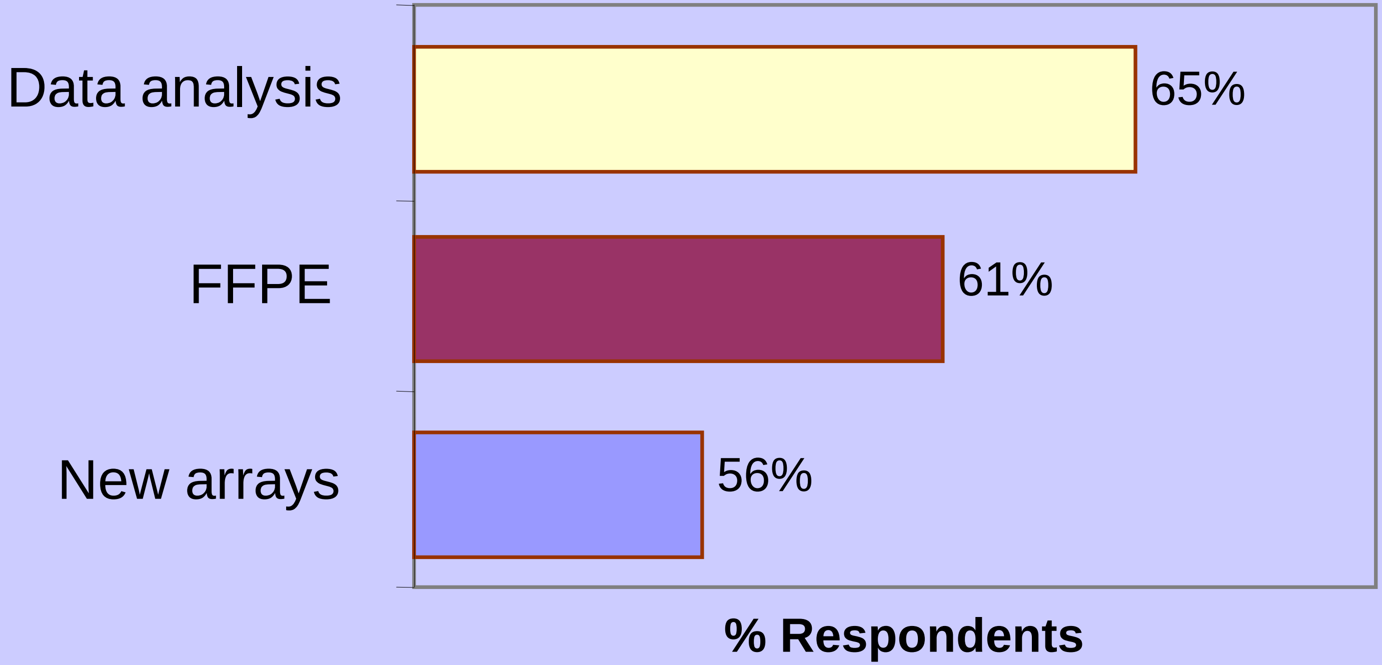
Future directions:
Clinical Applications

- 50% of respondents (n=60) interested in running arrays for clinical applications (experimental and FDA approved)
- 30% have investigated guidelines/regulations
- 10% familiar with NIST programs that address FDA requirements

Summary: 2008 MARG survey

1. Expression profiling still most frequently used array application.
2. Affymetrix system most widely-used platform, followed by Agilent and in-house spotted arrays. Illumina Bead Arrays close behind.
3. 91%of respondents expect array throughput to increase or stay the same in 2008.
4. As in previous MARG surveys, bioinformatics continues to be primary challenge for MA users and core facilities.
5. Only 9% of respondents expect HT sequencing to completely replace microarrays.
6. About half of respondents are interested in running arrays for clinical applications, same as in 2005, but only 30% have investigated guidelines

Suggested New MARG Projects



Acknowledgements

MARG members

Susan Hester, Chair
Christina Harrington
Nadereh Jafari
Steve Potter
Rick Jensen
Jay Tiesman
Herbert Auer
Laura Reid
Nancy Denslow

ABRF:

Support for Survey Monkey