

Minimal Information About Microarray Experiments (MIAME):

Concept definitions, mapping to MAGE Object Model (MAGE-OM) and relationship with MGED ontology.

[Draft 2 for Version 1.0:](#)

MIAME version 1.1, March

MAGE-OM version October 1, 2001

MGED BioMaterial ontology version 13

(Revision will be made as the MGED ontology evolves)

MIAME, MAGE-OM and MGED ontology mapping.....	2
1. Array design	2
1.1. Array related information	2
1.2. Reporter related information	4
1.2.1. For each reporter type.....	4
1.2.2. For each reporter	4
1.3. Features related information	6
1.3.1. For each feature type	6
1.3.2. For each feature	6
1.4. Composite sequence related information	7
1.4.1. For each composite sequence	7
1.5. Control elements related information	8
2. Experiment design.....	9
2.1. Experimental design	9
2.2. Sample	11
2.2.1. Bio-source properties	11
2.2.2. Biomaterial manipulation	15
2.2.3. Hybridizationextract preparation.....	18
2.2.4. Sample labeling.....	19
2.2.5. Spiking control.....	20
2.3. Hybridizations	21
2.4. Measurements	22
2.4.1. Raw data	22
2.4.2. Image analysis and quantitation.....	22
2.4.3. Normalized and summarized data.....	23
MIAME Glossary	25

MIAME, MAGE-OM and MGED ontology mapping

The boundaries between MIAME concepts, the MIAME-compliant MAGE-OM and the MGED ontology- that try to define and structure the MIAME concepts- is neither well defined nor easy to understand.

In order to provide some help, these pages contain explanatory documentation for the MIAME concepts, how its requirements map to the MAGE-OM and where a MGED ontology inclusion is required.

At the present time the MGED ontology covers only experimental sample (BioMaterial). Work is in progress. Microarray descriptions that still require inclusion into the ontology are specified.

<u>MIAME</u>	<u>Description</u>	<u>MGED Ontology</u>	<u>MAGE-Object Model</u>	<u>When applicable</u>	<u>Notes</u>	<u>Allowed values</u>
1. Array design	The layout or conceptual description of array that can be implemented as one or more physical arrays. The array design specification consists of the description of the common features of the array as the whole, and the description of each array design elements (e.g., each spot). MIAME distinguishes between three levels of array design elements: feature (the location on the array), reporter (the nucleotide sequence present in a particular location on the array), and composite sequence (a set of reporters used collectively to measure an expression of a particular gene)		ArrayDesign_package	When an array design is novel and cannot refer to manufacturer	Array design should be provided by the array providers and manufactures, in which case the user will only need to reference an existing design	
1.1. Array related information	Description of the array as the whole					
<u>Array design</u>	Given name for the array		Name	When an	Should be consistent	Design name,

<u>name</u>	design, that helps to identify a design between others (e.g: EMBL yeast 12K ver1.1)		is an attribute of ArrayDesign_package	array design is novel and cannot refer to manufacturer	with the design name given for the array copy in the Experiment design	number of features, version (e.g: EMBL yeast 12K ver1.1)
<u>Platform type</u>	The technology type used to place the biological sequence on the array	<i>MGED controlled vocabulary to be developed for FeatureGroup TechnologyType</i>	TechnologyType is an association with FeatureGroup, class of ArrayDesign_package	When an array design is novel and cannot refer to manufacturer		in situ synthesized, spotted cDNA, etc
<u>Surface and coating specification</u>	Type of surface and name for the type of coating used	<i>MGED controlled vocabulary to be developed for PhysicalArrayDesign SurfaceType</i>	SurfaceType is an association with PhysicalArrayDesign, a class of ArrayDesign_package OntologyEntry class in Description_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with TechnologyType	SurfaceType = glass, membrane, etc name of coating type (e.g. amino silane)
<u>Array dimensions</u>	The physical dimension of the array support (e.g. of slide)	<i>MGED controlled vocabulary to be developed for ArrayGroup Substrate type</i>	Width and Length are attributes of ArrayGroup, class of Array_package	When an array design is novel and cannot refer to manufacturer		width, length
<u>Number of elements on the array</u>	The number of features on the array		NumberOfFeatures is an attribute of ArrayDesign, class of Array_package	When an array design is novel and cannot refer to manufacturer		number of elements
<u>Production protocol</u>	A description of how the array was manufactured	<i>MGED controlled vocabulary to be developed for Protocol type, Hardware and Software type</i>	Protocol_package ProtocolApplication is an association with ArrayManufacture, class of Array_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with Feature Location and Zone	Protocol= description, printing hardware, printing software

<u>Provider</u>	The primary contact (manufacturer) for the information on the array design		DesignProvider as an association with ArrayDesign, class of ArrayDesign_package	Always		Contact details of manufacturer
1.2. Reporter related information	Information on the nucleotide sequence present in a particular location on the array					
1.2.1. For each reporter type						
<u>Reporter type</u>	Physical nature of the reporter (e.g. PCR product, synthesized oligonucleotide)	<i>MGED controlled vocabulary to be developed for DesignElementGroup type</i>	Types is an association with DesignElementGroup, class of Array_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with TechnologyType	Types= empty, PCR, synthesized oligonucleotide, plasmid, colony, etc
<u>Single or double stranded</u>	Whether the reporter sequences are single or double stranded	<i>MGED controlled vocabulary to be developed for DesignElementGroup type</i>	Types is an association with DesignElementGroup, class of Array_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with element Type	Types= single, double
1.2.2. For each reporter						
<u>Reporter sequence information</u>	The nucleotide sequence information for reporter: sequence accession number (from DDBJ/EMBL/GenBank), the sequence itself (if known) or a reference sequences (e.g. for oligonucleotides) and PCR primers pair information (if relevant)	<i>MGED controlled vocabulary to be developed for DatabaseEntry</i>	ImmobilizedCharacteristics is an association with Reporter, class of DesignElement_package DatabaseEntry is a class of Description_package	When elements are NOT composite and when array design is novel and cannot refer to manufacturer	Should be consistent with element type and clone	sequence annotation, sequence, sequence accession number, PCR primer pair

<u>Reporter approximate length</u>	The approximate length of the reporter's sequence			When the exact reporter sequence is NOT known		Number of bases
<u>Clone information</u>	For each reporter, the identity of the clone along with information on the clone provider, the date obtained, and availability	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	<p>ImmobilizedCharacteristics is an association with Reporter, class of DesignElement_package</p> <p>BioMaterial Is associated with ManufactureLIMS, class of Array_package</p> <p><u>OntologyEntry</u>, associated to Biosource a class in BioMaterial_package</p> <p>DatabaseEntry is a class of Description_package</p>	When elements are obtained from clones and when an array design is novel and cannot refer to manufacturer	Should be consistent with element type	clone ID, provider, date obtained, availability
<u>Reporter generation protocol</u>	A description of how the reporters were generated	<i>MGED controlled vocabulary to be developed for Protocol type</i>	ProtocolApplications is an association with ArrayManufacture, class of Array_package	When an array design is novel and cannot refer to manufacturer		Protocol

1.3. Features related information	Information on the location of the reporters on the array					
1.3.1. For each feature type						
<u>Element dimensions</u>	The physical dimensions of each features		FeatureWidth, FeatureLength and FeatureHeight are attribute of FeatureGroup, class of ArrayDesign_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with array dimensions and number of array elements	Width, length, height, diameter
<u>Attachment</u>	How the element (reporter) sequences are physically attached to the array (e.g. covalent, ionic)	<i>MGED controlled vocabulary to be developed DesignElementGroup type</i>	Types is an association with DesignElementGroup, class of Array_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with element generation protocol	covalent, ionic, hydrophobic, etc
1.3.2. For each feature						Normally given as a spread-sheet or tab-delimited file
<u>Reporter and location</u>	The arrangement and the system used to specify the location of each features on the array (e.g. grid, row, column, zone)		FeatureLocation and Position are associations with Feature, class of DesignElement_package Zone, ZoneLayout and ZoneGroup are classes of ArrayDesign_package	When an array design is novel and cannot refer to manufacturer	Should be consistent with array dimensions and NumberOfFeatures	row, column, x microns, y microns, zone

1.4. Composite sequence related information	Information on the set of reporters used collectively to measure an expression of a particular gene					
1.4.1. For each composite sequence						
<u>Composite sequence information</u>	The set of reporters contained in the composite sequence. The nucleotide sequence information for each composite element: number of oligonucleotides, oligonucleotide sequences (if given), and the reference sequence accession number (from relevant databases)	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	BiologicalCharacteristics Is an association with CompositeSequence, class of DesignElement_package ImmobilizedCharacteristics is an association with Reporter, class of DesignElement_package ReporterCompositeMap is an association with CompositeSequence, class of DesignElement_package DatabaseEntry is a class of Description_package	When elements ARE composite and when array design is novel and cannot refer to manufacturer	Should be consistent with element type	oligonucleotide sequences, number of oligonucleotides, reference sequence
<u>Gene name</u>	The gene represented at each composite sequence: name and links to appropriate databases (e.g. SWISS-PTOR or organism specific database)	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	BiologicalCharacteristics Is an association with CompositeSequence, class of DesignElement_package DatabaseEntry	When an array design is novel and cannot refer to manufacturer	Should be consistent with clone and composite sequence information	Gene name, accession number, annotation

			is a class of Description_package			
<u>Qualifier, value, source</u> (may use more than once)	Describe any further information about the array in a structured manner	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	OntologyEntry and DatabaseEntry are class in Description_package NameValueType is also a top level class	When additional information is available that would be useful to base queries on		Qualifier= name Value= value Source= database entry or ontology entry
1.5. Control elements related information	Array elements that have an expected value and/or are used for normalization					
<u>Control elements position</u>	The position of the control features on the array		ControlFeatures is an association with DesignElement_package	When any elements on the array were used as controls	Should be consistent with QualityControlDescription	row, column, x microns, y microns, zone
<u>Control type</u>	The type of control used for the normalization and their qualifier	<i>MGED controlled vocabulary to be developed for DesignElement controlType</i>	ControlType is an association with DesignElement_package	When any elements on the array were used as controls	Should be consistent with QualityControlDescription	control type (spiking, negative, positive), control qualifier (endogenous, exogenous)

<u>MIAME</u>	<u>Description</u>	<u>MGED Ontology</u>	<u>MAGE-Object Model</u>	<u>When applicable</u>	<u>Notes</u>	<u>Allowed values</u>
2. Experiment design	Experiment is a set of one or more hybridizations that are in some way related (e.g., related to the same publication MIAME distinguishes between: the experiment design (the design, purpose common to all hybridisations performed in the experiment), the sample used (sample characteristics, the extract preparation and the labeling), the hybridisation (procedures and parameters) and the data (measurements and specifications)					
2.1. Experimental design	Design and purpose common to all hybridisations performed in the experiment		Experiment_package	Always	Experiment represents the container for all the related BioAssays (hybridizations)	
<u>Author, laboratory, and contact</u>	Person(s) and organization (s) names and details (address, phone, FAX, email, URL)	<i>MGED controlled vocabulary to be developed for Contact roles</i>	AuditandSecurity_package	Always		Contact details
<u>Experiment type (s)</u>	A controlled vocabulary that classify an experiment	<i>MGED controlled vocabulary to be developed for ExperimentalDesign type</i>	Type is an association with ExperimentalDesign, class of Experiment_package	Always	Type should be consistent with ExperimentalFactor (s)	Type list = time course, dose response, comparison (disease vs normal, treated vs untreated), temperature shock, gene knock out, gene knock in (transgenic), ect.

<u>Experimental factor (s)</u>	Parameter (s) or condition (s) tested in the experiment	<i>MGED controlled vocabulary to be developed for ExperimentalFactor category</i>	ExperimentalFactor is a class of Experiment_package	Always	ExperimentalFactor (s) should be consistent with Type (s)	Biological factor= time, dose, genetic variation (knock out, knock in) compound, temperature Methodological factor= Protocol difference (extraction, hybridization, labeling, scanning)
<u>Number of hybridisations</u>	Number of hybridizations performed in the experiment		Relationship between Experimental class of experiment_package and PhysicalBioAssay class of BioAssayData_package	Always	Should be consistent with Type (s)	Single, multiple
<u>Common reference</u>	A hybridization to which all the other hybridisations have been compared	<i>MGED controlled vocabulary be developed for Common reference type</i>	Captured by the relationships among BioAssays and BioAssayData	Always		Yes, no, type (e.g. pairwise comparison, circular comparison)
<u>Quality control steps</u>	Measures taken to ensure or measure quality: replicates (number and description), dye swap (for two channel platforms) or others (unspecific binding, low complexity regions, polyA tails)	<i>MGED controlled vocabulary be developed for ReplicateDescription</i>	QualityControlDescription from Description_package associated to ExperimentalDesign, class of Experiment_package ReplicateDescription from Description_package associated to ExperimentalDesign, class of Experiment_package	When these have been done		Text description. biological, technical

<u>Experiment description</u>	Free text description of the experiment and link to an electronic publication in a peer-reviewed journal	<i>MGED controlled vocabulary to be developed for BibliographicReferences parameters and DatabaseEntry</i>	Experiment_package and BQS_package DatabaseEntry is a class of Description_package	When additional information is available and an electronic publication exists	Should be consistent with ExperimentalDesign	Text description, citation, URL, database entry
<u>Qualifier, value, source (may use more than once)</u>	Describe any further information about the experiment set in a structured manner	<i>MGED controlled vocabulary to be developed for DatabasEntry type</i>	OntologyEntry and DatabaseEntry are class in Description_package NameValueType is also a top level class	When additional information is available that would be useful to base queries on		Qualifier= name Value= value Source= database entry or ontology entry
2.2. Sample	The biological material from which the nucleic acids have been extracted for subsequent labelling and hybridisation. MIAME distinguishes between: source of the sample (bio-source), its treatment, the extract preparation, and its labeling	BioMaterial Ontology	BioMaterial_package. BioMaterial is the biological material used in the experiment: Biosource (the primary source of the nucleic acid used to generate labelled material for the microarray experiment); Biosample (the Biosource after any treatment); LabelledExtract (the biosample after labeling for detection of the nucleic acids.)	Always	Should be consistent with the Experiment_package, Array_package, BioMaterial_package and BioAssay_data	For recommendations see also www.mged.org/ontology
2.2.1. Bio-source properties	Information on the source of the sample		(BioMaterial) Biosource			
<u>Organism</u>	The genus and species (and subspecies) of the organism from which the BioMaterial is derived [MGED Ontology Definition]	Organism is a <u>BiosourceOntology Entry</u> in BioMaterial	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	Always		Organism= genus, species, subspecies from NCBI taxonomy

		Ontology				
<u>Contact details for sample</u>	The resource (e.g, company, hospital, geographical location) used to obtain or purchase the BioMaterial and the type of specimen [MGED Ontology Definition]	BioMaterialProvider is a <u>BiosourceOntology Entry</u> in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When BioMaterial was prepared or grown outside of the laboratory listed for the author		Biosource provider= details, contact. Type of specimen= tumor biopsy, paraffin section, stool sample
<u>Cell type</u>	Cell type used in the experiment if non mixed. If mixed the targeted cell type should be used [MGED Ontology Definition]	CellType is a <u>BiosourceOntology Entry</u> in BioMaterial Ontology . <i>MGED controlled vocabulary to be developed for BioSource characteristics CellType</i>	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	Always	Should be consistent with organism and targetedCellType	Cell type= Term (epithelial, hepatic.), source of term (ontology, dictionary, or controlled vocabulary) e.g: Mouse Anatomical Dictionary, FlyBase, CBIL controlled vocabulary, ATCC
<u>Sex</u>	Term applied to any organism able to undergo sexual reproduction in order to differentiate the individuals or type involved. Sexual reproduction is defined as the ability to exchange genetic material with the potential of recombinant progeny [MGED Ontology Definition]	Sex is a <u>BiosourceOntology Entry</u> , in BioMaterial Ontology .	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When applicable	Should be consistent with organism	Sex= Mating type alpha, F ⁺ , F ⁻ , Hfr, Mating type a, Mixed sex, Unknown sex
<u>Age</u>	The time period elapsed since an identifiable point in the life cycle of an organism. (If a developmental stage is specified, the identifiable point would be the beginning of that stage.	Age is a <u>BiosourceOntology Entry</u> in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When applicable	Should be consistent with organism	Age = combination of real number (measurement) and initial time point e.g.: coitus, birth, planting, beginning of stage

	Otherwise the identifiable point must be specified such as planting) [MGED Ontology Definition]					
<u>Developmental stage</u>	The developmental stage of the organism's life cycle during which the BioMaterial was extracted [MGED Ontology Definition]	DevelopmentalStage is a <u>BiosourceOntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	For multicellular species	Should be consistent with organism	Developmental stage = term, source of term (ontology, dictionary, or controlled vocabulary)
<u>Organism part</u>	The part or tissue of the organism's anatomy from which the BioMaterial was derived [MGED Ontology Definition]	OrganismPart is a <u>BiosourceOntologyEntry</u> in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	For multicellular species	Should be consistent with organism	Organism part = term, source of term (ontology, dictionary, or controlled vocabulary)
<u>Strain or line</u>	Animals or plants that have a single ancestral breeding pair or parent as a result of brother x sister or parent x offspring matings [MGED Ontology Definition]	StrainOrLine is a <u>BiosourceOntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When known	Should be consistent with organism	Strain or line = term, source of term (ontology, dictionary, or controlled vocabulary). E.g.: Jax mouse strains cultivar= NCBI taxonomy
<u>Genetic variation</u>	The genetic modification introduced into the organism from which the BioMaterial was derived. Examples of genetic variation include specification of a transgene or the gene knocked-out [MGED Ontology Definition]	GeneticVariation is a <u>BiosourceOntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When the source organism is genetically modified	Should be consistent with organism	Genetic variation = term, source of term (ontology, dictionary, or controlled vocabulary)
<u>Individual number</u>	Identifier or number of the individual organism from which the BioMaterial was derived. For patients, the identifier must be approved by Institutional Review Boards (IRB, review and monitor	Individual is an <u>OntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When the organism can be distinguished on an individual basis with a unique ID	Should be consistent with organism	Individual = ID

	biomedical research involving human subjects) or appropriate body [MGED Ontology Definition]					
<u>Individual genetic characteristics</u>	The genotype of the individual organism from which the BioMaterial was derived [MGED Ontology Definition]	IndividualGeneticC haracteristics is a <u>BiosourceOntology Entry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for BioSource IndividualGeneticC haracteristics</i>	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When applicable	Should be consistent with organism	Individual genetic characteristics= allele, genotype, haplotype, polymorphisms. term, source of term (ontology, dictionary, or controlled vocabulary)
<u>Disease state</u>	The name of the pathology diagnosed in the organism from which the BioMaterial was derived. The disease state is normal if no disease has been diagnosed [MGED Ontology Definition]	DiseaseState is an <u>OntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When applicable	Should be consistent with organism	If no disease then value "normal". Disease state= disease= term, source of term (ontology, dictionary, or controlled vocabulary)
<u>Targeted cell type</u>	The targeted cell type is the cell of primary interest. The BioMaterial may be derived from a mixed population of cells although only one cell type is of interest [MGED Ontology Definition]	TargetedCellType is a <u>BiosourceOntology Entry</u> in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for BioSource characteristics cell type</i>	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When the BioMaterial is derived from a mixed population of cells	Should be consistent with organism and cell type	Targeted cell type= term, source of term (ontology, dictionary, or controlled vocabulary) e.g: Mouse Anatomical Dictionary, FlyBase, CBIL controlled vocabulary

<u>Cell line</u>	The identifier for the immortalized cell line if one was used to derive the BioMaterial [MGED Ontology Definition]	CellLine is a <u>BiosourceOntologyEntry</u> , in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When the BioMaterial is derived from the immortalized cell line	Should be consistent with organism and cell type	Cell line= term, source of term (ontology, dictionary, or controlled vocabulary). E.g.: Hela, Caco-2
2.2.2. Biomaterial manipulation	Information on the treatment applied to the biomaterial					
<u>Growth conditions</u>	A description of the isolated environment used to grow organisms or parts of the organism [MGED Ontology Definition]	CultureCondition is an class of BioMaterialManipulation, in BioMaterial Ontology	<u>OntologyEntry</u> , associated to Biosource a class in BioMaterial_package	When known		Culture condition= atmosphere, contaminant organism, density range, generations, host, humidity, light, medium, nutrients, temperature

<p><u>In vivo treatment</u></p>	<p>The manipulation of the organism for the purposes of generating one of the variables under study and the documentation of the set of steps taken in the treatment</p>	<p>Treatment is an class of BioMaterialManipulation, in BioMaterial Ontology</p> <p><i>MGED controlled vocabulary to be developed for Treatment actions (e.g. grow, wait, add)</i></p> <p>Protocol Is an <u>OntologyEntry</u>, in BioMaterial Ontology</p> <p><i>MGED controlled vocabulary to be developed for Protocol type</i></p>	<p>Treatment is a class in BioMaterial_package</p> <p><u>OntologyEntry</u> associated to BioMaterial in BioMaterial_package</p>	<p>When the sample has been treated or manipulated in vivo for the study purpose</p>	<p>Should be consistent (where appropriate) with ExperimentType, ExperimentalFactors</p> <p>Should be consistent with Protocol_package</p>	<p>Protocol= citation, name, description, hardware, software</p>
---------------------------------	--	---	---	--	---	--

<p><u>In vitro treatment</u></p>	<p>The manipulation of the cell culture condition for the purposes of generating one of the variables under study and the documentation of the set of steps taken in the treatment</p>	<p>Treatment is an class of BioMaterialManipulation, in BioMaterial Ontology</p> <p><i>MGED controlled vocabulary to be developed for Treatment actions (e.g. grow, wait, add)</i></p> <p>Protocol Is an <u>OntologyEntry</u>, in BioMaterial Ontology</p> <p><i>MGED controlled vocabulary to be developed for Protocol type</i></p>	<p>Treatment is a class in BioMaterial_package</p> <p><u>OntologyEntry</u> associated to BioMaterial in BioMaterial_package</p>	<p>When the sample has been treated or manipulated in vitro for the study purpose</p>	<p>Should be consistent (where appropriate) with ExperimentType, ExperimentalFactors</p> <p>Should be consistent with Protocol_package</p>	<p>Protocol= citation, name, description, hardware, software</p>
<p><u>Treatment type</u></p>	<p>The type of manipulation applied to the BioMaterial for the purposes of generating one of the variables under study [MGED Ontology Definition]</p>	<p>Treatment type are sub-classes of Treatment, a class of BioMaterialManipulation, in BioMaterial Ontology</p>	<p>Treatment is a class in BioMaterial_package</p> <p><u>OntologyEntry</u> associated to BioMaterial in BioMaterial_package</p>	<p>When the sample has been treated or manipulated for the study purpose</p>	<p>Should be consistent (where appropriate) with ExperimentType, ExperimentalFactors and Treatment</p>	<p>Treatment type= behavioural stimulus, compound based treatment, infection, modification (genetic, somatic), starvation, heat shock, cold shock</p>

<u>Compound</u>	A drug, solvent, chemical, etc., that can be measured [MGED Ontology Definition]	Compound is an <u>BiosourceOntologyEntry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	Compound is a class in BioMaterial_package DatabaseEntry is a class of Description_package <u>OntologyEntry</u> , associated to BioMaterial in BioMaterial_package	When the sample has been treated or manipulated for the study purpose with a compound	Should be consistent with Treatment	Compound= protocol, compound, database entry, measurement, delivery method (e.g. intraperitoneal, gavage)
<u>Separation technique</u>	Technique to separate tissues or cells from a heterogenous sample (e.g. trimming, microdissection, FACS)	Protocol Is an <u>OntologyEntry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for Protocol type</i>	Treatment is a class in BioMaterial_package <u>OntologyEntry</u> associated to BioMaterial in BioMaterial_package	When the cells or tissue are separated from a heterogenous sample		Protocol= description, hardware, software
2.2.3. Hybridization extract preparation	Information on the extract preparation for each extract prepared from the sample		Biosample, the biosource after any treatment.			
<u>Extraction method</u>	The protocol used to extract nucleic acids from the sample	Protocol Is an <u>OntologyEntry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for Protocol type</i>	Extraction is a ProtocolType, <u>OntologyEntry</u> , associated to BioMaterial in BioMaterial_package Treatment is a class in BioMaterial_package	Always	Should be consistent with Protocol_package	Protocol,

<u>Nucleic acid type</u>	The type of nucleic acid extracted (e.g. total RNA, mRNA)	<i>MGED controlled vocabulary to be developed for BioMaterial material type and BioSample type (to describe the role that the BioSample hold in the treatment hierarchy)</i>	Extract is a Biosample_type <u>OntologyEntry</u> , associated to Biosample in BioMaterial_package	Always		Polymer type= total RNA, mRNA, DNA
<u>Amplification method</u>	The method used to amplify the nucleic acid extracted	Protocol Is an <u>OntologyEntry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for Protocol type</i>	Treatment is a class in BioMaterial_package	When applicable	Should be consistent with Protocol_package	Protocol,, RNA polymerases, PCR
2.2.4. Sample labeling	Information on the labeling preparation for each labelled extract		LabelledExtract, the biosample after labeling for detection of the nucleic acids			
<u>Amount of nucleic acid labeled</u>	The amount of nucleic acid labeled	Protocol Is an <u>OntologyEntry</u> , in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for Protocol type</i>	Labeling is a ProtocolType, <u>OntologyEntry</u> , associated to BioMaterial in BioMaterial_package Treatment is a class in BioMaterial_package		Should be consistent with Protocol_package	Protocol
<u>Label used</u>	The name of the label used	<i>MGED controlled vocabulary to be developed for label type</i>	Compound is an <u>BiosourceOntologyEntry</u> , associated to BioMaterial in BioMaterial_package	Always		Label= Cy3, Cy5, etc.

<u>Label incorporation method</u>	The label incorporation method used	Protocol Is an <u>OntologyEntry</u> in BioMaterial Ontology <i>MGED controlled vocabulary to be developed for Protocol type</i>	Labeling is a ProtocolType, <u>OntologyEntry</u> associated to BioMaterial in BioMaterial_package Treatment is a class in BioMaterial_package	Always	Should be consistent with Protocol_package	Protocol
2.2.5. Spiking control	External controls added to the hybridisation extract (s)					
<u>Spiking control feature</u>	Position of the feature (s) on the array expected to hybridise to the spiking control		ControlFeatures is an association with DesignElement_package	When applicable	Should be consistent with QualityControlDescription	row, column, x microns, y microns, zone
<u>Spike type and qualifier</u>	The type of spike used (e.g. oligonucleotide, plasmid DNA, transcript) and its qualifier (e.g. concentration, expected ratio, labelling methods)	<i>MGED controlled vocabulary to be developed for DesignElement controlType</i>	ControlType is an association with DesignElement_package	When applicable	Should be consistent with QualityControlDescription	spike type (e.g. oligonucleotide, plasmid DNA, transcript), qualifier (e.g. concentration, expected ratio, labelling methods)
<u>Qualifier, value, source (may use more than once)</u>	Describe any further information about the sample in a structured manner	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	OntologyEntry and DatabaseEntry are class in Description_package NameValueType is also a top level class	When additional information is available that would be useful to base queries on		Qualifier= name Value= value Source= database entry or ontology entry

2.3. Hybridizations	Procedures and parameters for each hybridization		The joining of the BioMaterial with an Array is a BioAssayCreation, a class of BioAssay_package	Always		
<u>Relationship between samples and arrays</u>	Relationship between the labelled extract (related to which sample which extract) and arrays (design, batch and serial number) in the experiment		BioAssays_package.	Always	Should be consistent with TechnologyType and QualityControlDescription	Which labelled extract (related to which sample which extract) was used on which array (array design, batch and serial number) during which hybridization
<u>Hybridization protocol</u>	Documentation of the set of steps taken in the hybridization, including: solution (e.g. concentration of solutes); blocking agent and concentration used; wash procedure; quantity of labelled target used; time; concentration; volume, temperature, and description of the hybridization instruments	<i>MGED controlled vocabulary to be developed for Protocol type, Hardware and Software type</i>	Protocol_package Hybridization is a class of BioAssay_package	Always		Protocol= description, hardware, software
<u>Qualifier, value, source (may use more than once)</u>	Describe any further information about the hybridization in a structured manner.	<i>MGED controlled vocabulary to be developed for DatabaseEntry type</i>	OntologyEntry and DatabaseEntry are class in Description_package NameValueType is also a top level class	When additional information is available that would be useful to base queries on		Qualifier= name Value= value Source= database entry or ontology entry

2.4. Measurements	MIAME distinguishes between three levels of data processing: image (raw data), image analysis and quantitation, gene expression data matrix (normalized and summarized data)					
2.4.1. Raw data	Each hybridization has at least one image					
<u>Scanner image file</u>	The TIFF file including header	<i>MGED controlled vocabulary to be developed for Image format</i>	Image is a class in BioAssay_package	Always	Should be consistent with BioAssay_package and Measurement_package	TIF, JPEG (Note: MGED does <i>not</i> have a consensus on image as part of MIAME)
<u>Scanning protocol</u>	Documentation of the set of steps taken for scanning the array and generating an image including: description of the scanning instruments and the parameter settings	<i>MGED controlled vocabulary to be developed for Protocol type Hardware and Software type</i>	Protocol_package ImageAcquisition is a class in BioAssay_package	Always.	Should be consistent with BioAssay_package .	Protocol= description, scanning hardware, scanning software, scan parameters (laser power, spatial resolution, pixel space, PMT voltage)
2.4.2. Image analysis and quantitation.	Each image has a corresponding image quantitation table, where a row represents a array design element and a column to a different quantitation types (e.g. mean or median pixel intensity)					
<u>Image analysis output</u>	The complete image analysis output for each image	<i>MGED controlled vocabulary to be developed for QuantitationType dataType and scale</i>	MeasuredBioAssayData is a class in BioAssayData_package	Always.	Should be consistent with Image in BioAssay_package	Normally given as a spread-sheet or tab-delimited file

<u>Image analysis protocol</u>	Documentation of the set of steps taken to quantify the image including: the image analysis software, the algorithm and all the parameters used	<i>MGED controlled vocabulary to be developed for Protocol type Hardware and Software type</i>	Protocol_package and BioAssayData_package	Always.	Should be consistent with Image in BioAssay_package	Protocol= description, image analysis hardware, image analysis software (specification, availability and version) algorithms, parameters
2.4.3. Normalized and summarized data	Several quantitation tables are combined using data processing metrics to obtain the 'final' gene expression measurement table (gene expression data matrix) associated with the experiment					For recommendations see also www.mged.org/normalization
<u>Data processing protocol</u>	Documentation of the set of steps taken to process the data, including: the normalization strategy and the algorithm used to allow comparison of all data	<i>MGED controlled vocabulary to be developed for NormalizationDescription, NodeValue dataType and scale of the value</i>	NormalizationDescription from Description_package associated to ExperimentalDesign, class of Experiment_package Transformation is a class of BioAssayData_package	When normalization has been performed		Protocol, normalization strategy (spiking, "housekeeping gene", total array), algorithm (linear regression, total intensity, ratio statistics, log (ratio) mean/median centring)
<u>Final gene expression table (s)</u>	Derived measurement value summarizing related elements and replicates, providing the type of reliability indicator used		DerivedBioAssayData is a class in BioAssayData_package ConfidenceIndicator is a class in QuantitationType_package Transformation is a class in BioAssayData_package Protocol_package	When a value used for a reliability indicator has been generated	Should be consistent with QualityControlDescription and ReplicateDescription	Replicates of the elements on the same or different arrays or hybridizations, as well as different elements related to the same entity (e.g. gene). Reliability indicator for each datapoint (e.g. standard deviation)

<p><u>Qualifier, value, source</u> (<u>may use more than once</u>)</p>	<p>Describe any further information about the measurements in a structured manner</p>	<p><i>MGED controlled vocabulary to be developed for DatabaseEntry type</i></p>	<p>OntologyEntry and DatabaseEntry are class in Description_package</p> <p>NameValueType is also a top level class</p>	<p>When additional information is available that would be useful to base queries on</p>		<p>Qualifier= name Value= value Source= database entry or ontology entry</p>
--	---	---	---	---	--	--

MIAME Glossary

MIAME concepts are listed in alphabetical order and definitions are provided.

Age	The time period elapsed since an identifiable point in the life cycle of an organism. (If a developmental stage is specified, the identifiable point would be the beginning of that stage. Otherwise the identifiable point must be specified such as planting) [MGED Ontology Definition]
Amount of nucleic acid labeled	The amount of nucleic acid labeled
Amplification method	The method used to amplify the nucleic acid extracted
Array design	The layout or conceptual description of array that can be implemented as one or more physical arrays. The array design specification consists of the description of the common features of the array as the whole, and the description of each array design elements (e.g., each spot). MIAME distinguishes between three levels of array design elements: feature (the location on the array), reporter (the nucleotide sequence present in a particular location on the array), and composite sequence (a set of reporters used collectively to measure an expression of a particular gene)
Array design name	Given name for the array design, that helps to identify a design between others (e.g: EMBL yeast 12K ver1.1)
Array dimensions	The physical dimension of the array support (e.g. of slide)
Array related information	Description of the array as the whole
Attachment	How the element (reporter) sequences are physically attached to the array (e.g. covalent, ionic)
Author, laboratory, and contact	Person(s) and organization (s) names and details (address, phone, FAX, email, URL)
Biomaterial manipulation	Information on the treatment applied to the biomaterial
Bio-source properties	Information on the source of the sample
Cell line	The identifier for the immortalized cell line if one was used to derive the BioMaterial [MGED Ontology Definition]
Cell type	Cell type used in the experiment if non mixed. If mixed the targeted cell type should be used [MGED Ontology Definition]
Clone information	For each reporter, the identity of the clone along with information on the clone provider, the date obtained, and availability
Common reference	A hybridization to which all the other hybridisations have been compared
Composite sequence information	The set of reporters contained in the composite sequence. The nucleotide sequence information for each composite element: number of oligonucleotides, oligonucleotide sequences (if given), and the reference sequence accession number (from relevant databases)
Composite sequence related information	Information on the set of reporters used collectively to measure an expression of a particular gene
Compound	A drug, solvent, chemical, etc., that can be measured [MGED Ontology Definition]
Contact details for sample	The resource (e.g. company, hospital, geographical location) used to obtain or purchase the BioMaterial and the type of specimen [MGED Ontology Definition]
Control elements position	The position of the control features on the array
Control elements related information	Array elements that have an expected value and/or are used for normalization
Control type	The type of control used for the normalization and their qualifier
Data processing protocol	Documentation of the set of steps taken to process the data, including: the normalization strategy and the algorithm used to allow comparison of all data

Developmental stage	The developmental stage of the organism's life cycle during which the BioMaterial was extracted [MGED Ontology Definition]
Disease state	The name of the pathology diagnosed in the organism from which the BioMaterial was derived. The disease state is normal if no disease has been diagnosed [MGED Ontology Definition]
Element dimensions	The physical dimensions of each features
Experiment description	Free text description of the experiment and link to an electronic publication in a peer-reviewed journal
Experiment design	Experiment is a set of one or more hybridizations that are in some way related (e.g., related to the same publication MIAME distinguishes between: the experiment design (the design, purpose common to all hybridisations performed in the experiment), the sample used (sample characteristics, the extract preparation and the labeling), the hybridisation (procedures and parameters) and the data (measurements and specifications)
Experiment type (s)	A controlled vocabulary that classify an experiment
Experimental design	Design and purpose common to all hybridisations performed in the experiment
Experimental factor (s)	Parameter (s) or condition (s) tested in the experiment
Extraction method	The protocol used to extract nucleic acids from the sample
Features related information	Information on the location of the reporters on the array
Final gene expression table (s)	Derived measurement value summarizing related elements and replicates, providing the type of reliability indicator used
Gene name	The gene represented at each composite sequence: name and links to appropriate databases (e.g. SWISS-PTOR or organism specific database)
Genetic variation	The genetic modification introduced into the organism from which the BioMaterial was derived. Examples of genetic variation include specification of a transgene or the gene knocked-out [MGED Ontology Definition]
Growth conditions	A description of the isolated environment used to grow organisms or parts of the organism [MGED Ontology Definition]
Hybridization protocol	Documentation of the set of steps taken in the hybridization, including: solution (e.g. concentration of solutes); blocking agent and concentration used; wash procedure; quantity of labelled target used; time; concentration; volume, temperature, and description of the hybridization instruments
Hybridizationextract preparation	Information on the extract preparation for each extract prepared from the sample
Hybridizations	Procedures and parameters for each hybridization
Image analysis and quantitation.	Each image has a corresponding image quantitation table, where a row represents a array design element and a column to a different quantitation types (e.g. mean or median pixel intensity)
Image analysis output	The complete image analysis output for each image
Image analysis protocol	Documentation of the set of steps taken to quantify the image including: the image analysis software, the algorithm and all the parameters used
In vitro treatment	The manipulation of the cell culture condition for the purposes of generating one of the variables under study and the documentation of the set of steps taken in the treatment
In vivo treatment	The manipulation of the organism for the purposes of generating one of the variables under study and the documentation of the set of steps taken in the treatment
Individual genetic characteristics	The genotype of the individual organism from which the BioMaterial was derived [MGED Ontology Definition]
Individual number	Identifier or number of the individual organism from which the BioMaterial was derived. For patients, the identifier must be approved by Institutional Review Boards (IRB, review and monitor biomedical research involving human subjects) or appropriate body [MGED Ontology Definition]
Label incorporation method	The label incorporation method used
Label used	The name of the label used

Measurements	MIAME distinguishes between three levels of data processing: image (raw data), image analysis and quantitation, gene expression data matrix (normalized and summarized data)
Normalized and summarized data	Several quantitation tables are combined using data processing metrics to obtain the 'final' gene expression measurement table (gene expression data matrix) associated with the experiment
Nucleic acid type	The type of nucleic acid extracted (e.g. total RNA, mRNA)
Number of elements on the array	The number of features on the array
Number of hybridisations	Number of hybridizations performed in the experiment
Organism	The genus and species (and subspecies) of the organism from which the BioMaterial is derived [MGED Ontology Definition]
Organism part	The part or tissue of the organism's anatomy from which the BioMaterial was derived MGED Ontology Definition]
Platform type	The technology type used to place the biological sequence on the array
Production protocol	A description of how the array was manufactured
Provider	The primary contact (manufacturer) for the information on the array design
Qualifier, value, source (may use more than once)	Describe any further information about the array in a structured manner
Quality control steps	Measures taken to ensure or measure quality: replicates (number and description), dye swap (for two channel platforms) or others (unspecific binding, low complexity regions, polyA tails)
Raw data	Each hybridization has at least one image
Relationship between samples and arrays	Relationship between the labelled extract (related to which sample which extract) and arrays (design, batch and serial number) in the experiment
Reporter and location	The arrangement and the system used to specify the location of each features on the array (e.g. grid, row, column, zone)
Reporter approximate length	The approximate length of the reporter's sequence
Reporter generation protocol	A description of how the reporters were generated
Reporter related information	Information on the nucleotide sequence present in a particular location on the array
Reporter sequence information	The nucleotide sequence information for reporter: sequence accession number (from DDBJ/EMBL/GenBank), the sequence itself (if known) or a reference sequences (e.g. for oligonucleotides) and PCR primers pair information (if relevant)
Reporter type	Physical nature of the reporter (e.g. PCR product, synthesized oligonucleotide)
Sample	The biological material from which the nucleic acids have been extracted for subsequent labelling and hybridisation. MIAME distinguishes between: source of the sample (bio-source), its treatment, the extract preparation, and its labeling
Sample labeling	Information on the labeling preparation for each labelled extract
Scanner image file	The TIFF file including header
Scanning protocol	Documentation of the set of steps taken for scanning the array and generating an image including: description of the scanning instruments and the parameter settings
Separation technique	Technique to separate tissues or cells from a heterogenous sample (e.g. trimming, microdissection, FACS)
Sex	Term applied to any organism able to undergo sexual reproduction in order to differentiate the individuals or type involved. Sexual reproduction is defined as the ability to exchange genetic material with the potential of recombinant progeny [MGED Ontology Definition]
Single or double stranded	Whether the reporter sequences are single or double stranded

Spike type and qualifier	The type of spike used (e.g. oligonucleotide, plasmid DNA, transcript) and its qualifier (e.g. concentration, expected ratio, labelling methods)
Spiking control	External controls added to the hybridisation extract (s)
Spiking control feature	Position of the feature (s) on the array expected to hybridise to the spiking control
Strain or line	Animals or plants that have a single ancestral breeding pair or parent as a result of brother x sister or parent x offspring matings [MGED Ontology Definition]
Surface and coating specification	Type of surface and name for the type of coating used
Targeted cell type	The targeted cell type is the cell of primary interest. The BioMaterial may be derived from a mixed population of cells although only one cell type is of interest [MGED Ontology Definition]
Treatment type	The type of manipulation applied to the BioMaterial for the purposes of generating one of the variables under study [MGED Ontology Definition]