ArrayExpress and Gene Expression Atlas: Mining Functional Genomics data

Emma Hastings
Functional Genomics Team
EBI-EMBL

emma@ebi.ac.uk
http://www.ebi.ac.uk/~emma/bcn_2012/
Talk structure

- Why do we need a database for functional genomics data?
- ArrayExpress database
  - Archive
  - Gene Expression Atlas
What is functional genomics?

- Functional genomics (FG) is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions.

- The aim of FG is to understand the **function of genes and other parts of the genome**

- High-throughput technologies such as **microarrays and high-throughput sequencing (HTS)** are frequently used in this field to interrogate the transcriptome
What biological questions is FG addressing?

• When and where are genes expressed?
• How do gene expression levels differ in various cell types and states?
• What are the functional roles of different genes and in what cellular processes do they participate?
• How are genes regulated?
• How do genes and gene products interact?
• How is gene expression changed in various diseases or following a treatment?
Components of a FG experiment

**Microarrays**
- Sample Prep: Convert to cDNA by RT and random priming and Label
- Procedure: Hybridization
- Acquisition: Scanning by laser
- Data: Raw: Image files (a few megabytes in size) with fluorescent intensities
  Processed: Text files with intensities of gene expression

**mRNA-Seq**
- Sample Prep: Fragment RNA
- Procedure: Ligate Adapters
- Cluster (solid phase clonal amplification)
- Acquisition: Sequence by Synthesis
- Data: Raw: FASTQ files of millions of sequences (1-10s of Gigabytes / lane)
  Processed: Mapping, Junction Detection, Quantification
Why do we need a database for functional genomics data?

**E-MEXP-2451**

**Transcription profiling of grape berries during ripening**

Grape berries undergo considerable physical and biochemical changes during the ripening process. Ripening is characterized by a number of changes, including the degradation of chlorophyll, an increase in berry deformability, a rapid increase in the level of hexoses in the berry vacuole, an increase in berry volume, the catabolism of organic acids, the development of skin colour, and the formation of compounds that influence flavour, aroma, and therefore, wine quality. **The aim of this work is to identify differentially expressed genes during grape ripening by microarray and real-time PCR techniques.** Using a custom array of new generation, we analysed the expression of 6000 grape genes from pre-veraison to full maturity, in *Vitis vinifera* cultivar Muscat of Hamburg, in two different years (2006 and 2007). Five time points per year and two biological replicates per stadium were considered. To reduced intra-plant and inter-plant biological variability, for each ripening stadium we collected around hundred berries from several bunch grapes of five plants of *V. vinifera* cv Muscat of Hamburg. We will use the real-time PCR technique to validate microarray data. Muscat of Hamburg. We will use the real-time PCR technique to validate microarray data.
Why do we need a database for functional genomics data?
ArrayExpress
www.ebi.ac.uk/arrayexpress/

• Is a public repository for FG data
• Provides easy access to well annotated data in a structured and standardized format
• Serves the scientific community as an archive for data supporting publications, together with GEO at NCBI and CIBEX at DDBJ
• Facilitates the sharing of experimental information associated with the data such as microarray designs, experimental protocols
• Based on community standards: MIAME guidelines & MAGE-TAB format for microarray, MINSEQE guidelines for HTS data (http://www.mged.org/minseqe/)
Data standardization efforts focus on answering 3 major questions:

1. What information do we need to capture?
2. What syntax (or file format) should we use to exchange data?
3. What semantics (or ontology) should we use to best describe its annotation?
What information do we need to capture?

MIAME checklist

- Minimal Information About a Microarray Experiment

- The 6 most critical elements contributing towards MIAME are:

1. Essential **sample annotation** including experimental factors and their values (e.g. compound and dose)
2. **Experimental design** including sample data relationships (e.g. which raw data file relates to which sample, ….)
3. Sufficient **array annotation** (e.g. gene identifiers, genomic coordinates, probe sequences or array catalog number)
4. Essential laboratory and data processing **protocols** (e.g. normalization method used)
5. **Raw data** for each hybridization (e.g. CEL or GPR files)
6. **Final normalized data** for the set of hybridizations in the experiment
What information do we need to capture?
MINSEQE checklist

- **Minimal Information about a high-throughput Nucleotide SEQuencing Experiment**

- The proposed guidelines for MINSEQE are (still work in progress):
  1. General information about the experiment
  2. Essential sample annotation including experimental factors and their values (e.g. compound and dose)
  3. Experimental design including sample data relationships (e.g. which raw data file relates to which sample, ….)
  4. Essential experimental and data processing protocols
  5. Sequence read data with quality scores, raw intensities and processing parameters for the instrument
  6. Final processed data for the set of assays in the experiment
What syntax (or file format) should we use to exchange data?

MAGE-TAB is a simple spreadsheet format that uses a number of different files to capture information about an experiment:

<table>
<thead>
<tr>
<th>File Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDF</td>
<td><strong>Investigation Description Format</strong> file, contains top-level information about the experiment including title, description, submitter contact details and protocols.</td>
</tr>
<tr>
<td>SDRF</td>
<td><strong>Sample and Data Relationship Format</strong> file contains the relationships between samples and arrays, as well as sample properties and experimental factors, as provided by the data submitter.</td>
</tr>
<tr>
<td>ADF</td>
<td><strong>Array Design Format</strong> file, describes the design of an array, i.e. the sequence located at each feature on the array and annotation of the sequences.</td>
</tr>
<tr>
<td>Data files</td>
<td><strong>Raw and processed data</strong> files. The ‘raw’ data files are the files produced by the microarray image analysis software, such as CEL files for Affymetrix or GPR files from GenePix. The processed data file is a ‘data matrix’ file containing processed values, as provided by the data submitter.</td>
</tr>
</tbody>
</table>
What semantics (or ontology) should we use to best describe its annotation?

- **Ontology**, which is a formal specification of terms in a particular subject area and the relations among them.

- Its purpose is to provide a basic, stable and unambiguous description of such terms and relations in order to avoid improper and inconsistent use of the terminology pertaining to a given domain.
Experimental factor ontology (EFO)
http://www.ebi.ac.uk/efo

- Application focused ontology modeling experimental factors (EFs) in AE
- Developed to:
  - Consistent curation – ensure that curators use same vocabulary
  - Query support (e.g., query for 'cancer' and get also ‘leukemia’)
- EFs are transformed into an ontological representation, forming classes and relationships between those classes
- EFO terms map to multiple existing domain specific ontologies, such as the Disease Ontology and Cell Type Ontology
Experimental Factor Example:

Experiment Factor: Growth condition
Experimental Factor Values: Flight vs Ground
Experimental factor ontology (EFO)

Example:

- acinar cell carcinoma
- adrenocortical carcinoma
- bladder carcinoma
- breast carcinoma
- cervical carcinoma
- esophageal carcinoma
- follicular thyroid carcinoma
- hepatocellular carcinoma
- pancreatic carcinoma
- papillary thyroid carcinoma
- renal carcinoma
- signet ring cell carcinoma
- uterine carcinoma
- adenocarcinoma
- adenoid cystic carcinoma
- endometrioid carcinoma
- gastric carcinoma
- hereditary leiomyomatosis and renal cell cancer
- hypopharyngeal carcinoma
Experimental factor ontology (EFO)

Example:
Searching ArrayExpress Archive
Simple query - EFO

- Matches to exact terms are highlighted in yellow
- Matches to child terms in the EFO are highlighted in pink
ArrayExpress – two databases
Browsing the ArrayExpress Archive

The ArrayExpress Archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MISEQE standards. Gene Expression Atlas contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.

Experiments Archive
30142 experiments, 868961 assays
Experiment, citation, sample and factor annotations
Browse experiments | platform designs | protocols
Query
Submitter/reviewer login
ArrayExpress Query Help

Gene Expression Atlas
3212 experiments, 92523 assays, 20018 conditions
Genes
Conditions
Gene Expression Atlas Home

News
• 25 Apr 2012 - ArrayExpress Archive interface updated
  We have released an update to the ArrayExpress Archive interface. The main change is in the layout of the experiment detail view - it is easier for users interested in sample information to find the SDRF viewer page. We have removed sample attribute and experimental factor value lists from the top page, making the experiment overview more compact and retaining only the most essential information.
• 21 Feb 2012 - New training course
  ArrayExpress: Submitting data using MAGE-TAB

Links
• ArrayExpress User Survey
• Help | Training | FAQ | Citing
• Submit Data (array based and re-sequencing)
• Programmatic Access | FTP Access
• Software Downloads
• EFO | Bioconductor Package | Quality Metrics
• ArrayExpress Scientific Advisory Board
• Funding
• Functional Genomics Group
ArrayExpress Archive – when to use it?

• Find FG experiments that might be relevant to your research

• Download data and re-analyze it.

  Often data deposited in public repositories can be used to answer different biological questions from the one asked in the original experiments.

• Submit microarray or HTS data that you want to publish.

  Major journals will require data to be submitted to a public repository like ArrayExpress as part of the peer-review process.
Expression Atlas – when to use it?

- Find out if the expression of a gene (or a group of genes with a common gene attribute, e.g. GO term) change(s) across all the experiments available in the Expression Atlas;

- Discover which genes are differentially expressed in a particular biological condition that you are interested in.
ARRAYEXPRESS AND GENE EXPRESSION ATLAS: MINING FUNCTIONAL GENOMICS DATA

Hands-on-session
ArrayExpress – two databases

Direct submissions
Import from external databases (GEO)

ArrayExpress Archive

Curation
Statistical analysis

Expression Atlas

Links to other EBI databases and analysis software, i.e. Bioconductor
What is the difference between them?

ArrayExpress Archive

- Central object: experiment
- Query to retrieve experimental information and associated data

Expression Atlas

- Central object: gene/condition
- Query for gene expression changes across experiments and across platforms
The ArrayExpress Archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. Gene Expression Atlas contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.
## Browsing the ArrayExpress Archive

The list of experiments retrieved can be printed, saved as Tab-delimited format or exported to Excel or as RSS feed. The total number of experiments and assay retrieved. The direct link to raw and processed data. An icon indicates that this type of data is available.

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Assays</th>
<th>Species</th>
<th>Date</th>
<th>Processed</th>
<th>Raw</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-GEO-24233</td>
<td>Deep transcriptional sequencing analysis of human prostate adenocarcinoma and reference samples</td>
<td>8</td>
<td>Homo sapiens</td>
<td>2011-01-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-19323</td>
<td>PTB CLIP-seq</td>
<td>4</td>
<td>Homo sapiens</td>
<td>2011-01-13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-11941</td>
<td>Transcription profiling by array of human primary hepatocytes treated with trovafloxacin</td>
<td>8</td>
<td>Homo sapiens</td>
<td>2010-11-25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-11940</td>
<td>Transcription profiling by array of human hepatocytes treated with doxorubicin.</td>
<td>10</td>
<td>Homo sapiens</td>
<td>2010-11-24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-11190</td>
<td>Transcription profiling by array of human liver biopsies and blood samples collected from CHC patients before a</td>
<td>78</td>
<td>Homo sapiens</td>
<td>2010-11-24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-11135</td>
<td>Transcription profiling by array of human cancer cell lines</td>
<td>204</td>
<td>Homo sapiens</td>
<td>2010-11-24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-20333</td>
<td>Transcription profiling by array of substantia nigra from individuals with Parkinsons disease</td>
<td>12</td>
<td>Homo sapiens</td>
<td>2010-11-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-ARMX-7</td>
<td>Transcription profiling of peritoneal fat of SHR/Ola (spontaneously hypertensive rat), BN-Lx (Brown Norway cor</td>
<td>258</td>
<td>Rattus norvegicus</td>
<td>2010-11-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-25025</td>
<td>Expression profile of third instar larval fat body and midgut tissues</td>
<td>2</td>
<td>Drosophila melanogaster</td>
<td>2010-11-02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-TABM-811</td>
<td>Transcription profiling of human malignant pheochromocytoma vs normal tissue from the adrenal medulla</td>
<td>2</td>
<td>Homo sapiens</td>
<td>2010-09-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MTAB-372</td>
<td>Transcription profiling by array of human CD138 purified bone marrow plasma cells of previously untreated myeloma patients</td>
<td>280</td>
<td>Homo sapiens</td>
<td>2010-09-12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-2488</td>
<td>Transcription profiling of human nucleus pulposus vs articular cartilage cell types</td>
<td>6</td>
<td>Homo sapiens</td>
<td>2010-09-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-TABM-997</td>
<td>Transcription profiling by array of mouse mammary epithelial stem cell tumours.</td>
<td>9</td>
<td>Mus musculus</td>
<td>2010-09-09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-TABM-654</td>
<td>Transcription profiling of mouse basal-like mammary tumors of luminal origin</td>
<td>12</td>
<td>Mus musculus</td>
<td>2010-09-09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-2746</td>
<td>Transcription profiling by array of Drosophila stage 14 oocytes</td>
<td>4</td>
<td>Drosophila melanogaster</td>
<td>2010-09-07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-2580</td>
<td>Transcription profiling of Drosophila early embryos and unfertilized eggs to analyze mRNA decay patterns in early embryos</td>
<td>24</td>
<td>Drosophila melanogaster</td>
<td>2010-09-07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-2750</td>
<td>Transcription profiling of Arabidopsis thaliana APT1::APRT transgenic plants vs wild type plants to identify the gene</td>
<td>6</td>
<td>Arabidopsis thaliana</td>
<td>2010-09-07</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Browsing the ArrayExpress Archive
Searching AE Archive
Simple query

➢ Search across all fields:
  • AE accession number e.g. E-MEXP-568
  • Secondary accession numbers e.g. GEO series accession GSE5389
  • Experiment name
  • Submitter's experiment description
  • Sample attributes, experimental factor and values, including species (e.g. GeneticModification, Mus musculus, DREB2C over-expression)
  • Publication title, authors and journal name, PubMed ID

➢ Synonyms for terms are always included in searches e.g. 'human' and 'Homo sapiens'
## Searching ArrayExpress Archive

### Simple query - lactococcus

<table>
<thead>
<tr>
<th>Accession</th>
<th>Title</th>
<th>Assays</th>
<th>Species</th>
<th>Date</th>
<th>Processed</th>
<th>Raw</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-GEO-34505</td>
<td>Comparative transcriptional analysis of bacillus subtilis cells overproducing...</td>
<td>29</td>
<td>Bacillus subtilis</td>
<td>2012-05-24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-3162</td>
<td>Transcription profile of dendritic cell line D1 treated with Lactococcus Lac...</td>
<td>8</td>
<td>Mus musculus</td>
<td>2012-03-28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-31032</td>
<td>Discovering novel bile protection systems in Bifidobacterium breve UCC200...</td>
<td>4</td>
<td>Bifidobacterium breve U...</td>
<td>2011-12-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-31943</td>
<td>The impact of a consortium of fermented milk strains on the gut microbiota...</td>
<td>140</td>
<td>Bacteria</td>
<td>2011-11-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-12962</td>
<td>Lactococcus lactis response to isoleucine starvation</td>
<td>12</td>
<td>Lactococcus lactis</td>
<td>2011-09-21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-30625</td>
<td>pBL1 impairs cellulose metabolism in Lactococcus lactis</td>
<td>0</td>
<td>Lactococcus lactis</td>
<td>2011-09-14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-30776</td>
<td>L. lactis NZ9000 (pNZbcApH6) Vs. L. lactis NZ9000 (pNZbP048)</td>
<td>2</td>
<td>Lactococcus lactis sub...</td>
<td>2011-07-19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-MEXP-2842</td>
<td>Transcription profiling by array of Staphylococcus aureus grown in cheese...</td>
<td>15</td>
<td>Staphylococcus aureus</td>
<td>2011-07-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-20379</td>
<td>S. oneidensis MR-1 with electrode, Fe(III)citrate and oxygen as electron...</td>
<td>9</td>
<td>Shewanella oneidensis M...</td>
<td>2011-02-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-26536</td>
<td>Multi-omics approach to study the growth efficiency and amino acid meta...</td>
<td>7</td>
<td>Lactococcus lactis sub...</td>
<td>2011-01-11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-26042</td>
<td>Host intelligence in response to bacteriophage c2 infection in Lactococcus...</td>
<td>2</td>
<td>Lactococcus lactis, Lact...</td>
<td>2010-12-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-20343</td>
<td>SES performance of S. oneidensis pure- and co-culture</td>
<td>4</td>
<td>Shewanella oneidensis M...</td>
<td>2010-12-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-GEO-21759</td>
<td>Genome Sequences of Lactococcus lactis MG1363 (Revised) and NZ9000 ...</td>
<td>2</td>
<td>Lactococcus lactis sub...</td>
<td>2010-11-11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

33 experiments, 579 assays.
## ArrayExpress Archive – experiment view

**Experiment E-MEXP-2842**  
Transcription profiling by array of *Staphylococcus aureus* grown in cheese matrix in mixed culture with *Lactococcus lactis* (15 assays)

<table>
<thead>
<tr>
<th><strong>Species</strong></th>
<th><em>Staphylococcus aureus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status</strong></td>
<td>Submitted on 19 August 2010, released on 1 July 2011, last updated on 27 July 2011</td>
</tr>
<tr>
<td><strong>Samples (16)</strong></td>
<td>Click for detailed sample information and links to data</td>
</tr>
<tr>
<td><strong>Platform (1)</strong></td>
<td>A-MEXP-1792 - STAPH-V2</td>
</tr>
<tr>
<td><strong>Protocols (9)</strong></td>
<td>Click for all experimental protocols</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Effect of the presence of <em>Lactococcus lactis</em> on <em>Staphylococcus aureus</em> transcriptome in cheese matrix. <em>S. aureus</em> was co-cultured with <em>L. lactis</em> LD61 in cheese matrix during 7 days. RNA samples were extracted at different time points (6 h, 8 h, 10 h, 24 h and 7 days) in order to monitor the dynamic response of <em>S. aureus</em> MW2 in cheese matrix in presence of <em>L. lactis</em></td>
</tr>
<tr>
<td><strong>Experiment types</strong></td>
<td>transcription profiling by array, co-expression, growth condition, in vitro, time series</td>
</tr>
<tr>
<td><strong>Contact</strong></td>
<td>Sergine Even <a href="mailto:sergine.even@rennes.inra.fr">sergine.even@rennes.inra.fr</a></td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td><em>Staphylococcus aureus</em> virulence and metabolism are dramatically affected by <em>Lactococcus lactis</em> in cheese matrix. Marina Cretenet, Sébastien Nouaille, Jennifer Thouin, Lucie Rault, Ludwig Stenz, Patrice François, Jacques-Antoine Hennekinne, Michel Piot, Marie Bernadette Maillard, Jacques Fauquant, Pascal Loubière, Yves Le Loir and Sergine Even. <em>Environmental Microbiology Reports</em></td>
</tr>
</tbody>
</table>

**MIAME**

- Platforms: 🟢 ✓ ✓ ✓ ✓ ✓
- Protocols: ✓ ✓ ✓ ✓ ✓
- Factors: ✓ ✓ ✓ ✓ ✓
- Processed: ✓ ✓ ✓ ✓ ✓
- Raw: ✓ ✓ ✓ ✓ ✓

**Files**

| **Investigation Description** | E-MEXP-2842.idf.txt |
| **Sample and Data Relationship** | E-MEXP-2842.sdrf.txt |

Click to browse all available files.

MIAME or MINSEQE scores show how much the experiment is standard compliant.

Sample information, experimental factor(s) and its values.

Link to files available.

This varies between sequencing and microarray data. For microarray experiments you also have array design file.
ArrayExpress Archive – sample information

Experiment E-MEXP-2842
Transcription profiling by array of Staphylococcus aureus grown in cheese matrix in mixed culture

<table>
<thead>
<tr>
<th>Species</th>
<th>Status</th>
<th>Samples (16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staphylococcus aureus</td>
<td>Submitted on 19 August 2010, released on 1 July 2011, last updated on 2</td>
<td>Click for detailed sample information and links to data</td>
</tr>
</tbody>
</table>

Experiment E-MEXP-2842
Transcription profiling by array of Staphylococcus aureus grown in cheese matrix in mixed culture with Lactococcus lactis (16 samples)

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Organism</th>
<th>Sample Characteristics</th>
<th>StrainOrLine</th>
<th>TIME</th>
<th>Factor Values</th>
<th>Raw</th>
<th>Processed</th>
<th>Processed Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1L12 - MX6h_3</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>6</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L12 - MX6h_2</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>6</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2L3 - MX6h_1</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>6</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L10 - MX7Days_14</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>7</td>
<td>days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L20 - MX7Days_15</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>7</td>
<td>days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2L11 - MX7Days_13</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>7</td>
<td>days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L14 - MX8h_6</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>8</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L4 - MX8h_5</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>8</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2L5 - MX8h_4</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>8</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L16 - MX10h_9</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>10</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1L6 - MX10h_8</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>10</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2L7 - MX10h_7</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>10</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW2 gDNA_16</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>16</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW2 gDNA_16</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>16</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW2 gDNA_16</td>
<td>Staphylococcus aureus</td>
<td>MW2</td>
<td>16</td>
<td>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EBI > ArrayExpress > Experiments > E-MEXP-2842 > Samples and Data
# ArrayExpress Archive – ADF file

## Experiment E-MEXP-2842

Transcription profiling by array of *Staphylococcus aureus* grown in cheese matrix in mixed culture with *Lactococcus lactis* (15 assays)

<table>
<thead>
<tr>
<th>Species</th>
<th><em>Staphylococcus aureus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Submitted on 19 August 2010, released on 1 July 2011, last updated on 27 July 2011</td>
</tr>
<tr>
<td>Samples (16)</td>
<td>Click for detailed sample information and links to data</td>
</tr>
<tr>
<td><strong>Platform (1)</strong></td>
<td>A-MEXP-1792 - STAPH-V2</td>
</tr>
<tr>
<td>Protocols (9)</td>
<td>Click for all experimental protocols</td>
</tr>
<tr>
<td>Description</td>
<td>Effect of the presence of <em>Lactococcus lactis</em> on <em>Staphylococcus aureus</em> transcriptome in cheese matrix. <em>S. aureus</em> was co-cultured with <em>L. lactis</em> LD61 in cheese matrix during 7 days. RNA samples were extracted at different time points (6 h, 8 h, 10 h, 24 h and 7 days) in order to monitor the dynamic response of <em>S. aureus</em> MW2 in cheese matrix in presence of <em>L. lactis</em></td>
</tr>
<tr>
<td>Experiment types</td>
<td>transcription profiling by array, co-expression, growth condition, in vitro, time series</td>
</tr>
<tr>
<td>Contact</td>
<td>Sergine Even <a href="mailto:sergine.even@rennes.inra.fr">sergine.even@rennes.inra.fr</a></td>
</tr>
<tr>
<td>Citation</td>
<td><em>Staphylococcus aureus</em> virulence and metabolism are dramatically affected by <em>Lactococcus lactis</em> in cheese matrix. Marina Cretenet, Sébastien Nouaille, Jennifer Thounin, Lucie Rault, Ludwig Stenz, Patrice François, Jacques-Antoine Hennekinne, Michel Piot, Marie Bernadette Maillard, Jacques Fauquant, Pascal Loubière, Yves Le Loir and Sergine Even. Environmental Microbiology Reports</td>
</tr>
</tbody>
</table>

### MIAME

<table>
<thead>
<tr>
<th>Platforms</th>
<th>Protocols</th>
<th>Factors</th>
<th>Processed</th>
<th>Raw</th>
</tr>
</thead>
</table>

### Files

- **Data Archives**
  - E-MEXP-2842.processed.1.zip, E-MEXP-2842.processed.2.zip, E-MEXP-2842.raw.1.zip
  - E-MEXP-2842.idf.txt
  - E-MEXP-2842.srf.txt

- **Browse all available files**
# ArrayExpress Archive – all files

## Experiment E-MEXP-2842

Transcription profiling by array of *Staphylococcus aureus* grown in cheese matrix in mixed culture with *Lactococcus lactis* (15 assays)

<table>
<thead>
<tr>
<th>Species</th>
<th><em>Staphylococcus aureus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Submitted on 19 August 2010, released on 1 July 2011, last updated on 27 July 2011</td>
</tr>
<tr>
<td>Samples (16)</td>
<td>Click for detailed sample information and links to data</td>
</tr>
<tr>
<td>Platform (1)</td>
<td>A-MEXP-1792 - STAPH-V2</td>
</tr>
<tr>
<td>Protocols (9)</td>
<td>Click for all experimental protocols</td>
</tr>
<tr>
<td>Description</td>
<td>Effect of the presence of <em>Lactococcus lactis</em> on <em>Staphylococcus aureus</em> transcriptome in cheese matrix. <em>S. aureus</em> was co-cultured with <em>L. lactis</em> LD61 in cheese matrix during 7 days. RNA samples were extracted at different time points (6 h, 8 h, 10 h, 24 h and 7 days) in order to monitor the dynamic response of <em>S. aureus</em> MW2 in cheese matrix in presence of <em>L. lactis</em></td>
</tr>
<tr>
<td>Experiment types</td>
<td>transcription profiling by array, co-expression, growth condition, in vitro, time series</td>
</tr>
<tr>
<td>Contact</td>
<td>Sergine Even <a href="mailto:sergine.even@rennes.inra.fr">sergine.even@rennes.inra.fr</a></td>
</tr>
<tr>
<td>Citation</td>
<td><em>Staphylococcus aureus</em> virulence and metabolism are dramatically affected by <em>Lactococcus lactis</em> in cheese matrix. Marina Cretenet, Sébastien Nouaille, Jennifer Thounin, Lucie Rault, Ludwig Stenz, Patrice François, Jacques-Antoine Hennekinne, Michel Piot, Marie Bernadette Maillard, Jacques Faquvant, Pascal Loubière, Yves Le Loir and Sergine Even. <em>Environmental Microbiology Reports</em></td>
</tr>
</tbody>
</table>

### MIAME
- **Platforms**: ✔️
- **Protocols**: ✔️
- **Factors**: ✔️
- **Processed**: ✔️
- **Raw**: ✔️

### Files
- **Data Archives**: E-MEXP-2842.processed.1.zip, E-MEXP-2842.processed.2.zip, E-MEXP-2842.raw.1.zip
- **Investigation Description**: E-MEXP-2842.idf.txt
- **Sample and Data Relationship**: E-MEXP-2842.sdrf.txt

[Browse all available files](#)
## ArrayExpress Archive – all files

### Experiment E-MEXP-2842

Transcription profiling by array of Staphylococcus aureus grown in cheese matrix in mixed culture with Lactococcus lactis

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-MEXP-2842.README.txt</td>
<td>4 KB</td>
<td>28 December 2011, 12:24</td>
</tr>
<tr>
<td>E-MEXP-2842.processed.1.zip</td>
<td>257 KB</td>
<td>28 December 2011, 12:24</td>
</tr>
<tr>
<td>E-MEXP-2842.processed.2.zip</td>
<td>351 KB</td>
<td>28 December 2011, 12:24</td>
</tr>
<tr>
<td>E-MEXP-2842.raw.1.zip</td>
<td>9.4 MB</td>
<td>28 December 2011, 12:24</td>
</tr>
<tr>
<td>E-MEXP-2842.idf.txt</td>
<td>10 KB</td>
<td>28 December 2011, 12:24</td>
</tr>
<tr>
<td>E-MEXP-2842.sdrof.txt</td>
<td>14 KB</td>
<td>28 December 2011, 12:24</td>
</tr>
</tbody>
</table>

### Array A-MEXP-1792

STAPH-V2

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-MEXP-1792.README.txt</td>
<td>1 KB</td>
<td>23 February 2010, 23:45</td>
</tr>
<tr>
<td>A-MEXP-1792.adf.txt</td>
<td>883 KB</td>
<td>10 January 2011, 21:16</td>
</tr>
<tr>
<td>A-MEXP-1792.adf.xls</td>
<td>1.2 MB</td>
<td>23 February 2010, 23:45</td>
</tr>
<tr>
<td>A-MEXP-1792.features.txt</td>
<td>381 KB</td>
<td>23 February 2010, 23:45</td>
</tr>
<tr>
<td>A-MEXP-1792.mageml.tar.gz</td>
<td>113 KB</td>
<td>23 February 2010, 23:48</td>
</tr>
<tr>
<td>A-MEXP-1792.reporters.txt</td>
<td>62 KB</td>
<td>23 February 2010, 23:45</td>
</tr>
</tbody>
</table>
Search AE Archive

Advanced query

➢ Combine search terms

  • Enter two or more keywords in the search box with the operators **AND**, **OR** or **NOT**. **AND** is the default search term; a search for kidney cancer will return hits with a match to ‘kidney’ **AND** ‘cancer’

  • Search terms of more than one word must be entered inside quotes otherwise only the first word will be searched for. E.g. “kidney cancer”

➢ Specify fields for searches

  • Particular fields for searching can also be specified in the format of `fieldname:value`
## Searching AE Archive
### Advanced query - fieldnames

<table>
<thead>
<tr>
<th>Field name</th>
<th>Searches</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>accession</strong></td>
<td>Experiment primary or secondary accession</td>
<td>accession:E-MEXP-568</td>
</tr>
<tr>
<td><strong>array</strong></td>
<td>Array design accession or name</td>
<td>array:AFFY-2 OR array:Agilent*</td>
</tr>
<tr>
<td><strong>ef</strong></td>
<td>Experimental factor, the name of the main variables in an experiment.</td>
<td>ef:celltype OR ef:compound</td>
</tr>
<tr>
<td><strong>efv</strong></td>
<td>Experimental factor value. Has EFO expansion.</td>
<td>efv:fibroblast</td>
</tr>
<tr>
<td><strong>expdesign</strong></td>
<td>Experiment design type</td>
<td>expdesign:&quot;dose response&quot;</td>
</tr>
<tr>
<td><strong>exptype</strong></td>
<td>Experiment type. Has EFO expansion.</td>
<td>exptype:RNA-seq</td>
</tr>
<tr>
<td><strong>gxa</strong></td>
<td>Presence in the Gene Expression Atlas. Only value is gxa:true.</td>
<td>ef:compound AND gxa:true</td>
</tr>
<tr>
<td><strong>pmid</strong></td>
<td>PubMed identifier</td>
<td>pmid:16553887</td>
</tr>
<tr>
<td><strong>sa</strong></td>
<td>Sample attribute values. Has EFO expansion.</td>
<td>sa:wild_type</td>
</tr>
<tr>
<td><strong>species</strong></td>
<td>Species of the samples. Has EFO expansion.</td>
<td>species:&quot;homo sapiens&quot; AND ef:cellline</td>
</tr>
</tbody>
</table>
Searching AE Archive
Advanced query – Examples

Experiment, citation, sample and factor annotations [clear]

species: "homo sapiens" AND ef:disease_state

Experiment, citation, sample and factor annotations [clear]

array:A-AFFY-* OR array:agilent*
ArrayExpress – two databases
What is the Gene Expression Atlas?

The Gene Expression Atlas (GXA) is a semantically enriched database of publicly available gene expression data. The data is re-analysed in-house to detect genes showing interesting expression patterns under the conditions of the original experiment.

Based on data from a subset of gene expression experiments from the ArrayExpress archive.
Atlas construction

- Curators select eligible experiments and annotate samples to a high standard using the **Experimental Factor Ontology** (EFO).

- Raw expression data from each **Affymetrix** microarray experiment is pre-processed in **R** using RMA (Robust Multichip Averaging), via the **oligo** package from **Bioconductor**.

- The **limma** Bioconductor package is used to discover differentially expressed genes. For each experiment:
  
  (i) **identify differentially expressed genes for each experimental factor**
  (ii) **for each gene found, identify experimental factor values where the gene’s mean expression level is significantly different from its overall mean across all factor values**

For example, in an experiment with samples from tumors and healthy tissue, we ask whether the expression level in the tumor samples is significantly up or down compared with the mean expression level for all samples.
Gene Expression Atlas

The ArrayExpress Archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. Gene Expression Atlas contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.

Experiments Archive
15782 experiments, 442575 assays

Gene Expression Atlas
5670 experiments, 138915 assays, 18346 conditions

Genes Conditions
up/down in

Any species

Gene Expression Atlas Home

News
• 20 Oct 2010 - Internship for a student project in human gene expression - Filled now
  This student project is now taken.
• 17 Sep 2010 - New Atlas Data Release 10.8
  A new release of Gene Expression Atlas has been made with 93 new experiments...more

Links
• ArrayExpress User Survey
• Old ArrayExpress Interface
• Help | Training | FAQ | Citing
• Submit Data (array based and re-sequencing)
• Programmatic Access | FTP Access
• Software Downloads and Statistics
• EFO | Bioconductor Package | Quality Metrics
• ArrayExpress Scientific Advisory Board
• Functional Genomics Group
**Atlas home page**

http://www.ebi.ac.uk/gxa/

<table>
<thead>
<tr>
<th>Genes</th>
<th>Organism</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(all genes)</td>
<td>up/down in (any)</td>
<td>(all conditions)</td>
</tr>
</tbody>
</table>

e.g. ASPM, "p53 binding"

e.g. liver, cancer, diabetes

**Query for genes**

**Query for conditions**

Restrict query by direction of differential expression

The ‘advanced query’ option allows building more complex queries.
The gene autocomplete lists hits matching all gene annotations such as gene names, uniprot ids, GO terms...etc
Saa4
Mus musculus

Saa4 is differentially expressed in 176 experiments: liver, kidney; ..., (64 more); 13 disease states: normal, LPS-induced acute lung injury, ..., (11 more); 22 cell types; 6 cell lines; 29 compound treatments and 17 other conditions.

Synonyms
Saa2-5, Saa2-4

Orthologs
SAA4 (Homo sapiens)  SAA4_BOVIN (Bos taurus)  (Compare orthologs)

Gene Ontology Term
acute-phase response, extracellular region, high-density lipoprotein particle

UniProt Accession
P31532

InterPro Term
Serum amyloid A protein

Search EBI eye
ENSMUSG00000040817

Show more properties

Experimental Factors
organism part
studied in E-TABM-1039, E-MEXP-114, E-GEOD-91, E-GEOD-33381, E-TABM-865, ..., (47 experiments)

Expression Profiles
2 experiments showing differential expression in "EFO_0000827: eye structure"

E-GEOD-6238: Transcription profiling of hippocampus and retina from two rare inbred mouse models of accelerated neurological senescence (SAMP8 and SAMP14) and a related wild type strain SAMR1 to study molecular senescence of the retina and hippocampus. PubMed 15960800

Experimental Factors
age | disease state | organ part | strain or line

show expression profile | experiment details

Emma Hastings
Box plot showing differential expression of saa4 across the organism parts studied.

The X-axis represents all samples in this study, grouped by experimental factor value, and the Y-axis represents the expression levels for Saa4 in each sample.
The condition box uses a controlled vocabulary called EFO to expand queries, query synonyms and to suggest query terms. EFO conditions can be browsed by clicking the "plus" sign and navigate down the tree of conditions.
## Atlas heatmap view

### Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Organism</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>LARP1B</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>EFCAB2</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>ELP2</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>TUBGCP3</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>TNPO1</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>NPHP1</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>LPP</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>ETV4</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
<tr>
<td>CACNA2D3</td>
<td>Homo sapiens</td>
<td>HeLa</td>
</tr>
</tbody>
</table>

**Legend:**
- 

- number of studies the gene is **over/under** expressed in (~ in experiment pop-ups indicates non-differential expression)

**E-GEO-20081:** Expression Profiling Reveals Unexpected Targets and Functions of the Human Steroid Receptor RNA Activator (SRA) Gene
Data submission to AE

ARRAYEXPRESS
NEEDS YOU
I have a question about how to submit data?
miamexpress@ebi.ac.uk

I have a question about ArrayExpress Archive or the Gene Expression Atlas?
arrayexpress@ebi.ac.uk
**Summary: ArrayExpress**

- **Search by keyword**
- **Search by experiment**
- **Search by gene name, species and experimental condition**
- **Browse results summary**
- **Link to sample properties and experiment design**
- **View expression under different conditions and profiles**
- **View experiment**
Find out more

• Visit our eLearning portal, **Train online**, at [http://www.ebi.ac.uk/training/online/](http://www.ebi.ac.uk/training/online/) for courses on ArrayExpress and Atlas

• Email us at: [miamexpress@ebi.ac.uk](mailto:miamexpress@ebi.ac.uk)

• Atlas mailing list: [arrayexpress-atlas@ebi.ac.uk](mailto:arrayexpress-atlas@ebi.ac.uk)