

EMBL-EBI workshop: Data and tools for transcriptomics and protein biology

Expression Atlas: gene expression results across species and conditions

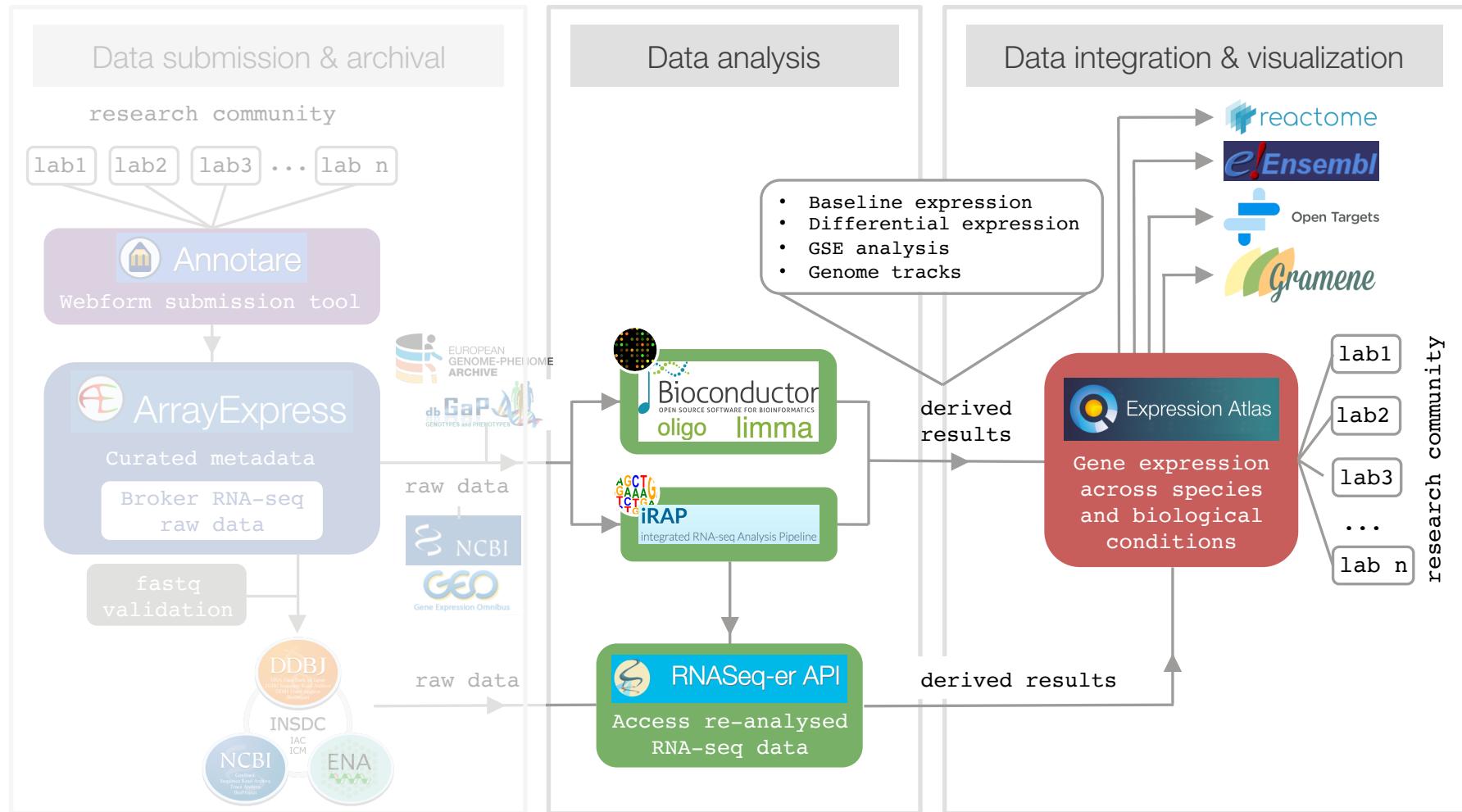
Laura Huerta, PhD

Senior Scientific Curator

lauhuema@ebi.ac.uk

Malta, 31 July 2018

Functional genomics resources at EMBL-EBI





What is Expression Atlas?



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)

Organism Any

Biological conditions

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

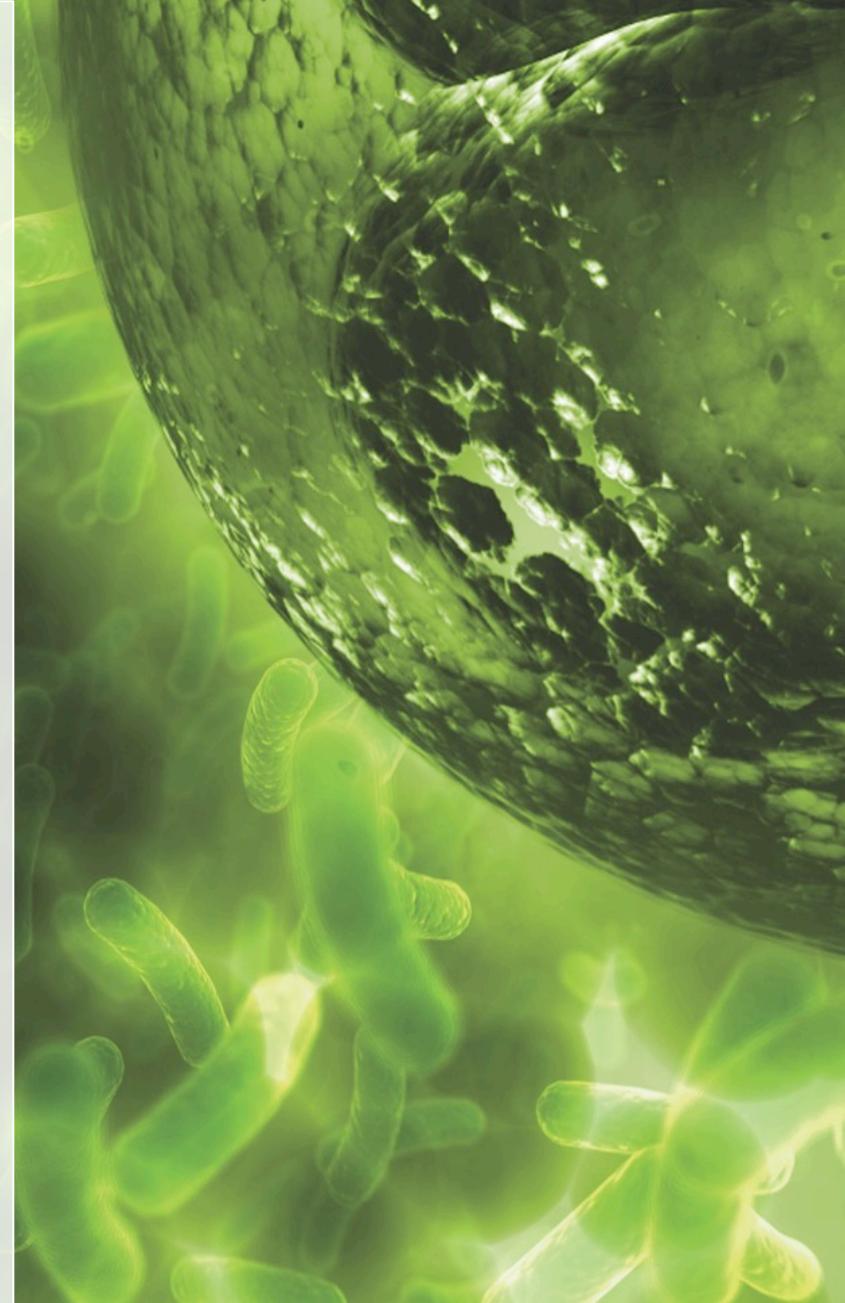
[Search](#) [Clear](#)

Open science resource for gene/protein expression across species and biological conditions

www.ebi.ac.uk/gxa/

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?

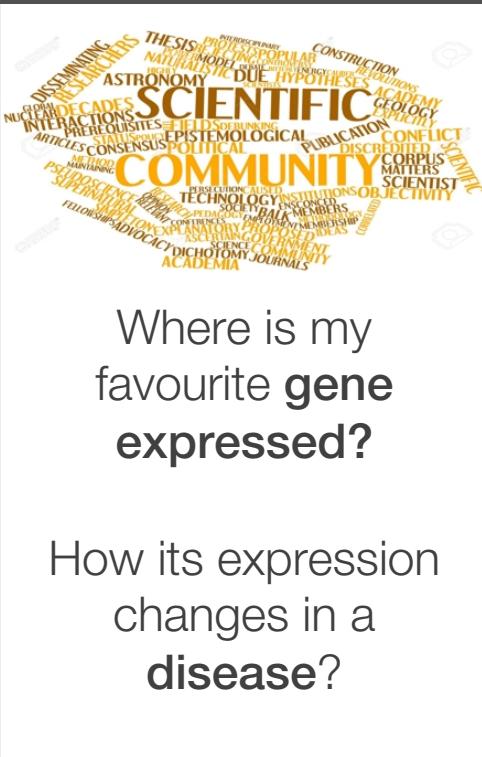
Big Data



‘Big’ Analysis



'Big' Results





How ‘big’ is Expression Atlas data analysis?



<https://commonfund.nih.gov/gtex>

Genetic effects on gene expression across human tissues

GTEX Consortium, Lead analysts:, Laboratory, Data Analysis & Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, Alexis Battle, Christopher D. Brown, Barbara E. Engelhardt & Stephen B. Montgomery

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Nature 550, 204–213 (12 October 2017) | doi:10.1038/nature24277

Received 08 September 2016 | Accepted 15 September 2017 | Published online 11 October 2017

doi:10.1038/nature24267

doi:10.1038/nature24265

doi:10.1038/nature24041



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



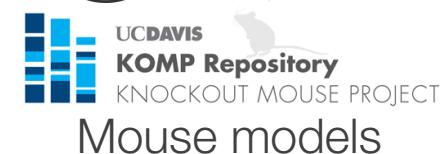
Cancer research



Zebrafish development

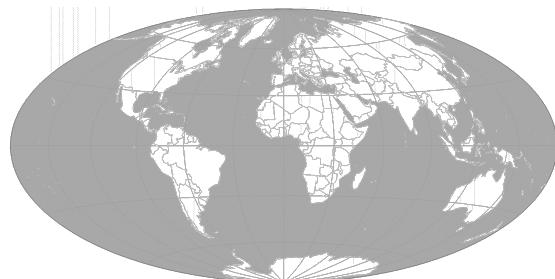


Prenatal human brain



Mouse models

Expression Atlas contains thousands of selected and curated datasets



> 3,300 datasets



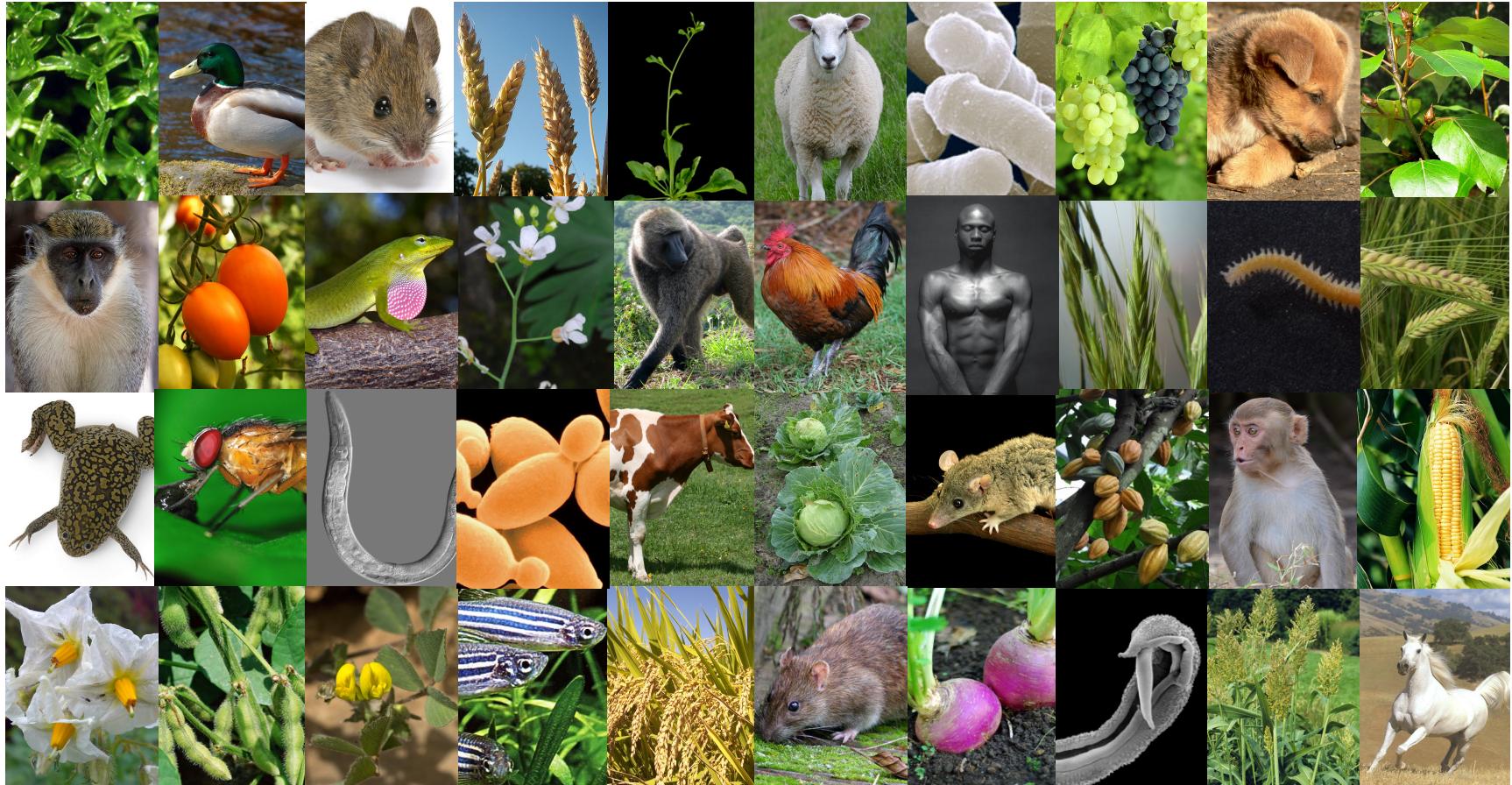
150 baseline expression



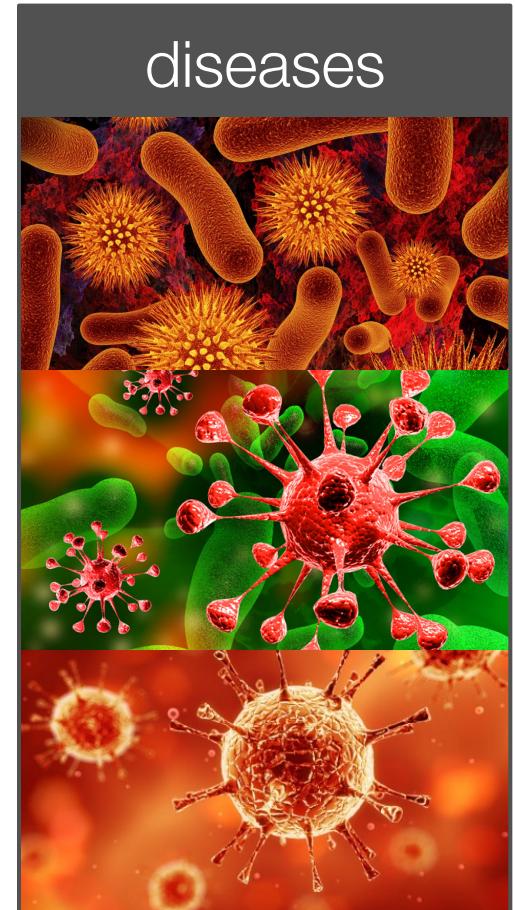
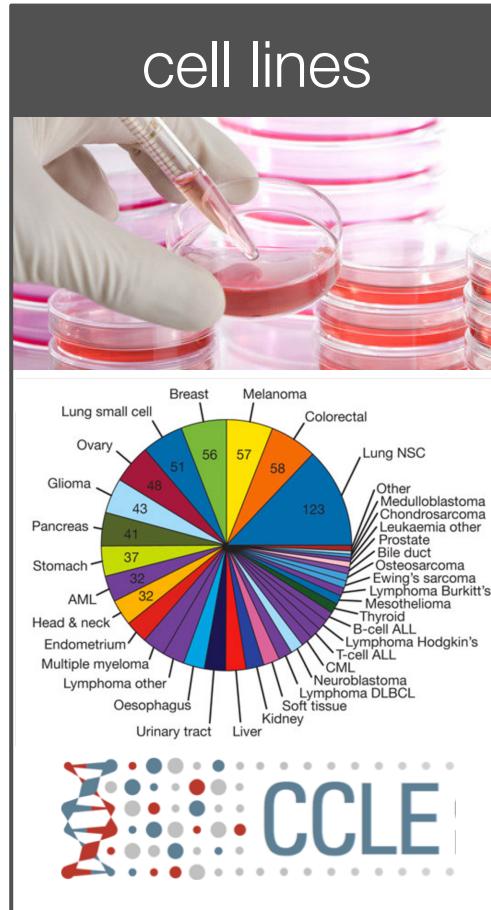
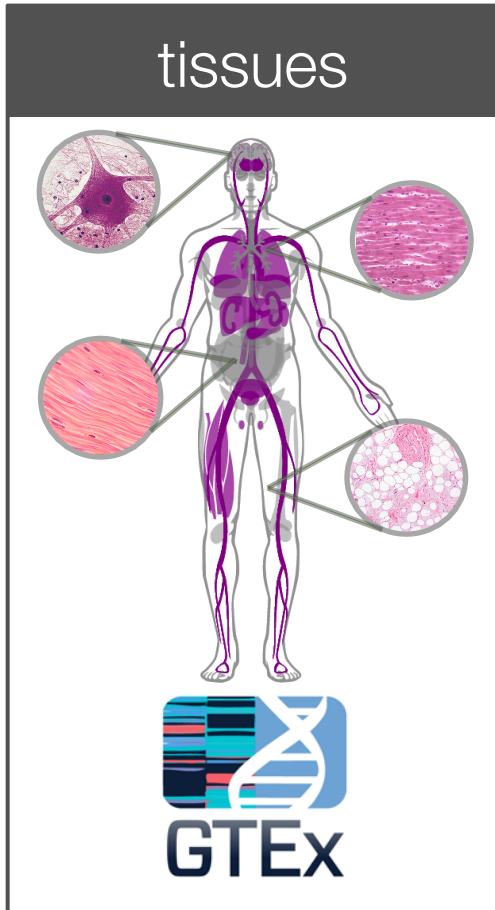
> 3,150 differential expression



... across more than 45 species ...



... under different biological conditions ...



... to help scientists find answers ...

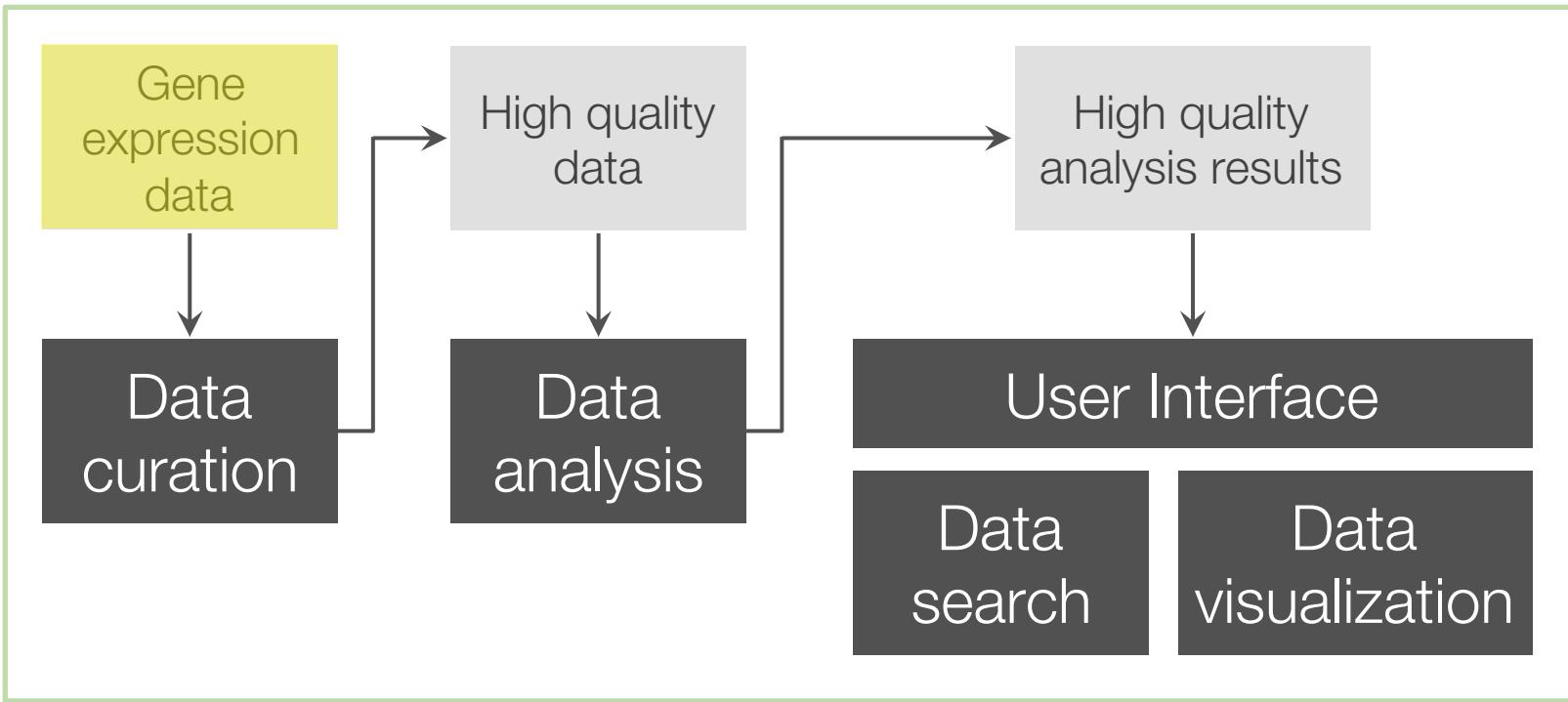
In which conditions
is my favourite gene
expressed?

What genes are
differentially expressed in
a particular condition?

What genes are
expressed in a
particular condition?

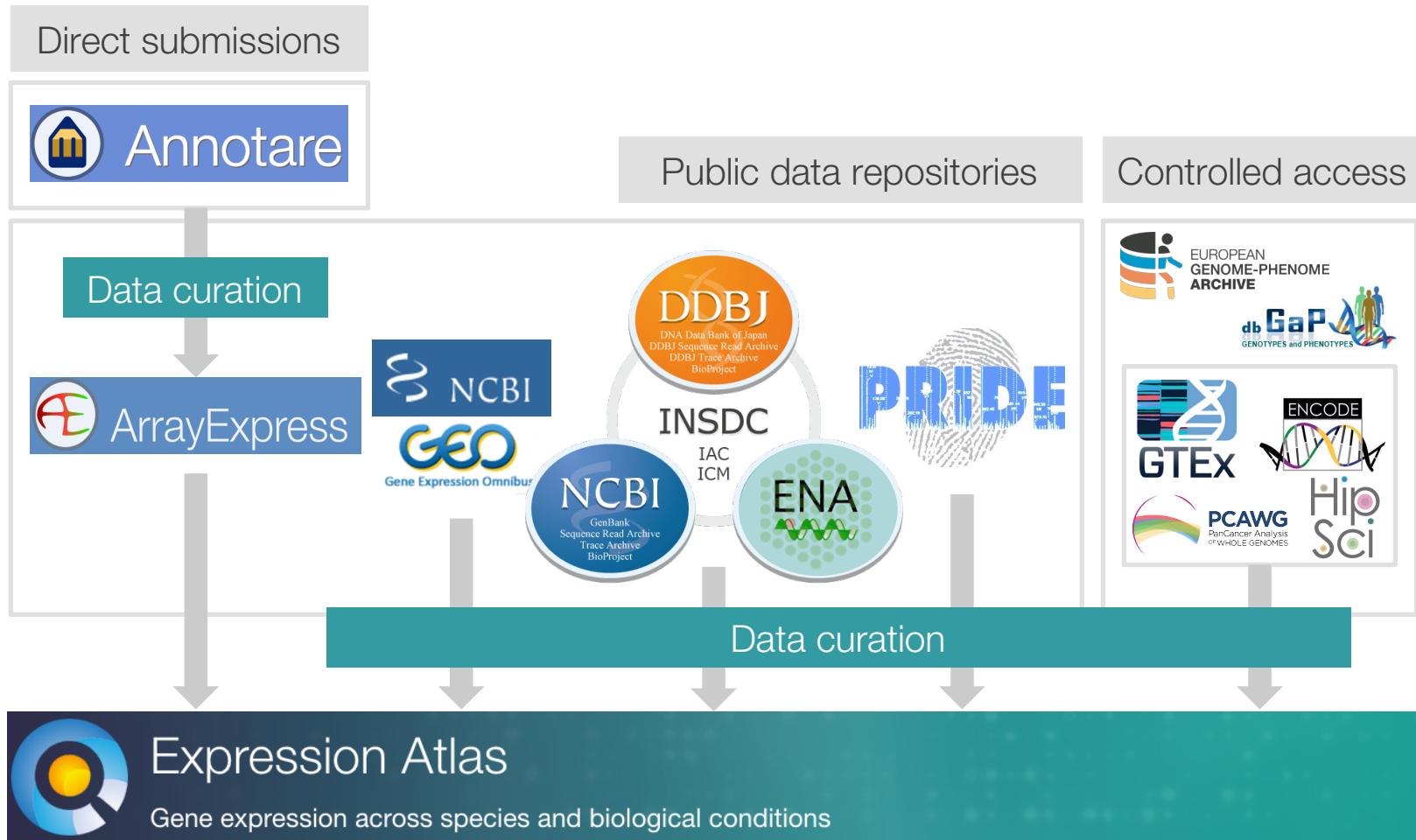


How does Expression Atlas do it?



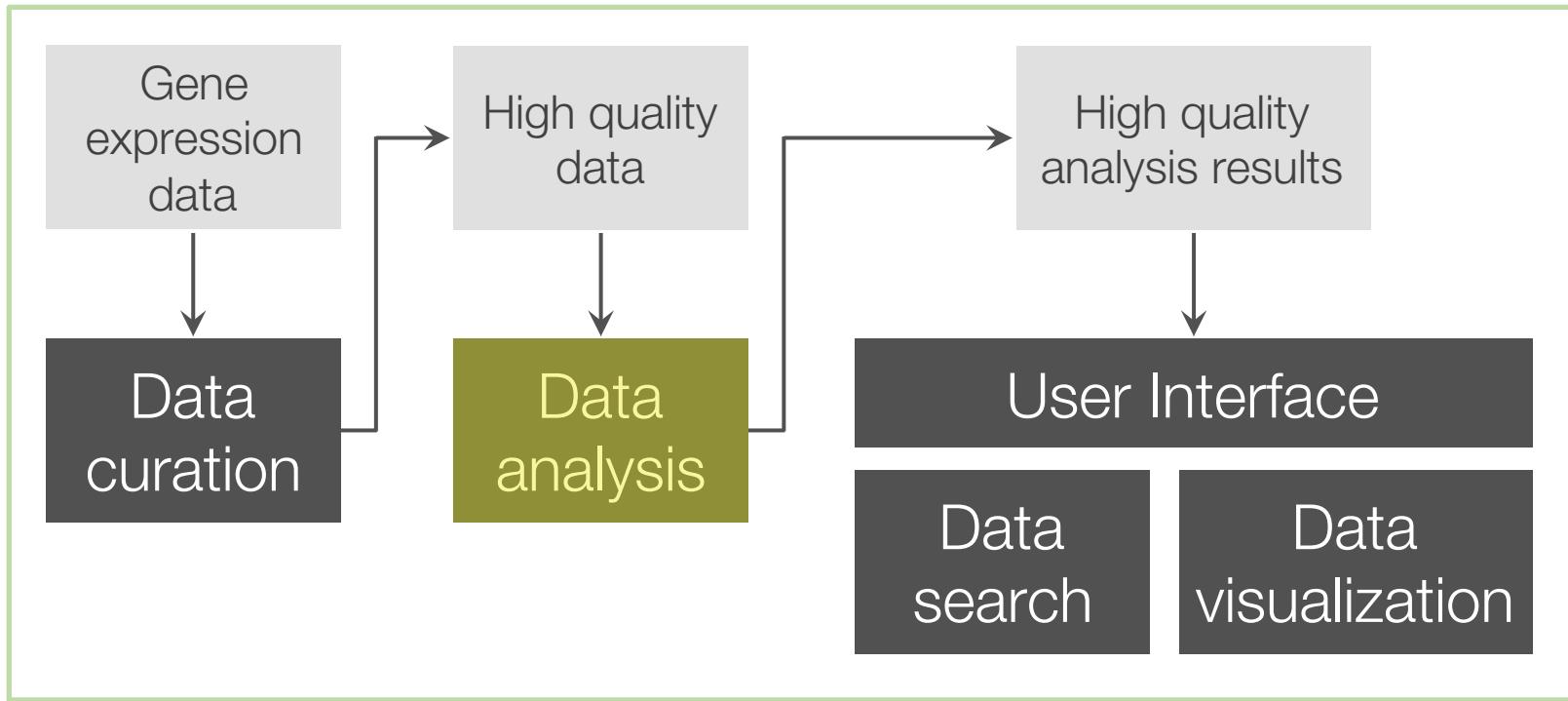
Discover and interpret gene expression analysis results quickly and easy

Where does expression data come from?





How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

RNA-sequencing workflow



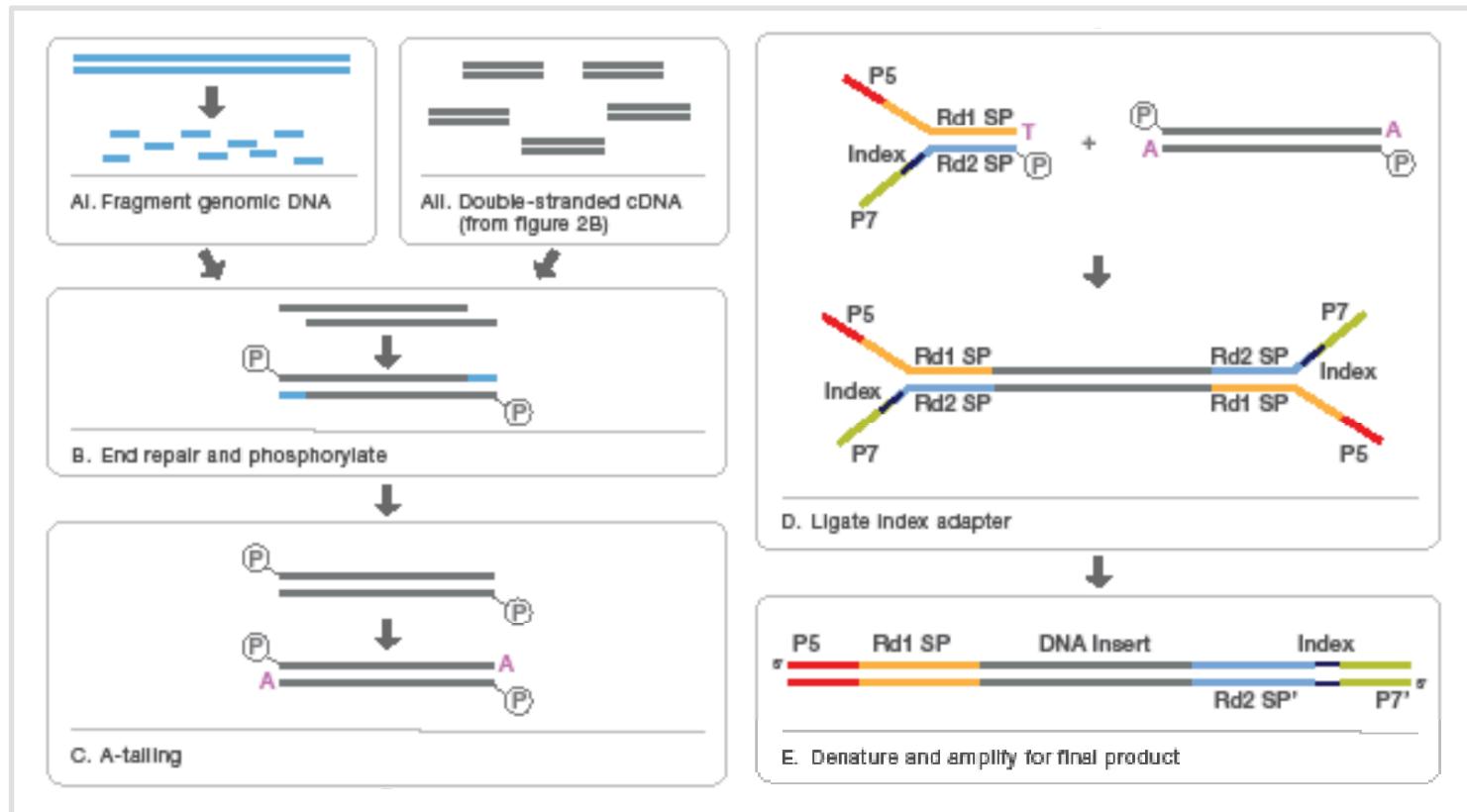
Library Prep



Sequencing



Data Analysis



RNA-sequencing workflow



Sequencing

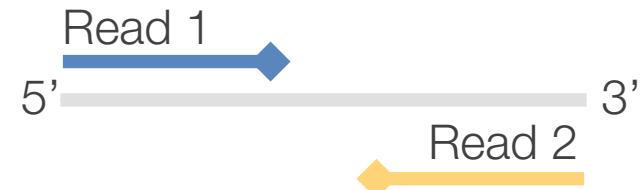
Single-end sequencing



my_sequence.fastq

@HWI-BRUNOP16X_0001:1:1:1466:1018#0/1
AAGGAAGTGTCTGCTGGCTAACACAGCNAGNCACGTGAC
+
aVfbe`^^^ TTTSSdfffffdffffabbZbbfebafbbbb

Paired-end sequencing



my_sequence_I.fastq

@HWI-BRUNOP16X_0001:1:1:1278:989#0/1
NAAATTCGAATTCTGTGAAGTAAGCATCTTCTTTGTCA-
+
BJJGGKIINN^^^^^QONTUQOOTTTRTOTY^^Y^\\/^\\

my_sequence_2.fastq

@HWI-BRUNOP16X_0001:1:1:1278:989#0/2
AACCCACACAGGAGAGCAGCCTTACAGATGCAAATACTGTG
+
]K f f f f f g g g h g e g g g g g g d g g g g g f g g g g g e g g g g h h

RNA-sequencing workflow



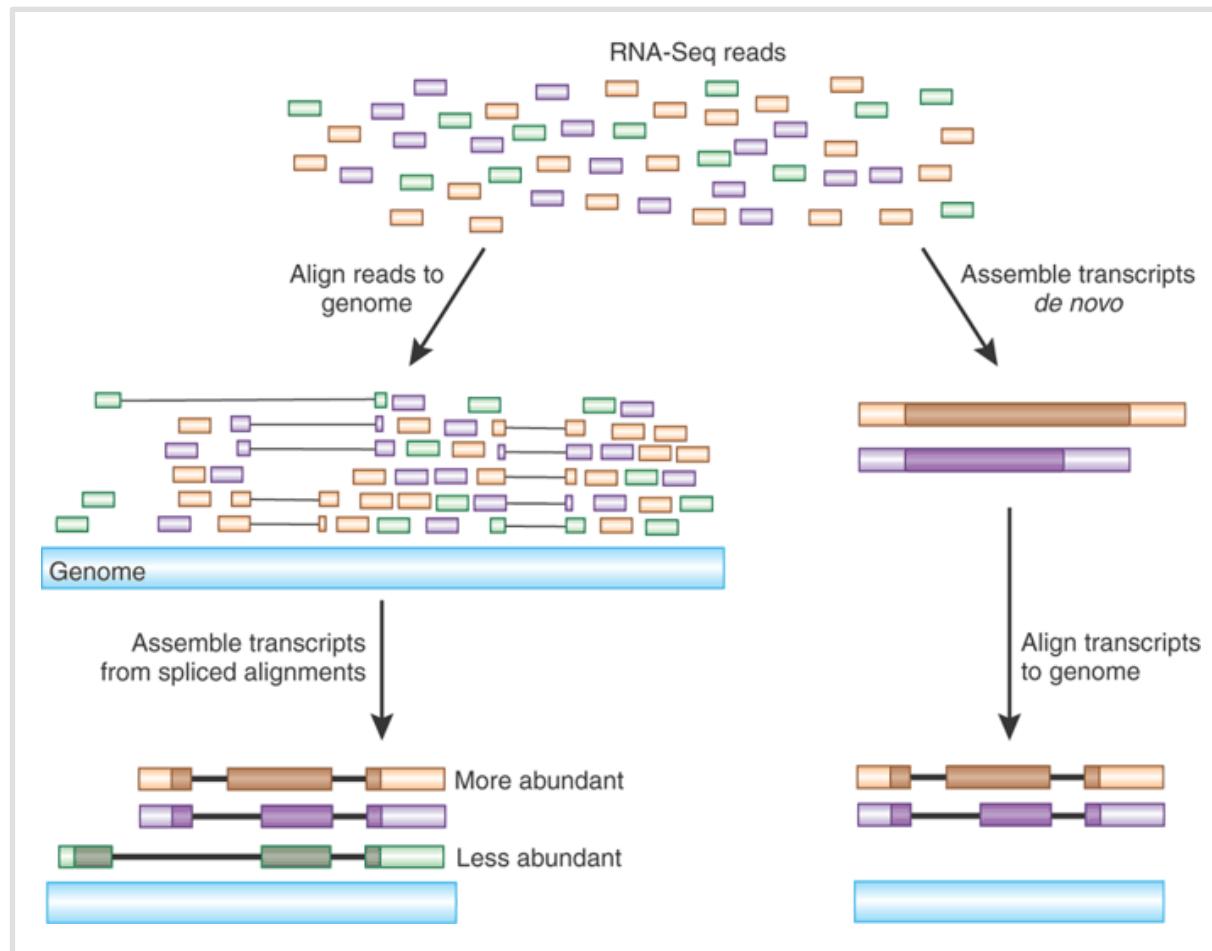
Library Prep



Sequencing



Data Analysis



Haas and Zody, 2010; doi:10.1038/nbt0510-421

RNA-sequencing workflow



Library Prep



Sequencing



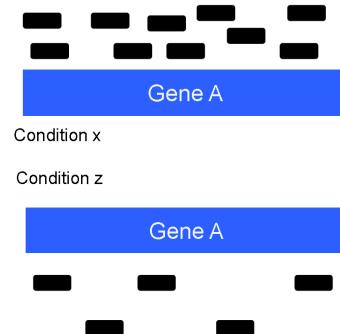
Data Analysis

Gene length normalisation



Gene	Raw reads	Length	Normalised Reads
A	10	2	5
B	5	1	5

Sequencing depth normalisation



Condition	Raw reads	Total mapped reads	Normalised Reads
x	10	1000	0.01
z	5	500	0.01

RNA-sequencing workflow



Library Prep



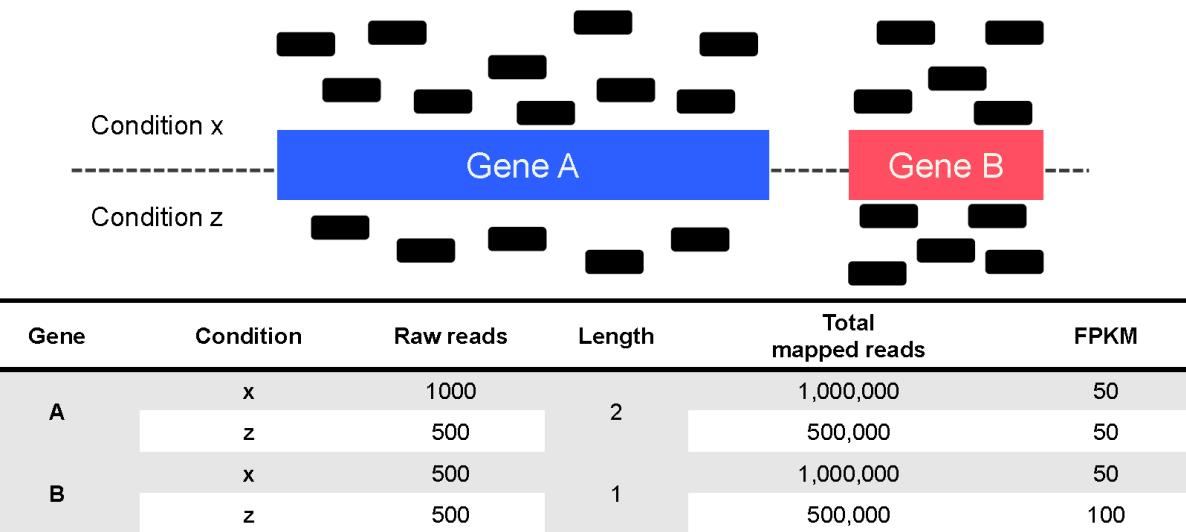
Sequencing



Data Analysis

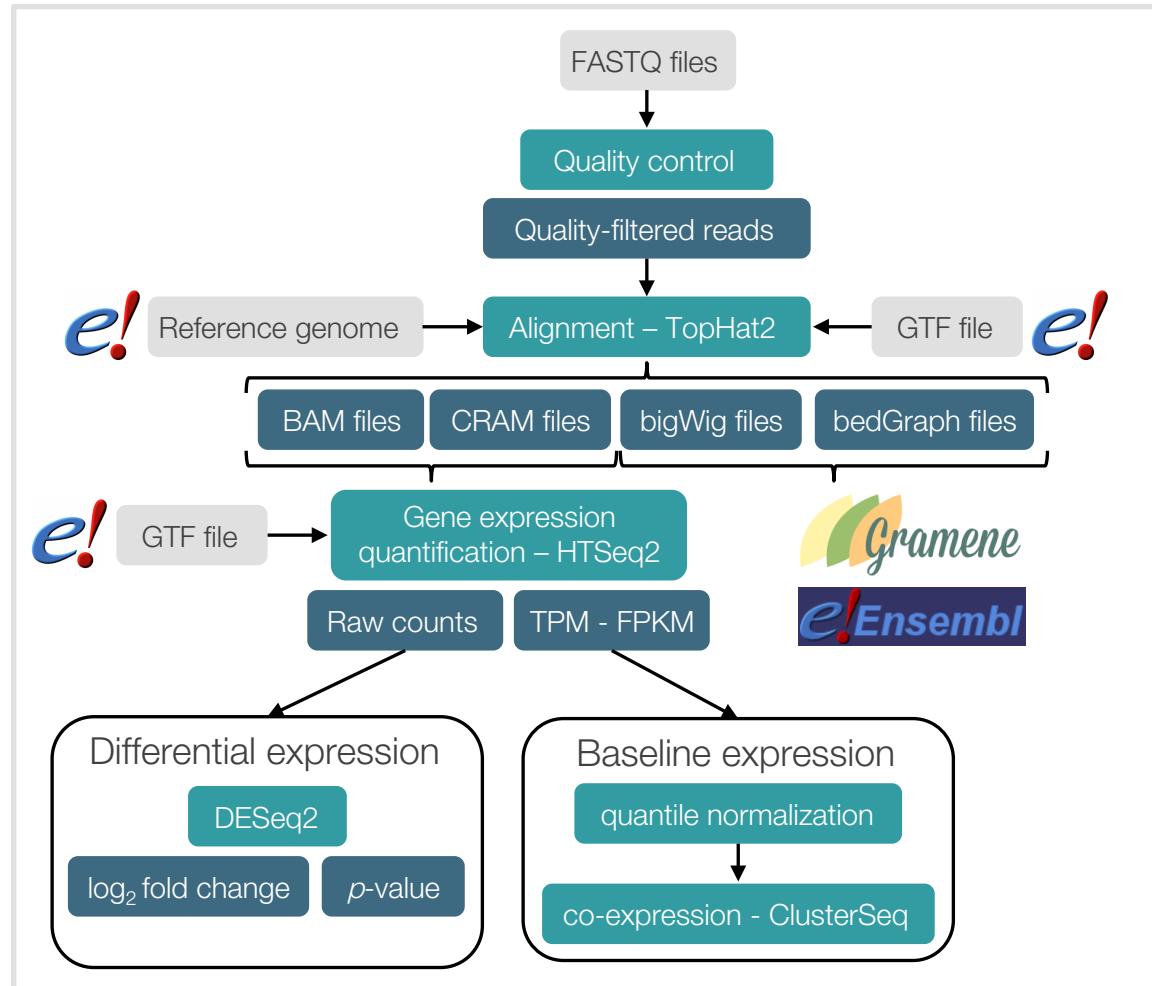
FPKM and TPM values

Normalised expression values for both gene length and total number of mapped reads





Expression Atlas – RNA-seq data analysis



RNA-seq data analysis

<https://www.biorxiv.org/content/early/2014/06/06/005991>

iRAP - an integrated RNA-seq Analysis Pipeline

Nuno A. Fonseca, Robert Petryszak, John Marioni, Alvis Brazma

doi: <https://doi.org/10.1101/005991>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

Info/History Metrics

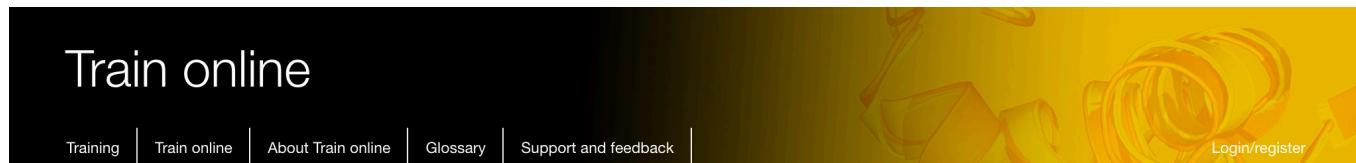
Preview PDF

Abstract

RNA-sequencing (RNA-Seq) has become the technology of choice for whole-transcriptome profiling. However, processing the millions of sequence reads generated requires considerable bioinformatics skills and computational resources. At each step of the processing pipeline many tools are available, each with specific advantages and disadvantages. While using a specific combination of tools might be desirable, integrating the different tools can be time consuming, often due to specificities in the formats of input/output files required by the different programs. Here we present iRAP, an integrated RNA-seq analysis pipeline that allows the user to select and apply their preferred combination of existing tools for mapping reads, quantifying expression, testing for differential expression. iRAP also includes multiple tools for gene set enrichment analysis and generates web browsable reports of the results obtained in the different stages of the pipeline. Depending upon the application, iRAP can be used to quantify expression at the gene, exon or transcript level. iRAP is aimed at a broad group of users with basic bioinformatics training and requires little experience with the command line. Despite this, it also provides more advanced users with the ability to customise the options used by their chosen tools.

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

EBI training: RNA-sequencing



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-sequencing (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



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>

EBI training: RNA-sequencing

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

- Introduction
- Real-time PCR
- Microarrays
- Next Generation Sequencing (NGS)
- RNA sequencing
 - Advantages of RNA-seq over hybridisation-based approaches
 - Applications of RNA-seq
 - Performing a RNA-seq experiment**
 - Design considerations
 - Number of replicates
 - Laboratory performance
 - Data analysis
 - Quality control
 - Read mapping or alignment
 - Quantification
 - Differential gene expression analysis
 - 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



All course materials in Train online
are free cultural works licensed
under a Creative Commons
Attribution-ShareAlike 4.0
International license

Performing a RNA-seq experiment

There is no common pipeline for all the different types of RNA-seq. Both the experimental design and the analysis procedures will be different depending on the organism being studied and the research goals. As the primary objective of many biological studies is a comparison of gene expression between samples, we outline a possible workflow for detecting differential expression using RNA-seq (Figure 7).

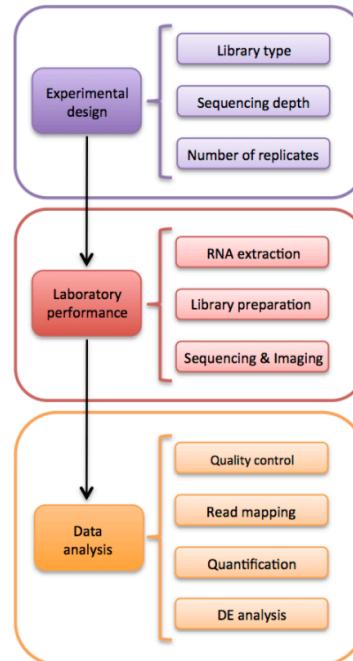


Figure 7 Overview of a RNA-seq experiment for detecting differential expression (DE).

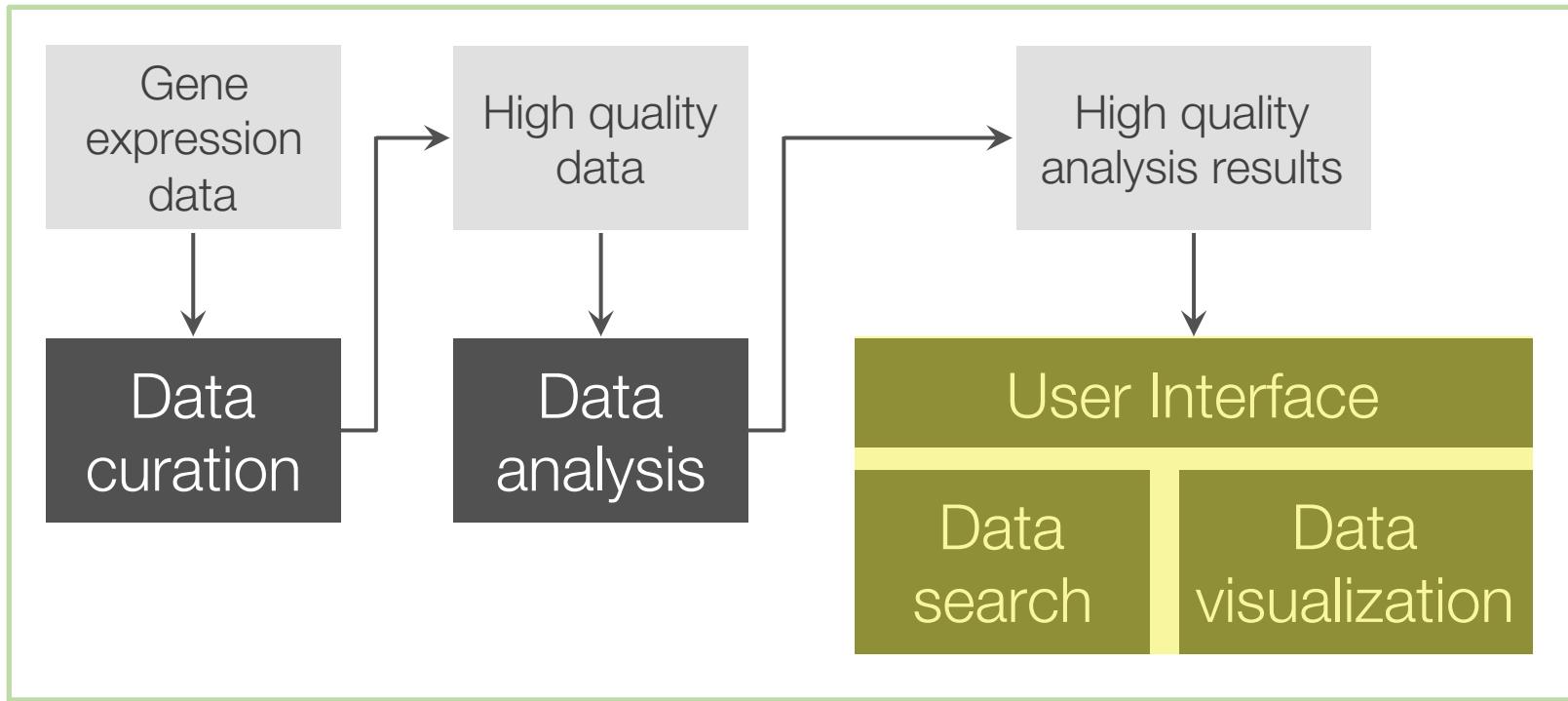
« Applications of RNA-seq

Next: Design considerations »

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



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



Let's try Expression Atlas

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
expressed?

Search | Gene set enrichment

Gene / Gene properties

UMOD x

umod

Umod

Umod1

umod.1

UMOD1

UMOD

UMOD1-AS1

symbol expression)

Organism

Homo sapiens

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

Any

Aegilops tauschii
Amborella trichopoda
Anas platyrhynchos
Anolis carolinensis
Anopheles gambiae
Arabidopsis lyrata
Arabidopsis thaliana

Biological conditions

Enter condition query...

Examples: lung, leaf, valproic acid, cancer

Search

Clear



Baseline expression results

ENSG00000169344 (UMOD) Homo sapiens uromodulin

Baseline expression

Differential expression

UMOD information

Show anatomograms

Filter your results

Homo sapiens

Organism part

Cell line

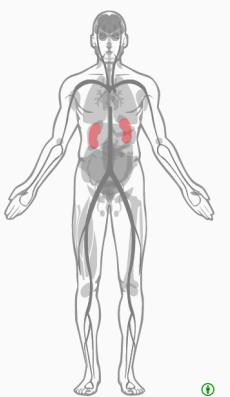
Cell type

Disease

Individual

Organism part

Showing 10 experiments:



①

19 NIH Epigenomics Roadmap

ENCODE (M. Snyder lab)

32 Uhlen's Lab

GTEX

Mammalian Kaessmann

Illumina Body Map

HDBR developing brain - 10 post conception weeks

68 FANTOM5 project - adult

Human Proteome Map - adult

Human Protein Atlas

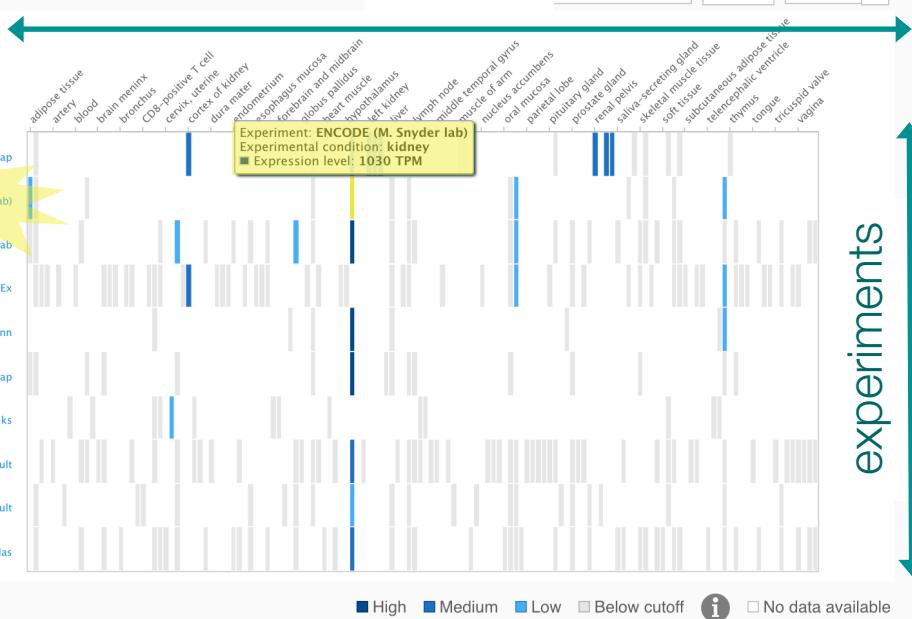
UMOD gene expression across human tissues from different experiments is integrated and visualised in one heatmap

tissues

By experiment type ▾

Filters

Download ▾





Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab for the RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

**UMOD gene expression results
in a particular experiment**





Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab for the ENCODE project

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

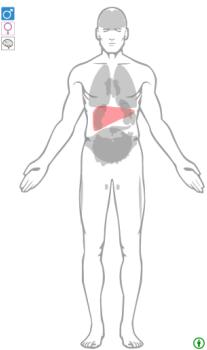
Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Results Experiment Design Supplementary Information Downloads

Genes

Showing 50 of 37,057 genes found:



Specific checked

Most specific

Expression value

[See distribution](#)

Data units

TPM
 FPKM

Organism parts

[Select](#)

Selected: 13 / 13





Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab for the ENCODE project

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

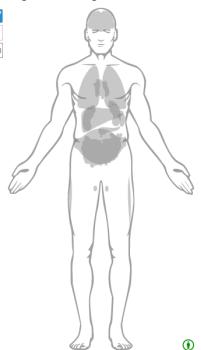
Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Results Experiment Design Supplementary Information Downloads

Genes

Showing 50 of 37,057 genes found:



Specific unchecked

Most specific

Expression value

0.5

See distribution

Data units

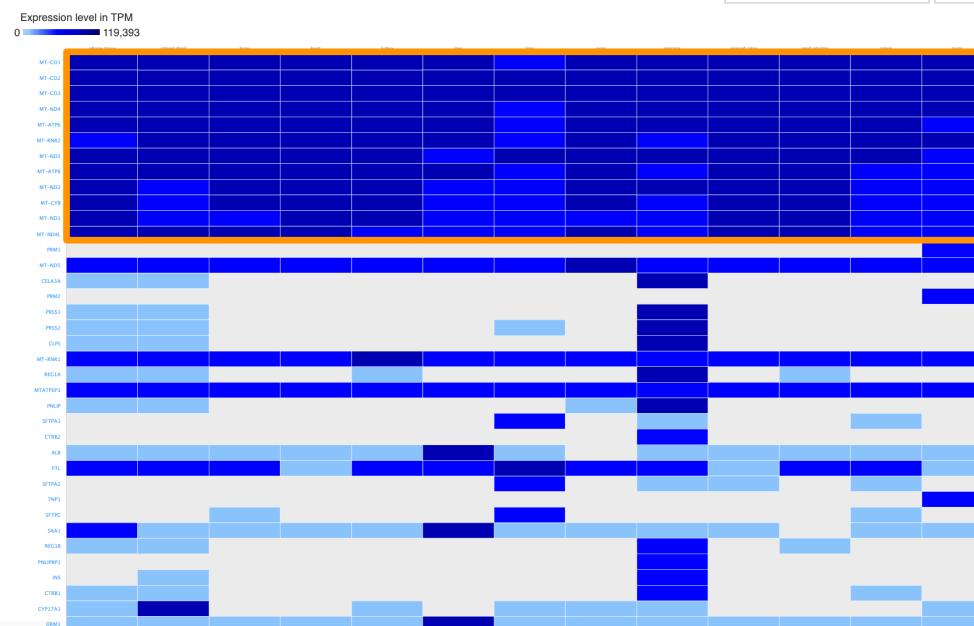
TPM

FPKM

Organism parts

Select

Selected: 13 / 13



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

First, genes with
high expression
in all tissues



Baseline experiment page

General information

Strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab for the ENCODE project

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Results

Experiment Design

Supplementary Information

Downloads

Download results

Genes

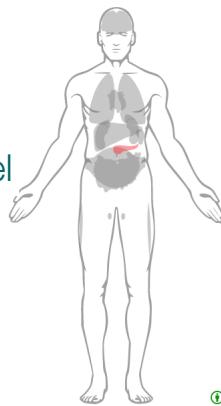
Showing 50 of 37,057 genes found:

Genes expressed above 0.5 TPM



Apply

Clear



Expression level in TPM
0 13,039

PRM1
PRM2
CTR2B
TNP1
PNLIPRP1
AMY2A
AC009078.2
BOD1L2
REG1CP
SMCP

Expression level of a gene

Gene name: CTR2B
Organism part: pancreas
Expression level: 5313 TPM
Number of biological replicates: 1



Minimum expression level

Switch data units

TPM
 FPKM

See sample annotations

Selected: 13 / 13



Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Reference(s): [25413365 \(Filter by genes in paper\)](#)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Search for a particular gene, e.g. *TSPAN6* and find genes with similar expression

Results

Experiment Design

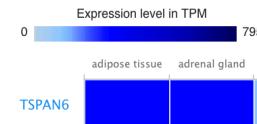
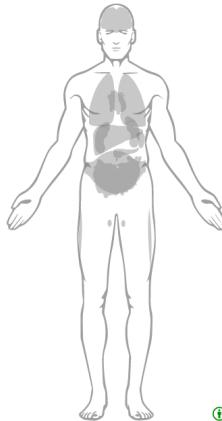
Supplementary Information

Downloads

Genes

Show boxplot and transcripts view

Showing 1 gene:



Select genome browser to view tracks ▾

Download ▾

[#] Add similarly expressed genes

TSPAN6

TSPAN6 (symbol)

tspan6 (symbol)

Tspan6 (symbol)

TSPAN6 (hgnc_symbol)

Tspan6 (rgd_symbol)

tetraspanin 6 (**tspan6**), mRNA [Source:RefSeq mRNA;Acc:NM_001045652] (description)

Tspan6 (mgf_symbol)

Sus scrofa tetraspanin 6 (**TSPAN6**), mRNA.
[Source:RefSeq mRNA;Acc:NM_001244061]
(description)

Apply

Clear

Most specific

Expression value

0.5



See distribution



Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Search for a particular gene, e.g. *TSPAN6* and find genes with similar expression





Baseline experiment page

e*Ensembl* BLAST/BLAT | BioMart | VEP | Tools | Downloads | Help & Docs | Blog

Human (GRCh38.p12) v

Location: X:100,627,109-100,639,991 Gene: TSPAN6

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
- Synteny
- Alignments (image)
- Alignments (text)
- Region Comparison
- Genetic Variation
- Resequencing
- Linkage Data
- Man

- Other genome browsers
- UCSC[®]
- NCBI[®]
- Ensembl GRCh37.p1

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Chromosome X: 100,627,109-100,639,991



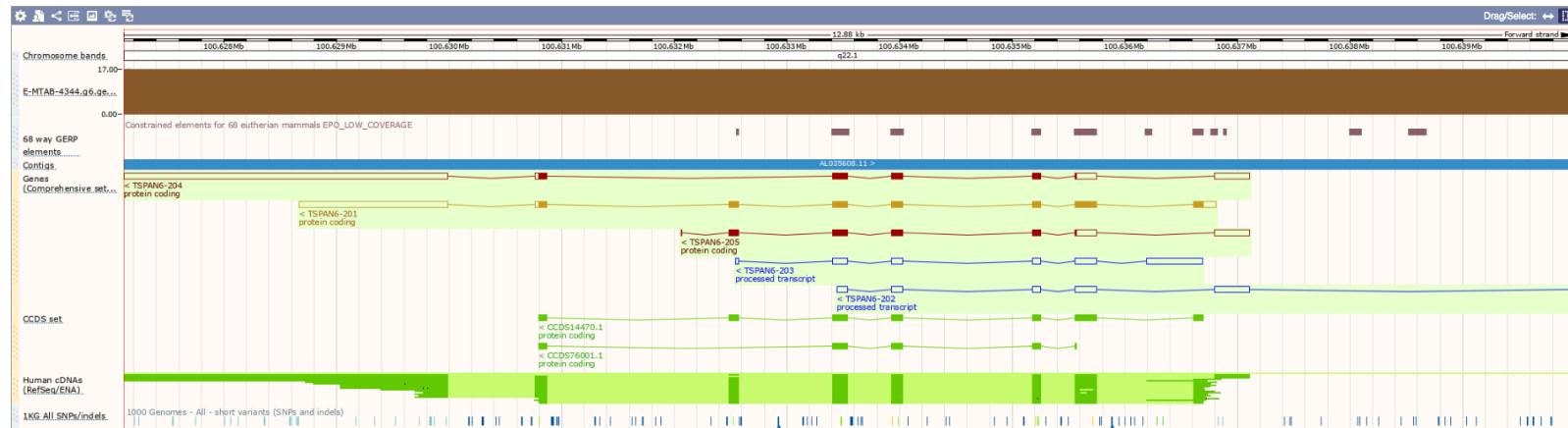
Location: X:100,627,109-100,639,991

Go

Gene:

Go

Drag>Select: ← →



1000 Genomes - All - short variants (SNPs and indels)



Baseline experiment page

Strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Reference(s): 25413365 (Filter by genes in paper)

Raw Data Provider: The ENCODE (Encyclopedia of DNA Elements) Consortium

Download gene quantification matrices
(TPM and FPKM) and R object

Results Experiment Design Supplementary Information **Downloads**

Expression values across all genes (TPM)

Expression values across all genes (FPKM)

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

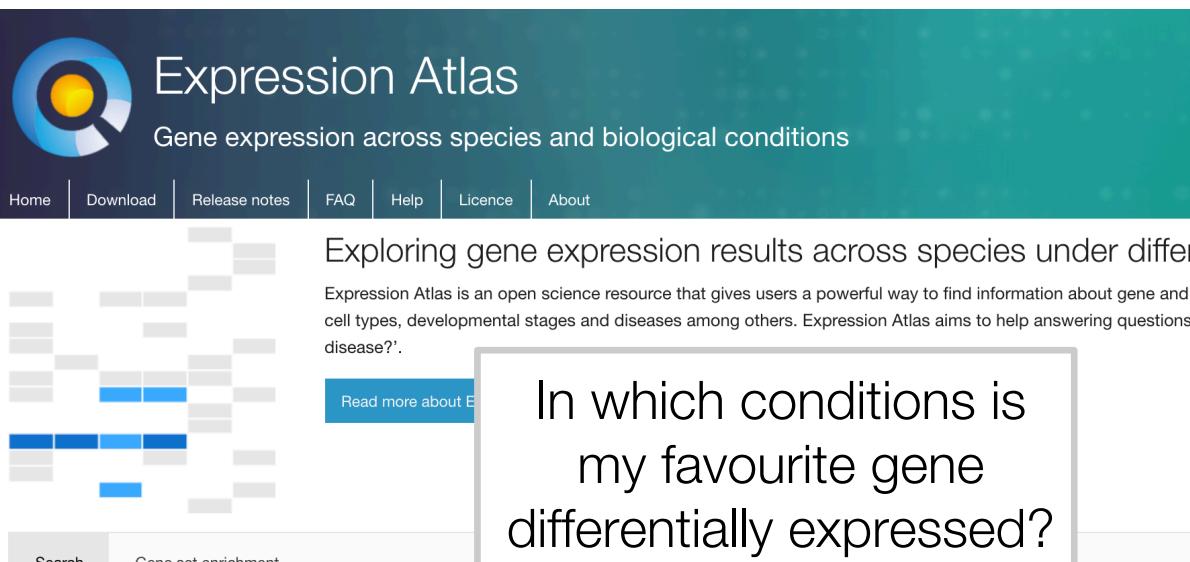


Let's try Expression Atlas

Expression Atlas

Gene expression across species and biological conditions

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



Discover and interpret gene expression analysis results quickly and easy

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?

Search | Gene set enrichment

Gene / Gene properties

CFHR2 x

CFHR2	symbol
Cfhr2	symbol
cfhr2	synonym
Cfhr2	
cfhr2	

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



Differential expression results

Results for CFHR2 (symbol)

Baseline expression Differential expression



Display log₂-fold change

Filter your results

Experimental variables

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

Log ₂ -fold	Species	Gene	Comparison	Experimental variables	Experiment name	
Adjusted p-value	Log ₂ -fold change					
1.575 × 10 ⁻¹⁷	6.5		'lung carcinoma' vs 'normal' in 'CD11b-CD11c- epithelial cell'	cell type, disease	Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq	
	CFH		Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse lung cancer by RNA-Seq 'lung carcinoma' vs 'normal' in 'lung cell'			
	CFH	Property	Test value (N=3)	Reference value (N=3)		
	CFH	cell type	lung cell	lung cell		
	CFH	disease	lung carcinoma	normal		
	CFH	organism	Mus musculus	Mus musculus		
	CFH	organism part	lung	lung		
	CFH	strain	C57BL/6	C57BL/6		
	CFHR2	'hepatobiliary carcinoma' vs 'normal'		block, disease	RNA-seq of blood platelets from six tumor types and healthy donors	
	Cfhr2	'lung carcinoma' vs 'normal' in 'lung cell'		cell type, disease	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



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

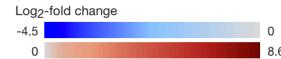
Experiment Design

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

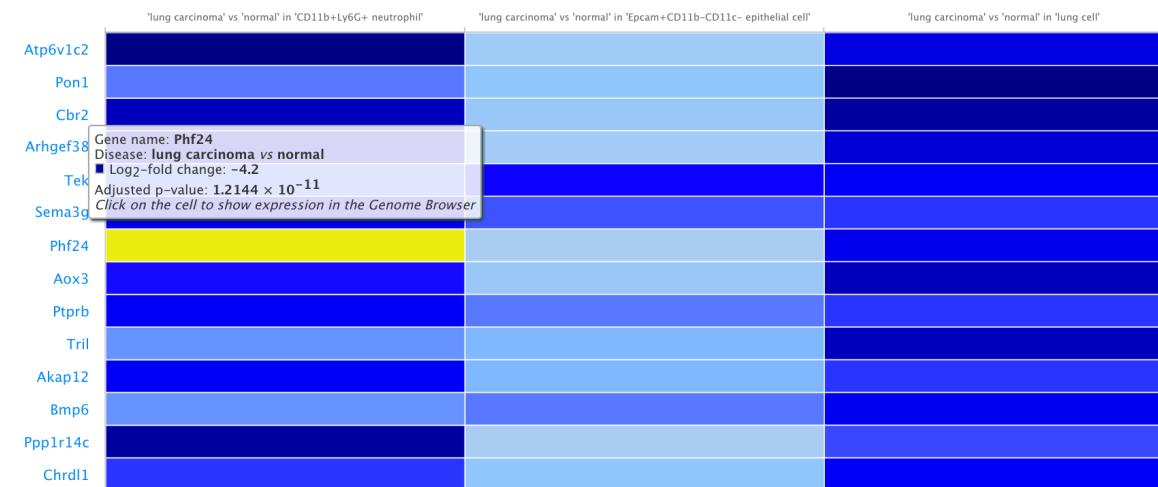
Genes

Showing 50 of 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

Select

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 tumor

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'

GSE GO terms enrichment analysis

GSE Interpro domains enrichment analysis

GSE Reactome pathways enrichment analysis

MA MA plot for the contrast across all genes

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

GSE GO terms enrichment analysis

GSE Interpro domains enrichment analysis

GSE Reactome pathways enrichment analysis

MA MA plot for the contrast across all genes

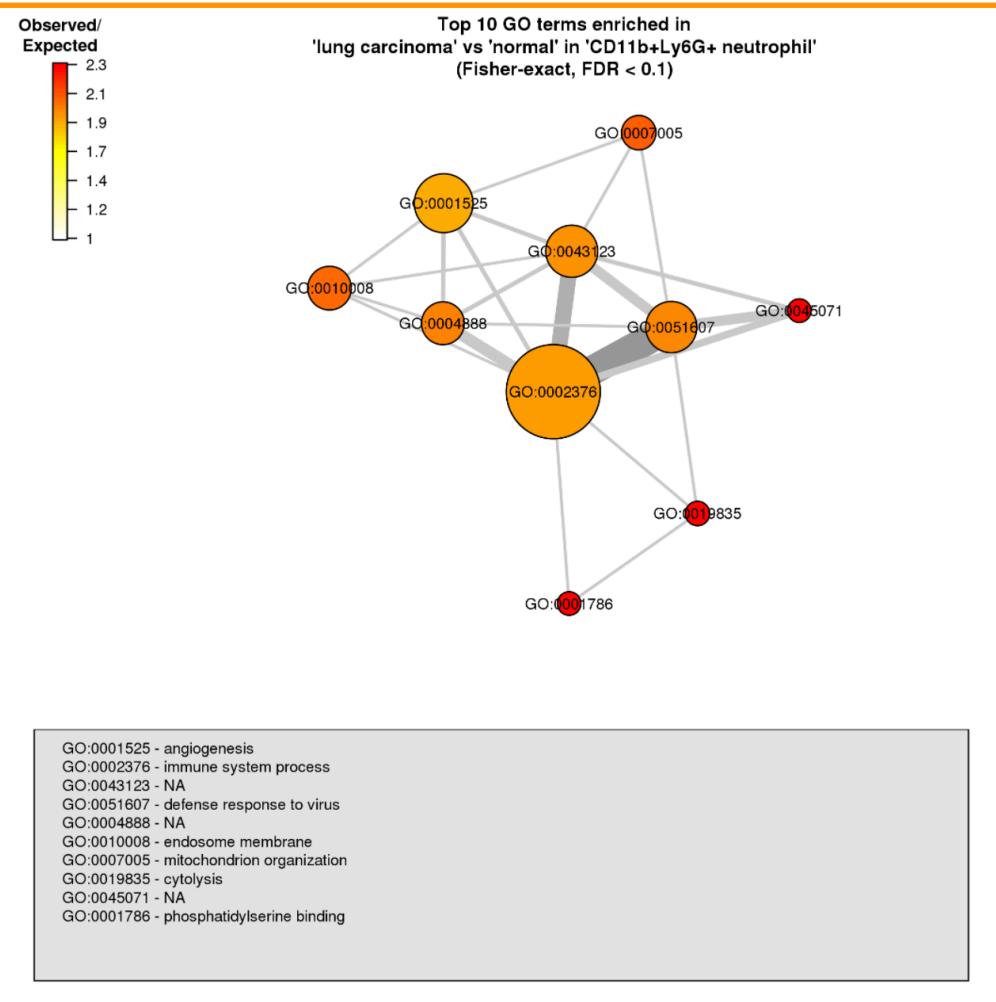
'lung carcinoma' vs 'normal' in 'lung cell'

GSE GO terms enrichment analysis

GSE Interpro domains enrichment analysis

GSE Reactome pathways enrichment analysis

MA MA plot for the contrast across all genes





Differential experiment page

Transcriptome analysis of isolated stromal cells and tumor epithelial cells in mouse

RNA-Seq mRNA differential

Organism: *Mus musculus*

Reference(s): [25704820](#) ([Filter by genes in paper](#))

Download gene quantification matrices
(TPM and FPKM) and R object

Results Plots Experiment Design Supplementary Information Downloads

All expression results in the experiment

All analytics for the experiment

All the raw counts for the experiment

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

Experiment Design (tsv)



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



Science People Resources Contact Blog Jobs Platform



Open Targets Platform

A partner
discovery
identification

Find new targets for drug discovery

CYP1A2



CYP1A2

cytochrome P450 family 1 subfamily A member 2

222 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...

Targets

CYP27B1 cytochrome P450 family 27 subfamily B member 1

CYP1A1 cytochrome P450 family 1 subfamily A member 1



Expression Atlas data in other resources

12 diseases associated with CYP1A2

[View CYP1A2 profile](#)

Filter by

Data type

selected ▾

- Genetic associations (28) ▾
- Somatic mutations (0)
- Drugs (0)
- Affected pathways (10) ▾
- RNA expression (12) ▾
 - Expression Atlas (12)
- Text mining (185) ▾
- Animal models (32) ▾

Therapeutic area

Table

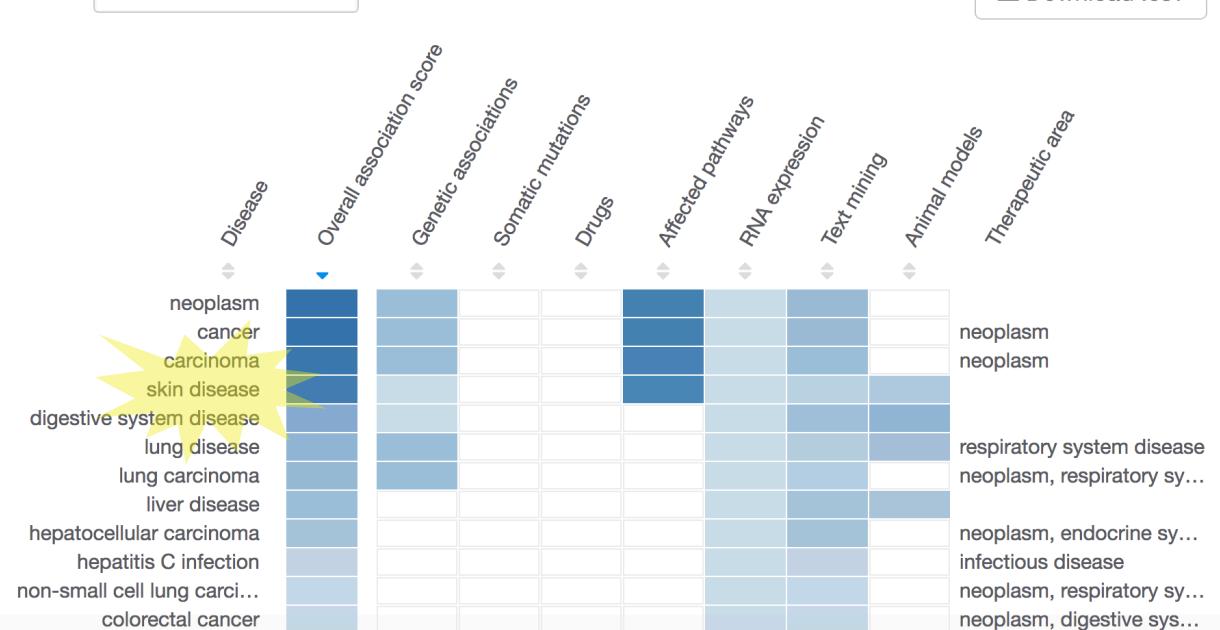
Bubbles

Tree

Showing 1 to 12 of 12 entries

Search:

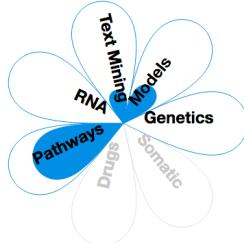
[Download .csv](#)





Expression Atlas data in other resources

Evidence for CYP1A2 in skin disease



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...

skin disease
Synonyms: SKIN AND SUBCUTANEOUS TISSUE DISORDERS, Cutaneous Disorder, Skin Disorder, Skin Diseases and Manifes...

Genetic associations

Somatic mutation

Drugs

Affected pathways

RNA expression

Text mining

Animal models



Expression Atlas data in other resources

RNA expression

Source: Expression Atlas

Showing 1 to 1 of 1 entries

Search:

[Download .csv](#)

Disease	Gene-disease evidence		Tissue/cell	Evidence source	\log_2 fold change	p-value	Percentile rank	Experiment overview and data	Publications
	Comparison	Activity							
psoriasis	'psoriasis; lesional skin' vs 'normal; normal'	unknown	skin	ECO_0000357	-4.7	5.47e-46	99	Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms	

Show entries

[Previous](#) [1](#) [Next](#)



Expression Atlas data in other resources

222 diseases associated with CYP1A2

View CYP1A2 profile

Filter by

Data type

- Genetic associations (28)
- Somatic mutations (0)
- Drugs (0)
- Affected pathways (10)
- RNA expression (12)
- Text mining (185)
- Animal models (32)

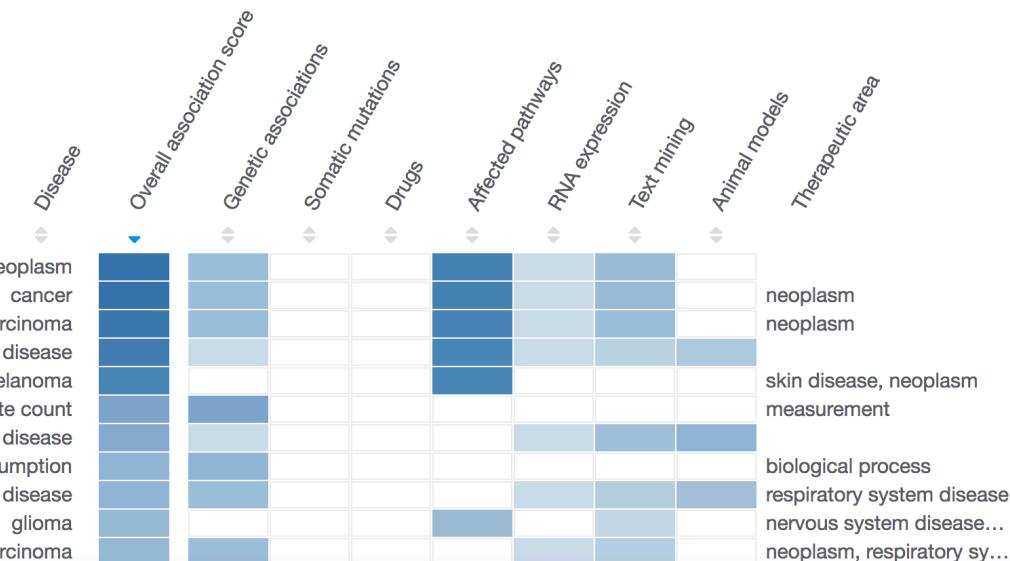
Therapeutic area

Table **Bubbles** Tree

Showing 1 to 50 of 222 entries

Search:

Download .csv





Expression Atlas data in other resources

CYP1A2

cytochrome P450 family 1 subfamily A member 2

| View associated diseases

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-hydroxylation. Caffeine is metabolized primarily by cytochrome CYP1A2 in the liver through an initial N3-demethylation. Also acts in the metabolism of aflatoxin B1 and acetaminophen. Participates in the bioactivation of carcinogenic aromatic and heterocyclic amines. Catalyzes the N-hydroxylation of heterocyclic amines and the O-deethylation of phenacetin.

(information provided by UniProt)

Synonyms: P3-450 | CP12 | 1.14.14.1 | Cytochrome P450-P3 | Cytochrome P450 1A2 | Cytochrome P450 4 | Cytochrome P(3)450 | CYP1A2 | Cholesterol 25-hydroxylase

Drugs

Protein Information

Pathways

Similar targets (based on diseases in common)

Variants, isoforms and genomic context

Protein interactions

RNA and protein baseline expression

Mouse phenotypes

Protein Structure

Gene Ontology

Gene tree

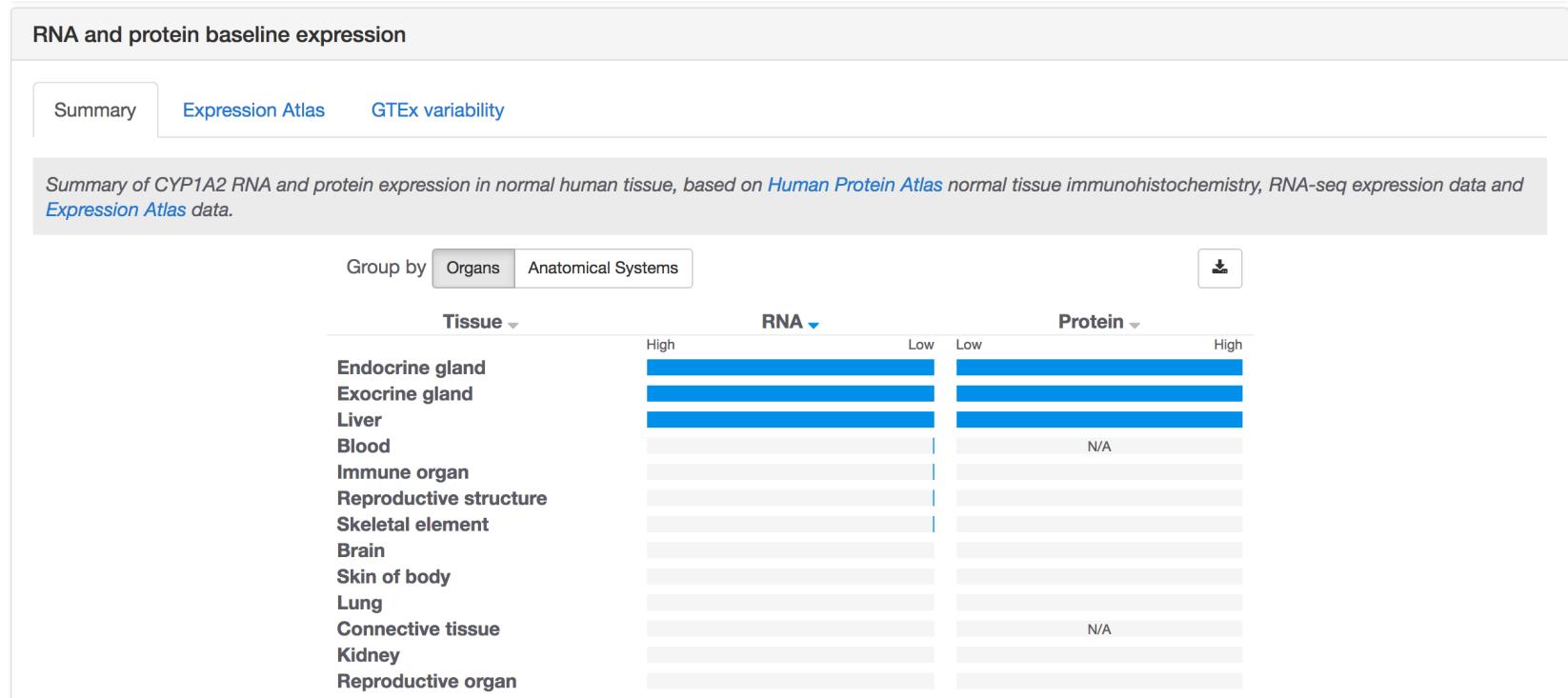
Bibliography

Cancer hallmarks

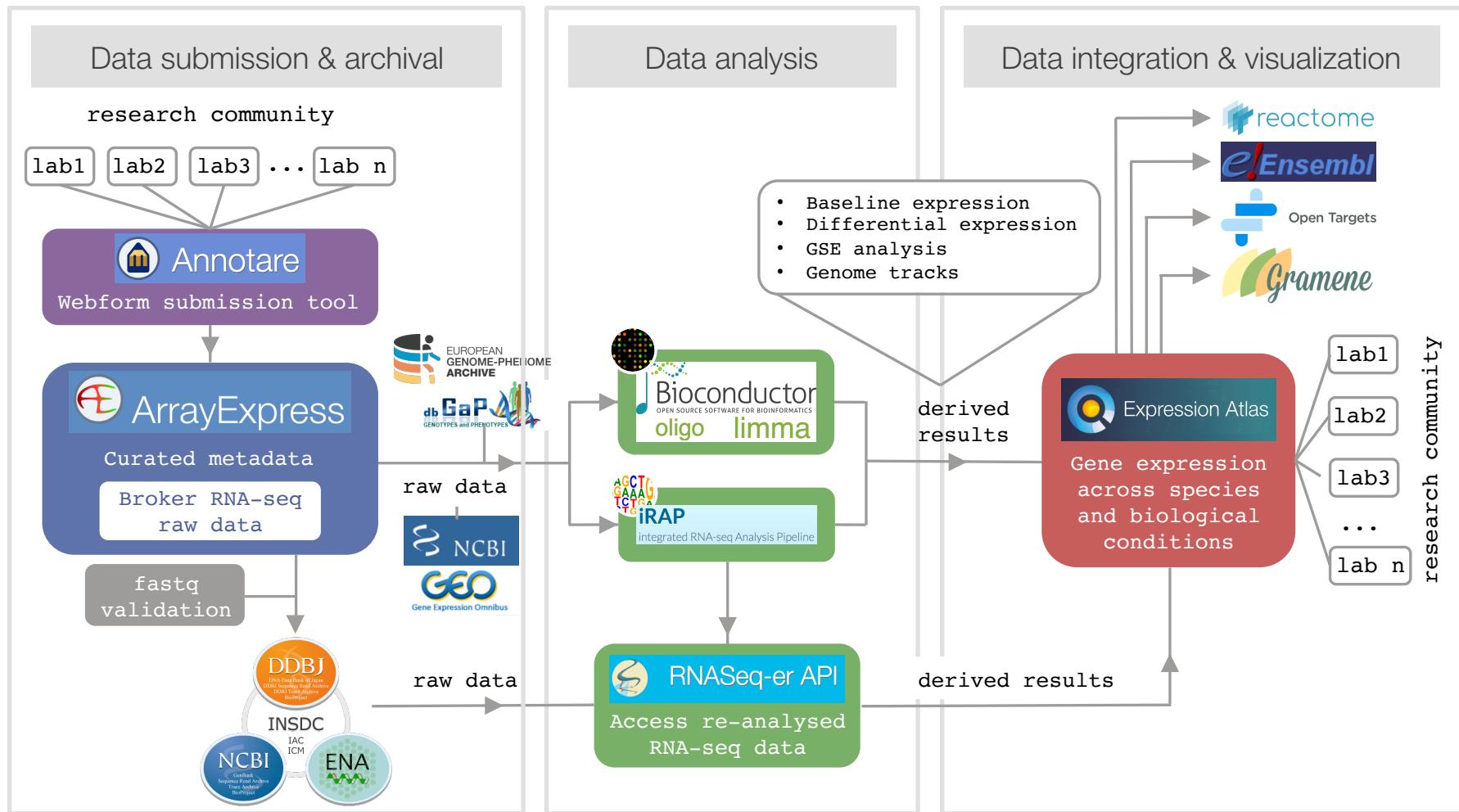
Cancer biomarkers



Expression Atlas data in other resources



Functional genomics resources at 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' (highlighted in yellow), 'Help', 'Licence', and 'About' (highlighted in yellow). The main content area has a teal background. A large heading 'About Expression Atlas' is centered above a section titled 'What is Expression Atlas?'. Below this is a paragraph of text. To the right of the text is a diagram consisting of four teal boxes in a 2x2 grid. Each box contains an icon and text: 'Data curation' (document with checkmark), 'Data analysis' (bar chart), 'Data search' (sunburst chart), and 'Data visualisation' (heat map). Below each box is a subtitle: 'High quality data', 'High quality results', 'Accessible results', and '... easy to interpret'.

Expression Atlas

Gene expression across species and biological conditions

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

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.

Data curation
High quality data

Data analysis
High quality results

Data search
Accessible results

Data visualisation
... easy to interpret

atlas-feedback@ebi.ac.uk



@ExpressionAtlas

EMBL-EBI workshop: Data and tools for transcriptomics and protein biology

Expression Atlas: gene expression results across species and conditions

Laura Huerta, PhD

Senior Scientific Curator

lauhuema@ebi.ac.uk

Malta, 31 July 2018