Gene Expression

Beginner

0.5 hour

This extended quick tour provides an introduction to Expression Atlas [2], an open resource for finding information about gene expression across species and biological conditions. Expression Atlas aims to help answer questions such as ‘where is a certain gene expressed?’ or ‘how does its expression change in a disease?’.

Learning objectives:

- Gain a basic understanding of what Expression Atlas is and how it works
- Learn how to use Expression Atlas to explore gene expression data
- Know where to find out more about Expression Atlas

What is Expression Atlas?

Expression Atlas [2] is an EMBL-EBI database and web-service that curates, re-analyses and displays gene expression [3] data across species and biological conditions such as different tissues, cell types, developmental stages and diseases.

Expression Atlas aims to help answer questions such as ‘where is my favourite gene expressed?’ or ‘how does its expression change in a disease?’. To achieve this objective, the Expression Atlas project involves data curation, data analysis and provides a web application to access and visualise publicly available data (Figure 1).
Data curation, analysis, search and visualization

All data are manually curated

Expression Atlas contains thousands of selected microarray [4] and RNA-sequencing (RNA-seq [5]) experiments from public repositories such as ArrayExpress [6], European Nucleotide Archive [7] (ENA) at EMBL-EBI and Gene Expression Omnibus [8] (GEO) at NCBI. Controlled access datasets from the European Phenome-Genome [9] Archive (EGA) and the database of Genotypes and Phenotypes [10] (dbGAP) are also selected and included in Expression Atlas. Current criteria for selection and inclusion of a gene expression dataset in Expression Atlas are:

- the study must be of general interest
- it must be performed on a species from which a good quality reference genome build is available
- for microarray data, it must be possible to re-annotate the array design against Ensembl [11]
- the study must include at least three biological replicates
- clear experimental variables must be available

The selected datasets are manually curated by PhD biologists. Curation [12] in Expression Atlas involves a critical review of each dataset to provide a comprehensive representation of gene expression data. We extract and structure information from the literature to enrich the annotation of each sample by adding more metadata [13].

All data are re-analysed using standardised methods
Expression Atlas has re-analysed more than 3,000 experiments. Microarray raw data [14] are analysed using different packages from Bioconductor [15] depending on the array platform used to perform the experiment.

More than 500 RNA-seq experiments have been re-analysed by Expression Atlas. RNA-seq data are analysed using the open source iRAP [16] pipeline, which is available through this github repository [17]. RNA-seq experiments in Expression Atlas include large landmark studies such as GTEx [18], CCLE [19], ENCODE [20] or HipSci [21] (Figure 2).

![Image of various logos including GTEx, CCLE, ENCODE, Gramene, FANTOM5, CCLE, THE HUMAN PROTEIN ATLAS, BL PRINT epigenome]

**Figure 2** Expression Atlas includes large landmark RNA-seq studies.

**Efficient search via ontology-driven query expansion**

Expression Atlas uses ontology-driven search expansion to make your searching more efficient. The terms used to annotate samples are mapped to ontology terms from the Experimental Factor Ontology [22] (EFO). Ontology mappings allow for much richer queries using the hierarchy within the ontology, for example, searching for lung carcinoma will return matches to the keyword and also results for different subtypes such as large cell lung carcinoma (Figure 3).
Easy interpretation of results through heatmap visualisation

Gene expression results are visualised using heatmaps in which gene expression values are converted into a colour-scale image. This provides a visual representation of the expression of multiple genes across different biological conditions (Figure 4A) or of a single gene across different experiments and biological conditions (Figure 4B). We will look at these heatmaps in more detail in the section Searching and visualising data in Expression Atlas.

Figure 3 Expression Atlas provides ontology-driven query expansion.

Figure 4 Visualisation of gene expression results through heatmaps.
What can I do with Expression Atlas?

Expression Atlas can be used to:

- Find in which conditions a particular gene, e.g. UMOD, is normally expressed (baseline results) and in which conditions this gene is differentially expressed (differential results) (Figure 5A)
- Find which genes are expressed in a particular condition, e.g. heart (baseline results) or which genes are differentially expressed, e.g. lung carcinoma (differential results) across experiments in different species (Figure 5B)
- Access analysis results of baseline or differential datasets for one species, e.g. Mus musculus (Figure 5C)
- Access analysis results of large-scale datasets, e.g. GTEx (Figure 5D)

Figure 5 Expression Atlas homepage.

During this course we will cover other functionalities of Expression Atlas such as:

- Exploring gene co-expression across tissues or cell lines
- Visualising biological variation among replicates
- Visualising gene expression in a genome browser
- Identifying Gene Ontology [23], InterPro [24] or Reactome [25] terms that are statistically over-represented in the set of differentially expressed genes
- Downloading gene expression results for each experiment
- Finding annotations for each gene from different resources such as Ensembl [26], UniProt [27], InterPro and Gene Ontology
Expression Atlas: Quick tour
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Searching and visualising data in Expression Atlas

In Expression Atlas there are three options for accessing gene expression analysis results:

- **Search by gene** [28]
- **Search by biological condition** [29]
- **Combined search** [30]

Searching by gene

To search for a particular gene e.g. *UMOD* use the Gene/Gene properties search box on the homepage (Figure 6A). Select one species from the dropdown menu if you want to limit the results to a particular species, e.g. *Homo sapiens* (Figure 6B).

![Search by gene annotations (e.g. gene name symbol, Ensembl gene ID, Gene Ontology ID)](image)

![Select one species from the dropdown menu](image)

**Figure 6** Search by gene name from Expression Atlas homepage.

You can use the following identifiers to search using the Gene/Gene properties search box:

- **Ensembl** [26] gene ID, e.g. ENSG00000169344
- **Interpro** [24] ID, e.g. IPR001507
- **Gene Ontology** [23] ID, e.g. GO:0072233
- Gene Ontology term, e.g. metanephric thick ascending limb development

The search results will show Baseline expression, Differential expression and Gene information tabs. We will look at these tabs in more detail on next few pages.

Baseline expression results

When you search for a gene (e.g. *UMOD*) in one species (e.g. *Homo sapiens*), you will see the Baseline expression results (Figure 7) heatmap. This heatmap displays all tissues studied (columns) in different experiments (rows) in which *UMOD* gene is expressed above the default minimum expression level of 0.5 TPM [31].
Figure 7 Search by gene: baseline expression results.

By default, the heatmap shows baseline expression across different tissues. However, you can use the filters to the left to explore UMOD gene expression in other biological conditions such as cell lines (Figure 7A).

Use the 'Filters' button above the heatmap to limit the results to a particular anatomical system. For example, you can filter the results in the heatmap to show UMOD gene expression only in tissues of the renal system (Figure 7B).

Just click on the title of any experiment in the baseline expression results (e.g. Uhlen's lab, Figure 7) to see the corresponding baseline experiment page.

Baseline experiment page

Each baseline experiment in Expression Atlas has its own Baseline experiment page. If you want to see gene expression results for all genes in the experiment, you have to remove UMOD from the Genes box and click Apply (Figure 8A).
RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

Figure 8 Baseline experiment page.

On a baseline experiment page, expression levels are displayed in a heatmap. The level of expression is relative to the colour intensity, as shown in the gradient bar above the heatmap (Figure 8F). Mouse over a cell in the heatmap to see expression values for each gene in each tissue (Figure 8I). Click on **Select genome browser to view tracks** to view gene expression results in the context of the genomic location of each gene (Figure 8G).

The left section of the experiment page gives you the following options:

- Search for a particular gene (Figure 8A)
- Uncheck the ‘Most specific’ option to show genes with highest expression across all tissues first (Figure 8B)
- Specify a different minimum expression level (the default is 0.5 TPMs) so that only genes expressed above this level are displayed (Figure 8C)
- Switch units between **TPM** [31] and **FPKM** [32] (Figure 8D)
- Select a subset of tissues under ‘Organism parts’ (Figure 8E)

When you search for a particular gene, e.g. **IGLC2**, you will be able to:

1. Visualise variation among biological replicates (Figure 9A). Click on **Switch to boxplot view** to display a plot per tissue with the maximum, upper quartile, median, lower quartile and minimum expression values for each set of biological replicates.
2. Explore co-expression (Figure 9B). Click on **Add similarly expressed genes** button below the heatmap to find genes with similar expression pattern.
Differential expression results

The **Differential expression results** for a particular gene (e.g. *UMOD*) and species (e.g. *Homo sapiens*) show all comparisons in which *UMOD* is differentially expressed (absolute value of $\log_2$ fold-change > 1 and adjusted $p$-value < 0.05). Mouse over the coloured cells to see *UMOD* gene expression values in each comparison (Figure 10). Mouse over the comparison name to display more information: experimental variables are shown in bold along with other characteristics that define each group of samples (Figure 10).
Figure 10 Search by gene: differential expression results.

Just click on the title of any experiment in the differential expression results (e.g., 'Gene array analysis of clear cell renal cell carcinoma tissue versus matched normal kidney tissue', Figure 10) to see the corresponding **differential experiment page**.

**Differential experiment page**

Each differential experiment in Expression Atlas has its own **Differential experiment page**. On this page, differentially expressed genes (rows) in each comparison studied (columns) are displayed in a heatmap by colour intensity, according to the gradient bar above the heatmap (Figure 11F). The heatmap ranks genes by absolute log2 fold-change. Blue cells indicate that the gene is down-regulated while red indicates up-regulated genes.

Mouse over a cell in the heatmap to see log2 fold-change and adjusted p-value (or t-statistic for microarray data) for differentially expressed gene in each comparison (Figure 11I). Click on **Select genome browser to view tracks** to view gene expression results in the context of the genomic location of each gene (Figure 11G). You can download the corresponding data using the buttons at the top right of the page (Figure 11H).

**Figure 11** Differential experiment page.

The left section of the experiment page gives you the following options:

- Search for a particular gene (Figure 11A)
- Uncheck the 'Most specific' option to show genes that are differentially expressed genes in all comparisons first (Figure 11B)
- Select only up-regulated genes, only down-regulated genes or both (Figure 11C)
- Specify different criteria for differential expression (the default is log2 fold-change > 1 and adjusted p-value < 0.05). Only genes satisfying both criteria are displayed (Figure 11D)
- Select a subset of comparisons (Figure 11E)

The differential experiment page has an additional tab called **Plots**. This includes **Enrichment analysis plots** to see if any Gene Ontology [23], InterPro [24] or Reactome [25] terms are significantly over-represented in the set of differentially expressed genes. Just click on the title to display the plot (Figure 12).
Gene array analysis of clear cell renal cell carcinoma tissue versus matched normal kidney tissue

Figure 12 Enrichment analysis results.

Gene information

When you search for a particular gene (e.g. UMOD) in one species (e.g. Homo sapiens), the Gene information tab will display annotations for UMOD gene from Gene Ontology [33], InterPro [34], and Ensembl [11] (Figure 13).
Searching by biological condition

You can also search Expression Atlas by using the **Biological condition** search box on the homepage. This type of search helps you find genes that are expressed in a particular condition, e.g. lung carcinoma (Figure 14A). You can limit the results to a particular species by selecting a species from the dropdown menu (Figure 14B). Your search is expanded using the Experimental Factor Ontology [35] (EFO), so that it will also return synonyms and child terms of lung carcinoma (Figure 14).

**Figure 13** Gene information in Expression Atlas.

**Figure 14** Searching by Biological condition from the Expression Atlas homepage.

When searching by biological condition, you will see both baseline expression (Figure 15A) and differential expression results (Figure 15B). Just click on the title of any experiment in the:

1. **Baseline expression results** (Figure 15A) to see the corresponding baseline experiment page [36]
2. **Differential expression results** (Figure 15B) to see the corresponding differential experiment page [37].

**Figure 15** Search by Biological condition: baseline and differential expression results.

Combined search
You can build combined searches by using both the Gene/Gene properties search box and the Biological condition search box on the homepage. For example, to find out if HOXB13 is differentially expressed in any cancer (Figure 16A) search for 'HOXB13' AND 'cancer' and look at the differential expression results (Figure 16B).

![Combined search example](image)

**Figure 16** Combined search: is HOXB13 gene differentially expressed in any cancer?

### Getting data from Expression Atlas

#### Downloading baseline and differential expression results

You can download baseline expression results by clicking on the Download button above the heatmap (Figure 17A) and differential expression results by clicking on the Download results button above the results table (Figure 17B).
Figure 17 Download baseline and differential expression results.

**Downloading results from an experiment page**

Use the **Downloads** tab on the **baseline experiment page** (Figure 18A) to download:

- Gene expression results in tab-delimited format
- A file containing the R object representing the experiment
- The results of hierarchical clustering [38] using the top 100 most variable genes across all tissues (or other condition) in the experiment

Use the **Downloads** tab on the **differential experiment page** (Figure 18B) to download:

- RNA-seq raw counts or normalised microarray intensity data
- All statistical analysis results for all comparisons in the experiment
- A file containing the R object representing the experiment
Figure 18 Download results from baseline and differential experiment page.

Get help and support on Expression Atlas

Support

- See if your question has been already answered in the Expression Atlas FAQs [39].
- If you cannot find the answer to your question, contact arrayexpress [40]-atlas [at] ebi.ac.uk (Expression Atlas helpdesk) or use the EBI support form [41] - selecting 'Expression Atlas' as the subject.
- For source code, feature requests and bug reports, see the Github repositories [42].

References


Related courses on Train online

- ArrayExpress: Discover functional genomics data quickly and easily [44]
- Functional genomics (I): Introduction and designing experiments [45]
- Functional genomics (II): Common technologies and data analysis methods [46]
- Functional genomics (III): Submitting your data and functional genomics databases [47]

Collaborators
Funding

- European Molecular Biology Laboratories (EMBL)
- WellcomeTrust [108437/Z/15/Z and WT101477MA]
- National Science Foundation (NSF) - Gramene [NFS IOS#1127112]
- Chan Zuckerberg Initiative
- Open Targets

Your feedback

Please tell us what you thought about this Quick tour. Your feedback is invaluable and helps us to improve our courses and enhance your learning experience.

Contributors

Laura Huerta [1]
EMBL-EBI
Senior Scientific Curator - Petryszak team: Gene Expression

Laura works as a Senior Scientific Curator to improve the content of functional genomics data in Expression Atlas. She is involved in developing and implementing metadata standards with a particular interest on data integration through ontology annotation. She manages training activities, delivering courses on functional genomics resources at EMBL-EBI and worldwide. She also interacts with software developers to improve user experience of Expression Atlas website. Laura joined EMBL-EBI in 2015 after receiving her PhD in Molecular Biology from the Polytechnic University of Valencia, focused on studying plant development regulated by hormones through the generation and analysis of transcriptomics data.

Source URL: https://www.ebi.ac.uk/training/online/course/expression-atlas-quick-tour-1

Links
[1] https://www.ebi.ac.uk/training/online/trainers/lauhuema
[2] https://www.ebi.ac.uk/gxa/home
[3] https://www.ebi.ac.uk/training/online/glossary/gene-expression