

ArrayExpress and Atlas practical: querying and exporting gene expression data at the EBI

Gabriella Rustici – gabry@ebi.ac.uk

This practical will introduce you to the data content and query functionality of ArrayExpress Archive and Atlas. We suggest using 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:

<http://www.ebi.ac.uk/training/online/>:

- For the Archive see:
<http://www.ebi.ac.uk/training/online/course/arrayexpress-exploring-functional-genomics-data-ar>
- For the Atlas see:
<http://www.ebi.ac.uk/training/online/course/arrayexpress-investigating-gene-expression-pattern-0>

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 Archive - 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 tumor. Mutations, epigenetic mis-regulation and genomic reorganisation are just some of the things that can be studied using this technology. The results obtained from these experiments will provide a new dimension in the study of cancer biology.

Imagine that you have just started a project working on human “prostate adenocarcinoma” and you want to find out all the experiments in the ArrayExpress Archive that use RNA sequencing assays to study this cancer.

Task

Use the ArrayExpress Archive (<http://www.ebi.ac.uk/arrayexpress/>) to find relevant experiments.

Ed's notes...

Best to use Experiments Archive (the search term suggestions work better here)




The **ArrayExpress Archive** is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. **Gene Expression Atlas** contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.

Experiments Archive

34154 experiments, 988741 assays

Experiment, citation, sample and factor annotations

[Browse experiments](#) | [platform designs](#) | [protocols](#) | [files](#) [Query](#)

[Submitter/reviewer login](#) [ArrayExpress Query Help](#)

Gene Expression Atlas

3596 experiments, 99866 assays, 20862 conditions

Genes: up/down in:

Any species: [Query](#)

[Gene Expression Atlas Home](#)

Then...limit search in the next window:

EMBL-EBI [Find](#) [Terms of Use](#) [Privacy](#) [Cookies](#)

Databases Tools Research Training Industry About Us Help Site Index 

Experiment, citation, sample and factor annotations [clear] Filter on [reset] [Options \[reset\]](#)

adenocarcinomas [Advanced query syntax](#) [All arrays](#) [RNA assay](#) by [Sequencing assay](#) [Query](#)

ArrayExpress data only [Submitter/reviewer login](#) [ArrayExpress Browser Help](#)

Accession	Title	Assays	Species	Date	Processed	Raw	Atlas
E-GEOD-29173	microRNA sequence and expression analysis in breast tumors by deep sequencing (miRNA sequence data)	243	Homo sapiens	2011-06-06	-	-	-
E-GEOD-29580	RNA-seq of two paired normal and cancer tissues in two stage III colorectal cancer patients	4	Homo sapiens	2011-05-27	-	-	✓
E-GEOD-18245	Elucidating the stromal expression pattern in response to tumor-derived Shh	4	Homo sapiens	2010-09-23	-	-	-
E-GEOD-24284	Deep sequencing analysis of transcription-induced chimeras in human prostate adenocarcinoma and reference sam...	14	Homo sapiens	2010-09-22	-	-	-
E-GEOD-24283	Deep transcriptional sequencing analysis of human prostate adenocarcinoma and reference samples	8	Homo sapiens	2010-09-22	-	-	✓
E-GEOD-23776	Alternative expression analysis by RNA sequencing (comparison of 5-FU sensitive and resistant colorectal cancer ce...	2	Homo sapiens	2010-08-25	-	-	-
E-GEOD-14092	An Integrated Network of Androgen Receptor and TMPRSS2-ERG Gene Fusion in Prostate Cancer Progression (II)	60	Homo sapiens	2010-05-17	-	-	-
E-GEOD-20592	Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collecti...	60	Homo sapiens	2010-04-27	-	-	-

29 experiments, 1382 assays.

[Terms of Use](#) [Privacy](#) [Cookies](#) [EBI Funding](#) [Contact EBI](#) © European Bioinformatics Institute 2013. EBI is an Outstation of the European Molecular Biology Laboratory.

The entries with the checks (right hand column) have some EBI analysis already performed.

Expand any entry. The GenomeSpace link will move the data into analysis platforms (like Galaxy). Raw data also available here. This also gives you experimental design. Link to the "Samples" to see what was sampled and how.

[E-GEOD-18245](#) Elucidating the stromal expression pattern in response to tumor-derived Shh

Samples (4) [Click for detailed sample information and links to data](#)

Protocols (6) [Click for all experimental protocols](#)

Description Most tumors of the upper gastrointestinal tract are known to depend on an excessive that in turn produce an unknown set of reciprocal signals that act as growth- or survival-promoting signals. In this study, we have generated a comprehensive set of human pancreatic adenocarcinoma cells were either grown alone, or in the presence of a Shh agonist (SMN2). In addition, 5E1 blocking antibody was added to the cells and after 5 days, RNA was isolated and analyzed by RNA-seq.

Experiment type RNA-seq of coding RNA

Contacts [henk roelink <roelink@berkeley.edu>](mailto:henk.roelink@berkeley.edu), [Maarten F Bijlsma](mailto:Maarten.F.Bijlsma@ebi.ac.uk)

MINSEQE 

Exp. design Protocols Factors Processed Seq. reads

Files [Data Archives](#) [Investigation Description](#) [Sample and Data Relationship](#) [Array Design](#) [Browse all available files](#)

E-GEOD-18245.raw.2.zip, E-GEOD-18245.idf.txt, E-GEOD-18245.sdrf.txt

Links [ENA - SRP002306](#), [GEO - GSE18245](#)
[Send E-GEOD-18245 data to GENOMESPACE](#)

Status Released on 23 September 2010, last updated on 6 November 2012

Exercise 2

Expression Atlas - Regulation of transcription

Scenario

Carcinoma of the prostate is the most frequently diagnosed neoplasm in men in industrialized countries. The androgen receptor (AR), a transcription factor that mediates the action of androgens in target tissues, is expressed in nearly all prostate cancers. During prostatic carcinogenesis, major changes in the androgen receptor pathways occur. Androgen receptor signaling in the nuclei of malignant cells directly stimulates growth of tumor cells.

Imagine that you have a mouse model for human prostate carcinoma, trying to elucidate the role of androgen receptor dependent transcription. You want to find out which mouse genes, annotated as members of the “androgen receptor signaling pathway”, are differentially expressed in prostate carcinoma and which of these are also transcription factors themselves, i.e. involved in regulation of transcription from RNA polymerase II promoter.

Task

Use the Expression Atlas database (<http://www.ebi.ac.uk/gxa/>) to search for such genes

Ed's Notes...

Start here (from the Gene Expression link on the EMBL-EBI home page)

ATLAS [about the project](#) | [faq](#) | [feedback](#) | [das](#) | [api](#) | [help](#)

Genes [?] up/down in Conditions [?] [advanced search](#)

e.g. ASPM, "p53 binding" e.g. liver, cancer, diabetes

returns.....

ATLAS

Genes [?] up/down in Conditions [?] View Heatmap List advanced search

e.g. ASPM, "p53 binding" e.g. liver, cancer, diabetes

Genes 1-25 of **25** total found (you can refine your query) • Download all results • [JSON](#) [XML](#)

Legend:   - number of studies the gene is over/under expressed in (~ in experiment pop-ups indicates non-differential expression)

Gene	Ontology				Keyv
	prostate adenocarcinoma	prostate carcinoma	prostate intraepithelial neopl...	prostate cancer	
Rnf6	1	1	1	1	1
Dnaja1	1	1	1	1	1
Med14	1	1	1	1	1
Ube3a	1	1	1	1	1

Click "refine your query" for suggestions...

ATLAS

Genes [?] up/down in Conditions [?] View Heatmap List advanced search

e.g. ASPM, "p53 binding" e.g. liver, cancer, diabetes

REFINE YOUR QUERY

Zinc finger, RING-type (3)
 Bromodomain (2)
 Helicase, C-terminal (2)
 Helicase, superfamily 1/2, ATP... (2)
 Zinc finger, C3HC4 RING-type (2)

nucleus (21)
 protein binding (19)
 positive regulation of transcr... (16)
 positive regulation of transcr... (15)

Genes 1-25 of **25** total found • Download all results

Legend:   - number of studies the gene is over,

Gene	Ontology				Keyv
	prostate adenocarcinoma	prostate carcinoma	prostate intraepithelial neopl...	prostate cancer	
Rnf6	1	1	1	1	1

The choices in "Refine your Query" column are identifiers that frequently appear among the genes from your previous search. (Choice marked by the arrow is "Positive regulation of transcription by RNA Polymerase II".)

15 genes are returned that are differentially regulated in mouse prostate cancer, that are members of the androgen signalling pathway and that are involved in positive regulation of RNA Pol II.

