

EMBL-EBI: Bioinformatics resources for exploring disease related data

ArrayExpress and Expression Atlas services at EMBL-EBI

Anja Fullgrabe, PhD

Scientific Curator

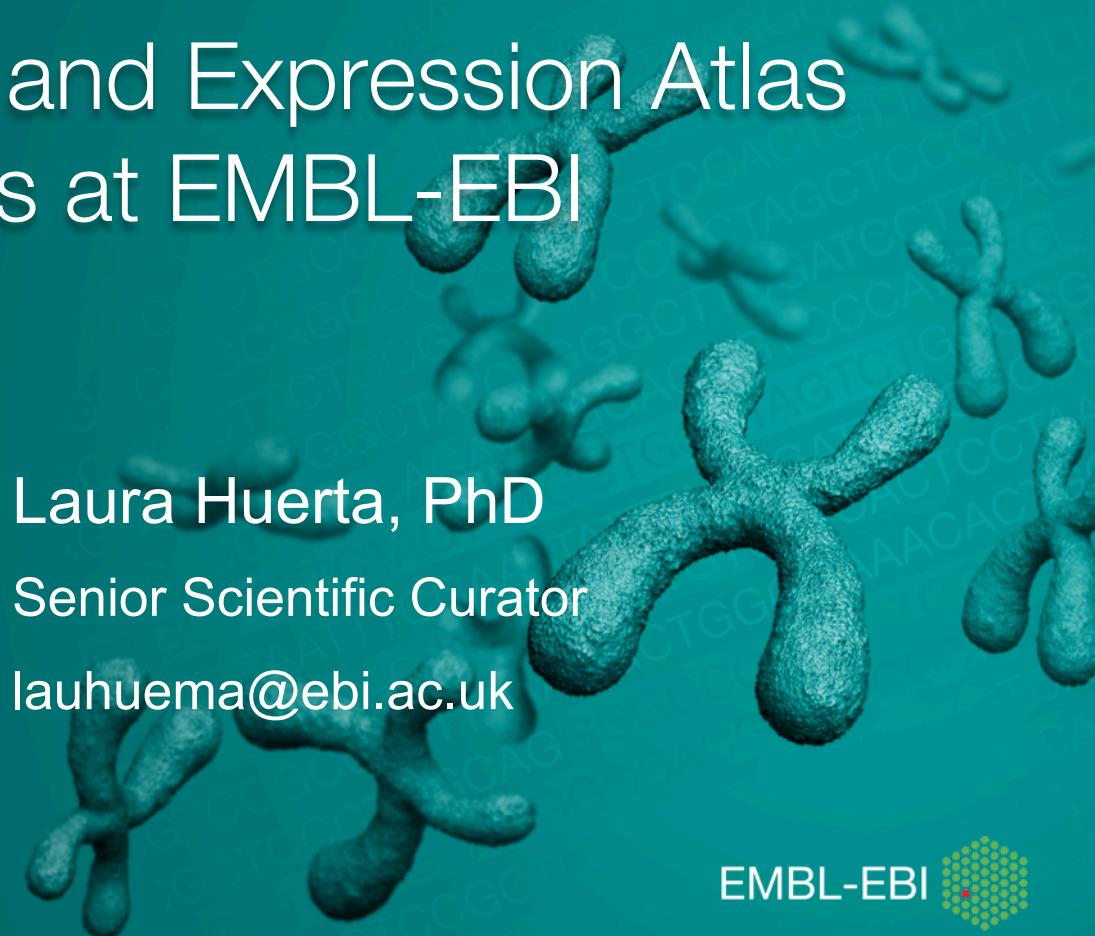
anjaf@ebi.ac.uk

30 October 2017

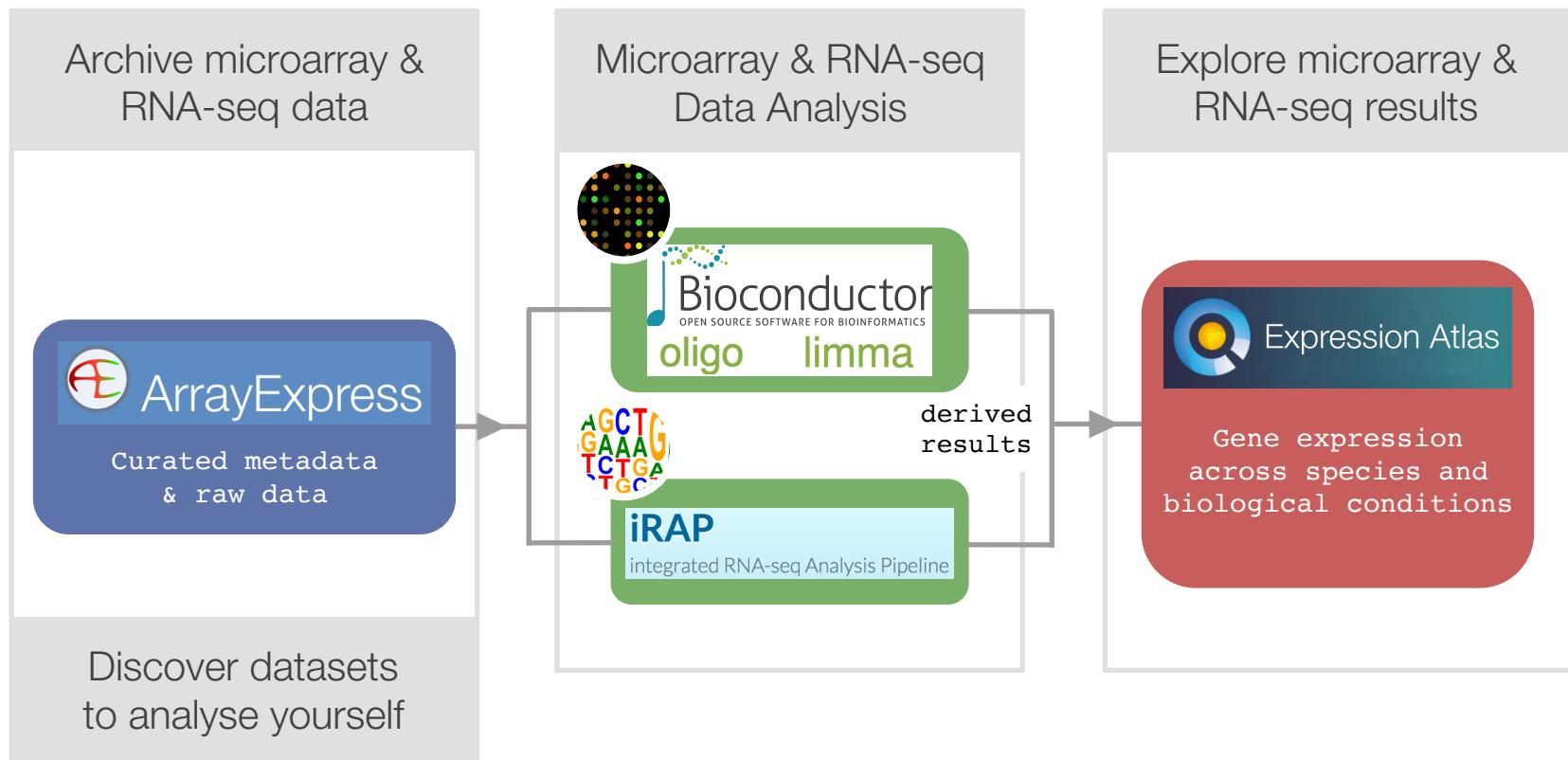
Laura Huerta, PhD

Senior Scientific Curator

lauhuema@ebi.ac.uk



Outline of the session



<https://www.ebi.ac.uk/~lauhuema/workshop/ExploringDiseaseData/>

What is functional genomics?

- Study function of genes and other parts of the genome
- Not whole genome sequencing for assembly into chromosomes
- Often involve microarrays or sequencing (NGS/HTS)
- Gene expression - where? how much? Changes? pathways?



vs

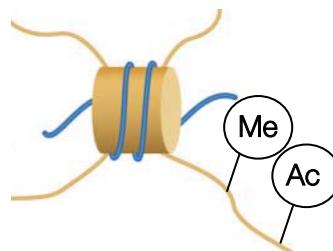
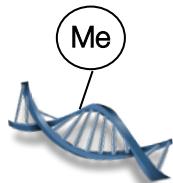


Human lung



Circadian cycle

- Gene/genome regulation - e.g. CpG/histone methylation, SNP typing



AAAT~~T~~CCCCGA
AAAT~~A~~ACCCGA

Experiment types in functional genomics

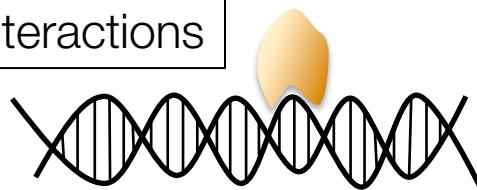
Transcriptomics



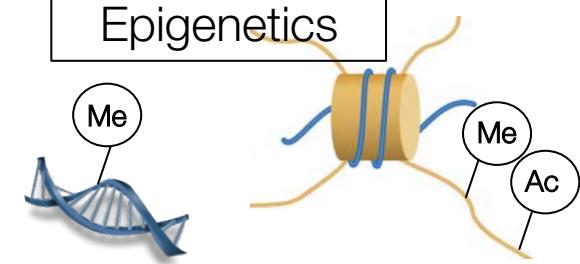
Genotyping

AAATT_{red}CCCGA
AAAT_{red}ACCCGA

DNA/RNA-protein interactions



Epigenetics



Example technologies:

RNA-based assays

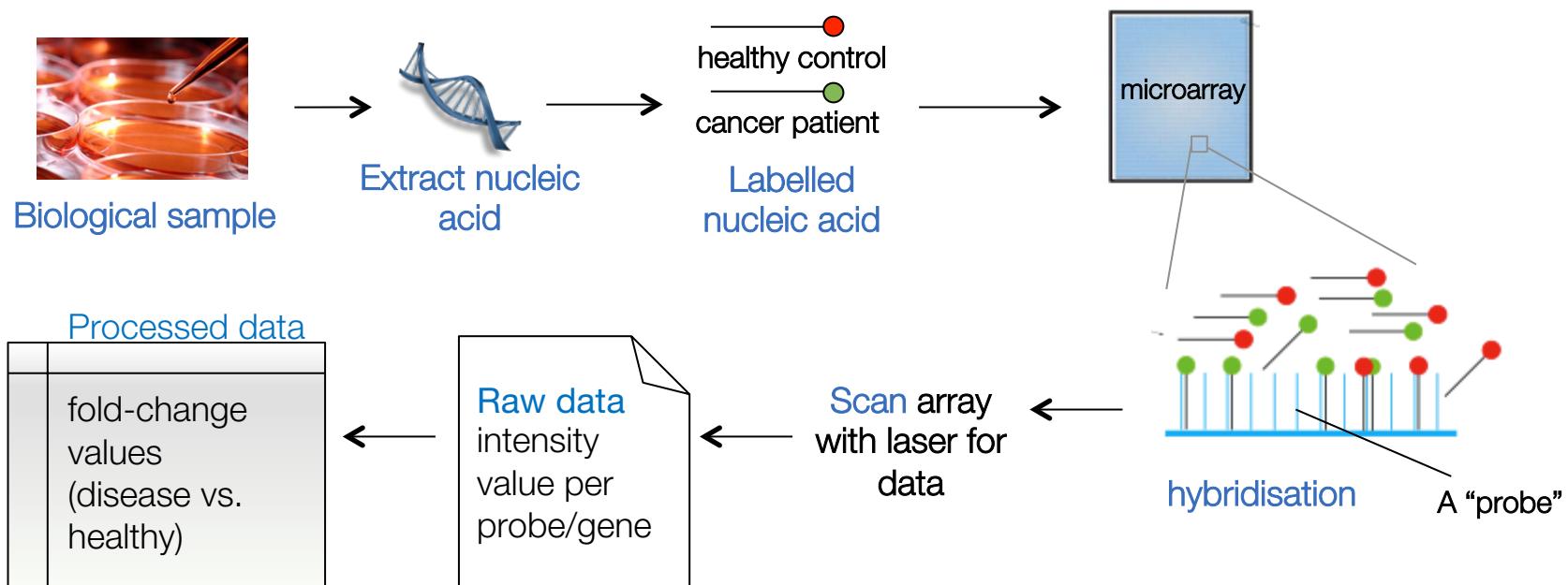
- Gene expression microarray
- RNA-sequencing
- microRNA profiling

DNA-based assays

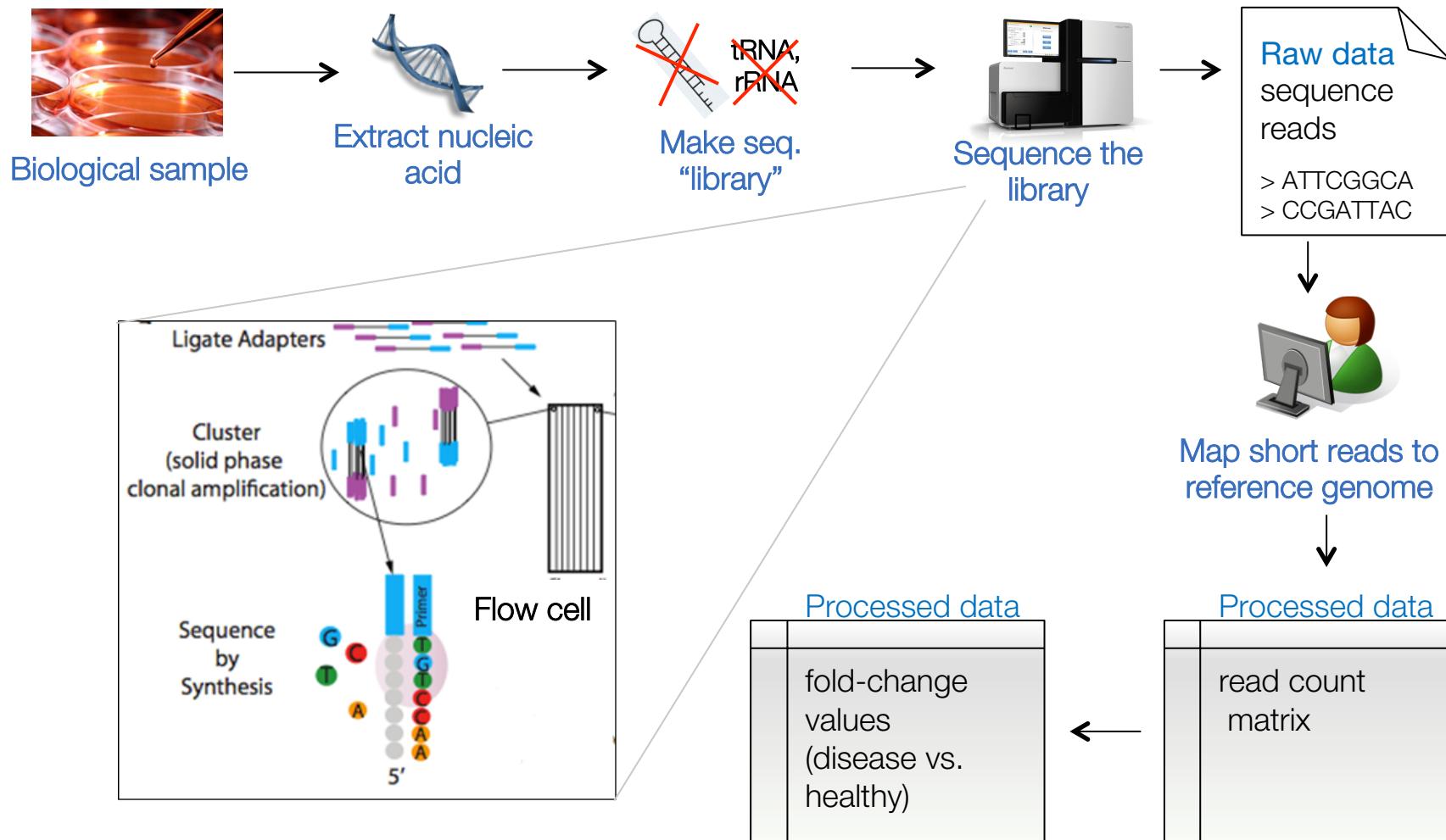
- Chromatin Immunoprecipitation (ChIP-seq or ChIP-on-chip)
- Comparative genomic hybridisation
- Methylation profiling

Microarray technology

- Microarray: single-stranded (usually DNA) short “probes” fixed on a solid surface (“chip”). One probe (<100bp) ~ one cDNA/trans./exon.
- Probe location on the chip and probe sequence must be known at the time of manufacturing



High-throughput sequencing technology



ArrayExpress - an archive for functional genomics data

<https://www.ebi.ac.uk/arrayexpress>

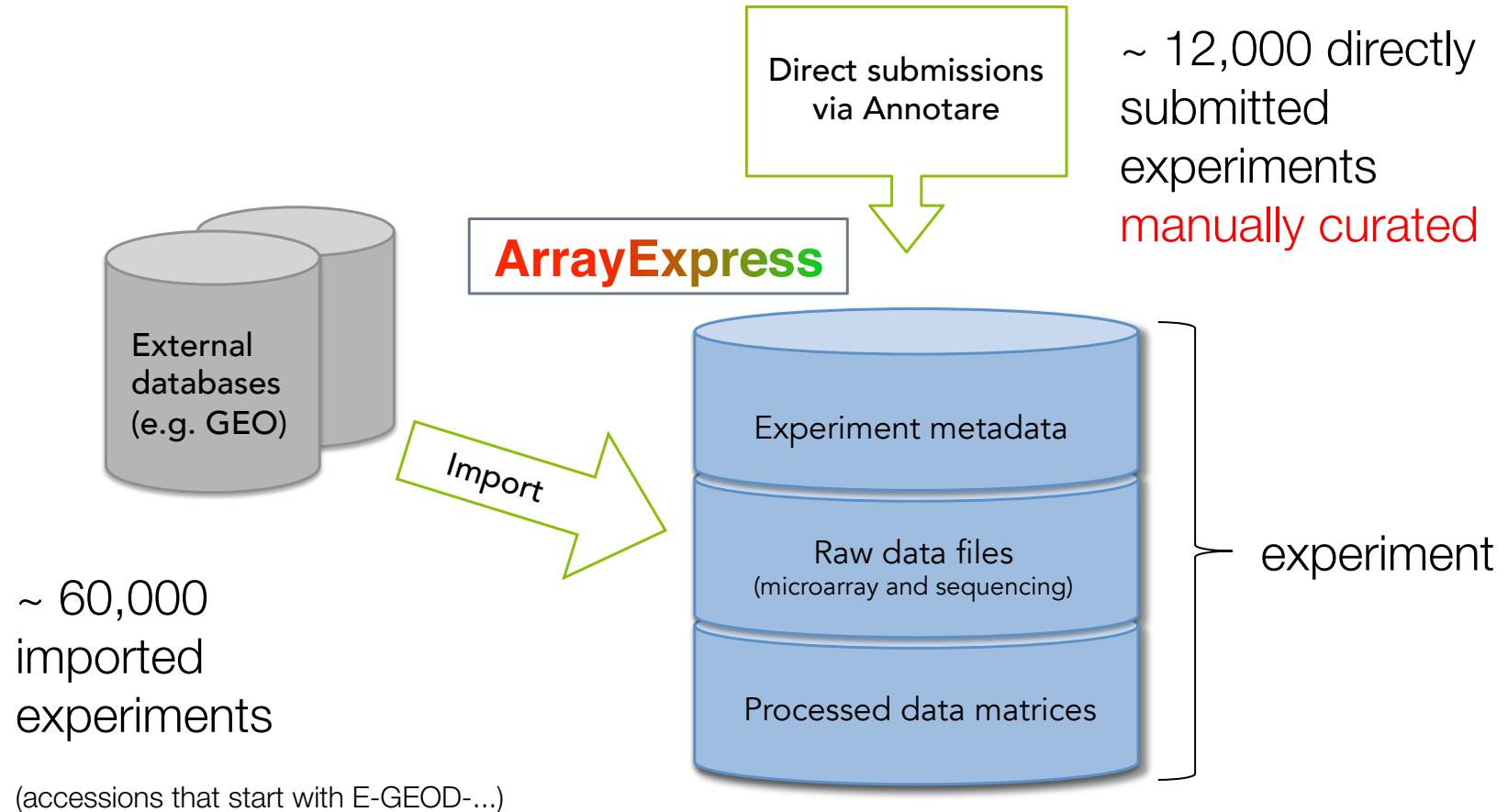
Started in 2000
“One-stop shop for microarray data”

A screenshot of a Nature journal article. The title is "One-stop shop for microarray data". It features a red header with the Nature logo and "International weekly journal of science". Below the header, there's a sidebar with links like "Journal content", "Commentary", and "Archive". The main text discusses the challenges of managing large genomic datasets and the creation of a public database.

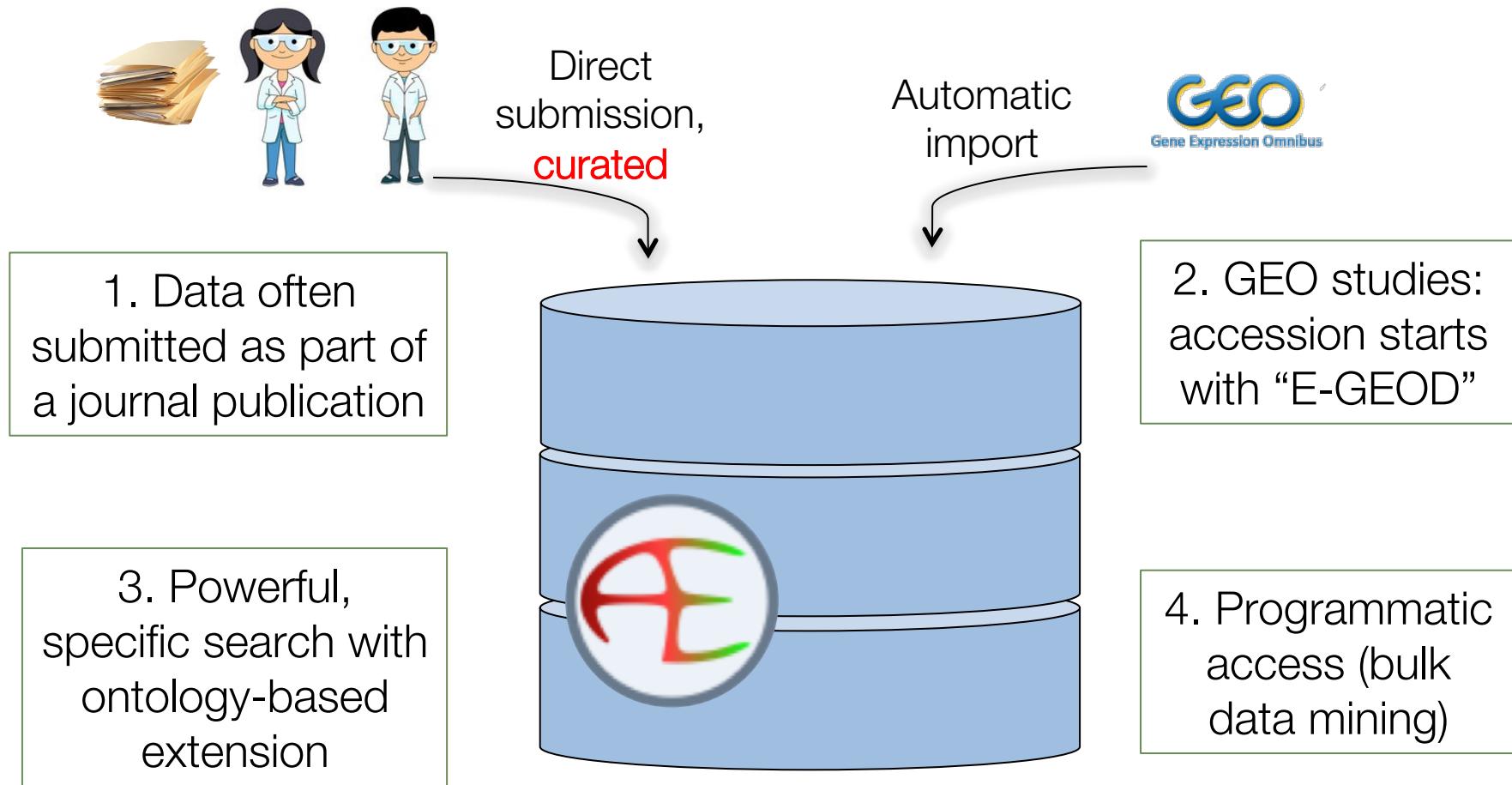
The homepage of the ArrayExpress website. It has a blue header with the ArrayExpress logo, search bar, and navigation links for Home, Browse, Submit, Help, and About ArrayExpress. The main content area features a section titled "ArrayExpress – functional genomics data" with a brief description and a "Browse ArrayExpress" button. To the right, there's a "Data Content" section showing statistics: 70453 experiments, 2230060 assays, and 45.89 TB of archived data. Below that is a "Latest News" section with a recent update about stopping imports from GEO. The bottom of the page is divided into several sections: "Links", "Tools and Access", and "Related Projects".

ArrayExpress – Data sources

In total over 70,000 experiments



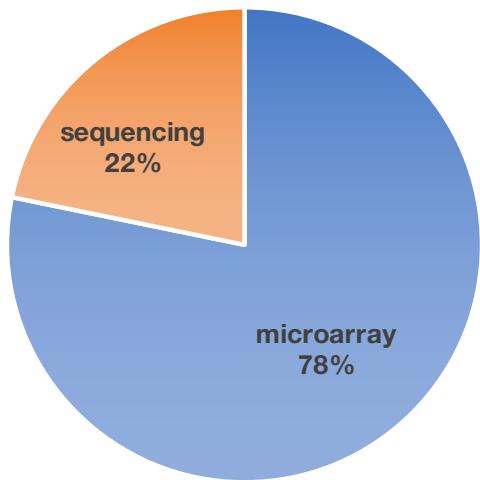
ArrayExpress – Key features



Main use: discover data sets to analyse yourself

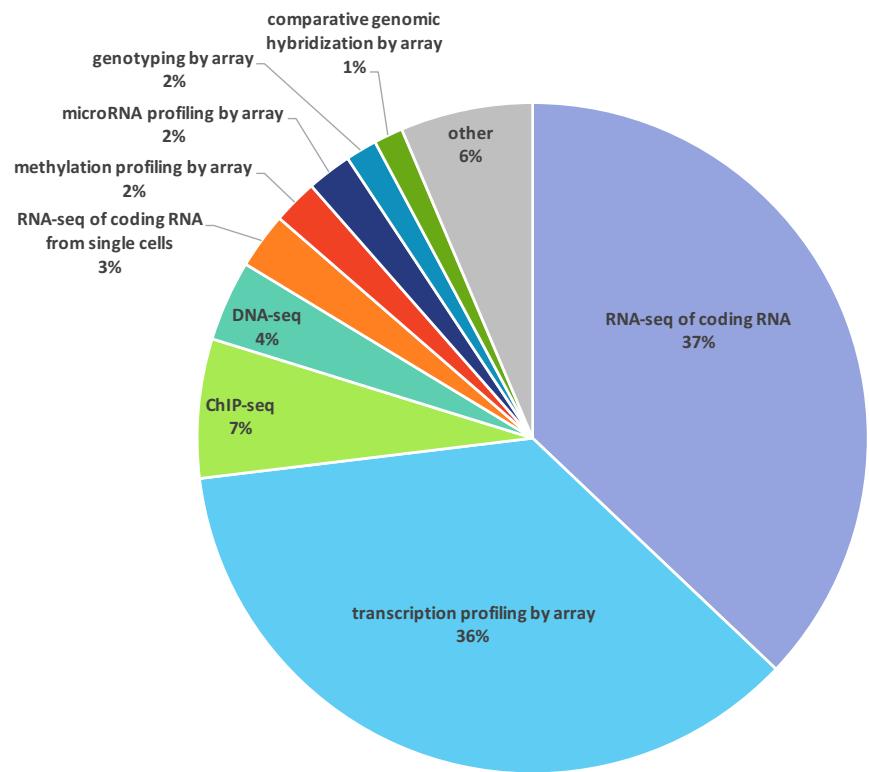
Microarray vs. sequencing

All experiments currently in ArrayExpress



Direct submissions in the last 1 year by experiment type

now ~ 50% sequencing



Browse ArrayExpress database

Type keywords and accessions here

Search Advanced

Feedback Login

Home Browse Submit Help About ArrayExpress

Filter search results

Filter options

Sortable headings

Page size 25 50 100 250 500

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-2480	Molecular profiling of Arabidopsis leaf growth in LD	transcription profiling by tiling array	Arabidopsis thaliana	25	Today			-	-
E-MTAB-2376	Transcription profiling by tiling array of Arabidopsis leaves from wild-type plants and CAF-1 mutants fas2-4 to study the loss of CAF-1 on the transcriptome of non-dividing cells	transcription profiling by tiling array	Arabidopsis thaliana	6	Today			7	-
E-MTAB-3976	Transcription profiling by array of mouse retinas from rodless, coneless, or rodless/coneless mice compared with wild type controls	transcription profiling by array	Mus musculus	16	Yesterday	-		4	-
E-MTAB-3960	SNP genotyping by array of four head and neck cancer cell lines in pre-clonal cells and clonal sub-lines to Identify biologically relevant alterations in fewer samples	genotyping by array	Homo sapiens	7	Yesterday	-		4	-
E-MTAB-3920	Transcription profiling of transgenic barley plants overexpressing the HvNAC005 gene	transcription profiling by array	Hordeum vulgare	16	Yesterday			3	-
E-MTAB-3808	Pancreatic ductal adenocarcinoma primary cultures enriched or not in cancer stem cells	RNA-seq of coding RNA	Homo sapiens	20	Yesterday	-		3	-
E-MTAB-3481	Analysis of whole-genome chromosomal alterations in primary cutaneous melanoma	comparative genomic hybridization by array	Homo sapiens	98	Yesterday	-		5	-
E-MTAB-173	Intestinal mucosa of Apc Min/+ mice fed with maslinic acid (MA) to study the underlying mechanisms by which MA inhibits intestinal tumorigenesis	trans prof trans prof	Mus musculus	10	31/10/2015	-		5	-

Accession number: Click to view experiment details

Experiment type and platform

Organism of the source material

Number of hybridisations/libraries

Experiment description

ARRAYEXPRESS / BROWSE / E-MTAB-5094

E-MTAB-5094 - Whole blood transcriptional profiling of healthy volunteers and rheumatoid arthritis patients treated with methotrexate or anti-TNF therapy after ex vivo whole blood stimulation with TNF

Samples and files →

Samples (32) [Click for detailed sample information and links to data](#)

Arrays (2) [A-GEOD-20844 - Agilent-072363 SurePrint G3 Human GE v3 8x60K Microarray 039494 \[Feature Number Version\]](#)
[A-MEXP-2104 - Agilent 028004 SurePrint G3 Human GE 8x60K Microarray \(dessen\)](#)

Protocols (7) [Click for detailed protocol information](#)

Description We tested the hypothesis that anti-tumor necrosis factor (TNF) therapy reduces TNF-inducible gene expression in blood. Whole peripheral blood from healthy volunteers and rheumatoid arthritis patients (treated with the monoclonal anti-TNF antibody adalimumab, the soluble TNF receptor etanercept or the standard therapy methotrexate) was incubated at 37 °C for three hours in the presence or absence of 10 ng/ml recombinant human TNF. Blood samples were then processed for RNA isolation and transcriptional profiling by gene microarray. This submission includes two human whole genome Agilent Array Designs: A-MEXP-2104 and A-GEOD-20844. Each of the individual raw array files are included as well as a single processed file representing the data matrix of all the merged and normalised data for the probes that are shared by the two array designs.

Description →

Experiment types transcription profiling by array, stimulus or stress design

Contact [Carolin Turner <carolin.turner@ucl.ac.uk>](mailto:Carolin.Turner@ucl.ac.uk)

Citation [Tumor Necrosis Factor \(TNF\) Bioactivity at the Site of an Acute Cell-Mediated Immune Response Is Preserved in Rheumatoid Arthritis Patients Responding to Anti-TNF Therapy](#), Rachel Byng-Maddick, Carolin T Turner, Gabriele Pollara, Matthew Ellis, Naomi J Guppy, Lucy C K Bell, Michael R Ehrenstein, Mahdad Noursadeghi. *Frontiers in Immunology* 8(932) (2017)

MIAME — * * * * *

Platforms Protocols Variables Processed Raw

Files

Investigation description	E-MTAB-5094.idf.txt
Sample and data relationship	E-MTAB-5094.sdrf.txt
Raw data (1)	E-MTAB-5094.raw.1.zip
Processed data (1)	E-MTAB-5094.processed.1.zip
Array designs	A-GEOD-20844.adf.txt, A-MEXP-2104.adf.txt

[Click to browse all available files](#)

Links to data →

The samples table

ARRAYEXPRESS / BROWSE / E-MTAB-5094 / SAMPLES AND DATA

E-MTAB-5094 - Whole blood transcriptional profiling of healthy volunteers and rheumatoid arthritis patients treated with methotrexate or anti-TNF therapy after ex vivo whole blood stimulation with TNF

Sample Attributes					Variables			Links to Data	
Source Name	disease	clinical history	individual	organism part	disease	clinical history	stimulus	Raw	Processed
Sample 1_Cy3	normal	healthy volunteer	1	blood	normal	healthy volunteer	none	Download	Download
Sample 3_Cy5	normal	healthy volunteer	3	blood	normal	healthy volunteer	none	Download	Download
Sample 5_Cy3	normal	healthy volunteer	1	blood	normal	healthy volunteer	tumor necrosis factor	Download	Download
Sample 7_Cy5	normal	healthy volunteer	3	blood	normal	healthy volunteer	tumor necrosis factor	Download	Download
Sample 2_Cy3	normal	healthy volunteer	2	blood	normal	healthy volunteer	none	Download	Download
Sample 4_Cy5	normal	healthy volunteer	4	blood	normal	healthy volunteer	none	Download	Download
Sample 6_Cy3	normal	healthy volunteer	2	blood	normal	healthy volunteer	tumor necrosis factor	Download	Download
Sample 8_Cy5	normal	healthy volunteer	4	blood	normal	healthy volunteer	tumor necrosis factor	Download	Download
Sample 18_Cy3	rheumatoid arthritis	adalimumab therapy	10	blood	rheumatoid arthritis	adalimumab therapy	none	Download	Download
Sample 19_Cy5	rheumatoid arthritis	adalimumab therapy	11	blood	rheumatoid arthritis	adalimumab therapy	none	Download	Download
Sample 10_Cy5	rheumatoid arthritis	methotrexate therapy	6	blood	rheumatoid arthritis	methotrexate therapy	none	Download	Download
Sample 12_Cy3	rheumatoid arthritis	methotrexate therapy	8	blood	rheumatoid arthritis	methotrexate therapy	none	Download	Download
Sample 14_Cy5	rheumatoid arthritis	methotrexate therapy	6	blood	rheumatoid arthritis	methotrexate therapy	tumor necrosis factor	Download	Download
Sample 16_Cy3	rheumatoid arthritis	methotrexate therapy	8	blood	rheumatoid arthritis	methotrexate therapy	tumor necrosis factor	Download	Download
Sample 29_Cy3	rheumatoid arthritis	etanercept therapy	13	blood	rheumatoid arthritis	etanercept therapy	tumor necrosis factor	Download	Download
Sample 32_Cy5	rheumatoid arthritis	etanercept therapy	16	blood	rheumatoid arthritis	etanercept therapy	none	Download	Download
Sample 27_Cy3	rheumatoid arthritis	etanercept therapy	15	blood	rheumatoid arthritis	etanercept therapy	none	Download	Download

1. Unique sample name for each biological replicate

2. Sample attributes describing the source material

3. Experimental variable(s): the condition that is studied

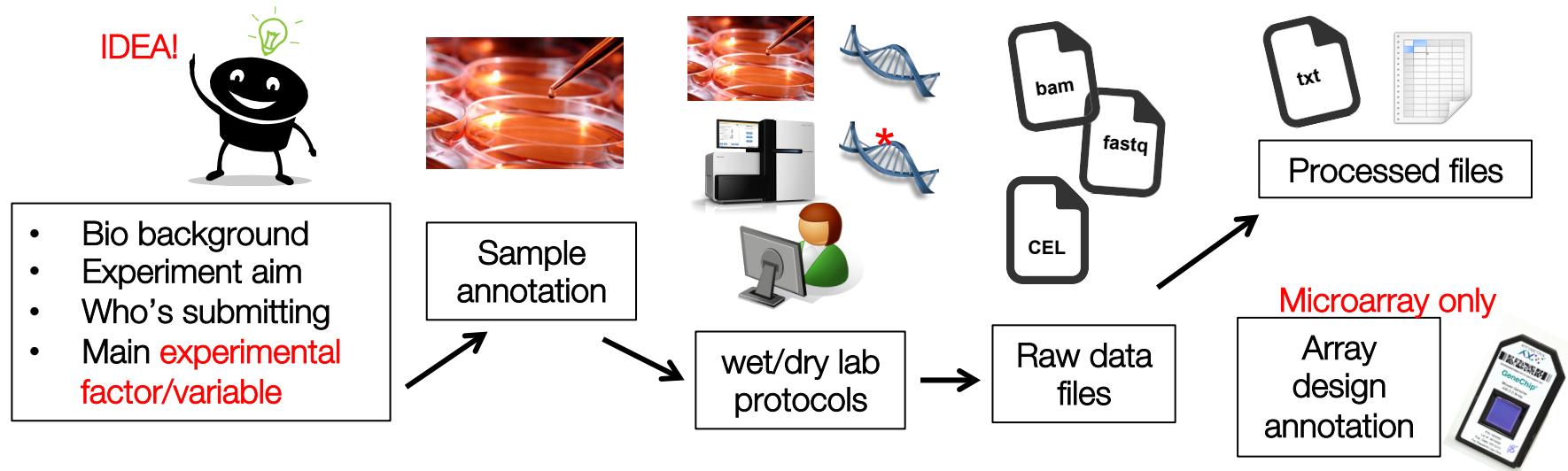
4. Links to download individual data files

Information captured per experiment

We apply community standards, when curating direct submissions

MIAME: Minimal Information about a Microarray Experiment

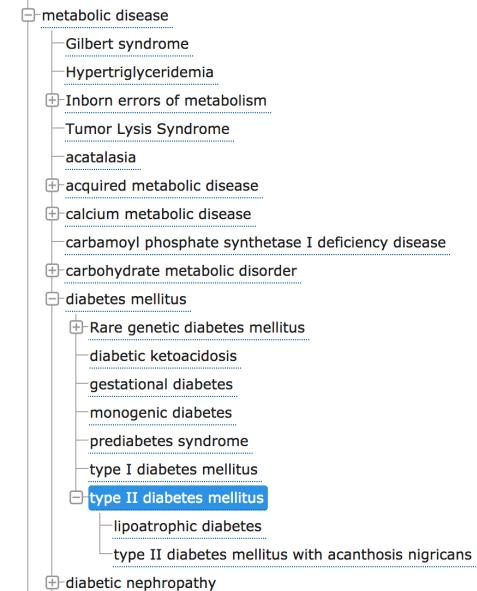
MINSEQE: Minimum Information about a high-throughput SEQuencing Experiment



Annotation with ontology terms

Sample description:
“blood collected from a
T2DM patient (54y, m)
after metformin treatment”

Ontology



Attribute	Disease	Age (year)	Sex	Organism Part	Compound
Value (usually ontology term)	type II diabetes mellitus	54	male	blood	metformin

The experimental variable

Example 1

Source	Organism	Disease	Sex	Cell type
Test Sample	Homo sapiens	diabetes mellitus	male	adipocyte
Control	Homo sapiens	normal	female	adipocyte

“What is the subject of the study?”

“What differs between test and control samples?”

Example 2

Source	Organism	Ecotype	Organism part	Compound
Test Sample 1	Arabidopsis	Col-0	leaf	sodium chloride
Test Sample 2	Arabidopsis	Col-0	root	sodium chloride
Control 1	Arabidopsis	Col-0	leaf	water
Control 2	Arabidopsis	Col-0	root	water

How to search ArrayExpress

Only RNA + sequencing
Only human data

Must be array
"HG U133 Plus 2"

Diabetes-related experiments

Methylation profiling experiments

Filter options

The screenshot shows the ArrayExpress homepage with a navigation bar at the top: Home, Browse (selected), Submit, Help, About ArrayExpress. Below the navigation is a sidebar titled "Filter search results". It includes dropdown menus for "By organism" (All organisms), "By experiment type" (All assays by molecule, All technologies), and "By array" (All arrays). There is also a checkbox for "ArrayExpress data only" and a "Reset filters" button. A "Filter" button is located at the bottom right of the sidebar. The main content area displays several experiment cards, such as "E-MTAB-3960" and "SNP genotyping by array of four head and neck cancer cell lines to Identify biologically relevant clonal sub-lines to Identify biologically relevant". A callout box points to the "Reset filters" button with the text "Find only curated direct submissions".

Free-text search

The screenshot shows the ArrayExpress search interface. In the top search bar, the word "diabet" is typed, and a dropdown menu shows suggestions: "diabetes", "diabetes mellitus", "gestational diabetes", "type I diabetes mellitus", "type II diabetes mellitus", "diabetes mellitus biomarker", "diabetes mellitus type 2 associated cataract", "diabetic", "diabetic nephropathy", "diabetic retinopathy", "diabetics", and "diabetogenic". To the right of the search bar are "Search" and "Advanced" buttons, along with "Feedback" and "Login" links. Below the search bar, there are sections for "experiments" and "Organism". The "experiments" section lists "Arabidopsis thaliana" with 25 results today. The "Organism" section lists "Arabidopsis thaliana" with 6 results today. A callout box points to the search suggestions with the text "EFO expansion".

Quick exercise – Search experiments



Find RNA expression arrays from human subjects with diabetes

- Filter box for: Homo sapiens; RNA assay; array assay
- Free text search for diabetes (more specific *evv:diabetes*)

E-TABM-987	Transcription profiling by array of mouse sciatic nerve endoneurium and dorsal root ganglia isolated from pre-symptomatic and early symptomatic diabetes mellitus type 1 (DM1) model animals	transcription profiling by array
E-GEOD-20067	Genome wide DNA methylation profiling of diabetic nephropathy in type 1 diabetes mellitus	methylation profiling by array
E-GEOD-21321	Blood microRNA profiles and upregulation of hsa-miR-144 in males with type 2 diabetes mellitus	<div style="background-color: yellow; padding: 5px;">Exact match to search term</div> <div style="background-color: lightgreen; padding: 5px;">Matched EFO synonyms to search term</div> <div style="background-color: orange; padding: 5px;">Matched EFO child term of search term</div>

Advanced search

Restrict search to certain fields:

“Experimental variable value should be *normal*”

evv:normal

Examples: E-MEXP-31, cancer, p53, Geuvadis

advanced search

Find the whole list of search fields and terms

Set conditions:

“attribute *disease* needs to be present” and

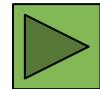
“only experiments with *processed data*”

sac:disease AND processed:true

Examples: E-MEXP-31, cancer, p53, Geuvadis

advanced search

Quick exercise - Advanced search



Find experiments that compare healthy ("normal") and rheumatoid arthritis patients.

evv:"rheumatoid arthritis" AND evv:normal

evv:"rheumatoid arthritis" AND (evv:normal OR evv:healthy)

Search results for evv:diabetes AND (evv:healthy OR evv:normal)

Showing 1 - 25 of 63 experiments

Welcome to ArrayExpress Help

How can we help you?



How to Search



Online Tutorials

Train online

Training | Train online | About Train online | Glossary | Support and Feedback | Log in/Register

ArrayExpress: Discover functional genomics data quickly and easily

Discover functional genomics data quickly and easily

ArrayExpress is a database of functional genomics data. This course will give you an overview of how these data are stored in ArrayExpress and will teach you how to effectively search and retrieve data from the ArrayExpress website.

Course published April 2016.

A basic understanding of microarray technology and its applications is beneficial. We recommend that you complete "Functional genomics: an introduction to EMBL-EBI resources" before taking this course.

Summary
Quick ArrayExpress Quiz

Train Online
ArrayExpress and
Annotare courses

Annotare
web-based
submission tool

New Experiment Submission

Select Experiment Design(s) binding site identification design case control design cell type comparison design compound treatment design development or differentiation design drug response design gene expression design gene perturbation design growth factor response design high throughput screening design proteome design sample collection design

Provide Experiment Details

How many samples does your experiment have?

What label did you use?

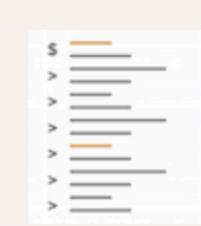
What array did you use?

If you cannot find your array design, try searching [here](#) or submit your array design to ArrayExpress.

OK



Submit your Data

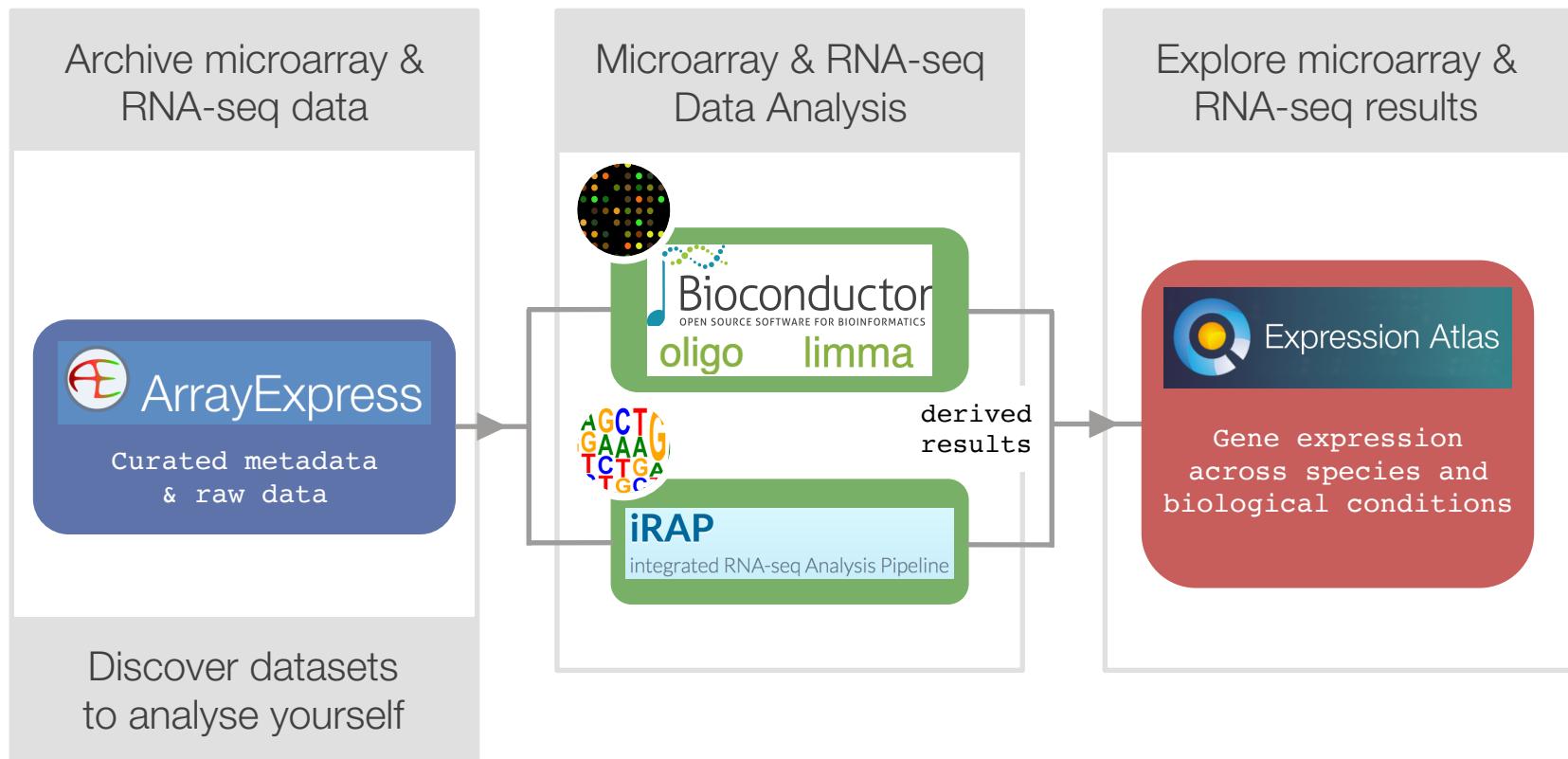


Programmatic Access

```
<?xml version='1.0' encoding='UTF-8'?>
<experiment>
  <@id>330451</@id>
  <@accession>E-MEXP-3429</@accession>
  <@name>GABA<sup>-</sup> receptor profiling by array of siRNA silencing of GABA<sub>A</sub>R in MCF10A cells</@name>
  <@lastUpdated>2015-10-20</@lastUpdated>
  <organism>Homo sapiens</organism>
  <@description>GABA<sub>A</sub> receptor profiling by array</@description>
  <experimentDesign>cellular modification/experimentDesign</experimentDesign>
  <@type>experimental</@type>
  <@status>published</@status>
  <@version>1</@version>
  <@title>GABA<sub>A</sub> receptor profiling by array of siRNA silencing of GABA<sub>A</sub>R in MCF10A cells</title>
  <@description>The ETS transcription factors EKLF and GATA control cell migration in breast epithelial cells through targeting a cohort of genes, independently from another family member STAT5, and thereby achieves biological specificity. Here, both direct and indirect target genes for GABA<sub>A</sub>R are identified by siRNA silencing of GABA<sub>A</sub>R. Therefore, although EKLF and GATA ultimately control cell migration, they do so by regulating distinct cohorts of target genes associated with actin cytoskeletal reorganization and cell migration control.</description>
  <@url>http://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-3429</url>
  <@doi>10.1101/234424</doi>
  <@category>cell line</category>
  <@category>characteristic</category>
  <@category>cell line</category>
  <@category>characteristic</category>
</experiment>
```

REST/JSON API
Retrieve metadata
and links to data

Outline of the session





What is Expression Atlas?

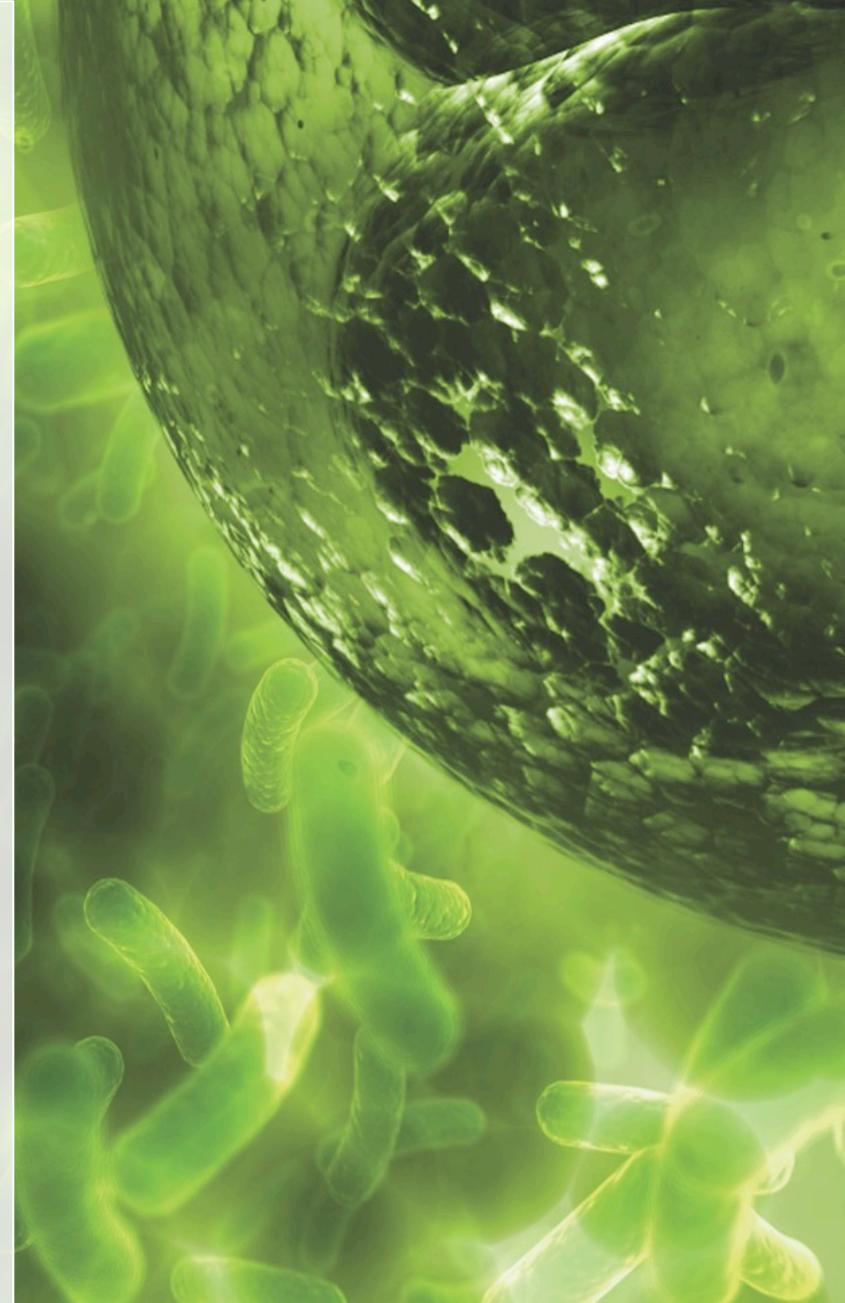


Reference resource for gene/protein expression
across species and biological conditions

www.ebi.ac.uk/gxa/home

OUR MISSION

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





What is Expression Atlas?

Big Data



Generate and
publish gene
expression data



GEO
Gene Expression Omnibus

ENA



EUROPEAN
GENOME-PHENOME
ARCHIVE

dbGaP
GENOTYPES and PHENOTYPES

'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?



<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



Genentech

Cancer research



Proteomics



Zebrafish development



Prenatal human brain



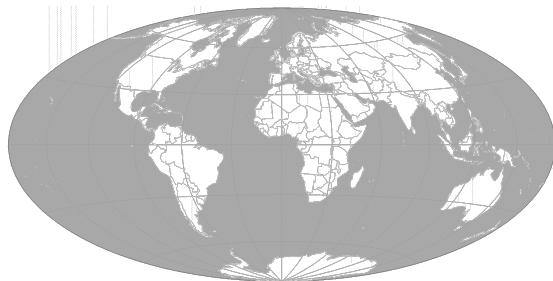
DMDD

Deciphering the Mechanisms
of Developmental Disorders



Mouse models

Expression Atlas contains thousands of selected and curated datasets



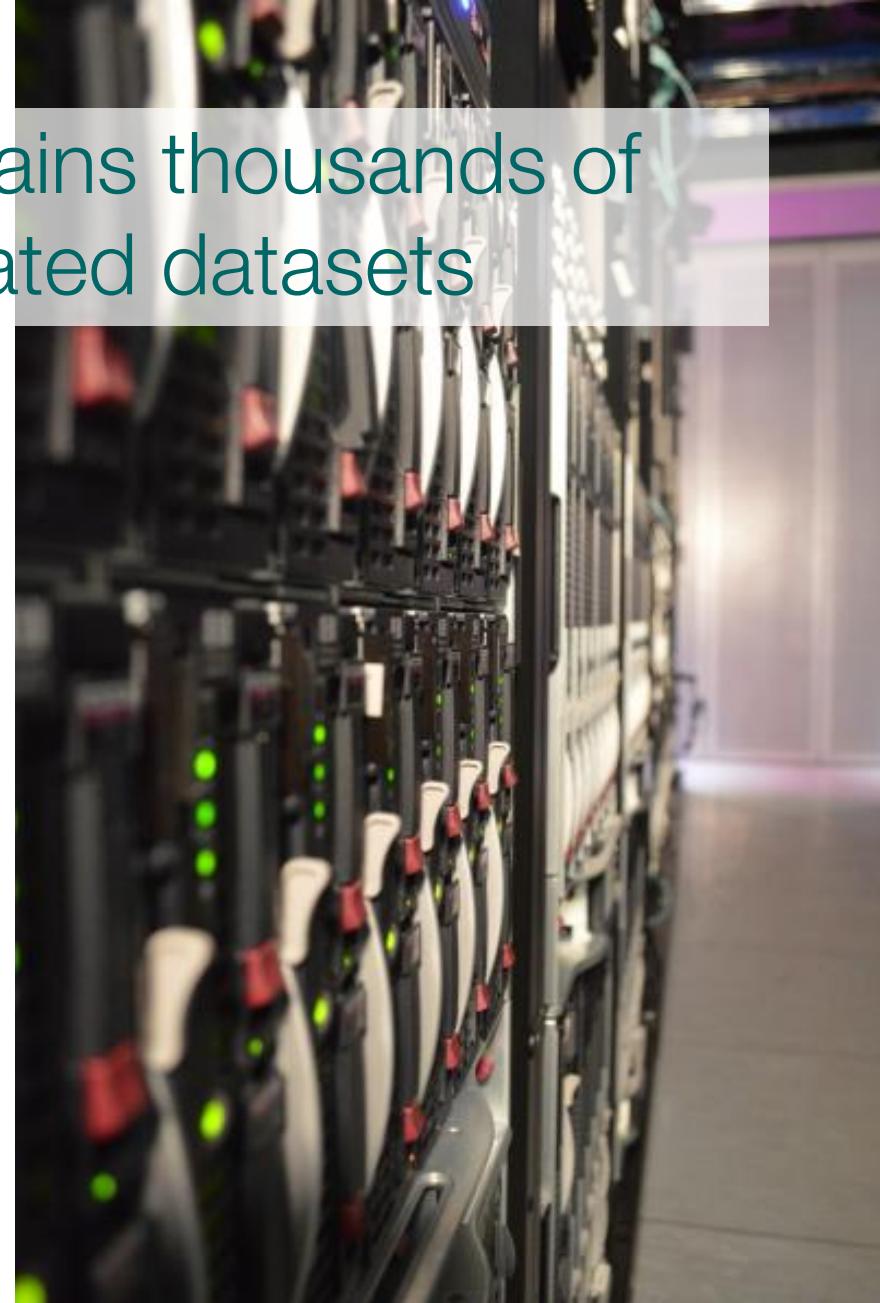
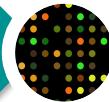
> 3,100 datasets



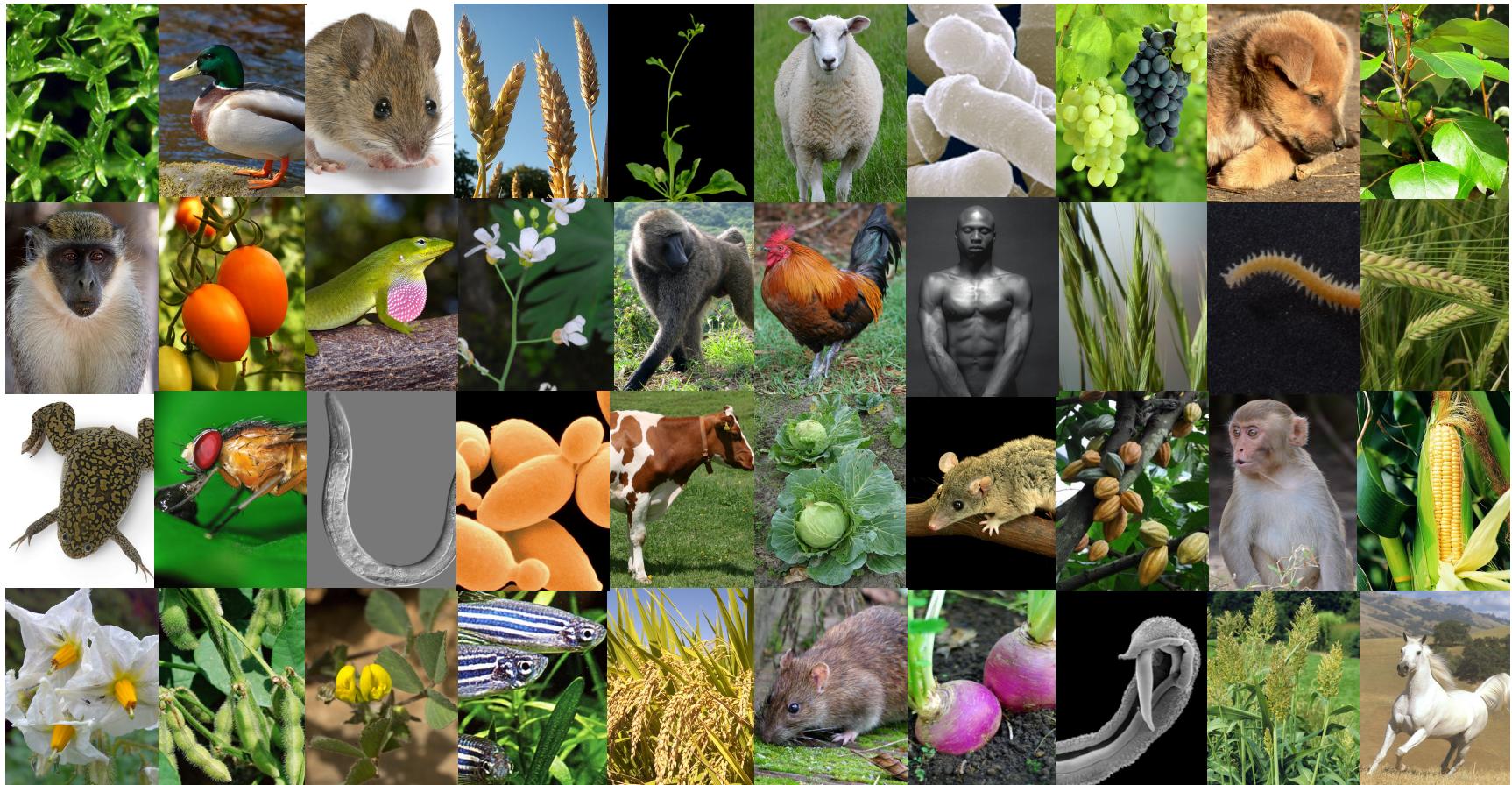
> 550 RNA-sequencing data



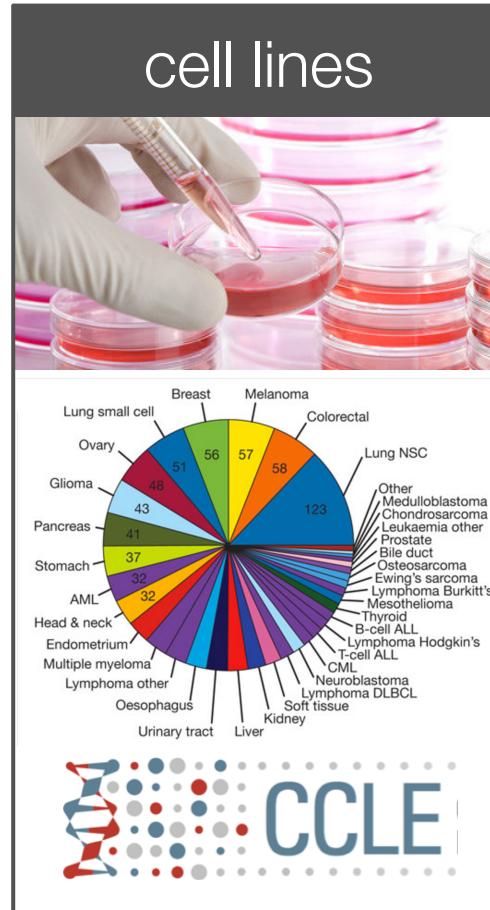
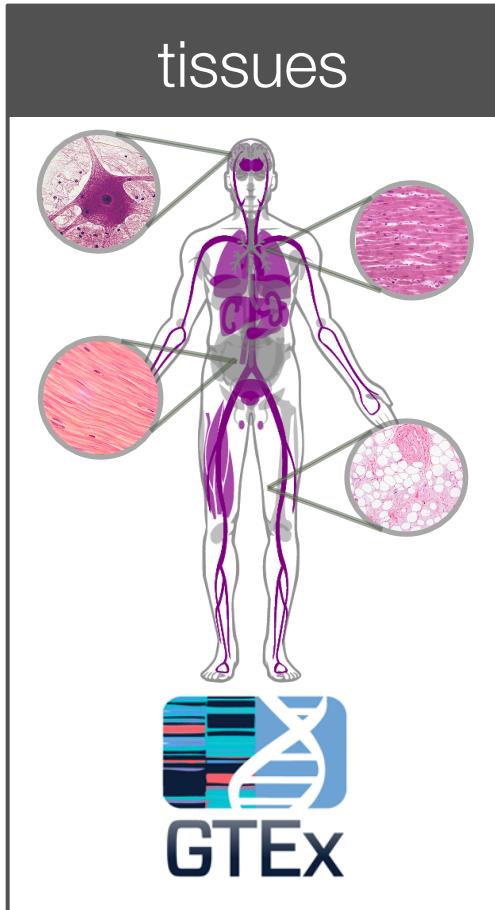
> 2,550 microarray data



... across more than 40 species ...



... under different biological conditions ...



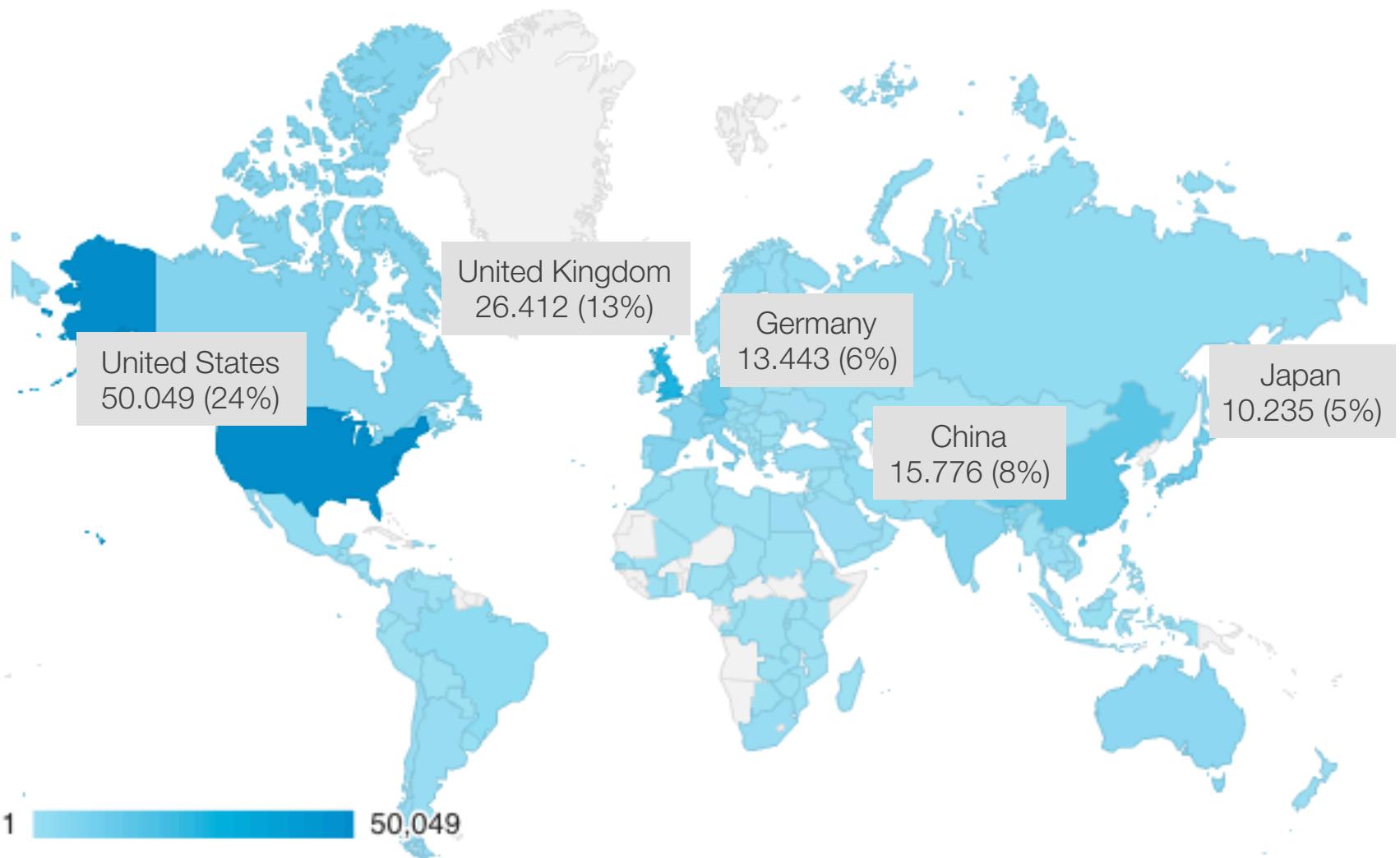
... to help scientists find answers ...

In which conditions
is my favourite gene
expressed?

What genes are
differentially expressed in
a particular condition?

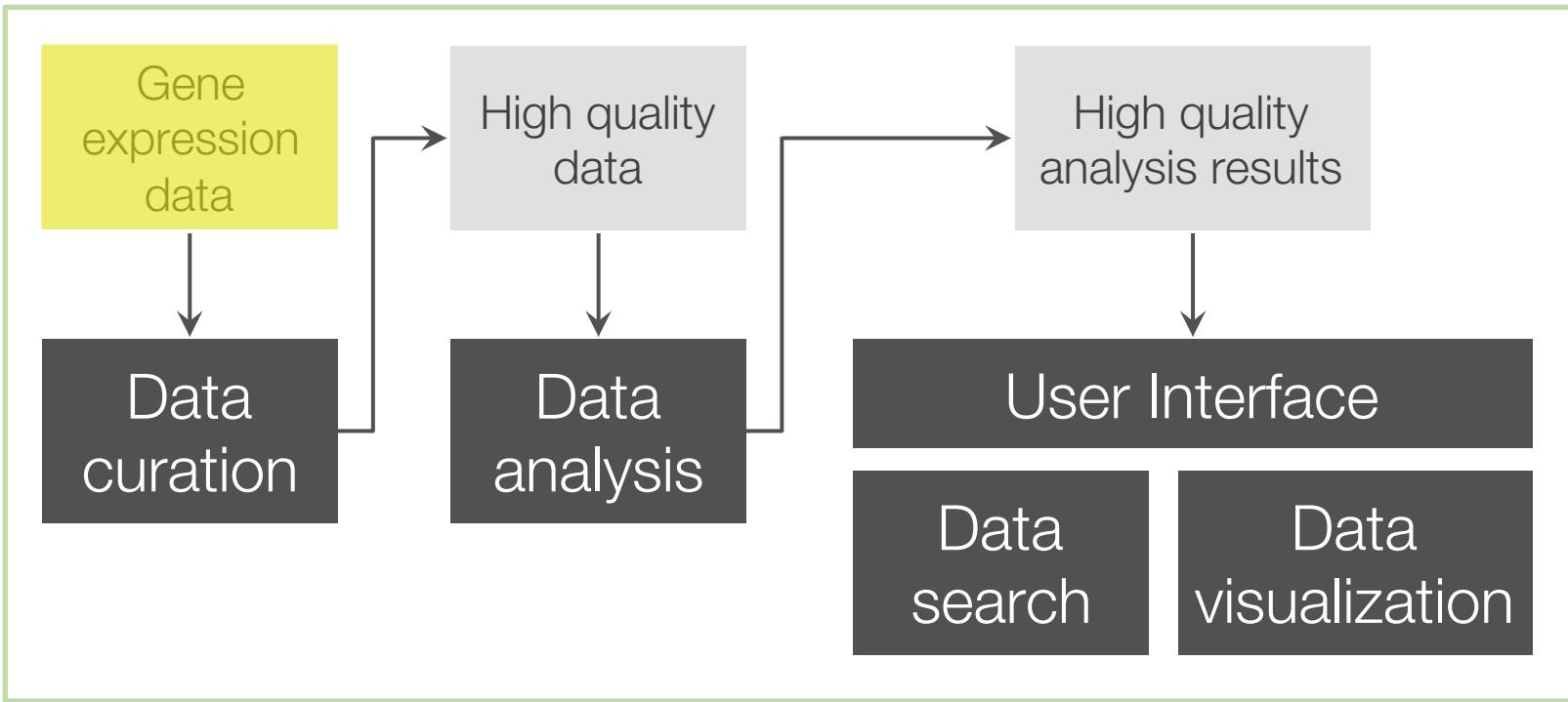
What genes are
expressed in a
particular condition?

... all across the world ...





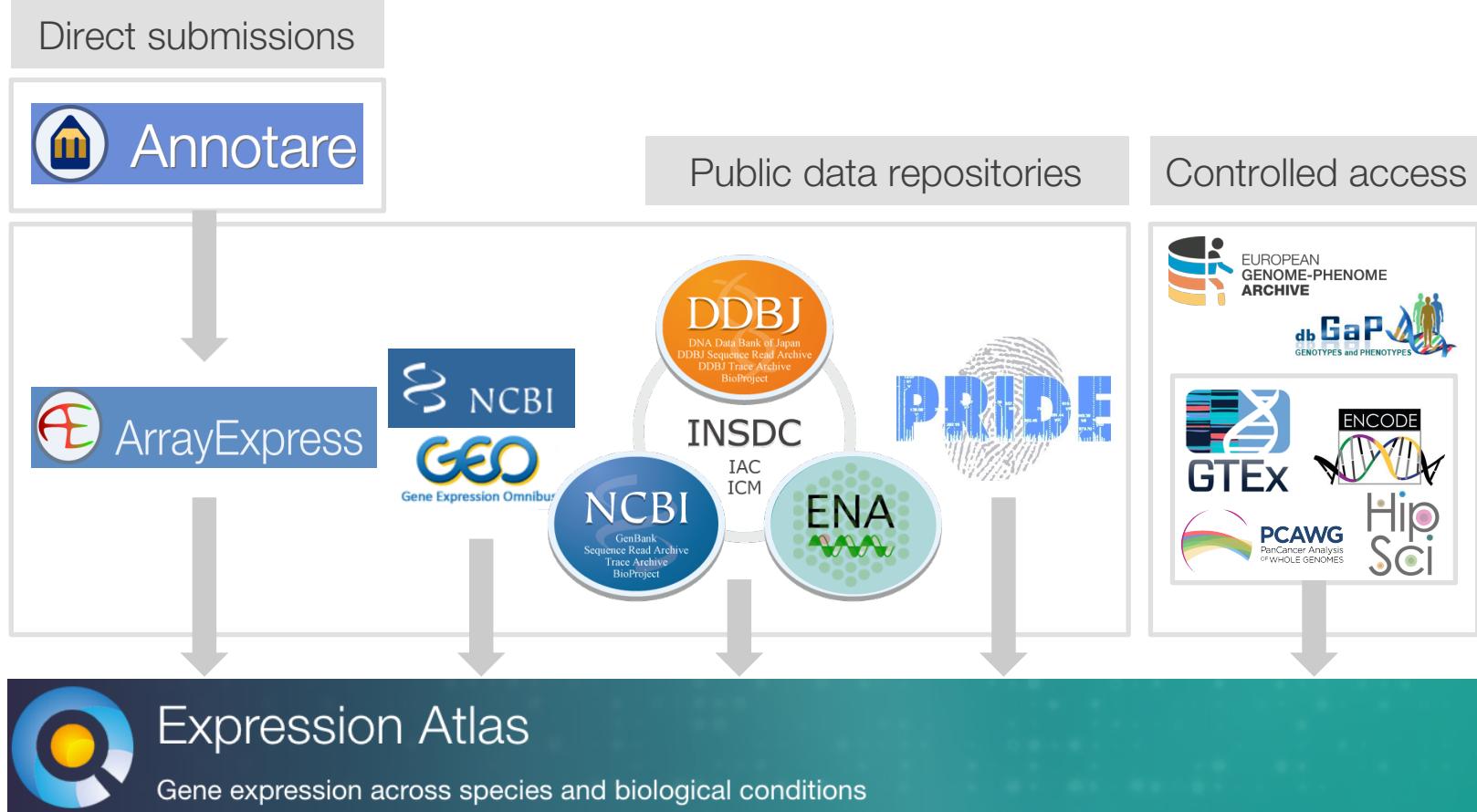
How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

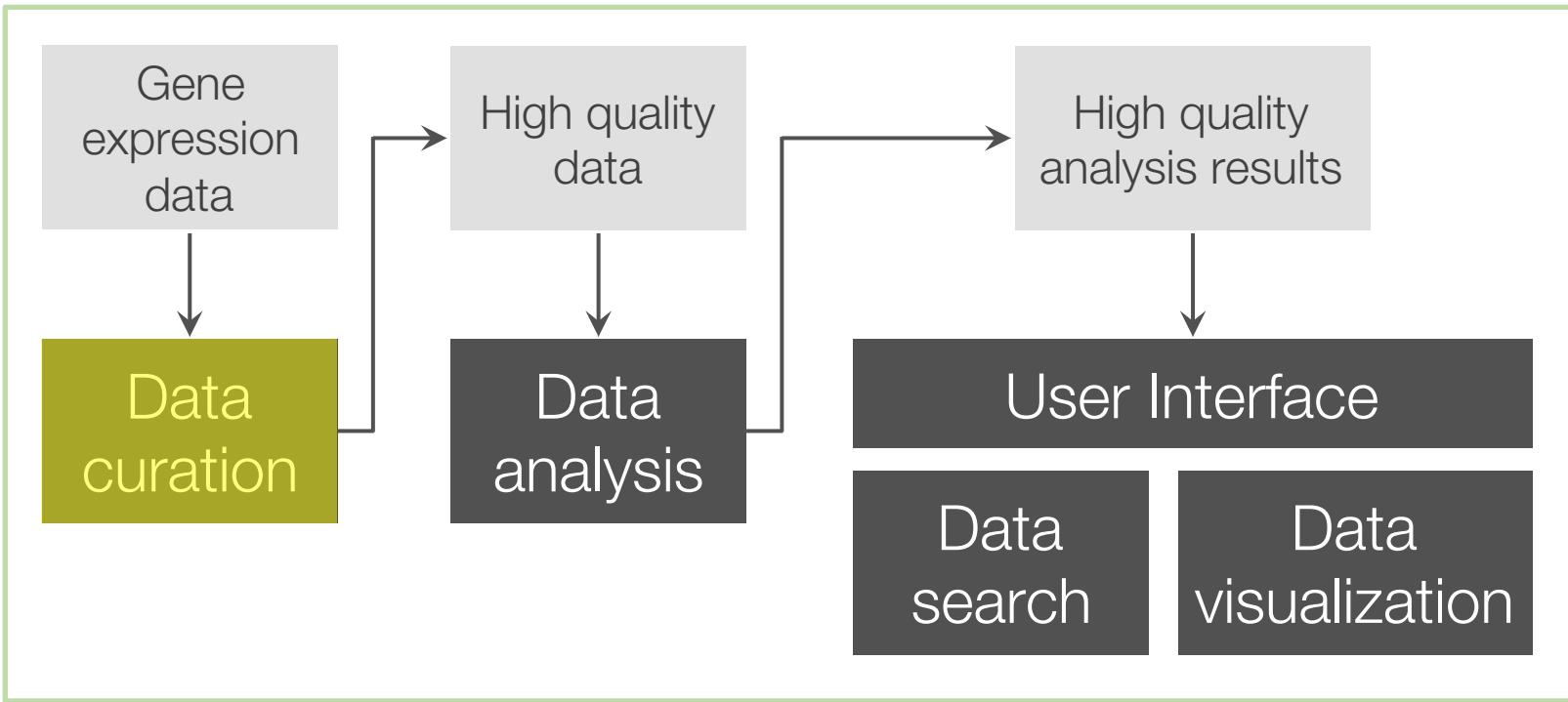


Where does expression data come from?





How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

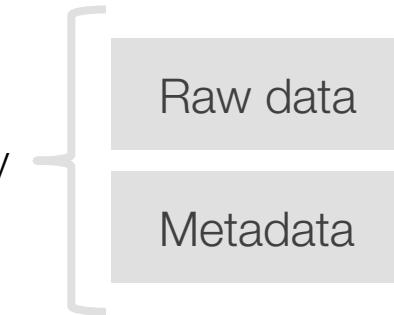


Expression Atlas – data curation

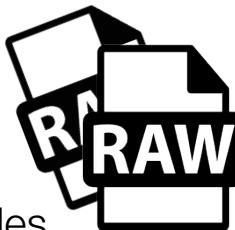
Biocuration



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



Raw data



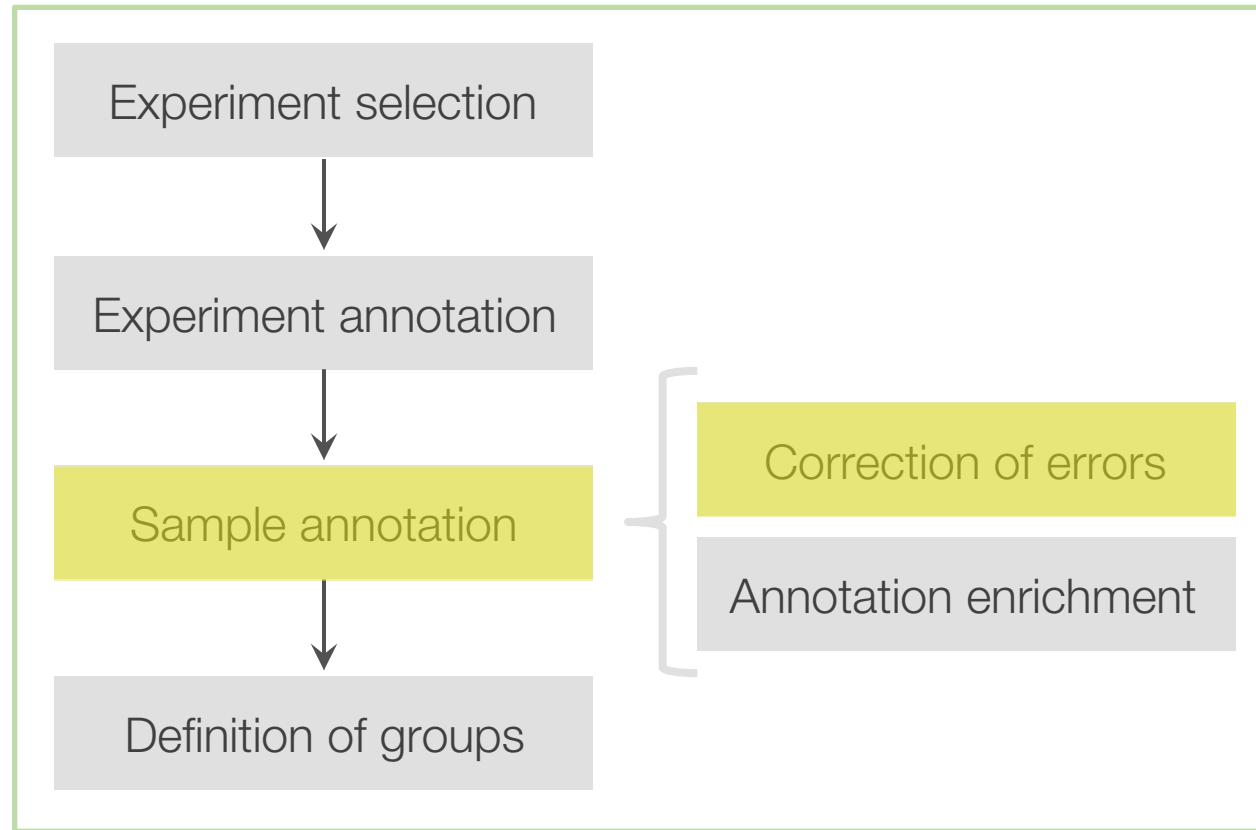
- unprocessed data files
- **Microarray:** files from the scanner (e.g. Affymetrix CEL files, Agilent feature extraction *txt* files, Illumina *idat* files)
- **Sequencing:** raw sequence read files (e.g. FASTQ files)

Metadata

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



Expression Atlas – data curation





Expression Atlas – data curation

Series GSE44596

Query DataSets for GSE44596

Status	Public on Feb 20, 2014
Title	The effect of Sparstolonin B (SsnB) on gene expression in HCAECs
Organism	<i>Homo sapiens</i>
Experiment type	Expression profiling by array
Summary	<p>Sparstolonin B is a novel bioactive compound isolated from <i>Sparganium stoloniferum</i>, an herb historically used in Traditional Chinese Medicine as an anti-tumor agent. SsnB has previously demonstrated anti-angiogenic properties. In functional assays, SsnB inhibited endothelial cell tube formation (Matrigel method) and cell migration (Transwell method) in a dose-dependent manner.</p> <p>We used microarrays to examine how SsnB affected the gene expression of human coronary artery endothelial cells (HCAECs), focusing in particular on pathways related to angiogenesis.</p>
Overall design	Three plates of HCAECs were exposed to 100 micromolar SsnB and three plates of HCAECs were exposed to Vehicle Control (1:1000 dilution of DMSO). After 24 hours, RNA was extracted for microarrays and gene expression was analyzed.
Contributor(s)	Bateman H, Barth J, Lessner S
Citation(s)	Bateman HR, Liang Q, Fan D, Rodriguez V et al. Sparstolonin B inhibits pro-angiogenic functions and blocks cell cycle progression in endothelial cells. <i>PLoS One</i> 2013;8(8):e70500. PMID: 23940584
Submission date	Feb 23, 2013
Last update date	May 05, 2017
Contact name	Henry Rhodes Bateman
E-mail	hbateman@uscmed.sc.edu
Organization name	University of South Carolina
Department	Cell Biology & Anatomy
Lab	Bldg 1, RM C-36
Street address	6439 Garners Ferry Road
City	Columbia
State/province	SC
ZIP/Postal code	29209
Country	USA
Platforms (1)	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
Samples (6)	GSM1087598 HCAEC_SsnB_rep1
≡ Less...	GSM1087599 HCAEC_SsnB_rep2
	GSM1087600 HCAEC_SsnB_rep3
	GSM1087601 HCAEC_DMSO_rep1
	GSM1087602 HCAEC_DMSO_rep2
	GSM1087603 HCAEC_DMSO_rep3





Expression Atlas – data curation

Sample annotation

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



GSM1087598 HCAEC_SsnB_rep1
GSM1087599 HCAEC_SsnB_rep2
GSM1087600 HCAEC_SsnB_rep3

{ Source name
Organism
Characteristics

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

GSM1087601 HCAEC_DMSO_rep1
GSM1087602 HCAEC_DMSO_rep2
GSM1087603 HCAEC_DMSO_rep3

{ Source name
Organism
Characteristics

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



Expression Atlas – data curation

Sample annotation



Correct errors

Hi, Laura

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

GSM1087598 HCAEC_SsnB_rep1

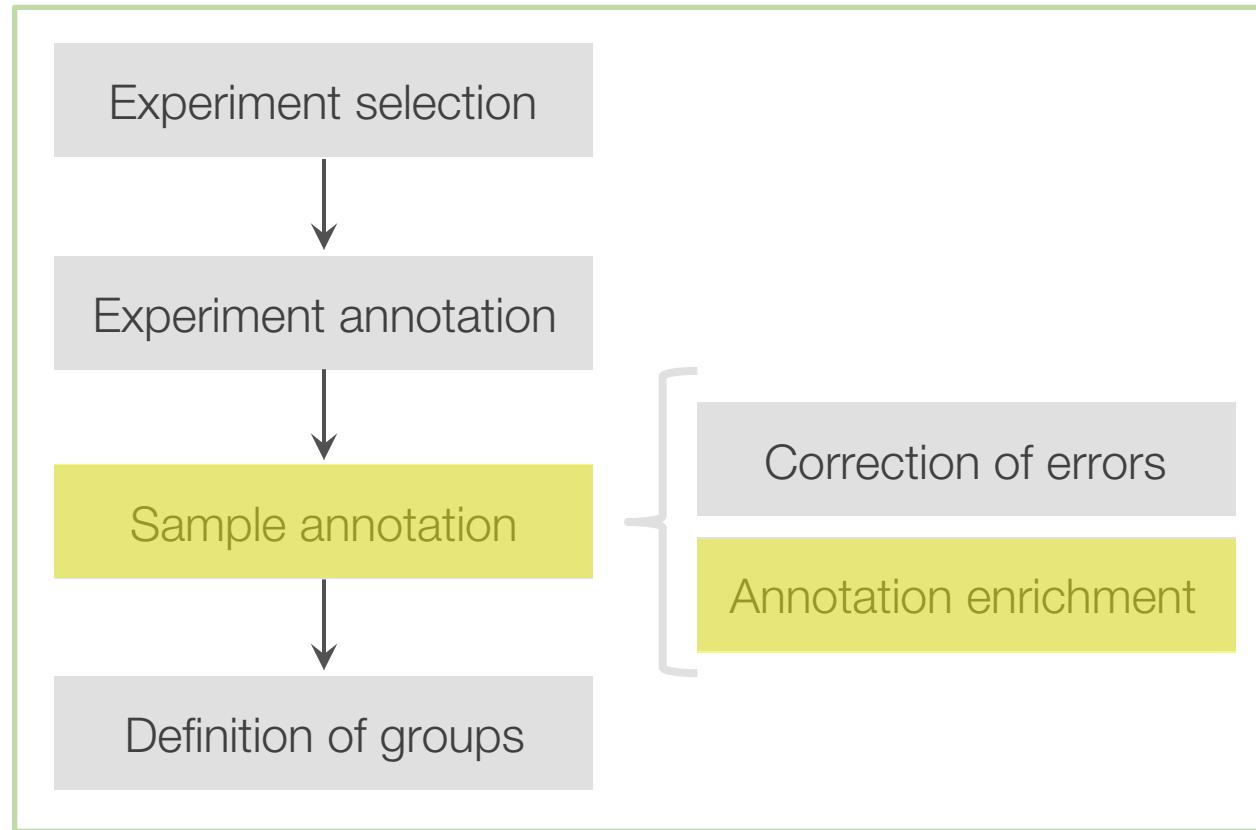
GSM1087599 HCAEC_SsnB_rep2

GSM1087600 HCAEC_SsnB_rep3

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



Expression Atlas – data curation





Expression Atlas – data curation

www.ncbi.nlm.nih.gov/geo



Series GSE81089 Query DataSets for GSE81089

Status	Public on Jun 01, 2016
Title	Next Generation Sequencing (RNAseq) of non-small cell lung cancer
Organism	<i>Homo sapiens</i>
Experiment type	Expression profiling by high throughput sequencing
Summary	Cancer testis antigens (CTAs) are of clinical interest as biomarkers and present valuable targets for immunotherapy. To comprehensively characterize the CTA landscape of non-small cell lung cancer (NSCLC), we compared RNASEq data of 199 NSCLC tissues to the normal transcriptome of 142 samples from 32 different normal organs. Of 232 CTAs currently annotated in the CTDatabase, 96 were confirmed in NSCLC. applied stringent criteria on CTAs, of which 55 genes v analysis revealed that CTA e expression is common. Imm expression of selected gene regulatory mechanism of CTA Cancer Genome Atlas. The pr was not confirmed, neither in analysis of 1117 NSCLC cases
Overall design	Fresh frozen tumor tissue surgically treated 2006-201 Sweden and 19 paired norm the regional lung cancer regi Several of the new CTAs are ; Sample characteristics values pTNM: decided by Hans Bruni Stage according to pTNM: 1= Histology diagnosis spring 20 1=squamous cell cancer 2=Ad Surgery date: the date when Age: age when surgery was p Vital date: day of death or lat Dead: 0=no 1=yes Smoking history : 1=current WHO performance status: Per Please note that the L608T_ data files) are associated with
Contributor(s)	Djureinovic D, Hallström BM, Horie M, Mattsson JS, La Fleur L, Fagerberg L, Brunnström H, Lindskog C, Madjar K, Rahnenführer J, Ekman S, Ståhle E, Koyi H, Brandén E, Edlund K, Hengstler JG, Lambe M, Saito A, Botling J, Pontén F, Uhlén M, Micke P
Citation missing	Has this study been published? Please login to update or notify GEO.

Profiling cancer testis antigens in non-small-cell lung cancer.
(PMID:27699219 PMCID:PMC5033889)

[Abstract](#) [Citations](#) [BioEntities](#) [Related Articles](#) [External Links](#)

[Djureinovic D¹, Hallström BM², Horie M³, Mattsson JS¹, La Fleur L¹, Fagerberg L² , Brunnström H⁴, Lindskog C¹ , Madjar K⁵, Rahnenführer J⁵ , Ekman S⁶, Ståhle E⁷ , Koyi H⁸, Brandén E⁸, Edlund K⁹, Hengstler JG⁹, Lambe M¹⁰, Saito A³, Botling J¹, Pontén F¹ , Uhlén M², Micke P¹](#)

[Affiliations](#) ▾

[JCI Insight](#) [07 Jul 2016, 1(10):e86837]

Type: research-article, Journal Article

machines (Illumina) using the standard Illumina RNAseq protocol with a read length of 2 × 100 bases. The raw data has been uploaded together with clinical information on GEO, with the accession number [GSE81089](http://www.ncbi.nlm.nih.gov/geo/) (<http://www.ncbi.nlm.nih.gov/geo/>).



Expression Atlas – data curation

Sample annotation



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

GSE81089



Search

Sample GSM2142642	
Status	Public on Jun 01, 2016
Title	matched sample_L511N
Sample type	SRA
Source name	Human non-malignant tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511N

non-small cell
lung carcinoma

Sample GSM2142481	
Status	Public on Jun 01, 2016
Title	L511T
Sample type	SRA
Source name	Human NSCLC tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511T stage tnmx: 5 histology: 2 surgery date: 2007-04-05 age: 56 gender: female vital date: 2013-04-28 dead: 0 smoking: 3 ps who: 0

non-small cell
lung carcinoma



Disease?



Expression Atlas – data curation

Sample annotation



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

→

Sample GSM2142642

Status	Public on Jun 01, 2016
Title	matched sample_L511N
Sample type	SRA
Source name	Human non-malignant tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511N lung

Sample GSM2142481

Status	Public on Jun 01, 2016
Title	L511T
Sample type	SRA
Source name	Human NSCLC tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511T lung stage tn: 5 histology: 2 surgery date: 2007-04-05 age: 56 gender: female vital date: 2013-04-28 dead: 0 smoking: 3 ps who: 0



Tissue?



Expression Atlas – data curation

Sample annotation



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

→

Sample GSM2142642

Status	Public on Jun 01, 2016
Title	matched sample_L511N
Sample type	SRA
Source name	Human non-malignant tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511N

non-tumor

Sample GSM2142481

Status	Public on Jun 01, 2016
Title	L511T
Sample type	SRA
Source name	Human NSCLC tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511T stage tnM: 5 histology: 2 surgery date: 2007-04-05 age: 56 gender: female vital date: 2013-04-28 dead: 0 smoking: 3 ps who: 0



Sampling site?



Expression Atlas – data curation

Sample annotation



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

→

Sample GSM2142642

Status	Public on Jun 01, 2016
Title	matched sample_L511N
Sample type	SRA
Source name	Human non-malignant tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511N

individual 511

Sample GSM2142481

Status	Public on Jun 01, 2016
Title	L511T
Sample type	SRA
Source name	Human NSCLC tissue
Organism	Homo sapiens
Characteristics	tumor (t) or normal (n): L511T stage tnM: 5 histology: 2 surgery date: 2007-04-05 age: 56 gender: female vital date: 2013-04-28 dead: 0 smoking: 3 ps who: 0

individual 511



Same patient?



Expression Atlas – data curation

Sample annotation

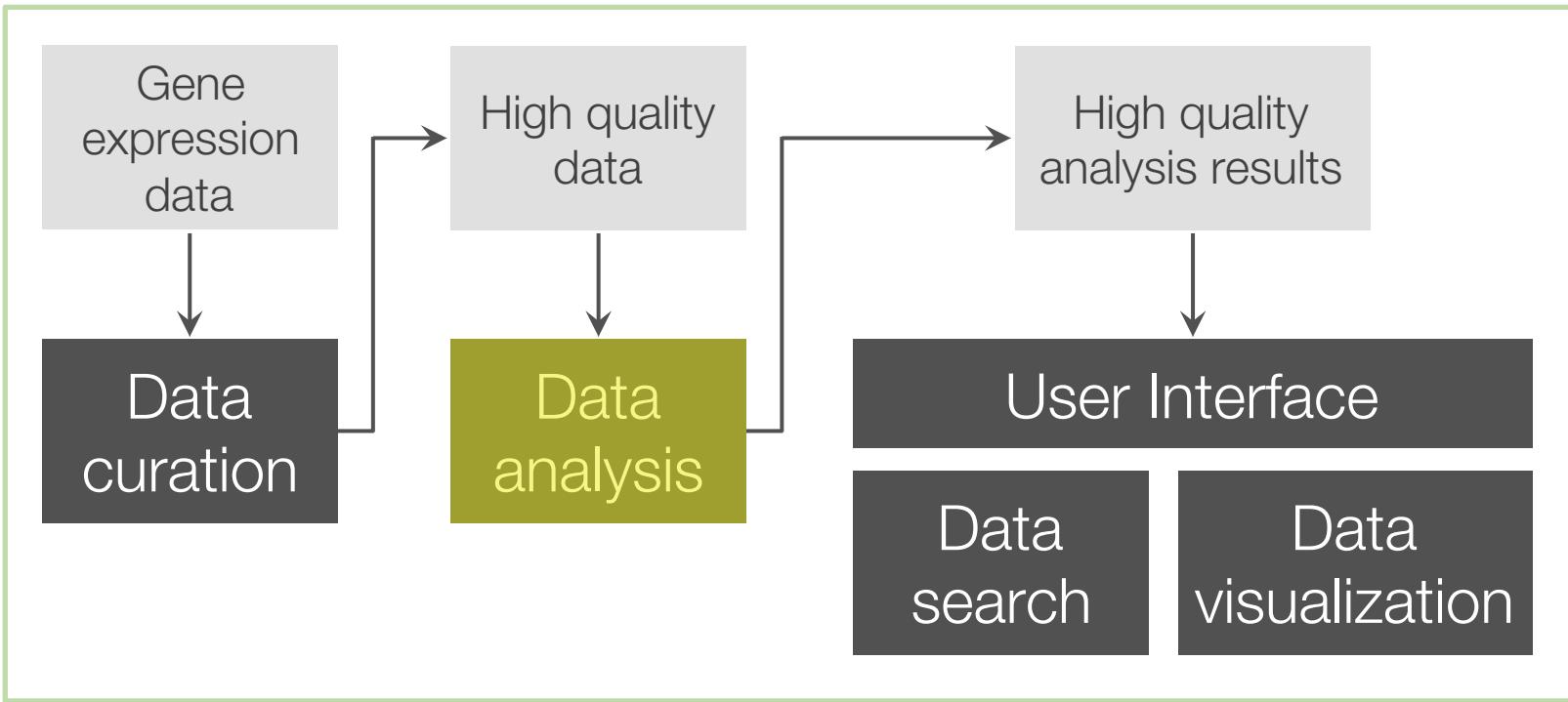


Adding value to your data

Run	Sample Characteristics									Experimental Variables
	age	clinical information	disease	individual	organism	organism part	sampling site	sex	sampling site	
SRR3474756	56 year	smoker	non-small cell lung carcinoma	L493	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3474757	77 year	smoker	non-small cell lung carcinoma	L496	Homo sapiens	lung	tumor tissue	male	tumor tissue	
SRR3474758	65 year	ex-smoker	non-small cell lung carcinoma	L504	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3474759	56 year	non-smoker	non-small cell lung carcinoma	L511	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3475321	56 year	non-smoker	non-small cell lung carcinoma	L511	Homo sapiens	lung	non-malignant tissue	female	non-malignant tissue	
SRR3474760	79 year	ex-smoker	non-small cell lung carcinoma	L529	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3474761	76 year	smoker	non-small cell lung carcinoma	L530	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3474762	74 year	ex-smoker	non-small cell lung carcinoma	L531	Homo sapiens	lung	tumor tissue	male	tumor tissue	
SRR3474763	69 year	ex-smoker	non-small cell lung carcinoma	L532	Homo sapiens	lung	tumor tissue	female	tumor tissue	
SRR3475322	69 year	non-smoker	non-small cell lung carcinoma	L532	Homo sapiens	lung	non-malignant tissue	female	non-malignant tissue	
SRR3474764	66 year	ex-smoker	non-small cell lung carcinoma	L534	Homo sapiens	lung	tumor tissue	male	tumor tissue	

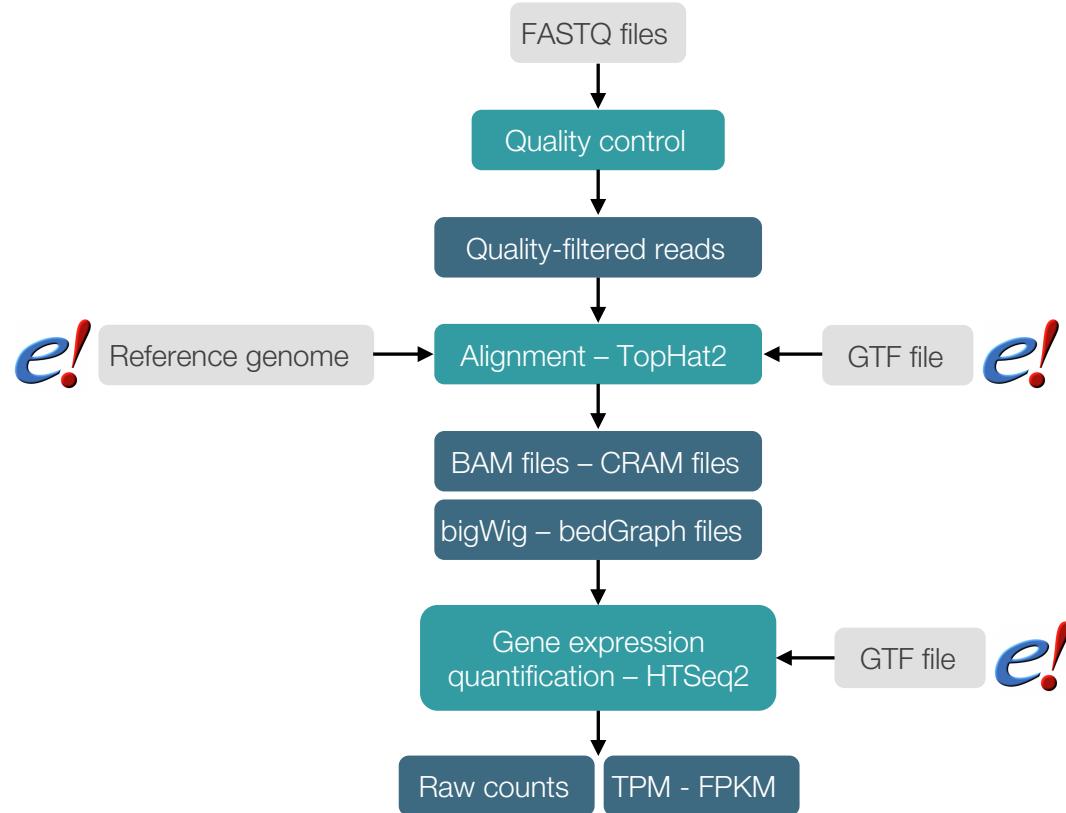


How does Expression Atlas do it?



Discover and interpret gene expression analysis results quickly and easy

RNA-seq data analysis



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



RNA-seq data analysis

Train online

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

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

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

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

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

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

About this course

Author(s): Laura Huerta, Melissa Burke

Gene Expression

Beginner

2 hours

[Start the course](#)

Learning objectives:

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

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



RNA-seq data analysis

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).

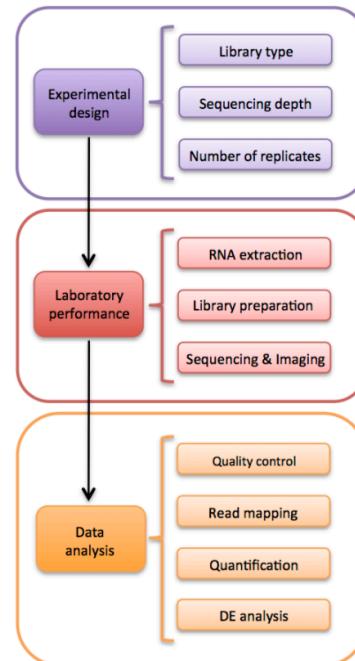


Figure 7 Overview of a RNA-seq experiment for detecting differential expression (DE).

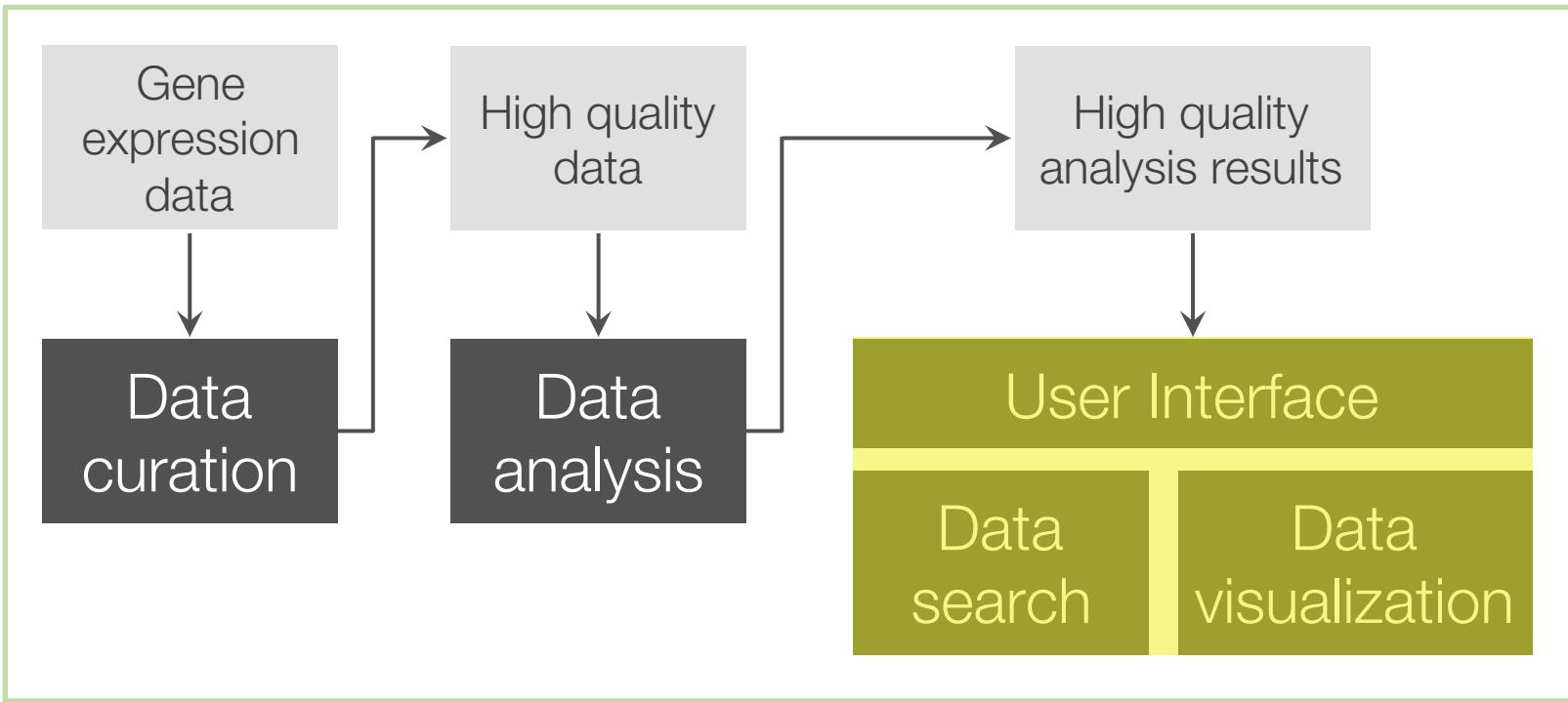
« Applications of RNA-seq

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



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
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

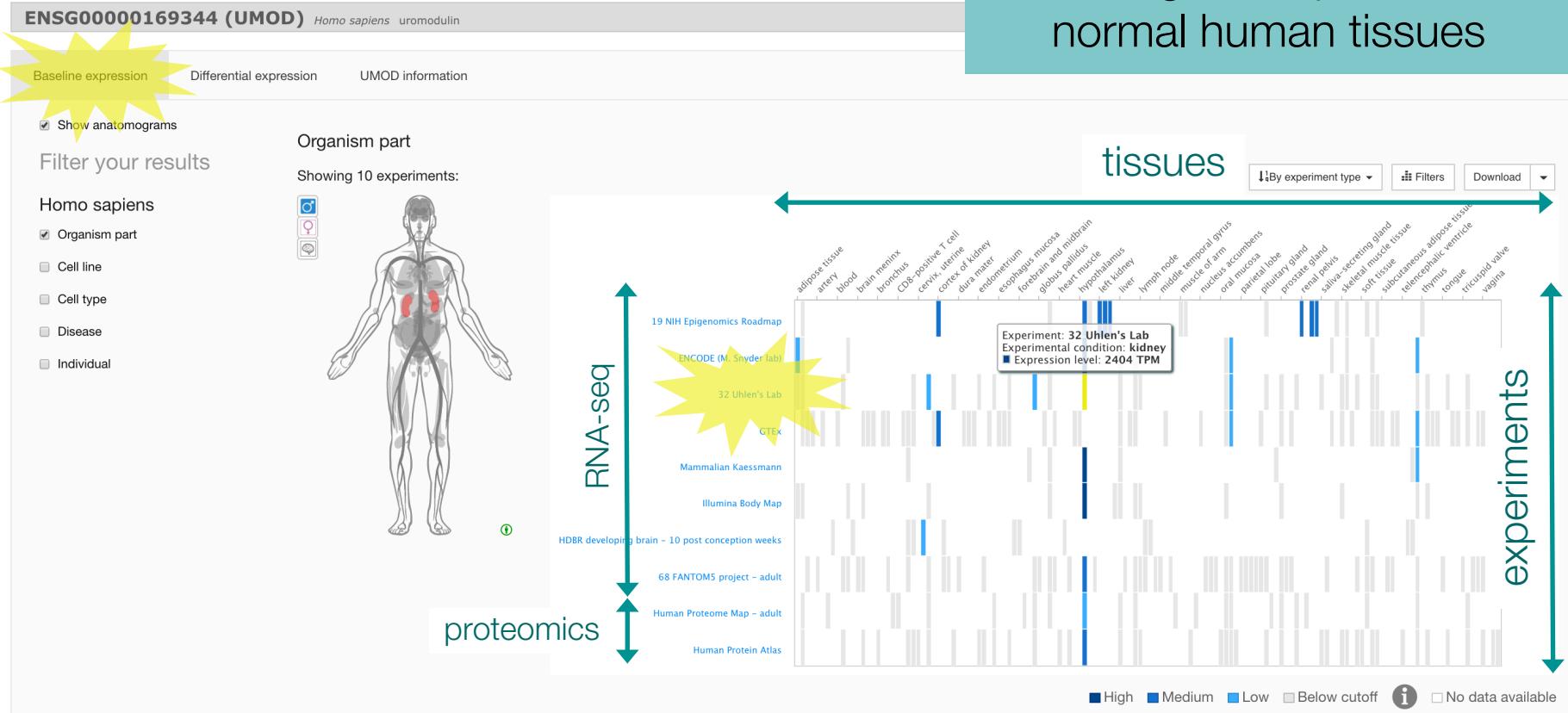
Biological conditions

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

Search | Clear



Look up expression of a gene: baseline





Look up expression of a gene: baseline

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

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

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





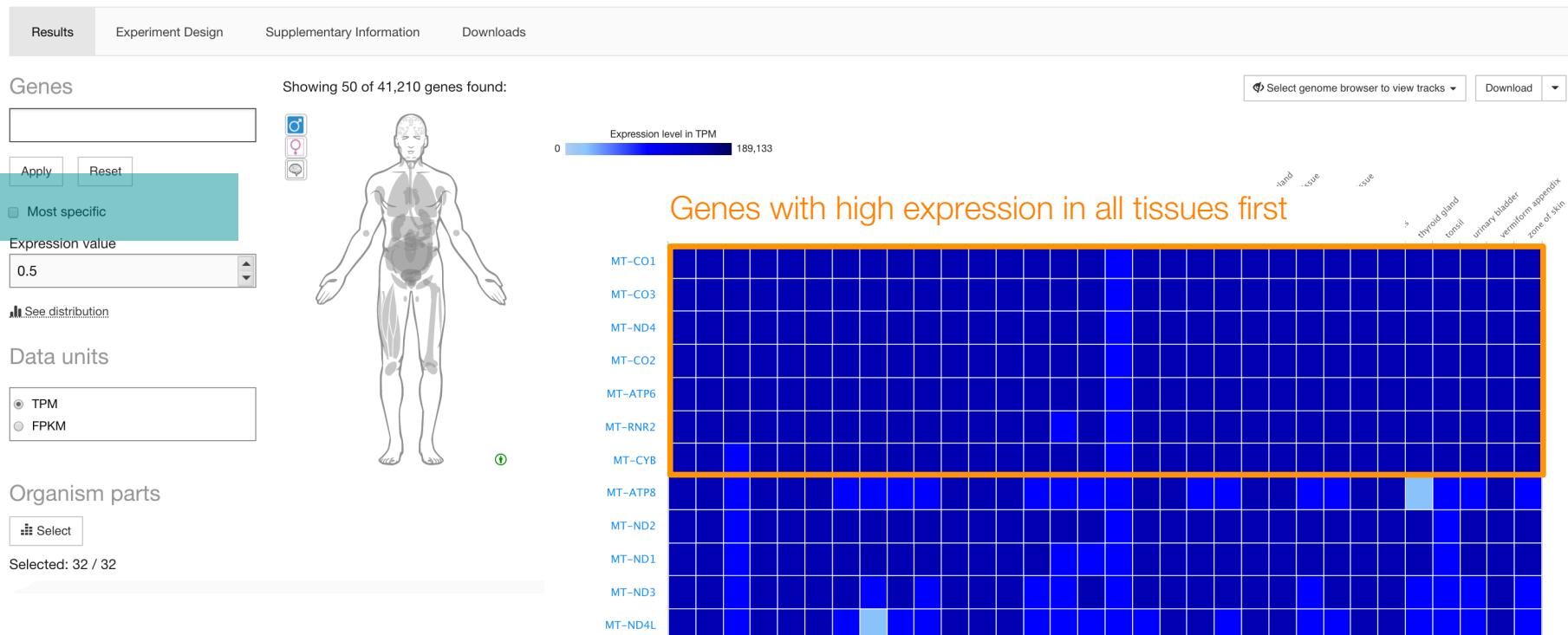
Look up expression of a gene: baseline

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

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





Look up expression of a gene: baseline

General information

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

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





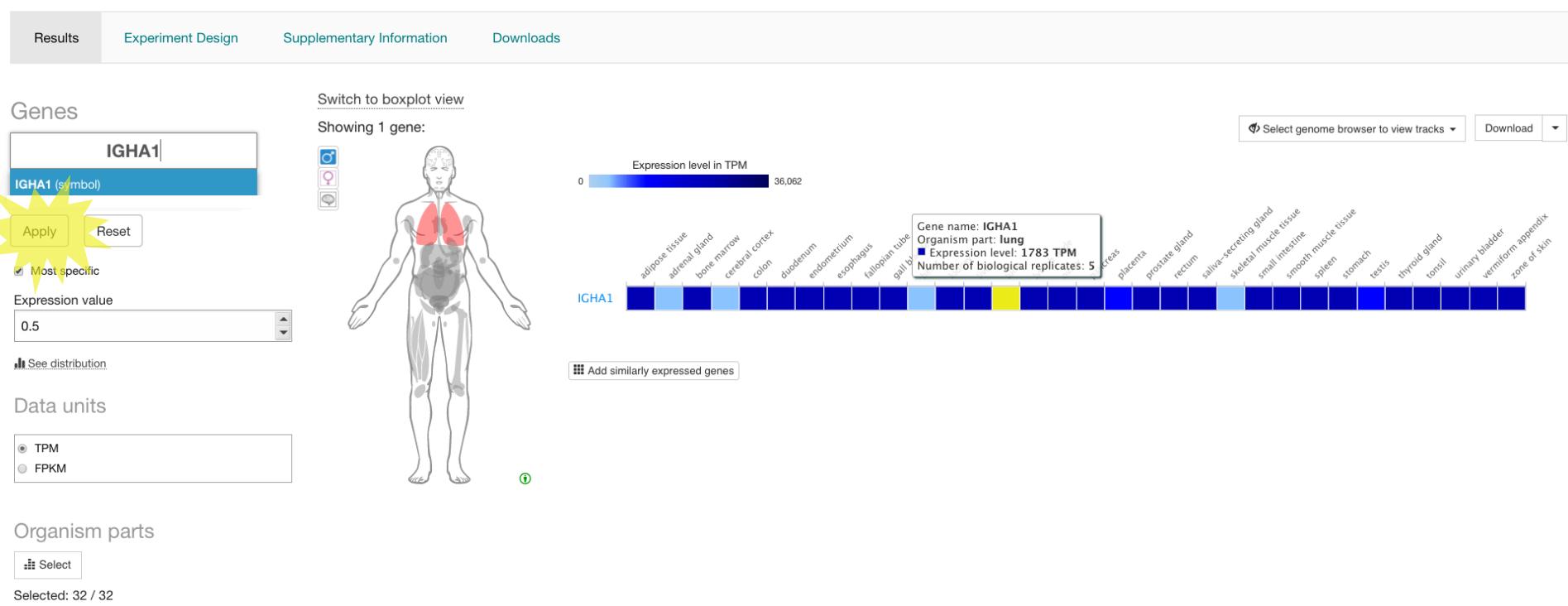
Look up expression of a gene: baseline

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

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas





Look up expression of a gene: baseline

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

RNA-Seq mRNA baseline

Organism: *Homo sapiens*

Raw Data Provider: The Human Protein Atlas

Results Experiment Design Supplementary Information Download

Genes

IGHA1

[Switch to boxplot view](#)

Showing 1 gene:



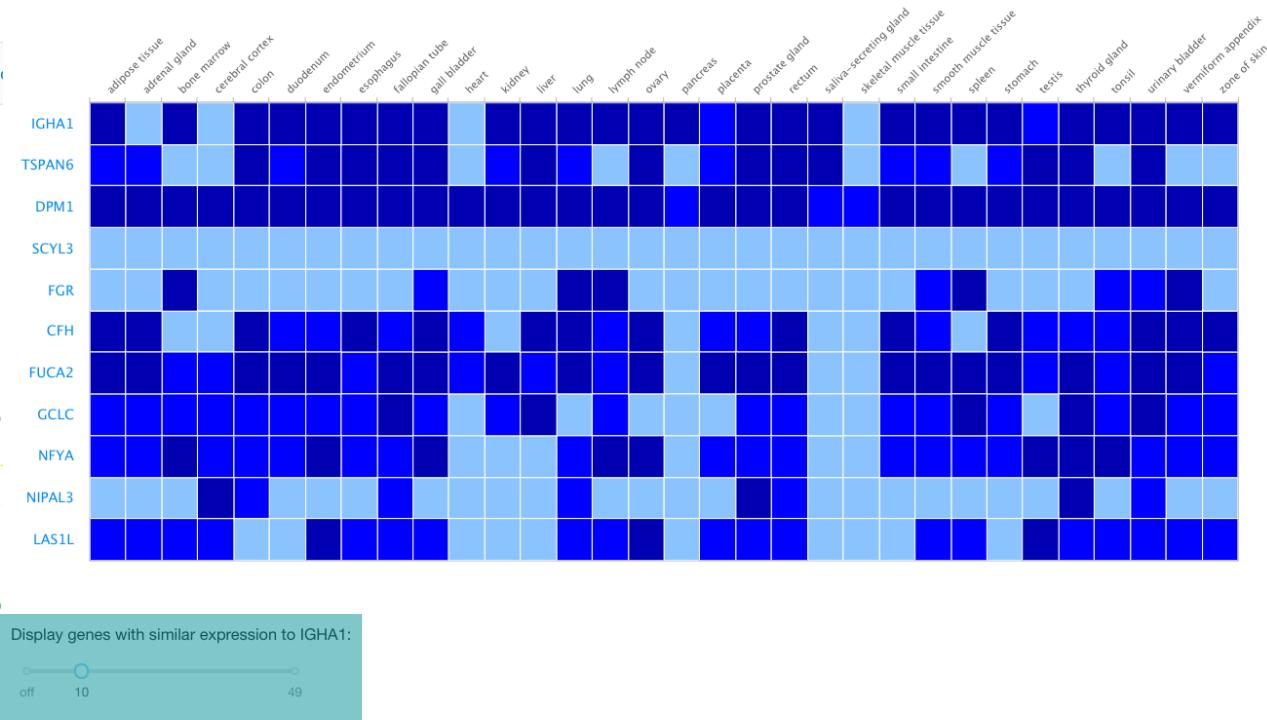
Most specific

Expression value: 0.5

[See distribution](#)

Data units

TPM FPKM





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

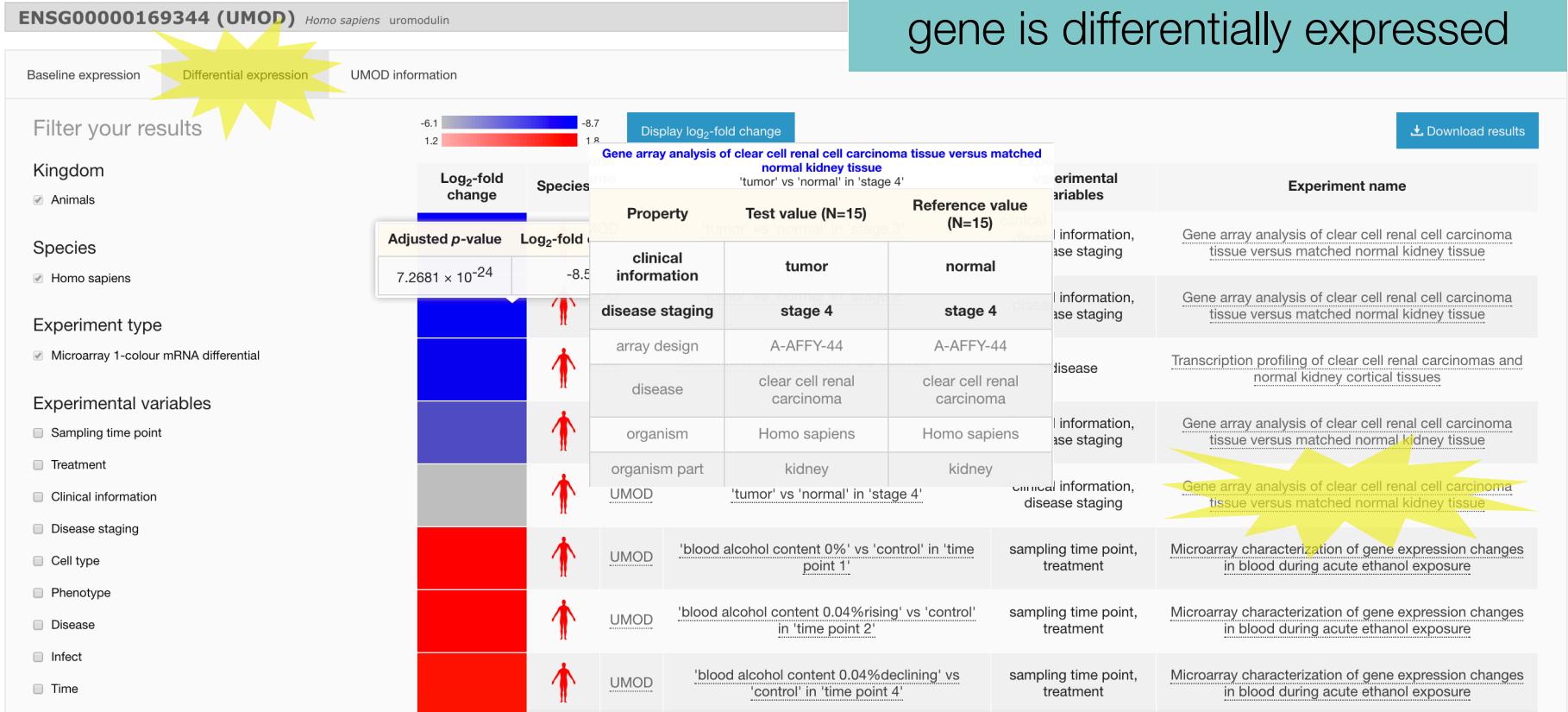
Biological conditions

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)





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



Expression Atlas – data curation

Ontology

www.ebi.ac.uk/efo



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

Controlled vocabulary

Hierarchy (relationship)

Smart search

Efficient search via ontology-driven query expansion



Biological conditions

lung car|

Ex lung carcinoid tumor

+ lung carcinoma

Complex queries



lung carcinoma

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



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

Log₂-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 β -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

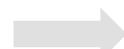
Differentially expressed genes in human lung carcinoma



Expression Atlas – data curation

Ontology

www.ebi.ac.uk/efo



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 



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



Let's try Expression Atlas

Differential expression

Hands-on activity

Find differentially expressed genes in patients with hepatocellular carcinoma



In pairs



More tips on how to use Expression Atlas

<http://goo.gl/c3ce23>

Train online

Training | Train online | About Train online | Glossary | Support and feedback | Login/register

Expression Atlas: Quick tour

What is Expression Atlas?
What can I do with Expression Atlas?
Searching and visualising data in Expression Atlas
Getting data from Expression Atlas
Get help and support on Expression Atlas
Your feedback
Contributors

CC BY SA

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

Expression Atlas: Quick tour



This extended quick tour provides an introduction to [Expression Atlas](#), an open resource for finding information about gene expression across species and biological conditions. Expression Atlas aims to help answer questions such as 'where is a certain gene expressed?' or 'how does its expression change in a disease?'. Undergraduate-level understanding of biology would be an advantage.

Last updated October 2017.

About this course

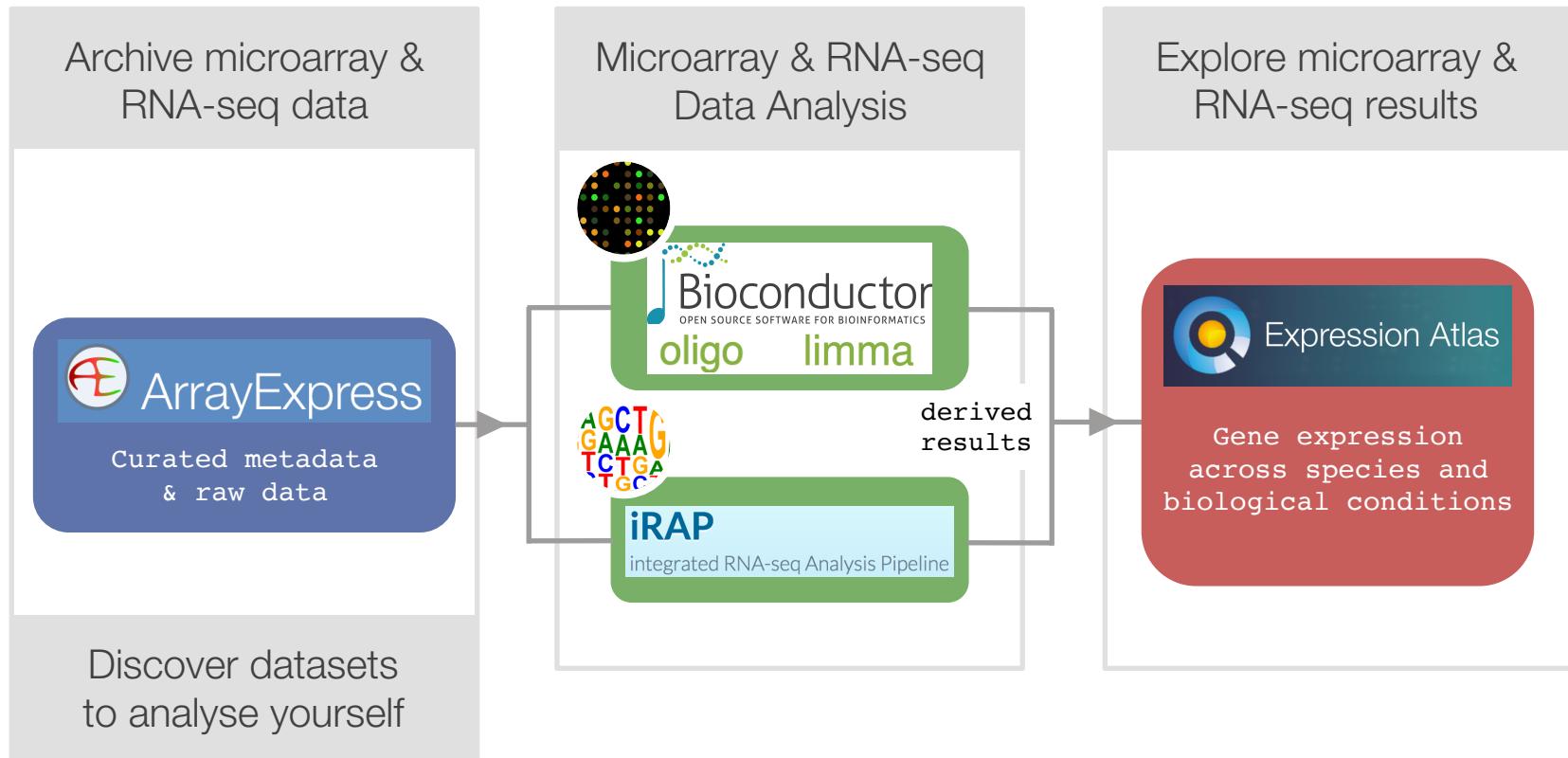
Author(s): Laura Huerta
Gene Expression
Beginner 0.5 hour

[Start the course](#)

Learning objectives:

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

Summary of the session



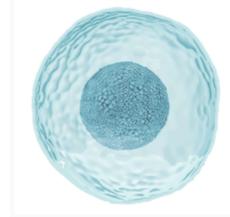


New project on single-cell...

 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	Analysis result	
 06-02-2017 Single-cell RNA-sequencing resolves a CD4+ T cell fate bifurcation <i>Mus musculus</i>	752	Analysis result	
 06-02-2017 Single-cell RNA-sequencing of <i>Platynereis dumerilii</i> larval brain cells <i>Platynereis dumerilii</i>	285	Analysis result	

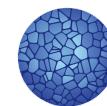
Marker genes search

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 document icon, 'High quality results' with a bar chart icon, 'Accessible results' with a sunburst chart icon, and '... easy to interpret' with a grid icon. A grey arrow points from the bottom of the first column to the bottom of the second column.

About Expression Atlas

What is Expression Atlas?

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

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

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

atlas-feedback@ebi.ac.uk



@ExpressionAtlas

EMBL-EBI: Bioinformatics resources for exploring disease related data

ArrayExpress and Expression Atlas services at EMBL-EBI

Anja Fullgrabe, PhD

Scientific Curator

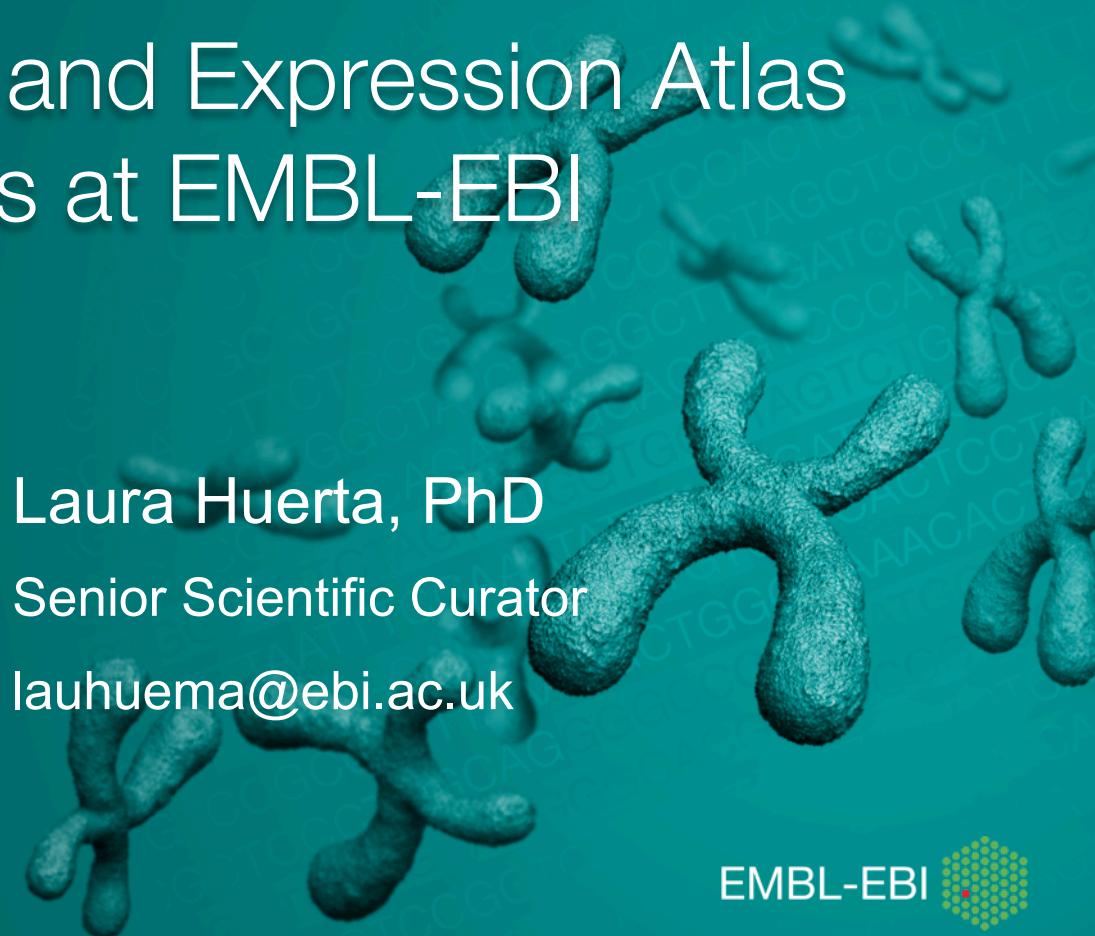
anjaf@ebi.ac.uk

30 October 2017

Laura Huerta, PhD

Senior Scientific Curator

lauhuema@ebi.ac.uk



Please fill in the survey before you go!

<https://www.surveymonkey.co.uk/r/7HFKTBJ>

[Exit this survey >>](#)

EMBL-EBI: Bioinformatics resources for exploring disease related data

Workshop Evaluation Form

EMBL-EBI collects feedback from every course and workshop we run. The survey is a way for you to inform us about the workshop you have participated in, what you enjoyed, what you found useful and how we can make improvements. This information is also used to inform the development of new courses and workshops.

Next