



# EMERALD



## Enhancing microarray data quality

The European Union FP6 Coordination Action (CA) EMERALD aims to establish and disseminate quality metrics (QM), microarray standards and best laboratory practices throughout the European microarray community. This will allow microarray data production to take full advantage of QA/QC, thereby significantly enhancing the quality of microarray data and setting a precedent for other array-based technologies. Data quality and meta data (documentation) are key to all microarray data generation and analysis, to ensure that maximum information can be extracted from the data. The need to reanalyse and reproduce data spawned a 'grassroots movement', now the MGED Society, that established guidelines for experiment description (MIAME) and a structured data exchange model (MAGE-ML). MGED initiatives have been predominantly focused on data context, and its scope has only recently been extended to include data content. Quality and integrity of microarray data compendia (e.g. in ArrayExpress) are major determinants for information and extraction model building. High quality data will constitute one of the pillars of systems biology. This CA is designed to structure and amalgamate ongoing efforts across the European community, in close association with MGED and the ERCC.

Summary of the

### EMERALD QC web survey

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.

The main conclusions of the survey were:

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 use 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.

A more extensive summary of the result from the survey can be found at our web page: <http://www.microarray-quality.org/>

---

Sign up for future issues of the EMERALD Newsletter at our web page:  
[www.microarray-quality.org](http://www.microarray-quality.org)

---

Relevant conference

### MGED12

The 12th International MGED Meeting, focusing on translational genomics and high throughput sequencing, will be held October 5 - 8, 2009, Phoenix, Arizona, USA. More information about the conference can be found at: <https://www.tgen.org/mged/index.cfm>

EMERALD will offer travel bursaries for PhD students from the European Union. They will be selected based on submitted abstracts having relevance to quality control and standardisation of microarray data (or high throughput sequencing) technology.

---

New on the web page

### Discussion forum

Through our webpage (<http://www.microarray-quality.org/>) a new discussion forum, related to the aims of the project, is now open. Categories are related to quality metrics, ontology development, external standards, best laboratory practices and new technologies.

The forum can be found at this address: <http://www.microarray-quality.org/phpBB3/index.php>