

ArrayExpress: Experimental variable

Last updated: 9 November 2017

Laura Huerta lauhuema@ebi.ac.uk

This practical will introduce you to an important concept in MIAME data standards: the experimental variable. We will also cover how this concept is represented in MAGE-TAB format. More information can be found on the following publications:

Brazma et al. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*, 29(4): 365-71

Rayner et al. 2006. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics*, 6(7): 489



Exercise 1

Metadata: MIAME standards & MAGE-TAB format

Scenario

The main goal of MIAME is to make microarray data as useful as possible. MIAME provides the conceptual structure for the representation of microarray data including:

1. general information about the investigation and its design
2. designs of the arrays used in the investigation
3. **characteristics of the samples** used in the investigation
4. experimental and data processing protocols
5. raw and processed (normalized, filtered and/or selected) data.

An important concept in MIAME is that of **experimental variables** that are the principal variables of interest in the investigation. For instance 'time' in time series investigations, 'dose' in dose response investigations, 'compound' in compound treatment investigations, or 'disease' (normal or otherwise) in disease studies. The same investigation may have several experimental variables; for example, compound, dose and time may all be experimental variables in a dose response investigation in which several compounds are added to the samples over a time course.

It is also important to distinguish between **characteristics of the samples** (or sample attributes) that are intrinsic properties of the sample such as 'organism', 'sex', 'age', 'organism part' or 'disease' and the **experimental variable** that is the factor that you are investigating, for example 'compound'.

In this activity you are going to work on annotating all sample metadata following the MAGE-TAB specification. In particular we will focus on collecting all the characteristics of the samples and defining the experimental variable. For that purpose, let's have a look at experiment GSE63569 from Gene Expression Omnibus (GEO) database at NCBI (<https://www.ncbi.nlm.nih.gov/geo/>).

Tasks – Introductory questions

- What is the experiment about?
- Is it a microarray or a RNA-sequencing experiment?
- In which organism is the experiment done?
- What is the biological material used in the experiment?
- How many samples are authors testing?
- What are the groups of samples compared?
- How many biological replicates are there in each group?
- Are there any technical replicates?

Tasks – Sample annotation: characteristics of the samples

Now that we have a better understanding of the experiment we are going to work on annotating the samples with all the relevant information that we can find.

First, let's create a table in a spreadsheet with as many rows as samples. Add an extra row to label each column. The first column will contain the sample identifiers, e.g. GSM1552789, GSM1552790, and so on. Then, let's start describing the samples by adding as many **characteristics** for each sample as we can. You might need to check the publication in order to add more characteristics.

Those are all the characteristics that I was able to find by using the information at GEO and by checking the publication (Hint: supplementary table S3 is very useful!):

- Characteristics [organism] > Homo sapiens
- Characteristics [individual]
- Characteristics [sex]
- Characteristics [organism part]
- Characteristics [cell type]
- Characteristics [karyotype]
- Characteristics [disease]
- Characteristics [genotype]
- Characteristics [clinical information]

Tasks – Sample annotation: the experimental variable/s

Let's define now what is/are the main variables that the authors are studying. We can summarize our sample annotation including all characteristics and experimental variable/s in a table such as this one:

Sample	Sample attributes			Experimental variables		
	Characteristics [organism]			FactorValue [...]		
GSM1552789	Homo sapiens					
GSM1552790	Homo sapiens					
GSM1552791	Homo sapiens					
GSM1552792	Homo sapiens					
GSM1552793	Homo sapiens					
GSM1552794	Homo sapiens					
GSM1552795	Homo sapiens					
GSM1552796	Homo sapiens					
GSM1552797	Homo sapiens					
GSM1552798	Homo sapiens					
GSM1552799	Homo sapiens					
GSM1552800	Homo sapiens					
GSM1552801	Homo sapiens					
GSM1552802	Homo sapiens					
GSM1552803	Homo sapiens					
GSM1552804	Homo sapiens					
GSM1552805	Homo sapiens					