### ArrayExpress and Expression Atlas practical: querying and exploring gene expression data at EMBL-EBI

### Last updated: 4 March 2014

### Gabriella Rustici – gabry@ebi.ac.uk

This practical will introduce you to the data content and query functionality of ArrayExpress and Expression Atlas databases. We suggest using Chrome or Firefox for this tuto­rial.

**Additional information on these two resources including dedicated courses and more exercises can be found on the EBI eLearning portal, Train Online.**

**ArrayExpress:** <http://www.ebi.ac.uk/training/online/course/arrayexpress-exploring-functional-genomics-data-ar>

**Expression Atlas:** http://www.ebi.ac.uk/training/online/course/expression-atlas-quick-tour/new-expression-atlas

***Please consider that the results you will obtain while doing the exercises might differ from what illustrated here due to a recent database update.***

**Exercise 1 **

**ArrayExpress - High-throughput sequencing example**

**Scenario**

High-throughput sequencing (HTS) is becoming a popular tool in cancer research to decipher the genetic make-up of a tumour. Mutations, epigenetic misregulation and genomic reorganisation are just some of the things that can be studied using this technology.

Imagine that you have just started a project working on human ‘prostate adenocarcinoma’ and you want to find out all RNA-seq experiments in ArrayExpress that study this cancer.

**Task**

Use the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/experiments/browse.html) to find relevant experiments.

**Exercise 2 **

**ArrayExpress – Effect of sodium dodecyl sulfate on human skin**

**Scenario**

Sodium dodecyl sulfate (SDS) / sodium lauryl sulfate (SLS) is a common ingredient in many household cleaning products. SDS is a known irritant and upon direct topical application, can cause skin inflammation in some individuals, with possible perturbations of gene expression underlying the inflammatory symptoms. You are interested in finding out experiments which studied how SDS affects gene expression in human skin or its *in vitro* models.

**Task**

Use the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/experiments/browse.html) to find relevant experiments.

**Exercise 3 **

**Expression Atlas – Expression in liver (Organism Part search)**

**Scenario**

Imagine that you are working with liver and you are interested in finding out what is the baseline expression of genes in this tissue, in human.

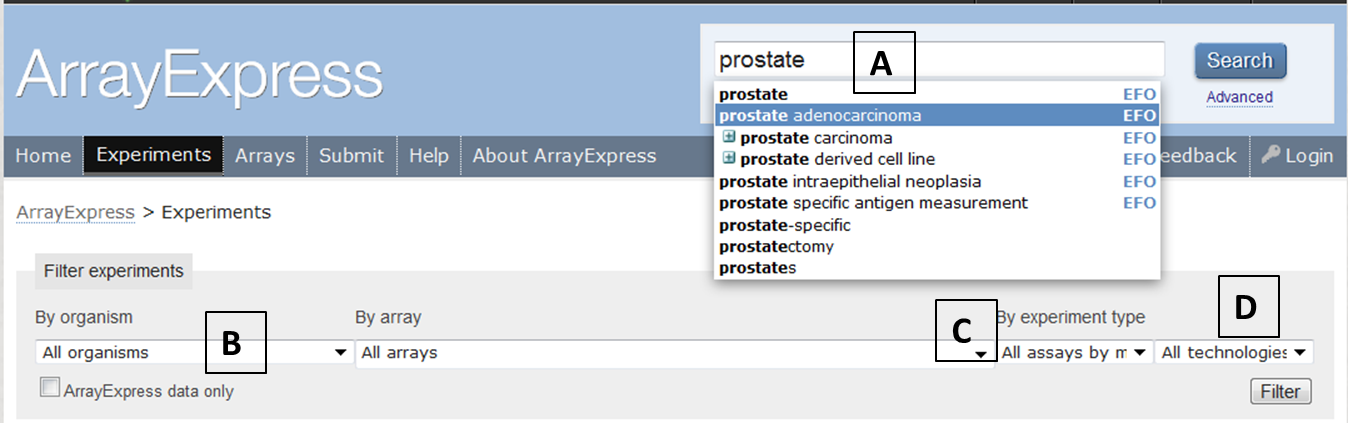
**Task**

Use the Expression [Atlas](http://www.ebi.ac.uk/gxa/" \t "_blank) database (<http://www.ebi.ac.uk/gxa/>) to find information about gene expression in liver.

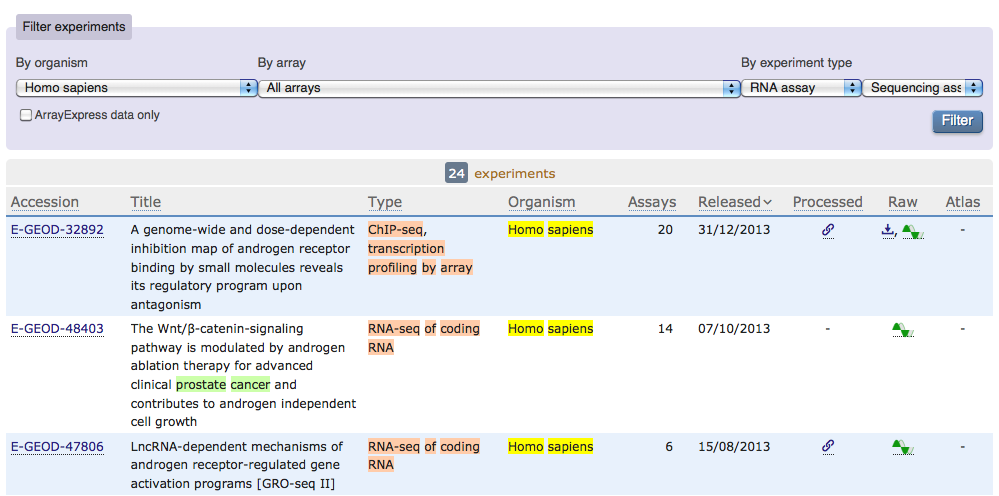
**Need some help?**

**Exercise 1**

1. Open the ArrayExpress experiments page, http://www.ebi.ac.uk/arrayexpress/experiments/browse.html so you can see both the search box and the drop-down filters.
2. Start typing ‘prostate’ in the experiment search box **[A]** and then select the EFO term ‘prostate adenocarcinoma’. Click ‘Search’.

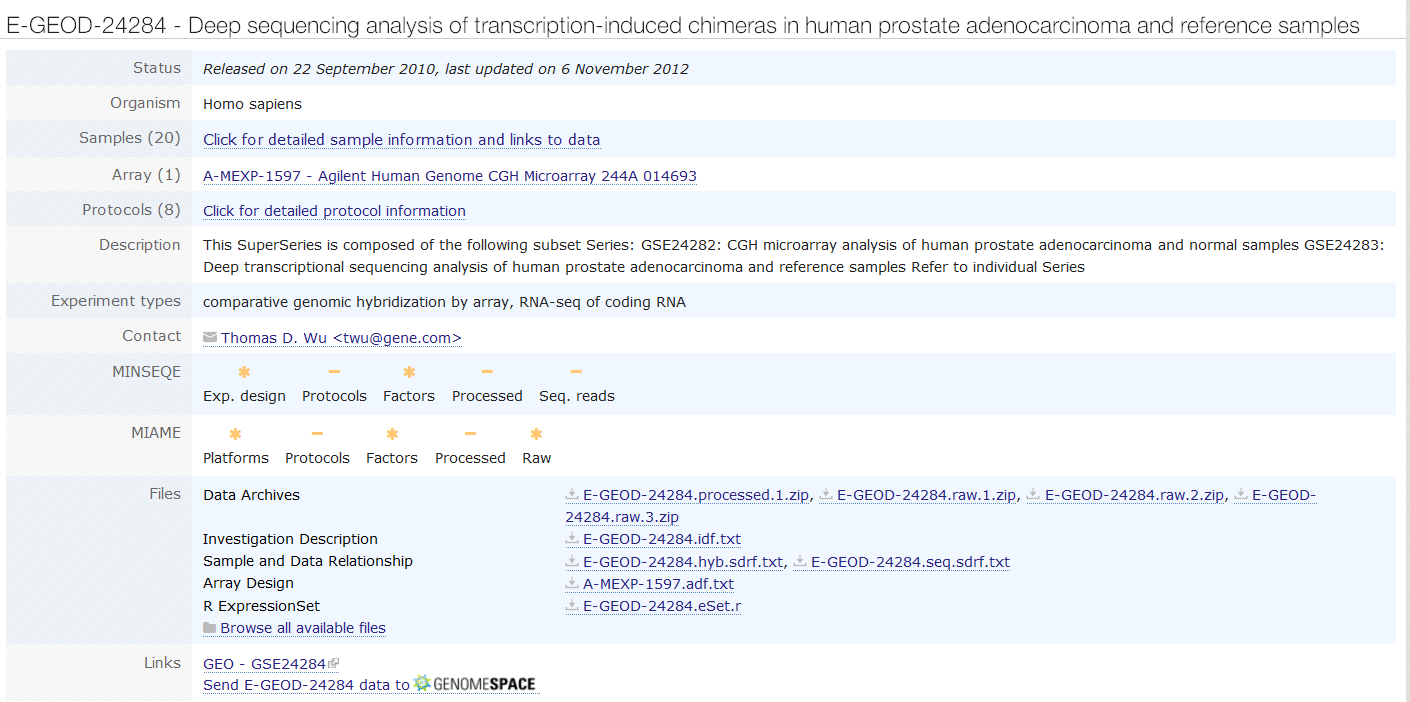


1. Filter your search by organism ‘Homo sapiens’ **[B]**, ‘RNA assay’ **[C]** and ‘Sequencing assay’ **[D],**  then click ‘Filter’. You do *not* need to touch the ‘All arrays’ option as it is only used when you want to filter for experiments done on a specific microarray platform (e.g. Affymetrix mouse 3’ IVT array).
2. The results would look like this:



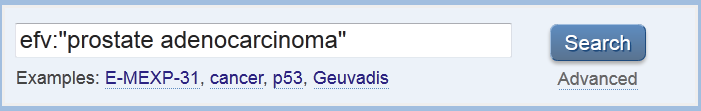
1. You can now look more carefully at individual experiments to identify the one that might be more relevant for your research. Let’s explore E-GEOD-24284.

Click on experiment E-GEOD-24284 and explore the information available for this experiment.



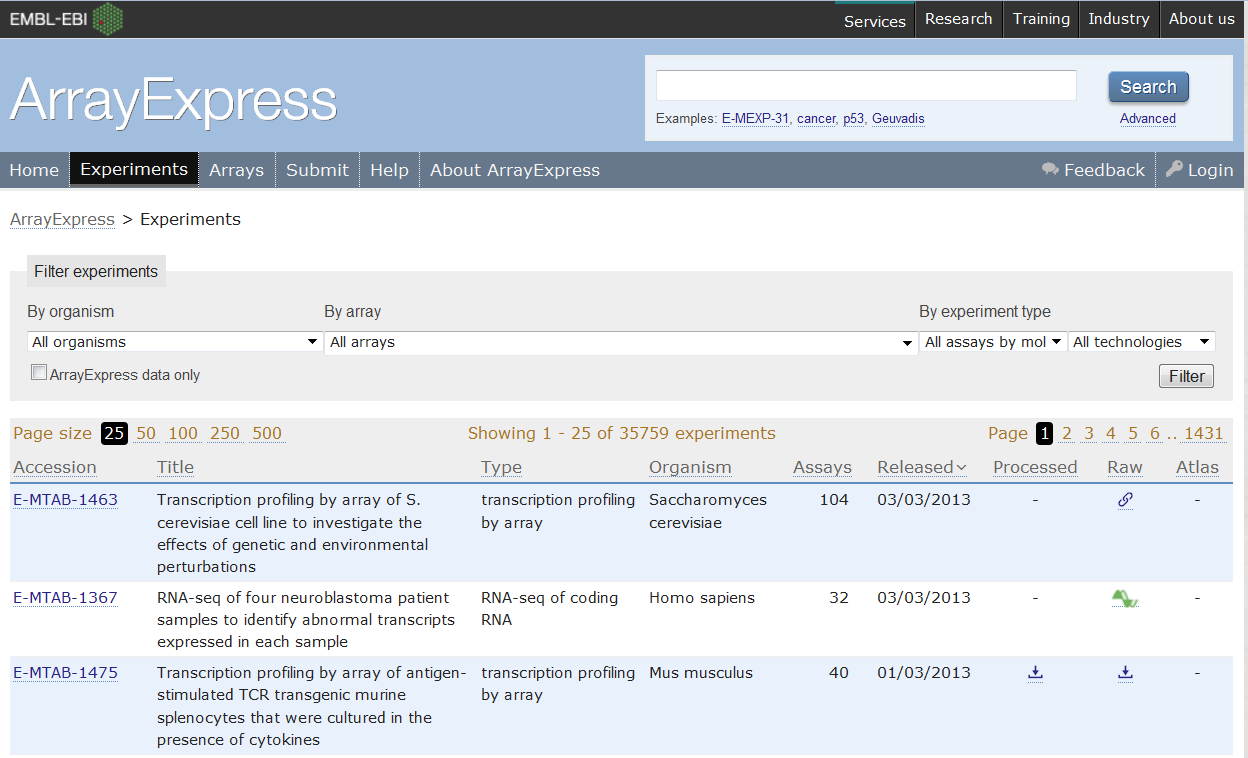
This is an interesting experiment because the investigator performed both microarray and sequencing analyses on the samples. In particular, take a look at the ‘Samples (20) Click for detailed sample information and links to data’ section (circled in red in the screenshot above) to find out more information about the samples analyzed in this experiment and how they relate to the data files.

1. To search specifically for experiments which have “prostate adenocarcinoma” as experimental factor value (e.g. experimental factor = ‘disease’, factor values ‘normal’ and ‘prostate adenocarcinoma’), and not just merely mentioning the disease in the experiment description, try searching with: *efv:’prostate adenocarcinoma’*:

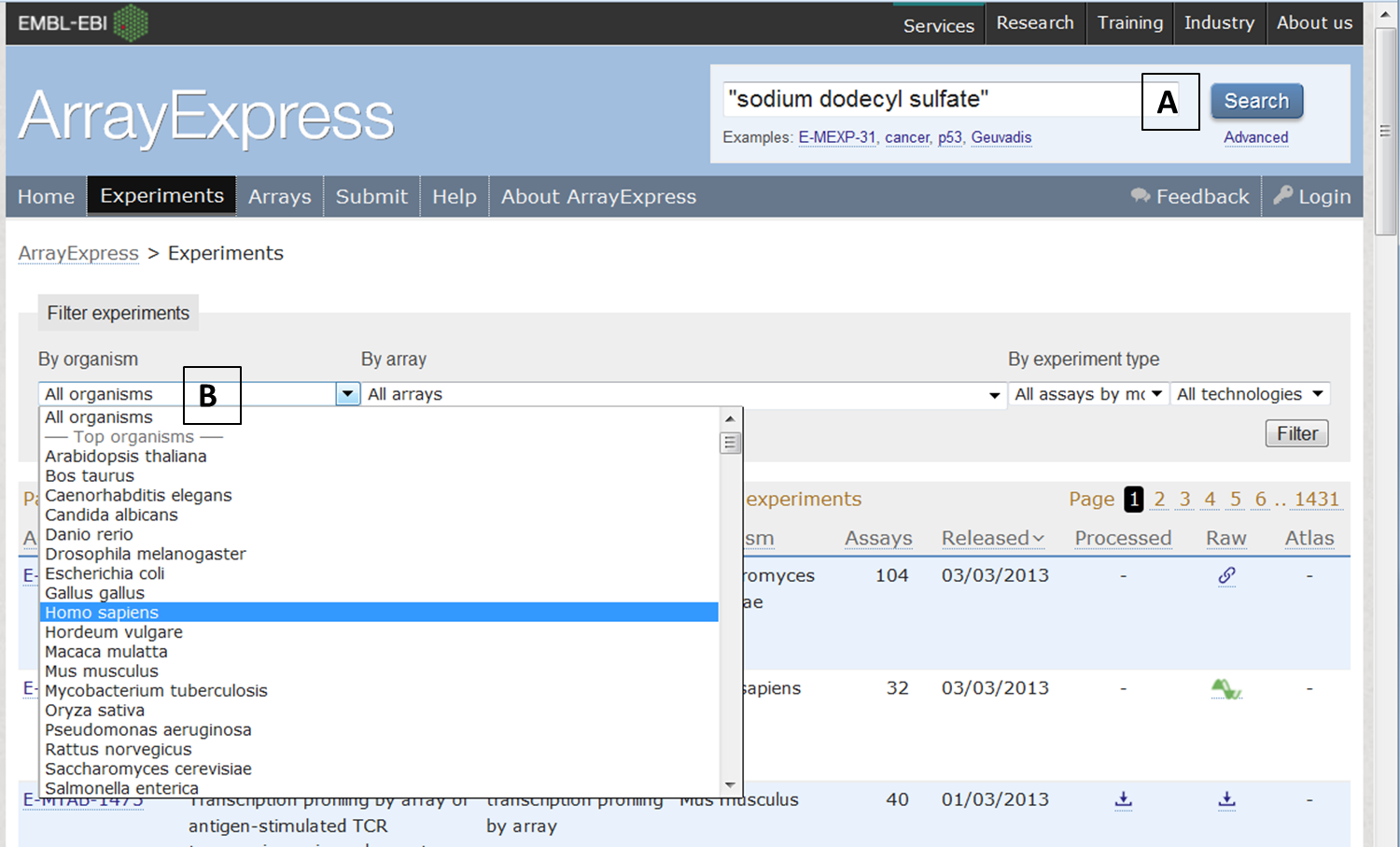


**Exercise 2**

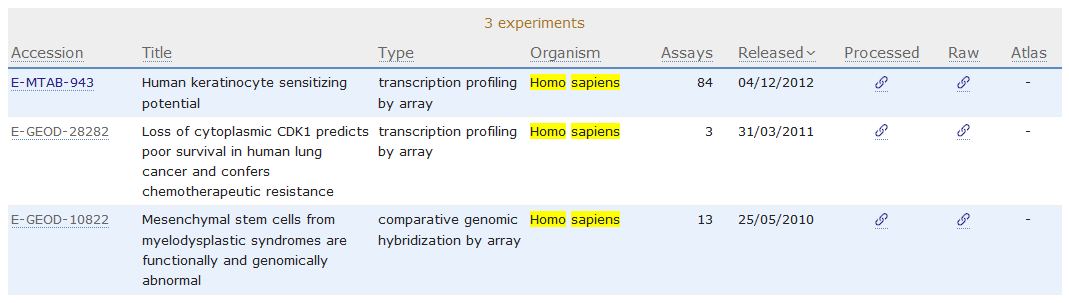
1. Open the ArrayExpress experiments page, http://www.ebi.ac.uk/arrayexpress/experiments/browse.html so you can see both the search box and the drop-down filters.



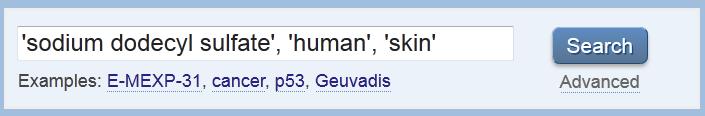
1. Type ‘sodium dodecyl sulfate’ in the search box (there is no auto-completion of this term yet because it is not in EFO), click ‘Search’ **[A].** Make sure you put the search term in quotes so you’re searching for the chemical’s name, rather than ‘sodium’, ‘dodecyl’ and ‘sulfate’ separately.
2. Filter the results by organism, select ‘Homo sapiens’ **[B]**, then click ‘Filter’:



1. Out of the few experiments returned, have a look at each of them to see which one studied the effect of SDS on skin, and which ones merely mentioned the term somewhere in the experiment (e.g. in the description or protocols):



Alternatively, you can find the experiment of interest by typing three search terms to start with: ‘sodium dodecyl sulfate’, ‘human’, ‘skin’:



It should return ‘E-MTAB-943’ only.

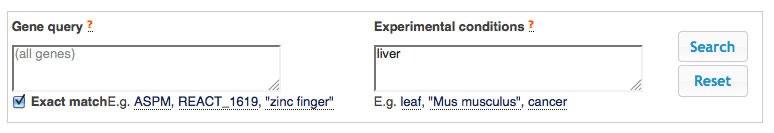
Remember that if your search terms are very specific, you may not get any results at all, so you may find it easier to start with a broader search and then narrow it down by adding more search terms later, or by using filters.

1. Sodium dodecyl sulfate is not yet in Experimental Factor Ontology, so when you searched for ‘sodium dodecyl sulfate’, its synonyms (e.g. SDS, SLS, ‘sodium lauryl sulfate’) or spelling variations (e.g. ‘sodium lauryl sulphate’) are NOT searched automatically.

Try searching for the alternatives terms of ‘sodium dodecyl sulfate’ and see what results are returned.

**Exercise 3**

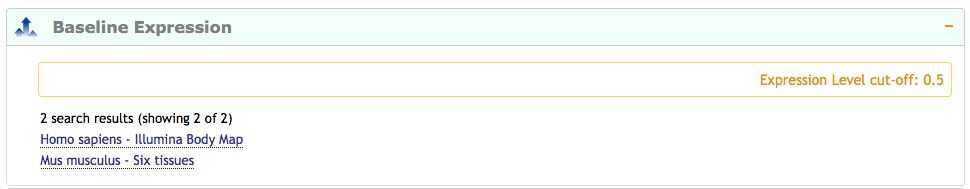
1. Open the Atlas homepage, http://www.ebi.ac.uk/gxa/.
2. Since you are interested in gene expression in liver, you could type, in the “Experimental conditions’ box, ‘liver’ and then click ‘Search’.



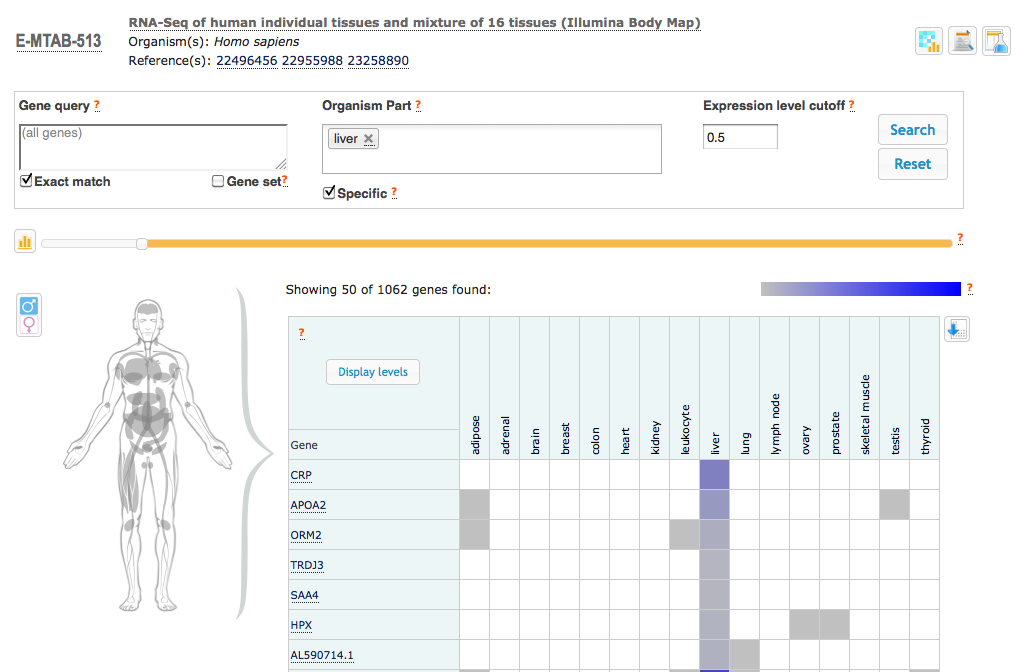
1. The results page contains two tabs: one for “Baseline expression’ and the other for ‘Differential expression’.



1. Expand the ‘Baseline expression’ tab by clinking on it. You will now see that there are two experiments where baseline expression in liver tissue was measured.
2. Click on the ‘Homo sapiens – Illumina Body Map’ link to explore this experiment.

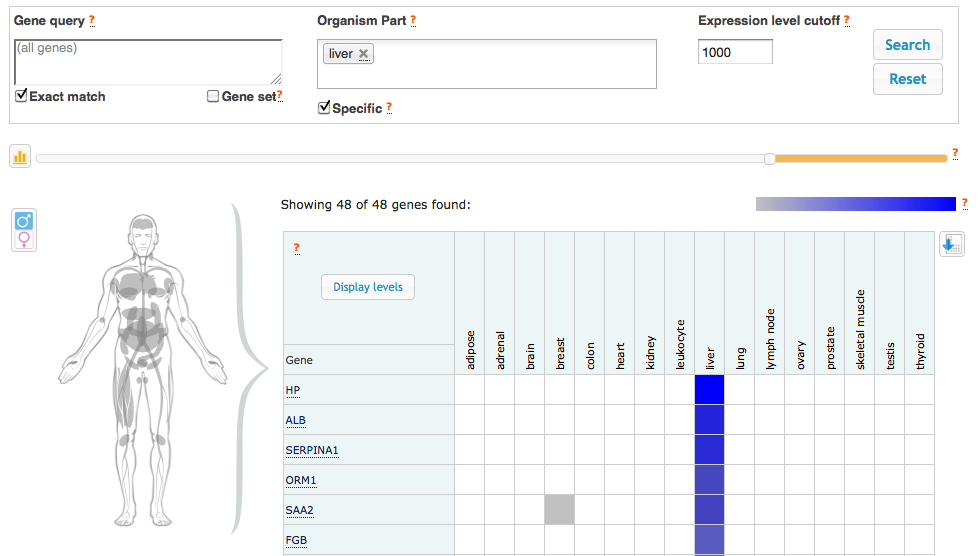


1. The experiment page is now showing 50 genes (out of 1062), which were detected as expressed in liver in the Illumina Body Map dataset.

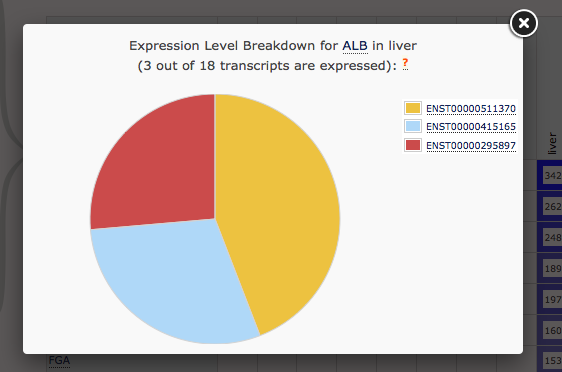


1. You can specify a minimum expression level to act as a cutoff, and only genes expressed above this level are displayed. This is to aid distinction between "background noise" and "real" gene expression. Our default FPKM cutoff is 0.5. To change it, enter a new value into the "Expression level cutoff" input box, or drag the slider until the "Expression level cutoff" input box shows your desired cutoff value.

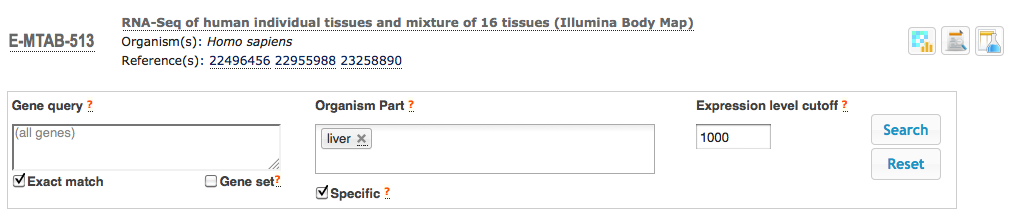
For example, let’s use a cut off of FPKM = 1000 and see how many genes we retrieve expressed above this FPKM level.



1. We identified 48 genes, expressed in tissue liver, above a cut off of 1000. You can show FPKM levels for each gene by clicking on the ‘Display levels’ button.
2. By clicking on a coloured cell in the heatmap, you can also get information regarding the expression of each transcript isoform for a given gene. For example, click on the coloured cell corresponding to ALB expression in liver and you will display information about the expression of the 3 ALB isoforms.



1. Additionally, on the experiment page, you have icons that you can click on to access information about the experiment such as the analysis methods used to process the data and details of the experimental design.



1. Keep exploring the page to see all the functionalities that are available to you.

