



Sixth Framework Programme for Quality of Life and  
Management of Living Resources

Project no. LSHG-CT-2006-037686

# EMERALD

Empowering the Microarray-Based  
European Research Area to Take a Lead in  
Development and Exploitation

EU Deliverable: D3.2

Due Date: 3<sup>rd</sup> to 7<sup>th</sup> of September 2007

Delivery Date: ??th October 2007

Version 1

Partner responsible: NTNU and VIB

## **Minutes of EMERALD associated activities at MGED10 3<sup>rd</sup>-7<sup>th</sup> September, 2007, Brisbane Australia**

### **Introduction:**

Microarray and Gene Expression Data society (MGED) is the main world wide organization working with development of microarray technology, standardization, quality control and data analysis of microarray data. Annually MGED organizes an international meeting gathering the most experienced researchers from the whole world. This year MGED10 took place in Brisbane Australia. Approximately 170 people attended the conference (approximately 1/3 US, 1/3 Asian and 1/3 European).

The meeting covered seven different topics with keynote, plenary lectures, shorter talks selected from submitted abstracts, four workshops and a MAGE Jamboree.

According to our project plan, and also taking into account the advise of the proposal reviewers EMERALD should team up with MGED, schedule meetings and promote the results of our work to the European microarray society. Even though MGED10 this year was hosted in Australia, we felt that our attendance was essential and indeed very useful for the project. First of all, by participating on this meeting we were able to get in close contact with MGED by informing the organization about our project effort and establishing an official connection.

Furthermore, EMERALD hosted a plenary session (EMERALD MICROARRAY DATA STANDARDS AND QUALITY CONTROL SYMPOSIUM) where our project was presented to the audience. Martin Kuiper presented the project objectives and status. He gave examples of how we will disseminate the results of the project through a series of workshops in relation to European meetings. Further, we had invited and given travel support for speakers to give plenary lectures. Cathy Ball talked about MGEDs work to establish microarray standards (MIAME), Mark Salit from the external controls consortium (ERCC) talked about how to use of RNA titration to evaluate microarray experiments and Federico Goodsaid from the US Food and Drug Administration (FDA) talked about FDAs mission with respect to microarray standards and what they do in relation to how microarray experiments and data have to be performed to be accepted by FDA. These projects are all located in the US and EMERALD has now through this symposium established a connection which may constitute a good platform for collaboration when similar initiatives are started in Europe as a part of the EMERALD effort.

In addition to the oral presentations we presented the project and disseminated results by a poster (see attachment 1) where we specifically presented some results from WP1 focusing on quality metrics and the development of additional MGED ontology (people responsible: Wolfgang Huber, Audrey Kauffmann, Helen Parkinson and James Malone all EBI). We also distributed a leaflet describing the EMERALD project, including all contact information for EMERALD (see attachment 2). Finally, EMERALD was responsible for two workshops. At Workshop A, Audrey Kauffmann presented a test version of the EMERALD Quality diagnostics (Bioconductor) program and the participants were able to try a test version on some real data. At Workshop C, Helen Parkinson, Anna Farne, and Alvis Brazma, provided a practical session on making efficient use of ArrayExpress and the participants were able to try how to download large datasets form AE which can be important in relation to QM. Furthermore an introduction to Expression profiler was given and the participants were able to use this program on real datasets.

All together EMERALD was very visible at the conference and we were able to communicate our mission and results of our work to the microarray community. An obvious drawback was that this meeting took place in Australia with less European attendance than if the meeting has been located in Europe. However, we feel this was the

right time and opportunity to present the project and our results to the microarray community. Moreover, to help European students afford to go to this meeting in Australia EMERALD offered two travel bursaries. Our participation was a success in relation to how much attention we had, and even with the somewhat lower European attendance this year many of the most important representatives from the European microarray community were present. The remaining workshops associated with EMERALD will be hosted in Europe and we will again be present at MGED11 in Trento, Italy in 2009.

## **1. Short review of the program (3<sup>rd</sup>-7<sup>th</sup> September, 2007):**

### **Main topics included:**

#### **1. GENOMIC CONTROL OF GENE EXPRESSION**

Keynote Lecture: John Mattick: The mammalian genome as an RNA machine.

Plenary Lecture: Jason Lieb: Genome-wide identification of active regulatory elements in the human genome by FAIRE.

Plenary Lecture: David Johnson: The comprehensive annotation of the functional regulatory elements of the human genome.

Plenary Lecture: M. K. Raghuraman: Genome wide studies of origin activation and its regulation.

Plenary Lecture: Terry Speed: A Method for the Detection of Alternative Splicing from Exon Array Data.

#### **2. ANALYTICAL APPROACHES FOR GENOME-WIDE DATA**

Shorter talks.

#### **3. CANCER GENOMICS**

Keynote Lecture: David Bowtell: Molecular subtype analysis of ovarian cancer.

Keynote Lecture: Richard Wooster: Genome wide analysis of somatic mutations in cancer

#### **4. SEQUENCE BASED ANALYSIS OF THE TRANSCRIPTOME**

Plenary Lecture: Piero Carninci: Sequence based surveys of transcriptional complexity and regulation of the mammalian transcriptome.

Plenary Lecture: Ruan Yijun: Next generation sequencing approaches for studying the mammalian genome and transcriptome

Plenary Lecture: Alistair Forrest: SOLiD Sequencing of the mammalian transcriptome

#### **5. FUNCTIONAL GENOMICS**

Plenary Lecture: John Quackenbush: Network Approaches to Systems Biology.

#### **6. BIOINFORMATIC TOOLS FOR ANALYSING AND MANAGING DATA**

Shorter talks.

#### **7. EMERALD MICROARRAY DATA STANDARDS AND QUALITY CONTROL SYMPOSIUM**

Main topics:

Cathy Ball (11.30-11.45), Discussion on future of Microarray Data Standards

Martin Kuiper (11.45-11.50), EMERALD QC Initiative

Mark Salit (11.50-12.20), External controls consortium (ERCC)

Federico Goodsaid (14.00-14.30), Microarray experiments and the FDA.

### **Workshops:**

Workshop A: Audrey Kauffmann, Quality diagnostics with Bioconductor.

Workshop B: Ronald Taylor, Network inference and analysis using SEBINI-CABIN.

Workshop C: Helen Parkinson, Misha Kapushesky, and Alvis Brazma, Making efficient use of ArrayExpress.

Workshop D: John Quackenbush and Roger Bumgarner, Statistical and Computational Approaches to Extracting Meaning from 'omics Data using MeV.

## 2. Minutes from the EMERALD MICROARRAY DATA STANDARDS AND QUALITY CONTROL SYMPOSIUM, 5<sup>th</sup> September 2007.

**EMERALD session** (chaired by Alvis Brazma):

- A. Session opened by Cathy Ball describing how MGED has put a focus on microarray standardization, how it has evolved during the last years and how MGED thinks it will evolve in the future.  
MGED has been the major player in the development of MIAME standards which must be considered a success because most major journals now require microarray experiments to be described according to the MIAME standards and deposited in a public repository like ArrayExpress and GEO to be accepted. Furthermore, MGED is heavily involved in developing ontologies for describing biology (e.g. diseases), experiments (including microarray experiments) and data analysis of -omic data. MGED collaborates with the US initiated MAQC and ERCC consortia and Ball expressed that MGED is very happy that Europe now has initiated a project that will join in this common effort of increasing the focus on standards and standardization of microarray technology and data analysis.
- B. Martin Kuiper gave a short introduction of the aims of the EMERALD project.  
Kuiper presented the project's different work packages and focused on how the project will disseminate its work and results through a series of workshops in Europe in the years to come. Kuiper also referred to the poster and the EMERALD leaflet we distributed at the conference for more information about the project and contact information.
- C. Mark Salit talked about their external control consortium (ERCC) and how they have used RNA titrations to validate Microarray Gene Expression Data. Salit mainly referred to the work they recently presented in the MAQC issue of Nature 2006.  
Salit also contributes to the EMERALD project by participating in our scientific advisory board (SAB).
- D. Federico Goodsaid explained the FDA view on microarray standards and what they do in relation to how microarray experiments and data have to be performed to be accepted by FDA.  
At the moment we do not have an official organization like the FDA in Europe. However, we think their expertise will be very important in Europe as well, and are very satisfied that a connection has been established. The FDA may prove to be an important collaborator especially with respect to our work package 2 (Standards, LGC) and work package 5 (Standards and European legislation, IRMM).



### **Selected abstracts for short talks during the EMERALD session:**

Two submitted abstracts were chosen by the MGED10 organizers for short oral presentation during our EMERALD session, on the basis of their relevance to the subject of our session (microarray standards and quality control).

- A. Aedin Culhane, Dana-Farber Cancer Institute/Harvard School of Public Health, US  
Title: "Biotic, an automated approach to evaluation and exclusion of poor quality gene expression data".

Culhane et al. propose a new approach to quality assessment of microarray data in which both biological variance and technical quality control are considered. BioTecQC is an intuitive approach which provides an automated way of excluding poor quality array data. It generates a ranked list of QC scores for arrays and then iteratively tests and excludes arrays with highly ranked poor QC scores if they are over-represented in the set of arrays which are in the outlying cluster of a hierarchical cluster analysis of the normalized data. Chulane et al. applied BioTecQC to 120 Affymetrix arrays in the MAQC study, and to breast cancer data in which they selectively disrupt biologically important and random probes. By BioTecQC they were able to successfully exclude poor quality arrays which contribute little to biological variance. Results of BioTecQC program are presented in a user-friendly html report with plots of excluded and retained arrays. Source code is available in R and Audrey Kauffmann et al. will consider to implement this into a future version of the EMERALD Quality diagnostics program.

- B. Laurent Gatto, DNAVision, Belgium  
Title: "How to objectively and efficiently assess the inter-lab quality of microarray data".

Gatto et al. have used the recently published MAQC reference data sets for the Affymetrix platform to assess quality metrics in terms of variability between and within laboratories. After two different normalizations (RMA and GCRMA), they performed intra- and inter-site analyses of the four reference RNA data sets using two independent procedures (parametric and non-parametric) to identify probe sets that show significant differences in expression. The 24 intra-site comparisons were used as controls to define the probe sets that can be effectively considered as differentially expressed. The results by Gatto et al. showed a good correlation for the intra-site analysis, in concordance with the results found by the MAQC. Then they used this list to infer how efficiently inter-site comparisons can recover the correct results. Surprisingly, they observed notable discrepancies for some inter-site comparisons although nearly all the data passed the Affymetrix quality requirements. Finally, Gatto et al. correlated their results to metrics that are widely used to assess the quality of the raw data. This approach enabled them to gain reliable insights in how to objectively assess the importance of various quality metrics as well as how to proceed to establish reliable thresholds in large inter-site projects.

### **3. Minutes from EMERALD associated Workshops, 6<sup>th</sup> and 7<sup>th</sup> September 2007 at Griffith University, Brisbane, Australia**

Two workshops were associated with EMERALD work/people. Approximately 30-35 people attended the workshops.

#### **Workshop A Quality diagnostics with Bioconductor**

Audrey Kauffmann presented a first version of the EMERALD Quality diagnostics (Bioconductor) program and the participants were able to try a test version on some real data. Some improvements were discussed e.g. to implement some of the tests previously presented at the meeting by Aedin Culhane. The workshop participants found the program potentially very useful, and look forward to the official release.

This program will first of all be an important help and tool for core facilities and research groups using microarray technology to evaluate their results. Today the availability of such programs is limited. Another use of such a program that has been discussed is to offer this as a tool for journal referees who today often have problems to evaluate the quality of microarray data.

Such systems for quality control of microarray data is one of EMERALD's main focus areas and this workshop was only our first attempt to introduce this program to the microarray community. Today the program handles two-colour arrays and Affymetrix data, however it will be possible to import data from other platforms like Illumina in near future. To disseminate a first official version of the EMERALD Quality diagnostics program to the European community a new workshop associated with MGED11 in Trento Italy in September 2008 is already planned and the program may be also be presented at other European meetings/workshops in late 2007 and in 2008.

#### **Workshop C, Making efficient use of ArrayExpress**

Helen Parkinson, Anna Farne, and Alvis Brazma provided a practical session on making efficient use of ArrayExpress (AE) and the participants were able to try how to download large datasets from AE which can be important in relation to QM. Further an introduction to Expression profiler was given and the participants were able to use with this program on real datasets.

ArrayExpress (EBI, UK) is today, together with GEO (US), the most important repository for microarray data, not just a repository but also with its possibilities for meta data analysis of submitted datasets through Expression profiler. EMERALD collaborates with EBI and ArrayExpress in two ways. First the EMERALD Quality diagnostics program may be offered as a tool to evaluate data submitted to AE. In addition our WP1 work with ontologies will be important in standardization of how e.g. microarray experiments are described. A workshop on this ontology work will be organized by EMERALD and hosted at EBI in November 2007.



#### **4. EMERALD travel bursary for MGED10**

Because MGED10 this year was held in Australia, EMERALD announced three travel bursaries (for up to Euro 1500) for PhD students from the European Union member countries. Our intention was to provide support for selected students to participate despite the increased expenses compared to if the meeting would have been held in Europe.

Only two applications for travel bursaries were received from applicants who fulfilled our requested requirements. Travel bursaries were awarded to Adam Kundewicz, Department of Biochemistry, University of Geneva, Switzerland who gave a talk titled "Linkage of neural specification and bHLH regulated transcription by a "ChIP-on-chip" strategy". The second bursary was given to Klemens Vierlinger, Austrian Research Centers GmbH - ARC, Division of Life Sciences, Seibersdorf, Austria, who gave a talk titled "SERPINA, for thyroid tumor papillary carcinoma classification".

#### **5. EMERALD consortium meeting**

A consortium meeting was held 5<sup>th</sup> September at the MGED10 venue. Five project partners were present including Martin Kuiper (VIB), Alvis Brazma (EBI), Arne K. Sandvik (NTNU), Carole Foy (LGC), Joaquin Dopazo (CIPF), Helen Parkinson (EBI), Vidar Beisvåg (NTNU), Audrey Kauffmann (EBI) and James Malone (EBI). In addition Mark Salit (NIST) (member of EMERALD SAB) participated at the meeting.

The program for the meeting included discussion of the current status of the project, a workshop overview and discussion on how to organize the future workshops, and other outreach activities.

In particular the participants discussed how the EMERALD project best can advertise its activities, and how to improve the website with current information. The status of the deliverables was assessed. Detailed minutes from these discussions can be found in Attachment 3.

The final conclusion of our consortium meeting was that the project is doing ok. Especially the development of quality metrics is well on track. Because of this we were already able to organize a first QM tutorial at this meeting.

#### **6. Summary**

EMERALDs participation to MGED10 was intended to fulfill part of the planned objectives for WS1: bringing together the microarray community and inform it about the scope of the project. The EMERALD session was well attended, and it was repeatedly stated that the objectives of MGED and EMERALD are very much aligned. EMERALD and EBI had further visibility through the workshops related to the EMERALD Quality diagnostics program and the use of ArrayExpress. Our EMERALD project poster was visited by many conference attendees, and our information leaflet was widely distributed. Very importantly, we were able to establish closer relations and specific collaboration with MGED society, ERCC, FDA and the MAQC project representatives. During the conference a consortium meeting was held. Five project partners were present and further project activities were discussed. Overall, this meeting was very useful for our project.

## Attachment 1: EMERALD poster presented at MGED10, Brisbane Australia, 2007



### Project objectives

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 predominantly been focused on data context, and its scope has only recently been extended to included 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 Europe community, in close association with MGED and the ERCC.

### Coordination and dissemination activities

Coordination activities are defined in six main areas relevant for microarray analysis: Development of quality metrics, ontology for data description, implementation of standards and best practices, selection of standards that are candidates for European Reference Materials, impact of QA/QC on data information content, and dissemination of QA/QC principles to novel experimental high-throughput techniques for the different -omics domains. These activities are made up of six work packages (WP).

**WP1: Quality Metrics and Ontologies (EBI).** The objective of this WP is to develop and disseminate quality metrics and tools for determining data quality and communicating data transformations. As part of the MGED ontology, a normalisation and transformation ontology (NTO) is being developed to describe data transformations (Figure 1). We are also developing a new Bioconductor package, named arrayQualityMetrics, that provides a HTML report with diagnostic plots for one or dual color microarray data (Figure 2-4). The quality report contains the evaluation of different categories of quality metrics. The individual array quality is controlled by M versus A plots. The existence of spatial effects is checked by image representations of the arrays. Scatter plots are used to assess the reproducibility of the experiments. Boxplots and density plots allows the control of the homogeneity between the experiments. The report also contains a study of the GC content effects and the mapping of the reporters to test the array platform quality. A heatmap representing the distance between the samples allows the evaluation of the biological signal to noise ratio. In the case of Affymetrix experiments, some quality controls usually used for this platform are added to the report, such as Relative Log Expression (RLE) or Normalized Unscaled Standard Error (NUSE) plots for instance. This report can be used as a first step of the microarray analysis or to compare the efficiency of different methods of normalization. The quality metrics report will also be useful to assess the quality of public data in the context of meta-analysis for instance. **People responsible: Wolfgang Huber, Audrey Kauffmann, Helen Parkinson and James Malone (EBI).**

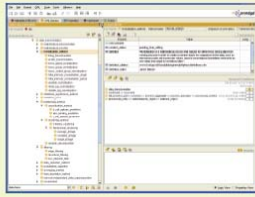


Figure 1. A Normalisation and Transformation Ontology (NTO).

As part of the MGED ontology, a normalisation and transformation ontology is being developed to describe data transformations. The ontology will cover aspects of microarray data such as normalisation techniques, quality metrics and quality control and data transformation. The development of this ontology will employ several strategies that will be the subject of workshop group discussion, and it will include analysis of current vocabularies and text mining of relevant literature.

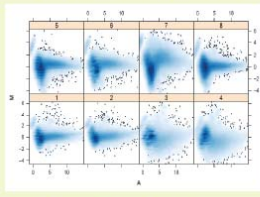


Figure 2. Represents MA-plot for each array.

MA-plots are useful for pairwise comparisons between arrays. Rather than comparing each array to every other array, here we compare each array to a single median "pseudo"-array. Typically, we expect the mass of the distribution in an MA-plot to be concentrated along the  $M = 0$  axis, and there should be no trend in the mean of M as a function of A. Note that a bigger width of the plot of the M-distribution at the lower end of the A scale does not necessarily imply that the variance of the M-distribution is larger at the lower end of the A scale. To visualize whether there is a trend in the variance of M as a function of A, consider plotting M versus rank(A).

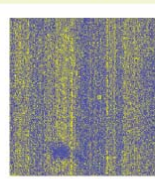


Figure 3. Intensity representation on the array (spatial plots).

False color representations of the spatial intensity distributions of each arrays. The color scale is shown in the panel on the right. The color scale was chosen proportional to the ranks. These graphical representation permit to show problems during the experimentation such as fingerprints, artificial gradient or dye specific failure for instance.

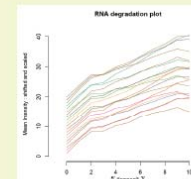


Figure 4. RNA digestion plot.

In this plot each array is represented by a single line. It is important to identify any array(s) that has a slope which is very different from the others. The indication is that the RNA used for that array has potentially been handled quite differently from the other arrays.

**WP2: Standards (LGC).** The objective of this work package is to plan and advocate the use of standards by the microarray community. This will involve the identification of suitable reference materials (spikes, reference RNAs), the assessment of analytical "best practice" guidelines and standardised approaches to experimental design and execution.

**WP3: Organisation and dissemination (NTNU).** The purpose of WP3 is to organise and structure the community "pull". First, we will identify and bring together the key players in the field of transcriptome microarray use and further development. We will disseminate the results of WP1 and WP2 to the community through a series of workshops. Updated information will be available through our web page: [www.microarray-quality.org](http://www.microarray-quality.org).

**WP4: Data Quality and Systems Biology (VIB).** WP4 will assess the impact of QM-based filtering and general QA/QC implementation on the performance of various mining and modelling approaches of such data compendia.

**WP5: Standards and European legislation (IRMM).** The purpose of WP5 is to take the QA/QC criteria analysed, developed and discussed in the previous 4 work packages and translate these into comparability criteria for microarray-relevant reference materials. These criteria will form the basis for independent projects, aimed at developing and distributing European reference materials.

**WP6: New Technologies (UU).** A survey of new applications and development efforts in microarray technologies will be performed, in order to identify key academic and commercial players (research groups, users, product and service providers).

### \*Project partners

Martin Kuiper - VIB, Belgium.  
Ame K. Sandvik - NTNU, Norway.  
Alvis Brazma - EBI, United Kingdom.  
Carole Foy - LGC, United Kingdom.

Joaquin Dopazo - CIPF, Spain.  
Laszlo Puskas - HAS, Hungary.  
Heinz Schimmel - IRMM, Belgium.  
Ulf Landegren - UU, Sweden.

If you are interested to participate, or have information relevant to this project, please contact:

**Project coordinator:**  
Martin Kuiper ([martin.kuiper@psb.ugent.be](mailto:martin.kuiper@psb.ugent.be))

**Project fellows:**  
Vidar Beisvåg ([vidar.beisvag@ntnu.no](mailto:vidar.beisvag@ntnu.no))  
Ewa Sugajska ([ewsug@psb.ugent.be](mailto:ewsug@psb.ugent.be))

### Funding

EMERALD is funded by the Sixth Framework Programme for the Quality of Life and Management of Living Resources. Project no. LSHG-CT-2006-037689. Scientific officer: Christina Kyriakopoulou ([@ec.europa.eu](mailto:@ec.europa.eu))

[www.microarray-quality.org](http://www.microarray-quality.org)

## Attachment 2: EMERALD leaflet distributed at MGED10, Brisbane Australia, 2007

### Project management

The project is managed by a project board which has representatives of the eight partners:

<b>Martin Kulp</b>	Flanders Institute for Biotechnology, VIB, Gent, Belgium.
<b>Arne K. Sandvik</b>	Norwegian University of Science and Technology, NTNU, Norway.
<b>Alvis Brazma</b>	European Bioinformatics Institute, EBI, United Kingdom.
<b>Carole Fay</b>	LGC, United Kingdom.
<b>Joaquín Dopazo</b>	Centro de Investigación Príncipe Felipe, Spain.
<b>László Puskas</b>	Biological Research Center of the Hungarian Academy of Sciences, Hungary.
<b>Heinz Schramel</b>	Institute for Reference Materials and Measurements, Belgium.
<b>Ulf Landegren</b>	Uppsala University, Sweden.

The project management is assisted by a scientific advisory board:

<b>Frank Holtege</b>	Utrecht University, Netherlands.
<b>Helen Causton</b>	Imperial College London, United Kingdom.
<b>Rafael Irizarry</b>	Johns Hopkins University, United States.
<b>Joerg Hübner</b>	German Cancer Research Center, DKFZ, Germany.
<b>Astrid Lagrèid</b>	Norwegian University of Science and Technology, Norway.
<b>Marc Salt</b>	National Institute of Standards and Technology, NIST, United States.
<b>Janet Warrington</b>	Affymetrix, Inc., United States.



[www.microarray-quality.org](http://www.microarray-quality.org)



## EMERALD

A European Project funded by the Sixth Framework Programme for the Quality of Life and Management of Living Resources  
Project no. LSHG-CT-2006-037689  
Scientific officer: Christina Kyriakopoulou (@ec.europa.eu)




### Contact:

**Project coordinator:**  
Martin Kulp  
VIB Department of Plant Systems Biology,  
UGent-VIB Research Building F5VM,  
Technologiepark 927  
BE-9052 GENT, Belgium  
Email: martin.kulp@psb.ugent.be  
Phone +32 93313805  
Fax +32 93313809

**Project fellows:**  
Vidar Belsvåg  
Dept. of Cancer Research and Molecular Medicine,  
Norwegian University of Science and Technology,  
Medisinsk Teknisk Forskningscenter  
Olav Kyrres gt.9,  
7489 Trondheim, Norway  
Email: vidar.belsvag@ntnu.no  
Phone +47 73598615  
Fax +47 72576400  
or  
Ewa Sugajska  
VIB Department of Plant Systems Biology,  
UGent-VIB Research Building F5VM,  
Technologiepark 927,  
BE-9052 GENT, Belgium  
Email: ewasug@psb.ugent.be  
Phone +32 93313823  
Fax +32 93313809

[www.microarray-quality.org](http://www.microarray-quality.org)



A European Project funded by the Sixth Framework Programme for the Quality of Life and Management of Living Resources




### European Project on Standards and Standardisation of Microarray Technology and Data Analysis

[www.microarray-quality.org](http://www.microarray-quality.org)

---

### Project objectives

This European Union Framework Program 6 Coordination Action (CA) will serve to establish and disseminate quality metrics (QM), microarray standards and best laboratory practices throughout the European microarray community. This will allow microarray data production governed by QA/QC, significantly enhancing the quality of microarray data and setting a precedent for other array-based technologies. Over the last 15 years microarray technology has proved the method of choice for capturing molecular biological data in a massively parallel fashion. 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. Very early in the development of microarray-based transcript profiling the microarray community has realised the importance of structured documentation accompanying microarray

[www.microarray-quality.org](http://www.microarray-quality.org)

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 predominantly been focused on data context, and has only recently been extended to included data content. Quality and integrity of microarray data compendia (e.g. in ArrayExpress) are major determinants for information extraction model building and high quality data will be one of the pillars of systems biology. This CA is designed to structure and amalgamate ongoing efforts across Europe, in close association with MGED and the ERCC.

### Coordination and dissemination activities

Coordination activities are defined in six main areas relevant for microarray analysis: Development of quality metrics, ontology for data description, implementation of standards and best practices, selection of standards that are candidates for European Reference Materials, impact on data information content, and dissemination of QA/QC principles to novel experimental high-throughput techniques for the different -omics domains. These activities are made up of six work packages (WP).

### Work packages

**WP1:** Quality Metrics and Ontologies (EB). The objective of this WP is to develop and disseminate quality metrics and tools for determining data quality and communicating data transformations. Ontology for describing microarray experiments and analysis will also be developed and disseminated.

**We are working on a website and a portal (www.microarray-quality.org). Please contact us with information relevant to the project.**

**WP2:** Standards (LGC). The objective of this work package is to plan and advocate the use of standards by the microarray community. This will involve the identification of suitable reference materials (spikes, reference RNAs), the assessment of analytical "best practice" guidelines and standardised approaches to experimental design and execution.

**WP3:** Organisation and dissemination (NTNU). The purpose of WP3 is to organise and structure the community "pull". First, we will identify and bring together the key players in the field of transcriptome microarray use and further development. We will disseminate the results of WP1 and WP2 to the community through a series of workshops.

**WP4:** Data Quality and Systems Biology (VIB). WP4 will assess the impact of QM-based filtering and general QA/QC implementation on the performance of various mining and modelling approaches of such data compendia.

**WP5:** Standards and European legislation (IRMM). The purpose of WP5 is to take the QA/QC criteria analysed, developed and discussed in the previous 4 work packages and translate these into commutability criteria for microarray-relevant reference materials. These criteria will form the basis for independent projects, aimed at developing and distributing European reference materials.

**WP6:** New Technologies (UU). A survey of new applications and development efforts in microarray technologies will be performed, in order to identify key academic and commercial players (research groups, users, product and service providers).

### EMERALD workshops

**WS1:** Introducing EMERALD to the community (EMERALD session in conjunction with MGED10, September 2007, Brisbane, Australia).

**WS2:** Ontology Workshop (planned 5-9 November 2007, Hinxton, UK).

**WS3:** Assessment of best practices, launch of a pilot QA/QC implementation (expected to be held in conjunction with MGED11, September 2008, Trento, Italy).

**WS4:** Presentation and dissemination of first QA/QC results (expected to be held in conjunction with MGED 11, September 2008, Trento, Italy).

**WS5:** Towards federal standards (planned 2008).

**WS6:** Data quality and Systems Biology (planned Autumn 2008 / Spring 2009)

**WS7:** Implications for new technologies (planned Spring 2009).

**WS8:** Dissemination of results to larger community (planned Autumn 2009).

[www.microarray-quality.org](http://www.microarray-quality.org)

### Web pages relevant for the project

EMERALD ([www.microarray-quality.org](http://www.microarray-quality.org))  
Microarray Gene Expression Data (MGED) Society ([www.mged.org](http://www.mged.org))  
National Institute of Standards and Technology (NIST) ([www.nist.gov/](http://www.nist.gov/))  
External RNA Control Consortium (ERCC) ([www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GenExpres/on/ERCC.htm](http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GenExpres/on/ERCC.htm))  
MicroArray Quality Control (MAQC) project ([www.eia.gov/nct/science/centers/toxicoinformatics/maqc/](http://www.eia.gov/nct/science/centers/toxicoinformatics/maqc/))