ExpressionAtlas package vignette

Maria Keays

2016-01-20

Expression Atlas

The [EMBL-EBI Expression Atlas](#) consists of hand-picked high quality datasets from [ArrayExpress](#) that have been manually curated and re-analyzed via the Expression Atlas analysis pipeline. The Expression Atlas website allows users to search these datasets for genes and/or experimental conditions, to discover which genes are expressed in which tissues, cell types, developmental stages, and hundreds of other experimental conditions.

The *ExpressionAtlas* R package enables download of pre-packaged data from Expression Atlas directly into an R session. Raw counts are provided for RNA-seq datasets, while normalized intensities are available for microarray experiments. Protocols describing how the data was generated are contained within the downloaded R objects, with more detailed information available on the [Expression Atlas website](#). Sample annotations are also included in the R object.

Searching and downloading Expression Atlas data

Searching

You can search for experiments in Atlas using the `searchAtlasExperiments()` function. This function returns a *DataFrame* (see [S4Vectors](#)) containing the results of your search. The first argument to `searchAtlasExperiments()` should be a character vector of sample properties, e.g. biological sample attributes and/or experimental treatments. You may also optionally provide a species to limit your search to, as a second argument.

```
suppressMessages( library( ExpressionAtlas ) )
atlasRes <- searchAtlasExperiments( "salt", "Oryza sativa" )
```

```
## Searching for Expression Atlas experiments matching your query
## http://www.ebi.ac.uk/arrayexpress/xml/v2/experiments?keywords=salt&gxa=TRUE&species=Oryza%20sativa .
## Query successful.
## Found 3 experiments matching your query.
```

```
atlasRes
```

```
## DataFrame with 3 rows and 4 columns
##       Accession                      Species
##     <character>                  <character>
## 1 E-GEOD-11175 Oryza sativa Japonica group
## 2  E-MTAB-1625 Oryza sativa Japonica Group
## 3  E-MTAB-1624 Oryza sativa Japonica Group
##                                    Type
##                             <character>
## 1 transcription profiling by array
## 2             RNA-seq of coding RNA
```

```
## 3 transcription profiling by array
##
##
## 1                                 Transcription profiling by array of wild type and drought and salt tolerance
## 2              RNA-seq of coding RNA in rice (Oryza sativa L. Japonica nipponbare) under salt stress o
## 3 Transcription profiling by array of rice (Oryza sativa L. Japonica Nipponbare) under salt stress o
```

The *Accession* column contains the ArrayExpress accession of each dataset – the unique identifier assigned to
it. The species, experiment type (e.g. microarray or RNA-seq), and title of each dataset are also listed.

Downloading the data

To download the data for any/all of the experiments in your results, you can use the function `getAtlasData()`.
This function accepts a vector of ArrayExpress accessions. The data is downloaded into a *SimpleList* object
(see package S4Vectors), with one entry per experiment, listed by accession.

For example, to download all the datasets in your results:

```r
allExps <- getAtlasData( atlasRes$Accession )
```

```
## Downloading Expression Atlas experiment summary from:
##  http://wwwdev.ebi.ac.uk/gxa/experiments/E-GEOD-11175/E-GEOD-11175-atlasExperimentSummary.Rdata ...
## Successfully downloaded experiment summary object for E-GEOD-11175
## Downloading Expression Atlas experiment summary from:
##  http://wwwdev.ebi.ac.uk/gxa/experiments/E-MTAB-1625/E-MTAB-1625-atlasExperimentSummary.Rdata ...
## Successfully downloaded experiment summary object for E-MTAB-1625
## Downloading Expression Atlas experiment summary from:
##  http://wwwdev.ebi.ac.uk/gxa/experiments/E-MTAB-1624/E-MTAB-1624-atlasExperimentSummary.Rdata ...
## Successfully downloaded experiment summary object for E-MTAB-1624
```

```r
allExps
```

```
## List of length 3
## names(3): E-GEOD-11175 E-MTAB-1625 E-MTAB-1624
```

To only download the RNA-seq experiment(s):

```r
rnaseqExps <- getAtlasData(
    atlasRes$Accession[
        grep(
            "rna-seq",
            atlasRes$Type,
            ignore.case = TRUE
        )
    ]
)
```

```
## Downloading Expression Atlas experiment summary from:
##  http://wwwdev.ebi.ac.uk/gxa/experiments/E-MTAB-1625/E-MTAB-1625-atlasExperimentSummary.Rdata ...
## Successfully downloaded experiment summary object for E-MTAB-1625
```

```
rnaseqExps
```

```
## List of length 1
## names(1): E-MTAB-1625
```

To access an experiment summary, use the accession:

```
mtab1624 <- allExps[[ "E-MTAB-1624" ]]
mtab1625 <- allExps[[ "E-MTAB-1625" ]]
```

Each dataset is also represented by a *SimpleList*, with one entry per platform used in the experiment. For RNA-seq data there will only ever be one entry, named **rnaseq**. For microarray data, there is one entry per array design used, listed by ArrayExpress array design accession (see below).

RNA-seq experiment summaries

Following on from above, **mtab1625** now contains a *SimpleList* object with a single entry named **rnaseq**. For RNA-seq experiments, this entry is a *RangedSummarizedExperiment* object (see package SummarizedExperiment).

```
sumexp <- mtab1625$rnaseq
sumexp
```

```
## class: RangedSummarizedExperiment
## dim: 91080 18
## metadata(4): pipeline filtering mapping quantification
## assays(1): counts
## rownames(91080): EPlOSAG00000000001 EPlOSAG00000000002 ...
##    OS12G0641500 OS12G0641600
## rowRanges metadata column names(0):
## colnames(18): ERR266221 ERR266222 ... ERR266237 ERR266238
## colData names(9): AtlasAssayGroup organism ... growth_condition
##    sampling_time
```

The matrix of raw counts for this experiment is stored in the *assays* slot:

```
head( assays( sumexp )$counts )
```

```
##                    ERR266221 ERR266222 ERR266223 ERR266224 ERR266225
## EPlOSAG00000000001         0         0         0         0         1
## EPlOSAG00000000002         0         0         0         0         0
## EPlOSAG00000000003         0         0         0         0         0
## EPlOSAG00000000004         0         0         0         0         0
## EPlOSAG00000000005         0         0         0         0         0
## EPlOSAG00000000006         0         0         0         0         0
##                    ERR266226 ERR266227 ERR266228 ERR266229 ERR266230
## EPlOSAG00000000001         0         0         0         0         1
## EPlOSAG00000000002         0         0         0         0         0
## EPlOSAG00000000003         0         0         0         0         0
## EPlOSAG00000000004         0         0         0         0         0
```

```
## EP1OSAG00000000005           0        0        0        0        0
## EP1OSAG00000000006           0        0        0        0        0
##                       ERR266231 ERR266232 ERR266233 ERR266234 ERR266235
## EP1OSAG00000000001           0        1        0        1        1
## EP1OSAG00000000002           0        0        0        0        0
## EP1OSAG00000000003           0        0        0        0        0
## EP1OSAG00000000004           0        0        0        0        0
## EP1OSAG00000000005           0        0        0        0        0
## EP1OSAG00000000006           0        0        0        0        0
##                       ERR266236 ERR266237 ERR266238
## EP1OSAG00000000001           0        0        0
## EP1OSAG00000000002           0        0        0
## EP1OSAG00000000003           0        0        0
## EP1OSAG00000000004           0        0        0
## EP1OSAG00000000005           0        0        0
## EP1OSAG00000000006           0        0        0
```

The sample annotations can be found in the *colData* slot:

```
colData( sumexp )
```

```
## DataFrame with 18 rows and 9 columns
##           AtlasAssayGroup                organism    cultivar
##               <character>             <character> <character>
## ERR266221              g5 Oryza sativa Japonica Group  Nipponbare
## ERR266222              g2 Oryza sativa Japonica Group  Nipponbare
## ERR266223              g2 Oryza sativa Japonica Group  Nipponbare
## ERR266224              g5 Oryza sativa Japonica Group  Nipponbare
## ERR266225              g3 Oryza sativa Japonica Group  Nipponbare
## ...                    ...                     ...         ...
## ERR266234              g3 Oryza sativa Japonica Group  Nipponbare
## ERR266235              g4 Oryza sativa Japonica Group  Nipponbare
## ERR266236              g4 Oryza sativa Japonica Group  Nipponbare
## ERR266237              g4 Oryza sativa Japonica Group  Nipponbare
## ERR266238              g6 Oryza sativa Japonica Group  Nipponbare
##                                       developmental_stage         age
##                                               <character> <character>
## ERR266221 seedling, two leaves visible, three leaves visible           2
## ERR266222 seedling, two leaves visible, three leaves visible           2
## ERR266223 seedling, two leaves visible, three leaves visible           2
## ERR266224 seedling, two leaves visible, three leaves visible           2
## ERR266225 seedling, two leaves visible, three leaves visible           2
## ...                                                   ...         ...
## ERR266234 seedling, two leaves visible, three leaves visible           2
## ERR266235 seedling, two leaves visible, three leaves visible           2
## ERR266236 seedling, two leaves visible, three leaves visible           2
## ERR266237 seedling, two leaves visible, three leaves visible           2
## ERR266238 seedling, two leaves visible, three leaves visible           2
##             time_unit           organism_part
##           <character>             <character>
## ERR266221        week shoot axis, vascular leaf
## ERR266222        week shoot axis, vascular leaf
## ERR266223        week shoot axis, vascular leaf
```

4

```
## ERR266224          week shoot axis, vascular leaf
## ERR266225          week shoot axis, vascular leaf
## ...                    ...                      ...
## ERR266234          week shoot axis, vascular leaf
## ERR266235          week shoot axis, vascular leaf
## ERR266236          week shoot axis, vascular leaf
## ERR266237          week shoot axis, vascular leaf
## ERR266238          week shoot axis, vascular leaf
##                         growth_condition sampling_time
##                              <character>   <character>
## ERR266221 300 millimolar sodium chloride             5
## ERR266222                 normal watering             5
## ERR266223                 normal watering             5
## ERR266224 300 millimolar sodium chloride             5
## ERR266225                 normal watering            24
## ...                                  ...           ...
## ERR266234                 normal watering            24
## ERR266235 300 millimolar sodium chloride             1
## ERR266236 300 millimolar sodium chloride             1
## ERR266237 300 millimolar sodium chloride             1
## ERR266238 300 millimolar sodium chloride            24
```

Information describing how the raw data files were processed to obtain the raw counts matrix are found in the *metadata* slot:

```
metadata( sumexp )
```

```
## $pipeline
## [1] "iRAP version 0.7.0p1 (http://nunofonseca.github.io/irap/)"
##
## $filtering
## [1] "Discard reads below minimum quality threshold"
## [2] "Check of bacterial contamination; discard offending reads"
## [3] "Discard reads with common uncalled characters (e.g. N)"
## [4] "Remove reads from pair-end libraries that were orphaned by filtering steps 1-3"
##
## $mapping
## [1] "Against genome reference (Ensembl Plants release: 26) tophat2 version: 2.0.12"
##
## $quantification
## [1] "htseq2 version: 0.6.1p1"
```

Single-channel microarray experiments

Data from a single-channel microarray experiment, e.g. E-MTAB-1624, is represented as one or more *ExpressionSet* object(s) in the SimpleList that is downloaded. *ExpressionSet* objects are indexed by the ArrayExpress accession(s) of the microarray design(s) used in the original experiment.

```
names( mtab1624 )
```

```
## [1] "A-AFFY-126"
```

```
affy126data <- mtab1624[[ "A-AFFY-126" ]]
affy126data
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 57381 features, 18 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: nippon_control_1hr_rep1 nippon_control_1hr_rep2 ...
##     nippon_salt_5hr_rep3 (18 total)
##   varLabels: AtlasAssayGroup organism ... sampling_time (9 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: AFFX-BioB-3_at AFFX-BioB-5_at ...
##     RPTR-Os-XXU09476-1_at (57381 total)
##   fvarLabels: probeSets
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

The matrix of normalized intensity values is in the *assayData* slot:

```
head( exprs( affy126data ) )
```

```
##                nippon_control_1hr_rep1 nippon_control_1hr_rep2
## AFFX-BioB-3_at                7.869421                8.365278
## AFFX-BioB-5_at                7.702652                8.020915
## AFFX-BioB-M_at                7.652985                8.156907
## AFFX-BioC-3_at                9.219287                9.556873
## AFFX-BioC-5_at                8.881807                9.186310
## AFFX-BioDn-3_at              11.617725               11.877930
##                nippon_control_1hr_rep3 nippon_control_24hr_rep1
## AFFX-BioB-3_at                8.637034                 8.403105
## AFFX-BioB-5_at                8.407186                 8.150995
## AFFX-BioB-M_at                8.348117                 7.998242
## AFFX-BioC-3_at                9.895652                 9.598754
## AFFX-BioC-5_at                9.553833                 9.354739
## AFFX-BioDn-3_at              12.193324                11.861900
##                nippon_control_24hr_rep2 nippon_control_24hr_rep3
## AFFX-BioB-3_at                 8.678257                 8.456243
## AFFX-BioB-5_at                 8.413489                 8.227663
## AFFX-BioB-M_at                 8.358092                 8.111307
## AFFX-BioC-3_at                 9.872856                 9.699042
## AFFX-BioC-5_at                 9.594959                 9.383014
## AFFX-BioDn-3_at               12.077460                11.959052
##                nippon_control_5hr_rep1 nippon_control_5hr_rep2
## AFFX-BioB-3_at                8.348849                8.643520
## AFFX-BioB-5_at                8.129436                8.374279
## AFFX-BioB-M_at                7.978514                8.285401
## AFFX-BioC-3_at                9.588422                9.828320
## AFFX-BioC-5_at                9.210903                9.512925
## AFFX-BioDn-3_at              11.853478               12.043559
```

```
##                    nippon_control_5hr_rep3 nippon_salt_1hr_rep1
## AFFX-BioB-3_at                    8.401530             8.331911
## AFFX-BioB-5_at                    8.193307             8.015213
## AFFX-BioB-M_at                    8.046037             7.944433
## AFFX-BioC-3_at                    9.685030             9.509499
## AFFX-BioC-5_at                    9.379879             9.194149
## AFFX-BioDn-3_at                  11.952693            11.800154
##                    nippon_salt_1hr_rep2 nippon_salt_1hr_rep3
## AFFX-BioB-3_at                 8.463545             8.901247
## AFFX-BioB-5_at                 8.290420             8.533720
## AFFX-BioB-M_at                 8.139875             8.462567
## AFFX-BioC-3_at                 9.676649             9.950136
## AFFX-BioC-5_at                 9.343052             9.714590
## AFFX-BioDn-3_at               12.043509            12.263983
##                    nippon_salt_24hr_rep1 nippon_salt_24hr_rep2
## AFFX-BioB-3_at                  8.363449              8.185702
## AFFX-BioB-5_at                  8.112024              7.828596
## AFFX-BioB-M_at                  8.115197              7.775437
## AFFX-BioC-3_at                  9.549797              9.300713
## AFFX-BioC-5_at                  9.276916              8.916605
## AFFX-BioDn-3_at                11.866490             11.647577
##                    nippon_salt_24hr_rep3 nippon_salt_5hr_rep1
## AFFX-BioB-3_at                  8.586542             8.562492
## AFFX-BioB-5_at                  8.207703             8.307241
## AFFX-BioB-M_at                  8.231843             8.078984
## AFFX-BioC-3_at                  9.624600             9.728441
## AFFX-BioC-5_at                  9.371714             9.430824
## AFFX-BioDn-3_at                11.908544            12.029436
##                    nippon_salt_5hr_rep2 nippon_salt_5hr_rep3
## AFFX-BioB-3_at                 8.366250             8.499076
## AFFX-BioB-5_at                 8.275820             8.155382
## AFFX-BioB-M_at                 8.114156             8.115064
## AFFX-BioC-3_at                 9.632023             9.531450
## AFFX-BioC-5_at                 9.358933             9.260982
## AFFX-BioDn-3_at               11.832181            11.987857
```

The sample annotations are in the *phenoData* slot:

```
pData( affy126data )
```

```
##                        AtlasAssayGroup                   organism
## nippon_control_1hr_rep1             g1 Oryza sativa Japonica Group
## nippon_control_1hr_rep2             g1 Oryza sativa Japonica Group
## nippon_control_1hr_rep3             g1 Oryza sativa Japonica Group
## nippon_control_24hr_rep1            g3 Oryza sativa Japonica Group
## nippon_control_24hr_rep2            g3 Oryza sativa Japonica Group
## nippon_control_24hr_rep3            g3 Oryza sativa Japonica Group
## nippon_control_5hr_rep1             g2 Oryza sativa Japonica Group
## nippon_control_5hr_rep2             g2 Oryza sativa Japonica Group
## nippon_control_5hr_rep3             g2 Oryza sativa Japonica Group
## nippon_salt_1hr_rep1                g4 Oryza sativa Japonica Group
## nippon_salt_1hr_rep2                g4 Oryza sativa Japonica Group
## nippon_salt_1hr_rep3                g4 Oryza sativa Japonica Group
```

```
## nippon_salt_24hr_rep1                       g6 Oryza sativa Japonica Group
## nippon_salt_24hr_rep2                       g6 Oryza sativa Japonica Group
## nippon_salt_24hr_rep3                       g6 Oryza sativa Japonica Group
## nippon_salt_5hr_rep1                        g5 Oryza sativa Japonica Group
## nippon_salt_5hr_rep2                        g5 Oryza sativa Japonica Group
## nippon_salt_5hr_rep3                        g5 Oryza sativa Japonica Group
##                              cultivar
## nippon_control_1hr_rep1     Nipponbare
## nippon_control_1hr_rep2     Nipponbare
## nippon_control_1hr_rep3     Nipponbare
## nippon_control_24hr_rep1    Nipponbare
## nippon_control_24hr_rep2    Nipponbare
## nippon_control_24hr_rep3    Nipponbare
## nippon_control_5hr_rep1     Nipponbare
## nippon_control_5hr_rep2     Nipponbare
## nippon_control_5hr_rep3     Nipponbare
## nippon_salt_1hr_rep1        Nipponbare
## nippon_salt_1hr_rep2        Nipponbare
## nippon_salt_1hr_rep3        Nipponbare
## nippon_salt_24hr_rep1       Nipponbare
## nippon_salt_24hr_rep2       Nipponbare
## nippon_salt_24hr_rep3       Nipponbare
## nippon_salt_5hr_rep1        Nipponbare
## nippon_salt_5hr_rep2        Nipponbare
## nippon_salt_5hr_rep3        Nipponbare
##                                                   developmental_stage
## nippon_control_1hr_rep1   seedling, two leaves visible, three leaves visible
## nippon_control_1hr_rep2   seedling, two leaves visible, three leaves visible
## nippon_control_1hr_rep3   seedling, two leaves visible, three leaves visible
## nippon_control_24hr_rep1  seedling, two leaves visible, three leaves visible
## nippon_control_24hr_rep2  seedling, two leaves visible, three leaves visible
## nippon_control_24hr_rep3  seedling, two leaves visible, three leaves visible
## nippon_control_5hr_rep1   seedling, two leaves visible, three leaves visible
## nippon_control_5hr_rep2   seedling, two leaves visible, three leaves visible
## nippon_control_5hr_rep3   seedling, two leaves visible, three leaves visible
## nippon_salt_1hr_rep1      seedling, two leaves visible, three leaves visible
## nippon_salt_1hr_rep2      seedling, two leaves visible, three leaves visible
## nippon_salt_1hr_rep3      seedling, two leaves visible, three leaves visible
## nippon_salt_24hr_rep1     seedling, two leaves visible, three leaves visible
## nippon_salt_24hr_rep2     seedling, two leaves visible, three leaves visible
## nippon_salt_24hr_rep3     seedling, two leaves visible, three leaves visible
## nippon_salt_5hr_rep1      seedling, two leaves visible, three leaves visible
## nippon_salt_5hr_rep2      seedling, two leaves visible, three leaves visible
## nippon_salt_5hr_rep3      seedling, two leaves visible, three leaves visible
##                           age time_unit            organism_part
## nippon_control_1hr_rep1    2       week shoot axis, vascular leaf
## nippon_control_1hr_rep2    2       week shoot axis, vascular leaf
## nippon_control_1hr_rep3    2       week shoot axis, vascular leaf
## nippon_control_24hr_rep1   2       week shoot axis, vascular leaf
## nippon_control_24hr_rep2   2       week shoot axis, vascular leaf
## nippon_control_24hr_rep3   2       week shoot axis, vascular leaf
## nippon_control_5hr_rep1    2       week shoot axis, vascular leaf
## nippon_control_5hr_rep2    2       week shoot axis, vascular leaf
## nippon_control_5hr_rep3    2       week shoot axis, vascular leaf
```

```
## nippon_salt_1hr_rep1         2       week shoot axis, vascular leaf
## nippon_salt_1hr_rep2         2       week shoot axis, vascular leaf
## nippon_salt_1hr_rep3         2       week shoot axis, vascular leaf
## nippon_salt_24hr_rep1        2       week shoot axis, vascular leaf
## nippon_salt_24hr_rep2        2       week shoot axis, vascular leaf
## nippon_salt_24hr_rep3        2       week shoot axis, vascular leaf
## nippon_salt_5hr_rep1         2       week shoot axis, vascular leaf
## nippon_salt_5hr_rep2         2       week shoot axis, vascular leaf
## nippon_salt_5hr_rep3         2       week shoot axis, vascular leaf
##                                      growth_condition sampling_time
## nippon_control_1hr_rep1               normal watering             1
## nippon_control_1hr_rep2               normal watering             1
## nippon_control_1hr_rep3               normal watering             1
## nippon_control_24hr_rep1              normal watering            24
## nippon_control_24hr_rep2              normal watering            24
## nippon_control_24hr_rep3              normal watering            24
## nippon_control_5hr_rep1               normal watering             5
## nippon_control_5hr_rep2               normal watering             5
## nippon_control_5hr_rep3               normal watering             5
## nippon_salt_1hr_rep1      300 millimolar sodium chloride          1
## nippon_salt_1hr_rep2      300 millimolar sodium chloride          1
## nippon_salt_1hr_rep3      300 millimolar sodium chloride          1
## nippon_salt_24hr_rep1     300 millimolar sodium chloride         24
## nippon_salt_24hr_rep2     300 millimolar sodium chloride         24
## nippon_salt_24hr_rep3     300 millimolar sodium chloride         24
## nippon_salt_5hr_rep1      300 millimolar sodium chloride          5
## nippon_salt_5hr_rep2      300 millimolar sodium chloride          5
## nippon_salt_5hr_rep3      300 millimolar sodium chloride          5
```

A brief outline of how the raw data was normalized is in the *experimentData* slot:

```
preproc( experimentData( affy126data ) )
```

```
## $normalization
## [1] "RMA using oligo (http://www.bioconductor.org/packages/release/bioc/html/oligo.html) version 1.2
```

Downloading a single Expression Atlas experiment summary

You can also download data for a single Expression Atlas experiment using the `getAtlasExperiment()` function:

```
mtab3007 <- getAtlasExperiment( "E-MTAB-3007" )
```

```
## Downloading Expression Atlas experiment summary from:
##  http://wwwdev.ebi.ac.uk/gxa/experiments/E-MTAB-3007/E-MTAB-3007-atlasExperimentSummary.Rdata ...
## Successfully downloaded experiment summary object for E-MTAB-3007
```