

# Introduction to Next Generation Sequencing

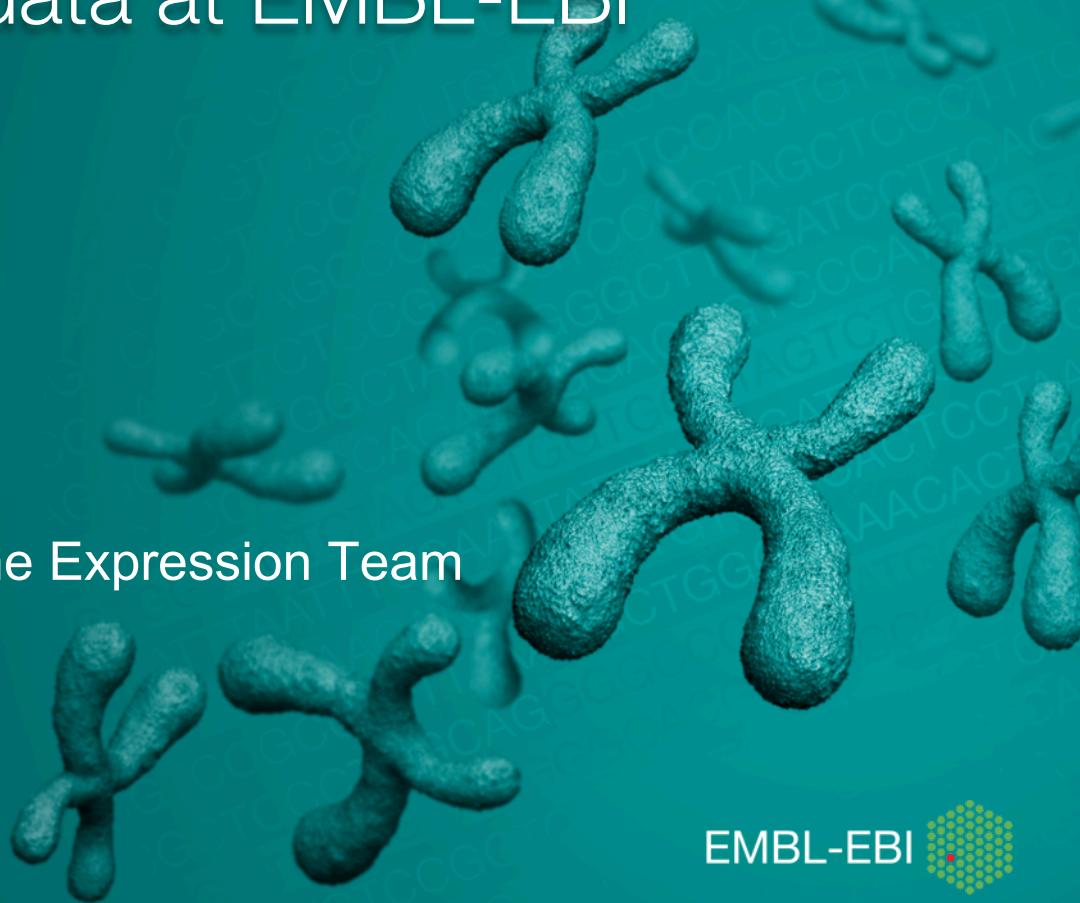
Public resources for accessing  
RNA-seq data at EMBL-EBI

Laura Huerta, PhD

Senior Scientific Curator, Gene Expression Team

[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)

10 October 2017





UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# *A bit about your speaker...*



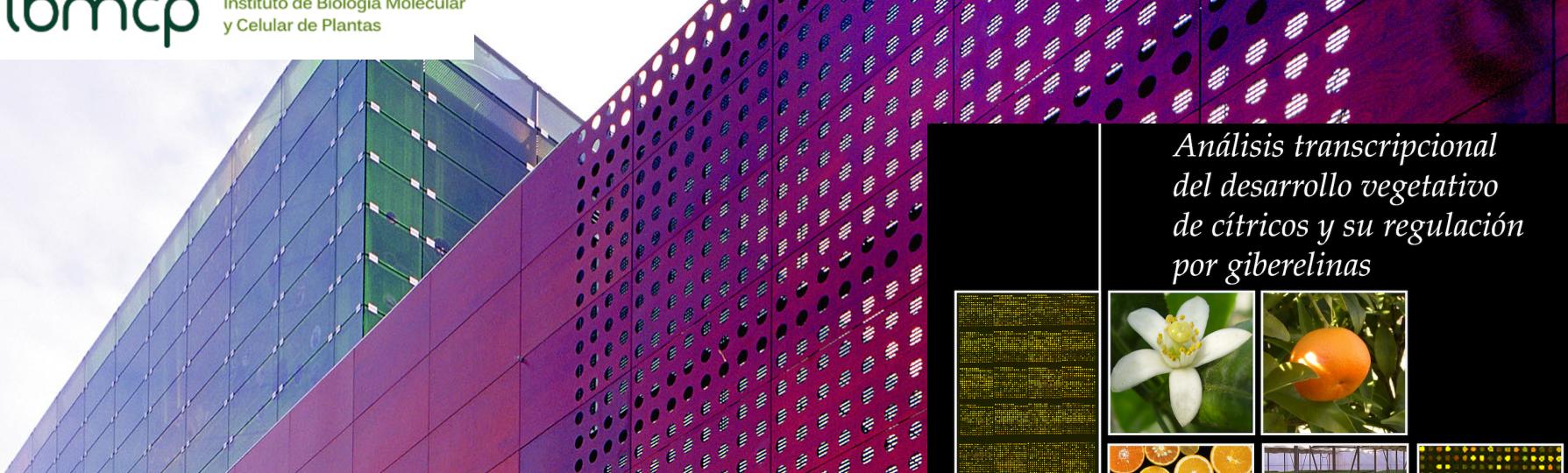
- *Agronomic Engineer, major in Biotechnology*

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

# *A bit about your speaker...*

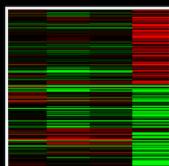
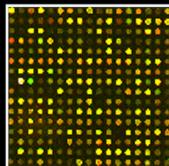


Instituto de Biología Molecular  
y Celular de Plantas



<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...*

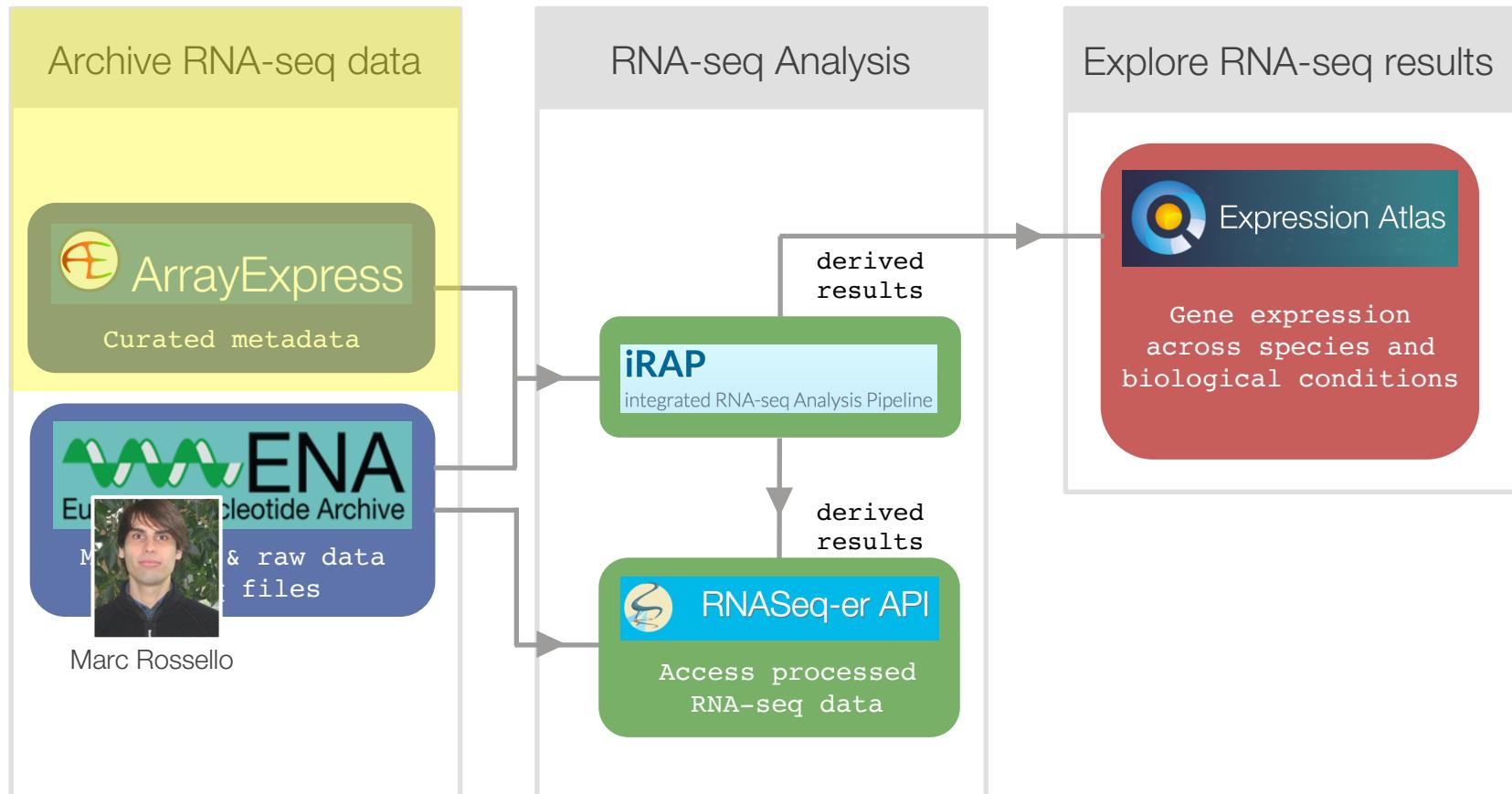




# *Outline of the session*

- ✓ RNA-seq: from raw data to analysis results
- ✓  ArrayExpress: archive RNA-seq data
- ✓  Expression Atlas: search & visualise RNA-seq results
  - ✓ iRAP: RNA-seq analysis pipeline
  - ✓ Expression Atlas hands-on activities
- ✓  RNASeq-er API: access RNA-seq results

# RNA-seq: from raw data to analysis results





# ArrayExpress: functional genomics data archive

 ArrayExpress

Home | Browse | Submit | Help | About ArrayExpress | Contact Us | Login

Search Examples: E-MEXP-31, cancer, p53, Geuvadis | [advanced search](#)

## 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](mailto:annotare@ebi.ac.uk), or send us questions about ArrayExpress at [arrayexpress@ebi.ac.uk](mailto: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

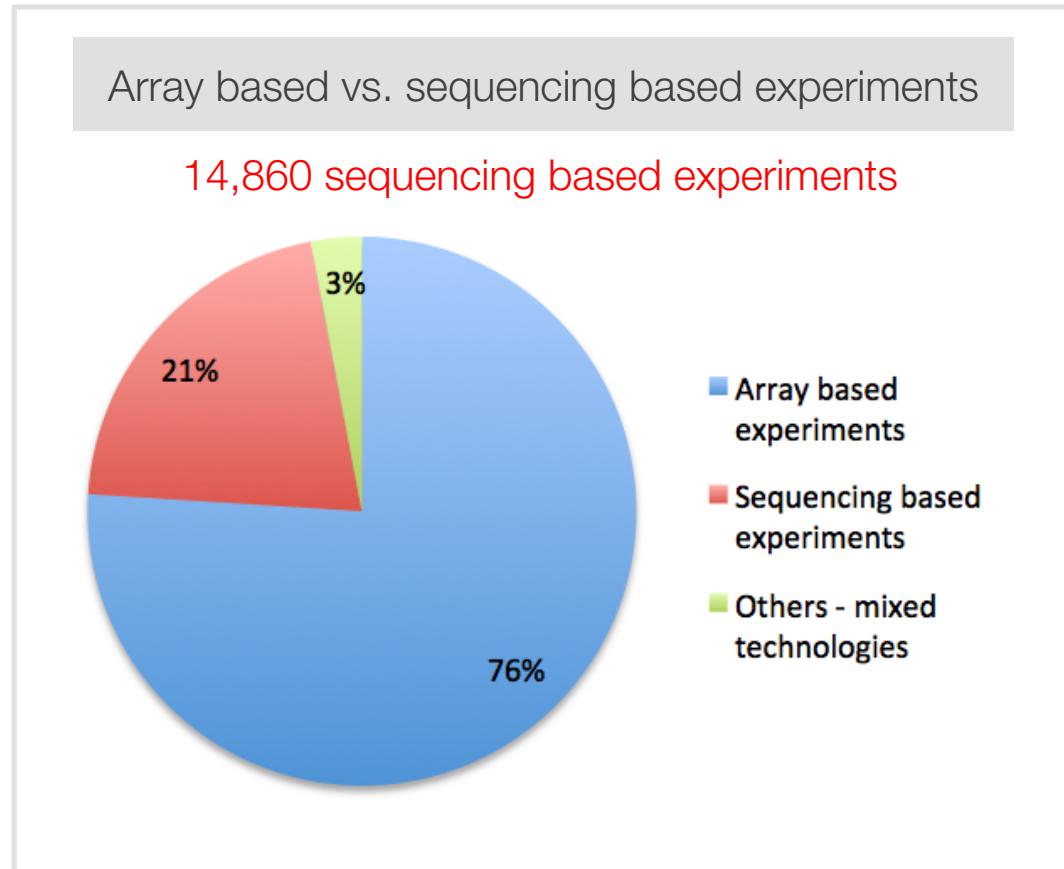
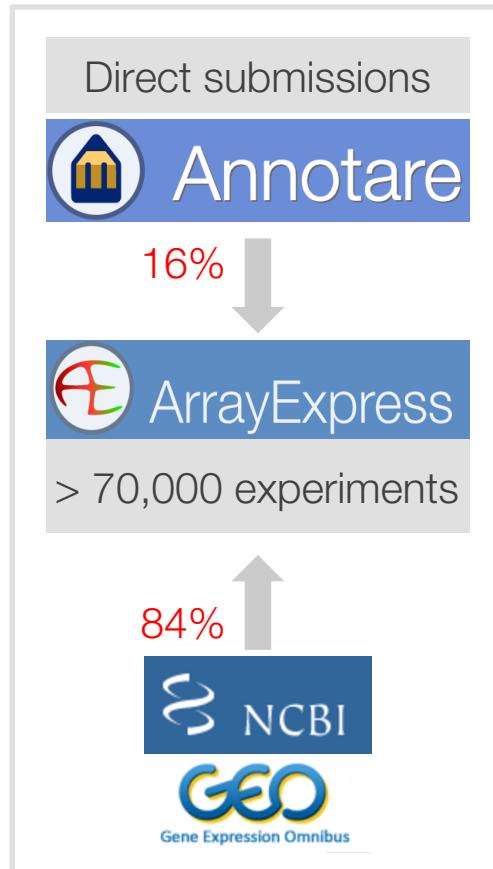
Discover up and down regulated genes in numerous experimental conditions in the [Expression Atlas](#).

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

[www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)

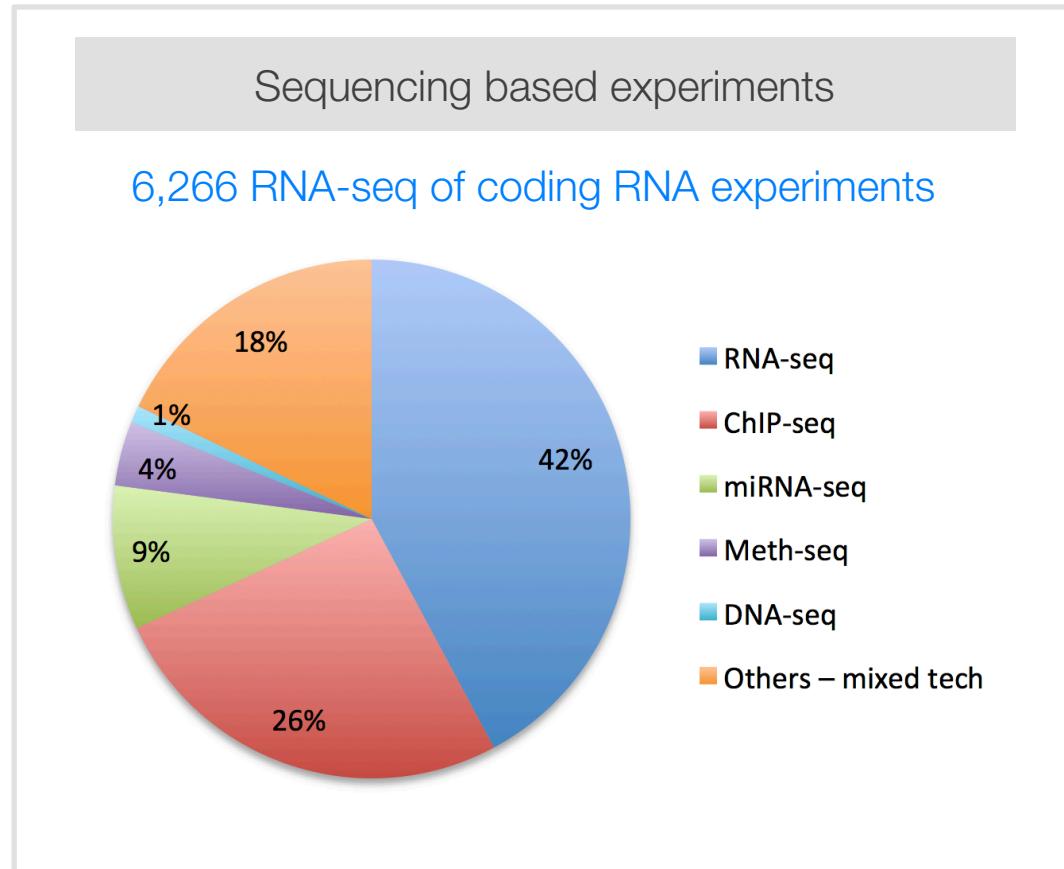
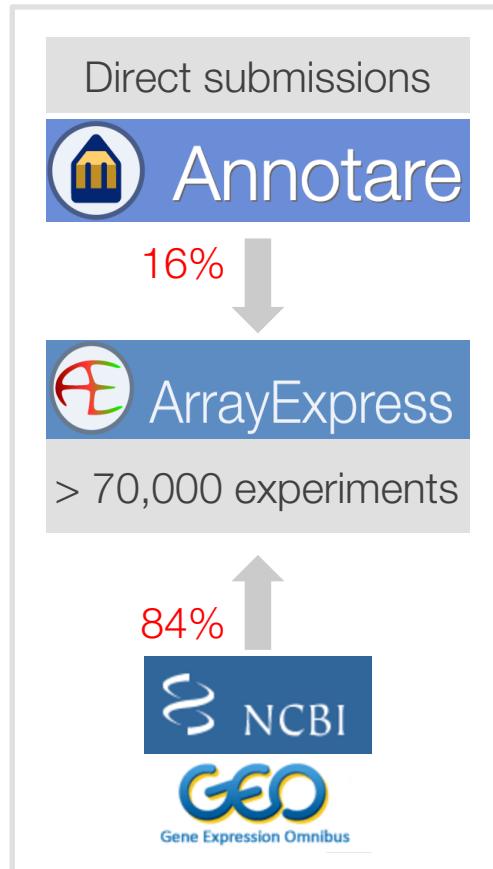


# ArrayExpress: functional genomics data archive





# ArrayExpress: functional genomics data archive





# ArrayExpress: functional genomics data archive

[Home](#)[Browse](#)[Submit](#)[Help](#)[About ArrayExpress](#)

## Experiment types in ArrayExpress

Experiment Type	Definition (EFO term accession)	Example
4C	EFO_0007690	E-MTAB-2180
antigen profiling	EFO_0000747	E-MTAB-3606
ATAC-seq	EFO_0007045	E-MTAB-3972
Bisulfite-seq	EFO_0003753	E-MTAB-1042
Capture-C	EFO_0007691	E-MTAB-4845
ChIP-chip by array	EFO_0002760	E-MTAB-2804
ChIP-chip by SNP array	EFO_0002764	E-GEOID-22306
ChIP-chip by tiling array	EFO_0002762	E-MTAB-1402
ChIP-seq	EFO_0002692	E-MTAB-3631
CLIP-seq	EFO_0003143	E-MTAB-1371
comparative genomic hybridization by array	EFO_0000749	E-MTAB-2293
DNA-seq	EFO_0002693	E-MTAB-3109
FAIRE-seq	EFO_0004428	E-MTAB-3199

[www.ebi.ac.uk/arrayexpress/help/experiment\\_types.html](http://www.ebi.ac.uk/arrayexpress/help/experiment_types.html)



# ArrayExpress: functional genomics data archive

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments involving human autoimmune disease samples*

main search box

The screenshot shows the ArrayExpress homepage with a blue header. On the left is the logo and the text "ArrayExpress". To the right are navigation links: Home, Browse (which is highlighted in a red box), Submit, Help, About ArrayExpress, Contact Us, and Login. Below the header is a search bar with a placeholder "Search" and a magnifying glass icon. Underneath it is a link "Examples: E-MEXP-31, cancer, p53, Geuvadis" and a link to "advanced search". At the bottom left, there is a yellow button with a funnel icon and the text "Filter search results". A red box surrounds this button and the search bar area. The text "filter search box" is written in red at the bottom left of the red box.

[www.ebi.ac.uk/arrayexpress/browse.html](http://www.ebi.ac.uk/arrayexpress/browse.html)



# ArrayExpress: functional genomics data archive

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments involving human autoimmune disease samples*

Filter search results

By organism:

Homo sapiens

By experiment type:

RNA assay

Sequencing assay

By array:

All arrays

ArrayExpress data only

[Reset filters](#)

[Filter](#)

"autoimmune dis

**autoimmune disease**

- Autoimmune Hepatitis
- Behcet's syndrome
- CNS demyelinating **autoimmune disease**
- Eosinophilia-Myalgia Syndrome
- Guillain-Barre syndrome
- Myasthenia gravis
- Sjogren syndrome
- Susac Syndrome
- Vitiligo
- Wegener's granulomatosis
- anti-neutrophil antibody associated vasculitis
- antiphospholipid antibodies
- arthritis
- autoimm

Efficient search via  
ontology-driven query  
expansion



# Ontology-driven query expansion

 **Ontology Lookup Service**

Home | **Ontologies** | Documentation | About

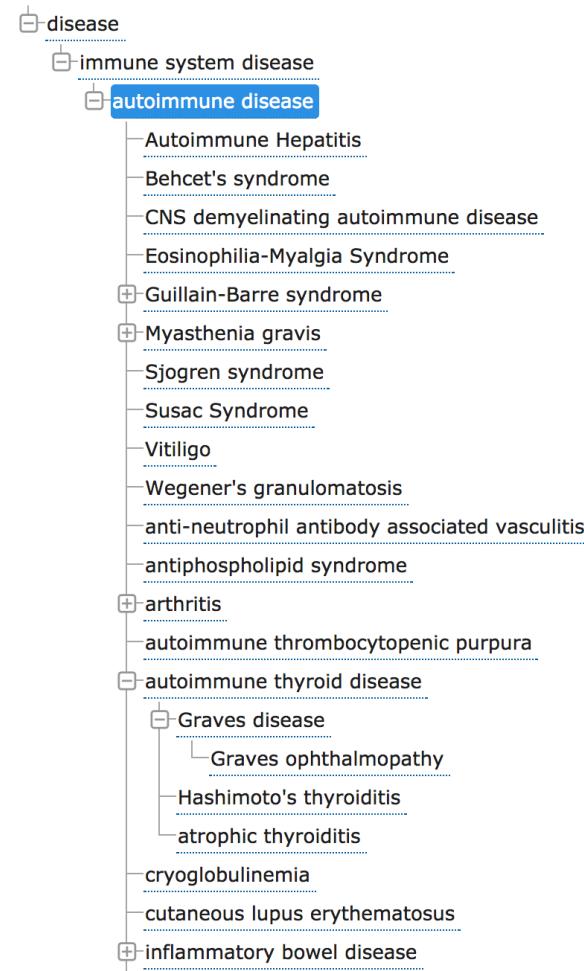
OLS > Experimental Factor Ontology | EFO > EFO:0005140 | 

## autoimmune disease

 [http://www.ebi.ac.uk/efo/EFO\\_0005140](http://www.ebi.ac.uk/efo/EFO_0005140) | 

Much richer queries by  
using the hierarchy within  
the ontology

Return subclasses when a  
higher level superclass is  
used in the query





# ArrayExpress: functional genomics data archive

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments involving human autoimmune disease samples*

Search results for "autoimmune disease"

Filtered by organism **Homo sapiens**, experiment type "sequencing assay", experiment type "rna assay"

Page 1 2

Showing 1 - 25 of 36 experiments

Page size 25 50 100 250 500

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEOD-60424	Next generation sequencing of human immune cell subsets across diseases	RNA-seq of coding RNA	Homo sapiens	134	06/01/2015	-		3054	
E-MTAB-1568	RNA-seq of coding RNA in human plasma cells isolated from individuals with colorectal cancer, ulcerative colitis and normal individuals	RNA-seq of coding RNA	Homo sapiens	9	29/03/2013	-		1078	-
E-GEOD-46579	A blood based 12-miRNA signature of Alzheimer patients	RNA-seq of non coding RNA	Homo sapiens	70	17/06/2013	-		824	-
E-GEOD-57945	Core Ileal Transcriptome in Pediatric Crohn Disease	RNA-seq of coding RNA	Homo sapiens	359	24/07/2014			623	-
E-MTAB-4304	RNA-Seq analysis of human intact and damaged osteoarthritic cartilage following total knee replacement	RNA-seq of coding RNA	Homo sapiens	22	06/04/2016	-		487	-
E-GEOD-83139	Single cell RNA-seq of human pancreatic endocrine cells from Juvenile, adult control and type 2 diabetic donors	RNA-seq of coding RNA	Homo sapiens	635	05/07/2016	-		328	-



# ArrayExpress: functional genomics data archive

## E-GEO-60424 - Next generation sequencing of human immune cell subsets across diseases

Status *Released on 6 January 2015, last updated on 25 February 2016*

Organism **Homo sapiens**

Samples (134) [Click for detailed sample information and links to data](#)  
*↳ found inside: multiple sclerosis, Type 1 Diabetes*

Protocols (2) [Click for detailed protocol information](#)

Description This study compared whole transcriptome signatures of 6 immune cell subsets and whole blood from patients with an array of immune-associated diseases. Fresh blood samples were collected from healthy subjects and subjects diagnosed **type 1 diabetes**, amyotrophic lateral sclerosis, and sepsis, as well as **multiple sclerosis** patients before and 24 hours after the first treatment with IFN-beta. At the time of blood draw, an aliquot of whole blood was collected into a Tempus tube (Invitrogen), while the remainder of the primary fresh blood sample was processed to highly pure populations of neutrophils, monocytes, B cells, CD4 T cells, CD8 T cells, and natural killer cells. RNA was extracted from each of these cell subsets, as well as the whole blood samples, and processed into RNA sequencing (RNAseq) libraries (Illumina TruSeq). Sequencing libraries were analyzed on an Illumina HiScan, with a target read depth of ~20M reads. Reads were demultiplexed, mapped to human gene models (ENSEMBL), and tabulated using HTSeq. Read count data were normalized by the TMM procedure (edgeR package). We performed whole genome RNAseq profiling of immune cell subsets and whole blood from subjects with an array of immune-associated diseases.

Experiment type **RNA-seq of coding RNA**

Contacts Scott Presnell <SPresnell@benaroyaresearch.org>, Carla J Greenbaum, Cate Speake, Damien Chaussabel, Elizabeth Whalen, Jane H Buckner, Kimm K O'Brien, Michael J Mason, Peter S Linsley, Quynh-Anh Nguyen, Scott R Presnell, Uma Malhotra, Vivian H Gersuk

Citation [Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis.](#) Linsley PS, Speake C, Whalen E, Chaussabel D., [Europe PMC 25314013](#)

MINSEQE \* \* \* - \*

Exp. design Protocols Variables Processed Seq. reads

Files Investigation description  [E-GEO-60424.idf.txt](#)  
Sample and data relationship  [E-GEO-60424.sdrf.txt](#)  
Additional data (1)  [E-GEO-60424.additional.1.zip](#)

Links [Click to browse all available files](#)

Links Expression Atlas - E-GEO-60424  
ENA - SRP045500, GEO - GSE60424  
Send E-GEO-60424 data to



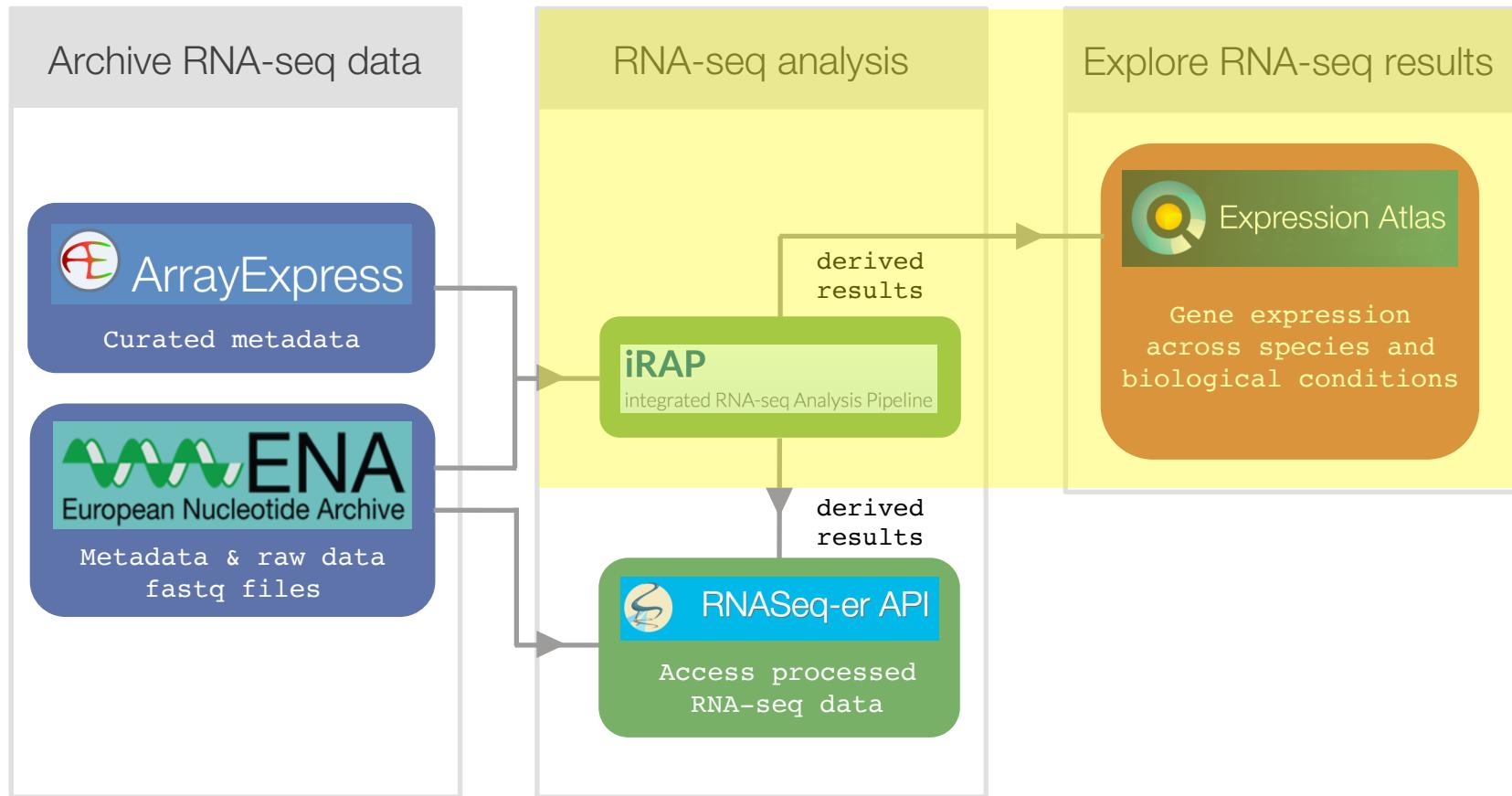
# ArrayExpress: functional genomics data archive

## E-GEOD-60424 - Next generation sequencing of human immune cell subsets across diseases

[Display full sample-data table](#)[Export table in Tab-delimited format](#)Page [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) .. [11](#)Showing **1 - 25** of **268** rowsPage size [25](#) [50](#) [100](#) [250](#) [500](#)

Source Name	name	Sample Attributes								Links to Data	
		age	cell type	disease	individual	organism	ethnic group	sex	clinical information	cell type	disease
GSM1479433 1	32 (year) whole blood	normal	44	Homo sapiens	hispanic	female			whole blood	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479433 1	32 (year) whole blood	normal	44	Homo sapiens	hispanic	female			whole blood	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479434 1	52 (year) whole blood	multiple sclerosis	31	Homo sapiens	white	female	smoker		whole blood	multi	<a href="#">View</a> <a href="#">Download</a>
GSM1479434 1	52 (year) whole blood	multiple sclerosis	31	Homo sapiens	white	female	smoker		whole blood	multi	<a href="#">View</a> <a href="#">Download</a>
GSM1479435 1	52 (year) whole blood	multiple sclerosis	33	Homo sapiens	white	female	smoker		whole blood	multi	<a href="#">View</a> <a href="#">Download</a>
GSM1479435 1	52 (year) whole blood	multiple sclerosis	33	Homo sapiens	white	female	smoker		whole blood	multi	<a href="#">View</a> <a href="#">Download</a>
GSM1479436 1	24 (year) whole blood	Type 1 Diabetes	34	Homo sapiens	white	female			whole blood	Type	<a href="#">View</a> <a href="#">Download</a>
GSM1479436 1	24 (year) whole blood	Type 1 Diabetes	34	Homo sapiens	white	female			whole blood	Type	<a href="#">View</a> <a href="#">Download</a>
GSM1479437 1	27 (year) whole blood	Type 1 Diabetes	37	Homo sapiens	white	female			whole blood	Type	<a href="#">View</a> <a href="#">Download</a>
GSM1479437 1	27 (year) whole blood	Type 1 Diabetes	37	Homo sapiens	white	female			whole blood	Type	<a href="#">View</a> <a href="#">Download</a>
GSM1479438 1	32 (year) neutrophils	normal	44	Homo sapiens	hispanic	female			neutrophils	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479438 1	32 (year) neutrophils	normal	44	Homo sapiens	hispanic	female			neutrophils	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479439 1	32 (year) monocytes	normal	44	Homo sapiens	hispanic	female			monocytes	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479439 1	32 (year) monocytes	normal	44	Homo sapiens	hispanic	female			monocytes	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479440 1	32 (year) B-Cells	normal	44	Homo sapiens	hispanic	female			B-Cells	norm	<a href="#">View</a> <a href="#">Download</a>
GSM1479440 1	32 (year) B-Cells	normal	44	Homo sapiens	hispanic	female			B-Cells	norm	<a href="#">View</a> <a href="#">Download</a>

# RNA-seq: from raw data to analysis results





# What is Expression Atlas?

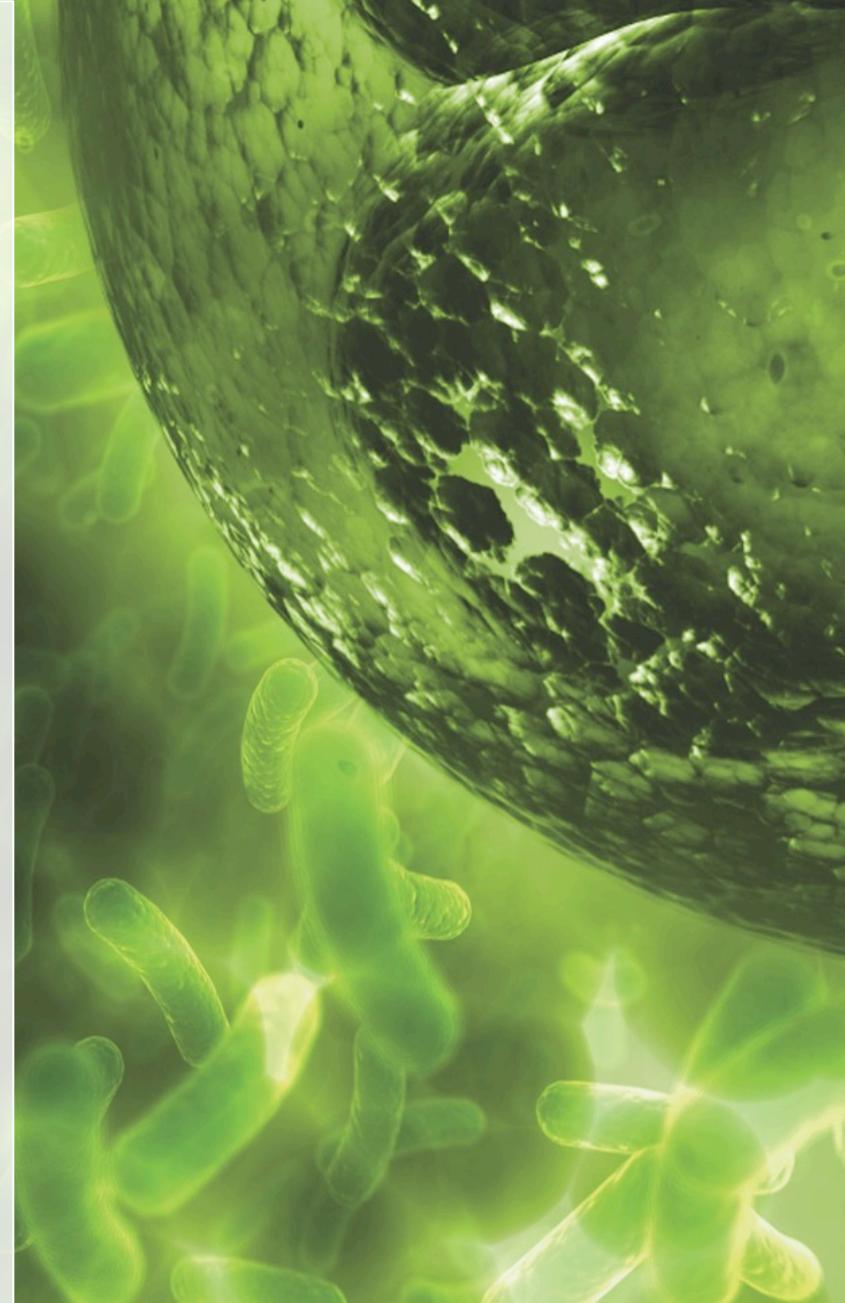


Reference resource for gene/protein expression  
across species and biological conditions

[www.ebi.ac.uk/gxa/home](http://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



ENA



Gene Expression Omnibus

## 'Big' Analysis



EBI infrastructure

## '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



# Large-scale RNA-seq experiments

Access gene expression results of large-scale datasets



Basic research



Key cell line  
models



Genentech

Cancer research



Proteomics



Zebrafish development



Prenatal human brain



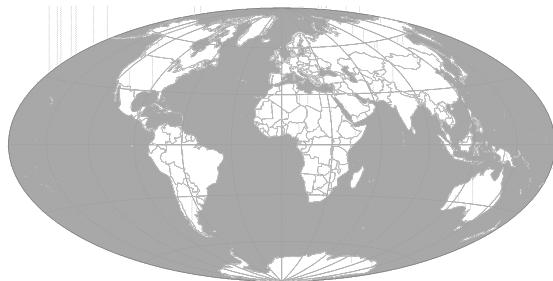
DMDD

Deciphering the Mechanisms  
of Developmental Disorders



Mouse models

# Expression Atlas contains thousands of selected and curated datasets



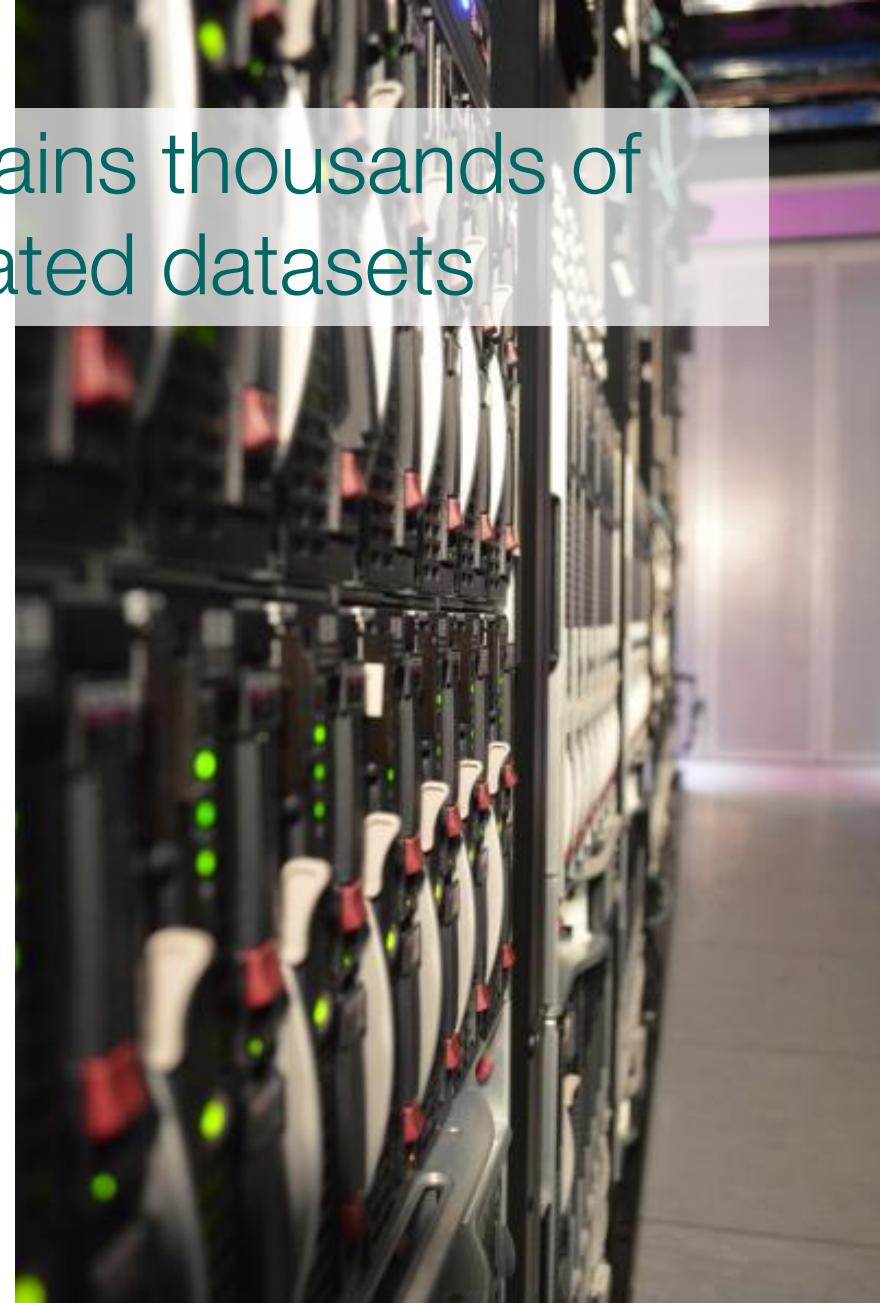
> 3,100 datasets



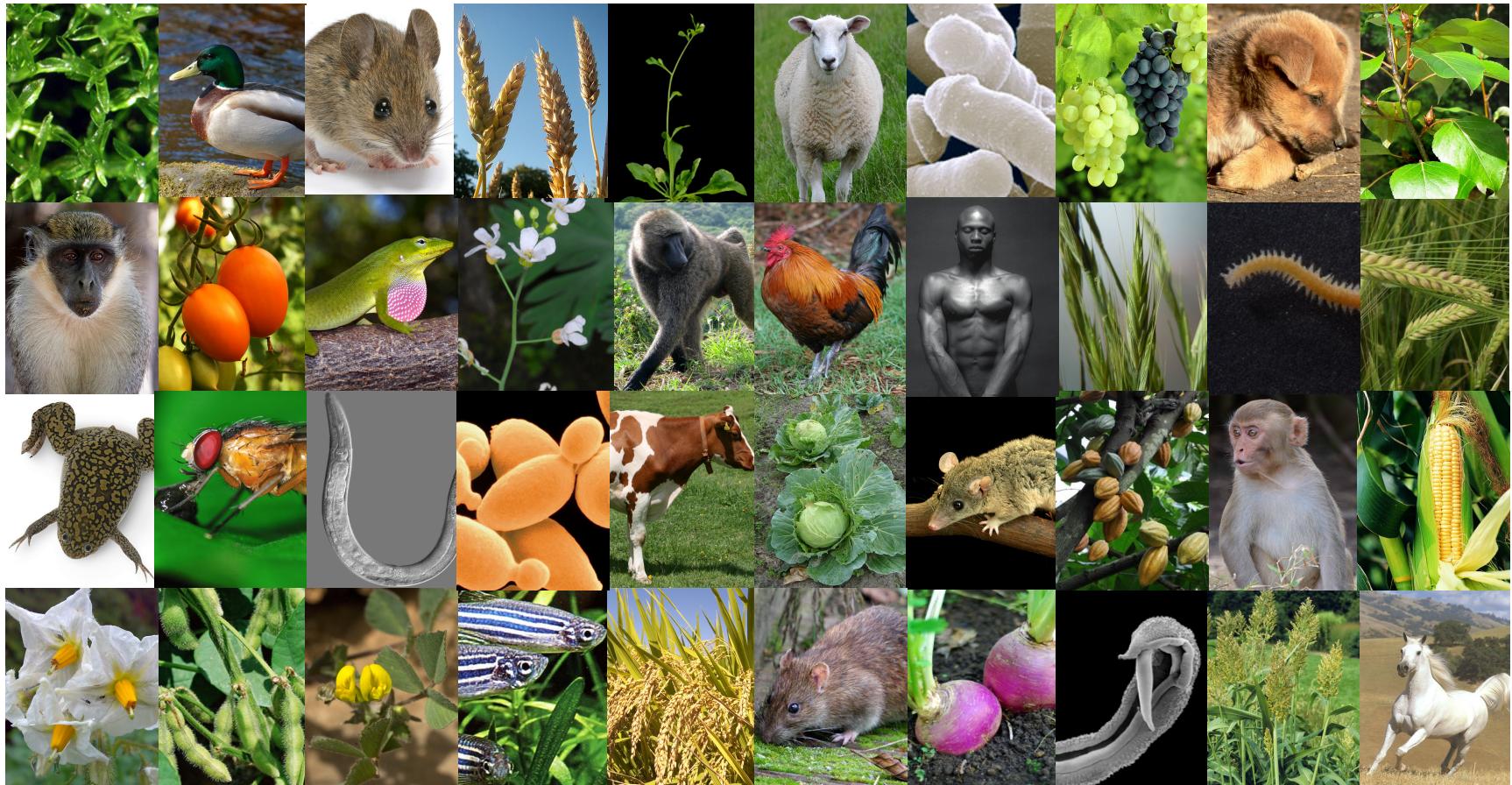
> 550 RNA-sequencing data



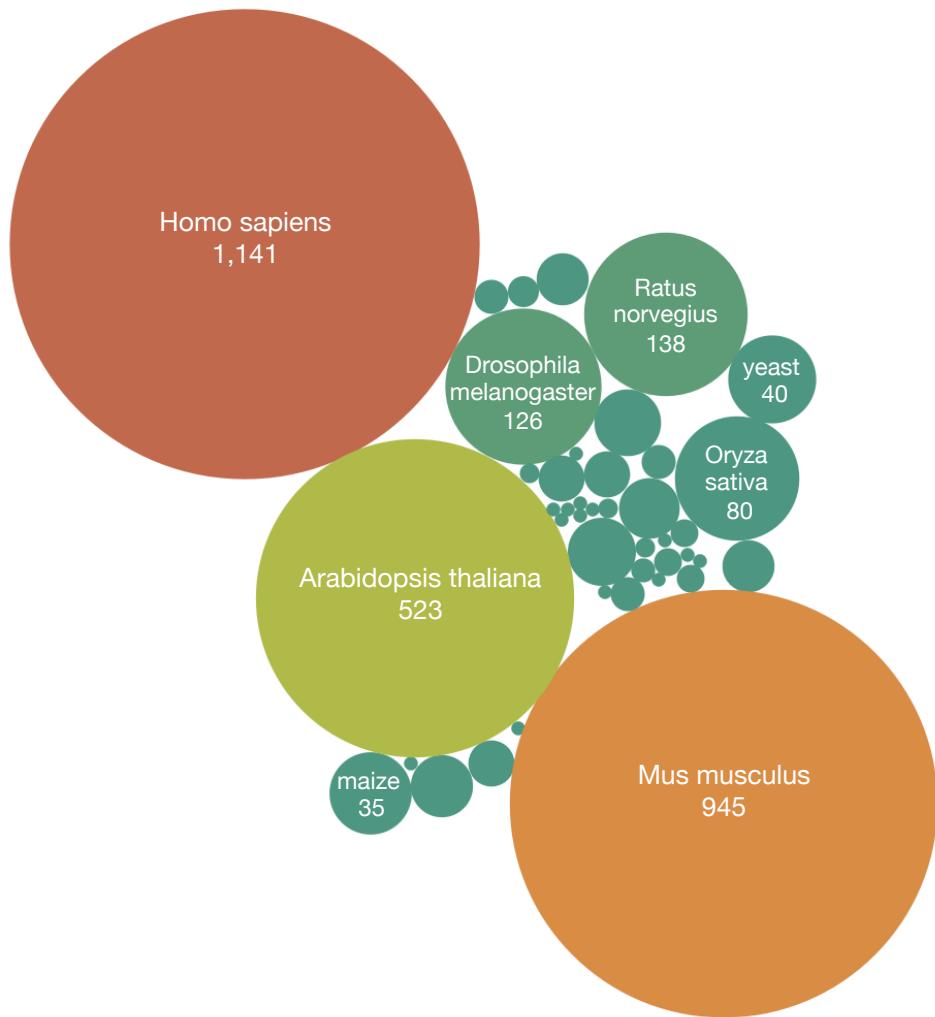
> 2,550 microarray data



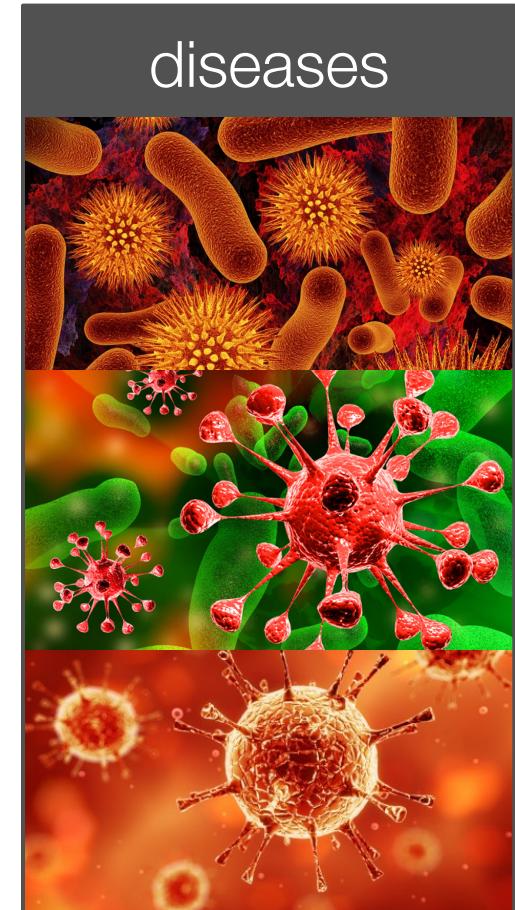
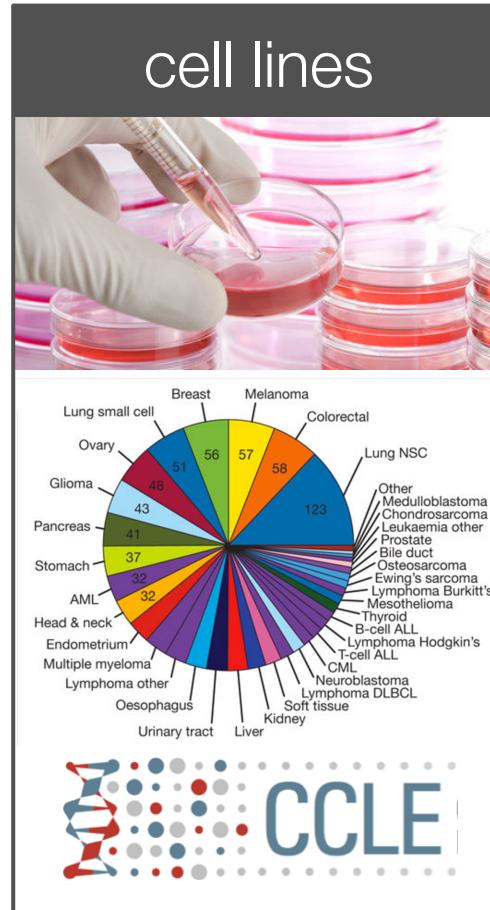
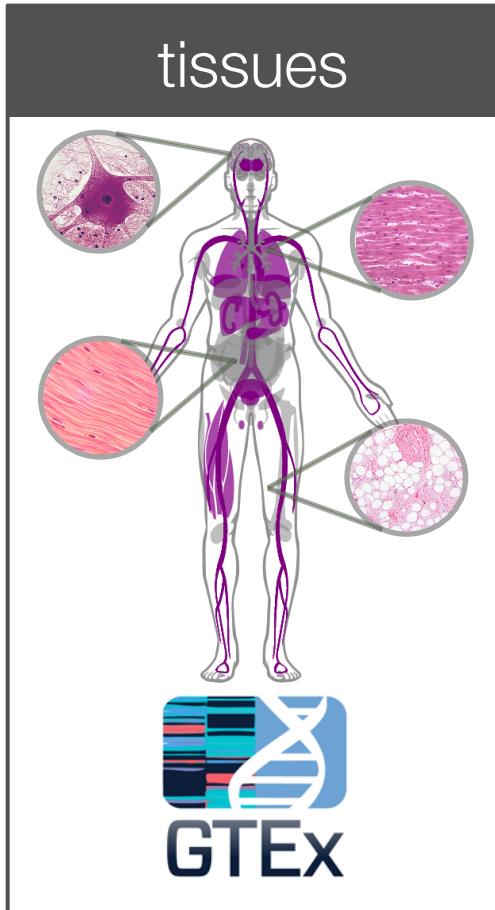
... across more than 40 species ...



... across more than 40 species ...



# ... under different biological conditions ...



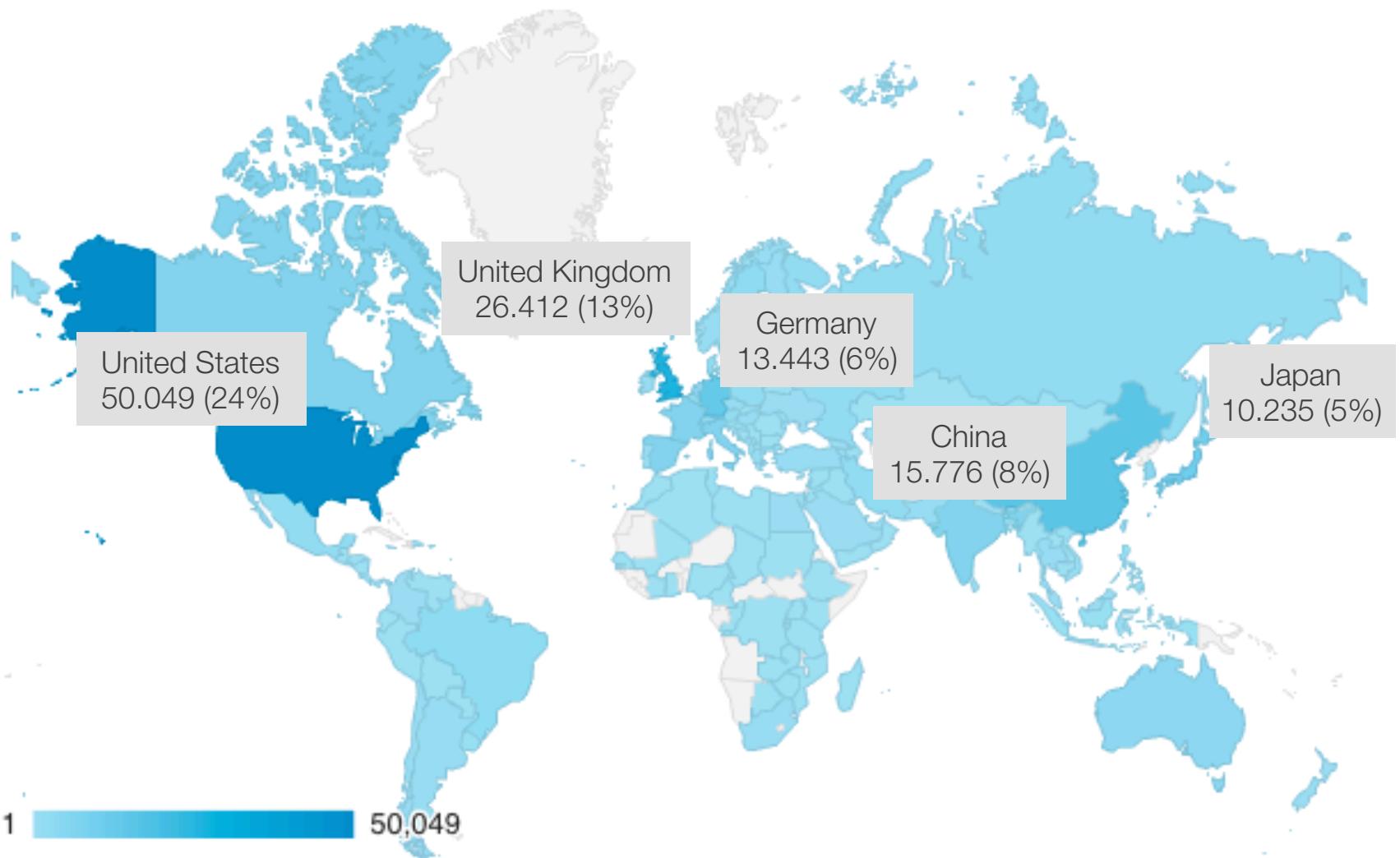
# ... to help scientists find answers ...

In which conditions  
is my favourite gene  
expressed?

What genes are  
differentially expressed in  
a particular condition?

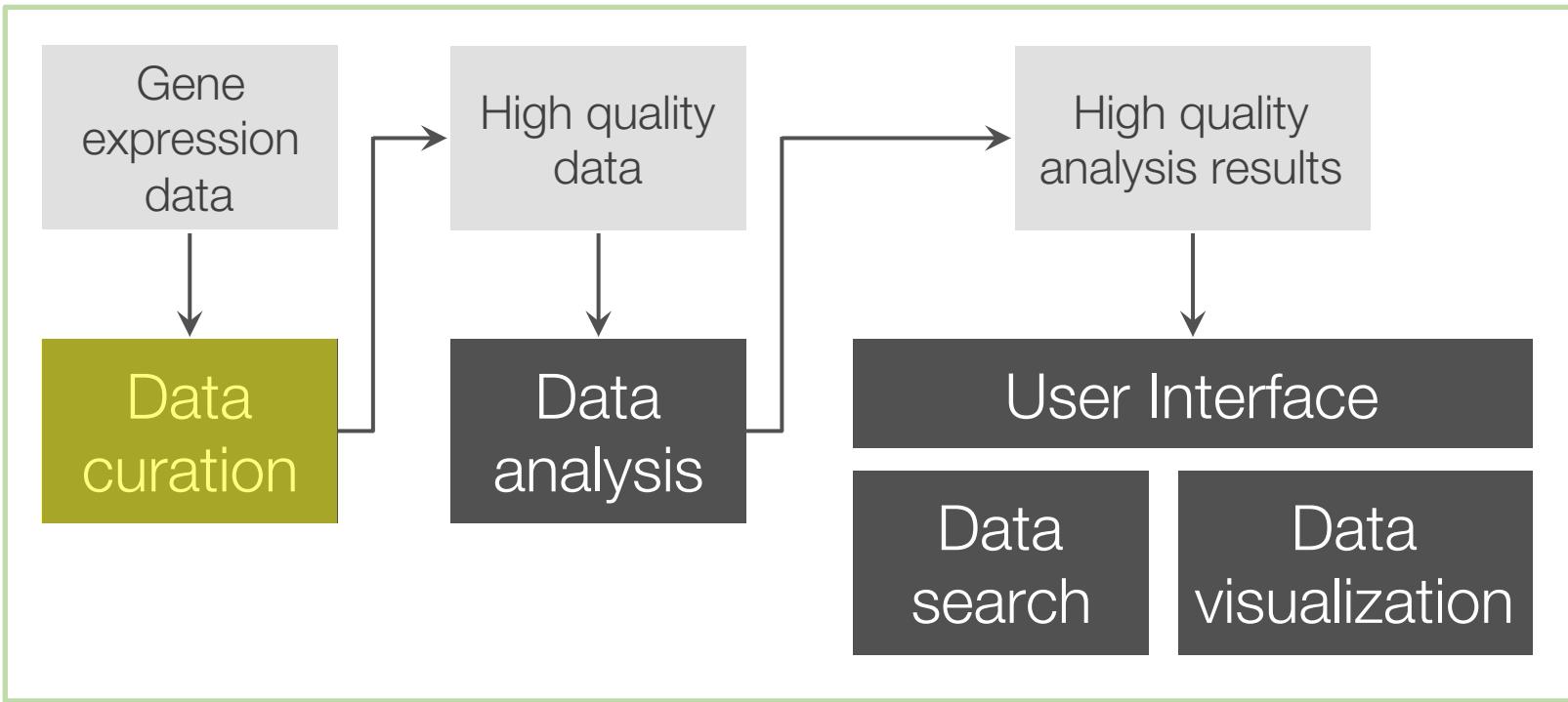
What genes are  
expressed in a  
particular condition?

... all across the world ...





# How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

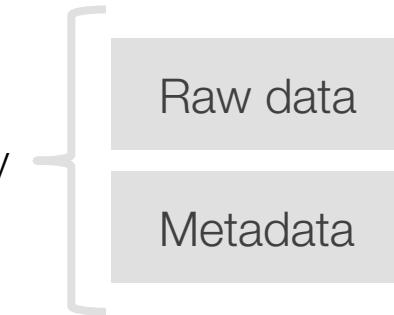


# Expression Atlas – data curation

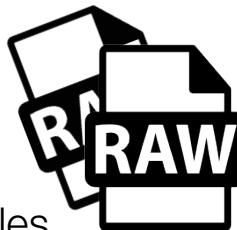
## Biocuration



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



### Raw data



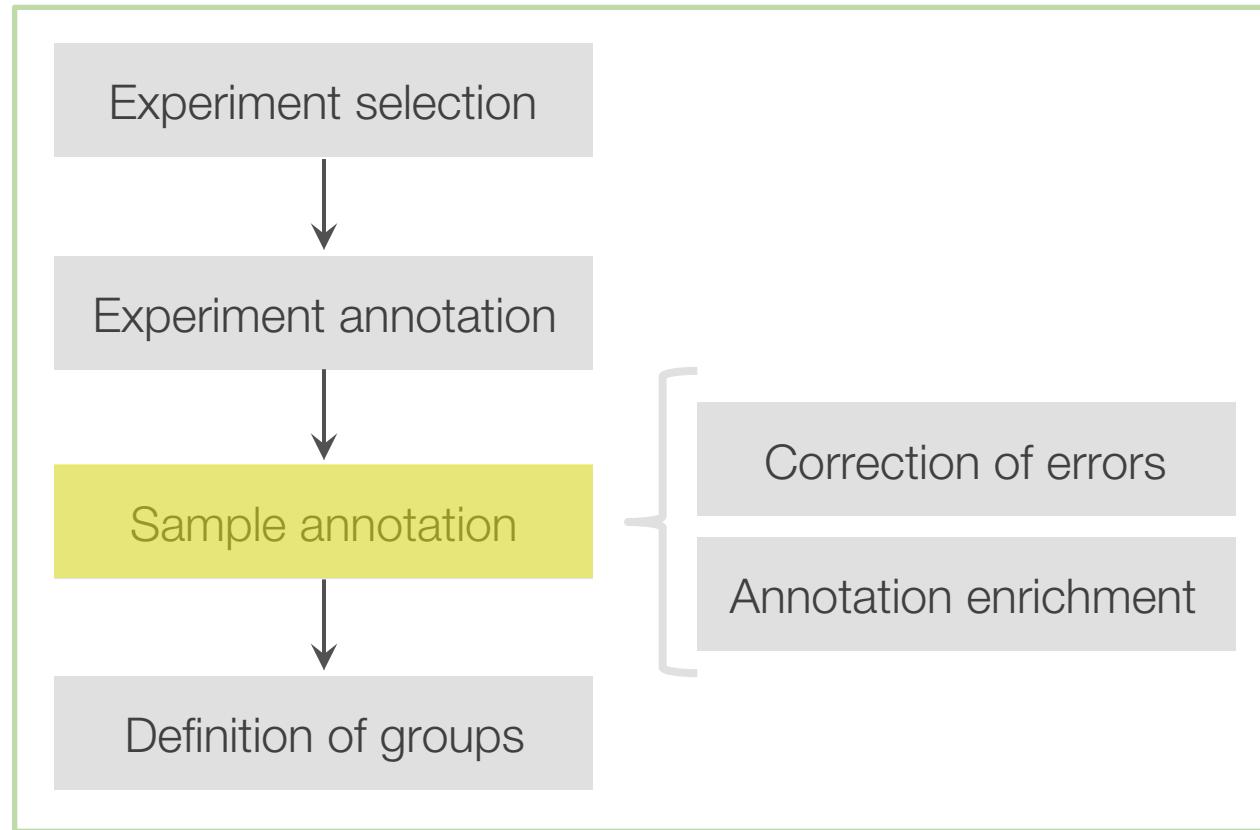
- unprocessed data files
- **Microarray:** files from the scanner (e.g. Affymetrix CEL files, Agilent feature extraction *txt* files, Illumina *idat* files)
- **Sequencing:** raw sequence read files (e.g. FASTQ files)

### Metadata

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



# Expression Atlas – data curation





# 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

7 samples x 3 biological replicates

### 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.

	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



- 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
SRR1799194	Kitaake 0 none
SRR1799197	Kitaake 0 none
SRR1799193	Kitaake 0 none
SRR1799210	EFR:XA21:GFP 0 none
SRR1799203	EFR:XA21:GFP 0 none
SRR1799205	EFR:XA21:GFP 0 none
SRR1799199	EFR:XA21:GFP 0.5 elf18
SRR1799204	EFR:XA21:GFP 0.5 elf18
SRR1799212	EFR:XA21:GFP 0.5 elf18
SRR1799208	EFR:XA21:GFP 1 elf18
SRR1799202	EFR:XA21:GFP 1 elf18
SRR1799198	EFR:XA21:GFP 1 elf18
SRR1799213	EFR:XA21:GFP 3 elf18
SRR1799201	EFR:XA21:GFP 3 elf18
SRR1799196	EFR:XA21:GFP 3 elf18
SRR1799209	EFR:XA21:GFP 6 elf18
SRR1799206	EFR:XA21:GFP 6 elf18
SRR1799195	EFR:XA21:GFP 6 elf18
SRR1799207	EFR:XA21:GFP 12 elf18
SRR1799200	EFR:XA21:GFP 12 elf18
SRR1799211	EFR:XA21:GFP 12 elf18



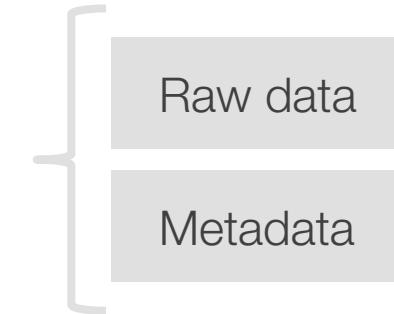


# Expression Atlas – data curation

PeerJ

✓ PEER-REVIEWED

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



Raw data

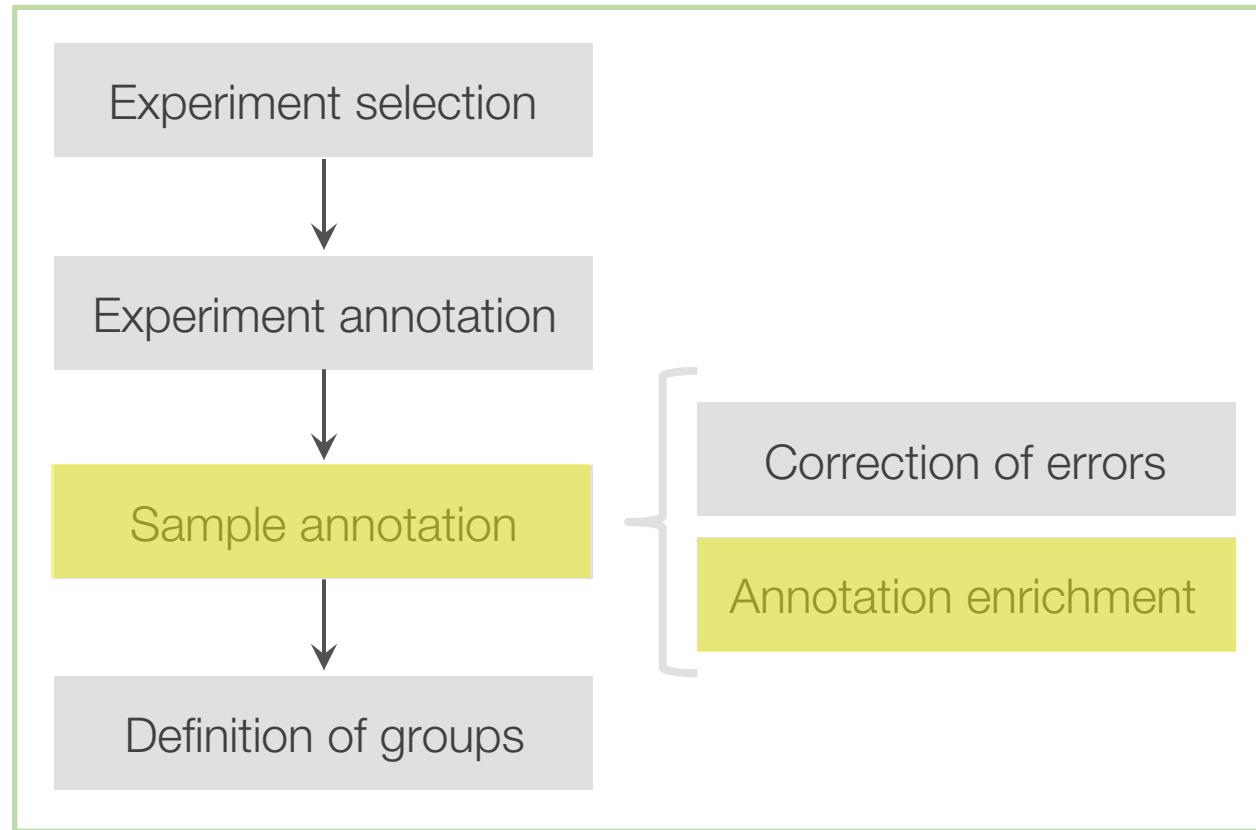
Metadata

Run
SRR1799194
SRR1799197
SRR1799193
SRR1799210
SRR1799203
SRR1799205
SRR1799199
SRR1799204
SRR1799212
SRR1799208
SRR1799202
SRR1799198
SRR1799213
SRR1799201
SRR1799196
SRR1799209
SRR1799206
SRR1799195
SRR1799207
SRR1799200
SRR1799211

FactorValue[genotype]	FactorValue[compound]	FactorValue[dose]	Unit[concentration unit]	FactorValue[time]	Unit[time unit]
wild type	none		0 nanomolar		0 hour
wild type	none		0 nanomolar		0 hour
wild type	none		0 nanomolar		0 hour
EFR:XA21:GFP	none		0 nanomolar		0 hour
EFR:XA21:GFP	none		0 nanomolar		0 hour
EFR:XA21:GFP	none		0 nanomolar		0 hour
EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
EFR:XA21:GFP	elf18		500 nanomolar		0.5 hour
EFR:XA21:GFP	elf18		500 nanomolar		1 hour
EFR:XA21:GFP	elf18		500 nanomolar		1 hour
EFR:XA21:GFP	elf18		500 nanomolar		1 hour
EFR:XA21:GFP	elf18		500 nanomolar		3 hour
EFR:XA21:GFP	elf18		500 nanomolar		3 hour
EFR:XA21:GFP	elf18		500 nanomolar		3 hour
EFR:XA21:GFP	elf18		500 nanomolar		6 hour
EFR:XA21:GFP	elf18		500 nanomolar		6 hour
EFR:XA21:GFP	elf18		500 nanomolar		6 hour
EFR:XA21:GFP	elf18		500 nanomolar		12 hour
EFR:XA21:GFP	elf18		500 nanomolar		12 hour
EFR:XA21:GFP	elf18		500 nanomolar		12 hour



# Expression Atlas – data curation





# 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

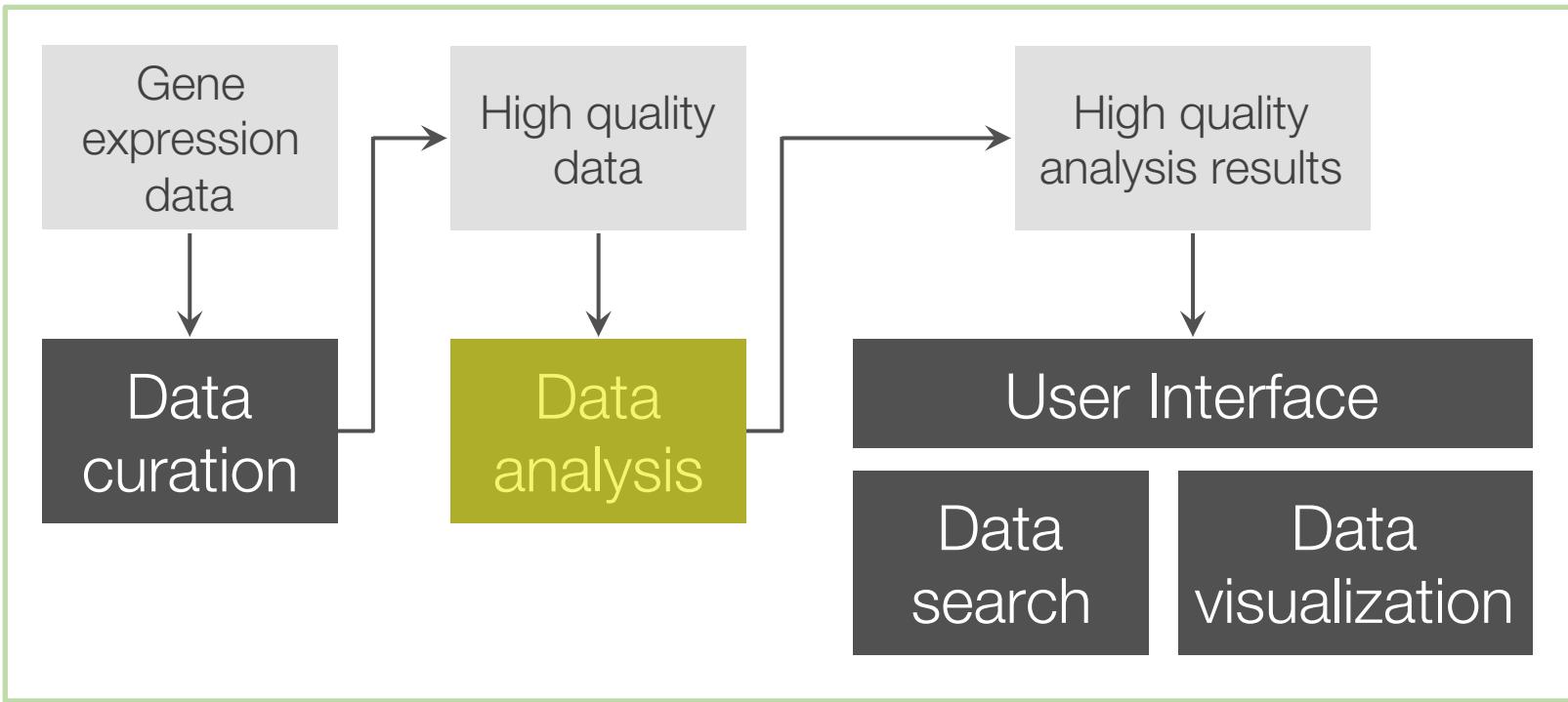
Characteristics[organism part]	Characteristics[cell line]	Characteristics[cell type]	Characteristics[disease]	Characteristics[growth condition]
liver		cholangiocyte	normal	1 day culture of primary hepatocyte
liver		hepatocyte	normal	3 day culture of primary hepatocyte
liver		hepatocyte	normal	7 day culture 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

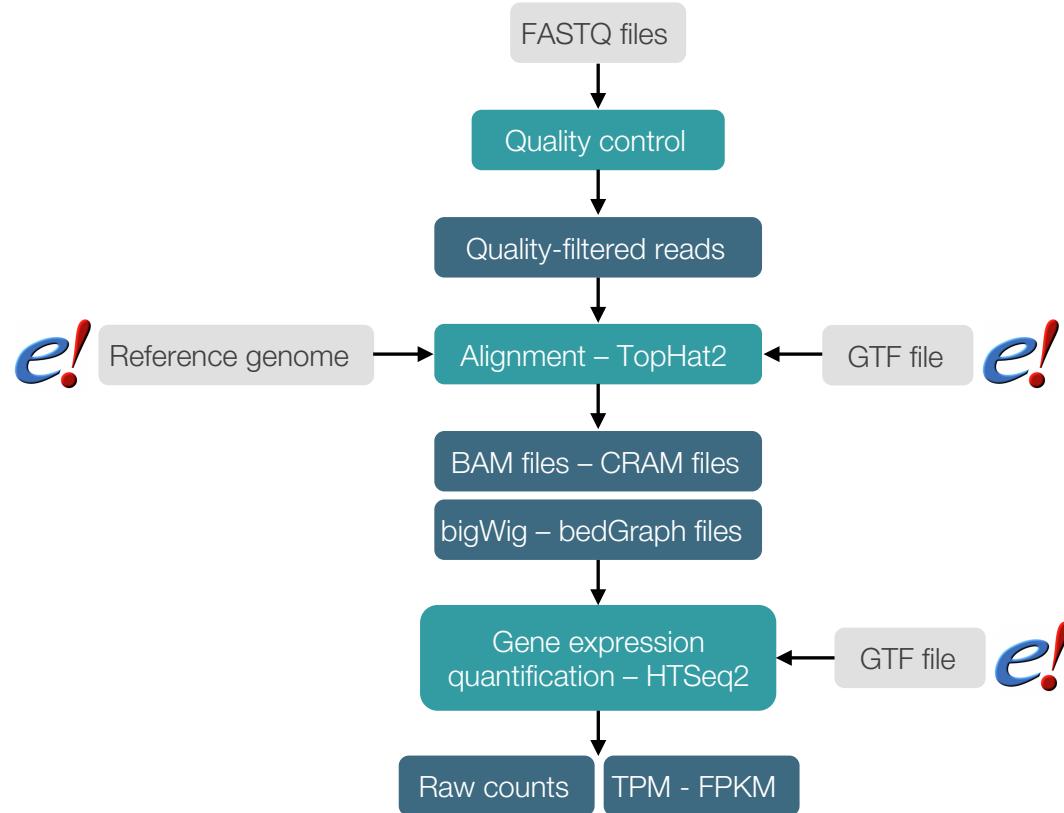


# How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

# RNA-seq data analysis



<https://github.com/nunofonseca/irap/>

## Pages

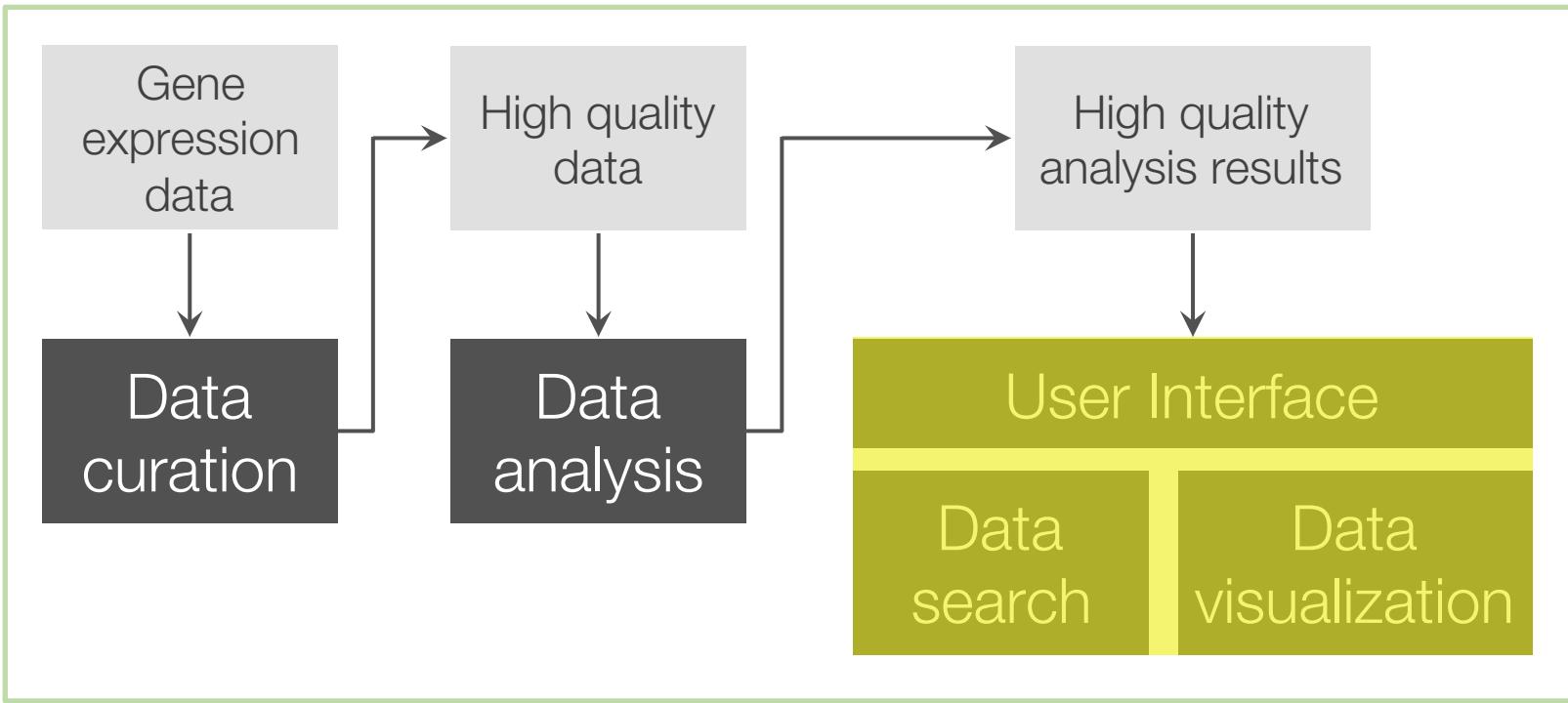
[New Page](#)

<a href="#">1 News</a>	Last updated on Jan 20
<a href="#">2 How to install and update iRAP</a>	Last updated on May 11
<a href="#">3 How to run iRAP</a>	Last updated on Jan 31
<a href="#">4 Output files and directory structure</a>	Last updated on Mar 23
<a href="#">5 Tools supported</a>	Last updated on Dec 14, 2016
<a href="#">6 Virtual Machines</a>	Last updated on Feb 16, 2016
<a href="#">7 Quick Example</a>	Last updated on Dec 20, 2016
<a href="#">8 Configuration file</a>	Last updated on Aug 21
<a href="#">9 Gene set analysis in iRAP</a>	Last updated on Mar 19, 2015
<a href="#">FAQ</a>	Last updated on Jun 3, 2016
<a href="#">GSE</a>	Last updated on Feb 16, 2016
<a href="#">iRAP single library</a>	Last updated on Dec 20, 2016
<a href="#">LSF</a>	Last updated on Feb 16, 2016

<https://github.com/nunofonseca/irap/wiki>



# How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy



# Let's try Expression Atlas

## Expression Atlas

Gene expression across species and biological conditions

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

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 does its expression change in a disease?'. [Read more about Expression Atlas](#)

Search Gene set enrichment

Gene / Gene properties

Enter gene query... Examples: REG1B, zinc finger, O14777 (UniProt), GO:0010468 (regulation of gene expression)

Organism

Any

Biological conditions

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

Search Clear

[www.ebi.ac.uk/gxa/home](http://www.ebi.ac.uk/gxa/home)



# Look up expression of a particular gene

## Expression Atlas

Gene expression across species and biological conditions

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

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 does its expression change in a disease?'. [Read more about Expression Atlas](#)

In which conditions is my favourite gene expressed?

Gene / Gene properties

UMOD x

Ex: umod  
Umodl  
Umod1  
umod.1  
UMODL1  
**UMOD**  
UMODL1-AS1

Organism

Homo sapiens

Mus musculus  
Arabidopsis thaliana  
Rattus norvegicus  
Drosophila melanogaster  
Oryza sativa Japonica Group

Any

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

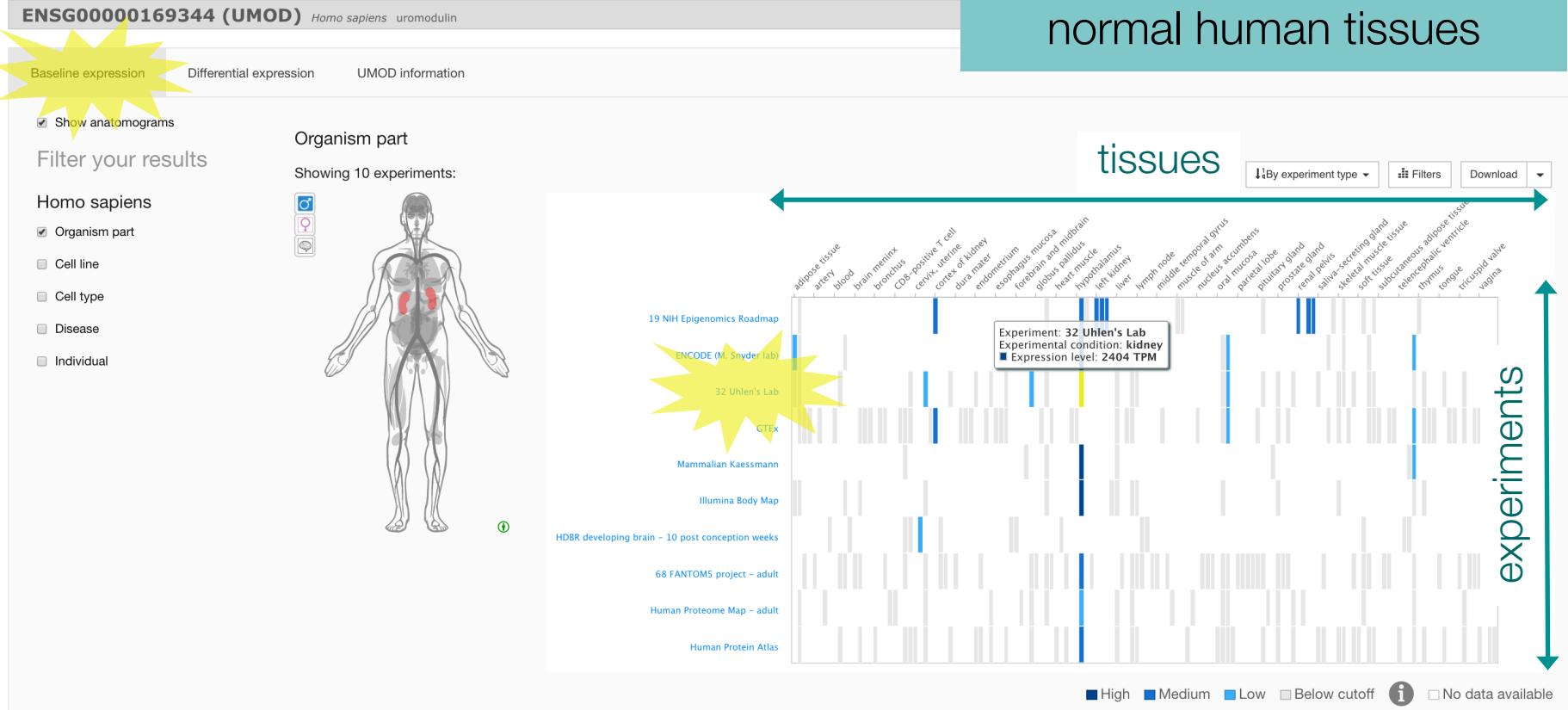
Biological conditions

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

Search | Clear



# Look up expression of a gene: baseline





# Look up expression of a gene: baseline

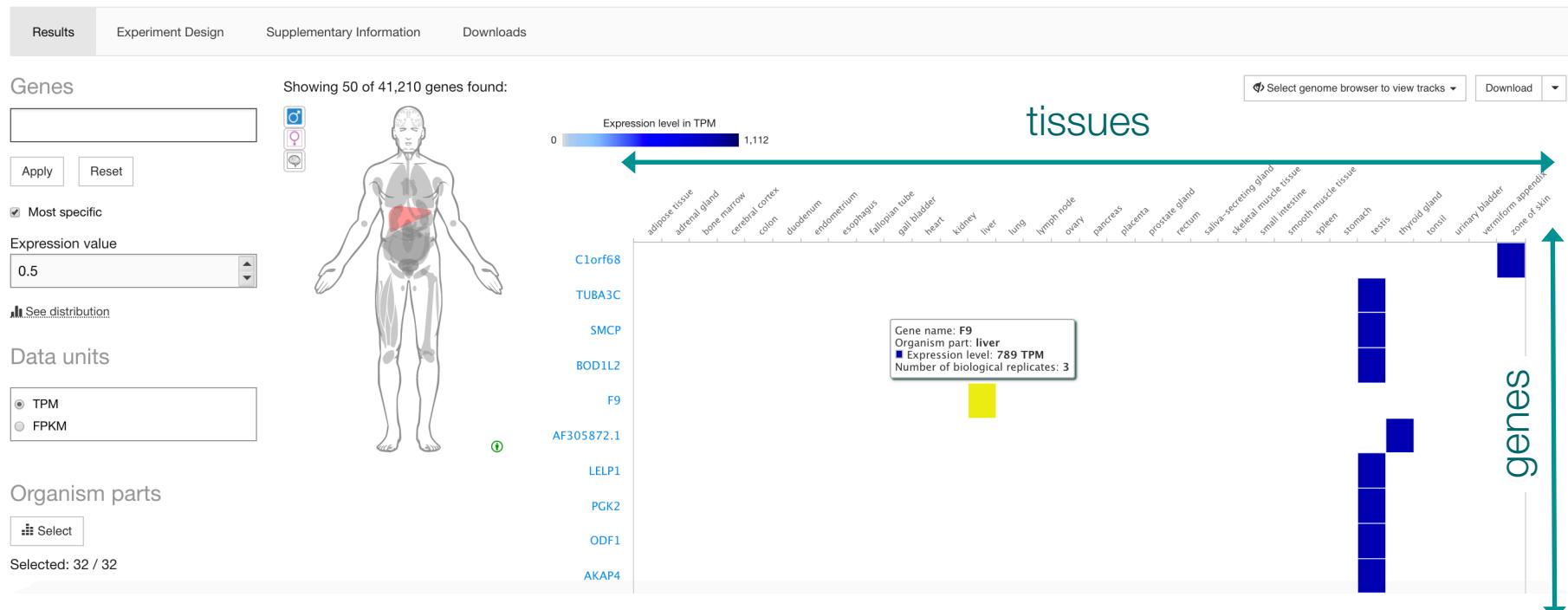
RNA-seq of coding RNA from tissue samples of 122 human individuals represented by 41,210 genes.

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

Gene expression results in a particular experiment





# Look up expression of a particular gene

## Expression Atlas

Gene expression across species and biological conditions

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

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 does its expression change in a disease?'. [Read more about Expression Atlas](#)

In which conditions is my favourite gene **differentially expressed?**

Gene / Gene properties

UMOD x

Ex: umod  
Umodl  
Umodl1  
umod.1  
UMODL1  
**UMOD**  
UMODL1-AS1

Organism

Homo sapiens

Mus musculus  
Arabidopsis thaliana  
Rattus norvegicus  
Drosophila melanogaster  
Oryza sativa Japonica Group

Any

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

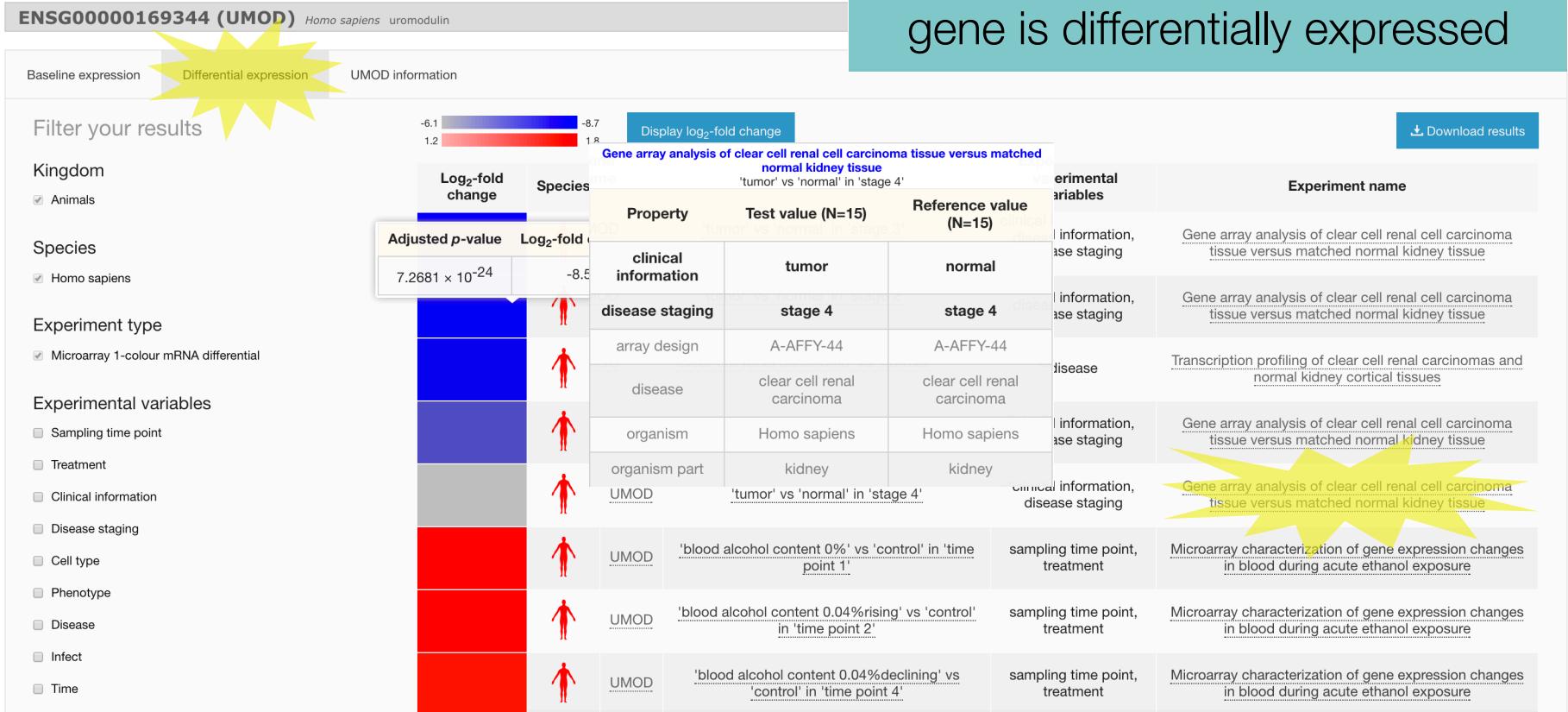
Biological conditions

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

Search | Clear



# Look up expression of a gene: differential





# Look up expression of a gene: differential

Gene array analysis of clear cell renal cell carcinoma tissue versus matched normal tissue

Microarray 1-colour mRNA

Organism: *Homo sapiens*

Array Design(s): Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133\_Plus\_2]

Reference(s): 24962026 (Filter by genes in paper)

Differentially expressed genes  
in a particular experiment





# Let's try Expression Atlas

Baseline expression

## Hands-on activity

Look up expression of TNNI3 gene in  
normal human tissues



In pairs



# Genes DE in a particular condition

**Expression Atlas**  
Gene expression across species and biological conditions

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

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 does its expression change in a disease?'.

Read more about Expression Atlas

Search Gene set enrichment

Gene / Gene properties

Enter gene query...  
Examples: REG1B, zinc finger, Q14777 (UniProt), GO:0010468 (regulation of gene expression)

Search Clear

Organism

- Homo sapiens
- Mus musculus
- Arabidopsis thaliana
- Rattus norvegicus
- Drosophila melanogaster
- Oryza sativa Japonica Group

✓ Any

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

Biological conditions

lung carcinoma x

Ex lung carcinoid tumor

lung carcinoma

- Carcinoma, Lewis Lung
- Combined Lung Carcinoma
- Lung Inflammatory Myofibroblastic Tumor
- Lung Lymphangioleiomyomatosis
- Lung Sclerosing Hemangioma
- Pancoast tumor
- adenosquamous lung carcinoma

What genes are differentially expressed in a particular condition



# Genes DE in a particular condition

Results for lung carcinoma AND Homo sapiens

Baseline expression      Differential expression

Filter your results

Kingdom  
 Animals

Species  
 Homo sapiens

Experiment type  
 Microarray 1-colour mRNA differential

Experimental variables

- Disease
- Cell line
- Passage
- Phenotype
- Growth condition
- Compound
- Organism part
- Protocol
- Time
- Genotype

Log<sub>2</sub>-fold change

Sp

-5.6      5.6

Transcription profiling by array of formalin-fixed paraffin-embedded or fresh frozen human tumor samples  
'lung cancer' vs 'normal' in 'lung; Fresh-frozen tissue'

Property	Test value (N=3)	Reference value (N=3)
disease	lung cancer	normal
organism part	lung	lung
protocol	Fresh-frozen tissue	Fresh-frozen tissue
array design	A-AFFY-37	A-AFFY-37
organism	Homo sapiens	Homo sapiens
	frozen tissue	part, protocol
	phenotype	
	Cancer-Associated Fibroblasts Support Lung Cancer Stemness through Paracrine IGF-II/IGF1R/Nanog Signaling	
	Transcription profiling by array of formalin-fixed paraffin-embedded or fresh frozen human tumor samples	
	Expression data from CL1-0 and CL1-5 lung cancer cell line	
	Transcription profiling by array of three non-small cell lung cancer cell lines upon TGF $\beta$ -induced epithelial-mesenchymal transition (EMT) against untreated counterparts to study EMT's role in tumour metastasis	
	Transcription profiling by array of formalin-fixed paraffin-embedded or fresh frozen human tumor samples	
	Expression data from CL1-0 and CL1-5 lung cancer cell line	
	Cancer-Associated Fibroblasts Support Lung Cancer Stemness through Paracrine IGF-II/IGF1R/Nanog Signaling	
	Transcription profiling by array of formalin-fixed paraffin-embedded or fresh frozen human tumor samples	
	Microarray analysis of xenograft models in use at the Developmental Therapeutics Program of the National Cancer Institute (DTP-NCI)	

Download results

Differentially expressed genes in human lung carcinoma



# Expression Atlas – data curation

## Ontology

[www.ebi.ac.uk/efo](http://www.ebi.ac.uk/efo)



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

Controlled vocabulary

Hierarchy (relationship)

OLS > Experimental Factor Ontology [EFO](#) > [EFO:0001071](#)  

## lung carcinoma

 [http://www.ebi.ac.uk/efo/EFO\\_0001071](http://www.ebi.ac.uk/efo/EFO_0001071) 



A carcinoma originating in the lung. Lung carcinomas usually arise from the epithelium that lines the bronchial tree (bronchogenic carcinomas), and are classified as small cell or non-small cell carcinomas. Non-small cell lung carcinomas are usually adenocarcinomas, squamous cell carcinomas, or large cell carcinomas. Metastatic carcinomas to the lung are also common, and can be difficult to distinguish from primary tumors.

A carcinoma that is located\_in the lungs and has\_symptom cough and has\_symptom chest discomfort or pain and has\_symptom weight loss and has\_symptom hemoptysis.

Tumors or cancer of the LUNG

**Synonyms:** Cancer, Pulmonary, Lung Cancer, Pulmonary Neoplasms, Lung Neoplasm, NEOPL PULM, cancer of lung, Neoplasm, Pulmonary, Pulmonary Neoplasm, Lung Neoplasms, Cancers, Pulmonary, **carcinoma OF LUNG**, Cancer, Lung, Neoplasm, Lung, Pulmonary Cancers, LUNG NEOPL, Cancer of the Lung, NEOPL LUNG, PULM NEOPL, Lung Cancers, Pulmonary Cancer, Neoplasms, Pulmonary, Neoplasms, Lung, Carcinoma of the Lung, Cancers, Lung



# Expression Atlas – data curation

## Ontology

[www.ebi.ac.uk/efo](http://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



# Let's try Expression Atlas

Differential expression

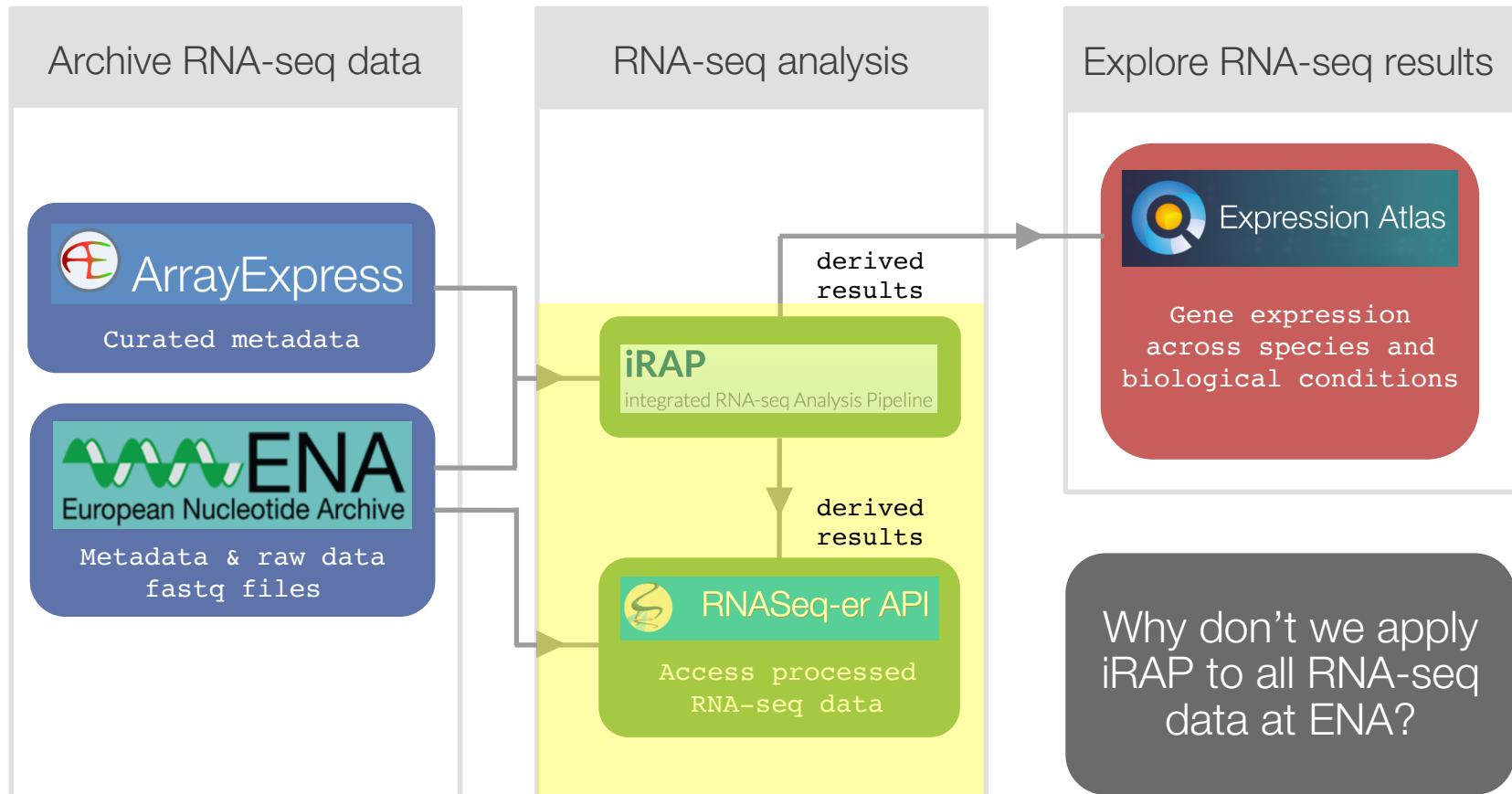
## Hands-on activity

Find differentially expressed genes in patients with hepatocellular carcinoma



In pairs

# RNA-seq: from raw data to analysis results





# RNASeq-er API



## RNASeq-er API

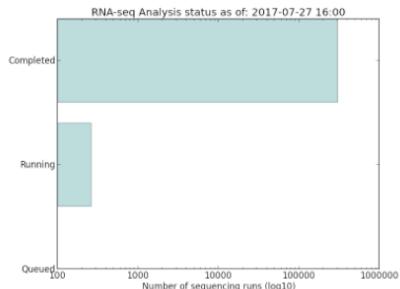
[Home](#) | [Documentation](#)

[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](#).

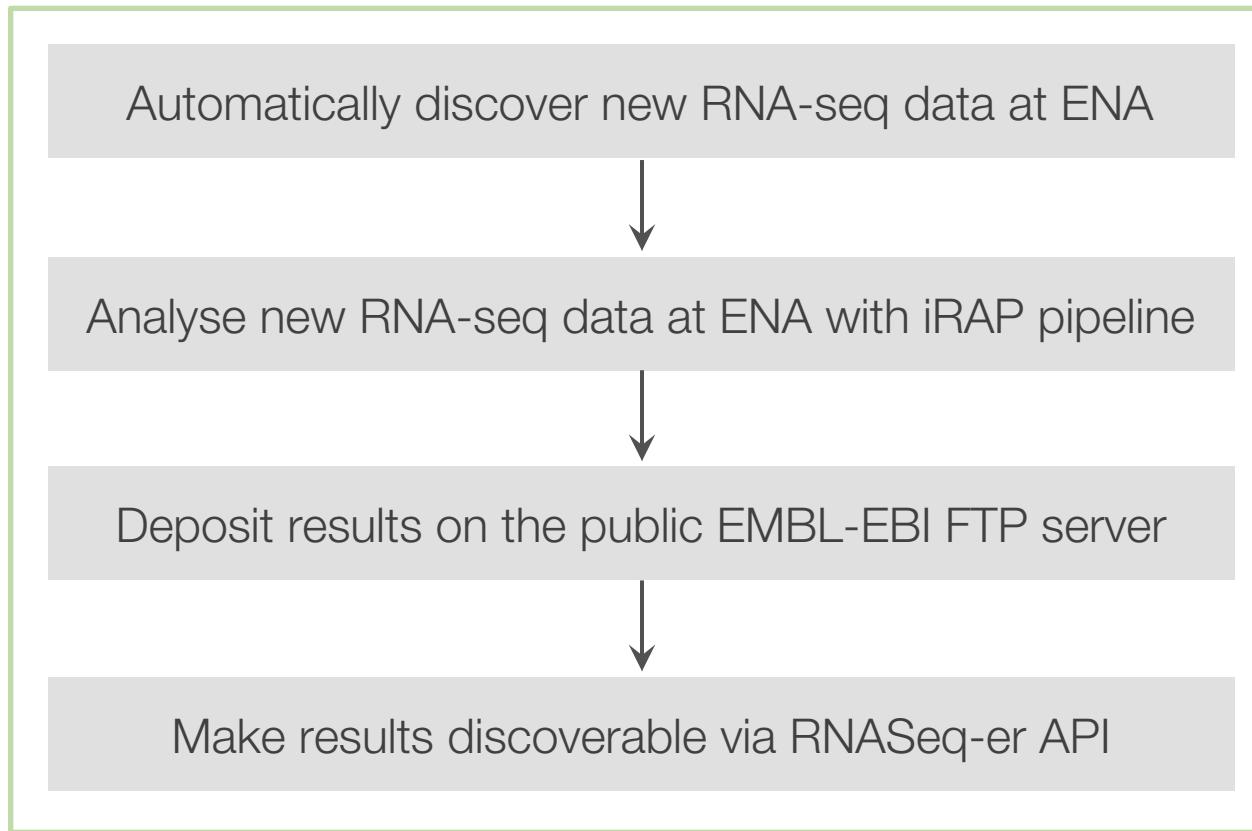
- [1. Get Started](#)
- [2. What does the RNASeq-er pipeline do?](#)
- [3. How is the RNASeq-er performed?](#)
- [4. How to use the RNASeq-er API?](#)
- [5. What are the main classes of API calls?](#)
  - [5.1. Analysis results Per Run](#)
    - [5.1.1. Making per-run API calls](#)
    - [5.1.2. Results of per-run API calls](#)
  - [5.2. Analysis results Per Study](#)
    - [5.2.1. Making per-study API calls](#)
    - [5.2.2. Results of per-study API calls](#)
  - [5.3. Sample Attributes Per Run](#)
    - [5.3.1. Making sample attributes per-run API calls](#)
    - [5.3.2. Results of sample attributes per-run API calls](#)
  - [5.4. Baseline expression By Gene](#)
    - [5.4.1. Making baseline expression per-gene API calls](#)
    - [5.4.2. Results of baseline expression per-gene API calls](#)
  - [5.5. Mapping Quality Statistics For All Organisms](#)
    - [5.5.1. Retrieving the mean and standard deviation of mapping quality for all organisms](#)



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



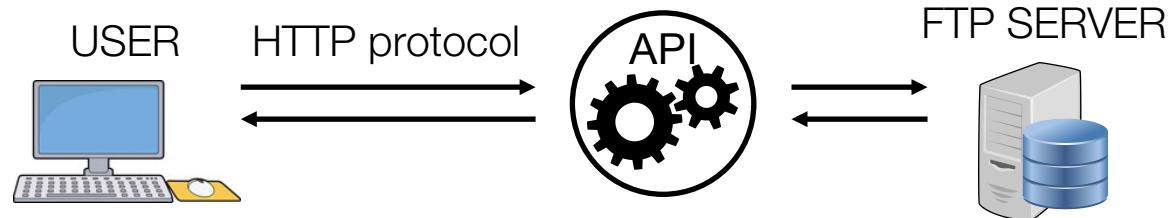
# RNASeq-er API



<https://doi.org/10.1093/bioinformatics/btx143>

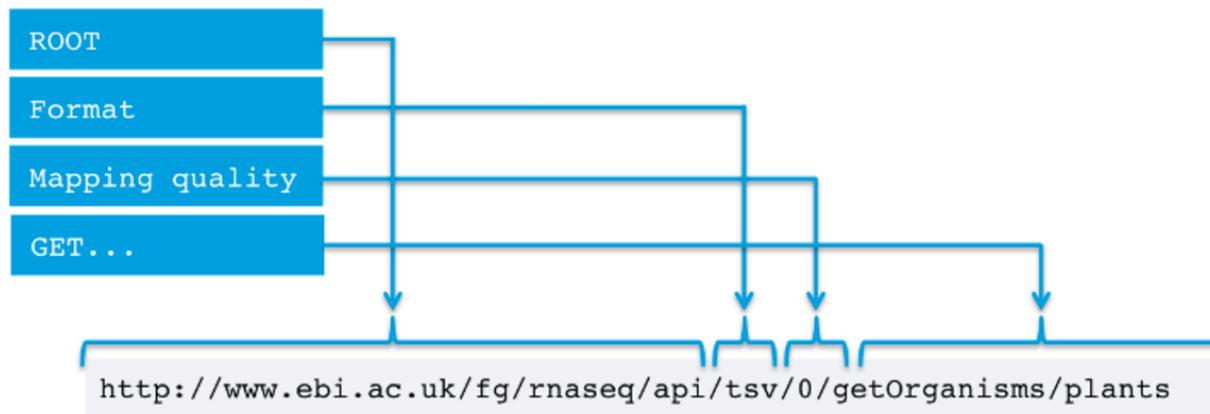


# RNASeq-er API



How to built API calls?

How to built the URL?





# RNASeq-er API

## Accessing the results of alignments

*Access alignments of RNA-seq runs involving human autism samples*

[https://www.ebi.ac.uk/fg/rnaseq/api/tsv/0/getRunsByOrganismCondition/homo\\_sapiens/autism](https://www.ebi.ac.uk/fg/rnaseq/api/tsv/0/getRunsByOrganismCondition/homo_sapiens/autism)

Secure | https://www.ebi.ac.uk/fg/rnaseq/api/tsv/0/getRunsByOrganismCondition/homo\_sapiens/autism

STUDY_ID	SAMPLE_IDS	BIOREP_ID	RUN_IDS	ORGANISM	REFERENCE_ORGANISM	STATUS	ASSEMBLY_USED	ENA_LAST_UPDATED	LAST_PROCESSED_DATE	CRAM_LOCATION	BEDGRAPH_LOCATION
SRP007483	SAMN00668205	SRR309133	SRR309133	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sat Aug 22 2015 10:42:42		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.bw	96			disease autism
SRP007483	SAMN00668205	SRR309134	SRR309134	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Aug 10 2015 23:07:36		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.bw	96			disease autism
SRP007483	SAMN00668206	SRR309135	SRR309135	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Aug 24 2015 23:30:14		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.bw	95			disease autism
SRP007483	SAMN00668206	SRR309136	SRR309136	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Jul 27 2015 07:34:38		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.bw	95			disease autism
SRP007483	SAMN00668207	SRR309137	SRR309137	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Tue Aug 11 2015 22:40:15		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.bw	94			disease autism
SRP007483	SAMN00668207	SRR309138	SRR309138	homo_sapiens	homo_sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sat Aug 15 2015 13:44:31		
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.cram				ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.bedgraph
							ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.bw	95			disease autism

*alignment results in EMBL-EBI FTP server*



# RNASeq-er API

## Accessing the results of alignments

*Access alignment results of all samples of a RNA-seq study*

<https://www.ebi.ac.uk/fg/rnaseq/api/tsv/0/getRunsByStudy/SRP007483>

Secure | <https://www.ebi.ac.uk/fg/rnaseq/api/tsv/0/getRunsByStudy/SRP007483>

STUDY_ID	SAMPLE_IDS	BIOREP_ID	RUN_IDS	ORGANISM	REFERENCE_ORGANISM	STATUS	ASSEMBLY_USED	ENA_LAST_UPDATED	LAST_PROCESSED_DATE	CRAM_LOCATION	BEDGRAPH_LOCATION
<b>BIGWIG LOCATION MAPPING_QUALITY</b>											
SRP007483	SAMN00668205	SRR309133	SRR309133	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sat Aug 22 2015 10:42:42		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309133/SRR309133.bw	96
SRP007483	SAMN00668205	SRR309134	SRR309134	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Aug 10 2015 23:07:36		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309134/SRR309134.bw	96
SRP007483	SAMN00668206	SRR309135	SRR309135	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Aug 24 2015 23:30:14		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309135/SRR309135.bw	95
SRP007483	SAMN00668206	SRR309136	SRR309136	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Jul 27 2015 07:34:38		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309136/SRR309136.bw	95
SRP007483	SAMN00668207	SRR309137	SRR309137	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Tue Aug 11 2015 22:40:15		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309137/SRR309137.bw	94
SRP007483	SAMN00668207	SRR309138	SRR309138	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sat Aug 15 2015 13:44:31		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309138/SRR309138.bw	95
SRP007483	SAMN00668208	SRR309139	SRR309139	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Thu Aug 13 2015 04:56:53		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309139/SRR309139.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309139/SRR309139.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309139/SRR309139.bw	94
SRP007483	SAMN00668208	SRR309140	SRR309140	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Aug 24 2015 21:30:41		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309140/SRR309140.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309140/SRR309140.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309140/SRR309140.bw	94
SRP007483	SAMN00668209	SRR309141	SRR309141	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Mon Jul 27 2015 00:06:37		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309141/SRR309141.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309141/SRR309141.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309141/SRR309141.bw	96
SRP007483	SAMN00668209	SRR309142	SRR309142	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sun Aug 16 2015 03:50:59		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309142/SRR309142.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309142/SRR309142.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309142/SRR309142.bw	96
SRP007483	SAMN00668210	SRR309143	SRR309143	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Sun Aug 23 2015 10:36:16		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309143/SRR309143.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309143/SRR309143.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309143/SRR309143.bw	96
SRP007483	SAMN00668210	SRR309144	SRR309144	homo sapiens	homo sapiens	Complete	GRCh38	Fri Jun 19 2015 20:17:06	Fri Jul 24 2015 20:40:35		
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309144/SRR309144.cram	ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309144/SRR309144.bedgraph
										ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/SRR309/SRR309144/SRR309144.bw	96



# RNASeq-er API

## Accessing sample metadata

<https://www.ebi.ac.uk/fg/rnaseq/api/tsv/getSampleAttributesPerRunByStudy/SRP007483>

STUDY_ID	RUN_ID	TYPE	VALUE_EFO_URL
SRP007483	SRR309133		clinical information brain region BA22 (temporal cortex) NA
SRP007483	SRR309133		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309133		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309133		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309133		clinical information brain region BA22 (temporal cortex) NA
SRP007483	SRR309134		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309134		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309134		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309134		clinical information brain region BA41 (temporal cortex) NA
SRP007483	SRR309134		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309134		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309134		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309137		clinical information brain region BA41 (temporal cortex) NA
SRP007483	SRR309137		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309137		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309137		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309138		clinical information brain region BA41 (temporal cortex) NA
SRP007483	SRR309138		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309138		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309138		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309139		clinical information brain region BA09 (frontal cortex) NA
SRP007483	SRR309139		disease autism http://www.ebi.ac.uk/efo/EFO_0003758
SRP007483	SRR309139		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309139		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309140		clinical information brain region BA09 (frontal cortex) NA
SRP007483	SRR309140		disease normal http://purl.obolibrary.org/obo/PATO_0000461
SRP007483	SRR309140		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309140		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309144		clinical information brain region BA41 (temporal cortex) NA
SRP007483	SRR309144		disease normal http://purl.obolibrary.org/obo/PATO_0000461
SRP007483	SRR309144		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309144		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955
SRP007483	SRR309144		clinical information brain region BA41 (temporal cortex) NA
SRP007483	SRR309144		disease normal http://purl.obolibrary.org/obo/PATO_0000461
SRP007483	SRR309144		organism Homo sapiens http://purl.obolibrary.org/obo/NCBITaxon_9606
SRP007483	SRR309144		organism part brain http://purl.obolibrary.org/obo/UBERON_0000955

samples from  
patients with autism

samples from  
normal individuals



# RNASeq-er API

## Accessing gene expression quantification results

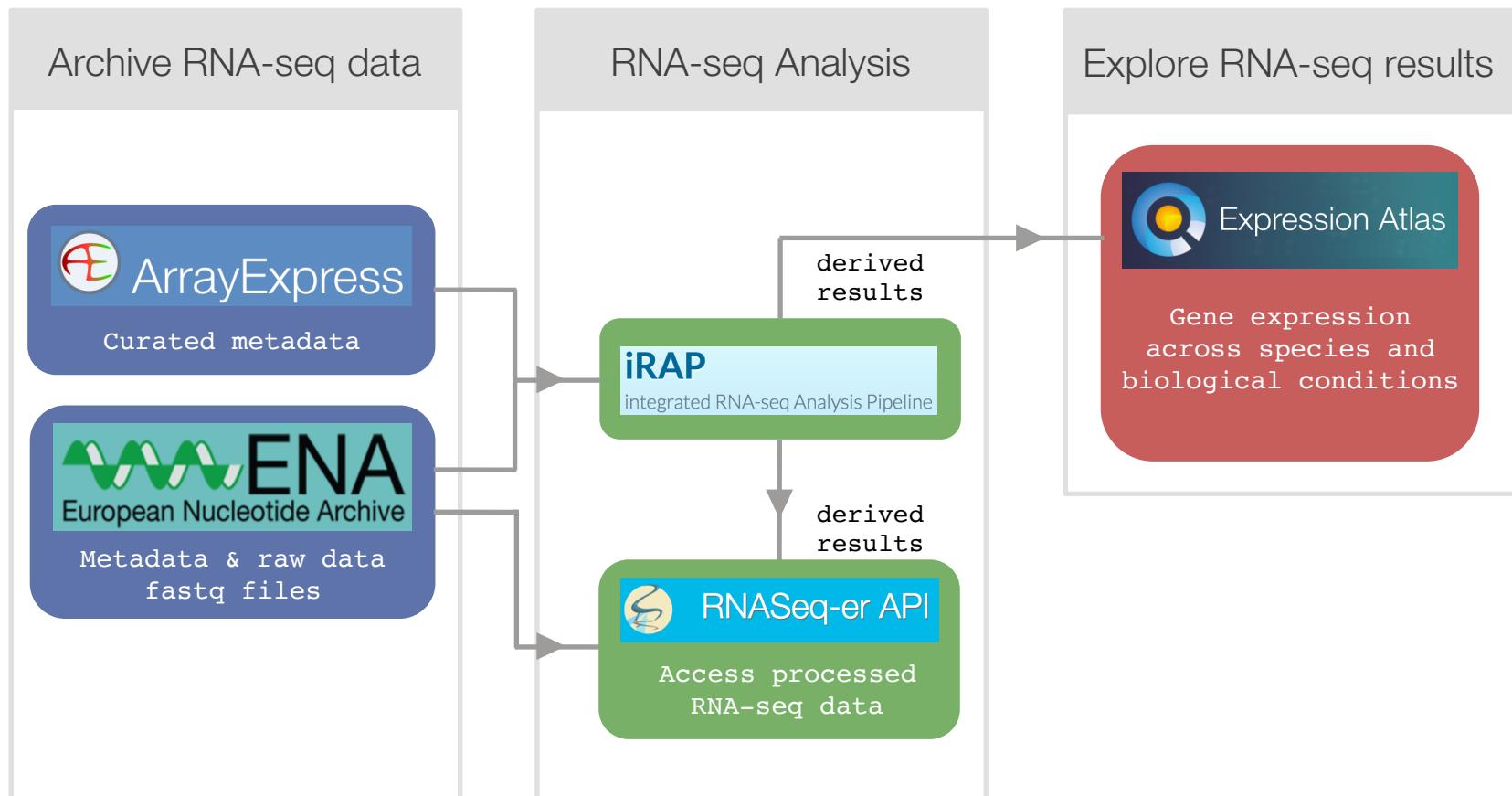
*Access gene expression quantification results of a RNA-seq study*

<https://www.ebi.ac.uk/fg/rnaseq/api/tsv/getStudy/SRP007483>

Secure | https://www.ebi.ac.uk/fg/rnaseq/api/tsv/getStudy/SRP007483

STUDY_ID	ORGANISM	REFERENCE_ORGANISM	ASSEMBLY_USED	GTF_USED	STATUS	GENES_FPKM_COUNTS_FTP_LOCATION	GENES TPM COUNTS FTP LOCATION	GENES_RAW_COUNTS_FTP_LOCATION	EXONS_FPKM_COUNTS_FTP_LOCATION																																																																																																																																								
SRP007483	homo_sapiens	homo_sapiens	GRCh38	Homo sapiens.GRCh38.79.gtf.gz	Complete																																																																																																																																												
<pre>ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/studies/ena/SRP007483/homo_sapiens/genes.fpkm.htseq2.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/genes.tpm.htseq2.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/genes.raw.htseq2.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/exons.fpkm.dexseq.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/exons.tpm.dexseq.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/exons.raw.dexseq.tsv ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/studies/ena/SRP007483/homo_sapiens/irap VERSIONS.tsv</pre> Thu Jun 30 2016 21:43:57																																																																																																																																																	
<i>gene expression quantification matrices</i>																																																																																																																																																	
<a href="ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/studies/ena/SRP007483/homo_sapiens/genes.tpm.htseq2.tsv">ftp://ftp.ebi.ac.uk/pub/databases/arrayexpress/data/atlas/rnaseq/studies/ena/SRP007483/homo_sapiens/genes.tpm.htseq2.tsv</a>																																																																																																																																																	
<table><thead><tr><th>Gene ID</th><th>SRR309133</th><th>SRR309134</th><th>SRR309135</th><th>SRR309136</th><th>SRR309137</th><th>SRR309138</th><th>SRR309139</th></tr><tr><th>SRR309140</th><th>SRR309141</th><th>SRR309142</th><th>SRR309143</th><th>SRR309144</th><th></th><th></th><th></th></tr></thead><tbody><tr><td>ENSG000000000003</td><td>3.13</td><td>3.04</td><td>3.39</td><td>3.33</td><td>5.97</td><td>7.32</td><td>3.35</td><td>3.75</td><td>3.25</td></tr><tr><td>ENSG000000000005</td><td>0</td><td>0</td><td>0.06</td><td>0.12</td><td>0.25</td><td>0.18</td><td>0.43</td><td>0.29</td><td>0.27</td></tr><tr><td>ENSG00000000419</td><td>7.47</td><td>7.95</td><td>16.86</td><td>15.17</td><td>21.79</td><td>22.01</td><td>24.15</td><td>24.09</td><td>42.88</td></tr><tr><td>ENSG00000000457</td><td>0.67</td><td>0.73</td><td>0.55</td><td>0.48</td><td>2.7</td><td>3.24</td><td>3.2</td><td>2.15</td><td>3.97</td></tr><tr><td>ENSG00000000460</td><td>0.21</td><td>0.32</td><td>0.1</td><td>0.21</td><td>1.06</td><td>1.14</td><td>0.66</td><td>0.52</td><td>0.53</td></tr><tr><td>ENSG00000000938</td><td>0.78</td><td>0.55</td><td>1.44</td><td>1.73</td><td>8.71</td><td>9.22</td><td>3.48</td><td>3.74</td><td>2.5</td></tr><tr><td>ENSG00000000971</td><td>0.88</td><td>1.09</td><td>6.58</td><td>6.57</td><td>5.12</td><td>4.69</td><td>1.85</td><td>1.76</td><td>1.28</td></tr><tr><td>ENSG00000001036</td><td>4.93</td><td>5.5</td><td>5.5</td><td>5.31</td><td>14.87</td><td>15.48</td><td>9.32</td><td>8.59</td><td>9.99</td></tr><tr><td>ENSG00000001084</td><td>5.49</td><td>5.32</td><td>5.74</td><td>6.07</td><td>11.03</td><td>10.79</td><td>8.32</td><td>7.28</td><td>8.88</td></tr><tr><td>ENSG00000001167</td><td>5.31</td><td>5.69</td><td>6.83</td><td>7.37</td><td>11.83</td><td>12.78</td><td>8.28</td><td>8.19</td><td>8.08</td></tr><tr><td>ENSG00000001460</td><td>0.78</td><td>0.69</td><td>1.34</td><td>1.04</td><td>6.2</td><td>6.48</td><td>5.02</td><td>4.67</td><td>8.36</td></tr><tr><td>ENSG00000001461</td><td>14.64</td><td>14.75</td><td>17.54</td><td>17.8</td><td>47.25</td><td>46.31</td><td>18.4</td><td>17.97</td><td>39.94</td></tr></tbody></table>										Gene ID	SRR309133	SRR309134	SRR309135	SRR309136	SRR309137	SRR309138	SRR309139	SRR309140	SRR309141	SRR309142	SRR309143	SRR309144				ENSG000000000003	3.13	3.04	3.39	3.33	5.97	7.32	3.35	3.75	3.25	ENSG000000000005	0	0	0.06	0.12	0.25	0.18	0.43	0.29	0.27	ENSG00000000419	7.47	7.95	16.86	15.17	21.79	22.01	24.15	24.09	42.88	ENSG00000000457	0.67	0.73	0.55	0.48	2.7	3.24	3.2	2.15	3.97	ENSG00000000460	0.21	0.32	0.1	0.21	1.06	1.14	0.66	0.52	0.53	ENSG00000000938	0.78	0.55	1.44	1.73	8.71	9.22	3.48	3.74	2.5	ENSG00000000971	0.88	1.09	6.58	6.57	5.12	4.69	1.85	1.76	1.28	ENSG00000001036	4.93	5.5	5.5	5.31	14.87	15.48	9.32	8.59	9.99	ENSG00000001084	5.49	5.32	5.74	6.07	11.03	10.79	8.32	7.28	8.88	ENSG00000001167	5.31	5.69	6.83	7.37	11.83	12.78	8.28	8.19	8.08	ENSG00000001460	0.78	0.69	1.34	1.04	6.2	6.48	5.02	4.67	8.36	ENSG00000001461	14.64	14.75	17.54	17.8	47.25	46.31	18.4	17.97	39.94
Gene ID	SRR309133	SRR309134	SRR309135	SRR309136	SRR309137	SRR309138	SRR309139																																																																																																																																										
SRR309140	SRR309141	SRR309142	SRR309143	SRR309144																																																																																																																																													
ENSG000000000003	3.13	3.04	3.39	3.33	5.97	7.32	3.35	3.75	3.25																																																																																																																																								
ENSG000000000005	0	0	0.06	0.12	0.25	0.18	0.43	0.29	0.27																																																																																																																																								
ENSG00000000419	7.47	7.95	16.86	15.17	21.79	22.01	24.15	24.09	42.88																																																																																																																																								
ENSG00000000457	0.67	0.73	0.55	0.48	2.7	3.24	3.2	2.15	3.97																																																																																																																																								
ENSG00000000460	0.21	0.32	0.1	0.21	1.06	1.14	0.66	0.52	0.53																																																																																																																																								
ENSG00000000938	0.78	0.55	1.44	1.73	8.71	9.22	3.48	3.74	2.5																																																																																																																																								
ENSG00000000971	0.88	1.09	6.58	6.57	5.12	4.69	1.85	1.76	1.28																																																																																																																																								
ENSG00000001036	4.93	5.5	5.5	5.31	14.87	15.48	9.32	8.59	9.99																																																																																																																																								
ENSG00000001084	5.49	5.32	5.74	6.07	11.03	10.79	8.32	7.28	8.88																																																																																																																																								
ENSG00000001167	5.31	5.69	6.83	7.37	11.83	12.78	8.28	8.19	8.08																																																																																																																																								
ENSG00000001460	0.78	0.69	1.34	1.04	6.2	6.48	5.02	4.67	8.36																																																																																																																																								
ENSG00000001461	14.64	14.75	17.54	17.8	47.25	46.31	18.4	17.97	39.94																																																																																																																																								

# *Summary of the session*



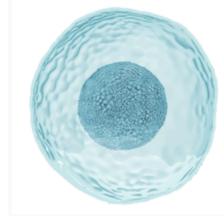


# New project on single-cell...

 Single Cell Expression Atlas

Single-cell gene expression across species

[Home](#) [Browse experiments](#) [Download](#) [Help](#) [About](#)



## Analysing gene expression for single cells experiments

The cell is a natural unit of biology, whose type and state can vary according to external influences or to internal processes. In multicellular organisms, all cells are derived from a single zygote which, through regulated programmes of proliferation and differentiation, generates all of the diverse cell types that populate the organism. Dysregulation of these programmes in single 'renegade' cells can lead to diseases such as cancers, neurological disorders and developmental disorders.

List of single-cell experiments

Total : 3 experiments

 06-02-2017 Single-cell RNA-seq analysis of human pancreas from healthy individuals and type 2 diabetes patients <i>Homo sapiens</i>	3514	<a href="#">Analysis result</a>	
 06-02-2017 Single-cell RNA-sequencing resolves a CD4+ T cell fate bifurcation <i>Mus musculus</i>	752	<a href="#">Analysis result</a>	
 06-02-2017 Single-cell RNA-sequencing of <i>Platynereis dumerilii</i> larval brain cells <i>Platynereis dumerilii</i>	285	<a href="#">Analysis result</a>	

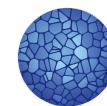
## Marker genes search

Gene ID, gene name or gene feature

Species

Any

Examples: REG1B, zinc finger, lung, leaf, valproic acid, cancer



HUMAN  
CELL  
ATLAS



# Expression Atlas: Who we are?

Robert  
Petryszak  
Team Leader



Irene  
Papatheodorou  
Co-ordinator



## Data curation

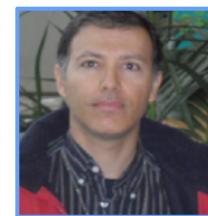


Laura Huerta

## Data analysis



Suhail Mohammed



Nuno Fonseca

## User Interface



Alfonso Fuentes



Elisabet Barrera



Nancy George



Anja Fullgrabe



Wojtek Bazant



Haider Iqbal

# 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 checkmark 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 fourth 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](mailto:atlas-feedback@ebi.ac.uk)



@ExpressionAtlas

# Introduction to Next Generation Sequencing

Public resources for accessing  
RNA-seq data at EMBL-EBI

Laura Huerta, PhD

Senior Scientific Curator, Gene Expression Team

[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)

10 October 2017

