

EMBL-EBI: Exploring genomic, protein and chemical data, Oviedo

Expression Atlas service at EMBL-EBI

Laura Huerta, PhD

Scientific Database Curator, Gene Expression Team

lauhuema@ebi.ac.uk

17 May 2017





UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

A bit about your speaker...



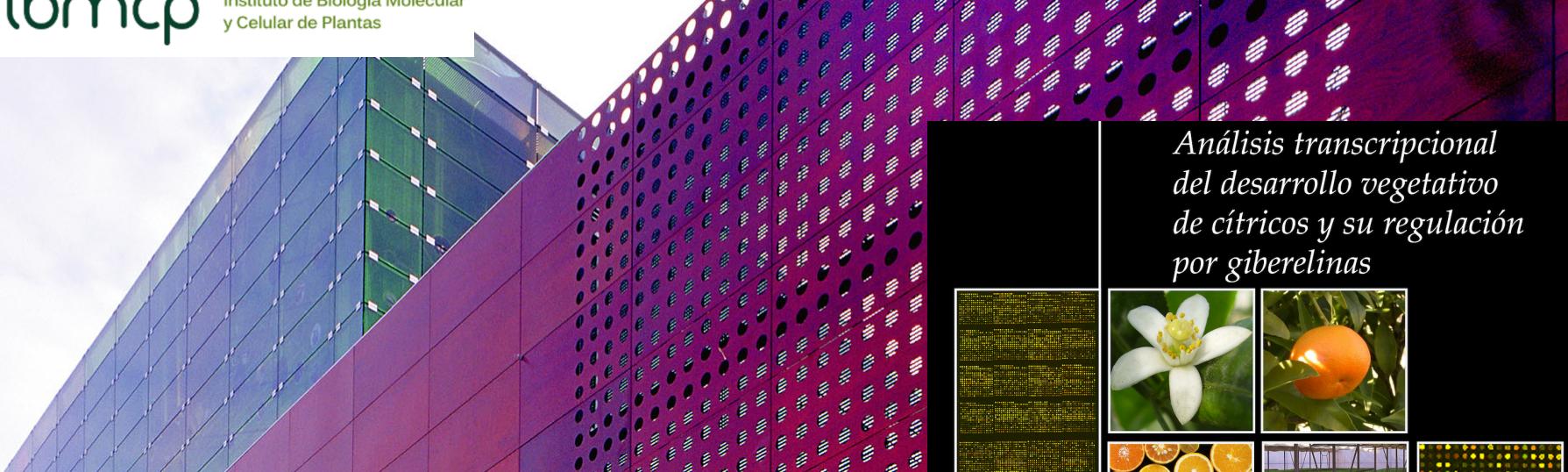
- *Agronomic Engineer, major in Biotechnology*

- *Cloning and characterisation of a GA 20-oxidase gene from Citrus sinensis (L.) Osbeck. Prof. José Luis García Martínez*

A bit about your speaker...

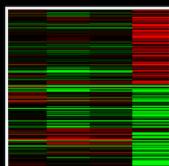
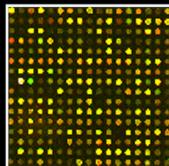


Instituto de Biología Molecular
y Celular de Plantas



<http://www.ibmcp.csic.es>

*Análisis transcripcional
del desarrollo vegetativo
de cítricos y su regulación
por giberelinas*



TESIS DOCTORAL
Laura Huerta Martínez

A bit about your speaker...





WELLCOME
GENOME
CAMPUS





Outline of the session

- ✓ What is Expression Atlas?
- ✓ Gene expression data
- ✓ Data curation
 - ✓ Hands-on exercise
- ✓ Data Analysis
- ✓ Expression Atlas: baseline
 - ✓ Hands-on exercise
- ✓ Expression Atlas: differential
 - ✓ Hands-on exercise
- ✓ Expression Atlas data integrated in other resources

Data resources at EMBL-EBI

Genes, genomes & variation

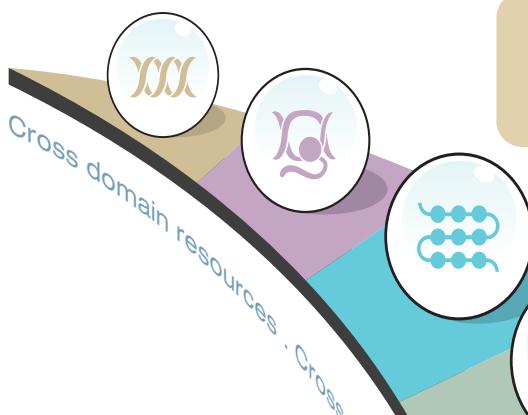
European Nucleotide Archive
European Variation Archive
Europea



Ensembl
Ensembl
Gene-phenome Archive
Genomes



GWAS Catalog
Metagenomics portal



Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor Ontology

Reactions, interactions & pathways

Intact
Reactome

MetaboLights

Gene, protein & metabolite expression

RNA Central
ArrayExpress

Expression Atlas



Metabolights
PRIDE

Protein sequences, families & motifs

InterPro
Pfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank



Chemical biology

ChEMBL
ChEBI



Systems biology

BioModels
Enzyme Portal
BioSamples

EMBL-EBI





What is Expression Atlas?

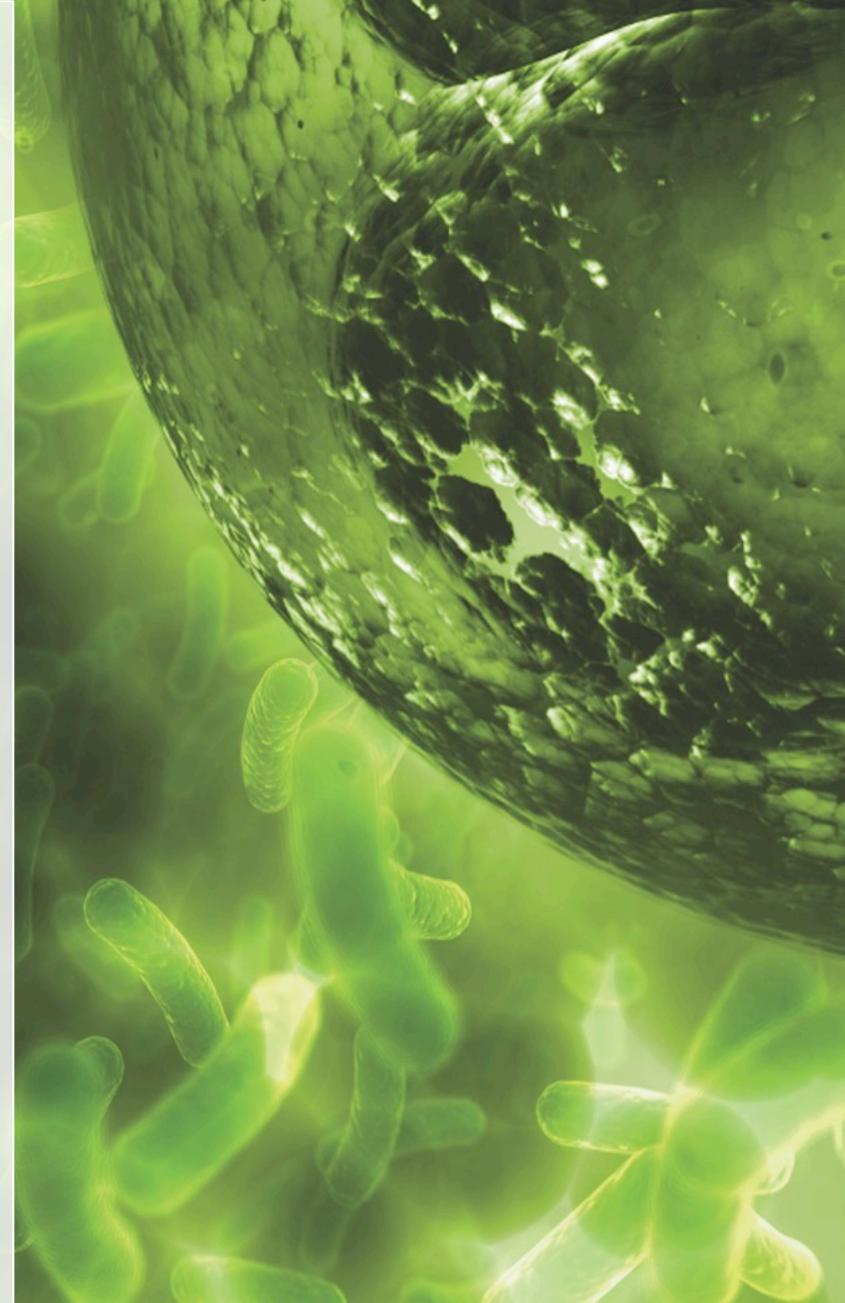


Reference resource for gene/protein expression
across species and biological conditions

www.ebi.ac.uk/gxa/home

OUR MISSION

To provide to the scientific community **freely available** information on the **abundance and localisation of RNA** (and proteins) across species and biological conditions





What is Expression Atlas?

Big Data



Generate and publish gene expression data



Big Analysis



Big Results



Where is my favourite gene expressed?

How its expression changes in a disease?



How ‘big’ is Expression Atlas data analysis?

... e.g. GTEx dataset



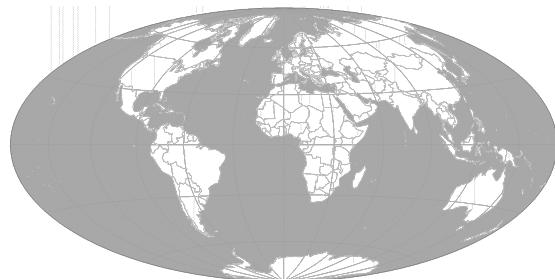
550 donors
53 tissues

19,000 paired
RNA-sequencing libraries

22 years
of processing time in a
single machine

78 terabytes
of EMBL-EBI storage capacity

Expression Atlas contains thousands of selected and curated datasets



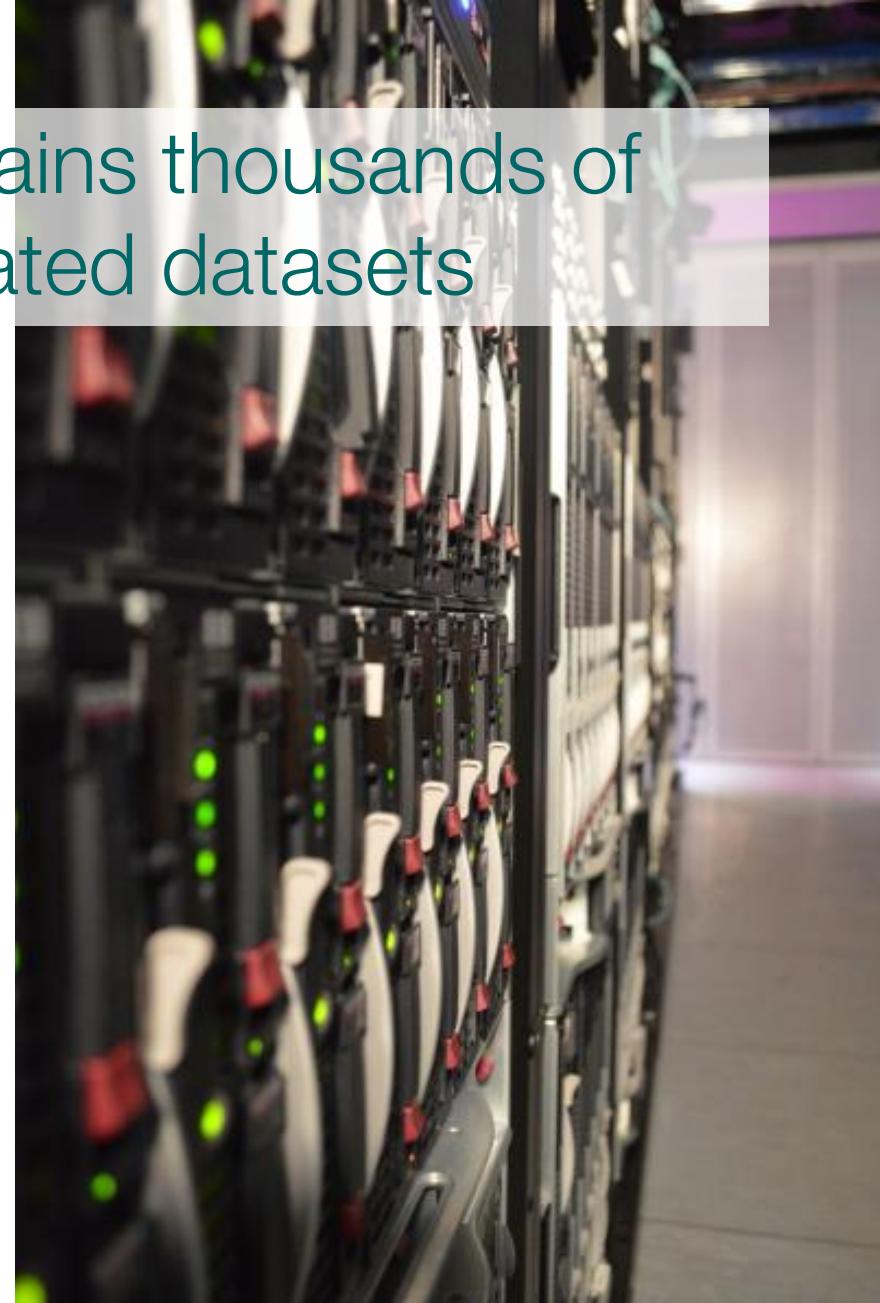
> 3,000 datasets



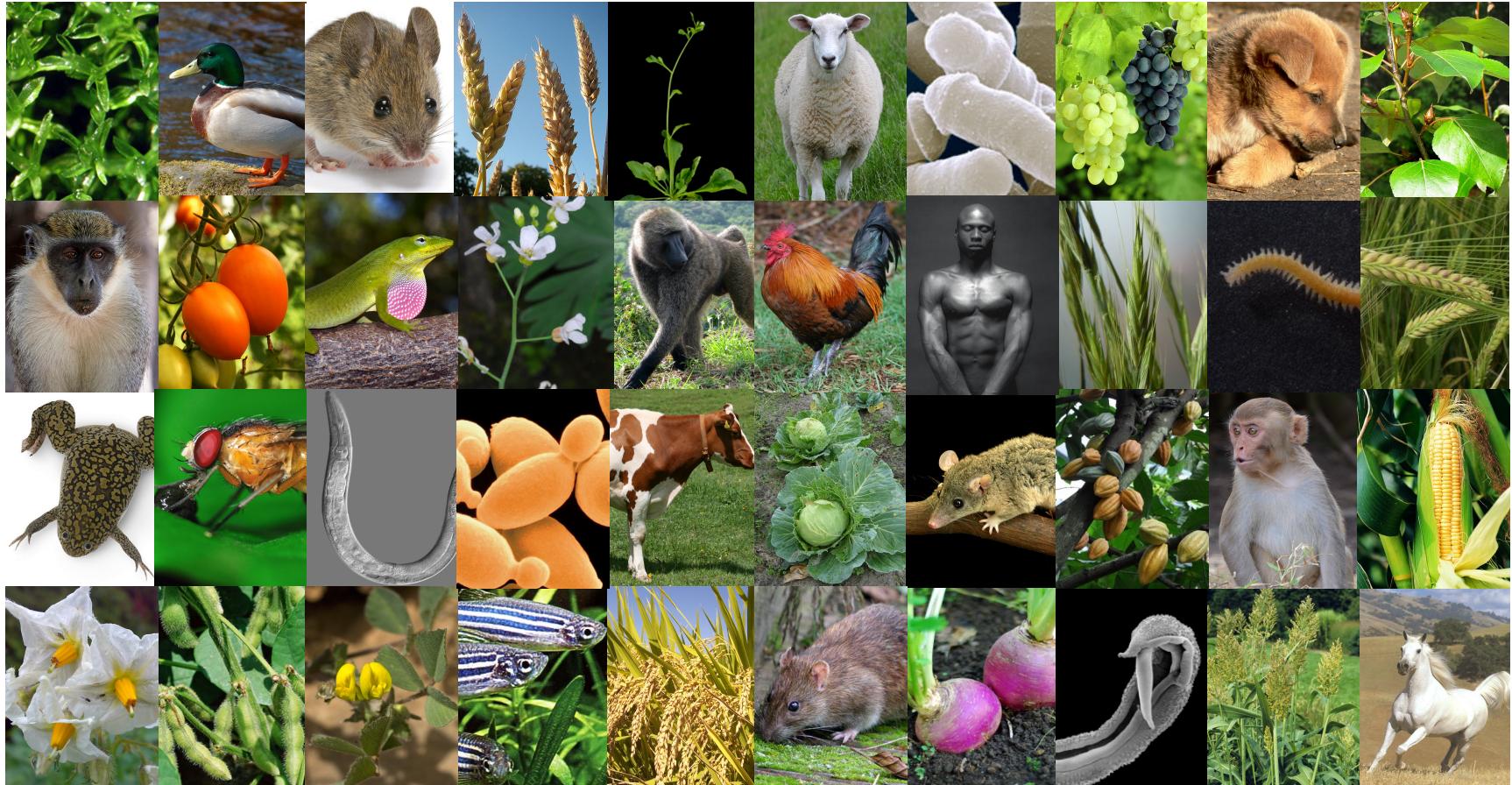
> 500 RNA-sequencing data



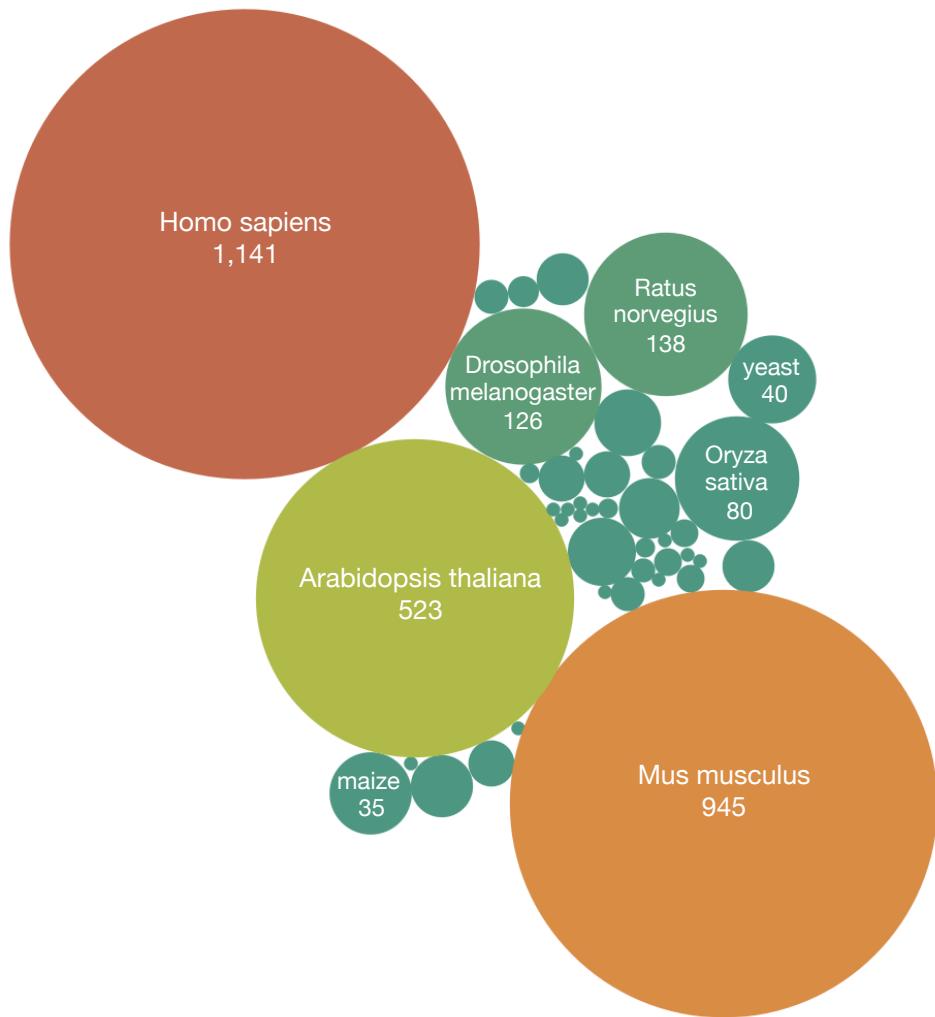
> 2,500 microarray data



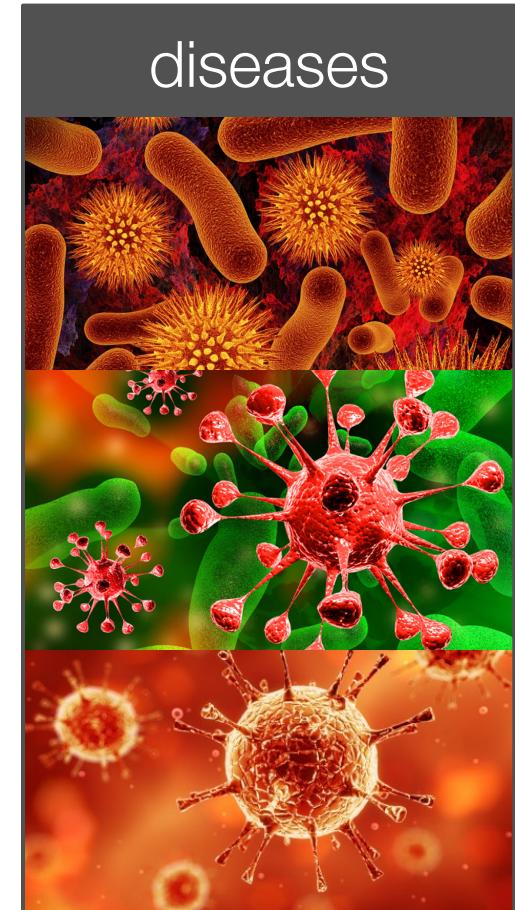
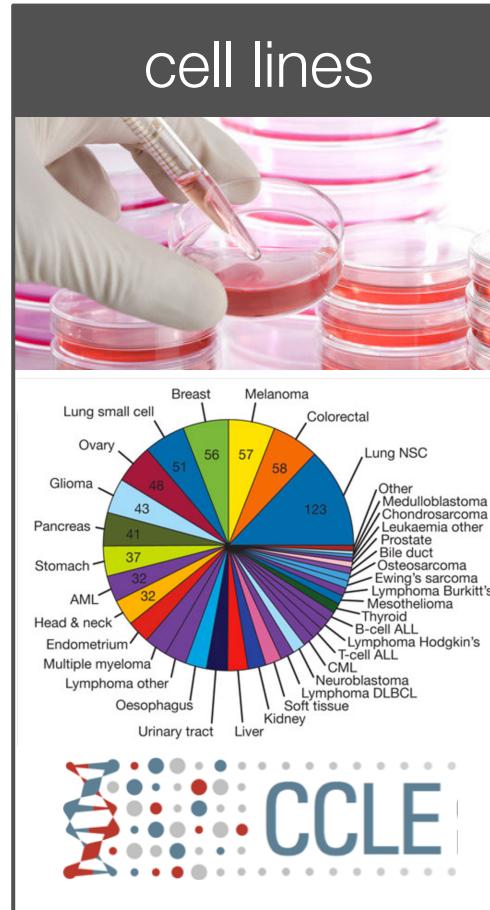
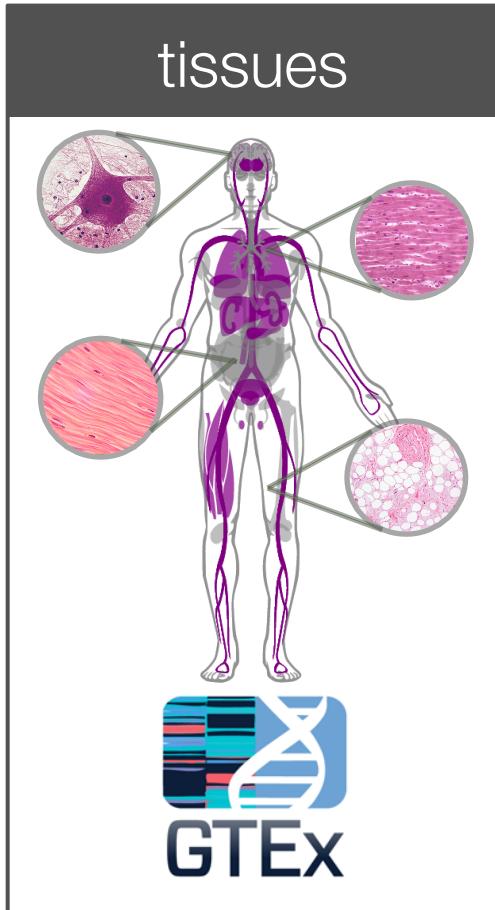
... across more than 40 species ...



... across more than 40 species ...



... under different biological conditions ...



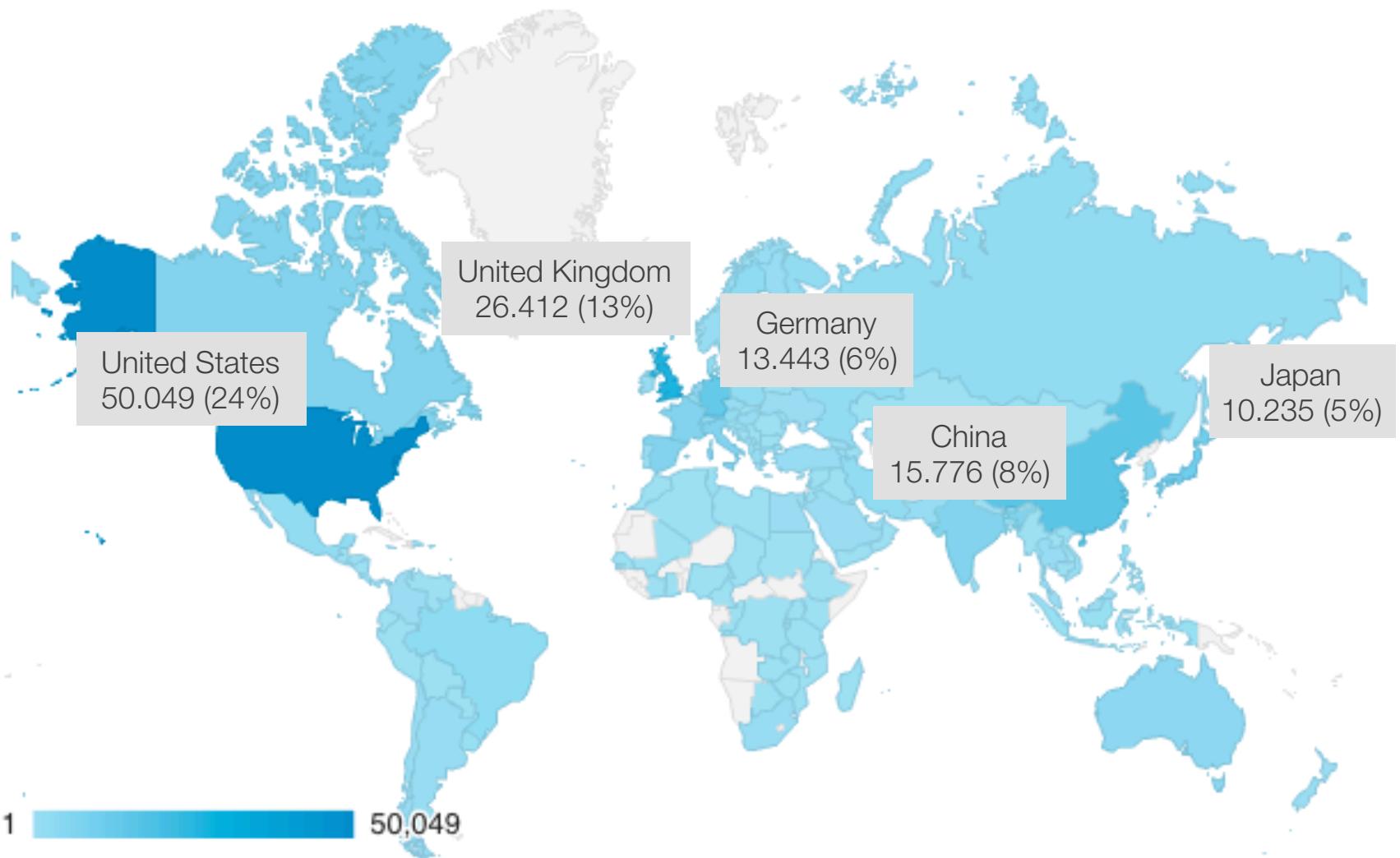
... to help scientists find answers ...

In which conditions
is my favourite gene
expressed?

What genes are
differentially expressed in
a particular condition?

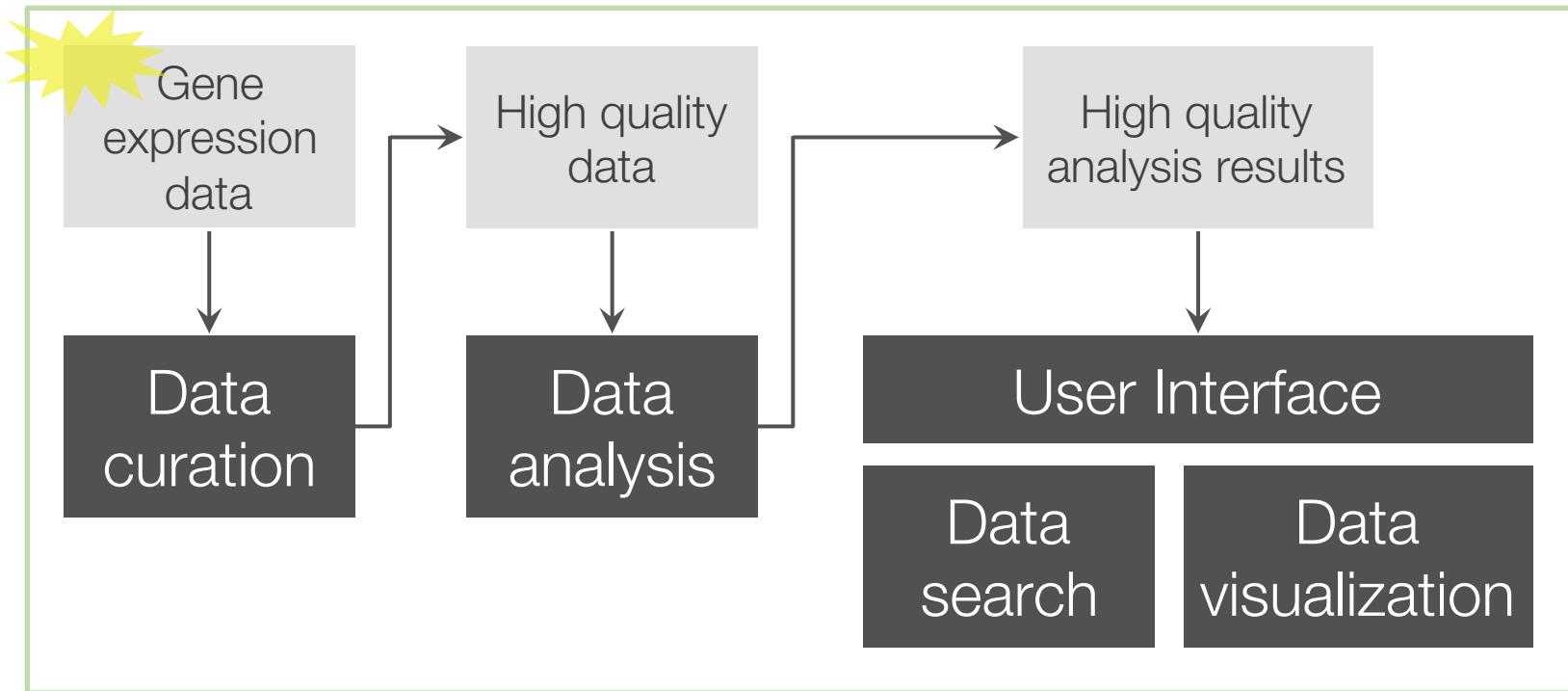
What genes are
expressed in a
particular condition?

... all across the world ...





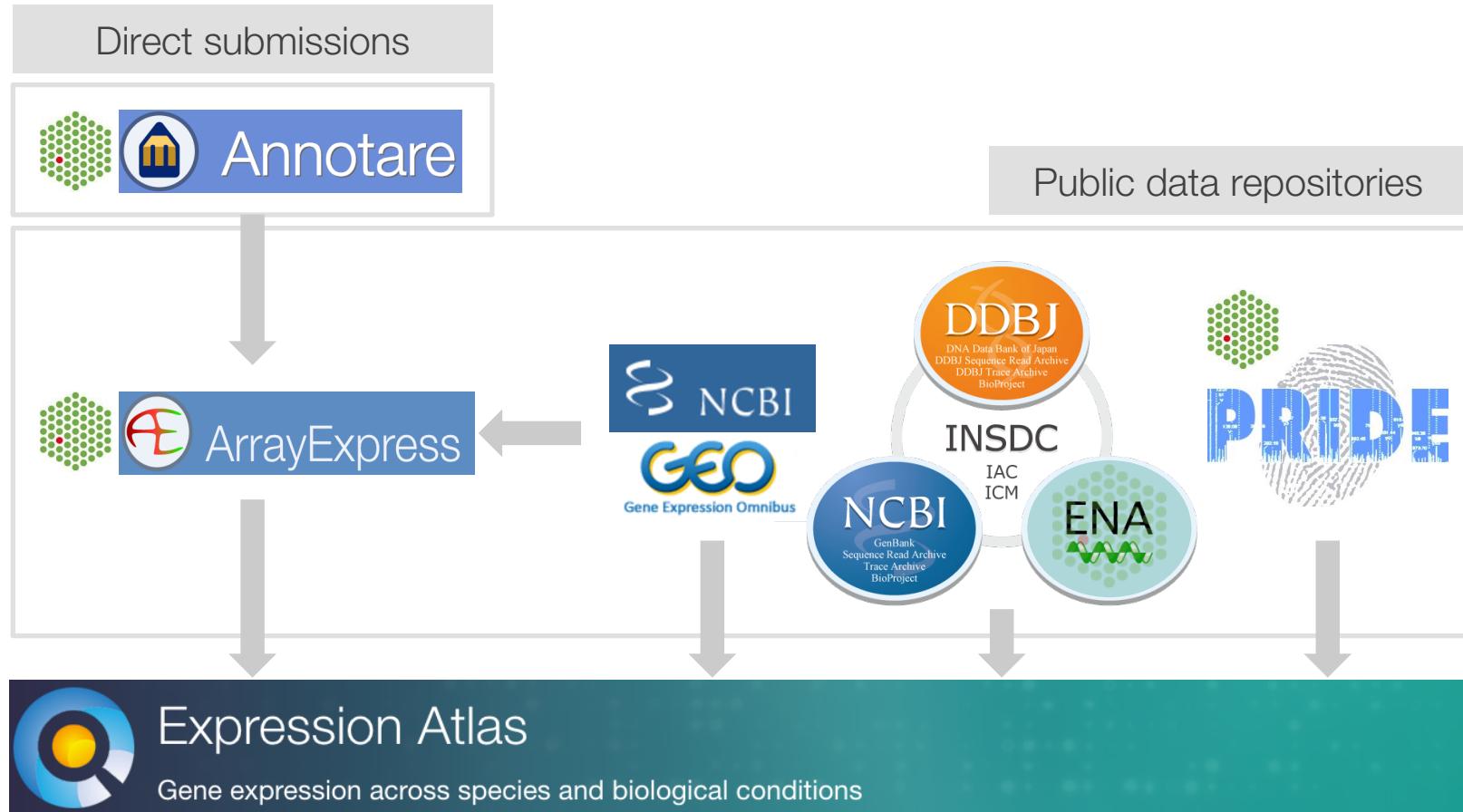
How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy



Where does expression data come from?





Annotare: webform submission tool

The screenshot shows the Annotare web application. On the left, there's a main panel titled "Submit your experiment" with a sub-section for "E-MTAB-9999". It features a flow diagram: two scientists at the top, a "Save as Spreadsheets" button, a "Submit to ArrayExpress" button, a central video camera icon, and three user profile icons below it. A large downward arrow points from the camera to the "ArrayExpress" logo at the bottom. On the right, there are two smaller panels: one for "Please log in to Annotare" with fields for email and password, and another for "Tweets" by @ArrayExpressEBI.

ArrayExpress: why and how to submit your data

Submit your experiment

E-MTAB-9999

Save as Spreadsheets Submit to ArrayExpress

ArrayExpress

Please log in to Annotare

Email address: lauhuema

Password: *****

Forgot your password? Log in

Don't have an account? Please register.

Tweets by @ArrayExpressEBI

ArrayExpress @ArrayExpressEBI It's Friday. There are those who often work at wknds to look after #AnnotareEBI submitters depositing data in ArrayExpress. Meet the team! 12 May

Yasset Perez-Riverol @ypriverol @OmicsDI nature.com/nbt/journal/v3... major coll. of 6 #EBI teams @pride_ebi @MetaboLights @ExpressionAtlas @ArrayExpressEBI @emblebies @ega_ebi 12 May

www.ebi.ac.uk/fg/annotare/login/



Annotare: webform submission tool

E-MTAB-5409

Experiment Description Samples and Data Experiment Description Preview Sample and Data Preview

General Information

Title * Loss of miR-22 in white adipose tissue

Contacts *

Publications

Description * Expression profiling of white adipose tissue isolated from WT and miR-22 KO animals
(at least 50 characters)

One-color microarray
- array design: A-GEOID-10528
- label: biotin
[change...](#)

Experiment Type * transcription profiling by array

[Add...](#)

E-MTAB-5409

Experiment Description Samples and Data Experiment Description Preview Sample and Data Preview

Create samples, add attributes and experimental variables

Describe protocols

Upload and assign data files

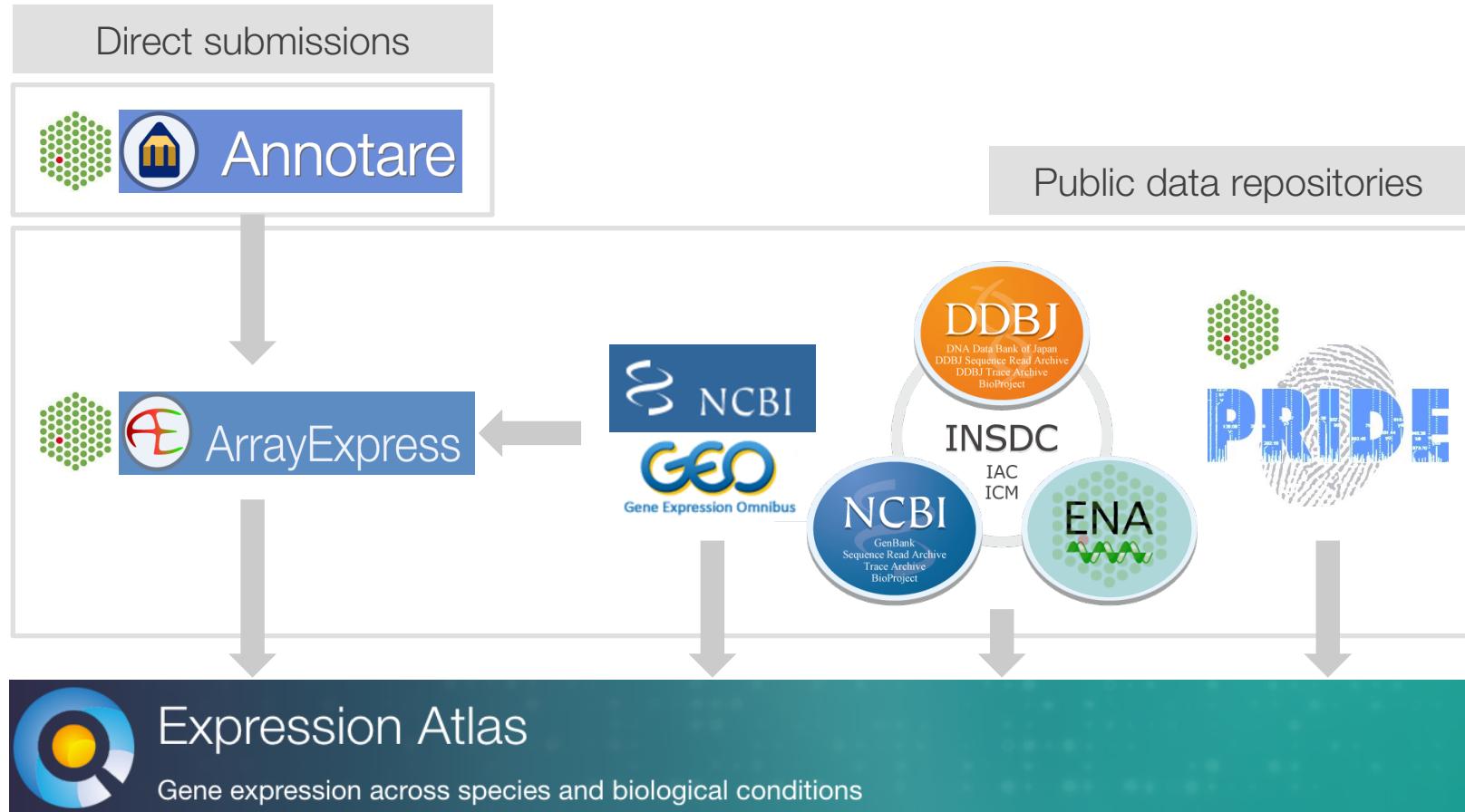
One-color microarray
- array design: A-GEOID-10528
- label: biotin
[change...](#)

Add Sample Attributes and Variables * [Add Samples](#) [Delete Samples](#) [Fill Down Value](#) [Import Values](#)

	Name	Material Type	Organism	Sex	Genotype	Organism Part
<input type="checkbox"/>	Sample 1	organism part	Mus musculus	male	wild type genotype	white adipose tissue
<input type="checkbox"/>	Sample 2	organism part	Mus musculus	male	wild type genotype	white adipose tissue
<input type="checkbox"/>	Sample 3	organism part	Mus musculus	male	wild type genotype	white adipose tissue
<input type="checkbox"/>	Sample 4	organism part	Mus musculus	male	miR-22 KO	white adipose tissue
<input type="checkbox"/>	Sample 5	organism part	Mus musculus	male	miR-22 KO	white adipose tissue
<input type="checkbox"/>	Sample 6	organism part	Mus musculus	male	miR-22 KO	white adipose tissue



Where does expression data come from?





ArrayExpress: public data repository

E-MTAB-5409 - Transcription profiling by array of mouse white adipose tissue from miR-22 knockouts and littermates

Status	Submitted on 9 February 2012, last updated on 13 March 2017, released on 13 March 2017
Organism	Mus musculus
Samples (6)	Click for detailed sample information and links to data
Array (1)	A-AFFY-130 - Affymetrix GeneChip Mouse Gene 1.0 ST Array [MoGene-1_0-st-v1]
Protocols (5)	Click for detailed protocol information

E-MTAB-5409 - Transcription profiling by array of mouse white adipose tissue from miR-22 knockouts and littermates

[Display full sample-data table](#)

[Export table in Tab-delimited format](#)

6 rows

Source Name	Sample Attributes					Variables	Assay	Links to Data
	organism	sex	genotype	organism part	strain			
Sample 1	Mus musculus	male	wild type genotype	white adipose tissue	C57Bl6/129SV	wild type genotype	WT_WAT_1	↓
Sample 2	Mus musculus	male	wild type genotype	white adipose tissue	C57Bl6/129SV	wild type genotype	WT_WAT_2	↓
Sample 3	Mus musculus	male	wild type genotype	white adipose tissue	C57Bl6/129SV	wild type genotype	WT_WAT_3	↓
Sample 4	Mus musculus	male	miR-22 knockout	white adipose tissue	C57Bl6/129SV	miR-22 knockout	miR22KO_WAT_1	↓
Sample 5	Mus musculus	male	miR-22 knockout	white adipose tissue	C57Bl6/129SV	miR-22 knockout	miR22KO_WAT_2	↓
Sample 6	Mus musculus	male	miR-22 knockout	white adipose tissue	C57Bl6/129SV	miR-22 knockout	miR22KO_WAT_3	↓



ArrayExpress: public data repository

ArrayExpress – functional genomics data

ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

[Browse ArrayExpress](#)

Latest News

25 April 2017 - **We are moving our helpdesk and submission email ticketing system**

If you are one of our users or submitters who have been in touch with us before, you may have got used to the email headers or format of our replies. Our ticketing system is coming up 10 years old, and it is time to migrate onto a more efficient system with new software and hardware. The migration will not affect our services, so you can continue to submit experiments via [Annotare](#), drop us submission-related queries at annotare@ebi.ac.uk, or send us questions about ArrayExpress at arrayexpress@ebi.ac.uk. The email headers or format will look a little different after the migration, but don't be alarmed, we are not spamming you!

Data Content

Updated today at 03:00

- 70003 experiments
- 2213515 assays
- 45.36 TB of archived data

Links

Information about how to search ArrayExpress, understand search results, how to submit data and FAQ can be found in our [Help section](#).

Find out more about the [Functional Genomics group](#).

Tools and Access

[Annotare](#): web-based submission tool for ArrayExpress.

[ArrayExpress Bioconductor package](#): an R package to access ArrayExpress and build data structures.

[Programmatic access](#): query and download data using web services or JSON.

[FTP access](#): data can be downloaded directly from our FTP site.

Related Projects

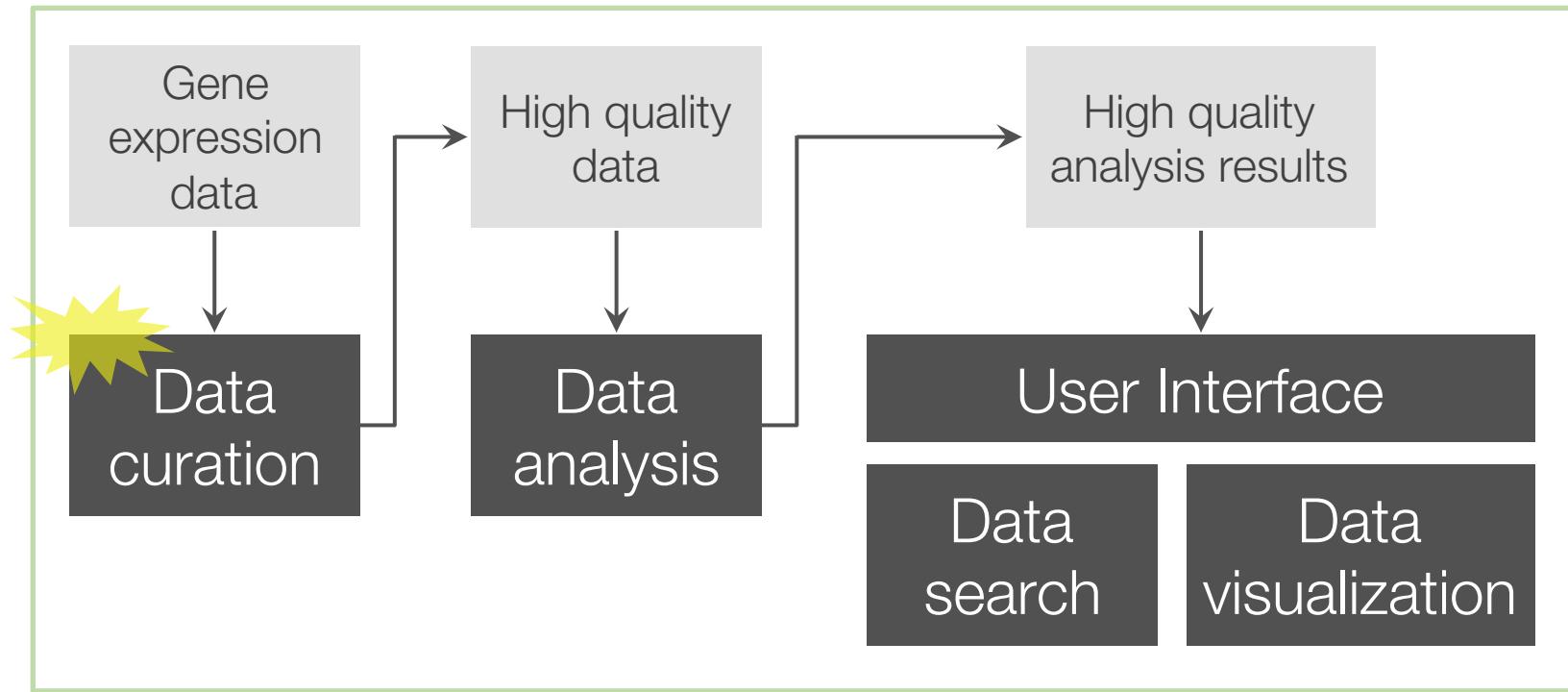
Discover up and down regulated genes in numerous experimental conditions in the [Expression Atlas](#).

Explore the [Experimental Factor Ontology](#) used to support queries and annotation of ArrayExpress data.

www.ebi.ac.uk/arrayexpress



How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easily

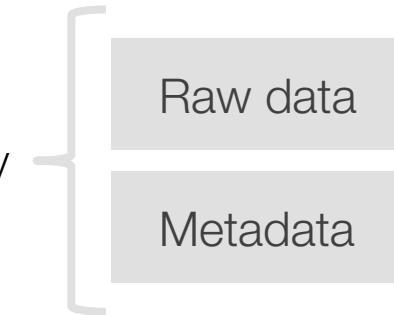


Expression Atlas – data curation

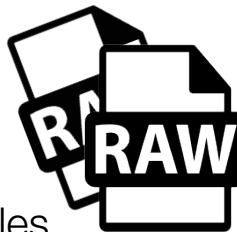
Biocuration



Translation and integration of information relevant to biology into a database or resource



Raw data



- unprocessed data files
- **Microarray:** files from the scanner (e.g. Affymetrix CEL files, Agilent feature extraction *txt* files, Illumina *idat* files)
- **Sequencing:** raw sequence read files (e.g. FASTQ files)

Metadata

- Experiment description
- Experiment title
- Sample annotation
- Protocols
- Publication details (if any)
- Author contact details



Expression Atlas – data curation

PeerJ

✓ PEER-REVIEWED

XA21-specific induction of stress-related genes following *Xanthomonas* infection of detached rice leaves

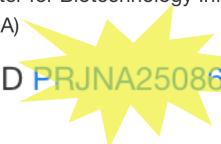


Data Availability

The following information was supplied regarding data availability:

The National Center for Biotechnology Information Sequence Read Archive (SRA)

BioProject ID [PRJNA250865](#).



Metadata

RNA sequencing sample treatment summary

Table summarizes the experimental setup including the genotypes, time of treatment, and type of treatment used for samples used in RNA sequencing. There were three replicates for each sample for a total of 21 sequenced samples.

21 sequenced samples

	A	B	C
1	Genotype	Time (hours)	Treatment
2	Kitaake	0	None
3	EFR::XA21::GFP	0	None
4	EFR::XA21::GFP	0.5	500 nM elf18
5	EFR::XA21::GFP	1	500 nM elf18
6	EFR::XA21::GFP	3	500 nM elf18
7	EFR::XA21::GFP	6	500 nM elf18
8	EFR::XA21::GFP	12	500 nM elf18



Expression Atlas – data curation

Raw data

BioProject ID [PRJNA250865](#).

SRA PRJNA250865
Create alert Advanced

Summary ▾ 20 per page ▾ Send to: ▾

[Oryza sativa Japonica Group strain:EFR-XA21 | cultivar:Kitaake Transcriptome or Gene expression - BioProject](#)
Rice RNAseq
Gene expression project
Accession: [PRJNA250865](#)

Items: 1 to 20 of 21



Expression Atlas – data curation

Raw data



- Genotype?
- Treatment?
- Time?

[SRX873376: Other Sequencing of Japanese rice](#)

1 ILLUMINA (Illumina HiSeq 2000) run: 29.9M spots, 9G bases, 5.3Gb downloads

Submitted by: DOE JOINT GENOME INSTITUTE (JGI)

Study: Oryza sativa Japonica strain:EFR-XA21 | cultivar:Kitaake Transcriptome or Gene expression

[PRJNA250865](#) • [SRP054056](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample: Oryza sativa cv. Kitaake EFR-XA21

[SAMN03003383](#) • [SRS843490](#) • [All experiments](#) • [All runs](#)

Organism: [Oryza sativa Japonica Group](#)

Library:

Name: NTBA
Instrument: Illumina HiSeq 2000
Strategy: RNA-Seq
Source: TRANSCRIPTOMIC
Selection: RT-PCR
Layout: PAIRED

Spot descriptor:



Runs: 1 run, 29.9M spots, 9G bases, [5.3Gb](#)

Run	# of Spots	# of Bases	Size	Published
SRR1799213	29,893,272	9G	5.3Gb	2015-02-20



Expression Atlas – data curation

PeerJ

✓ PEER-REVIEWED

XA21-specific induction of stress-related genes following *Xanthomonas* infection of detached rice leaves



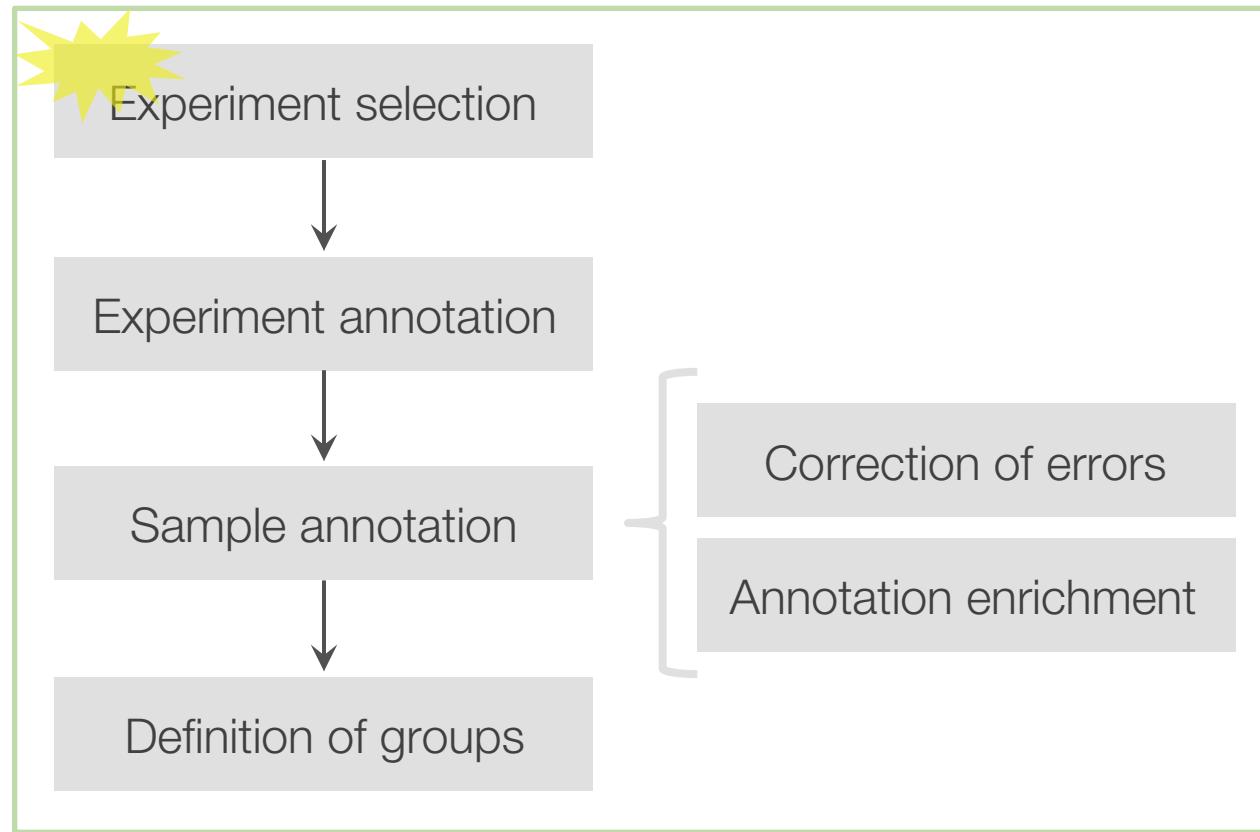
Raw data

Metadata

Run	Description	FactorValue[genotype]	FactorValue[compound]	FactorValue[dose]	Unit[concentration unit]	FactorValue[time]	Unit[time unit]
SRR1799194	Kitaake 0 none	wild type	none		0 nanomolar		0 hour
SRR1799197	Kitaake 0 none	wild type	none		0 nanomolar		0 hour
SRR1799193	Kitaake 0 none	wild type	none		0 nanomolar		0 hour
SRR1799210	EFR:XA21:GFP 0 none	EFR:XA21:GFP	none		0 nanomolar		0 hour
SRR1799203	EFR:XA21:GFP 0 none	EFR:XA21:GFP	none		0 nanomolar		0 hour
SRR1799205	EFR:XA21:GFP 0 none	EFR:XA21:GFP	none		0 nanomolar		0 hour
SRR1799199	EFR:XA21:GFP 0.5 elf18	EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
SRR1799204	EFR:XA21:GFP 0.5 elf18	EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
SRR1799212	EFR:XA21:GFP 0.5 elf18	EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
SRR1799208	EFR:XA21:GFP 1 elf18	EFR:XA21:GFP	elf18		500 nanomolar		1 hour
SRR1799202	EFR:XA21:GFP 1 elf18	EFR:XA21:GFP	elf18		500 nanomolar		1 hour
SRR1799198	EFR:XA21:GFP 1 elf18	EFR:XA21:GFP	elf18		500 nanomolar		1 hour
SRR1799213	EFR:XA21:GFP 3 elf18	EFR:XA21:GFP	elf18		500 nanomolar		3 hour
SRR1799201	EFR:XA21:GFP 3 elf18	EFR:XA21:GFP	elf18		500 nanomolar		3 hour
SRR1799196	EFR:XA21:GFP 3 elf18	EFR:XA21:GFP	elf18		500 nanomolar		3 hour
SRR1799209	EFR:XA21:GFP 6 elf18	EFR:XA21:GFP	elf18		500 nanomolar		6 hour
SRR1799206	EFR:XA21:GFP 6 elf18	EFR:XA21:GFP	elf18		500 nanomolar		6 hour
SRR1799195	EFR:XA21:GFP 6 elf18	EFR:XA21:GFP	elf18		500 nanomolar		6 hour
SRR1799207	EFR:XA21:GFP 12 elf18	EFR:XA21:GFP	elf18		500 nanomolar		12 hour
SRR1799200	EFR:XA21:GFP 12 elf18	EFR:XA21:GFP	elf18		500 nanomolar		12 hour
SRR1799211	EFR:XA21:GFP 12 elf18	EFR:XA21:GFP	elf18		500 nanomolar		12 hour



Expression Atlas – data curation





Expression Atlas – data curation

Experiment selection

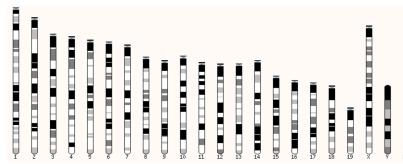


Raw data



Experiment type

www.ensembl.org



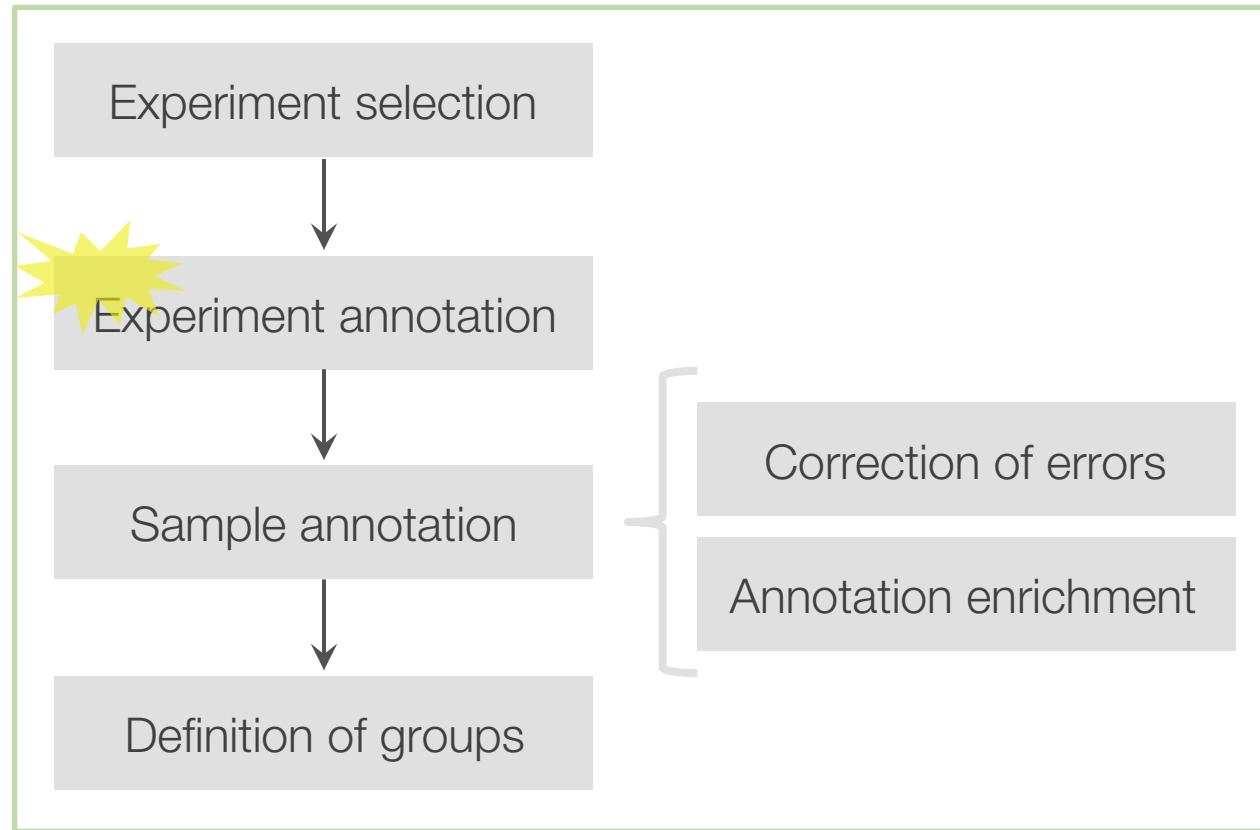
Reference genome



Biological replicates



Expression Atlas – data curation





Expression Atlas – data curation

Experiment annotation



Clear experiment intent

<title> "rice transcriptome project" →

"Transcription profiling by high throughput sequencing of rice oscerk1 mutants treated with two fungus-secreted substances"

Informative title



Protocol description



Europe PMC

□ Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.
(PMID:25954001 PMID:PMC4547484)

Abstract · Citations · BioEntities · Related Articles · External Links

GTEx Consortium
+ Collaborators (139)

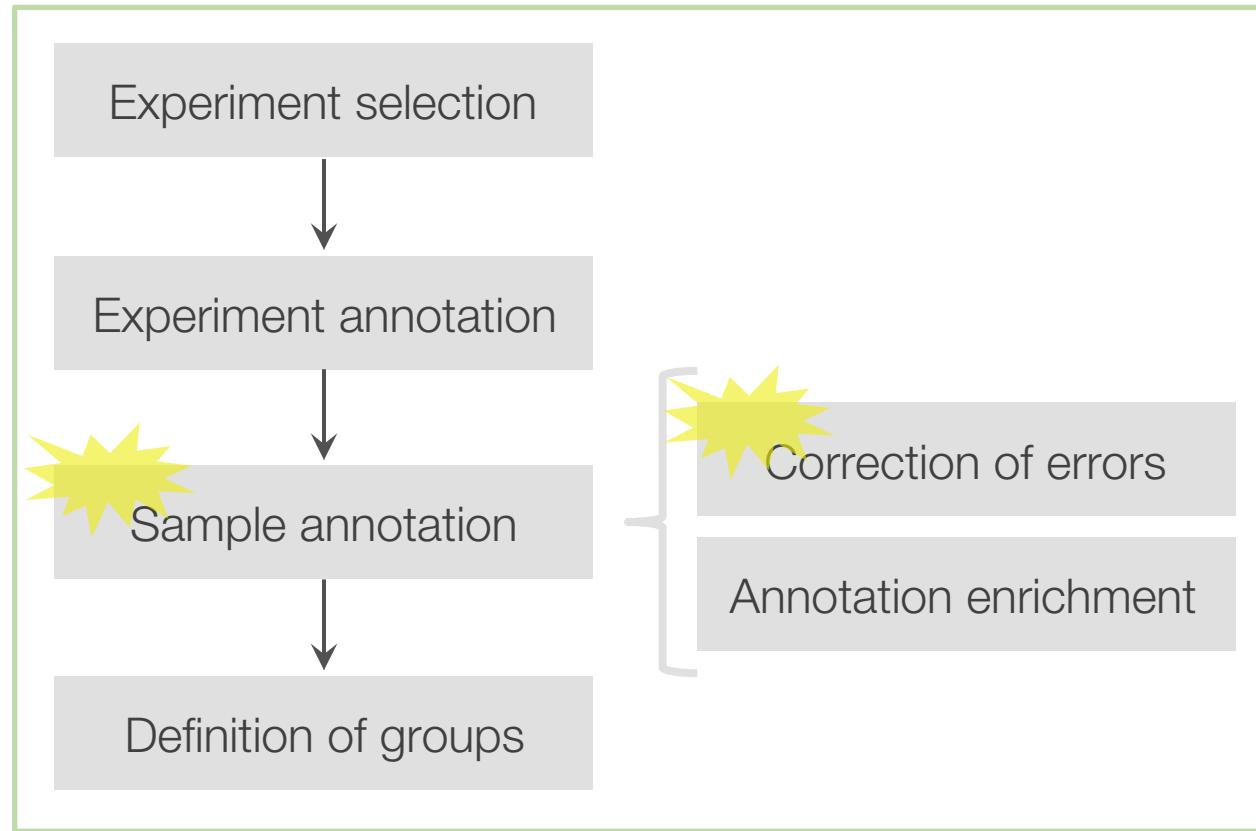
Science (New York, N.Y.) [07 May 2015, 348(6235):648-660]

Type: Research Support, Non-U.S. Gov't, research-article, journal Article, Research Support, N.I.H., Extramural
DOI: 10.1126/science.1262110

Publication details



Expression Atlas – data curation





Expression Atlas – data curation

Sample annotation

Hands-on activity

Let's become a biocurator for Expression
Atlas (at least during 20 min...)



In pairs



Expression Atlas – data curation

Sample annotation

<https://www.ncbi.nlm.nih.gov/geo>



GSE44596



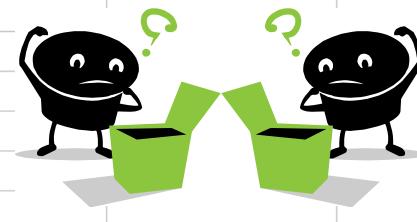
- ✓ What is the experiment about?
- ✓ In which organism is the experiment done?
- ✓ What is the biological material used in the experiment?
- ✓ How many samples are authors testing?
- ✓ Is it a microarray or a RNA-sequencing experiment?
- ✓ What is the main variable that authors are studying?
- ✓ What are the two groups of samples compared?
- ✓ How many biological replicates are there in each group?
- ✓ Are there any technical replicates?



Expression Atlas – data curation

Sample annotation

	Sample characteristics		Experimental variables		
Assay Name	Characteristics [organism]	Characteristics [cell type]	FactorValue [compound]	FactorValue [dose]	Unit [concentration unit]
GSM1087598	Homo sapiens	coronary artery endothelial cell			
GSM1087599	Homo sapiens	coronary artery endothelial cell			
GSM1087600	Homo sapiens	coronary artery endothelial cell			
GSM1087601	Homo sapiens	coronary artery endothelial cell			
GSM1087602	Homo sapiens	coronary artery endothelial cell			
GSM1087603	Homo sapiens	coronary artery endothelial cell			



GSM1087598 HCAEC_SsnB_rep1
GSM1087599 HCAEC_SsnB_rep2
GSM1087600 HCAEC_SsnB_rep3

{
Source name
Organism
Characteristics

HCAECs exposed to SsnB
Homo sapiens
cell type: Primary Human Coronary Artery Endothelial Cells
treatment: control

GSM1087601 HCAEC_DMSO_rep1
GSM1087602 HCAEC_DMSO_rep2
GSM1087603 HCAEC_DMSO_rep3

{
Source name
Organism
Characteristics

HCAECs exposed to Vehicle Control
Homo sapiens
cell type: Primary Human Coronary Artery Endothelial Cells
treatment: Sparstololin B



Expression Atlas – data curation

Sample annotation



Correct errors

Hi, Laura

The sample labels for each file are correct, e.g., these are all results for SsnB-treated cells.

GSM1087598 HCAEC_SsnB_rep1

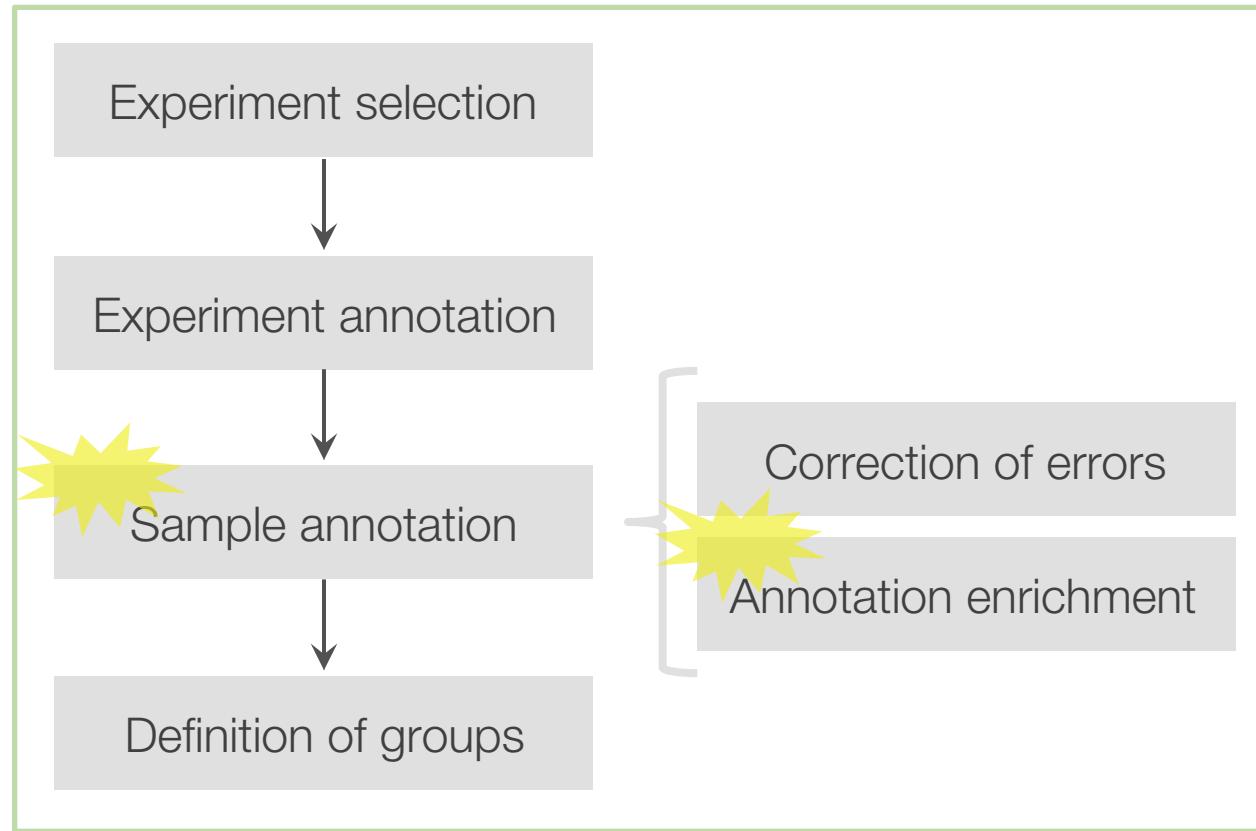
GSM1087599 HCAEC_SsnB_rep2

GSM1087600 HCAEC_SsnB_rep3

The information listed under 'Characteristics' appears to be inverted; that is, 'treatment: control' are actually the SsnB-treated cells and 'treatment: Sparstolonin B' are actually the DMSO vehicle controls.



Expression Atlas – data curation





Expression Atlas – data curation

Sample annotation



Adding value to your data

Characteristics[organism]

Mus musculus
Mus musculus

We enrich sample annotation by adding more metadata to:

- 1) better describe the biology of the sample
- 2) better browse the data

Characteristics[cell line]	Characteristics[cell type]
	cholangiocyte
	hepatocyte
	hepatocyte
	hepatocyte
	hepatocyte
Hep-53.4	epithelial
Hepa 1-6	epithelial
	hepatocyte
	hepatic stellate cell
J774.2	Kupffer cell
	endothelial cell of sinusoid

Characteristics[growth condition]

1 day culture of primary hepatocyte
3 day culture of primary hepatocyte
7 day culture of primary hepatocyte
primary hepatocyte

freshly isolated hepatocyte



Expression Atlas – data curation

Sample annotation



Adding value to your data

Characteristics[organism]	Characteristics[strain]	Characteristics[sex]	Characteristics[age]	Unit[time unit]
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	female	8 to 10	week
Mus musculus	C57L/J	female	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	BALB/c	female	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week
Mus musculus	C57BL/6J	male	8 to 10	week

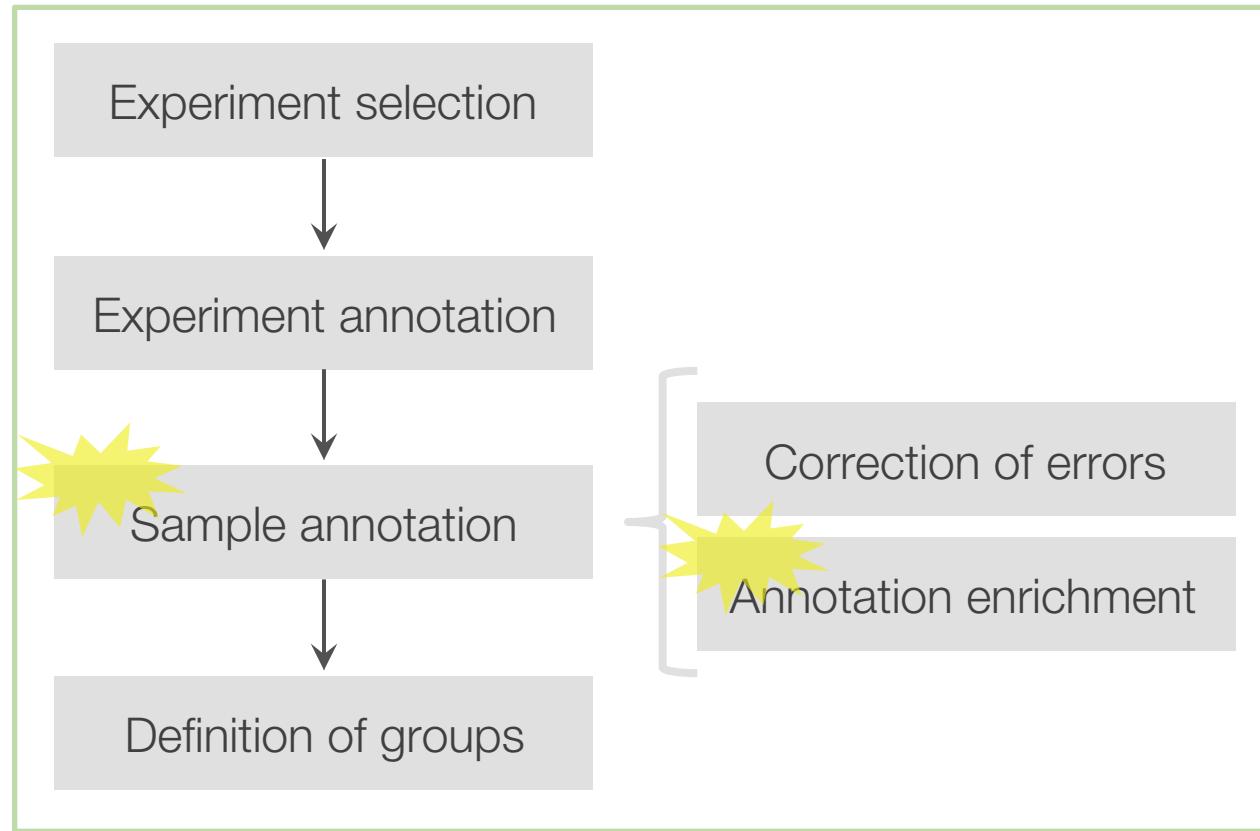
Characteristics[organism part]	Characteristics[cell line]	Characteristics[cell type]	Characteristics[disease]	Characteristics[growth condition]
liver		cholangiocyte	normal	1 day culture of primary hepatocyte
liver		hepatocyte	normal	3 day culture of primary hepatocyte
liver		hepatocyte	normal	7 day culture of primary hepatocyte
liver		hepatocyte	normal	primary hepatocyte
liver	Hep-53.4	epithelial	hepatocellular carcinoma of the mouse	
liver	Hepa 1-6	epithelial	hepatocellular carcinoma of the mouse	
liver		hepatocyte	normal	
liver		hepatic stellate cell	normal	
blood	J774.2	Kupffer cell	cancer	
liver			normal	
liver			normal	
liver		endothelial cell of sinusoid	normal	

We enrich sample annotation by adding more metadata to:

- 1) better describe the biology of the sample
- 2) better browse the data



Expression Atlas – data curation





Expression Atlas – data curation

Ontology

www.ebi.ac.uk/efo



A systematic way to name and organise entities, establishing relationships between the entities

Controlled vocabulary

Hierarchy (relationship)

Smart search

Efficient search via ontology-driven query expansion



Biological conditions

lung car|

Ex lung carcinoid tumor

+ lung carcinoma

Complex queries

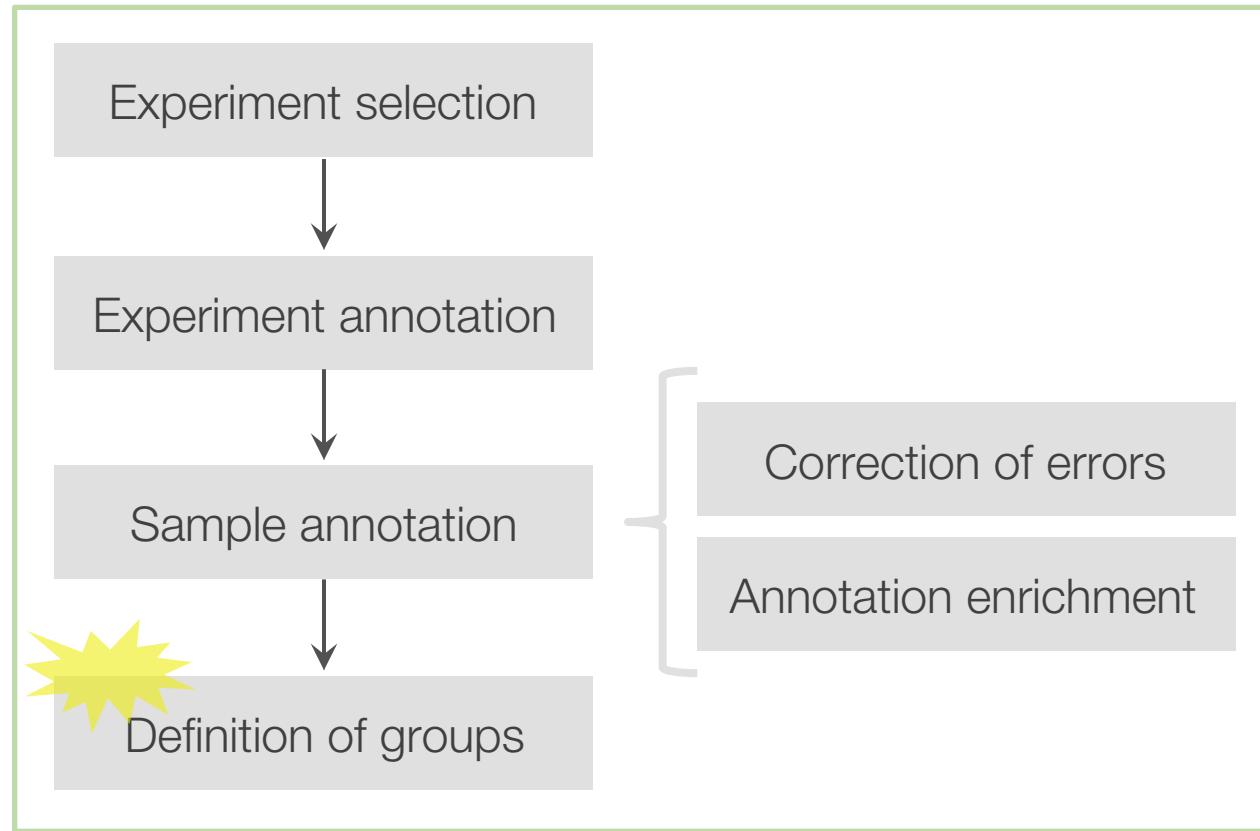


lung carcinoma

- bronchogenic carcinoma
- large cell lung carcinoma
 - Lung Sarcomatoid Carcinoma
 - Lymphoepithelioma-Like Lung Carcinoma
- lung adenocarcinoma
- non-small cell lung carcinoma
 - pulmonary mucoepidermoid carcinoma
 - pulmonary neuroendocrine tumor
 - small cell lung carcinoma
 - squamous cell lung carcinoma

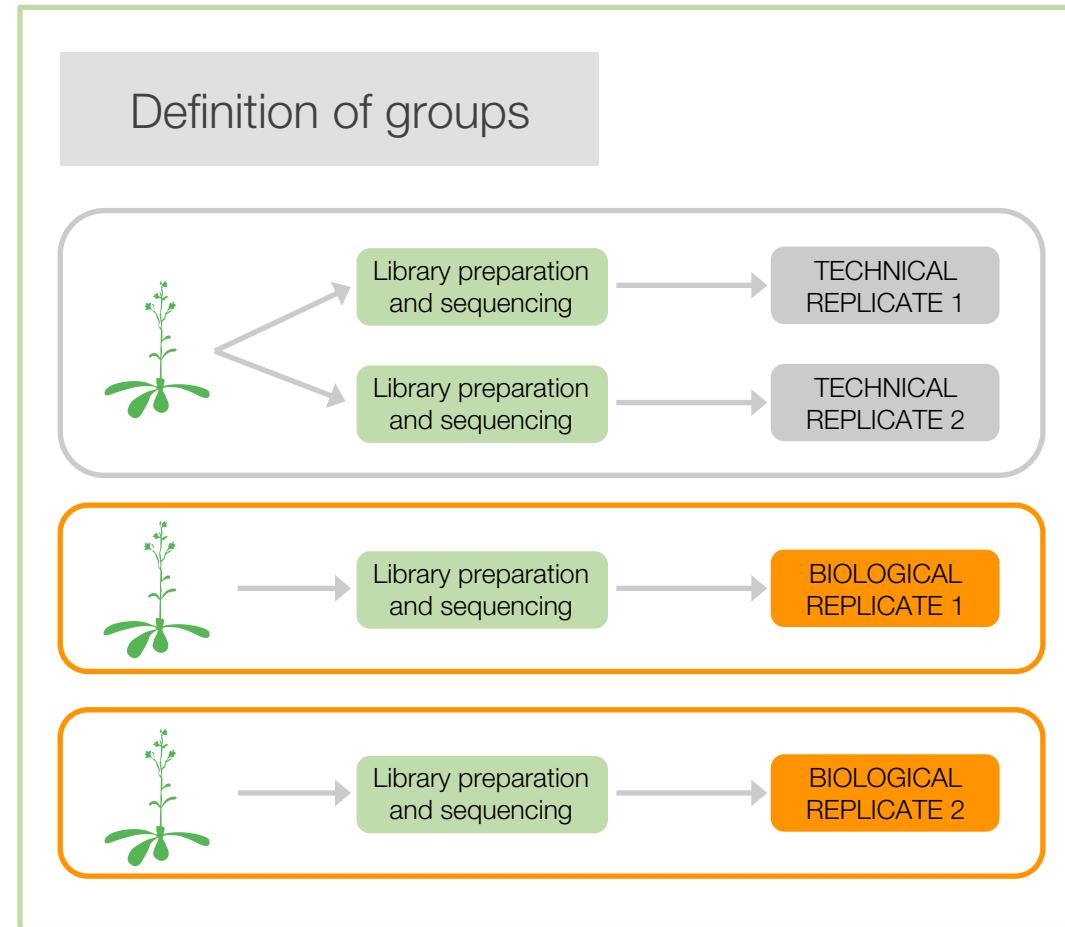


Expression Atlas – data curation



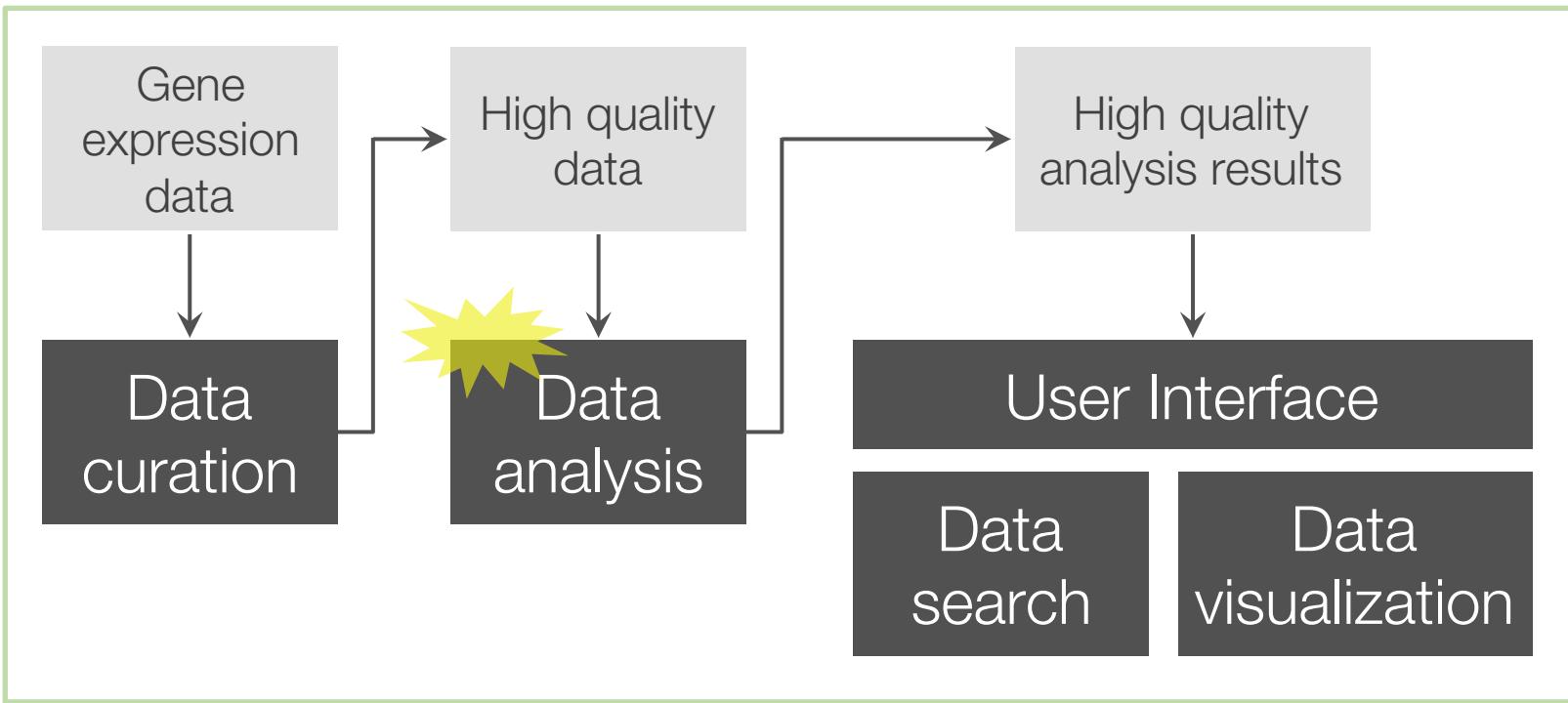


Expression Atlas – data curation





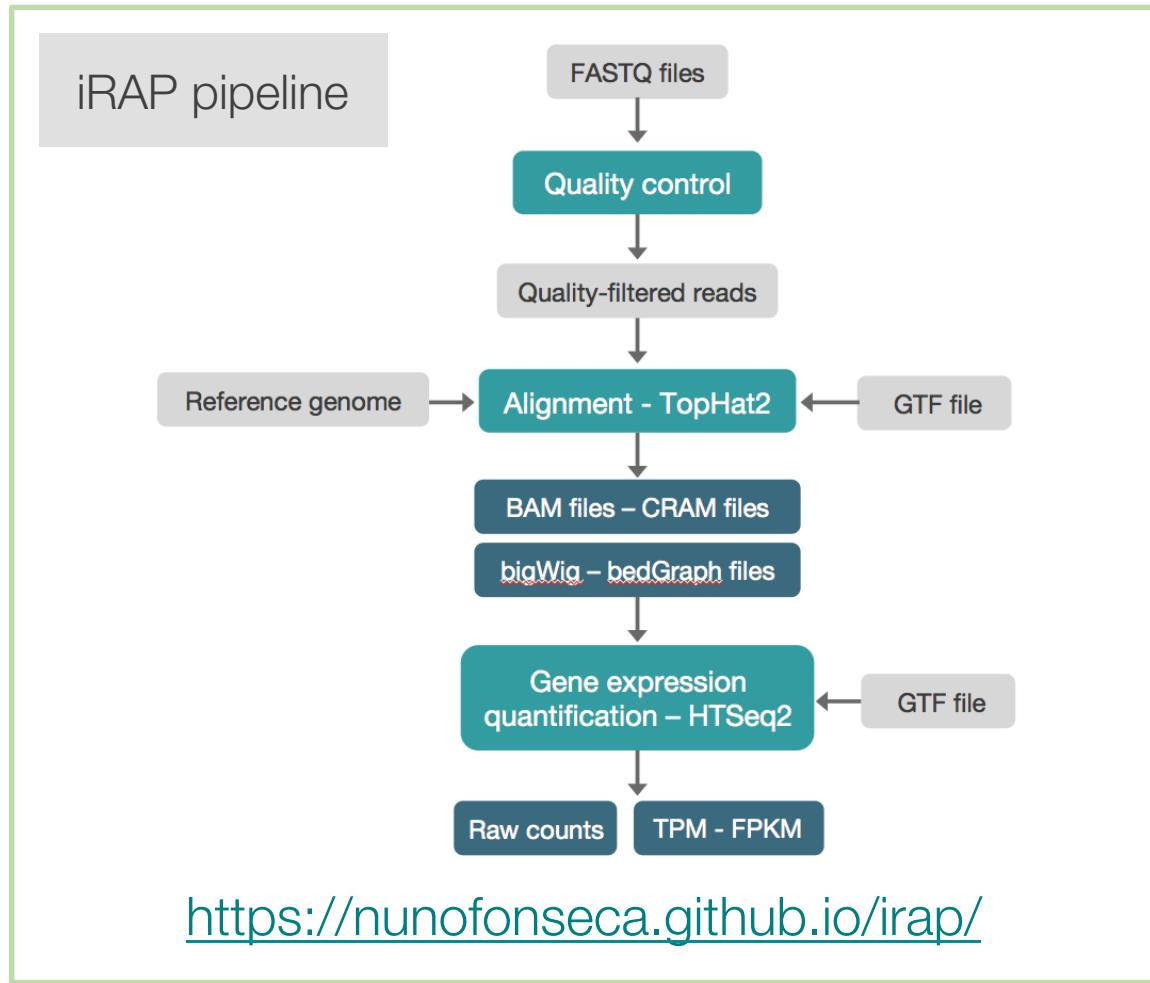
How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy



Expression Atlas – data analysis





RNA-seq data analysis

Train online

[Training](#)[Train online](#)[About Train online](#)[Glossary](#)[Support and feedback](#)[Login/register](#)

Functional genomics (II):
Common technologies and
data analysis methods

[Introduction](#)[Real-time PCR](#)[Microarrays](#)[Next Generation Sequencing \(NGS\)](#)[RNA sequencing](#)[Biological interpretation of gene expression data](#)[Genotyping, epigenetic and DNA/RNA-protein interaction methods](#)[Summary](#)[Quiz: Check your learning](#)[Your feedback](#)[Learn more](#)[References](#)[Contributors](#)

Functional genomics (II): Common technologies and data analysis methods

This is the second part of our functional genomics course. This course will introduce you to common technologies in functional genomics studies, including microarrays and next generation sequencing (NGS), with a special focus on RNA-seq. We will look at important considerations when designing your experiments, data analysis methods, and discuss when to use one technology over another.

We recommend that you look at [Functional genomics \(I\): Introduction and designing experiments](#) before starting this course.

About this course

Author(s): Laura Huerta, Melissa Burke

Gene Expression

Beginner

2 hours

[Start the course](#)

Learning objectives:

- Evaluate different functional genomics technologies and know when to use them
- List some of the considerations when designing your own functional genomics experiments using these technologies
- Describe the steps involved in data analysis of microarray and RNA-seq experiments

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

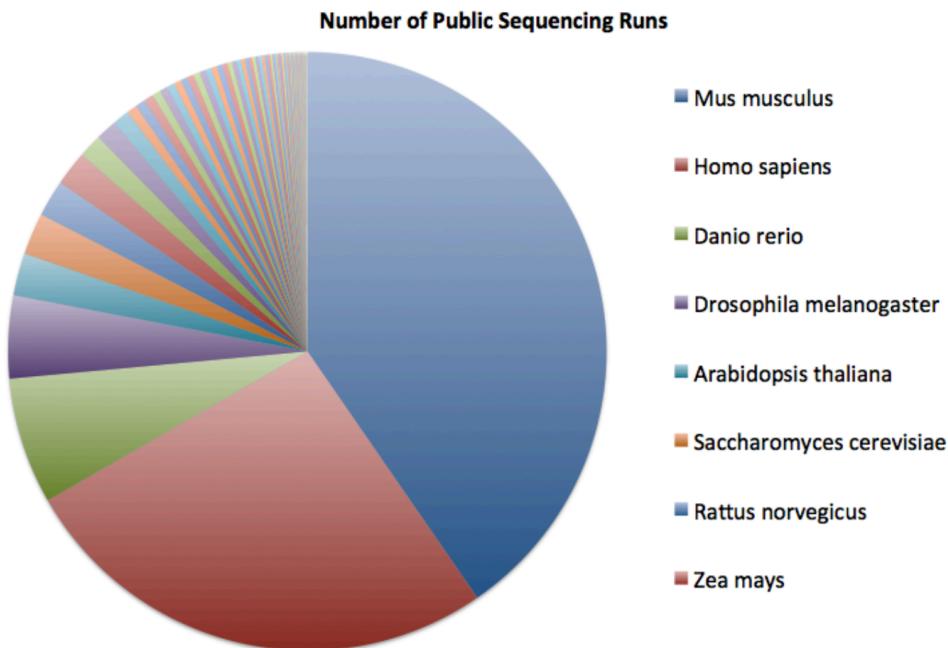


RNASeq-er API

> 270,000 analysed runs

> 15,000 studies

> 300 species (163 genome references)



800,000 runs not analysed from 6,500 species, not available via API (no reference genome/gene annotation available in public databases)



RNASeq-er API

RNASeq-er API

Home Documentation Feedback

Welcome to RNASeq-er API - a gateway to systematically updated analysis of public RNA-Seq data

The RNASeq-er REST API provides easy access to the results of the systematically updated and continually growing analysis of public RNA-seq data in European Nucleotide Archive (ENA). The analysis of each sequencing run is performed by the EMBL-EBI's Gene Expression Team using the [iRAP pipeline](#).

Numbers and mapping quality of sequencing runs processed and available via RNASeq-er API to date

Source of Genome Reference

ensembl	fungi	metazoa	plants	protists	wbps
Mus musculus Homo sapiens Danio rerio	Saccharomyces cerevisiae	Drosophila melanogaster	Arabidopsis thaliana	Caenorhabditis elegans	

Mean Pct o... 0.00 98.00

RNA-seq Analysis status as of: 2017-05-16 15:00

Completed
Running
Queued

Number of sequencing runs (log10)

← Undo → Redo ← Reset

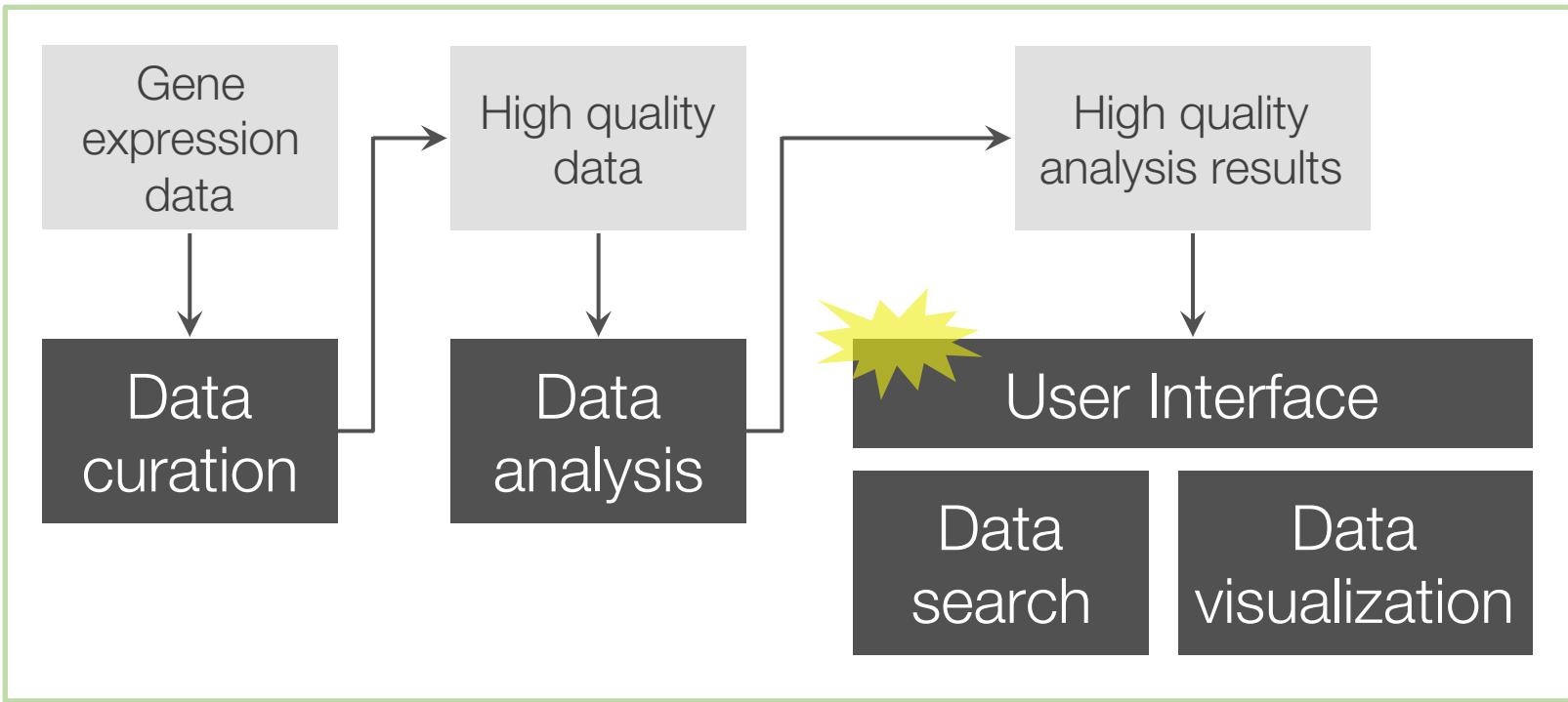
Share Download Full Screen

493 views | more by this author

<https://www.ebi.ac.uk/fg/rnaseq/>



How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy



Let's try Expression Atlas

www.ebi.ac.uk/gxa/home

Expression Atlas

Enter gene query...

Search

Examples: ASPM, Apoptosis, ENSMUSG00000021789, zinc finger

Home Release notes FAQ Download Help Licence About

Feedback

Expression Atlas: Differential and Baseline Expression

The Expression Atlas provides information on gene expression patterns under different biological conditions. Gene expression data is re-analysed in-house to detect genes showing interesting baseline and differential expression patterns. [Read more about Expression Atlas...](#)

Search...

Gene / Gene property ?

Enter gene query...

Organism

Any

E.g. REG1B, zinc finger

Sample properties ?

Enter condition query...

Search

Reset

E.g. lung, leaf, valproic acid, cancer

iRAP: RNA-seq analysis tool

iRAP is a flexible pipeline for RNA-seq analysis that integrates many existing tools for filtering and mapping reads, quantifying expression and testing for differential expression. iRAP is used to process all RNA-seq data in Expression Atlas.

Publications

RNA-Seq Gene Profiling - A Systematic Empirical Comparison (*PLoS One*, 2014).

Expression Atlas update — an integrated database of gene and protein expression in humans, animals and plants (*Nucleic Acids Research*, 2016).

Browse...

Baseline Experiments

See all baseline expression data sets in Expression Atlas.

Plant Experiments

See all expression data sets in plants in Expression Atlas.

All Experiments

Scroll through the complete list of all data sets in Expression Atlas.



Let's try Expression Atlas

Expression Atlas

Gene expression across species and biological conditions

Home | Download | Release notes | FAQ | Help | Licence | About

Search this project

Enter gene query... Examples: ASPM, Apoptosis, ENSMUSG00000021789, zinc finger

Exploring gene expression results across species under different biological conditions

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others. Expression Atlas aims to help answering questions such as 'where is a certain gene expressed?' or 'how its expression changes in a disease?'.

[Read more about Expression Atlas](#)

Search Gene set enrichment

Gene / Gene properties
CFHR2 [X](#)
Examples: REG1B, zinc finger

Organism
Homo sapiens

Biological conditions
Enter condition query... Examples: lung, leaf, valproic acid, cancer

[Search](#) [Clear](#)

wwwdev.ebi.ac.uk/gxa/home



Let's try Expression Atlas

Expression Atlas

Gene expression across species and biological conditions

Home | Download | Release notes | FAQ | Help | Licence | About

Exploring gene expression results across species under different biological conditions

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others. Expression Atlas aims to help answering questions such as 'where is a certain gene expressed?' or 'how its expression changes in a disease?'.

[Read more about Expression Atlas](#)

In which conditions is my favourite gene expressed?

Organism

- ✓ Homo sapiens
- Mus musculus
- Arabidopsis thaliana
- Rattus norvegicus
- Drosophila melanogaster
- Zea mays

Any

- Anas platyrhynchos
- Anolis carolinensis
- Anopheles gambiae
- Arabidopsis lyrata
- Arabidopsis thaliana
- Aspergillus fumigatus
- Bos taurus
- Brachypodium distachyon
- Brassica oleracea
- Brassica rapa
- Caenorhabditis elegans

Biological conditions

Enter condition query...

Examples: lung, leaf, valproic acid, cancer

Discover and interpret gene expression analysis results quickly and easy



Baseline expression results

Baseline expression

Differential expression

Show anatomograms

Filter your results

Homo sapiens

Organism part

Cell line

Danio rerio

Developmental stage

Mus musculus

Organism part

Cell type

Developmental stage

Homo sapiens — Organism part



Showing 7 of 7 experiments found:



All organisms and all conditions in which *CFHR2* gene is expressed

Danio rerio — Developmental stage

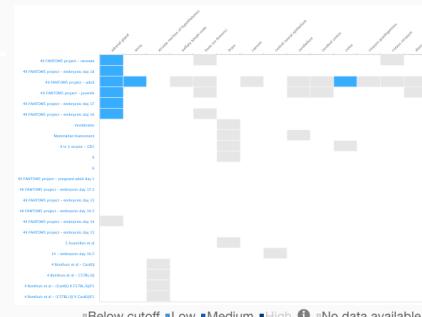
Showing 1 of 1 experiment found:

By experiment type • Filters • Download table content

Mus musculus — Organism part



Showing 23 of 23 experiments found:



By experiment type • Filters • Download table content



Baseline expression results

Baseline expression

Differential expression

Show anatomograms

Filter your results

Homo sapiens

Organism part

Cell line

Danio rerio

Developmental stage

Mus musculus

Organism part

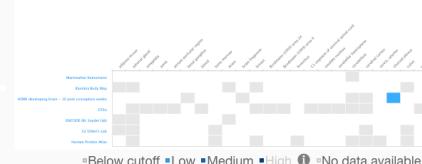
Cell type

Developmental stage

Homo sapiens — Organism part



Showing 7 of 7 experiments found:



Danio rerio — Developmental stage

Showing 1 of 1 experiment found:

Experiment: 18 White et al
Experimental condition: larval day 5
Expression level: 26 FPKM

By experiment type • Filters • Download table content

Mus musculus — Organism part



Showing 23 of 23 experiments found:



Expression levels are visualised using heatmaps

Below cutoff

Low

Medium

High



No data available



Baseline expression results

Results for CFHR2

Baseline expression Differential expression

Show anatomograms

Organism part

Filter your results

Homo sapiens

Organism part

Cell line

Danio rerio

Developmental stage

Mus musculus

Organism part

Cell type

Developmental stage

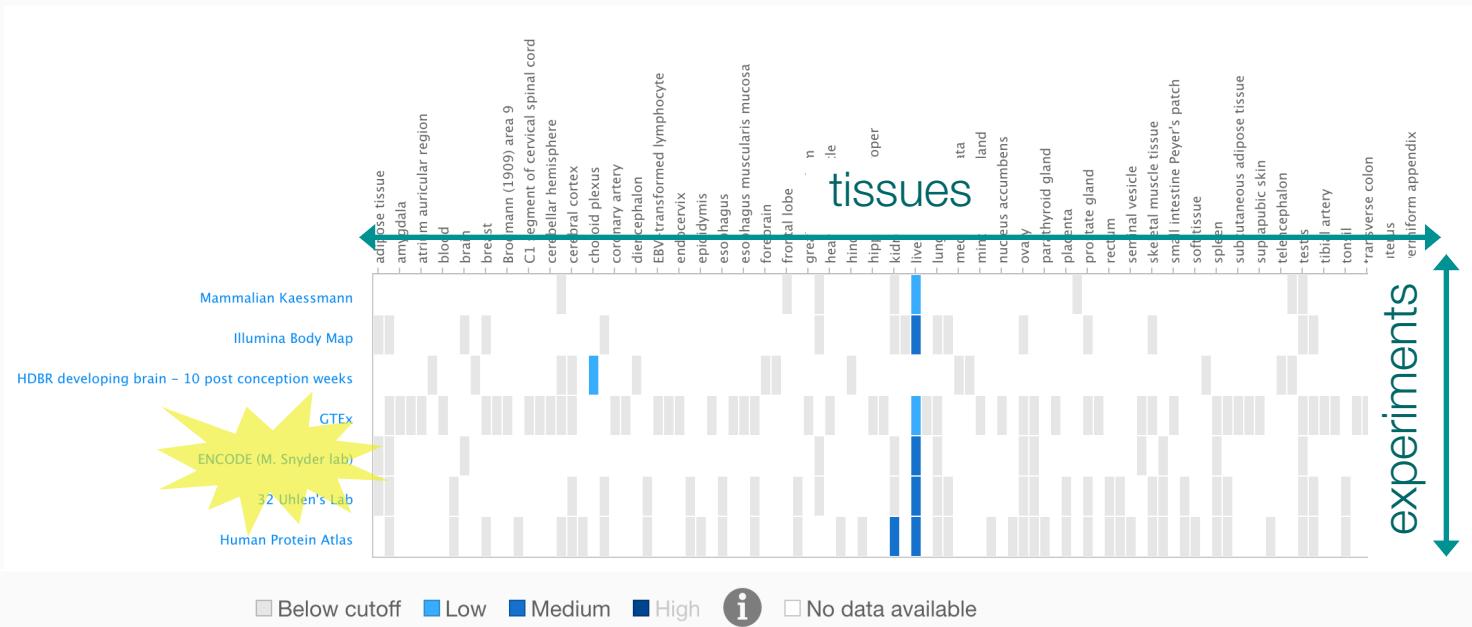
HEATMAP: For each species, *CFHR2* gene expression values for the selected condition (columns) in all experiments (rows) in which *CFHR2* gene is expressed

Showing 7 of 7 experiments found:

By experiment type ▾

Filters

Download table content



Below cutoff Low Medium High No data available



Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues
RNA-Seq mRNA baseline
Organism: *Homo sapiens*
Raw Data Provider: The Human Protein Atlas

Results Experiment Design Supplementary Information Download

Genes



Showing 50 of 34,068 genes found:
Expression level in FPKM
0 3.659

PRK2
MHN2
THPL1
PRK27
ANHBD7
FLC2
CAUS14
F2
C10orf68
TUBA1C
SMCP
F9
RODIL2
CYP1A2
ULP1
KRT77
SLC10A1
LCZ28
HAO1
FATE1
CTD-212N2R3.1
PCK2
SERPINA11
GDF1
AKAP1
SPATA8
PRB2
SLC10A1
ERVH46-1
SERPINA12
TSHR
SLC12A1
LCZ1C
S100G
LCZ1A
C16orf82
SP92
MBL2
C10orf62
UBR4N1
LCZ6A
HBLC
ACTL7A
TH
HMGB4
DDX4
C8orf12
ACSRG2
SLC34A1
PRK517

Specific checked

Most specific

Expression value

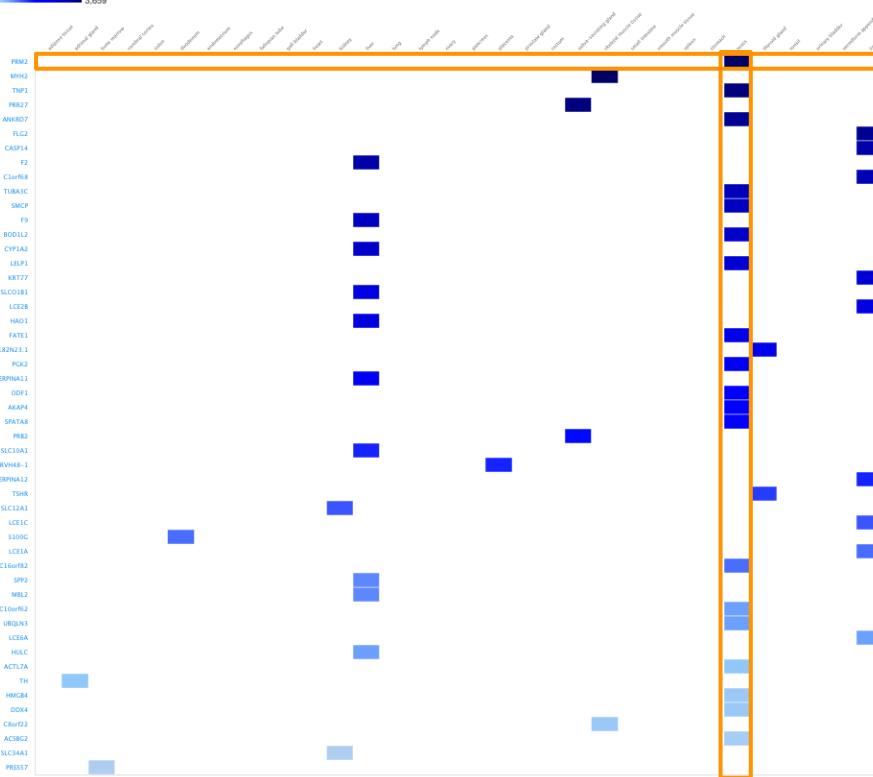
See distribution

Experimental variables

Select

Selected: 32 / 32

HEATMAP: top 50 genes (rows)
most specifically expressed
across all tissues (columns)





Baseline experiment page

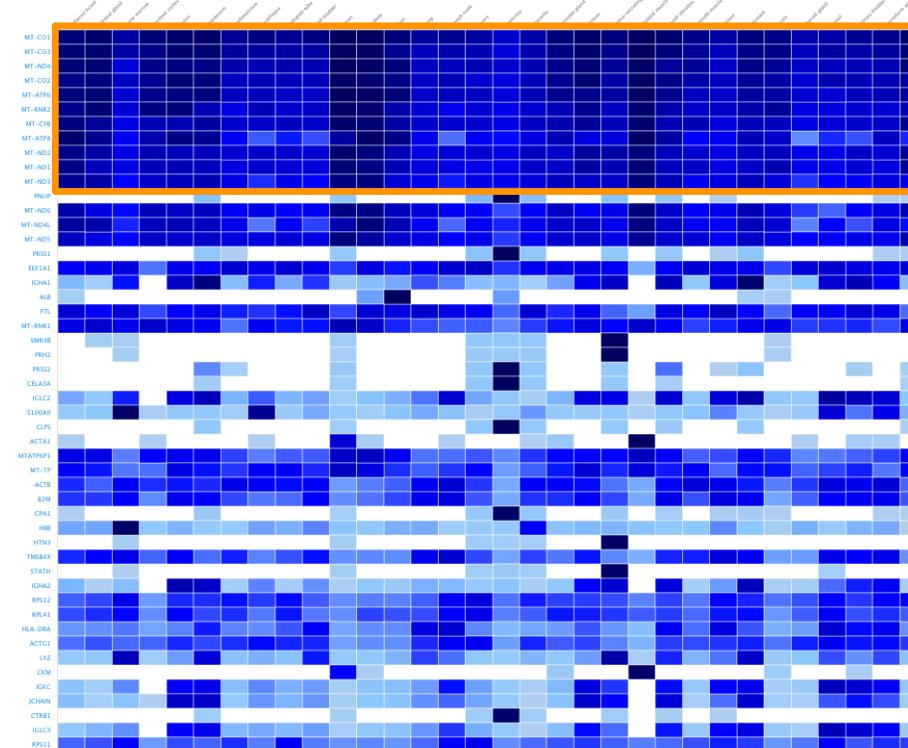
RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues
RNA-Seq mRNA baseline
Organism: Homo sapiens
Raw Data Provider: The Human Protein Atlas

Results Experiment Design Supplementary Information Download

Genes



Showing 50 of 34,068 genes found:
Expression level in FPKM
0 95.074



Genes with highest expression first

Specific unchecked

Most specific

Expression value

See distribution

Experimental variables

Select

Selected: 32 / 32

HEATMAP: top 50 genes (rows)
most highly expressed across all
tissues studied (columns)



Baseline experiment page

General information

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

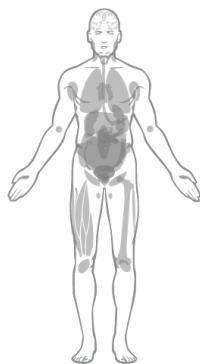
Results

Experiment Design

Supplementary Information

Downloads

Genes



Most specific

Expression value

0.5

See distribution

Experimental variables

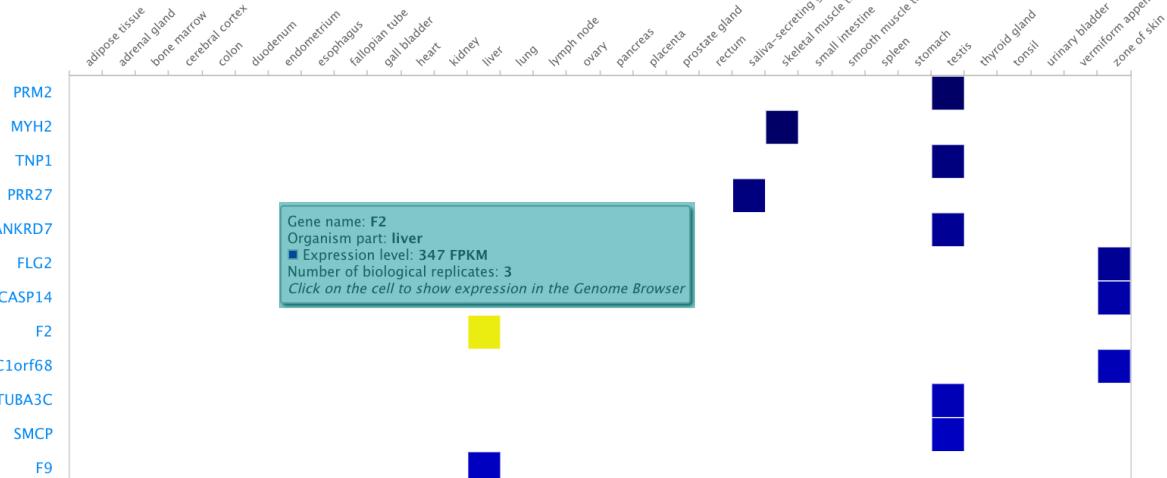
Select

Selected: 32 / 32

Genes expressed above 0.5 FPKM

Showing 50 of 33,970 genes found:

Expression level in FPKM
0 3,659



Download displayed results

Download table content



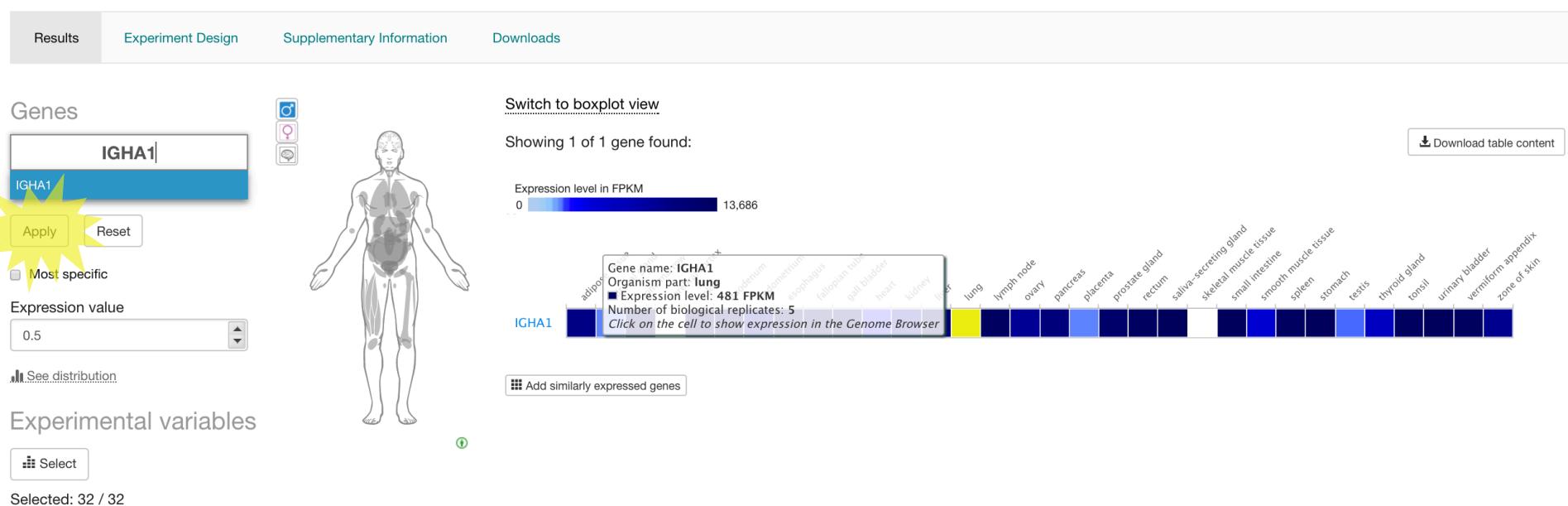
Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

Results Experiment Design Supplementary Information

Genes

IGHA1

IGHA1

Apply Reset

Most specific

Expression value

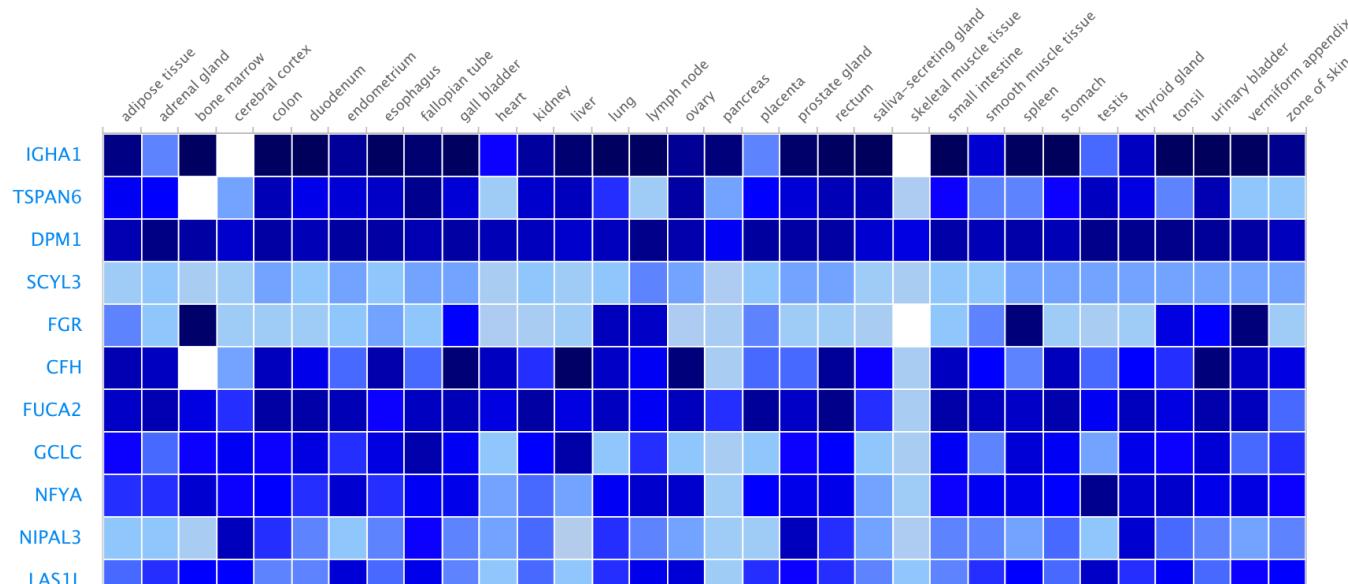
0.5

See distribution

Experimental variables

Select

Selected: 32 / 32



Display genes with similar expression to IGH1:

off 10 49



Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





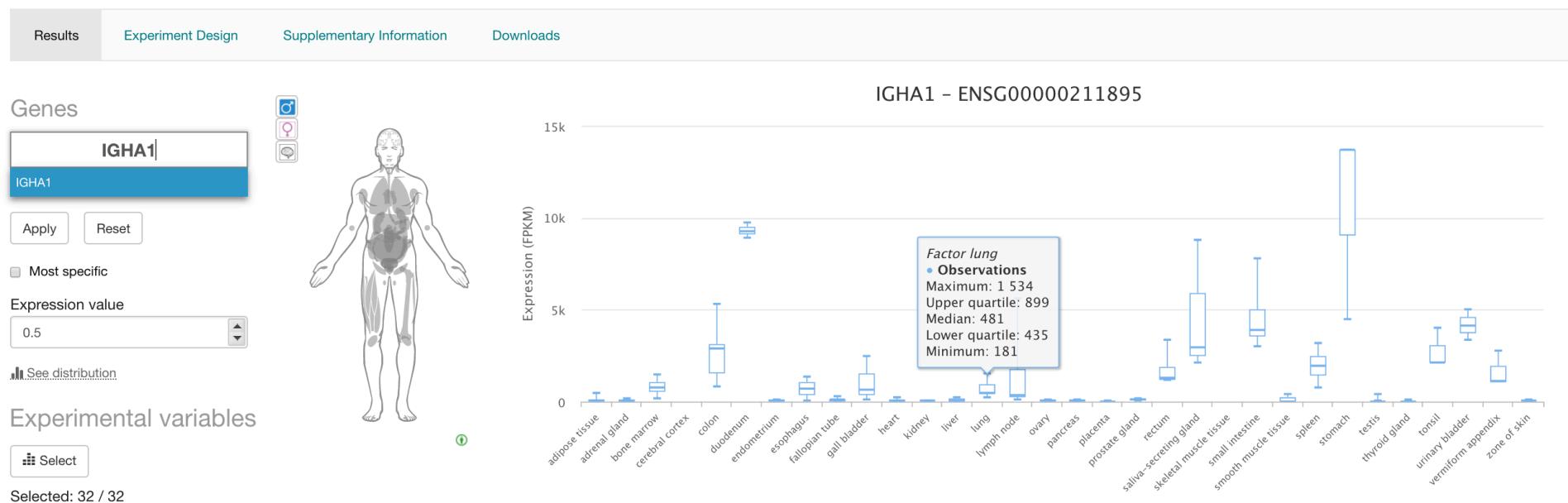
Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas



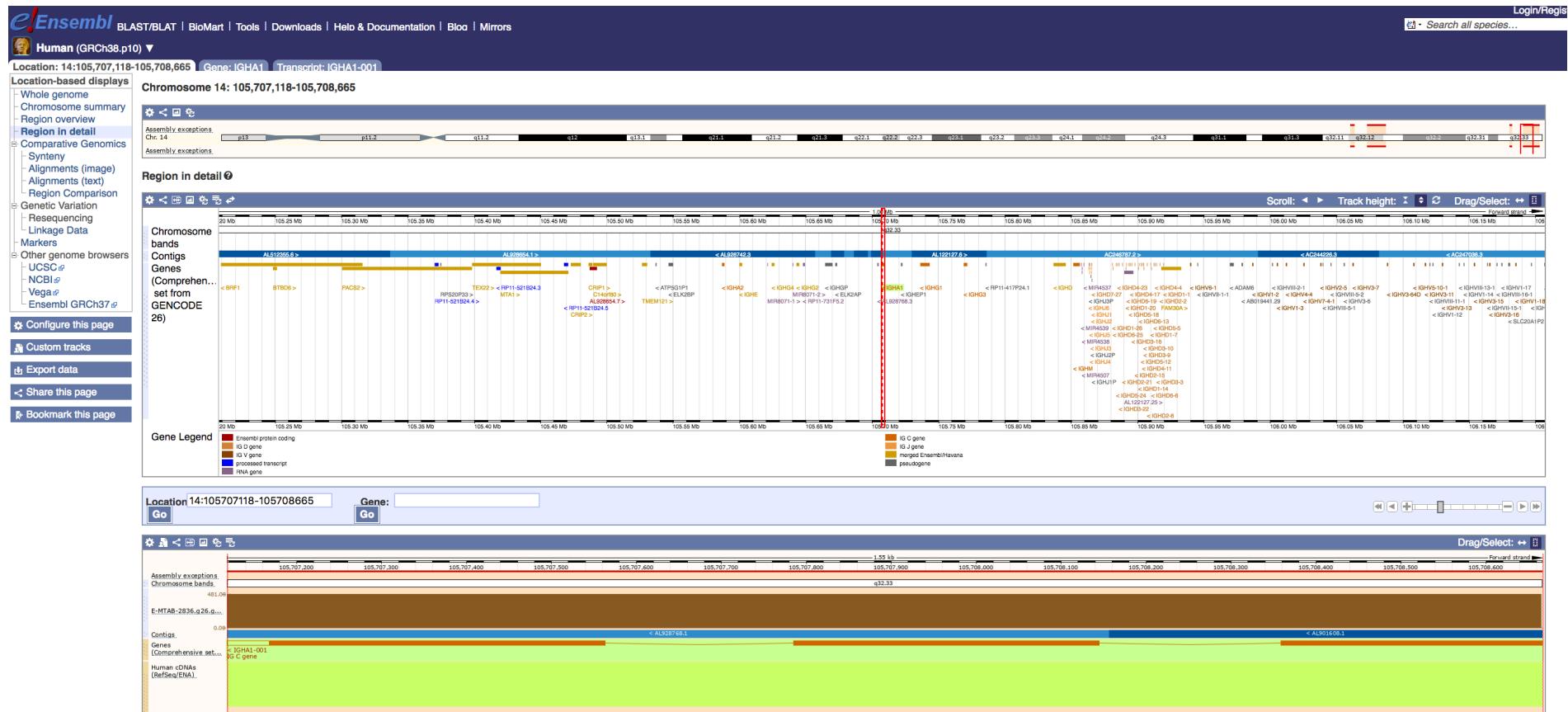
Experimental variables

Select

Selected: 32 / 32



Baseline experiment page





Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

Results Experiment Design Supplementary Information Downloads

Analysis Methods

Pipeline version	iRAP 0.5.1p1
Analyzed Libraries	Paired-end only
Filtering Step 1	Discard reads below minimum quality threshold
Filtering Step 2	Check of bacterial contamination; discard offending reads
Filtering Step 3	Discard reads with common uncalled characters (e.g. N)
Filtering Step 4	Remove reads from pair-end libraries that were orphaned by filtering steps 1-3
Read Mapping	Against genome reference (Ensembl release: 79) tophat2 version: 2.0.12
Quantification	htseq2 version: 0.6.1p1
Normalized Counts per Gene	(FPKMs) are calculated from the raw counts by iRAP. These are averaged for each set of technical replicates, and then quantile normalized within each set of biological replicates using limma. Finally, they are averaged for all biological replicates (if any)

Resources

 ArrayExpress: experiment E-MTAB-2836



Baseline experiment page

RNA-seq of coding RNA from tissue samples of 122 human individuals representing 32 different tissues

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

Results

Experiment Design

Supplementary Information

Downloads

Expression values across all genes

Summary of the expression results for this experiment ready to view in R

- Heatmap of aggregated expression data

Experiment Design (tsv)



Let's try Expression Atlas

Baseline expression

Hands-on activity

Look up expression of *TNNI3* gene in
normal human tissues



In pairs



Let's try Expression Atlas

Expression Atlas

Gene expression across species and biological conditions

Home | Download | Release notes | FAQ | Help | Licence | About

Exploring gene expression results across species under different biological conditions

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others. Expression Atlas aims to help answering questions such as 'where is a certain gene expressed?' or 'how its expression changes in a disease?'.

[Read more about Expression Atlas](#)

In which conditions is my favourite gene differentially expressed?

Organism

- ✓ Homo sapiens
- Mus musculus
- Arabidopsis thaliana
- Rattus norvegicus
- Drosophila melanogaster
- Zea mays

Any

- Anas platyrhynchos
- Anolis carolinensis
- Anopheles gambiae
- Arabidopsis lyrata
- Arabidopsis thaliana
- Aspergillus fumigatus
- Bos taurus
- Brachypodium distachyon
- Brassica oleracea
- Brassica rapa
- Caenorhabditis elegans

Biological conditions

Enter condition query...

Examples: lung, leaf, valproic acid, cancer

Discover and interpret gene expression analysis results quickly and easy



Differential expression results

Results for CFHR2 (symbol)

Baseline expression Differential expression

Display log₂-fold change -1.1 -2.7 6.5 1.2

Filter your results

Adjusted p-value Log₂-fold change

1.575 × 10⁻¹⁷ 6.5

CFHR2

'lung carcinoma' vs 'normal' in 'lung cell'

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

All organisms and all conditions in which *CFHR2* gene is differentially expressed

Experimental variables

- Growth condition
- Time
- Genotype
- Treatment
- Cell type
- Diet
- Disease
- Sampling time
- Infect
- Phenotype

Adjusted p-value Log₂-fold change

1.575 × 10⁻¹⁷ 6.5

Cfhr2

'lung carcinoma' vs 'normal' in 'lung cell'

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

Properties

Property	Test value (N=3)	Reference value (N=3)
cell type	lung cell	lung cell
disease	lung carcinoma	normal
organism	Mus musculus	Mus musculus
organism part	lung	lung
strain	C57BL/6	C57BL/6



Differential experiment page

General information

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): 25704820 (Filter by genes in paper)

Results

Plots

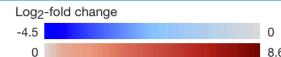
Experiment Design

Supplementary Information

Downloads

Differentially expressed genes

Showing 50 of 7,557 genes found:



Genes

Download displayed results

Most specific

Criteria to select DE genes

Log₂-fold change

Adjusted p-value

Comparisons

Selected: 3 / 3





Differential experiment page

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): 25704820 (Filter by genes in paper)

Results

Plots

Experiment Design

Supplementary Information

Downloads

Genes

Showing 50 of 7,557 genes found:

[Download table content](#)

Log₂-fold change

0 9.4

Specific unchecked

Most specific

- Up- or downregulated
- Upregulated only
- Downregulated only

Log₂-fold change

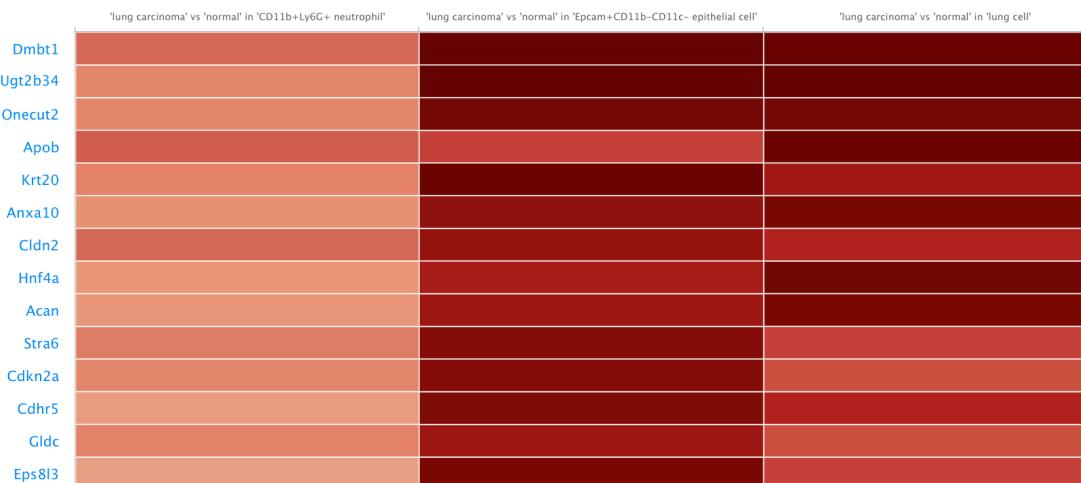
1.0

Adjusted p-value

0.05

Comparisons

Selected: 3 / 3





Differential experiment page

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): 25704820 (Filter by genes in paper)

Results

Plots

Experiment Design

Supplementary Information

Downloads

Genes

Showing 50 of 3,792 genes found:

[Download table content](#)

Log₂-fold change

-5.5 0



Specific unchecked

Most specific

Up- or downregulated
 Upregulated only
 Downregulated only

Log₂-fold change

1.0

Adjusted p-value

0.05

Comparisons

Selected: 3 / 3



Differential experiment page

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): 25704820 (Filter by genes in paper)

Results

Plots

Experiment Design

Supplementary Information

Downloads

'lung carcinoma' vs 'normal' in 'CD11b+Ly6G+ neutrophil'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes

'lung carcinoma' vs 'normal' in 'Epcam+CD11b-CD11c- epithelial cell'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes

'lung carcinoma' vs 'normal' in 'lung cell'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes



piano

Platform for integrative analysis of omics data

Bioconductor version: Release (3.5)

Piano performs gene set analysis using various statistical methods, from different gene level statistics and a wide range of gene-set collections. Furthermore, the Piano package contains functions for combining the results of multiple runs of gene set analyses.

Author: Leif Varemo <piano.rpkg@gmail.com> and Intawat Nookaew <piano.rpkg@gmail.com>

Maintainer: Leif Varemo <piano.rpkg@gmail.com>

Citation (from within R, enter `citation("piano")`):

Varemo L, Nielsen J and Nookaew I (2013). "Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods." *Nucleic Acids Research*, **41**(8), pp. 4378-4391. doi: [10.1093/nar/gkt111](https://doi.org/10.1093/nar/gkt111).



Differential experiment page

Transcriptome analysis of isolated stromal cells and tu

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): 25704820 (Filter by genes in paper)

Results Plots Experiment Design Supplementary Information Downloads

'lung carcinoma' vs 'normal' in 'CD11b+Ly6G+ neutrophil'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes

'lung carcinoma' vs 'normal' in 'Epcam+CD11b-CD11c- epithelial cell'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes

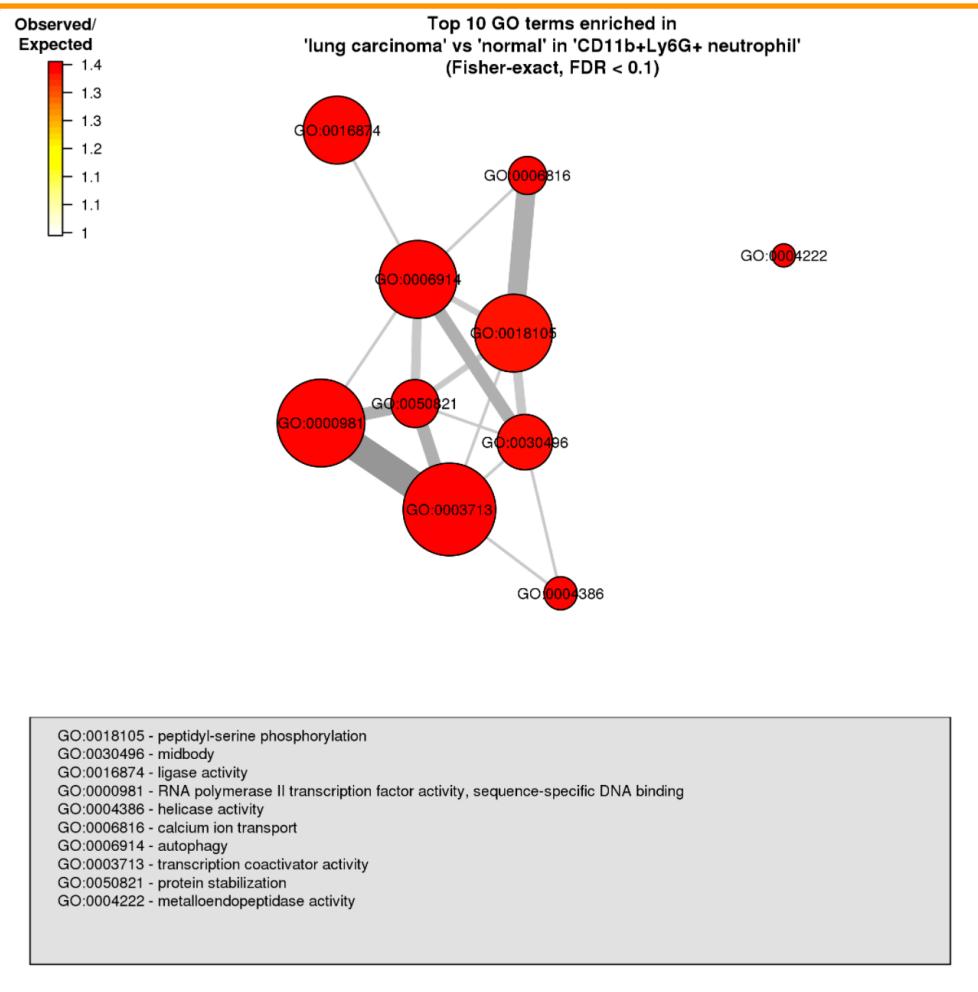
'lung carcinoma' vs 'normal' in 'lung cell'

GO terms enrichment analysis

Interpro domains enrichment analysis

Reactome pathways enrichment analysis

MA plot for the contrast across all genes





More information about *CFHR2* gene

Results for CFHR2 (symbol) AND Homo sapiens

ENSG00000080910 (CFHR2) Homo sapiens complement factor H related 2

Baseline expression

Differential expression

CFHR2 information

Synonym	HFL3, FHR2, CFHL2
Ortholog	ENSPTRG00000001804 , ENSPPYG00000000390 , ENSCING00000023181 , ENSOANG00000030269 , ENSTNIG00000012331 , ENSTNIG00000012333 , C54G4.4 (caenorhabditis elegans), ENSCSAG00000013569 (<i>chlorocebus sabaeus</i>), ENSCSAVG00000010763
Gene Ontology	extracellular region
InterPro	Sushi/SCR/CCP domain (domain)
Ensembl Family	COMPLEMENT FACTOR H RELATED PRECURSOR FHR H FACTOR H FACTOR
Ensembl Gene	ENSG00000080910
Ensembl Transcript	ENST00000496448 , ENST00000485647 , ENST00000476712 , ENST00000489703 , ENST00000473386 , ENST00000367415
Ensembl Protein	ENSP00000476677 , ENSP00000356385
Entrez	3080
UniProt	V9GYE7 , P36980
Gene biotype	protein_coding



More information about *CFHR2* gene

Ensembl

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh38.p10) ▾

Location: 1:196,943,772-196,959,226 Gene: CFHR2

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles
- Sequence
- Secondary Structure
- Comparative Genomics
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Ensembl protein families

Ontologies

- GO: Cellular component
- GO: Molecular function
- GO: Biological process

Phenotypes

Genetic Variation

- Variant table
- Variant image
- Structural variants

Gene expression

Regulation

External references

Supporting evidence

Description complement factor H related 2 [Source:HGNC Symbol;Acc:[HGNC:4890](#)]

Synonyms FHR2, HFL3, CFHL2

Location Chromosome 1: 196,943,772-196,959,226 forward strand.
GRCh38:CM000663.2

About this gene This gene has 6 transcripts ([splice variants](#), [16 orthologues](#), [11 paralogues](#)), is a member of [1 Ensembl protein family](#) and is associated with [3 phenotypes](#).

Transcripts [Hide transcript table](#)

Show/hide columns (1 hidden)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
CFHR2-001	ENST00000367415.6	1059	270aa	Protein coding	CCDS30959	P36980	NM_005666 NP_005657	TSL:1 GENCODE basic APPRIS P2
CFHR2-005	ENST00000476712.6	1020	254aa	Protein coding	-	V9GYE7	-	TSL:3 GENCODE basic APPRIS ALT2
CFHR2-006	ENST00000496448.5	811	No protein	Processed transcript	-	-	-	TSL:2
CFHR2-002	ENST00000473386.1	599	No protein	Processed transcript	-	-	NM_001312672	TSL:1
CFHR2-004	ENST00000489703.1	424	No protein	Processed transcript	-	-	-	TSL:3
CFHR2-003	ENST00000485647.1	376	No protein	Processed transcript	-	-	-	TSL:5



Let's try Expression Atlas

Differential expression

Hands-on activity

Find differentially expressed genes in patients with hepatocellular carcinoma

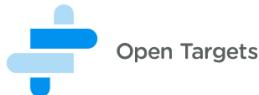


In pairs



Expression Atlas data in other resources

www.opentargets.org



About Us Projects People News Jobs Contact Open Targets Platform ➔

Welcome to Open Targets

For biomedical researchers who need to identify a biological target for a new therapy, Open Targets is a public-private initiative to generate evidence on the validity of therapeutic targets based on genome-scale experiments and analysis. Open Targets is working to create an R&D framework that applies to a wide range of human diseases, and is committed to sharing its data openly with the scientific community.

[About Open Targets ➔](#)



Copyright (c) 2016 Open Targets. All rights reserved.

Site maintained by the Web Development team at EMBL-EBI | [Terms of Use](#) | [Privacy](#) | [Cookies](#)



Expression Atlas data in other resources



Open Targets

About Us

Projects

People

News

Jobs

Contact

Open Targets Platform ➔

Welcome to Open Targets

For biomedical researchers who need to identify a biological target for a
of therapeutic targets based on genome-scale experiments and analysis
human diseases, and is committed to sharing its data openly with the sc

About Open Targets ➔



Copyright (c) 2016 Open Targets. All rights reserved.

Site maintained by the Web Development team at EMBL-EBI | Terms of Use | Privacy | Co



Open Targets Platform

Find new targets for drug discovery

CYP1A2



CYP1A2

cytochrome P450 family 1 subfamily A member 2

156 diseases associated

Target

Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics. Most active in catalyzing 2-hydrox...



Expression Atlas data in other resources

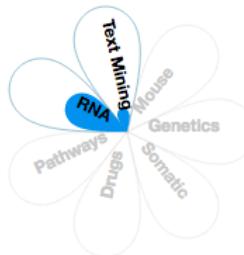
156 diseases associated with CYP1A2

 View CYP1A2 profile



Expression Atlas data in other resources

Evidence for CYP1A2 in hepatocellular carcinoma



CYP1A2

cytochrome P450 family 1 subfamily A member 2

Synonyms: P3-450, CP12

Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including many drugs and environmental pollutants.

hepatocellular carcinoma

Synonyms: Cancer, Liver, Hepatocellular Carcinomas, Liver Cancer, Adult, NEOPL HEPATIC, Hepatomas, Primary Car...

Tumors or cancer of the LIVER.. A primary malignant neoplasm of epithelial liver cells. It ranges from a well-differentiated tumor with EPITHELIAL CELLS indistinguishable from normal HEPATOCYTES to a ...

Genetic associations

Somatic mutations

Drugs

Affected pathways

RNA expression

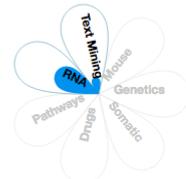
Text mining

Animal models



Expression Atlas data in other resources

Evidence for CYP1A2 in hepatocellular carcinoma



CYP1A2
cytochrome P450 family 1 subfamily A member 2
Synonyms: P3-450, CP12
Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrel...

hepatocellular carcinoma
Synonyms: Cancer, Liver, Hepatocellular Carcinomas, Liver Cancer, Adult, NEOPL HEPATIC, Hepatomas, Primary Car...
Tumors or cancer of the LIVER.. A primary malignant neoplasm of epithelial liver cells. It ranges from a well-differentiated tumor with EPITHELIAL CELLS indistinguishable from normal HEPATOCYTES to a ...

Genetic associations

Somatic mutations

Drugs

Affected pathways

RNA expression

Source: Expression Atlas

Showing 1 to 4 of 4 entries

Search:



Gene-disease evidence

Disease	Comparison	Activity	Tissue/cell	Evidence source	Log ₂ fold change	P-value	Percentile rank	Experiment overview and data	Publications
hepatocellular carcinoma	'liver tumor' vs 'adjacent non-tumor tissue'	decreased ↘	liver	differential gene expression evidence from RNA-seq experiment	-10.1	3.32e-63	100	Transcriptome sequencing of human hepatocellular carcinoma ↗	
hepatocellular carcinoma	'tumor' vs 'paramutormal tissue'	decreased ↘	liver	differential gene expression evidence from RNA-seq experiment	-3.3	3.67e-6	98	Transcriptomic characterization of Hepatocellular Carcinoma with CTNNB1 mutation ↗	



Expression Atlas data in other resources

Source: Expression Atlas

Showing 1 to 4 of 4 entries

Search:



Disease	Gene-disease evidence							Experiment overview and data	Publications
	Comparison	Activity	Tissue/cell	Evidence source	Log ₂ fold change	P-value	Percentile rank		
hepatocellular carcinoma	'liver tumor' vs 'adjacent non-tumor tissue'	decreased ↘	liver	differential gene expression evidence from RNA-seq experiment	-10.1	3.32e-63	100	Transcriptome sequencing of human hepatocellular carcinoma ↗	
hepatocellular carcinoma	'tumor' vs 'paratumoral tissue'	decreased ↘	liver	differential gene expression evidence from RNA-seq experiment	-3.3	3.67e-6	98	Transcriptomic characterization of Hepatocellular Carcinoma with CTNNB1 mutation ↗	
hepatocellular carcinoma	'minor tumor' vs 'normal' in 'IL28B rs8099917 TG/GG'	decreased ↘	liver	differential gene expression evidence from microarray experiment	-3.3	4.93e-3	99	Hepatic gene expression of HCV related Hepatocellular carcinoma and non-cancerous tissue with IL28B rs8099917 TT genotype and TG/GG genotype ↗	



New project on single-cell...

EMBL-EBI

Services

Research

Training

About us



EMBL-EBI  Hinxton ▾



Single-cell Expression Atlas

Single-cell gene expression across species



Analysing single-cell gene expression under different biological conditions

The Single-Cell Expression Atlas provides information on single-cell gene expression patterns under different biological conditions.

Browse experiments

Total : 3 experiments

B 06-02-2017 Single-cell RNA-sequencing resolves a CD4+ T cell fate bifurcation 752 assays – *Mus musculus*

B 06-02-2017 Single-cell RNA-seq analysis of human pancreas from healthy individuals and type 2 diabetes patients 3514 assays – *Homo sapiens*

Publications

 Expression Atlas update – an integrated database of gene and protein expression in humans, animals and plants.
Nucleic Acids Research, 19 October 2015.

 RNA-Seq Gene Profiling – A Systematic Empirical Comparison
PLoS One, 30 September 2014.

 Expression Atlas update – a database of gene and transcript expression from microarray and sequencing-based functional genomics experiments.
Nucleic Acids Research, 4 December 2013.



New project on single-cell...

[Home](#) [EMBL-EBI](#) [Services](#) [Research](#) [Training](#) [About us](#) [Search](#)

EMBL-EBI Hinxton ▾



Malaria infected mouse cells - Teichmann et al

Single-cell RNA-sequencing resolves a CD4+ T cell fate bifurcation

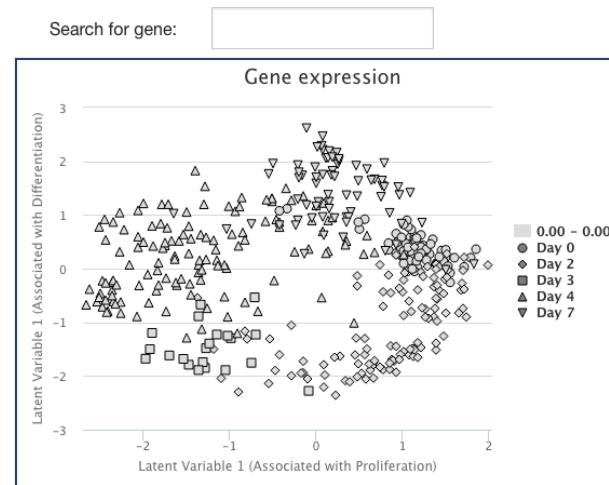
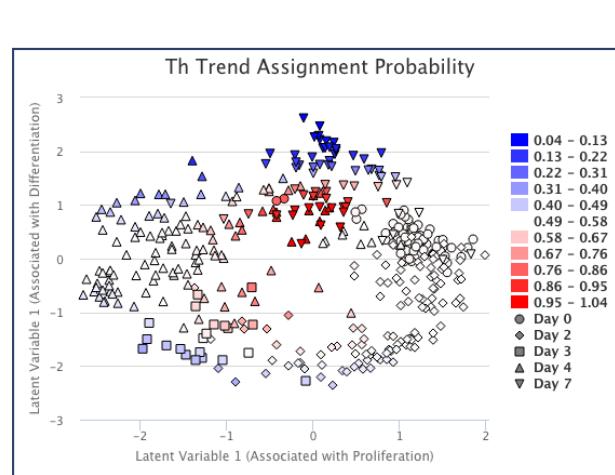
Experiment information

Organism: *Mus musculus*

752 single cells

Last updated: 06-02-2017

Reference plot and gene expression





Expression Atlas: Who we are?

Robert
Petryszak
Team Leader



Irene
Papatheodorou
Co-ordinator

Data curation



Laura Huerta



Amy Tang



Anja Fullgrabe

Data analysis



Maria Keays



Nuno Fonseca



Alfonso Fuentes



Elisabet Barrera



Wojtek Bazant

User Interface



EMBL-EBI

Getting help and sending feedback

The screenshot shows the Expression Atlas homepage. At the top left is a blue circular logo with a yellow and white 'Q'. To its right is the text 'Expression Atlas' and 'Gene expression across species and biological conditions'. Below this is a horizontal navigation bar with six items: 'Home', 'Download', 'Release notes', 'FAQ' (which is highlighted in green), 'Help', 'Licence', and 'About' (which is also highlighted in green). The main content area has a teal background. It features a large heading 'About Expression Atlas', followed by 'What is Expression Atlas?'. Below this is a paragraph about the project's mission to provide freely available information on gene and protein expression across various biological conditions. Another paragraph explains the project's aim to answer questions like 'where is my favourite gene expressed?' and 'how its expression changes in a disease?'. A final paragraph describes the project's objective of involving data curation, analysis, and development of a web application. To the right of the text is a diagram consisting of four teal boxes labeled 'Data curation', 'Data analysis', 'Data search', and 'Data visualisation'. Below these are four smaller teal boxes: 'High quality data' with a document icon, 'High quality results' with a bar chart icon, 'Accessible results' with a sunburst chart icon, and '... easy to interpret' with a grid icon. A grey arrow points from the bottom of the first column to the bottom of the second column.

About Expression Atlas

What is Expression Atlas?

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression. Our mission is to provide the scientific community with freely available information on the abundance and localisation of RNA (and proteins) across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others.

Expression Atlas aims to help answering questions such as 'where is my favourite gene expressed?' or 'how its expression changes in a disease?'

To achieve this objective, our project involves data curation, data analysis and the development of a web application to access and visualise publicly available data.

atlas-feedback@ebi.ac.uk



@ExpressionAtlas

EMBL-EBI: Exploring genomic, protein and chemical data, Oviedo

Expression Atlas service at EMBL-EBI

Laura Huerta, PhD

Scientific Database Curator, Gene Expression Team

lauhuema@ebi.ac.uk

17 May 2017

