The MIAME Checklist for array based CGH experiments (draft December 2005)

The purpose of this checklist is to guide authors, journal editors and referees in helping them to ensure that the data supporting published results based on array CGH experiments are made publicly available in a format that enables unambiguous interpretation of the data and potential verification of the conclusions (see [1]). For more detail regarding the rationale of MIAME see [2]. MGED strongly recommends that the data is made publicly available through one of the public repositories for microarray data (see [3]).

Experiment Design:

- The goal of the experiment – one line maximum (e.g., the title from the related publication)
- A brief description of the experiment (e.g., the abstract from the related publication)
- Keywords, for example, *deletion, gain, amplification, polymorphisms detection, strain determination* (the use of MGED ontology terms is recommended).
- Experimental factors - the parameters or conditions tested, for example, *disease state, strain, species isolate* (the use of MGED ontology terms is recommended).
- Experimental design - relationships between samples, treatments, extracts, labeling, and arrays (e.g., a diagram or table).
- Quality control steps taken (e.g., replicates or dye swaps).
- Links to the publication, any supplemental websites or database accession numbers.

Samples used, extract preparation and labelling:

- The origin of each biological sample (e.g., name of the organism, the provider of the sample) and its characteristics (e.g., gender, age, developmental stage, strain, or disease state). This information must not compromise the patients anonymity – legal and ethical requirements take the precedence.
- Manipulation of biological samples and protocols used (e.g., *stress, drugs*).
- Experimental factor value for each experimental factor, for each sample (e.g., '*disease state* = AML' for a sample in a time course experiment).
- Technical protocols for preparing the hybridization extract and labeling.
- External controls (spikes), if used.

Hybridization procedures and parameters:

- The protocol and conditions used for hybridization, blocking and washing, including any post-processing steps such as staining.

Measurement data and specifications:

- Data
  - The raw data, i.e. scanner or imager and feature extraction output (providing the images is optional). The data should be related to the respective array designs (typically each row of the imager output should be related to a feature on the array – see Array Designs).
  - The normalized and summarized data, i.e., set of quantifications from several arrays upon which the authors base their conclusions (for gene expression experiments also known as gene expression data matrix and may consist of averaged normalized log ratios). The data should be related to the respective array designs (typically each row of the summarized data will be related to one biological annotation, such as clone name or specific oligo).
- Data extraction and processing protocols,
  - Image scanning hardware and software, and processing procedures and parameters.
  - Normalization, transformation and data selection procedures and parameters.
  - Definition of chromosomal aberration, polymorphisms or strain determination used for data interpretation

Array Design:

- General array design, including the platform type (whether the array is a spotted glass array, an *in situ* synthesized array, etc.); surface and coating specifications and spotting protocols used (for custom made arrays), or product identifiers (the name or make, catalogue reference numbers) for commercially available arrays.
- Array feature and reporter annotation, normally represented as a table (for instance see Tables 1, 2 below), including
  - For each feature (spot) on the array, its location on the array (e.g., metacolumn, metarow, column, row) and the reporter present in the location (note that the same reporter may be present on several features).
  - For each reporter unambiguous characteristics of the reporter molecule, including
    - Reporter role – control or measurement
    - The sequence for oligonucleotide based reporters
    - The source, preparation and the database accession number for long (e.g., BAC, cDNA or PCR product based) reporters
    - Primers for PCR product based reporters
  - Appropriate biological annotation for each reporter, for instance a gene identifier or name (note that different reporters can have the same biological annotation)
- Principal array organism(s)

Oligonucleotide arrayCGH description file example:

Feature				Reporter			Genome location	
Coordinates on Array				Biosequence Type	Sequence	Reporter Usage	Chromosome	Chromosomal position start in kb Database name with built or freeze number
Meta Col	Meta Row	Col	Row					
1	1	1	1	Labelled oligo	AAAAAAAAAAAAAAAAAA	Control Positive	_	_
1	1	2	1	Cot-1	_	Control Negative	_	_
…	…	…	…	…	…	…	…	…
4	6	10	8	Oligo	ATGTCCGTTGAATTGG	Experimental	11	11811
…	…	…	…	…	…	…	…	…
4	6	11	8	Oligo	AGTGGCGAGGAGGAGGAC	Experimental	15	12460
4	6	12	8	Oligo	CCACCACCAAGACCTACTCC	Experimental	15	12478

BAC, PAC, fosmid, PCR amplicons, cDNA and/or cosmid arrayCGH description file example

Feature Coordinates on Array				Reporter				Genome location					
Meta Col	Meta Row	Col	Row	Biosequence type	Unique identifier*	Clone ID (Use Int. clone name)	Reporter Usage	Mapping type†	Accession 1†	Accession 2*	Chromosome	Chromosomal position start or midpoint in kb Database name with built or freeze number	Chromosomal position end in kb*
1	1	1	1	Total human DNA	1	_	Control Positive	N/A			_	_	_-
1	1	2	1	BAC	2	RP11-219o07	Experimental	STS			1	39885	39886
1	1	3	1	Cosmid	3	9H11	Experimental	BAC ends			5	43633	43633
…	…	…	…	…		…	…				…	…	…
4	8	24	12	BAC	5689	RP11-154N07	Experimental	Fingerprint			16	59000	59002

* column not required to fill out

† Mapping type and accession numbers correspond; give accession numbers or primer sequences used for mapping the spotted probe; e.g. for end sequenced BAC's; BAC end 1 and BAC end 2, Fingerprint, Marker, Cytoband or Free text (e.g. when numeric positions given under "chromosomal position" have been estimated by e.g. FISH).

References

[1] Cech, T.R., Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. Available at www.nap.edu/books/0309088593/html

[2] A. Brazma, P Hingamp, J Quackenbush, G Sherlock, P Spellman, C Stoeckert, J Aach, W Ansorge, C A Ball, H C Causton, T Gaasterland, P Glenisson, F C P Holstege, I F Kim, V Markowitz, J C Matese, H Parkinson, A Robinson, U Sarkans, S Schulze-Kremer, J Stewart, R Taylor, J Vilo & M Vingron. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, Nature Genetics, vol 29 (December 2001), pp 365 - 371.

[3] C.A. Ball, A. Brazma, H. Causton, S. Chervitz, R. Edgar, P. Hingamp, J.C. Matese, H. Parkinson, J. Quackenbush, M. Ringwald, S. Sansone, G. Sherlock, P. Spellman, C. Stoeckert, Y. Tateno, R. Taylor, J. White and N. Winegarden. Submission of Microarray Data to Public Repositories. PLoS Biology, 2, e317: 1276-1277.