

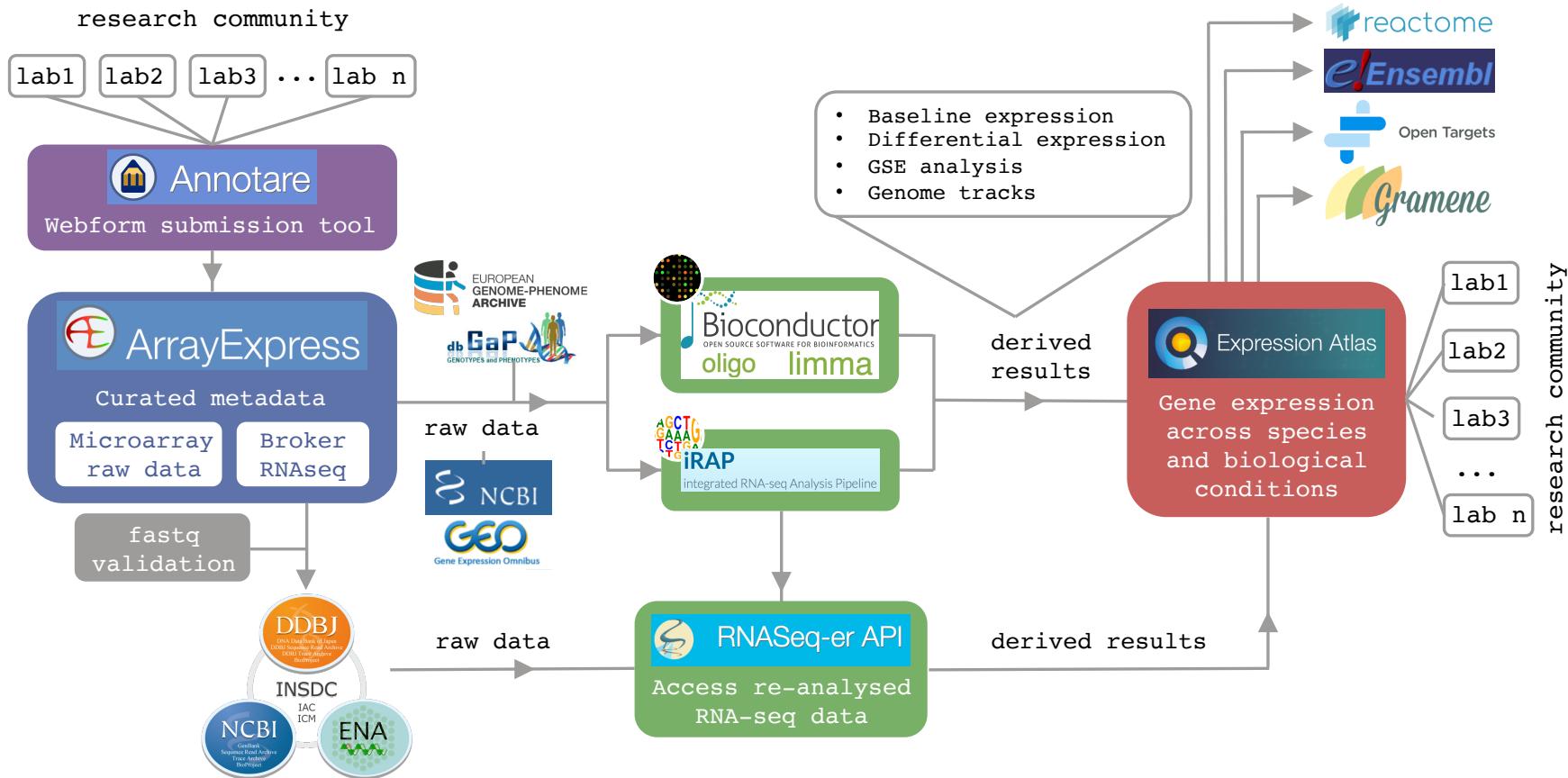
# EMBL-EBI workshop: Tools for functional genomics data

Discover functional genomics data  
with ArrayExpress

Laura Huerta, PhD  
Senior Scientific Curator  
[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)  
24 May 2018



# Functional genomics resources at EMBL-EBI



# *Outline of the session*

- ✓ What is ArrayExpress?
  - ✓ Data standards: MIAME & MINSEQE
  - ✓ Data format: MAGE-TAB
    - ✓ Hands-on exercise
  - ✓ Experimental variable
    - ✓ Hands-on exercise
- ✓ Data mining with ArrayExpress
  - ✓ Hands-on exercise



# What is ArrayExpress?

**commentary**

## One-stop shop for microarray data

Is a universal, public DNA-microarray database a realistic goal?

**Alvis Brazma, Alan Robinson,  
Graham Cameron  
and Michael Ashburner**

NATURE | VOL 403 | 17 FEBRUARY 2000 | [www.nature.com](http://www.nature.com)

*“The EBI, in collaboration with the German Cancer Research Centre, is developing ArrayExpress”*



Martin Vingron



Alvis Brazma



# What is ArrayExpress?

**commentary**

## One-stop shop for microarray data

Is a universal, public DNA-microarray database a realistic goal?

**Alvis Brazma, Alan Robinson,  
Graham Cameron  
and Michael Ashburner**

Of the techniques that are being used to obtain the massive data sets of the molecules of life, the most visible is the DNA sequencing of the human genome. Following on from the publication of the human chromosome 22 sequence<sup>1</sup>, a rough draft of the whole human genome should be available by the spring. But such advances can create the false impression that everything about life at the molecular level will soon be understood.

In reality, genome projects simply transfer digital information from DNA to computer file; this genetic 'parts-list' is a long way from providing an understanding of function. It took hundreds of years to advance from a fairly detailed understanding of human anatomy to any real understanding of function. Knowing the genome sequence and

experiment looking at 40,000 genes from 10 different samples, under 20 different conditions, produces at least 8,000,000 pieces of information. Currently, these data are scattered among various independent Internet sites, or may not be publicly available at all, although conclusions drawn from the data will have been published. Details about how experiments were carried out are often incomplete. Yet the amount of information being produced in this way is set to explode as the cost of microarray technology falls.

### The need for a public repository

It is time to create a public repository for microarray data, with standardized annotation (see Box 2, overleaf). But this is a complex and ambitious project, and is one of the biggest challenges that bioinformatics has yet faced. Major difficulties stem from the detail required to describe the conditions of an experiment, and the relative and imprecise

One difficulty concerns the inherent fuzziness of gene-expression data. Essentially all current expression measurements are relative: we can tell which genes are expressed differently in an experiment only in comparison with another experiment, or in relation to another gene in the same experiment. Such methods tell us little about how many copies of a messenger RNA are present. Moreover, the transcription levels reported are an average over the whole cell population sampled.

Consequently, gene-expression measurements from different technologies, or even from the same technology but from different laboratories, may not be quantitatively comparable. Two steps should allow data from different sources to be compared. First, relatively raw data should be stored to obviate any variation owing to, say, data-normalization methods. Second, standard sets of control probes and samples should be designed and used in experiments to give reference points

*"It's time to create a public repository for microarray data, with standardized annotation" (2000)*



# What is ArrayExpress?

**commentary**

## One-stop shop for microarray data

Is a universal, public DNA-microarray database a realistic goal?

Important tasks to be undertaken include:

1. agreement on the essential information that should be reported for a microarray experiment
2. definition of ontologies and an extensible, structured document format to capture these data and their semantics
3. production of a database to store these documents
4. development of tools for searching documents in a database and using the semantic context to allow comparisons and sophisticated queries.



# ArrayExpress: functional genomics data archive

Home    Browse    Submit    Help    About ArrayExpress    Search    Examples: E-MEXP-31, cancer, p53, Geuvadis    advanced search    Contact Us    Login

## ArrayExpress – functional genomics data

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

[Browse ArrayExpress](#)

### Latest News

13 October 2017 - **ArrayExpress is stopping import of GEO data**

Unfortunately, we are stopping the regular imports of Gene Expression Omnibus (GEO) data into ArrayExpress. We will keep using data from GEO to build our added value database [Expression Atlas](#), and the reprocessed and additionally annotated data for selected datasets will be available from there.

### Links

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

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

### Tools and Access

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

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

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

### Data Content

Updated today at 03:00

- 70487 experiments
- 2231358 assays
- 46.08 TB of archived data

### Related Projects

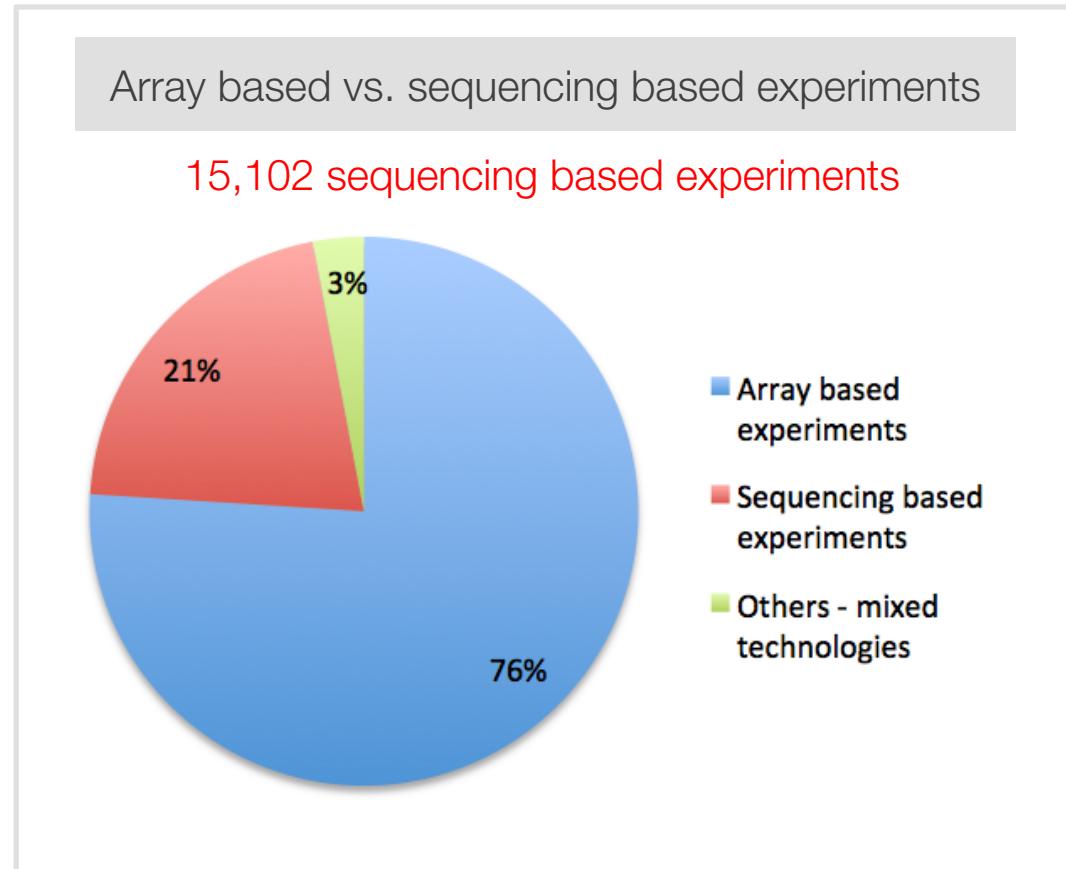
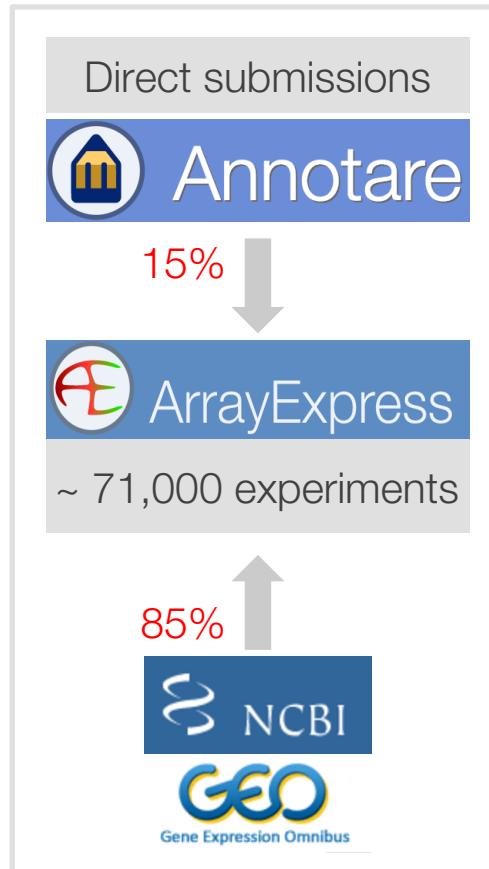
Discover up and down regulated genes in numerous experimental conditions in the [Expression Atlas](#).

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

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



# ArrayExpress: functional genomics data archive

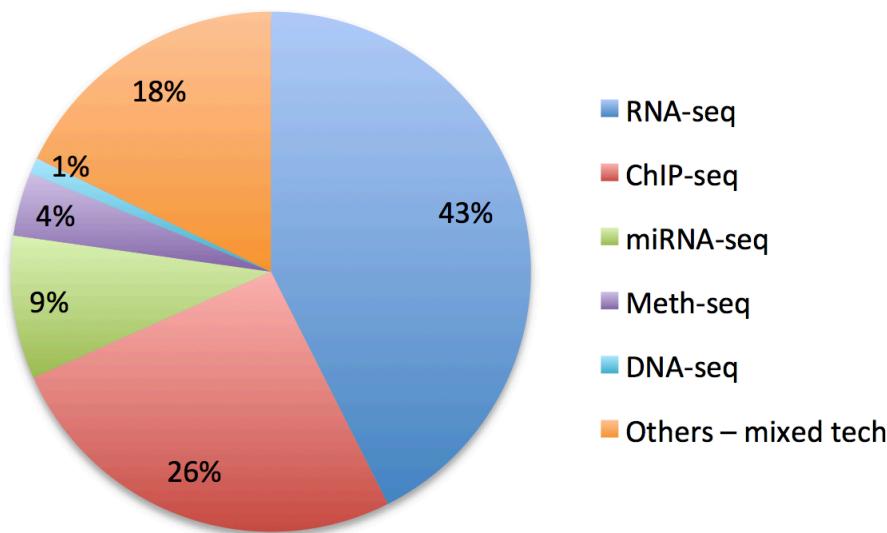




# ArrayExpress: functional genomics data archive

Sequencing based experiments

6,463 RNA-seq of coding RNA experiments





# ArrayExpress: functional genomics data archive

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

## Experiment types in ArrayExpress

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

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



# How standards are used in ArrayExpress

MIAME



Guidelines

MAGE-TAB



Format

EFO



Terminology



# ArrayExpress: MIAME compliant

*Nature Genetics* **29**, 365 - 371 (2001)  
doi:10.1038/ng1201-365

## **Minimum information about a microarray experiment (MIAME) —toward standards for microarray data**

Alvis Brazma<sup>1</sup>, Pascal Hingamp<sup>2</sup>, John Quackenbush<sup>3</sup>, Gavin Sherlock<sup>4</sup>, Paul Spellman<sup>5</sup>,  
Chris Stoeckert<sup>6</sup>, John Aach<sup>7</sup>, Wilhelm Ansorge<sup>8</sup>, Catherine A. Ball<sup>4</sup>, Helen C. Causton<sup>9</sup>,  
Terry Gaasterland<sup>10</sup>, Patrick Glenisson<sup>11</sup>, Frank C.P. Holstege<sup>12</sup>, Irene F. Kim<sup>4</sup>, Victor  
Markowitz<sup>13</sup>, John C. Matese<sup>4</sup>, Helen Parkinson<sup>1</sup>, Alan Robinson<sup>1</sup>, Ugis Sarkans<sup>1</sup>, Steffen  
Schulze-Kremer<sup>14</sup>, Jason Stewart<sup>15</sup>, Ronald Taylor<sup>16</sup>, Jaak Vilo<sup>1</sup> & Martin Vingron<sup>17</sup>

*“raw data is not enough to interpret the results  
and to verify the conclusions based on microarray  
data analysis”*



# Metadata is stored in MAGE-TAB format

## Microarray data

MAGE-TAB specification defines 4 types of files:

1. *Investigation Description Format (IDF)*
2. *Sample and Data Relationship Format (SDRF)*
3. *Array Design Format (ADF)*
4. *Raw and processed data files*



# Metadata is stored in MAGE-TAB format

Metadata in MAGE-TAB format (generated at submission)

**I**nvestigation  
**D**escription  
**F**ormat

Accession: E-MTAB-9999  
Title: "Transcription profiling of...  
Description: "In this experiment...  
Contacts: "r.e.searcher@lab...  
Protocol: "Growth protocol...  
Citation: "Dynamics of..."

**S**ample and  
**D**ata  
**R**elationship  
**F**ormat

Sample Name	Attributes	Assay Name	File
Sample 1	skin, epithelial cell	Hyb_sample1	S1.CEL
Sample 2	skin, epithelial cell	Hyb_sample2	S2.CEL
...			

Characteristics of the samples

Experimental variables

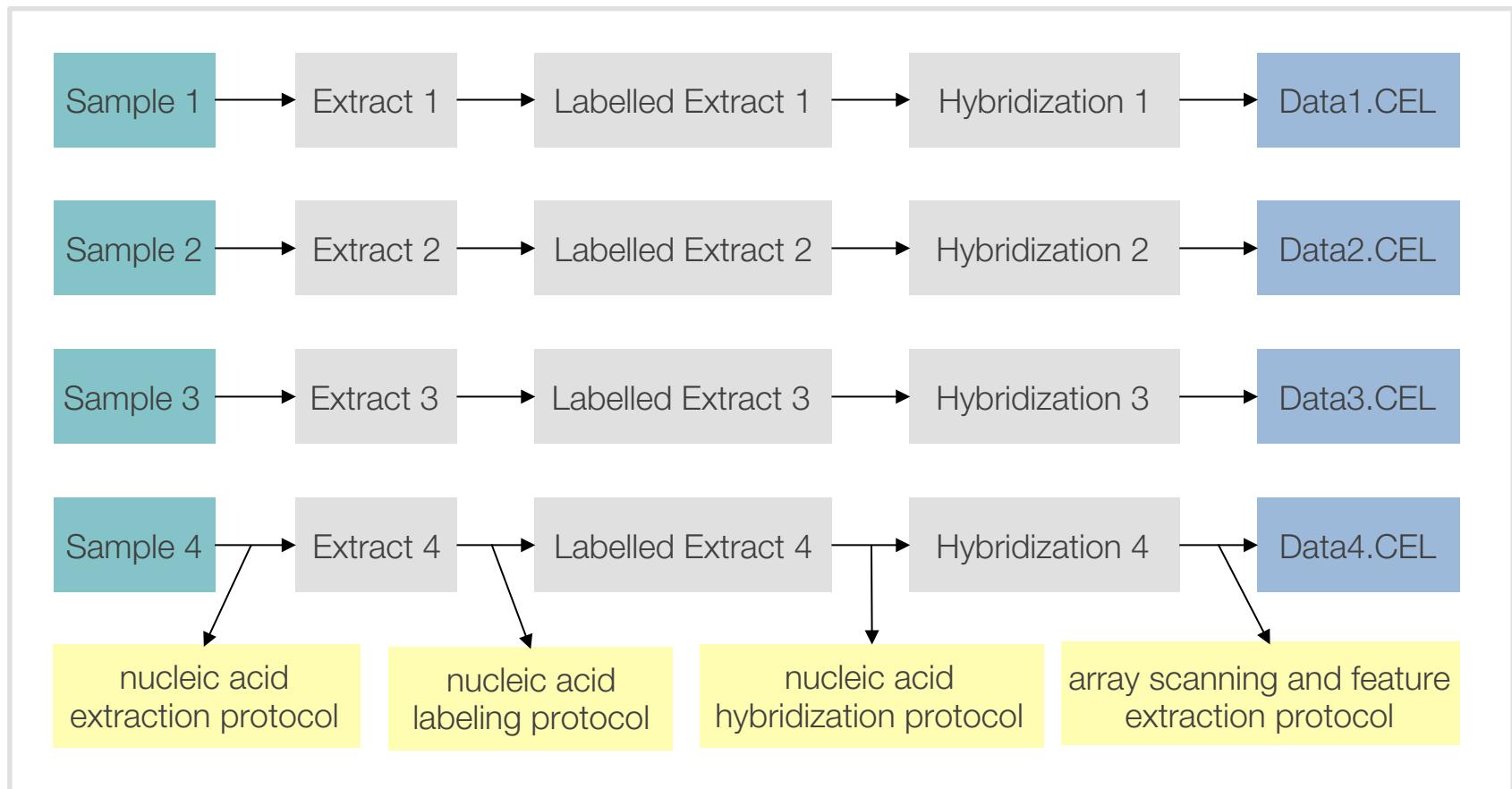
Relationship samples -> files

**IDF** describes the experiment

**SDRF** describes the individual samples and how they relate to the data files



# Metadata is stored in MAGE-TAB format





# Metadata is stored in MAGE-TAB format

SRDF and IDF files

## Hands-on activity

Go to experiment E-GEOD-49515 and download and open the corresponding SRDF and IDF files



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



# Metadata is stored in MAGE-TAB format

## Microarray data

MAGE-TAB specification defines 4 types of files:

1. *Investigation Description Format (IDF)*
2. *Sample and Data Relationship Format (SDRF)*
3. *Array Design Format (ADF)*
4. *Raw and processed data files*



# Metadata is stored in MAGE-TAB format

Array specification in MAGE-TAB format (generated by submitter/curator)

## Array Design Format

Accession: A-AFFY-44

Organism: Homo sapiens

Description: Affymetrix GeneChip Human Genome

Reporter Name	Sequence
200020_at	TTACGT...
200021_at	CGGTAA...

DB reference
NM_007375
NM_005507

Layout of the array

Location of each sequence

Sequence annotation

**ADF** describes the array and lists all probes

[www.ebi.ac.uk/arrayexpress/help/adf\\_submissions\\_overview.html](http://www.ebi.ac.uk/arrayexpress/help/adf_submissions_overview.html)



# Metadata is stored in MAGE-TAB format

Array Design (ADF) file

## Hands-on activity

Go to experiment E-GEOD-49515 and  
download and open the ADF file



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



# Metadata is stored in MAGE-TAB format

## Microarray data

MAGE-TAB specification defines 4 types of files:

1. *Investigation Description Format (IDF)*
2. *Sample and Data Relationship Format (SDRF)*
3. *Array Design Format (ADF)*
4. *Raw and processed data files*

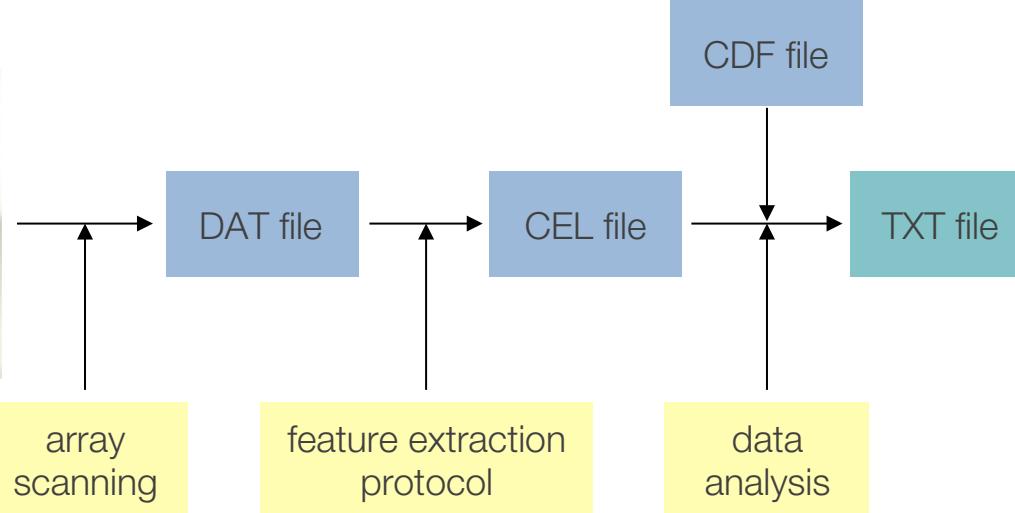


# Raw and processed data files

Microarray data

**Raw data:** "native" files generated by the microarray scanner software.

Hybridised  
GeneChip





# Common microarray raw data files

Manufacturer	Typical raw data format	How to open / Analysis software examples
Affymetrix	.CEL	R packages (affy, limma, oligo...)
Agilent	<i>feature extraction file</i> (tab-delimited text file per hybridisation)	Spreadsheet software (Excel, OpenOffice, etc.)
GenePix (scanner)	.gpr (tab-delimited text file per hybridisation)	Spreadsheet software (Excel, OpenOffice, etc.)
Illumina	.idat	R packages (e.g. illuminaio)
Illumina	.txt (tab-delimited text matrix for all samples)	Spreadsheet software (Excel, OpenOffice, etc.)
Nimblegen	NimbleScan, .pair (tab-delimited text matrix for all samples)	Spreadsheet software (Excel, OpenOffice, etc.)

[www.ebi.ac.uk/fg/annotare/help/accepted\\_raw\\_ma\\_file\\_formats.html](http://www.ebi.ac.uk/fg/annotare/help/accepted_raw_ma_file_formats.html)



# Metadata is stored in MAGE-TAB format

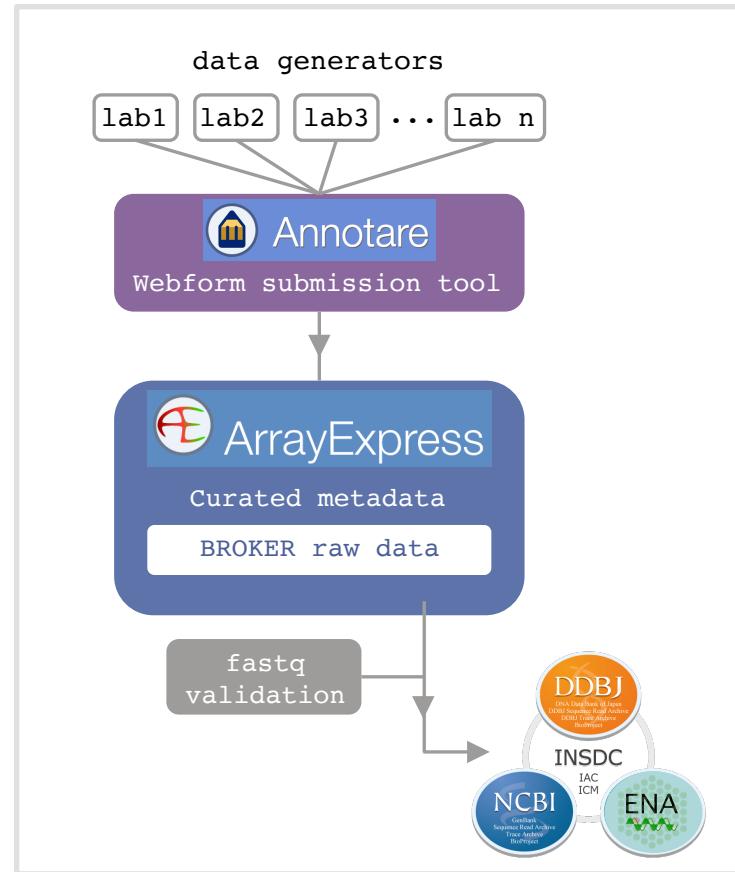
RNA-seq data

MAGE-TAB specification defines 3 types of files:

1. *Investigation Description Format (IDF)*
2. *Sample and Data Relationship Format (SDRF)*
3. *Array Design Format (ADF)*
4. *Raw and processed data files*



# RNA-seq raw data is deposited at ENA





# Raw data – stable archiving at INSDC

- ✓ INSDC (3 partners) runs the sequence read archive (SRA)
- ✓ ArrayExpress brokers raw NGS data to ENA & GEO to NCBI SRA

INSDC – International Nucleotide Sequence Database Collaboration

The slide features a central graphic representing the International Nucleotide Sequence Database Collaboration (INSDC). It consists of three logos arranged around a grayscale world map: the NCBI logo (blue square with a white DNA helix icon), the ENA logo (European Nucleotide Archive, with a green DNA helix icon), and the DDBJ logo (DNA Data Bank of Japan, with an orange DNA helix icon). Below the map, a dark gray rectangular box contains the text "Daily data exchange - mirroring".



# Accession number conventions

## E-ERAD-475 RNA-seq of zebrafish developmental stages

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	ERP014517 at ENA	
-----------------	------------------	----------------------------	----------------------	---------------	--------	-----------------	------------------	------------------	--

PRJEB12982	SAMEA3892004	ERS1079138	ERX1512938	ERR1442561	7955	Danio rerio	Illumina HiSeq 2500	PAIRED	File 1 File 2
PRJEB12982	SAMEA3892005	ERS1079139	ERX1512939	ERR1442562	7955	Danio rerio	Illumina HiSeq 2500	PAIRED	File 1 File 2
PRJEB12982	SAMEA3892006	ERS1079140	ERX1512940	ERR1442563	7955	Danio rerio	Illumina HiSeq 2500	PAIRED	File 1 File 2
PRJEB12982	SAMEA3892007	ERS1079141	ERX1512941	ERR1442564	7955	Danio rerio	Illumina HiSeq 2500	PAIRED	File 1 File 2

ENA objects	Accession examples
study	ERP014517
samples	ERS1079138
experiments (RNA-seq lib)	ERX1512938
runs	ERR1442561

SRA objects	DDBJ	ENA	NCBI
study	DRP	ERP	SRP
sample	DRS	ERS	SRS
experiment	DRX	ERX	SRX
run	DRR	ERR	SRR



# FASTQ format: RNA-seq raw data files

FASTQ file = FASTA + Quality

A FASTQ file is the common file format for sharing sequencing read data combining both the sequence and an associated per base quality score.

```
@SRR014849.1 EIXKN4201CFU84 length=93
GGGGGGGGGGGGGGGGCTTTTTGTTTGGAACCGAAAGG
GTTCGAATTCAAACCCCTTCGGTTCCAACCTTCAA
AGCAATGCCAATA
+SRR014849.1 EIXKN4201CFU84 length=93
3+&#""""""""7F@71,'";C?,B;?6B;:EA1EA
1EA5'9B:?:#9EA0D@2EA5':>5?:%A;A8A;?9B;D@
/=<?7=9<2A8==
```

*@title and optional description  
sequence line(s)  
+optional repeat of title line  
quality line(s)*

For each read:

1. @ Read ID
2. Nucleotide sequence of the read
3. +
4. Quality score for each nucleotide of the read

[en.wikipedia.org/wiki/FASTQ\\_format](https://en.wikipedia.org/wiki/FASTQ_format)



# Single-end and paired-end sequencing

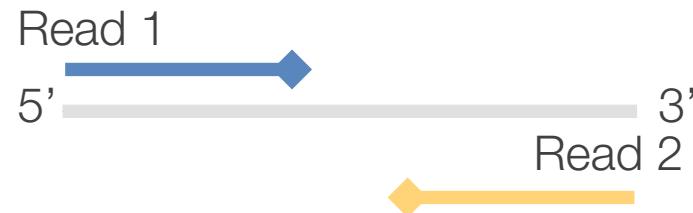
## Single-end sequencing



**my\_sequence.fastq**

```
@HWI-BRUNOP16X_0001:1:1:1466:1018#0/1  
AAGGAAGTGCTTGTCTGGCTAACACAGCNAGNCACGTGAC  
+  
aVfbe`^__^_TTTSSdfffffdfffabbZbbfebafbbbbbb
```

## Paired-end sequencing



**my\_sequence\_1.fastq**

```
@HWI-BRUNOP16X_0001:1:1:1278:989#0/1  
NAAATTCGAATTCTGTGAAGTAAGCATCTTCTTGTCA  
+  
BJJGGKIIINN^^^^^QQNTUQOOTTTRTOTY^^Y^\\^____
```

**my\_sequence\_2.fastq**

```
@HWI-BRUNOP16X_0001:1:1:1278:989#0/2  
AACCCACACAGGAGAGCAGCCTACAGATGCAAATACTGTG  
+  
]K___fffffggghgeggggggdggggggfgggggeggggghh
```



# ArrayExpress: Experimental variable

Experimental variable



It is the main factor that you are investigating

Transcription profiling of blood from smokers, non-smokers and former smokers to identify gene expression signature for cigarette smoke exposure response

organism > Homo sapiens

age > 22, 57, 43, 39 year, etc.

sex > male OR female

organism part > blood

disease > normal

ethnic group > Caucasian OR African American

clinical history > smoker, former smoker OR non-smoker



# ArrayExpress: Experimental variable

Experimental variable



It is the main factor that you are investigating

Transcription profiling of blood from smokers (with or without COPD), non-smokers and former smokers to identify gene expression signature for cigarette smoke exposure response

organism > Homo sapiens

age > 22, 57, 43, 39 year, etc.

sex > male OR female

organism part > blood

ethnic group > Caucasian OR African American

disease > normal OR COPD

clinical history > smoker, former smoker OR non-smoker



# ArrayExpress: Experimental variable

MIAME & MAGE-TAB

## Hands-on activity

Go to experiment E-GEO-49515, find the experimental variable and the groups of samples compared. Download the SDRF file and figure out which file belongs to each category



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



# ArrayExpress: Experimental variable

	A	B	C
1	Source Name	Array Data File	FactorValue [disease]
2	GSM1200316 1	GSM1200316_PAN-PBMC-S03.CEL	pancreatic cancer
3	GSM1200315 1	GSM1200315_PAN-PBMC-S02.CEL	pancreatic cancer
4	GSM1200314 1	GSM1200314_PAN-PBMC-S01.CEL	pancreatic cancer
5	GSM1200313 1	GSM1200313_NOR-PBMC-S10.CEL	normal
6	GSM1200312 1	GSM1200312_NOR-PBMC-S09.CEL	normal
7	GSM1200311 1	GSM1200311_NOR-PBMC-S08.CEL	normal
8	GSM1200310 1	GSM1200310_NOR-PBMC-S07.CEL	normal
9	GSM1200309 1	GSM1200309_NOR-PBMC-S06.CEL	normal
10	GSM1200308 1	GSM1200308_NOR-PBMC-S05.CEL	normal
11	GSM1200307 1	GSM1200307_NOR-PBMC-S04.CEL	normal
12	GSM1200306 1	GSM1200306_NOR-PBMC-S03.CEL	normal
13	GSM1200305 1	GSM1200305_NOR-PBMC-S02.CEL	normal
14	GSM1200304 1	GSM1200304_NOR-PBMC-S01.CEL	normal
15	GSM1200303 1	GSM1200303_HCC-PBMC-S10.CEL	hepatocellular carcinoma
16	GSM1200302 1	GSM1200302_HCC-PBMC-S09.CEL	hepatocellular carcinoma
17	GSM1200301 1	GSM1200301_HCC-PBMC-S08.CEL	hepatocellular carcinoma
18	GSM1200300 1	GSM1200300_HCC-PBMC-S07.CEL	hepatocellular carcinoma
19	GSM1200299 1	GSM1200299_HCC-PBMC-S06.CEL	hepatocellular carcinoma
20	GSM1200298 1	GSM1200298_HCC-PBMC-S05.CEL	hepatocellular carcinoma
21	GSM1200297 1	GSM1200297_HCC-PBMC-S04.CEL	hepatocellular carcinoma
22	GSM1200296 1	GSM1200296_HCC-PBMC-S03.CEL	hepatocellular carcinoma
23	GSM1200295 1	GSM1200295_HCC-PBMC-S02.CEL	hepatocellular carcinoma
24	GSM1200294 1	GSM1200294_HCC-PBMC-S01.CEL	hepatocellular carcinoma
25	GSM1200293 1	GSM1200293_GAS-PBMC-S03.CEL	gastric cancer
26	GSM1200292 1	GSM1200292_GAS-PBMC-S02.CEL	gastric cancer
27	GSM1200291 1	GSM1200291_GAS-PBMC-S01.CEL	gastric cancer

# *Outline of the session*

- ✓ What is ArrayExpress?
  - ✓ Data standards: MIAME & MINSEQE
  - ✓ Data format: MAGE-TAB
    - ✓ Hands-on exercise
  - ✓ Experimental variable
    - ✓ Hands-on exercise
- ✓ Data mining with ArrayExpress
  - ✓ Hands-on exercise



# Data mining with ArrayExpress

## ArrayExpress

Home Browse Submit Help About ArrayExpress

Search Examples: E-MEXP-31, cancer, p53, Geuvadis advanced search

Contact Us Login

### ArrayExpress – functional genomics data

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

[Browse ArrayExpress](#)

#### Latest News

17 January 2018 - **Start early: Build up your ArrayExpress submission with Annotare as your electronic "lab book"**

If you are doing functional genomics research, [sign up now for an Annotare account](#), and start building an ArrayExpress submission piece by piece as you go through the stages of your experiment. Having your data already archived, will save you time and stress when finalising your paper submission.

Many journals now require accession numbers and access to data at the time of manuscript submission. Annotare provides you with an accession number usually within 24 hours after submission. However, we can only give out reviewer access accounts after the datasets have been curated and approved for deposition in ArrayExpress. The curation team will do their best to ensure timely curation and loading of your data into ArrayExpress. Keep in mind that the processing can take 10 working days or longer in busy times.

## Data Content

Updated today at 03:00

- 70894 experiments
- 2258250 assays
- 46.37 TB of archived data

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



# Data mining with ArrayExpress

ArrayExpress

Search Examples: E-MEXP-31, cancer, p53, Geuvadis advanced search

Home Browse Submit Help About ArrayExpress Contact Us Login

Filter search results

Page 1 2 3 4 5 6 .. 2840 Showing 1 - 25 of 70979 experiments

Sortable headings

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-6777	Haploid genetic screen for components of MLKL-mediated cell death	DNA-seq	Homo sapiens	2	Today	-		-	-
E-MTAB-6755	RNA-seq analysis of Arabidopsis sulki1, Kas-2 and Ler/Kas-2 near isogenic line (NIL)	RNA-seq of coding RNA	Arabidopsis thaliana	9	Today	-		-	-
E-MTAB-6605	RNA-Seq of human aortic endothelial cells with lysophosphatidylinositol against untreated controls	RNA-seq of coding RNA	Homo sapiens	6	Today	-		1	-
E-MTAB-6604	RNA-Seq of human aortic endothelial cells with lysophosphatidylcholine against untreated controls	RNA-seq of coding RNA	Homo sapiens	2	Today	-		1	-
E-MTAB-5776	Microarray transcription profiling of healthy and PVYNTN infected potato tuber tissues (necrotic and non-necrotic)	transcription profiling by array	Solanum tuberosum	16	Today			-	-

Accession number

Experiment type

Number of hybridisations, sequencing libraries



# Data mining with ArrayExpress

Browse ArrayExpress

## Hands-on activity

Find the biggest experiments (in terms of number of assays). Now display the top 100 most viewed experiments on one page



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



# Data mining with ArrayExpress

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments related to human autoimmune disease*

main search box

filter search box

The screenshot shows the ArrayExpress homepage. At the top right is a search bar with a placeholder 'Search' and a magnifying glass icon. Below it is an 'advanced search' link. A red box highlights the search bar area. At the bottom left, there is a yellow button labeled 'Filter search results' with a funnel icon. A red box highlights this button. The navigation bar at the top includes links for Home, Browse (which is currently selected), Submit, Help, About ArrayExpress, Contact Us, and Login.

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



# Data mining with ArrayExpress

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments related human autoimmune disease*

Filter search results

By organism:

Homo sapiens

By experiment type:

RNA assay

Sequencing assay

By array:

All arrays

ArrayExpress data only

Reset filters

Filter

"autoimmune dis

autoimmune disease

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

Efficient search via  
ontology-driven query  
expansion



# Ontology-driven query expansion

**Ontology Lookup Service**

Home | **Ontologies** | Documentation | About

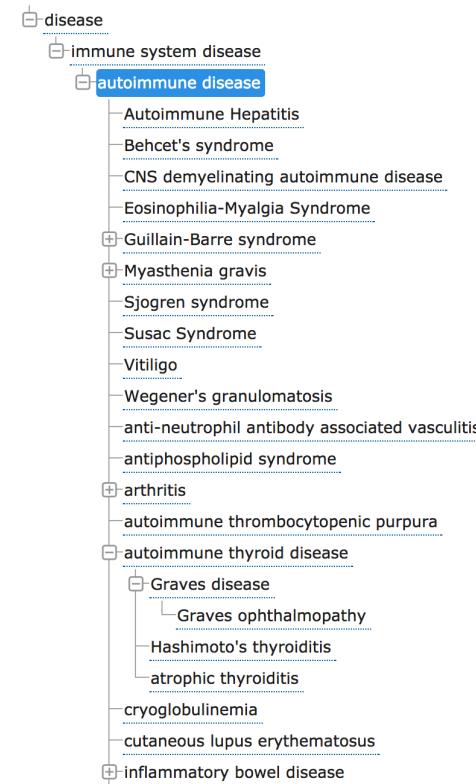
OLS > Experimental Factor Ontology > EFO:0005140

## autoimmune disease

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

Much richer queries by  
using the hierarchy within  
the ontology

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



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



# Data mining with ArrayExpress

Data mining: Discover datasets to analyse yourself

*Find RNA-seq experiments related human autoimmune disease*

Search results for "autoimmune disease"

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

Page **1** 2

Showing 1 - 25 of 39 experiments

Page size **25** 50 100 250 500

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views 	Atlas
E-GEO-60424	Next generation sequencing of human immune cell subsets across diseases	RNA-seq of coding RNA	Homo sapiens	134	06/01/2015	-	 	3631	
E-MTAB-1568	RNA-seq of coding RNA in human plasma cells isolated from individuals with colorectal cancer, ulcerative colitis and normal individuals	RNA-seq of coding RNA	Homo sapiens	9	29/03/2013	-	 	1140	-
E-GEO-46579	A blood based 12-miRNA signature of Alzheimer patients	RNA-seq of non coding RNA	Homo sapiens	70	17/06/2013	-	 	922	-
E-MTAB-4304	RNA-Seq analysis of human intact and damaged osteoarthritic cartilage following total knee replacement	RNA-seq of coding RNA	Homo sapiens	22	06/04/2016	-	 	773	-
E-GEO-57945	RNA-seq of 359 treatment-naive pediatric patients with Crohn's disease, patients with ulcerative colitis and control individuals	RNA-seq of coding RNA	Homo sapiens	359	24/07/2014		 	744	-
E-GEO-83139	Single cell RNA-seq of human pancreatic endocrine cells from Juvenile, adult control and type 2 diabetic donors	RNA-seq of coding RNA	Homo sapiens	635	05/07/2016	-	 	424	-



# Data mining with ArrayExpress

Samples  
and files

Experiment  
description

Experiment  
type

MINSEQE  
compliance score

Links to  
data

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

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

Organism Homo sapiens

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

Protocols (2) Click for detailed protocol information

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

Experiment type RNA-seq of coding RNA

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

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

MINSEQE \* \* \* - \*

Exp. design Protocols Variables Processed Seq. reads

Files

Investigation description  
Sample and data relationship  
Additional data (1)

↳ E-GEO-60424.idf.txt

↳ E-GEO-60424.sdrf.txt

↳ E-GEO-60424.additional.1.zip

Links

Expression Atlas - E-GEO-60424

ENA - SRP045500, GEO - GSE60424

Send E-GEO-60424 data to GENOME SPACE

MAGET-TAB  
format



# Data mining with ArrayExpress

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

[Display full sample-data table](#)

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

Page 1 2 3 4 5 6 .. 11

Showing 1 - 25 of 268 rows

Page size 25 50 100 250 500

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

Unique sample name for each biological replicate

Sample attributes describing the source material

Experimental variables

Links to download data



# Data mining with ArrayExpress

Data mining: Discover datasets to analyse yourself

*Find RNA expression arrays from human patients with diabetes*

## Search results for diabetes

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

Page 1 2 3 4 5 6 .. 11

Showing 1 - 25 of 255 experiments

Page size 25 50 100 250 500

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-2976	Global gene expression of bone marrow multipotent mesenchymal stromal cells isolated from <b>type 1 diabetes</b> patients and healthy donors	transcription profiling by array	<b>Homo sapiens</b>	8	01/12/2014		-	272	-
E-MTAB-2902	Transcriptional profiling by array of microRNAs of human peripheral blood mononuclear cells - PBMC - from control individuals for <b>Type 2 diabetes mellitus</b> patients	microRNA profiling by array	<b>Homo sapiens</b>	9	01/04/2015		255	-	
E-MTAB-2899	Transcriptional profiling by array of microRNAs from human peripheral blood mononuclear cells - PBMC - of <b>Type 2 diabetes mellitus</b> patients	microRNA profiling by array							
E-MTAB-2896	Transcriptional Profiles in Peripheral Blood Mononuclear Cells - PBMC - from individual controls for <b>Type 2 Diabetes Mellitus</b>	transcription profiling by array							

Exact match to search term

Matched EFO synonyms to search term

Matched EFO child term of search term



# ArrayExpress – Advanced search

diabetes

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

advanced search

Search in all experimental fields,  
e.g. experiment description,  
protocols, publication title...

evv:diabetes

advanced search

fieldname:value

Limit your search only to  
experiments in which “diabetes” is  
the value of the experimental variable

## Search results for evv:diabetes

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

3 experiments

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-5061	Single-cell RNA-seq analysis of human pancreas from healthy individuals and type 2 diabetes patients	RNA-seq of coding RNA from single cells	Homo sapiens	3514	22/09/2016			3091	
E-MTAB-5060	Whole-islet RNA-sequencing analysis of human pancreas from healthy individuals and type 2 diabetes patients	RNA-seq of coding RNA	Homo sapiens	7	22/09/2016			903	
E-GEO-60424	Next generation sequencing of human immune cell subsets across diseases	RNA-seq of coding RNA	Homo sapiens	134	06/01/2015	-		3163	

Export table in Tab-delimited format

Export matching metadata in XML format

Subscribe to RSS feed matching this search



# ArrayExpress – Advanced search

Field name	Search scope	Example use case
<b>accession</b>	Experiment primary or secondary accession	accession:E-MTAB-1234
<b>array</b>	Array design accession or name	array:A-AFFY-33
<b>ev (or ef)</b>	Experimental variable (or factor), the name of the main variable under study in an experiment. E.g. if the variable is "sex" in a human study, the researchers would be comparing between male and female samples, and "sex" is not merely an attribute the samples happen to have. Has <a href="#">EFO expansion</a> .	ev:genotype
<b>evv (or efv)</b>	The value of an experimental variable (or factor). E.g. The values for "genotype" factor can be "wild type", "p53-/-". Has <a href="#">EFO expansion</a> .	evv:"wild type"
<b>expdesign</b>	Experiment design type, related to the questions being addressed by the study, e.g. "time series design", "stimulus or stress design", "genetic modification design". Has <a href="#">EFO expansion</a> .	expdesign:"time series"
<b>exptype</b>	Experiment type, related to the assay technology used. See the full <a href="#">list of experiment types in ArrayExpress</a> . Has <a href="#">EFO expansion</a> .	exptype:"RNA-seq of coding RNA"
<b>gxa</b>	Presence/absence of an ArrayExpress experiment in the <a href="#">Expression Atlas</a> . Use values "true" and "false" respectively.	gxa:true
<b>pmid</b>	PubMed identifier for a publication.	pmid:16553887
<b>sa</b>	Sample attribute values. Has <a href="#">EFO expansion</a> .	sa:fibroblast
<b>sac</b>	Sample attribute category. Find experiments that have a specific sample attribute defined, e.g. "age", "strain". Has <a href="#">EFO expansion</a> .	sac:age
<b>organism</b>	Species of the samples. Can use common name (e.g. "mouse") or binomial nomenclature/Latin names (e.g. "Mus musculus"). Has <a href="#">EFO expansion</a> .	organism:"homo sapiens"



# ArrayExpress – Advanced search

Data mining: Discover datasets to analyse yourself

*Find experiments comparing healthy (normal) and rheumatoid arthritis patients*

evv: "rheu

**rheumatic**

**rheumatic disease**

**rheumatic fever**

**rheumatic heart disease**

**Rheumatic Nodule**

**rheumatoid**

**rheumatoid arthritis**

Felty's syndrome

chronic childhood arthritis

pauciarticular juvenile **rheumatoid arthritis**

systemic juvenile idiopathic arthritis

**rheumatoid factor measurement**

**rheumatoid factor seropositivity measurement**

**rheumatology**

Search results for evv: "rheumatoid arthritis"

Filtered by organism **Homo sapiens**

Showing 1 - 25 of 51 experiments



# ArrayExpress – Advanced search

Data mining: Discover datasets to analyse yourself

*Find experiments comparing healthy (normal) and rheumatoid arthritis patients*

Search results for evv: "rheumatoid arthritis" AND evv:normal

Filtered by organism Homo sapiens

21 experiments

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

Filtered by organism Homo sapiens

Showing 1 - 25 of 29 experiments



# ArrayExpress – Advanced search

Data mining: Discover datasets to analyse yourself

*Find RNA expression arrays from human cancer samples  
excluding all experiments performed on cell lines*

ArrayExpress

evv: "cancer" AND (evv:normal OR evv:healthy) NOT sac:  
Examples: E-MEXP-31, cancer, p53, Geuvadis [advanced search](#)

Home Experiments Arrays Protocols Files Users Submit Help About ArrayExpress Contact Us [Logout \[curator\]](#)

[Filter search results](#) [Show more data from EMBL-EBI](#)

Search results for evv: "cancer" AND (evv:normal OR evv:healthy) NOT sac:"cell line"

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

Page **1** 2 3 4 5 6 .. 29 Showing 1 - 25 of 717 experiments Page size 25 50 100 250 500



# Let's try ArrayExpress

Browsing ArrayExpress

## Hands-on activity

Discover interesting datasets related to  
your research area



In pairs

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



# Let's try ArrayExpress

 **ArrayExpress**

"Barrett's esophagus" AND ev:"disease staging" 

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

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

 Filter search results  Show more data from EMBL-EBI

Search results for "Barrett's esophagus" AND ev:"disease staging"

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

1 experiment

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-MTAB-4054	Whole transcriptome profiling of Esophageal adenocarcinoma and Barrett's	RNA-seq of coding RNA	Homo sapiens	63	23/07/2017	-		177	-

 Export table in Tab-delimited format  Export matching metadata in XML format  Subscribe to RSS feed matching this search



# Let's try ArrayExpress

## E-MTAB-4054 - Whole transcriptome profiling of Esophageal adenocarcinoma and Barrett's

Status	<i>Submitted on 11 November 2015, released on 23 July 2017, last updated yesterday</i>
Organism	Homo sapiens
Samples (51)	<a href="#">Click for detailed sample information and links to data</a>
Protocols (4)	<a href="#">Click for detailed protocol information</a>
Description	RNA-seq was performed on esophageal adenocarcinoma (EAC), Barrett's without dysplasia, Barrett's with low-grade dysplasia (LGD) and normal squamous esophagus tissue to find early alterations in the transcriptome level turning Barrett's dysplastic.
Experiment types	RNA-seq of coding RNA, disease state design
Contact	<a href="mailto:j.maag@garvan.org.au">✉ Jesper LV Maag &lt;j.maag@garvan.org.au&gt;</a>
Citation	<a href="#">Novel Aberrations Uncovered in Barrett's Esophagus and Esophageal Adenocarcinoma Using Whole Transcriptome Sequencing</a> . Maag J, Fisher OM, Levert-Mignon AJ, Kaczorowski DC, Thomas ML, Hussey D, Watson D, Wettstein A, Bobryshev YV, Edwards M, Dinger ME, Lord RV, , Europe PMC 28751461
MINSEQE	— — — —
	Exp. design   Protocols   Variables   Processed   Seq. reads
Files	Investigation description <a href="#"> E-MTAB-4054.idf.txt</a> Sample and data relationship <a href="#"> E-MTAB-4054.sdrf.txt</a> <a href="#">Click to browse all available files</a>
Links	<a href="#">ENA - ERP013206</a> <a href="#">Send E-MTAB-4054 data to  GENOMESPACE</a>



# Let's try ArrayExpress

E-MTAB-4054 - Whole transcriptome profiling of Esophageal adenocarcinoma and Barrett's

Display summary				Export table in Tab-delimited format		
Source Name	Protocol REF	Performer	Assay Name	Comment[technical replicate group]	Technology Type	Comment[ENA_EX]
N09	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N09_R		sequencing assay	ERX1220581
N10	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N10_R		sequencing assay	ERX1220582
N10	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N10_R		sequencing assay	ERX1220582
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_Human_CCGTCC_R	group1	sequencing assay	ERX1220524
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_Human_CCGTCC_R	group1	sequencing assay	ERX1220524
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_R	group1	sequencing assay	ERX1220525
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_R	group1	sequencing assay	ERX1220525
LNC14	P-MTAB-47385	Dominik C. Kaczorowski	LNC_14_Human_GTGAAA_R	group10	sequencing assay	ERX1220566
LNC14	P-MTAB-47385	Dominik C. Kaczorowski	LNC_14_Human_GTGAAA_R	group10	sequencing assay	ERX1220566
				group10	sequencing assay	ERX1220567
				group10	sequencing assay	ERX1220567
				group11	sequencing assay	ERX1220575
				group11	sequencing assay	ERX1220575
				group11	sequencing assay	ERX1220576
				group11	sequencing assay	ERX1220576
				group12	sequencing assay	ERX1220579
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_Human_CCGTCC_R	group12	sequencing assay	ERX1220579
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_R	group12	sequencing assay	ERX1220580
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_R	group12	sequencing assay	ERX1220580
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_Human_ATGTCA_R	group2	sequencing assay	ERX1220526
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_Human_ATGTCA_R	group2	sequencing assay	ERX1220526
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_R	group2	sequencing assay	ERX1220527
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_R	group2	sequencing assay	ERX1220527
B07	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B07_Human_GATCAG_R	group3	sequencing assay	ERX1220528
B07	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B07_Human_GATCAG_R	group3	sequencing assay	ERX1220528

12 samples (2 normal, 5 Barrett's esophagus without dysplasia and 5 esophageal adenocarcinoma)



# Let's try ArrayExpress

E-MTAB-4054 - Whole transcriptome profiling of Esophageal adenocarcinoma and Barrett's

Display summary		Showing 76 - 100 of 126 rows			Export table in Tab-delimited format	
Source Name	Protocol REF	Performer	Assay Name	Comment[technical replicate group]	Technology Type	Comment[ENA_EX]
N09	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N09_R		sequencing assay	ERX1220581
N10	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N10_R		sequencing assay	ERX1220582
N10	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N10_R		sequencing assay	ERX1220582
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_Human_CCGTCC_R	group1	sequencing assay	ERX1220524
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_Human_CCGTCC_R	group1	sequencing assay	ERX1220524
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_R	group1	sequencing assay	ERX1220525
B05	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B05_R	group1	sequencing assay	ERX1220525
LNC14	P-MTAB-47385	Do			sequencing assay	ERX1220566
LNC14	P-MTAB-47385	Do			sequencing assay	ERX1220566
LNC14	P-MTAB-47385	Do			sequencing assay	ERX1220567
LNC14	P-MTAB-47385	Dominik C. Kaczorowski	LNC_14_R	group10	sequencing assay	ERX1220567
N04	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N04_Human_AGTTCC_R	group11	sequencing assay	ERX1220575
N04	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N04_Human_AGTTCC_R	group11	sequencing assay	ERX1220575
N04	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N04_R	group11	sequencing assay	ERX1220576
N04	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N04_R	group11	sequencing assay	ERX1220576
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_Human_CCGTCC_R	group12	sequencing assay	ERX1220579
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_Human_CCGTCC_R	group12	sequencing assay	ERX1220579
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_R	group12	sequencing assay	ERX1220580
N08	P-MTAB-47385	Dominik C. Kaczorowski	LNC_N08_R	group12	sequencing assay	ERX1220580
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_Human_ATGTCA_R	group2	sequencing assay	ERX1220526
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_Human_ATGTCA_R	group2	sequencing assay	ERX1220526
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_R	group2	sequencing assay	ERX1220527
B06	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B06_R	group2	sequencing assay	ERX1220527
B07	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B07_Human_GATCAG_R	group3	sequencing assay	ERX1220528
B07	P-MTAB-47385	Dominik C. Kaczorowski	LNC_B07_Human_GATCAG_R	group3	sequencing assay	ERX1220528



# Let's try ArrayExpress

Disease cohort	Number of samples	Number of technical replicates
normal	17	2
Barrett's esophagus non-dysplastic	14	5
Barrett's esophagus low-grade dysplasia	8	0
esophageal adenocarcinoma	12	5
<b>Total number of samples</b>	<b>51</b>	12
Number of files (paired-end libraries)	102	24
<b>Total number of files</b>	<b>126</b>	



# Learn more about ArrayExpress

Train online

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

ArrayExpress: Discover functional genomics data quickly and easily

Overview

What is ArrayExpress?

How are datasets submitted and stored in ArrayExpress?

How to search ArrayExpress

Experiment and Sample information

Files and Download

Next steps (towards data analysis)

Summary

Quiz: ArrayExpress Quiz

Your feedback

Learn more

Get help and support on ArrayExpress

References

Contributors



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

ArrayExpress: 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.

## About this course

Author(s): Anja Füllgrabe

Gene Expression

Beginner

2 hours

[Start the course](#)

## Learning objectives:

- Describe what ArrayExpress is and when to use it
- Evaluate how functional genomics data is stored within ArrayExpress
- Search ArrayExpress to find information on functional genomics data
- Retrieve/download data from ArrayExpress

## Rating:



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

# EMBL-EBI workshop: Tools for functional genomics data

Discover functional genomics data  
with ArrayExpress

Laura Huerta, PhD  
Senior Scientific Curator  
[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)  
24 May 2018

