EMBL-EBI

ArrayExpress: data mining

Last updated: 9 November 2017
Laura Huerta lauhuema@ebi.ac.uk

This practical will introduce you to the data content and functionality of ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) at EMBL-EBI. More information can be found on the dedicated help page: http://www.ebi.ac.uk/arrayexpress/help/index.html and on the EBI train online course: http://www.ebi.ac.uk/training/online/course/arrayexpress-discover-functional-genomics-data-qui

Exercise 2

Discover interesting datasets to your research area

Scenario

You are working on a research project to improve early detection and treatment of patients with esophageal adenocarcinoma. The strongest risk factor for esophageal adenocarcinoma is the presence of Barrett's esophagus, a condition in which columnar intestinal-type mucosa replaces the normal squamous lining of the esophagus in response to gastroesophageal reflux. Barrett's esophagus without dysplasia (non-dysplastic BE) can progress through the stages of low-grade dysplasia to high-grade dysplasia to invasive esophageal adenocarcinoma.

In order to better understand esophageal adenocarcinoma development, you would like to find if there are any studies in ArrayExpress performing RNA-sequencing at different stages of the Barrett's disease spectrum.
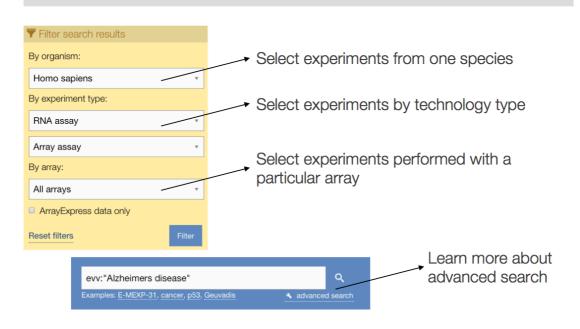
Tasks – Search for experiments in ArrayExpress

You can start your search by typing directly into the Search box or go to **Browse** (http://www.ebi.ac.uk/arrayexpress/browse.html) to combine the **Search** box with the **Filter search results** box.

The option you choose depends on what kind of criteria you would like to use to find interesting experiments. As a general rule, use the search box for attributes not listed in the filter box. For example, use the filters to select data sets from one species, and use

the search box so specify the experimental variable(s) you are interested in. Here you have an example on how to find microarray experiments form human patients with Alzheimer's disease.

1. How many functional genomics experiments are in ArrayExpress related to Barrett's esophagus disease?

2. Is there any of them performed at DNA level, e.g. ChIP-seq or genotyping?

3. How many of them are transcriptomics experiments? What about RNA-sequencig experiments? And microarray experiments?

4. Is there any of the RNA-seq ones studying different stages of the disease?

Tasks – View your favourite experiment

Now that we have found RNA-seq experiments related to Barrett's esophagus disease and in particular one experiment that studies different stages of the disease, let's have a closer look at it.

5. Can you find any information on how the experiment was designed? For example, what are the disease-specific patient cohorts studied?

6. How many samples were analysed in that experiment?

Disease-cohort	Number of samples
normal	
Barrett's esophagus non-dysplastic	
Barrett's esophagus low-grade dysplasia	
esophageal adenocarcinoma	

7. What other information can you find about the samples?

8. Does the experiment involve technical replicates? For how many samples are there technical replicates? Can you find what kind of technical replicates they are?

Disease-cohort	Number of technical replicates
normal	
Barrett's esophagus non-dysplastic	
Barrett's esophagus low-grade dysplasia	
esophageal adenocarcinoma	

9. Are the sequencing libraries single or paired-end? Taking into account the library layout, the number of samples and the number of technical replicates can you calculate how many FASTQ files can be expected for this experiment?

10. What is/are the variable/s of the experiment?

11. What files can you download?

12. Can you find any information about what genes are differentially expressed across Barrett's disease stages?