

Gabriella Rustici, Anna Farne, Eleanor Williams, Helen Parkinson (version 2, 17/09/2007)

ArrayExpress: data submission

ArrayExpress (AE) accepts submission of all types of array data, microarray designs, and supporting protocols and sample information. The types of microarray data we support include gene expression, ChIP-on-chip, comparative genomic hybridization (CGH) and protein arrays. Both Affymetrix and 2-channel data is supported. Follow this tutorial to submit your own data.

You will learn about:

- How to submit data to the database via web tools
- How to generate templates for large submissions for data submission

Contents:

- 1 How to submit data to the database
- 2 Data submission using MIAMExpress
- 3 Data submission using Tab2MAGE
- 4 Glossary

1 How to submit data to ArrayExpress

There are three ways to submit data to AE:

1. Using MIAMExpress web based submission tool (batch upload and web forms)
2. Using Tab2MAGE spreadsheet submission system
3. Exporting from an external database. This approach is suitable for programmers and will not be included in this tutorial. More information can be found in the AE documentation http://www.ebi.ac.uk/miamexpress/help/pipeline_help.html

For a fully MIAME-compliant submission, using MIAMExpress or Tab2MAGE, you will need to provide the following information:

- Protocols describing your sample growth/treatment and processing steps
- Array design (platform) used and layout information for the design (unless commercial)
- A description of your experiment
- Details of all the samples used in your experiment
- Raw data files (e.g. CEL, gpr, txt) for each hybridization performed
- Normalized/transformed data for each hybridization

Supported files and technology types:

Technology	Data Type	File format	Quantity
Affymetrix	Raw	.CEL and .EXP	1 CEL per hybridization - required 1 EXP per hybridization - optional
	Normalized	.CHP and/or .txt	1 CHP or 1 .txt per hybridization
	Combined data files	.txt	1 .txt per experiment
Other	Raw	.gpr or .txt	1 .gpr or .txt per hybridization - required
	Normalized	.txt	1 .txt per hybridization
	Combined data files	.txt	1 .txt per experiment

The following cases are not supported by MIAMExpress and Tab2MAGE should be used instead:

- Missing raw data files for some or all hybridizations
- More than 1 raw or 1 normalized data file per hybridization (e.g. multiple scans per hybridization, or 2-channel ImaGene data, where you have a separate data file for each channel)
- More than 1 combined data file per experiment
- Nimblegen and Illumina data

2 Data submission using MIAMExpress

The MIAMExpress help pages can be found at <http://www.ebi.ac.uk/miamexpress/help>

2.1 Login

Go to MIAMExpress main page at <http://www.ebi.ac.uk/miamexpress/> and then click the 'Submissions' link to get to the login page (Fig. 1).

Welcome To MIAMExpress

MIAMExpress is a web-based tool for gene expression data submissions to [ArrayExpress](#). For more information, and before starting your submission, please read the [MIAMExpress Help Documentation](#).

Login Name:

Password:

[Try MIAMExpress](#)

[Sign up new submitter](#)

[Forgot your password?](#)

Powered by MySQL

Accession Number Assignment - Current Practice:

Presently, the curation team assigns accession numbers only to Array, Protocol and Experiment submissions that will be immediately released to the public, or are included in accepted papers.

However, if during the paper reviewing process you are asked to provide the reviewers with accession numbers or proof of submission, please contact the [curation team](#).

Arrays, Protocols and Experiments will be finally released into the [ArrayExpress](#) public domain on publication of the paper.

MIAMExpress

[Login](#)

[Help Pages](#)

[Contact Team](#)

***** NEW *****

Try our MIAMExpress Batch Upload Tool from the Experiment Submissions page once logged in.

[HELP](#)

Fig. 1

If this is your first MIAMExpress submission click the 'Sign up new submitter' link to create an account for yourself. Then login to MIAMExpress and select the type of submission you want to make (Fig. 2).

If this is your first time start with Protocols (Section 2.2), otherwise go to Experiment submission (Section 2.4).

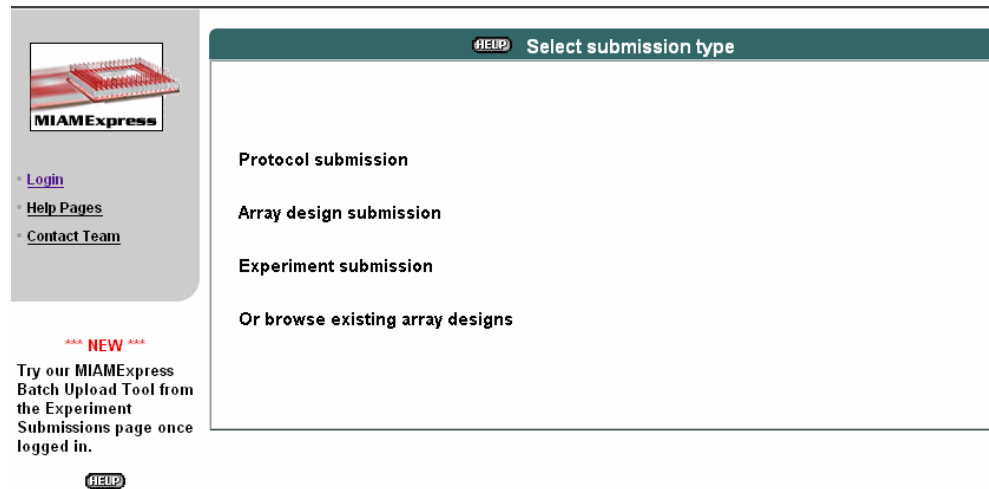


Fig. 2

2.2 Protocol submission

There are several types of protocols you can submit (Fig. 3). Standard Affymetrix protocols for Labeling, Hybridization, Scanning and Normalization are already in MIAMExpress and do not need to be re-submitted.

Change the automatically generated protocol names, like 'SAMPLETREAT12345', to something meaningful, e.g. 'RNA extraction'. You only need to submit your protocols once and you can then reference them for future submissions.

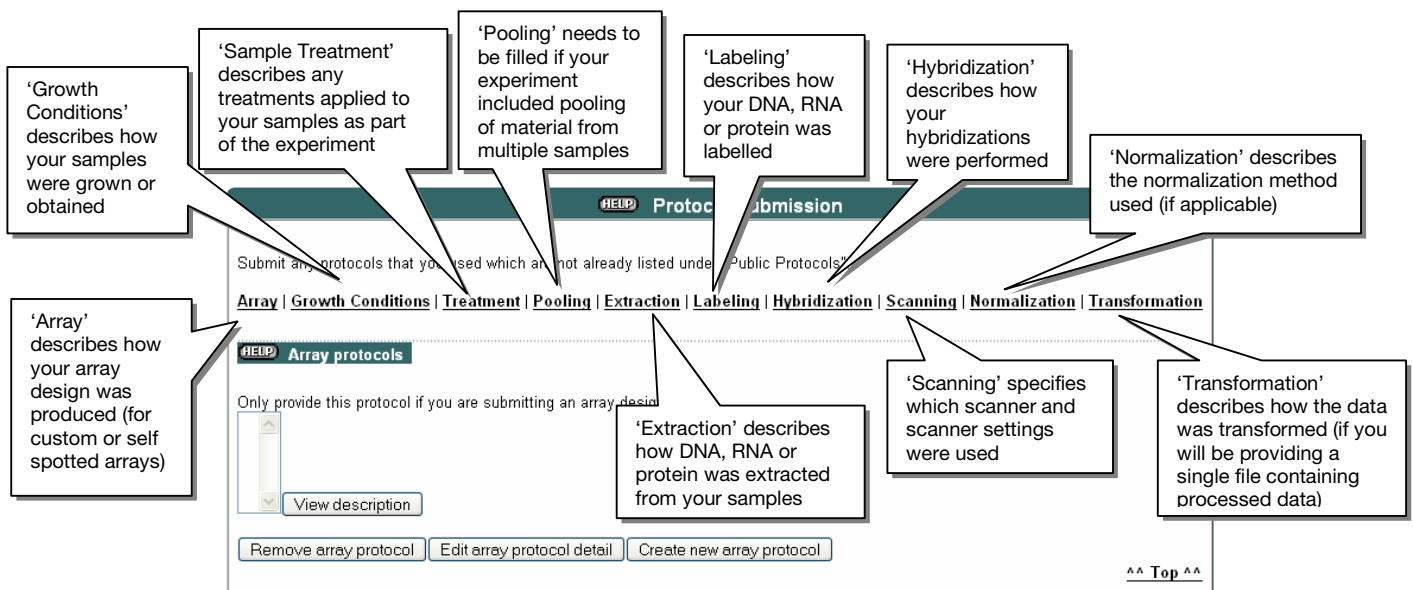


Fig. 3

2.3 Array design submission

After submitting your protocols, go back to 'Submission types' using the navigation link at the top of the page and make an array design submission (Fig. 2). This is needed only if you used a self-spotted or custom array design which has never been submitted to AE. For this tutorial we will assume that you used a commercial array design.

Instructions on preparing and submitting an array design can be found at http://www.ebi.ac.uk/miamexpress/help/array_designs.html.

2.4 Experiment submission using Batchloader

After submitting your protocols or array design, return to 'Submission types' using the navigation link at the top of the page and start the experiment submission (Fig. 2).

The experiment submission is where you first describe your experiment, then samples and hybridizations and finally upload your data files. You will select the protocols that you have already created, the name of the array design used and link these to the data files.

When you start an experiment submission, use the Batch Upload Tool. The Batchloader is a stand-alone Java application which is designed to provide fast submission of microarray experiments to AE. Click on the 'Start MIAMExpress Batch Upload Tool' link and download the application to your machine (Fig. 4).

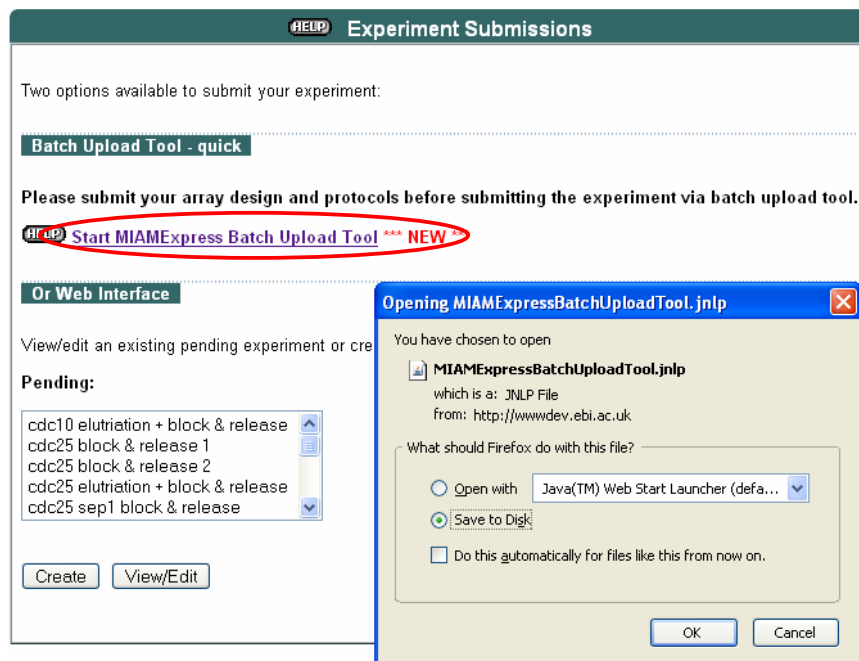


Fig. 4

Once downloaded, double click on the new desktop Batchloader icon and login in using the MIAMExpress user name and password. Then select a new experiment submission.

The Batchloader consists in a series of worksheets or tabs that need to be filled in a specific temporal order, from Experiment->Hybridization (Fig. 5). Errors are checked and you'll be prompted for suggestions. Full instructions on using the Batch Upload Tool can be found at http://www.ebi.ac.uk/miamexpress/help/miamexpress_batchloader.html

You can start adding general details of your experiment in the 'Experiment tab' (Fig. 5). Any field in red is mandatory. Use this tutorial to guide you through the process. Help is available by clicking on the '?' icon.

Experiment tab

Add general information about the experiment, the technology and the variables (Fig. 5).

The screenshot shows the 'Experiment Info' tab in the MIAMExpress Batch Upload interface. The interface is divided into several sections with callout boxes providing instructions:

- Experiment name:** Elutriation 1. Callout: "Enter a unique experiment name".
- Time Series:** . Callout: "Enter details of your publication".
- Experiment design types:**
 - Biological design:** behavior, cell cycle design, cell type comparison, cellular process design. Callout: "Enter details of your samples".
 - Methodology design:** normalization testing, quality control testing, reference, replicate, software variation. Callout: "Create links between extracts and samples, providing protocols for extraction and pooling".
 - Technology design:** CGH, chromatin immunoprecipitation, SNP profiling, tiling path, transcription profiling. Callout: "Link labelled extracts to array designs, upload data files and provide protocols for hybridisation, scanning and data normalisation".
- Experimental factors:** species, strain, temperature, time, other. Callout: "Select experiment design terms from the three available lists".
- Experiment Description:** Cell cycle program in fission yeast using a mechanical synchron... Callout: "Select experimental factors".
- Date for public release:** 01/06/2004. Callout: "Add an experiment description e.g. publication abstract".
- Affymetrix Submission Only:**
 - Affymetrix submission?:** . Callout: "For Affymetrix users, check and select an extraction protocol".
 - Please select extraction protocol:** Pombe RNA extraction.
- Quality related indicators:**
 - Quality Controls:** biological replicates, dye swap, spike. Callout: "Save and go to the next tab".
 - Description:** [Empty text area].

Buttons at the bottom right: Save, Save & Next.

Fig. 5


When completed this tab, you can save and go to the next tab, repeating the same operation for all tabs. In the 'Publication' tab enter the details of your publication, according to its status (in press, submitted, published, etc.).

The 'Samples', 'Extracts', 'Labeled extracts' and 'Hybridizations' tabs allow you to describe all your samples in a spreadsheet-like format, one sample per row.

Samples tab

In the 'Samples' tab (Fig. 6) you must provide a detailed description of each sample, specifying sample attributes (e.g. organism, genotype), experimental factors (e.g. age, sex), and selecting growth and/or treatment protocols previously created via the MIAMExpress web interface (as shown in Fig. 3).

The Batchloader toolbar icons allow to auto-increment values in name column, insert, delete, copy and paste rows as well as obtain a graphical representation of the experiment during submission



?	Name	Organism	Sex	Sex Other	Sample Provider	Sample Type
1	Sample_1	--Select Organism--				other
2	Sample_2	--Select Organism--				other

Fig. 6

Extracts tab

In the 'Extracts' tab you create links between extracts and their samples, providing protocols for extractions and any pooling (Fig. 7). Affymetrix users will skip this tab.

?	Extract Name	Sample Name	Extraction protocol	Pooling protocol
1	Extract1_1	Sample_1	Pombe RNA extraction	===== No Protocol =====
2	Extract2_2	Sample_2	Pombe RNA extraction	===== No Protocol =====

Fig. 7

Labeled Extracts tab

In the 'Labeled extracts' tab, you can create as many labeled extracts as needed, link them to an extract, e.g. if you have dye swap, provide a labeling protocol and specify the dye used. Clicking on the 'visualize' icon (Fig. 8 - red circle) on the Batchloader toolbar produces a node-edge diagram of the experiment so far submitted (Fig. 8).

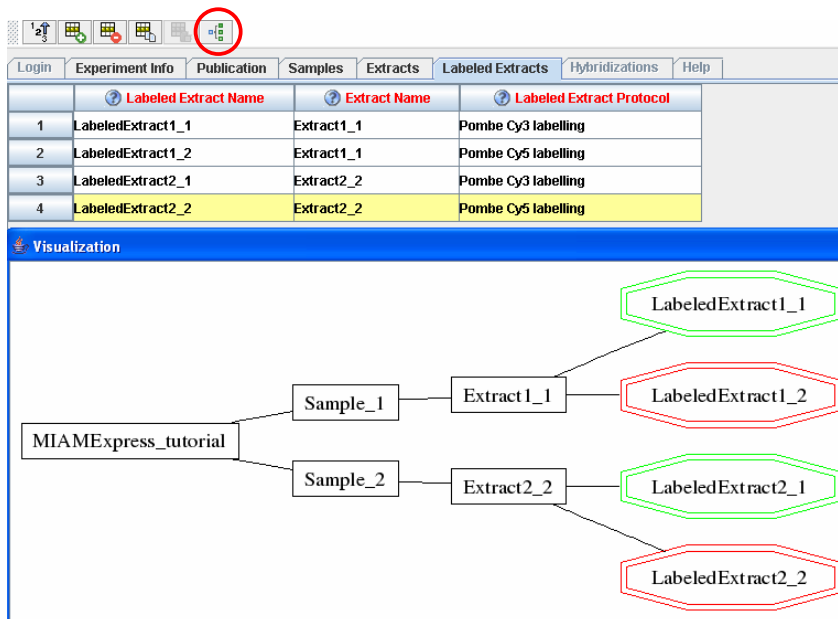


Fig. 8

Hybridisations tab

In the 'Hybridizations' tab, you link labeled extracts to array designs to create the hybridizations, upload data files (raw data files are required; 'per-hybridization' normalized files can be uploaded here if you have them) and provide protocols for hybridization, scanning and data normalization, if 'per-hybridization' normalized files are provided (Fig. 9).

	Hybridization Name	Raw Data File (Paste file path ...)	Optional .EXP file (Paste ...)	Normalized Data File (Pas...)
1	Hybridization_1			
2	Hybridization_2			

Fig. 9

Once you have uploaded the data files, click 'Save' at the bottom of the tab and wait until the data is uploaded onto the EBI server. Click on 'finish' to go to the final part of the submission where you can upload a file combining data from all your hybridizations.

2.5 Complete your submission

If you have generated a data file combining information from all your hybridizations, you can upload it at this stage. If you supply a combined data file, you must link it to the data transformation protocol used (Fig. 10).

Select final gene expression data matrix

Select the protocol

Fig. 10

On completion of a submission you will see a visualization of the experiment and then return to the login tab.

After completing your experiment submission you will be notified by email that your experiment has been submitted. Your submission will then be put into curation status and the curation team will contact you with any questions. An accession number will be sent to you when all the required information has been provided. The experiment is then loaded into AE and you will be provided with a 'reviewer login' for viewing the data before it is made publicly available.

3 Data submission using Tab2MAGE

Tab2MAGE is a format for describing a microarray experiment in a spreadsheet. You can submit a Tab2MAGE spreadsheet to AE, along with your data files, using our Tab2MAGE submission system. More help on making a Tab2MAGE submission can be found at http://www.ebi.ac.uk/miamexpress/help/tab2mage_help.html

3.1 Login

Go to the Tab2MAGE submissions page at <http://www.ebi.ac.uk/cgi-bin/microarray/tab2mage.cgi>

Create a login if this is your first Tab2MAGE submission. Click on 'Experiment list' and then click on 'Create new' to start a new experiment submission.

3.2 Create a template

You'll now generate a template using the web form provided. Select one or more terms from each of the lists (Fig. 11). The template will contain suggested columns for annotation. You can use these, and add more if needed. Click 'Save experiment' at the bottom of the page and move to the next step.

Create experiment

Please enter some summary information for your experiment by

Name:

Affymetrix experiment?

If you wish to have a spreadsheet template to be constructed for you, please select one or more items from each of the following lists. Please note that you can omit this step if you have already created your own spreadsheet. Once you have entered the appropriate information, please click the button at the bottom of the page.

Biological design:	Methodology:	Technology used:
<ul style="list-style-type: none"> Genetic characteristics Genetic modification Growth condition Organism part comparison RNAi Sex Species Strain or line Time series Translational design 	<ul style="list-style-type: none"> All pairs Array platform comparison Dye swap Ex vivo In vitro In vivo Loop Normalization testing Quality control testing Reference 	<ul style="list-style-type: none"> CGH Chromatin immunoprecipitation SNP profiling Tiling path design Transcription profiling
MGED Ontology link	MGED Ontology link	MGED Ontology link
Materials used:	Organisms used:	
<ul style="list-style-type: none"> Cell culture Organism part Whole organism 	<ul style="list-style-type: none"> Mytilus californianus Pan troglodytes Phodopus sungorus Physcomitrella patens Plasmodium falciparum Rattus norvegicus Rhizobium leguminosarum Saccharomyces cerevisiae Schistosoma japonicum Schizosaccharomyces pombe 	

If your organism is not listed, please [contact us](#). Note that you may still upload files without using this template generation tool, by leaving all the selection boxes blank.

Fig. 11

Now click 'Generate template' (Fig. 12). This will include in the template all the terms selected in the previous window. Save the template locally, as a .txt file, and edit in Excel (or similar).

Experiment created

Successfully created experiment "Tab2MAGE_tutorial".

Experiment designs: Time series
Reference
Transcription profiling

Materials used: Cell culture

Species studied: Schizosaccharomyces pombe

If you wish to use the automatically generated spreadsheet template, please click on the button below to open the template in a new window. You will then be able to save the template to your hard drive.

Generate template

Please use the following links to upload your data files and completed spreadsheet.

Upload files Submit experiment Experiment list

Click on 'Generate template' to obtain a spreadsheet template which can be saved on the disk and edited with any spreadsheet program (Excel, etc.)

Click on 'Upload files' to upload your completed spreadsheet and all your data files in one or more zipped archives

Click on 'Submit experiment' to send your submission to the curation team

Click on this button to view a list of the experiments submitted so far

Fig. 12

The Tab2MAGE spreadsheet structure is split into three main sections: **Experiment**, **Protocol** and **Hybridization**.

Experiment section

The 'Experiment section' holds all of the top-level information about an experiment. It consists of two columns; the left column contains a series of predefined row tags, while the right column contains the actual values pertaining to your experiment (Fig. 13).

Experiment section	
domain	ebi.ac.uk
accession	E-EXML-1
quality_control	dye_swap_quality_control
experiment_design_type	strain_or_line_design
name	<your experiment title>
description	<short description of your experiment>
release_date	2004-08-30
submission_date	2004-07-28
submitter	John Falstaff
organization	Windsor Laboratories
publication_title	<your manuscript title>
authors	John Falstaff; Robin Goodfellow
journal	Nature Genetics
volume	12
issue	4
pages	123-456
year	2004
pubmed_id	12345678

Fig. 13

Protocol section

The 'Protocol section' contains only columns for accession ID, name and text of the protocol, with an optional 'parameters' column (Fig. 14).

Protocol section			
accession	text	name	parameters
P-EXML-1	Cells were grown in YPD (1% yeast extract/2% peptone/2% glucose) to an OD600 of approximately 0.8	Yeast growth	growth temperature (degree_C); pH
P-EXML-2	<protocol text>	Yeast cell harvesting	pellet weight (mg)
P-EXML-3	<protocol text>	Cell lysis and RNA prep	
P-EXML-4	<protocol text>	cDNA labeling	
P-EXML-5	<protocol text>	Hybridization	hyb temp (degree_C); hyb volume (uL)
P-EXML-6	<protocol text>	Scanning	
P-EXML-7	<protocol text>	Image analysis	
P-EXML-8	<protocol text>	Normalization	

Fig. 14

Hybridization section

The 'Hybridization section' describes how each sample-hybridization-protocol relates and includes names of the data file which will be supplied with this spreadsheet. The columns can appear in any order.

- Names of materials and processes - these columns are provided so that you can give unique names to each of the materials used in your experiment, including samples, extracts, labeled extracts as well as hybridization and normalization events (Fig. 15).

Hybridization section						
BioSource	Sample	Extract	LabeledExtract	Dye	Hybridization	Normalization
S288C	S288C sample	S288C extract	S288C LE	Cy3	S288C Hyb	S288C Norm
S288C	S288C sample	Reference extract	Reference LE	Cy5	S288C Hyb	S288C Norm
Sigma1278b	Sigma1278b sample	Sigma1278b extract	Sigma1278b LE	Cy3	Sigma1278b Hyb	Sigma1278b Norm
Sigma1278b	Sigma1278b sample	Reference extract	Reference LE	Cy5	Sigma1278b Hyb	Sigma1278b Norm
W303a	W303a sample	W303a extract	W303a LE	Cy3	W303a Hyb	W303a Norm
W303a	W303a sample	Reference extract	Reference LE	Cy5	W303a Hyb	W303a Norm

Fig. 15

- Material types - indicate the type of each material used (e.g. whole organism, total RNA). These are curated terms from a controlled vocabulary (ontology) and will be corrected during curation (Fig. 16).

Hybridization section			
BioSourceMaterial	SampleMaterial	ExtractMaterial	LabeledExtractMaterial
whole_organism	whole_organism	total_RNA	synthetic_DNA
whole_organism	organism_part	total_RNA	synthetic_RNA
organism_part	cell	polyA_RNA	synthetic_RNA

Fig. 16

- Material characteristics - columns are used to provide sample property information e.g. BioMaterial Characteristics [Sex], [Age], [Genotype] etc. The terms in brackets come from a

controlled vocabulary, and a common set of columns are provided by default. Delete if not needed and add as many as needed, if those provided are not useful (Fig. 17).

Hybridization section		
BioMaterialCharacteristics[Genotype]	BioMaterialCharacteristics[Organism]	BioMaterialCharacteristics[StrainOrLine]
CAD1::myc9:TRP1	Saccharomyces cerevisiae	S288C
CAD1::myc9:TRP1	Saccharomyces cerevisiae	W303a
RTG3::myc18:TRP1	Saccharomyces cerevisiae	S288C
RTG3::myc18:TRP1	Saccharomyces cerevisiae	W303a

Fig. 17

- Protocol columns are a way to reference protocols from the Protocol section (Fig. 14). These might vary between samples e.g. samples might be grown differently. If protocols are the same, simply fill down the column (Fig. 18).

Hybridization section				
Protocol[<i>grow</i>]	Protocol[<i>treatment</i>]	Protocol[<i>extraction</i>]	Protocol[<i>labeling</i>]	Protocol[<i>hybridization</i>]
P-EXML-1	P-EXML-2	P-EXML-3	P-EXML-4	P-EXML-5
P-EXML-1	P-EXML-2	P-EXML-3	P-EXML-4	P-EXML-5
P-EXML-1	P-EXML-2	P-EXML-3	P-EXML-4	P-EXML-5
P-EXML-1	P-EXML-2	P-EXML-3	P-EXML-4	P-EXML-5

Fig. 18

- Experimental factors are the variables being investigated, e.g. genotype in an experiment where wildtype vs mutants are assayed. They can be sample properties, e.g. Factor[Genotype], compounds or other treatments, e.g. Factor[Compound]. The values in brackets will be curated. Each hybridization must be associated with at least one factor (Fig. 19).

Hybridization section		
FactorValue[Organism]	FactorValue[Sex]	FactorValue[OrganismPart]
Homo sapiens	male	liver
Homo sapiens	male	kidney
Homo sapiens	female	liver
Homo sapiens	female	kidney

Fig. 19

- Data files and array information - the final class of headings is used to specify the types of data provided, both raw (e.g. .CEL files or .gpr files) and normalized data (e.g. .txt). The accession number (if known) or the name of the array design must be provided as is the Array serial number, if known (e.g. barcode) (Fig. 20).

Hybridization section			
File[<i>raw</i>]	File[<i>normalized</i>]	Array[<i>accession</i>]	Array[<i>serial</i>]
Data1.txt	NormData1.txt	A-EXML-1	244532
Data2.txt	NormData2.txt	A-EXML-1	244533
Data3.txt	NormData3.txt	A-EXML-1	244534
Data4.txt	NormData4.txt	A-EXML-1	244535

Fig. 20

Once the spreadsheet has been completed, click on the 'Upload files' link (Fig. 12) and upload it together with the related data files. The data files can be submitted as one or more zipped archives (Fig. 21). Once the files have been uploaded, the experiment can be submitted using the 'Submit experiment' link on the same page (Fig. 21).

File upload

Uploading files for your experiment "Tab2MAGE_tutorial"

Please upload the following:

- One completed plain-text spreadsheet file containing your experiment annotation
- One or more zipped (.zip) or tar-gzipped (.tar.gz, .tgz) archives containing all your data files.

See the online [Tab2MAGE documentation](#) for help on filling out your spreadsheet. Please also see these [real-world example spreadsheets](#) for tips on specific experiment types.

Supported data file formats are described in the [Tab2MAGE help notes](#).

Please note that all uploaded file archives must have unique names. To replace a previously uploaded spreadsheet, simply rename the file and upload it again using the form below. This will substitute your new spreadsheet for the old one.

As many data file archives may be uploaded as are needed; the form below adds new data file uploads without replacing the old files. *Note:* if two different data file archives contain files with the same names, then when those archives are unpacked on our server, files in the later upload will overwrite files contained in earlier uploads.

Once you have uploaded these files you will be able to finalize your submission.

Upload type	Select file	Files already uploaded
Upload/replace Spreadsheet:	<input type="text"/> <input type="button" value="Browse..."/>	
Upload/add Data file archive:	<input type="text"/> <input type="button" value="Browse..."/>	

Click here to upload the selected files:

If you are not yet ready to upload files, please use the link below to return to the experiment listing.

Callout boxes:
1. Enter the location on your computer of the spreadsheet file (points to the first 'Browse...' button)
2. Enter the location on your computer of data file archive (points to the second 'Browse...' button)

Fig. 21

Your submission will be processed in the same way as described for a MIAMExpress submission.

4 Glossary

BioSource (or Sample)

A biological starting material used in the study, e.g. a mouse, a tumor sample, a bacterial culture, a group of seedlings. Enter a unique name for each BioSource used and remember to give biological replicates different names e.g. 'patient 1 - tumor sample 1', 'treated culture A'.

BioMaterial Characteristics

A characteristic of the BioSource. You can include as many of these columns as you need to describe your BioSources. Each column must have a different category in the column title, e.g. BioMaterialCharacteristics[Sex], and must contain the values for this category, e.g. male.

Experiment

The complete set of hybridizations performed in a study.

Experimental Factor (or Factor Value)

A property that varies between samples and it is important in the interpretation of your data (e.g. time, compound, genotype, etc.). You need a separate FactorValue[*factor*] column for each factor in the experiment, e.g. FactorValue[Compound]. In each FactorValue column enter the value relevant to that particular sample, e.g. doxycycline.

Extract

The RNA, DNA or protein extracted from a Sample. Enter a unique name for each Extract used.

Hybridization

A single array or chip which has one or two LabeledExtracts hybridized to it.

LabeledExtract

RNA, DNA or protein labeled with a particular dye such as biotin or Cy3. You can create multiple LabeledExtracts from the same Extract but remember to give them different names if they were labeled with different dyes, e.g. 'extract A cy3', 'extract A cy5'.

Sample

A biological material used in the study, e.g. a mouse, a tumor sample, a bacterial culture, a group of seedlings. You'll need at least one sample for each condition studied. If your experiment includes biological replicates create a sample for each biological replicate. If your experiment uses a common reference create this as a sample too.

Further reading

1. MIAMExpress help pages: <http://www.ebi.ac.uk/miamexpress/help>
2. Instructions on preparing and submitting an array design can be found at http://www.ebi.ac.uk/miamexpress/help/array_designs.html
3. Full instructions on using the Batch Upload Tool: http://www.ebi.ac.uk/miamexpress/help/miamexpress_batchloader.html
4. Tab2MAGE submission help pages: http://www.ebi.ac.uk/miamexpress/help/tab2mage_help.html

What to do next

Once you have read this tutorial, you might want to test your understanding by trying the related online quiz or reflective tasks. Please see the EBI moodle at www.ebi.ac.uk/training/ for these and other eLearning resources.