

Gabriella Rustici (v3, 31/03/2009)

ArrayExpress Atlas for beginners

The ArrayExpress (AE) database is a public curated repository of **transcriptomics** data [1]. As of March 2009 the database holds over 7700 separate studies.

The ArrayExpress Atlas is a semantically enriched database of meta-analysis based summary statistics over a curated subset of AE data. It allows exploring in which conditions a gene is expressed or which genes are **differentially expressed** in a particular condition, tissue, cell type, etc. As of March 2009, the Atlas holds approximately 870 experiments, corresponding to 4647 conditions.

This tutorial assumes that you have completed the “ArrayExpress for Beginners” tutorial.

You will learn about:

- The ArrayExpress Atlas
- How to query the Atlas

Contents:

- 1 What is the ArrayExpress Atlas and how to use it
- 2 How to query the Atlas by gene
- 3 How to query the Atlas by condition
- 4 The ArrayExpress Atlas advance searches

1 What is the ArrayExpress Atlas and how to use it

The AE Atlas provides a statistically robust framework for integrating gene expression data across different platforms at a **meta-analytical** level. A simple interface allows the user to identify **differentially expressed** genes in conditions of interest, providing both gene and condition queries.

The Atlas uses data from the AE Warehouse which have been carefully re-annotated for comparability and have had their **gene annotation** updated using Ensembl (<http://www.ensembl.org>) or Uniprot (<http://www.uniprot.org>). This means that information such as GO terms, **HUGO names** (<http://www.genenames.org/>) and other functional annotation have been added, when available. In addition, we are building an **experimental factor** ontology, or **EFO** (<http://www.ebi.ac.uk/microarray-srv/efo>), and using it to structure the data annotations in the Atlas. This ontology is used to enhance Atlas queries, through the EBI Ontology Look-Up service (<http://www.ebi.ac.uk/ontology-lookup/>).

Currently, 3700 **experiments** from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) have been imported into AE; therefore the Atlas also provides a way to integrate gene expression data from the two databases.

For constructing the Atlas, the data is taken as normalized by the submitters/authors. For every **experiment**, the strength of differential expression of genes across conditions across **experiments** is computed and **meta-analytical** statistics across **experiments** are calculated. The statistical models and computations are performed primarily with the aid of the **Bioconductor** package LIMMA [2].



This work is licensed under the Creative Commons Attribution-Share Alike 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

The resulting AE Atlas presents itself as a two-dimensional matrix, akin to a **data matrix**: one dimension (rows) corresponds to genes, and the second dimension (columns) corresponds to conditions, rather than samples. The matrix entries are **p-values** together with a sign, indicating the significance and direction of differential expression.

The AE Atlas user interface can be found at <http://www.ebi.ac.uk/microarray-as/atlas/>.

2 How to query the ArrayExpress Atlas by gene

The Atlas can be queried by gene(s), direction of differential expression, organism and **experimental condition(s)** (Fig. 1). For additional information on all query options go to http://www.ebi.ac.uk/microarray/doc/atlas/faq.html#faq_list_expt_conds.

The screenshot shows the ATLAS query interface. At the top left is the 'ATLAS' logo. Below it are three main input sections: 'Genes', 'Organism', and 'Conditions'. The 'Genes' section has a text input field containing '(all genes)' and a dropdown menu for 'up/down in' set to '(any)'. Below the 'Genes' input is a tooltip: 'Please enter a gene name, synonym, Ensembl or UniProt identifier, GO category, etc.' with an example: 'e.g. ASPM, "p53 binding"'. The 'Organism' section has a dropdown menu set to '(any)'. The 'Conditions' section has a text input field containing '(all conditions)' and a 'Search Atlas' button. Below the 'Conditions' input is a tooltip: 'Please enter an experimental condition or tissue, etc. Start typing and autosuggest will help you narrow down your choice.' with an example: 'e.g. liver, cancer, diabetes'. To the right of the 'Search Atlas' button is a 'hide help' link and an 'advanced search' link. A callout box labeled 'Advanced query interface' points to the 'advanced search' link. Other callouts include 'Query for gene' pointing to the 'Genes' input, 'Select an organism' pointing to the 'Organism' dropdown, and 'Query for conditions' pointing to the 'Conditions' input. At the top right of the interface are links: 'about the project | faq | feedback | blog | services api | help'.

Fig. 1: The Atlas query interface (<http://www.ebi.ac.uk/microarray-as/atlas/>)

More specifically, querying for gene(s) will identify all genes whose annotation matches your query. Your search can be restricted only to genes belonging to a given organism and also by direction of differential expression: only up-regulated ('up in'), only down-regulated ('down in') or both ('up/down in'). The 'Conditions' parameter will identify all **experiments** in which the conditions that match your query appear.

Any of the genes, organism or conditions boxes can be left blank but not all three. You can leave the 'Conditions' box blank. As result, the found genes will be ranked for differential expression across all conditions in the Atlas. The user might choose to specify a particular condition of interest and leave the 'Genes' box blank instead. As result, your condition query will be applied to all genes in the Atlas. We will see an example of 'condition query' in part 2 of this tutorial.

The 'Genes' and 'Conditions' boxes have auto-complete functions. Start typing and a list of suggested terms will appear. To select one of the suggested terms click on it. If you do not want to select any, click on the 'hide suggestions' link at the bottom of the list (Fig.2).

The screenshot shows the ATLAS search interface. The 'Genes' input field contains 'sa', which has triggered an auto-complete dropdown menu. The dropdown lists suggestions grouped by category: gene names (Sacm1l, Sa1, sae1, Sac1, Sac), interpro terms (SAB, SAC3/GANP/Nin1/mts3/eIF-3 p25, Saccharopine dehydrogenase), GO terms (saccharopine dehydrogenase (NAD+, L-glutamate-forming) activity..., saccharopine dehydrogenase (NADP+, L-lysine-forming) activity...), and diseases (Saccharopinuria, Sad1/UNC-like, C-terminal, Saethre-Chotzen syndrome, Saethre-Chotzen syndrome with eyelid anomalies...). A 'hide suggestions' link is at the bottom of the dropdown. The search criteria are: Organism: (any), Conditions: (all conditions). A 'Search Atlas' button is visible. A tooltip on the right side of the dropdown reads: 'database of meta-analysis based ArrayExpress Archive, servicing patterns as well as broader ng genes/samples.' with links for 'show help' and 'advanced search'.

Fig. 2: An auto-complete feature example. The 'Genes' auto-complete function groups the suggested terms together first by gene and then lists general textual matches in alphabetical order (categories include, Interpro term, GO term, disease, etc.).

To familiarise you with the Atlas query interface, we will first perform a simple gene query.

1. Open the Atlas homepage, at <http://www.ebi.ac.uk/microarray-as/atlas/>, in a Web browser
2. In the 'Genes' box, on the left-hand side of the page, type the following gene names, e.g. **saa4 mat1a** (as shown in Fig. 3).
3. In the direction of differential expression drop down menu leave the default 'up/down in' selected
4. Don't select any specie in the 'Organism' dropdown menu
5. Leave the 'Conditions' box, on the right-hand side of the page, blank and click the 'Search Atlas' button.

The screenshot shows the ATLAS search interface with a gene query. The 'Genes' input field contains 'saa4 mat1a' with a tooltip below it that reads 'e.g. ASPM, "p53 binding"'. The search criteria are: Organism: (any), Conditions: (all conditions). A 'Search Atlas' button is visible. A tooltip on the right side of the dropdown reads: 'e.g. liver, cancer, diabetes' with links for 'show help' and 'advanced search'.

Fig. 3: The Atlas query interface (<http://www.ebi.ac.uk/microarray-as/atlas/>); example of a gene query

If your query matches more than one gene, the results will be displayed in a heatmap (Fig. 4). All the genes found are listed in the first column of the heatmap, the conditions go across the columns and the heatmap entries provide the number of times a gene was observed significantly up- or down-regulated in the corresponding condition. The heatmap cell color ranges from red, i.e., up-regulated, to blue, i.e., down-regulated.

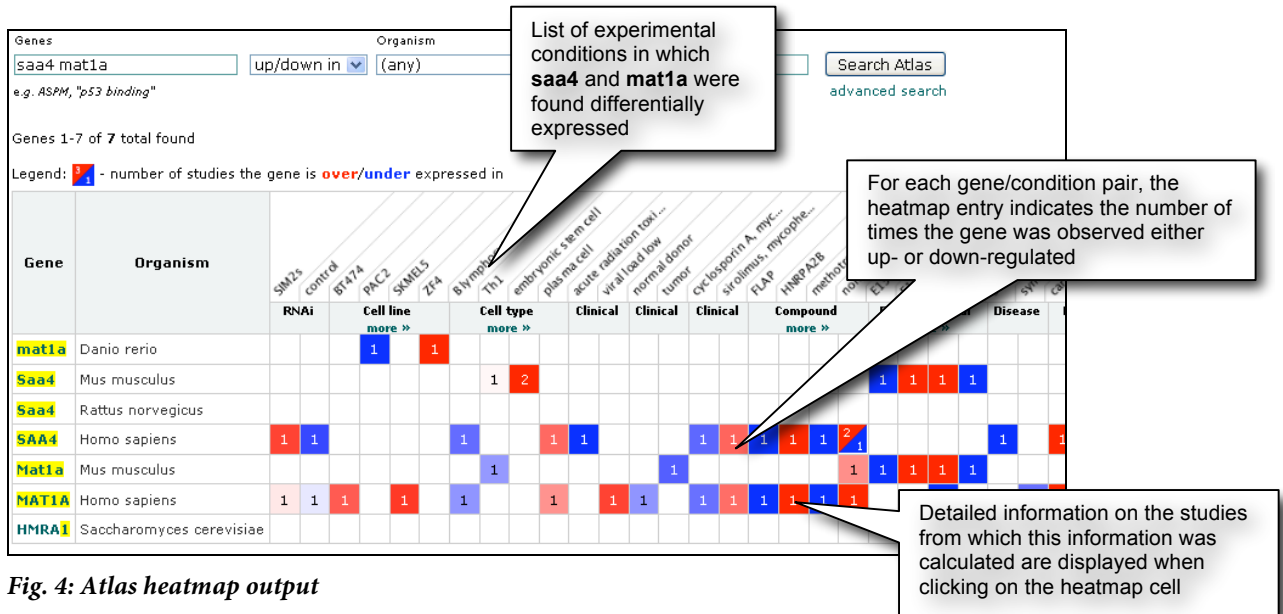


Fig. 4: Atlas heatmap output

Scroll to the right of the heat map to the condition 'Organism Part'. In this example, we see that mouse **saa4** was found up-regulated 13 times in liver samples, and down-regulated 5 times in kidney samples. Similarly, mouse **mat1a** was found up-regulated 10 times in liver samples, and down-regulated 4 times in kidney samples.

Clicking on a cell in the heatmap will show more information about the expression of the gene in the associated experiments. For example, click on the cell with the number 13 in it, corresponding to **saa4** expression in liver samples (Fig. 5).

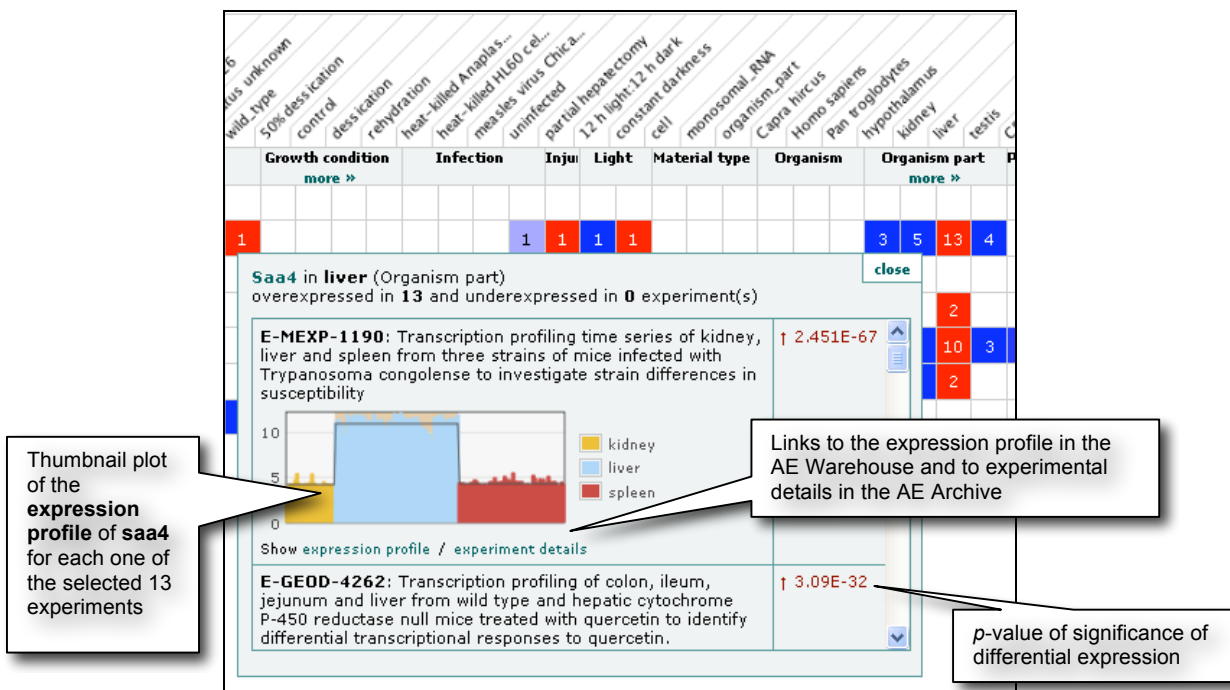


Fig. 5: Detailed view showing the over-expression of **saa4** in liver samples in 13 experiments.

The window that opens up contains detailed information on the 13 experiments in which **saa4** was observed to be up-regulated in liver samples. A thumbnail plot of the **expression profile** of **saa4** in each one of the selected 13 studies is showing, with links to the expression profile in the AE

Warehouse and to experimental details in the AE Archive (Fig. 5). For each experiment, a brief description and the p -value of significance of differential expression are also displayed.

If your search identifies only one gene, or you click on one gene from the gene list returned by a broader search, you will be redirected to the gene view page for that particular gene.

From the heatmap in Fig. 4, click on the mouse *saa4* gene name. The gene view page (Fig. 6) provides a one page summary of all the information present in the Atlas for a given gene.

The following information is displayed:

- summary of terms and external databases cross-references, including synonyms, InterPro and GO terms, Uniprot and Ensembl identifiers;
- expression heat map listing all the conditions in which the gene was observed differentially expressed; and
- thumbnail images of gene expression profiles for the studies in which the gene was found to be differentially expressed.

Saa4
 Synonyms: Saa4, Saa5
 InterPro Term: Serum amyloid A protein
 GO Terms: acute-phase response, extracellular region, lipid transporter activity
 Uniprot: P31532
 Search EB-eye: ENSMUSG0000040017

Expression Summary
 49 factor values, click each to filter

Factor Value	Factor	Up/Down
Liver	Organism part	↑
Kidney	Organism part	↓
Testis	Organism part	↓
Hypothalamus	Organism part	↓
Spleen	Organism part	↓
Colon	Organism part	↓
Ovary	Organism part	↓
Hippocampus	Organism part	↓
E23	Developmental stage	↓
Embryonic stem cell	Cell type	↓
Ileum	Organism part	↓
Jejunum	Organism part	↓
Quadriceps skeletal muscle	Organism part	↓
BA/Ca3	Strain or Line	↓
57BL/6J	Strain or Line	↓
Anterior tibialis	Organism part	↓
Pancreas	Organism part	↓
Submandibular gland	Organism part	↓
Embryo	Developmental stage	↓
Adult	Developmental stage	↓
Partial hepatectomy	Injury	↓
Medial gastrocnemius	Organism part	↓
E13.5	Developmental stage	↓
E15.5	Developmental stage	↓
Skeletal muscle	Organism part	↓
Regulatory cells anti-TCOS-t...	Cell type	↓
Main olfactory epithelium	Organism part	↓
E12	Developmental stage	↓
E8.5	Developmental stage	↓
Wild_type	Genotype	↓
Bmal1-/-	Genotype	↓

Expression Profiles
 38 experiments showing differential expression

E-MEXP-1190: Transcription profiling time series of kidney, liver and spleen from three strains of mice infected with *Trypanosoma congolense* to investigate strain differences in susceptibility

Experimental Factors
 Time | Organism Part | Infection | Strain Or Line

E-GE00-4262: Transcription profiling of colon, ileum, jejunum and liver from wild type and hepatic cytochrome P-450 reductase null mice treated with quercetin to identify differential transcriptional responses to quercetin.

Experimental Factors
 Organism Part | Genotype | Compound Treatment

E-MEXP-114: Transcription profiling of hypothalamus, liver, kidney, o and mice

Experimental Factors
 Sex | Organism | Organism Part

E-MEXP-114 Gene expression profile for (Saa4)
 The currently selected experimental factor is highlighted. To see the expression profile for another factor just click on the factor name.

Click the thumbnail image to see the profile in the AE Warehouse

Fig. 6: Gene view page for *Mus musculus saa4*

The mouse gene *Saa4* shows the highest level of differential expression in the experiment E-MEXP-1190 and it is up-regulated in liver and down-regulated in kidney.

Click on any of the conditions in the expression summary heatmap and you will display only the experiments where this gene is up- or down-regulated in the selected condition. Clicking on the thumbnail for the experiment will take you to the expression profile for the gene in the AE Warehouse (Fig. 6).

3 How to query the ArrayExpress Atlas by condition

We will now perform a simple search querying for a single condition (Fig. 7).

1. Open the Atlas homepage, at <http://www.ebi.ac.uk/microarray-as/atlas/>, in a Web browser
2. Leave the 'Genes' box, blank
3. Leave the default 'up/down in' selected, to include both up- and down-regulated genes
4. Select species, e.g. *Homo sapiens*, in the 'Organism' dropdown menu
5. Type 'cancer', in the 'Conditions' box
6. Select the 'table' option and click the 'Search Atlas' button.

The screenshot shows the Atlas search interface. At the top left is the 'ATLAS' logo in red. To the right are links: 'about the project | faq | feedback | blog | web services api | help'. Below this is a search form with four main sections: 'Genes', 'up/down in', 'Organism', and 'Conditions'. The 'Genes' section has a text input containing '(all genes)' with a small example 'e.g. ASPM, "p53 binding"' below it. The 'up/down in' section is a dropdown menu. The 'Organism' section has a dropdown menu with 'Homo sapiens' selected. The 'Conditions' section has a text input containing 'cancer' with a small example 'e.g. liver, cancer, diabetes' below it. To the right of the 'Conditions' input is a 'Search Atlas' button. Below the 'Search Atlas' button are links for 'show help' and 'advanced search'.

Fig. 7: The Atlas query interface (<http://www.ebi.ac.uk/microarray-as/atlas/>); example of a condition query

When querying for a term like 'cancer', your original query will be automatically expanded using the EFO ontology developed by AE Curation Team (<http://www.ebi.ac.uk/microarray-srv/efo>). The search will now include all child terms available for your original query making the output of your search more comprehensive. For example, in EFO, the term 'cancer' has three children: 'sarcoma', 'chordoma' and 'carcinoma'.

The list of all terms used in the expanded query is shown by clicking on the '+' sign next to the 'Your query has been expanded via EFO,.....' (Fig. 8).

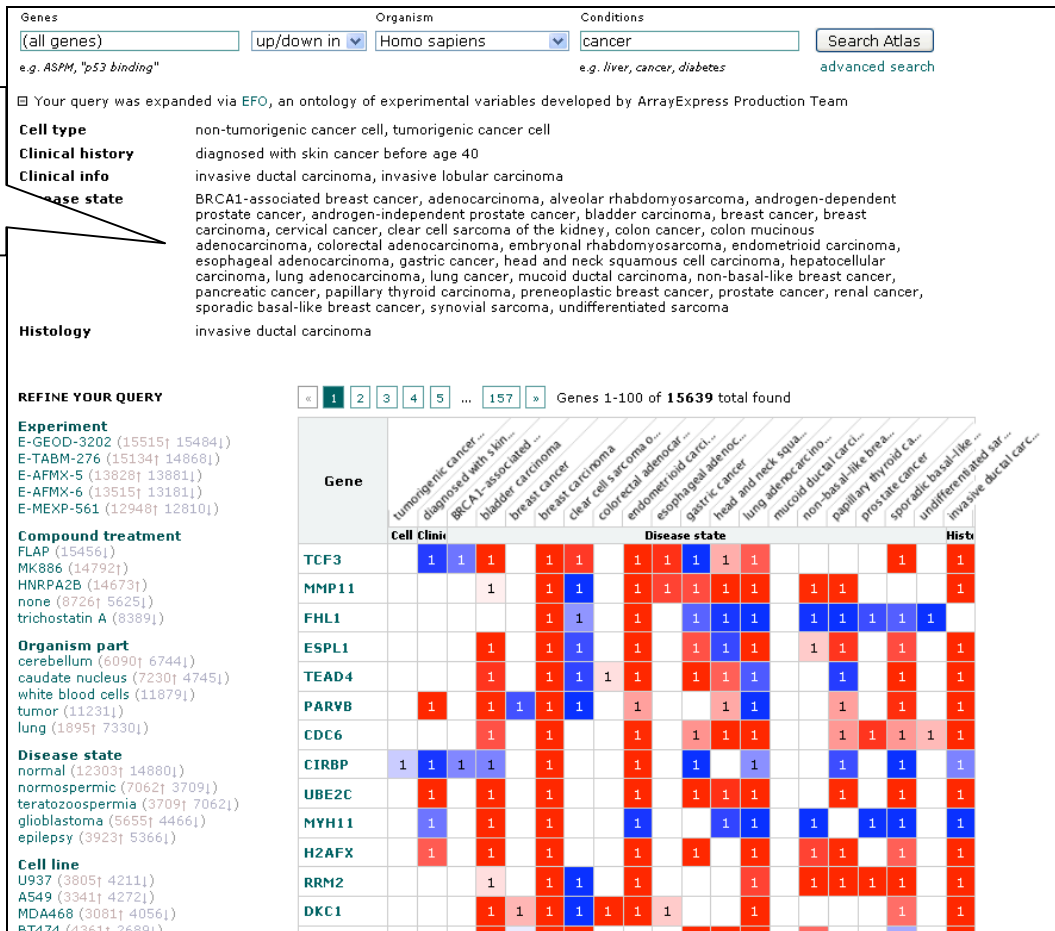


Fig. 8: Example of a condition query using the term 'cancer'. The query was expanded using EFO.

In this case, many genes/conditions are returned and understanding the results can be a daunting task. Filters are available on the left hand side of the heatmap to help you refine your search, by filtering on terms which are enriched in your results. Terms are grouped by category (e.g. 'Compound treatment', 'Organism part', 'Disease state'); for each category, the five most frequently occurring terms are listed.

For example, we could filter the results using the terms 'Leukemia, acute myeloid' and 'Chromosomal rearrangement'. This reduces the number of genes retrieved to just 14. The filters in use are now listed at the top of the page, below the parameters used in the starting query (Fig. 9).

Additional filtering conditions can also be added to the query using the available drop down menus: 'Gene property', 'Experimental factor' and 'Organism' (Fig. 9).

List of all filtering criteria currently in use

Additional filters can be added selecting from the available drop down menus

Find genes matching all of the following conditions

organism: Homo sapiens

is up/down in: any condition

has: gene disease: "Leukemia, acute myeloid"

has: gene keyword: "Chromosomal rearrangement"

Add conditions to the query

Gene property | Experimental factor | Organism

New Query | Search Atlas

simple search

Your query was expanded via EFO, an ontology of experimental variables developed by ArrayExpress Production Team

REFINE YOUR QUERY

Experiment
E-AFMX-5 (141 141)
E-AFMX-6 (141 141)
E-GEOD-3202 (141 141)
E-MEXP-455 (141 141)
E-MEXP-561 (141 141)

Compound treatment
none (121 71)
FLAP (141)
HNRPA2B (141)
MK886 (141)
trichostatin A (111)

Organism part
skeletal muscle (91 61)
caudate nucleus (71 61)
cerebellum (91 41)
blood (71 51)
atrial myocardium (61 51)

Disease state
normal (121 141)
glioblastoma (101 31)
epilepsy (31 91)
chronic lymphocytic leukemia (61 51)
acute promyelocytic leukemia (71 31)

Cell line
MDA468 (81 61)
MCF7 (81 41)
T47D (51 71)
BT-474 (51 61)
BT474 (61 51)

Gene Keyword
Phosphoprotein (11)
Proto-oncogene (10)

Genes 1-14 of 14 total found

Gene	Disease state											Hist
	bladder carcinoma	breast carcinoma	clear cell sarcoma	colorectal adenocarcinoma	endometrial adenocarcinoma	gastric cancer	lung adenocarcinoma	non-basal-like breast carcinoma	papillary thyroid carcinoma	prostate cancer	sporadic basal-like carcinoma	
MLF1	1		1									1
CHIC2		1										1
ETV6	1					1	1					1
SH3GL1	1			1				1				1
ARHGEF12	1		1					1			1	1
LPP	1			1					1			
WHSC1L1	1											1
PICALM	1	1		1	1							1
MLLT10	1											1
NPM1	1								1	1		1
NUP214	1											1
RUNX1	1	1	1		1	1		1			1	1
SEPT9	1			1				1				1
NSD1	1	1						1				1

Processing time: 1.429 secs.

Fig. 9: Example of a condition query using the term 'cancer'; additional filters on the terms 'Leukemia, acute myeloid' and 'Chromosomal rearrangement' have been added selecting from the menu of the left hand side of the heatmap.

4 The ArrayExpress Atlas advance searches

We will now use the Atlas advance search to perform a more complex query.

1. Open the Atlas homepage, at <http://www.ebi.ac.uk/microarray-as/atlas/>, in a Web browser
2. Click on the 'advance search' link, on the right hand side of the page.

Three drop-down menus are available: 'Gene property', 'Experimental factor' and 'Organism' and they can be used to select your searching criteria (Fig. 10).

Find genes matching all of the following conditions

Empty query

Add conditions to the query

Gene property | Experimental factor | Organism

New Query | Search Atlas

simple search

Fig. 10: Atlas advance search interface.

Let us search for genes which are up-regulated in 'hepatocellular carcinoma' AND down-regulated in 'liver' AND down-regulated in 'normal' (Fig. 11)

- From the drop-down menu 'Experimental factor', select 'disease state'. The selected filter will now appear above, under 'Find genes matching all the following conditions'. Restrict the search to the gene(s) which 'is up in' and type 'hepatocellular carcinoma' in the text box on the right (auto-complete function available).
- From the drop-down menu 'Experimental factor', select 'organism part'. This time restrict the search to the gene(s) which are 'down in' and type 'liver' in the text box.
- Finally, from the dropdown menu 'Experimental factor', select again 'disease state' and restrict the search to the gene(s) which 'is down in' and type 'normal' in the text box.
- Click on 'Search Atlas'.

A heatmap of all genes which match these searching criteria is retrieved and can be browsed as previously explained (Fig. 11).

Find genes matching all of the following conditions

is up in	disease state	"hepatocellular carcinoma"	-
is down in	organism part	liver	-
is down in	disease state	normal	-

Add conditions to the query

Gene property Experimental factor Organism simple search

Genes 1-6 of 6 total found

Legend: ■ - number of studies the gene is over expressed in ■ - number of studies the gene is under expressed in

Gene	Organism	hepatocellular carcinoma		
		Disease	normal	liver
Ehd4	Mus musculus	1	2	8
Pitpnm1	Mus musculus	1	1	10
Usp12	Mus musculus	1	2	1
Trim24	Mus musculus	1	1	1
Per3	Mus musculus	1	1	4
Dbp	Mus musculus	1	2	5

Processing time: 1.304 secs.

Fig.11: Advance search output.

4 How to access ArrayExpress Atlas with Web Services

The ArrayExpress Atlas provides a simple Web Services API that allows you to query and retrieve Atlas results programmatically, via SOAP. It provides a more complete data access scenario than the web interfaces currently available.

We will now perform a simple search as above, querying for genes and for conditions. This tutorial will use Java for the Web Services client implementation.

The AE Atlas Web Services WSDL is at <http://www.ebi.ac.uk/microarray-as/atlas/services/AtlasWebService?wsdl>. There are four basic methods there:

- query(String q_gene, String q_expt, String q_orgn, String q_updn)
- batchQuery(String[] q_genes, String[] q_expts, String q_orgn, String q_updn)
- getAtlasGene(String geneIdentifier)
- getAtlasExperiment(String exptAccession)

The main workhorse method of Atlas queries is the “query” method:

```
/**
 *
 * @param q_gene: null | any keyword for a gene (Tp53, apoptosis..etc) (null means all genes (q_expt must be set))
 * @param q_expt: null | any keyword for a condition (e.g. kidney, cancer..etc) (null means all conditions (q_gene must be set))
 * @param q_orgn: any | <Homo sapiens, Mus musculus...etc>
 * @param q_updn: null | up | down
 *
 * Columns (keys) returned from Atlas are as in the Atlas web interface:
 *   ef - experimental factor name (e.g., organismpart, diseasestate)
 *   gene_id - internal gene id
 *   updn_pvaladj - adjusted p-value
 *   efv - experimental factor value (e.g., Cerebellum)
 *   experiment_id - internal experiment id
 *   gene_species - species
 *   experiment_description - experiment description
 *   experiment_accession - external experiment accession number
 *   gene_identifier - external gene identifier (usually Ensembl transcript ID, but sometimes Uniprot)
 *   gene_highlights - internal info why the gene hit occurred
 *   updn - direction of expression change (+1, -1)
 *   gene_name - gene name/synonym
 */
public List<HashMap> query(String q_gene, String q_expt, String q_orgn, String q_updn)
```

Its parameters are fairly self-explanatory, however, there is a lot of control you can exert over the query via the `q_gene` and `q_expt` parameters. In fact, these parameters support a Lucene-like syntax for more complex querying on various attributes of genes and experiments in the Atlas. To find out what attributes genes and experiments have, it is useful to check into the `getAtlasGene` and `getAtlasExperiment` methods. For example:

```
ae3.service.webservices.AnyType2AnyTypeMapEntry[] getAtlasGeneRS = atlasWSproxy.getAtlasGene("ENSG00000148965")
```

This call returns a data structure that looks like this (simplified):

```
{
  'gene_synonym' => 'SAA4 hCG_2039479',
  'gene_ensprotein' => 'ENSP00000278222',
  'gene_refseq' => 'NM_006512',
  'gene_ensgene' => 'ENSG00000148965',
  'gene_experiment' => '188506905 281162231 ...',
  'gene_interproterm' => 'Serum amyloid A protein',
  'gene_goid' => 'GO:0005319 GO:0005576 GO:0006953',
  'gene_interproid' => 'IPR000096',
  'gene_identifier' => 'ENSG00000148965',
  'gene_unigene' => 'Hs.654493',
  'gene_enstranscript' => 'ENST00000278222',
  'gene_id' => '153076970',
  'gene_uniprot' => 'Q6FHJ4',
  'gene_species' => 'HOMO SAPIENS',
  'gene_name_exact' => 'SAA4 SAA4 hCG_2039479',
  'gene_keyword' => 'HDL Acute phase',
  'gene_goterm' => 'acute-phase response extracellular region ...',
  'gene_embldb' => 'BC007026 CH471064 CR541758 M81349 ...',
  'gene_name' => 'SAA4',
  'timestamp' => 'Thu Aug 07 22:21:09 BST 2008'
}
```

Looking at this structure, you should note that any of these fields can be used to query the Atlas. For example, if you have a list of Ensembl protein IDs and you want to retrieve all Atlas information on them, you can do it via a call like this:

```
ae3.service.webservices.AnyType2AnyTypeMapEntry[][] getAtlasResultsRS = atlasWSproxy.query(
    "gene_ensprotein:(ENSP00000278222 ENSP00000356379)", "", "any", "");
```

Using `getAtlasExperiment` you will obtain a similar view into the experiment/condition aspect of Atlas. Here's a brief overview of the data structure that is returned (partial extract here) when executing a call like:

```
ae3.service.webservices.AnyType2AnyTypeMapEntry[] getAtlasExperimentRS = atlasWSproxy.getAtlasExperiment("E-GEOD-3189");
{
  'aer_bdg_bioassay_count' => '69',
  'aer_total_hybs' => '69',
  'dwe_txt_accession' => 'E-GEOD-3189',
  'aer_txt_saac_cat' => 'DiseaseState DiseaseState ...',
  'dwe_ba_diseasestate' => 'malignant melanoma benign nevi ...',
  'aer_txt_bi_publication' => 'Clin Cancer Res',
  'aer_expid' => '1481436957',
  'aer_releasedate' => '2007-08-31',
  'dwe_id' => '313671463',
  'aer_txt_bi_authors' => 'Dmitri Talantov, Abhijit Mazumder, ...',
  'aer_txt_expname' => 'Transcription profiling of human tissue ...',
  'aer_txt_aerdes_type' => 'co-expression disease state',
  'aer_txt_arraydes_name' => "Affymetrix GeneChip\{ae} Human ...",
  'timestamp' => 'Wed Aug 27 13:17:18 BST 2008',
  'aer_txt_fv_factorname' => 'DiseaseState DiseaseState ...',
  'aer_txt_bi_title' => 'Novel genes associated with malignant ...',
  'aer_bi_year' => '2005',
  'dwe_bs_DISEASESTATE' => 'malignant melanoma benign nevi ...'
}
```

Again, you can use any combination of these fields for restricting queries to Atlas over experiments/conditions. For example:

```
ae3.service.webservices.AnyType2AnyTypeMapEntry[][] getAtlasResultsRS = atlasWSproxy.query(
    "gene_ids:ENSG00000097046 gene_name:saa4",
    "aer_txt_saac_cat:DiseaseState",
    "Homo sapiens",
    "up");
```

This query will look for genes defined by the given Ensembl transcript IDs and a gene name, such that they were up-regulated in Human samples experiments that deal with disease state.

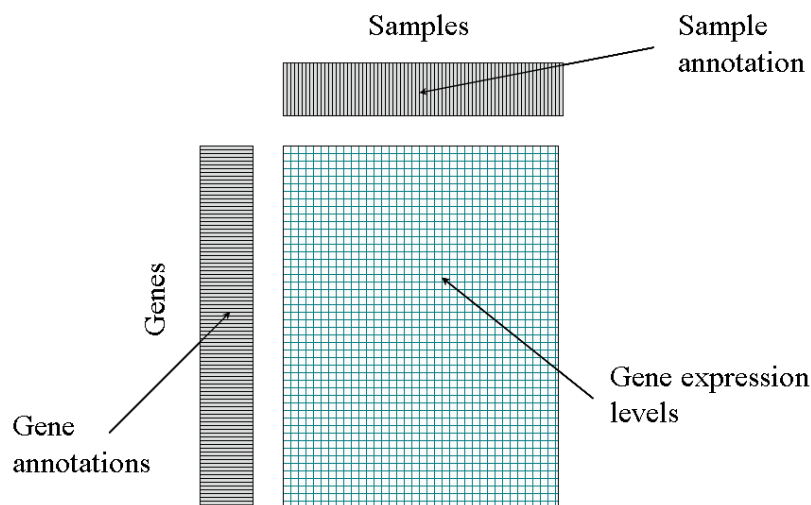
Glossary

Bioconductor

Bioconductor is an open source and open development software project for the analysis and comprehension of genomic data (<http://www.bioconductor.org/>).

Data matrix

In a gene expression **data matrix**, each row represents a gene and each column represents an experimental sample or array. An entry in the **data matrix** usually represents the expression level or expression ratio of a gene in a given sample or array. In addition to numerical values, the matrix can also contain additional columns for **gene annotation** or additional rows for **sample annotation**.



Differentially expressed

A gene is **differentially expressed** when its expression values under two or more conditions are statistically significantly different.

EFO

EFO stands for **Experimental Factor** Ontology and it is an application focused ontology modelling the **experimental factors** in AE (<http://www.ebi.ac.uk/microarray-srv/efo>).

Experiment

The complete set of hybridizations performed in a study.

Experimental factor or condition

A property that varies between samples and it is important in the interpretation of your data (e.g. time, compound, genotype, etc.).

Gene Annotation

Gene annotation is the process of attaching biological information to genes.

Gene expression profile

A **gene expression profile** describes the (relative) expression levels of a gene across a set of **experimental conditions**

HUGO names

The HUGO gene nomenclature committee approves a gene name and symbol (short-form abbreviation) for each known human gene. All approved symbols are stored in the HGNC database (<http://www.genenames.org/>). Each symbol is unique and each gene is only given one approved gene symbol.

Mat1a

methionine adenosyltransferase I, alpha Gene. More information about this gene can be found in Ensembl at http://www.ensembl.org/Mus_musculus/Gene/Summary?g=ENSMUSG00000037798

Meta-analytical

Defines a statistical method that combines the results of a number of different studies in order to provide a larger sample size for evaluation and to produce a stronger conclusion than can be provided by any single study

p-value

The probability of an event or outcome in a statistical experiment

Saa4

Serum amyloid A-4 protein precursor. More information about this gene can be found in Ensembl at http://www.ensembl.org/Mus_musculus/geneview?gene=ENSMUSG00000040017.

Further reading

1. Brazma, A., et al., *ArrayExpress--a public repository for microarray gene expression data at the EBI*. Nucleic Acids Res, 2003. **31**(1): p. 68-71.
2. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Stat Appl Genet Mol Biol, 2004. **3**(12): p. Article3.