

Summary of the results from the EMERALD web survey

Objectives

The goal of this survey was to gain insight into procedures, platforms, and needs of the microarray users community. The focus was both on commercial GeneChip arrays and all sources of cDNA and oligonucleotide microarrays. The survey was geared to gather information anonymously from academic, pharmaceutical, and commercial laboratories, which use microarray technologies routinely.

Results

How many and who attended the survey:

The number of persons who completed a survey was 56 and more than 90% of them are associated with an academic institution (or hospital or University medical center). Three fourth of the participants are post doctoral fellows, staff scientists or principal investigators. Some are lab directors, graduate students or research assistants. About 70% of the respondents are located in Europe (one third of these in UK), and 20% of them in North America. The remaining 10% are located in Asia, Australia and South America.

How do these people use microarray in their research:

Half of the participants study both DNA (SNP) and RNA (expression), the remaining half study only RNA. Forty percent of the participants just work with end results.

The majority (appr. 65%) of participants uses Affymetrix technology and custom (homemade) two color arrays (appr. 45%). About 32% use Agilent, 20% Illumina and only 4% Nimblegen. This indicates that most labs use more than one platform.

Forty percent of the participants rate their yearly sample throughput to be 10 – 100 samples, and about 30% runs 100 - 500 samples.

Seventy five percent of the participants use standard protocols recommended by the manufacturers.

About 80% of the participants perform the experiments by them selves or through an academic in house facility, and about 20% uses services offered by companies (like Nimblegen or LC Sciences etc.). Regarding the quality of the data obtained by core facilities or companies, about 75% are satisfied. However, 25% state that they are not satisfied.

How do you report microarray data:

Eighty five percent of the participants claim that they follow the MIAME recommendations. Sixty five percent upload their data to ArrayExpress, 45% to GEO and 10% to other repositories. Eighty percent find the repositories useful for gathering information from other experiments and about 75% are positive to add metadata to these repositories with the use of controlled vocabularies instead of free text. However, only 50% would supply metadata on e.g. pre-processing in MAGE-ML format.

Quality related questions:

Ninety five percent of the participants are generally satisfied with the quality of the data produce by themselves. However, 60% have had results that were not top-level.

Sixty five percent use some kind of QA/QC implemented in their core facility or lab, which includes e.g.: standardized protocols, only designated qualified technicians perform the microarray hybridization, RNA/DNA integrity/quantity tested by Bioanalyzer (e.g. RIN values obtained) or spectroscopy (NanoDrop), self-self hybridizations that should give no fold change, spike-in controls added, monitoring of ozone levels, checking of dye incorporation rates, box plot analysis of signal variance, spatial distribution etc.

About 65% would consider to use external standards (Spike in RNAs) for quality assessment whereas 20% would not. Fifty percent of the participants would consider doing weekly/monthly production pipeline checks. However, 60% think the price will be prohibitive.

Hybridization quality is assessed by: visual inspection of div plots (Bioconductor), software from the platform vendor (Affymetrix, Illumina), spike-in visualization, QC/QA program offered through the EMERALD project or other software like Partek GS and Agilent QC, dye-swap experiments, replicates, evaluation by eye and experience.

Ninety percent of the participants would consider using a quality metrics software tool to check quality, and 70% would be willing to use/change to another protocol or production line if these were recommended or if it were shown that these would perform better than others.

In general 65% of the participants prefer to use a QC system that offers both visual plots and absolute scores.

If there is a problem with the data the participants do the following: recheck RNA quality, cDNA yield, discard data and redo hybridization, use algorithms to “correct” for problem, ignore bad hybridizations,

Conclusions

In general, microarray users seem to be satisfied with the quality of the data they obtain, either by themselves or through a core facility. Most participants use commercial platforms like Affymetrix, Agilent and Illumina and run about 100 samples a year (this is not much, indicating that many scientists do their analysis by themselves instead of as a service in a core facility). However, still a quite large proportion of the arrays processed are homemade two-color arrays.

Almost all participants comply with MIAME and submit their data to a public repository and most are positive to reporting using a controlled vocabulary instead of free text.

For quality control, most of the participants seem to use regular Bioconductor packages and visual inspections of plots/graphs. But the participants are open to use other software if it is recommended or shown to be better. Some uses spike-in RNA as control and many are positive to that, but for some the added price may be prohibitive. When the quality is not good (which most report having seen occasionally), they try to correct for this or reject the sample. The most important factor giving bad results seems to be RNA quality.