

Supplementary code for article: BgeeDB, an R package for retrieval of curated expression datasets and for gene list enrichment tests

Andrea Komljenović, Julien Roux, Marc Robinson-Rechavi and Frederic Bastian

September 20, 2016

```
source("https://bioconductor.org/biocLite.R")
biocLite("BgeeDB")

## 
## The downloaded binary packages are in
## /var/folders/8b/_0z6s_zn661674ylzs59d1n00000gn/T//Rtmpphy5NIV/downloaded_packages

# additional packages for installation
biocLite(c("edgeR", "Mfuzz", "biomaRt"))

## 
## The downloaded binary packages are in
## /var/folders/8b/_0z6s_zn661674ylzs59d1n00000gn/T//Rtmpphy5NIV/downloaded_packages

library(BgeeDB)

# list the species
listBgeeSpecies()

## 
## Querying Bgee to get release information...
## 
## Building URL to query species in Bgee release 13_2...
## 
## Submitting URL to Bgee webservice... (http://r13_2.bgee.org/?page=species&display_type=tsv)
## 
## Query to Bgee webservice successful!

##      ID          GENUS SPECIES_NAME COMMON_NAME AFFYMETRIX    EST IN_SITU
## 1   6239  Caenorhabditis      elegans     c.elegans      TRUE FALSE    TRUE
## 2   7227 Drosophila melanogaster fruitfly      fruitfly      TRUE TRUE    TRUE
## 3   7955        Danio        rerio  zebrafish      zebrafish      TRUE TRUE    TRUE
## 4   8364       Xenopus    tropicalis    xenopus      xenopus     FALSE TRUE    TRUE
## 5   9031        Gallus       gallus     chicken      chicken     FALSE FALSE   FALSE
## 6   9258 Ornithorhynchus    anatinus    platypus      platypus     FALSE FALSE   FALSE
## 7   9544        Macaca     mulatta    macaque      macaque     FALSE FALSE   FALSE
## 8   9593       Gorilla     gorilla     gorilla      gorilla     FALSE FALSE   FALSE
## 9   9597         Pan    paniscus     bonobo      bonobo     FALSE FALSE   FALSE
## 10  9598         Pan troglodytes chimpanzee chimpanzee     FALSE FALSE   FALSE
## 11  9606        Homo     sapiens     human      human      TRUE TRUE   FALSE
```

```

## 12 9823           Sus      scrofa      pig     FALSE FALSE FALSE
## 13 9913           Bos      taurus      cow     FALSE FALSE FALSE
## 14 10090          Mus      musculus   mouse    TRUE  TRUE  TRUE
## 15 10116          Rattus   norvegicus rat     FALSE FALSE FALSE
## 16 13616          Monodelphis domestica opossum FALSE FALSE FALSE
## 17 28377          Anolis   carolinensis anolis  FALSE FALSE FALSE
##     RNA_SEQ
## 1     TRUE
## 2     FALSE
## 3     FALSE
## 4     TRUE
## 5     TRUE
## 6     TRUE
## 7     TRUE
## 8     TRUE
## 9     TRUE
## 10    TRUE
## 11    TRUE
## 12    TRUE
## 13    TRUE
## 14    TRUE
## 15    TRUE
## 16    TRUE
## 17    TRUE

```

```

# specify species and data type
bgee_affymetrix <- Bgee$new(species="Mus_musculus", dataType="affymetrix", release="13.2")

```

```

##
## Querying Bgee to get release information...
##
## Building URL to query species in Bgee release 13_2...
##
## Submitting URL to Bgee webservice... (http://r13\_2.bgee.org/?page=species&display\_type=tsv)
##
## Query to Bgee webservice successful!
##
## API key built: 696a8ee5f27f212d71349c131d413fb0de7947e688c95fcdd001689789c7333dbdf6fad8d4d48c239a3e0

```

```

# retrieve annotation of all mouse affymetrix datasets in Bgee
annotation_bgee_mouse_affymetrix <- getAnnotation(bgee_affymetrix)

```

```

##
## Downloading annotation files...
##
## Saved annotation files in /Users/jroux/Dropbox/RpackagePaper/Mus_musculus_Bgee_13_2 folder.

```

```

str(annotation_bgee_mouse_affymetrix)

```

```

## List of 2
## $ sample.annotation : 'data.frame': 6077 obs. of 20 variables:
##   ..$ Experiment.ID : chr [1:6077] "GSE3594" "GSE3594" "GSE3594" "GSE3594"

```

```

##   ..$ Chip.ID : chr [1:6077] "GSM82907" "GSM82908" "GSM82959"
##   ..$ Anatomical.entity.ID : chr [1:6077] "UBERON:0000007" "UBERON:0000007"
##   ..$ Anatomical.entity.name : chr [1:6077] "pituitary gland" "pituitary gla
##   ..$ Stage.ID : chr [1:6077] "MmusDv:0000052" "MmusDv:0000052"
##   ..$ Stage.name : chr [1:6077] "8 weeks (mouse)" "8 weeks (mou
##   ..$ IQRray.score : num [1:6077] NA ...
##   ..$ MAS5.percent.present : num [1:6077] 41.1 46.6 44.9 45 51.5 ...
##   ..$ Normalization.type : chr [1:6077] "MAS5" "MAS5" "MAS5" "MAS5" ...
##   ..$ Scan.date : chr [1:6077] NA NA NA NA ...
##   ..$ Chip.type.ID : chr [1:6077] "A-AFFY-6" "A-AFFY-6" "A-AFFY-6"
##   ..$ CDF.name : chr [1:6077] "MG_U74Av2" "MG_U74Av2" "MG_U74
##   ..$ Chip.type.name : chr [1:6077] "Affymetrix GeneChip Murine Gen
##   ..$ IQRray.score.threshold.for.the.chip.type : num [1:6077] 26.7 26.7 26.7 26.7 26.7 ...
##   ..$ MAS5.percent.present.threshold.for.the.chip.type: num [1:6077] 26.7 26.7 26.7 26.7 26.7 ...
##   ..$ Data.source : chr [1:6077] "GEO" "GEO" "GEO" "GEO" ...
##   ..$ Data.source.URL : chr [1:6077] "http://www.ncbi.nlm.nih.gov/geo/q
##   ..$ Bgee.normalized.data.URL : chr [1:6077] "ftp://lausanne.isb-sib.ch/pub/
##   ..$ Bgee.normalized.data.file : chr [1:6077] "Mus_musculus_probesets_GSE3594
##   ..$ Bgee.raw.file.URL : chr [1:6077] "ftp://lausanne.isb-sib.ch/pub/
## $ experiment.annotation:'data.frame': 694 obs. of 11 variables:
##   ..$ Experiment.ID : chr [1:694] "GSE3594" "GSE7759" "GSE10744" "GSE16496" ...
##   ..$ Chip.count : int [1:694] 129 110 108 102 91 87 69 67 65 63 ...
##   ..$ Condition.count : int [1:694] 34 13 6 46 45 7 10 21 6 25 ...
##   ..$ Organ.count : int [1:694] 34 3 6 46 45 7 1 21 1 7 ...
##   ..$ Stage.count : int [1:694] 1 5 1 1 2 1 10 3 6 12 ...
##   ..$ Data.source : chr [1:694] "GEO" "GEO" "GEO" "GEO" ...
##   ..$ Data.source.URL : chr [1:694] "http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35
##   ..$ Bgee.normalized.data.URL: chr [1:694] "ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1
##   ..$ Bgee.raw.files.URL : chr [1:694] "ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_
##   ..$ Experiment.name : chr [1:694] "Transcription profiling of twenty four mouse neural tiss
##   ..$ Experiment.description : chr [1:694] "The current model to explain the organization of the man

# retrieve annotation of samples and experiments
sample_annotation <- annotation_bgee_mouse_affymetrix$sample.annotation
experiment_annotation <- annotation_bgee_mouse_affymetrix$experiment.annotation

# list experiments including a zygote sample
selected_experiments <-
  unique(sample_annotation$Experiment.ID[sample_annotation$Stage.ID == "UBERON:0000106"])
experiment_annotation[experiment_annotation$Experiment.ID %in% selected_experiments,]

##      Experiment.ID Chip.count Condition.count Organ.count Stage.count
## 24        GSE1749         42             5          4          5
## 32        E-MEXP-51         35             9          5          7
## 113       GSE18290        12             4          3          4
##   Data.source
## 24          GEO
## 32  ArrayExpress
## 113         GEO
##                                         Data.source.URL
## 24  http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1749
## 32  http://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-51
## 113 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18290
##
```

```

## 24      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymetrix/
## 32      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymetrix/
## 113     ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymetrix/
##                                         Bgee.raw.files.URL
## 24      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/mas5_files/GSE1749/
## 32      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/
## 113     ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/GSE18290/
## 
## 24                               Transcription profiling of mouse embryo
## 32  Transcription profiling of mouse pre-implantation development over twelve time points from the g
## 113
## 
## 24  Studies using low-resolution methods to assess gene expression during preimplantation mouse deve
## 32
## 113

```

stages sampled in each of these experiments

```
unique(sample_annotation[sample_annotation$Experiment.ID %in% selected_experiments, c(1,6)])
```

##	Experiment.ID	Stage.name
## 1536	GSE1749	life cycle
## 1540	GSE1749	blastula stage
## 1543	GSE1749	zygote stage
## 1546	GSE1749	Theiler stage 02 (mouse)
## 1550	GSE1749	Theiler stage 03 (mouse)
## 1836	E-MEXP-51	zygote stage
## 1840	E-MEXP-51	post-juvenile adult stage
## 1847	E-MEXP-51	Theiler stage 03 (mouse)
## 1849	E-MEXP-51	Theiler stage 02 (mouse)
## 1862	E-MEXP-51	Theiler stage 04 (mouse)
## 1866	E-MEXP-51	Theiler stage 05 (mouse)
## 1868	E-MEXP-51	Theiler stage 06 (mouse)
## 3447	GSE18290	blastula stage
## 3449	GSE18290	zygote stage
## 3452	GSE18290	Theiler stage 02 (mouse)
## 3456	GSE18290	Theiler stage 03 (mouse)

List all samples from E-MEXP-51 in Bgee

```
sample_annotation[sample_annotation$Experiment.ID == "E-MEXP-51",]
```

##	Experiment.ID	Chip.ID	Anatomical.entity.ID
## 1836	E-MEXP-51	Zygote1	CL:0000365
## 1837	E-MEXP-51	Zygote2	CL:0000365
## 1838	E-MEXP-51	Zygote3	CL:0000365
## 1839	E-MEXP-51	Zygote4	CL:0000365
## 1840	E-MEXP-51	GVoocyte1	CL:0000654
## 1841	E-MEXP-51	GVoocyte2	CL:0000654
## 1842	E-MEXP-51	GVoocyte3	CL:0000654
## 1843	E-MEXP-51	GVoocyte4	CL:0000654
## 1844	E-MEXP-51	MIIoocyte1	CL:0000655
## 1845	E-MEXP-51	MIIoocyte2	CL:0000655
## 1846	E-MEXP-51	MIIoocyte3	CL:0000655
## 1847	E-MEXP-51	16cell1	UBERON:0000085

```

## 1848 E-MEXP-51      16cell2      UBERON:0000085
## 1849 E-MEXP-51      Early2-cell1  UBERON:0000922
## 1850 E-MEXP-51      Early2-cell2  UBERON:0000922
## 1851 E-MEXP-51      Late2-cell1  UBERON:0000922
## 1852 E-MEXP-51      Late2-cell2  UBERON:0000922
## 1853 E-MEXP-51      Mid2-cell1  UBERON:0000922
## 1854 E-MEXP-51      Mid2-cell2  UBERON:0000922
## 1855 E-MEXP-51      Mid2-cell3  UBERON:0000922
## 1856 E-MEXP-51      4Cell1      UBERON:0000922
## 1857 E-MEXP-51      4cell2      UBERON:0000922
## 1858 E-MEXP-51      4cell3      UBERON:0000922
## 1859 E-MEXP-51      8cell1      UBERON:0000922
## 1860 E-MEXP-51      8cell2      UBERON:0000922
## 1861 E-MEXP-51      8cell3      UBERON:0000922
## 1862 E-MEXP-51      EarlyBlastocyst1 UBERON:0000922
## 1863 E-MEXP-51      EarlyBlastocyst2 UBERON:0000922
## 1864 E-MEXP-51      EarlyBlastocyst3 UBERON:0000922
## 1865 E-MEXP-51      EarlyBlastocyst4 UBERON:0000922
## 1866 E-MEXP-51      MidBlastocyst1 UBERON:0000922
## 1867 E-MEXP-51      MidBlastocyst2 UBERON:0000922
## 1868 E-MEXP-51      LateBlastocyst1 UBERON:0000922
## 1869 E-MEXP-51      LateBlastocyst2 UBERON:0000922
## 1870 E-MEXP-51      LateBlastocyst3 UBERON:0000922

##   Anatomical.entity.name    Stage.ID        Stage.name
## 1836          zygote UBERON:0000106      zygote stage
## 1837          zygote UBERON:0000106      zygote stage
## 1838          zygote UBERON:0000106      zygote stage
## 1839          zygote UBERON:0000106      zygote stage
## 1840      primary oocyte UBERON:0000113 post-juvenile adult stage
## 1841      primary oocyte UBERON:0000113 post-juvenile adult stage
## 1842      primary oocyte UBERON:0000113 post-juvenile adult stage
## 1843      primary oocyte UBERON:0000113 post-juvenile adult stage
## 1844 secondary oocyte UBERON:0000113 post-juvenile adult stage
## 1845 secondary oocyte UBERON:0000113 post-juvenile adult stage
## 1846 secondary oocyte UBERON:0000113 post-juvenile adult stage
## 1847          morula MmusDv:0000006 Theiler stage 03 (mouse)
## 1848          morula MmusDv:0000006 Theiler stage 03 (mouse)
## 1849          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1850          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1851          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1852          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1853          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1854          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1855          embryo MmusDv:0000005 Theiler stage 02 (mouse)
## 1856          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1857          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1858          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1859          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1860          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1861          embryo MmusDv:0000006 Theiler stage 03 (mouse)
## 1862          embryo MmusDv:0000008 Theiler stage 04 (mouse)
## 1863          embryo MmusDv:0000008 Theiler stage 04 (mouse)
## 1864          embryo MmusDv:0000008 Theiler stage 04 (mouse)
## 1865          embryo MmusDv:0000008 Theiler stage 04 (mouse)

```

```

## 1866      embryo MmusDv:0000009 Theiler stage 05 (mouse)
## 1867      embryo MmusDv:0000009 Theiler stage 05 (mouse)
## 1868      embryo MmusDv:0000010 Theiler stage 06 (mouse)
## 1869      embryo MmusDv:0000010 Theiler stage 06 (mouse)
## 1870      embryo MmusDv:0000010 Theiler stage 06 (mouse)
##      IQRray.score MAS5.percent.present Normalization.type      Scan.date
## 1836      47883.54          32.01           gcRMA 02/27/ 16:31:23
## 1837      48592.69          33.04           gcRMA 02/27/ 16:19:35
## 1838      46959.46          29.06           gcRMA 02/20/ 17:56:57
## 1839      48053.21          30.73           gcRMA 02/20/ 18:08:12
## 1840      50769.96          33.42           gcRMA 12/21/ 14:48:27
## 1841      51133.57          34.18           gcRMA 12/21/ 15:00:32
## 1842      53766.59          38.92           gcRMA 12/23/ 15:51:49
## 1843      51714.74          38.11           gcRMA 12/30/ 13:57:52
## 1844      47625.00          30.18           gcRMA 02/20/ 19:22:46
## 1845      46662.15          29.14           gcRMA 02/20/ 19:45:42
## 1846      48439.18          31.95           gcRMA 03/06/ 14:36:39
## 1847      54791.02          41.80           gcRMA 04/26/ 20:26:07
## 1848      55477.89          42.30           gcRMA 04/26/ 19:57:23
## 1849      47968.59          34.32           gcRMA 02/27/ 14:52:08
## 1850      47188.40          33.64           gcRMA 03/06/ 16:20:26
## 1851      52613.70          41.14           gcRMA 12/30/ 14:13:53
## 1852      51534.73          40.46           gcRMA 12/30/ 14:26:59
## 1853      50929.56          39.04           gcRMA 03/14/ 16:51:24
## 1854      48914.31          36.87           gcRMA 03/14/ 16:11:45
## 1855      50681.62          39.25           gcRMA 04/27/ 17:11:34
## 1856      54046.04          41.60           gcRMA 04/25/ 14:33:00
## 1857      56463.66          43.79           gcRMA 04/26/ 20:36:26
## 1858      55470.60          42.93           gcRMA 04/26/ 20:10:43
## 1859      49949.18          37.08           gcRMA 03/06/ 14:22:14
## 1860      51584.40          39.13           gcRMA 03/06/ 16:07:51
## 1861      50419.38          35.91           gcRMA 02/27/ 15:04:39
## 1862      53141.50          40.88           gcRMA 04/26/ 19:33:43
## 1863      56462.55          44.11           gcRMA 04/26/ 19:44:03
## 1864      53715.91          42.78           gcRMA 04/27/ 17:46:32
## 1865      54169.92          41.01           gcRMA 04/25/ 15:03:49
## 1866      53501.41          41.45           gcRMA 04/25/ 15:41:24
## 1867      51745.67          40.51           gcRMA 04/27/ 16:59:27
## 1868      54152.55          43.47           gcRMA 04/27/ 17:24:54
## 1869      53305.39          42.83           gcRMA 04/25/ 14:52:10
## 1870      55872.16          45.41           gcRMA 04/27/ 17:34:59
##      Chip.type.ID  CDF.name           Chip.type.name
## 1836      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1837      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1838      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1839      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1840      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1841      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1842      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1843      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1844      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1845      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1846      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1847      A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2

```

```

## 1848 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1849 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1850 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1851 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1852 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1853 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1854 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1855 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1856 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1857 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1858 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1859 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1860 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1861 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1862 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1863 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1864 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1865 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1866 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1867 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1868 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1869 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
## 1870 A-AFFY-6 MG_U74Av2 Affymetrix GeneChip Murine Genome U74Av2
##     IQRray.score.threshold.for.the.chip.type
## 1836                               44245.73
## 1837                               44245.73
## 1838                               44245.73
## 1839                               44245.73
## 1840                               44245.73
## 1841                               44245.73
## 1842                               44245.73
## 1843                               44245.73
## 1844                               44245.73
## 1845                               44245.73
## 1846                               44245.73
## 1847                               44245.73
## 1848                               44245.73
## 1849                               44245.73
## 1850                               44245.73
## 1851                               44245.73
## 1852                               44245.73
## 1853                               44245.73
## 1854                               44245.73
## 1855                               44245.73
## 1856                               44245.73
## 1857                               44245.73
## 1858                               44245.73
## 1859                               44245.73
## 1860                               44245.73
## 1861                               44245.73
## 1862                               44245.73
## 1863                               44245.73
## 1864                               44245.73
## 1865                               44245.73

```

```

## 1866          44245.73
## 1867          44245.73
## 1868          44245.73
## 1869          44245.73
## 1870          44245.73
##      MAS5.percent.present.threshold.for.the.chip.type Data.source
## 1836          26.68 ArrayExpress
## 1837          26.68 ArrayExpress
## 1838          26.68 ArrayExpress
## 1839          26.68 ArrayExpress
## 1840          26.68 ArrayExpress
## 1841          26.68 ArrayExpress
## 1842          26.68 ArrayExpress
## 1843          26.68 ArrayExpress
## 1844          26.68 ArrayExpress
## 1845          26.68 ArrayExpress
## 1846          26.68 ArrayExpress
## 1847          26.68 ArrayExpress
## 1848          26.68 ArrayExpress
## 1849          26.68 ArrayExpress
## 1850          26.68 ArrayExpress
## 1851          26.68 ArrayExpress
## 1852          26.68 ArrayExpress
## 1853          26.68 ArrayExpress
## 1854          26.68 ArrayExpress
## 1855          26.68 ArrayExpress
## 1856          26.68 ArrayExpress
## 1857          26.68 ArrayExpress
## 1858          26.68 ArrayExpress
## 1859          26.68 ArrayExpress
## 1860          26.68 ArrayExpress
## 1861          26.68 ArrayExpress
## 1862          26.68 ArrayExpress
## 1863          26.68 ArrayExpress
## 1864          26.68 ArrayExpress
## 1865          26.68 ArrayExpress
## 1866          26.68 ArrayExpress
## 1867          26.68 ArrayExpress
## 1868          26.68 ArrayExpress
## 1869          26.68 ArrayExpress
## 1870          26.68 ArrayExpress
##      Data.source.URL
## 1836          <NA>
## 1837          <NA>
## 1838          <NA>
## 1839          <NA>
## 1840          <NA>
## 1841          <NA>
## 1842          <NA>
## 1843          <NA>
## 1844          <NA>
## 1845          <NA>
## 1846          <NA>
## 1847          <NA>

```



```

## 1866 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymet
## 1867 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymet
## 1868 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymet
## 1869 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymet
## 1870 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/affymet
##                                     Bgee.normalized.data.file
## 1836 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1837 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1838 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1839 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1840 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1841 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1842 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1843 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1844 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1845 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1846 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1847 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1848 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1849 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1850 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1851 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1852 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1853 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1854 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1855 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1856 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1857 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1858 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1859 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1860 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1861 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1862 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1863 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1864 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1865 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1866 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1867 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1868 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1869 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
## 1870 Mus_musculus_probesets_E-MEXP-51_A-AFFY-6_gcRMA.tsv
##
##                                         Bgee.raw
## 1836      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Zygote
## 1837      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Zygote
## 1838      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Zygote
## 1839      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Zygote
## 1840      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/GVoocyte
## 1841      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/GVoocyte
## 1842      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/GVoocyte
## 1843      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/GVoocyte
## 1844      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/MIIoocyte
## 1845      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/MIIoocyte
## 1846      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/MIIoocyte
## 1847      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/16cel

```

```

## 1848      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/16cel...
## 1849      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Early2-cell...
## 1850      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Early2-cell...
## 1851      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Late2-cell...
## 1852      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Late2-cell...
## 1853      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Mid2-cell...
## 1854      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Mid2-cell...
## 1855      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/Mid2-cell...
## 1856      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/4Cell...
## 1857      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/4cell...
## 1858      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/4cell...
## 1859      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/8cell...
## 1860      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/8cel...
## 1861      ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/8cel...
## 1862  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/EarlyBlastocyst...
## 1863  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/EarlyBlastocyst...
## 1864  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/EarlyBlastocyst...
## 1865  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/EarlyBlastocyst...
## 1866  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/MidBlastocyst...
## 1867  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/MidBlastocyst...
## 1868  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/LateBlastocyst...
## 1869  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/LateBlastocyst...
## 1870  ftp://lausanne.isb-sib.ch/pub/databases/Bgee/affymetrix_data/cel_files/E-MEXP-51/LateBlastocyst...

data_E_MEXP_51 <- getData(bgee_affymetrix, experimentId="E-MEXP-51")

##
## Downloading expression data for the experiment E-MEXP-51 ...
##
## Saved expression data file in /Users/jroux/Dropbox/RpackagePaper/Mus_musculus_Bgee_13_2 folder. Now ...
##
## Read 95.1% of 315595 rows
## Read 315595 rows and 12 (of 12) columns from 0.050 GB file in 00:00:03
##
## Saving all data in .rds file...

head(data_E_MEXP_51)

##   Experiment.ID Chip.ID Probeset.ID          Gene.ID
## 1       E-MEXP-51 16cell1 100001_at ENSMUSG00000002033
## 2       E-MEXP-51 16cell1 100002_at ENSMUSG00000006522
## 3       E-MEXP-51 16cell1 100003_at ENSMUSG00000030592
## 4       E-MEXP-51 16cell1 100004_at ENSMUSG00000037461
## 5       E-MEXP-51 16cell1 100005_at ENSMUSG00000017386
## 6       E-MEXP-51 16cell1 100006_at ENSMUSG00000031673
##   Anatomical.entity.ID Anatomical.entity.name        Stage.ID
## 1           UBERON:0000085             morula MmusDv:0000006
## 2           UBERON:0000085             morula MmusDv:0000006
## 3           UBERON:0000085             morula MmusDv:0000006
## 4           UBERON:0000085             morula MmusDv:0000006
## 5           UBERON:0000085             morula MmusDv:0000006
## 6           UBERON:0000085             morula MmusDv:0000006
##   Stage.name Log.of.normalized.signal.intensity
```

```

## 1 Theiler stage 03 (mouse) 2.06
## 2 Theiler stage 03 (mouse) 2.06
## 3 Theiler stage 03 (mouse) 2.25
## 4 Theiler stage 03 (mouse) 6.48
## 5 Theiler stage 03 (mouse) 4.85
## 6 Theiler stage 03 (mouse) 2.06
##   Detection.flag Detection.quality State.in.Bgee
## 1      absent     high quality Used in no-expression call
## 2      absent     high quality Used in no-expression call
## 3      absent     high quality Used in no-expression call
## 4      present    high quality Used in expression call
## 5      absent     high quality Used in no-expression call
## 6      absent     high quality Used in no-expression call

```

```

data_E_MEXP_51_formatted <- formatData(bgee_affymetrix,
                                         data_E_MEXP_51,
                                         callType="all",
                                         stats="intensities")

```

```

##
## Extracting expression data matrix...
##
## Extracting features information...
##
## Extracting samples information...

```

```
data_E_MEXP_51_formatted
```

```

## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 9017 features, 35 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: 16cell1 16cell2 ... Zygote4 (35 total)
##   varLabels: Chip.ID Anatomical.entity.ID ... Stage.name (5 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: 100001_at 100002_at ...
##   AFFX-TransRecMur/X57349_M_at (9017 total)
##   fvarLabels: Probeset.ID Gene.ID
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:

```

```

# matrix of expression intensities
head(exprs(data_E_MEXP_51_formatted))

```

```

##          16cell1 16cell2 4Cell1 4cell2 4cell3 8cell1 8cell2 8cell3
## 100001_at    2.06    2.06    2.06    2.06    2.06    2.06    2.06    2.21
## 100002_at    2.06    2.06    2.06    2.06    2.06    2.06    2.06    2.06
## 100003_at    2.25    2.06    2.87    4.09    2.08    2.06    2.28    2.06
## 100004_at    6.48    6.94    5.69    5.68    5.77    7.41    7.41    6.16

```

```

## 100005_at    4.85    4.85    4.85    4.85    4.85    4.47    4.84    4.84
## 100006_at    2.06    2.06    2.06    2.06    2.06    2.06    2.06    2.06
##          Early2-cell1 Early2-cell2 EarlyBlastocyst1 EarlyBlastocyst2
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.06      2.06      2.06      2.06
## 100004_at      5.63      5.68      6.75      6.43
## 100005_at      4.85      4.85      4.85      4.85
## 100006_at      2.06      2.06      2.06      2.06
##          EarlyBlastocyst3 EarlyBlastocyst4 GVooocyte1 GVooocyte2 GVooocyte3
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.45      2.06      2.06      2.06
## 100004_at      6.35      5.68      5.68      4.23
## 100005_at      4.85      4.85      4.85      4.85
## 100006_at      2.06      2.06      2.06      2.06
##          GVooocyte4 Late2-cell1 Late2-cell2 LateBlastocyst1
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.06      2.06      2.06      2.06
## 100004_at      4.25      5.68      5.68      6.65
## 100005_at      4.85      5.29      5.16      4.85
## 100006_at      2.06      2.06      2.06      2.06
##          LateBlastocyst2 LateBlastocyst3 Mid2-cell1 Mid2-cell2 Mid2-cell3
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.09      2.78      2.06      2.06
## 100004_at      5.68      6.52      5.68      5.42
## 100005_at      4.85      4.85      7.36      4.85
## 100006_at      2.06      2.06      2.06      2.06
##          MidBlastocyst1 MidBlastocyst2 MIIoocyte1 MIIoocyte2 MIIoocyte3
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.06      2.06      2.06      2.06
## 100004_at      6.21      5.68      5.68      5.68
## 100005_at      4.83      4.85      4.85      4.85
## 100006_at      2.06      2.06      2.06      2.06
##          Zygote1 Zygote2 Zygote3 Zygote4
## 100001_at      2.06      2.06      2.06      2.06
## 100002_at      2.06      2.06      2.06      2.06
## 100003_at      2.06      2.06      2.06      2.06
## 100004_at      5.63      5.68      5.68      5.02
## 100005_at      4.85      4.85      4.85      6.70
## 100006_at      2.06      2.06      2.06      2.06

# annotation of samples
pData(data_E_MEXP_51_formatted)

```

```

##                               Chip.ID Anatomical.entity.ID
## 16cell1                  16cell1      UBERON:0000085
## 16cell2                  16cell2      UBERON:0000085
## 4Cell1                   4Cell1       UBERON:0000922
## 4cell2                   4cell2       UBERON:0000922
## 4cell3                   4cell3       UBERON:0000922

```

```

## 8cell1           8cell1           UBERON:0000922
## 8cell2           8cell2           UBERON:0000922
## 8cell3           8cell3           UBERON:0000922
## Early2-cell1    Early2-cell1    UBERON:0000922
## Early2-cell2    Early2-cell2    UBERON:0000922
## EarlyBlastocyst1 EarlyBlastocyst1 UBERON:0000922
## EarlyBlastocyst2 EarlyBlastocyst2 UBERON:0000922
## EarlyBlastocyst3 EarlyBlastocyst3 UBERON:0000922
## EarlyBlastocyst4 EarlyBlastocyst4 UBERON:0000922
## GVocyte1         GVocyte1         CL:0000654
## GVocyte2         GVocyte2         CL:0000654
## GVocyte3         GVocyte3         CL:0000654
## GVocyte4         GVocyte4         CL:0000654
## Late2-cell1     Late2-cell1     UBERON:0000922
## Late2-cell2     Late2-cell2     UBERON:0000922
## LateBlastocyst1 LateBlastocyst1 UBERON:0000922
## LateBlastocyst2 LateBlastocyst2 UBERON:0000922
## LateBlastocyst3 LateBlastocyst3 UBERON:0000922
## Mid2-cell11     Mid2-cell11    UBERON:0000922
## Mid2-cell12     Mid2-cell12    UBERON:0000922
## Mid2-cell13     Mid2-cell13    UBERON:0000922
## MidBlastocyst1  MidBlastocyst1 UBERON:0000922
## MidBlastocyst2  MidBlastocyst2 UBERON:0000922
## MIIoocyte1       MIIoocyte1       CL:0000655
## MIIoocyte2       MIIoocyte2       CL:0000655
## MIIoocyte3       MIIoocyte3       CL:0000655
## Zygote1          Zygote1          CL:0000365
## Zygote2          Zygote2          CL:0000365
## Zygote3          Zygote3          CL:0000365
## Zygote4          Zygote4          CL:0000365
## Anatomical.entity.name   Stage.ID
## 16cell1          morula          MmusDv:0000006
## 16cell2          morula          MmusDv:0000006
## 4Cell1           embryo          MmusDv:0000006
## 4cell2           embryo          MmusDv:0000006
## 4cell3           embryo          MmusDv:0000006
## 8cell1           embryo          MmusDv:0000006
## 8cell2           embryo          MmusDv:0000006
## 8cell3           embryo          MmusDv:0000006
## Early2-cell1    embryo          MmusDv:0000005
## Early2-cell2    embryo          MmusDv:0000005
## EarlyBlastocyst1 embryo          MmusDv:0000008
## EarlyBlastocyst2 embryo          MmusDv:0000008
## EarlyBlastocyst3 embryo          MmusDv:0000008
## EarlyBlastocyst4 embryo          MmusDv:0000008
## GVocyte1         primary oocyte  UBERON:0000113
## GVocyte2         primary oocyte  UBERON:0000113
## GVocyte3         primary oocyte  UBERON:0000113
## GVocyte4         primary oocyte  UBERON:0000113
## Late2-cell1     embryo          MmusDv:0000005
## Late2-cell2     embryo          MmusDv:0000005
## LateBlastocyst1 embryo          MmusDv:0000010
## LateBlastocyst2 embryo          MmusDv:0000010
## LateBlastocyst3 embryo          MmusDv:0000010

```

```

## Mid2-cell11                     embryo MmusDv:0000005
## Mid2-cell12                     embryo MmusDv:0000005
## Mid2-cell13                     embryo MmusDv:0000005
## MidBlastocyst1                  embryo MmusDv:0000009
## MidBlastocyst2                  embryo MmusDv:0000009
## MIIoocyte1                      secondary oocyte UBERON:0000113
## MIIoocyte2                      secondary oocyte UBERON:0000113
## MIIoocyte3                      secondary oocyte UBERON:0000113
## Zygote1                         zygote UBERON:0000106
## Zygote2                         zygote UBERON:0000106
## Zygote3                         zygote UBERON:0000106
## Zygote4                         zygote UBERON:0000106
##
## Stage.name
## 16cell1                          Theiler stage 03 (mouse)
## 16cell2                          Theiler stage 03 (mouse)
## 4Cell1                           Theiler stage 03 (mouse)
## 4cell2                           Theiler stage 03 (mouse)
## 4cell3                           Theiler stage 03 (mouse)
## 8cell1                           Theiler stage 03 (mouse)
## 8cell2                           Theiler stage 03 (mouse)
## 8cell3                           Theiler stage 03 (mouse)
## Early2-cell1                     Theiler stage 02 (mouse)
## Early2-cell2                     Theiler stage 02 (mouse)
## EarlyBlastocyst1                 Theiler stage 04 (mouse)
## EarlyBlastocyst2                 Theiler stage 04 (mouse)
## EarlyBlastocyst3                 Theiler stage 04 (mouse)
## EarlyBlastocyst4                 Theiler stage 04 (mouse)
## GVoocyte1                        post-juvenile adult stage
## GVoocyte2                        post-juvenile adult stage
## GVoocyte3                        post-juvenile adult stage
## GVoocyte4                        post-juvenile adult stage
## Late2-cell1                      Theiler stage 02 (mouse)
## Late2-cell2                      Theiler stage 02 (mouse)
## LateBlastocyst1                  Theiler stage 06 (mouse)
## LateBlastocyst2                  Theiler stage 06 (mouse)
## LateBlastocyst3                  Theiler stage 06 (mouse)
## Mid2-cell11                      Theiler stage 02 (mouse)
## Mid2-cell12                      Theiler stage 02 (mouse)
## Mid2-cell13                      Theiler stage 02 (mouse)
## MidBlastocyst1                  Theiler stage 05 (mouse)
## MidBlastocyst2                  Theiler stage 05 (mouse)
## MIIoocyte1                       post-juvenile adult stage
## MIIoocyte2                       post-juvenile adult stage
## MIIoocyte3                       post-juvenile adult stage
## Zygote1                          zygote stage
## Zygote2                          zygote stage
## Zygote3                          zygote stage
## Zygote4                          zygote stage

```

```

# annotation of probesets
head(fData(data_E_MEXP_51_formatted))

```

```

##             Probeset.ID          Gene.ID
## 100001_at    100001_at ENSMUSG00000002033

```

```

## 100002_at    100002_at ENSMUSG00000006522
## 100003_at    100003_at ENSMUSG000000030592
## 100004_at    100004_at ENSMUSG000000037461
## 100005_at    100005_at ENSMUSG000000017386
## 100006_at    100006_at ENSMUSG000000031673

# specify species and data type
bgee_rnaseq <- Bgee$new(species="Macaca_mulatta", dataType="rna_seq", release="13.2")

##
## Querying Bgee to get release information...
##
## NOTE: the file describing Bgee species information for release 13_2 was found in the download directo
##
## API key built: 696a8ee5f27f212d71349c131d413fb0de7947e688c95fcdd001689789c7333dbdf6fad8d4d48c239a3e0

# retrieve annotation of RNA-seq samples and experiments
annotation_bgee_macaque_rna_seq <- getAnnotation(bgee_rnaseq)

##
## Downloading annotation files...
##
## Saved annotation files in /Users/jroux/Dropbox/RpackagePaper/Macaca_mulatta_Bgee_13_2 folder.

sample_annotation <- annotation_bgee_macaque_rna_seq$sample.annotation
experiment_annotation <- annotation_bgee_macaque_rna_seq$experiment.annotation

# list experiments including both brain and liver samples
selected_experiments <-
  intersect(unique(sample_annotation$Experiment.ID[sample_annotation$Anatomical.entity.ID
                                                 == "UBERON:0000955"]),
            unique(sample_annotation$Experiment.ID[sample_annotation$Anatomical.entity.ID
                                                 == "UBERON:0002107"]))
experiment_annotation[experiment_annotation$Experiment.ID %in% selected_experiments,]

##   Experiment.ID Library.count Condition.count Organ.count Stage.count
## 1      GSE41637          27           9         9         1
## 2      GSE30352          13           8         7         3
##   Data.source                         Data.source.URL
## 1      GEO http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41637
## 2      GEO http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30352
##
## 1 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/rna_seq/Mac
## 2 ftp://lausanne.isb-sib.ch/pub/databases/Bgee/bgee_v13_1/download/processed_expr_values/rna_seq/Mac
##                                         Experiment.name
## 1 Evolutionary dynamics of gene and isoform regulation in mammalian tissues
## 2             The evolution of gene expression levels in mammalian organs
##
## 1
## 2 Changes in gene expression are thought to underlie many of the phenotypic differences between speci

```

```
# check whether experiments include biological replicates
sample_annotation[sample_annotation$Experiment.ID %in% selected_experiments &
  (sample_annotation$Anatomical.entity.ID == "UBERON:0000955"
  | sample_annotation$Anatomical.entity.ID == "UBERON:0002107"), 1:6]
```

	Experiment.ID	Library.ID	Library.secondary.ID	Anatomical.entity.ID
## 10	GSE41637	GSM1020693	SRX196317	UBERON:0000955
## 11	GSE41637	GSM1020702	SRX196326	UBERON:0000955
## 12	GSE41637	GSM1020711	SRX196335	UBERON:0000955
## 25	GSE41637	GSM1020697	SRX196321	UBERON:0002107
## 26	GSE41637	GSM1020706	SRX196330	UBERON:0002107
## 27	GSE41637	GSM1020715	SRX196339	UBERON:0002107
## 35	GSE30352	GSM752631	SRX081922	UBERON:0000955
## 36	GSE30352	GSM752632	SRX081923	UBERON:0000955
## 39	GSE30352	GSM752640	SRX081931	UBERON:0002107
## 40	GSE30352	GSM752641	SRX081932	UBERON:0002107
	Anatomical.entity.name		Stage.ID	
## 10	brain	UBERON:0018241		
## 11	brain	UBERON:0018241		
## 12	brain	UBERON:0018241		
## 25	liver	UBERON:0018241		
## 26	liver	UBERON:0018241		
## 27	liver	UBERON:0018241		
## 35	brain	MmulDv:0000028		
## 36	brain	MmulDv:0000028		
## 39	liver	MmulDv:0000028		
## 40	liver	MmulDv:0000028		

```
data_GSE41637 <- getData(bgee_rnaseq, experimentId="GSE41637")
```

```
##
## Downloading expression data for the experiment GSE41637 ...
##
## Saved expression data file in /Users/jroux/Dropbox/RpackagePaper/Macaca_mulatta_Bgee_13_2 folder. No
##
Read 35.5% of 816642 rows
Read 84.5% of 816642 rows
Read 816642 rows and 13 (of 13) columns from 0.126 GB file in 00:00:04
##
## Saving all data in .rds file...
```

```
data_GSE41637_formatted <- formatData(bgee_rnaseq,
                                         data_GSE41637,
                                         callType="all",
                                         stats="counts")
```

```
##
## Extracting expression data matrix...
##
## Extracting features information...
##
## Extracting samples information...
```

```

data_GSE41637_formatted

## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 30246 features, 27 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1020693 GSM1020694 ... GSM1020719 (27 total)
##   varLabels: Library.ID Anatomical.entity.ID ... Stage.name (5
##     total)
##   varMetadata: labelDescription
## featureData
##   featureNames: ENSMMUG000000000001 ENSMMUG000000000002 ...
##     ENSMMUG00000037447 (30246 total)
##   fvarLabels: Gene.ID
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:

# for simplicity, keep only one sample per condition
data_E_MEXP_51_formatted <-
  data_E_MEXP_51_formatted[, !duplicated(pData(data_E_MEXP_51_formatted)[2:5])]

# order developmental stages
data_E_MEXP_51_formatted <- data_E_MEXP_51_formatted[, c(5,8,9,3,2,1,4,7,6)]

# filter out rows with no variance
data_E_MEXP_51_formatted <-
  data_E_MEXP_51_formatted[apply(exprs(data_E_MEXP_51_formatted), 1, sd) != 0, ]

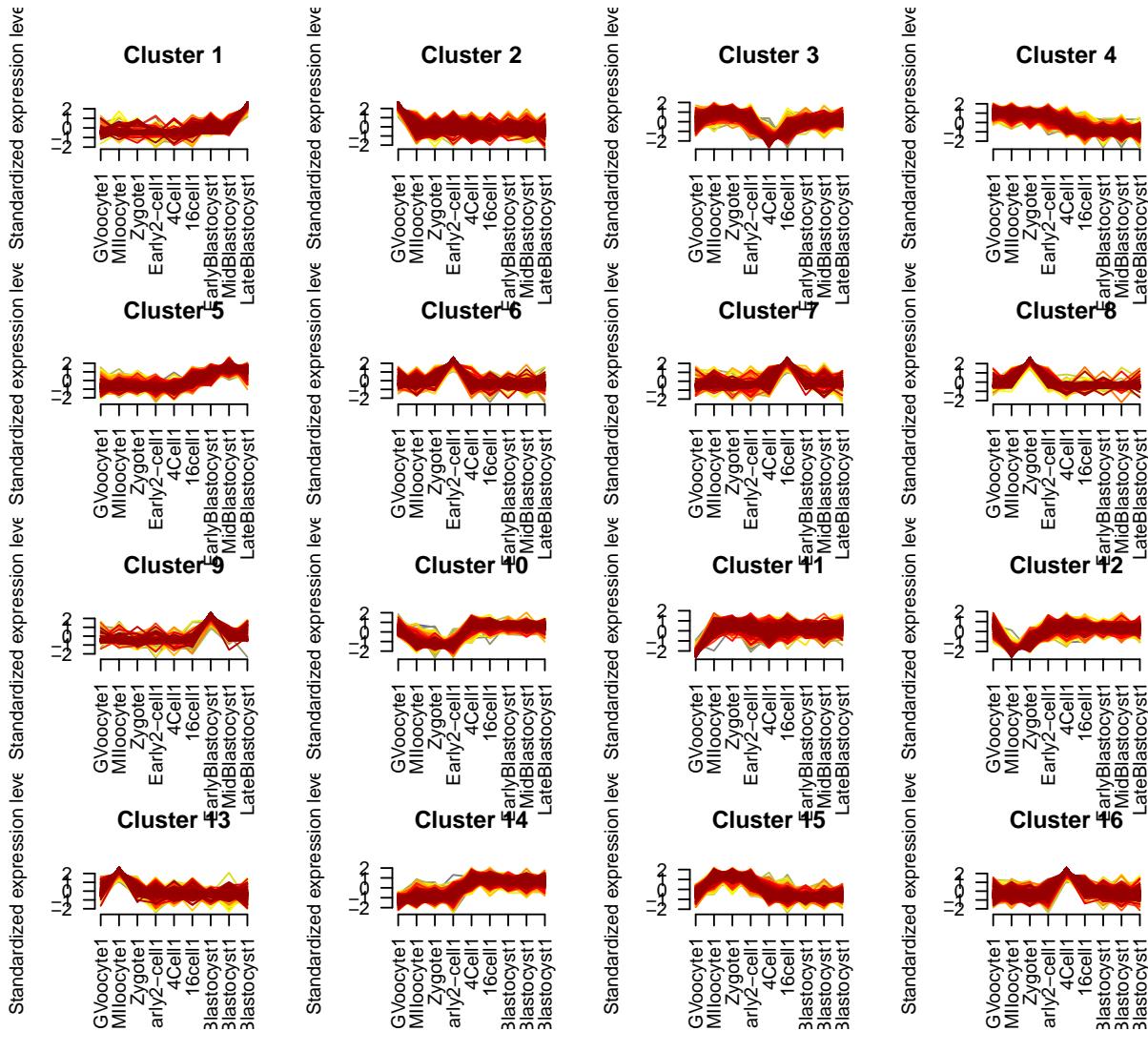
library(Mfuzz)

## Loading required package: e1071

# standardize matrix of expression data
z.mat <- standardise(data_E_MEXP_51_formatted)
# cluster data into 16 clusters
clusters <- mfuzz(z.mat, centers=16, m=1.25)

# visualizing clusters
mfuzz.plot2(z.mat,
            cl=clusters,
            mfrow=c(4,4),
            colo="fancy",
            time.labels=row.names(pData(z.mat)),
            las=2, xlab="", ylab="Standardized expression level",
            x11 = FALSE)

```



```
# differential expression analysis with edgeR
library(edgeR)
```

```
## Loading required package: limma

## 
## Attaching package: 'limma'

## The following object is masked from 'package:BiocGenerics':
## 
##     plotMA

# subset the dataset to brain and liver
brain_liver <-
  data_GSE41637_formatted[, pData(data_GSE41637_formatted)$Anatomical.entity.name
                           %in% c("brain", "liver")]

# filter out very lowly expressed genes
```

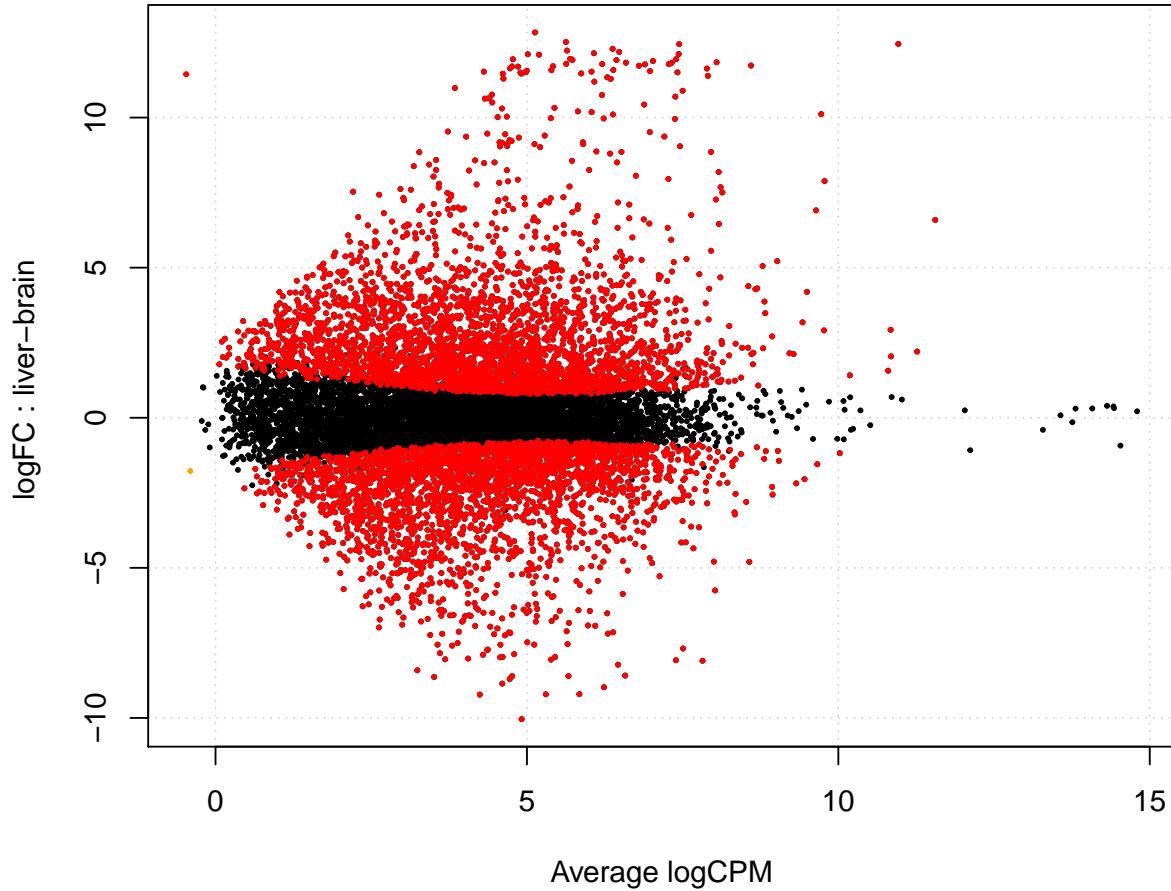
```

brain_liver_filtered <- brain_liver[rowSums(cpm(brain_liver) > 1) > 3, ]

# create edgeR DGEList object
dge <- DGEList(counts= brain_liver_filtered,
                 group=pData(brain_liver_filtered)$Anatomical.entity.name)
dge <- calcNormFactors(dge)
dge <- estimateCommonDisp(dge)
dge <- estimateTagwiseDisp(dge)
de <- exactTest(dge, pair=c("brain","liver"))
de.genes <- topTags(de, n=nrow(de))$table

# MA plot with DE genes highlighted
plotSmear(dge, de.tags=rownames(de.genes)[de.genes$FDR < 0.01], cex=0.3)

```



```
bgee_topanat <- Bgee$new(species="Danio_rerio", release="13.2")
```

```
##
```

```
## NOTE: You did not specify any data type. The argument dataType will be set to c("rna_seq","affymetrix")
```

```

## 
## Querying Bgee to get release information...
## 
## NOTE: the file describing Bgee species information for release 13_2 was found in the download directo...
## 
## API key built: 696a8ee5f27f212d71349c131d413fb0de7947e688c95fcdd001689789c7333dbdf6fad8d4d48c239a3e0...
myTopAnatData <- loadTopAnatData(bgee_topanat)

## 
## Building URLs to retrieve organ relationships from Bgee.....
## URL successfully built (http://r13\_2.bgee.org/?page=dao&action=org.bgee.model.dao.api.ontologycomm)
## Submitting URL to Bgee webservice (can be long)
## Got results from Bgee webservice. Files are written in "/Users/jroux/Dropbox/RpackagePaper/Danio_...
## 
## Building URLs to retrieve organ names from Bgee.....
## URL successfully built (http://r13\_2.bgee.org/?page=dao&action=org.bgee.model.dao.api.anatdev.Anat...)
## Submitting URL to Bgee webservice (can be long)
## Got results from Bgee webservice. Files are written in "/Users/jroux/Dropbox/RpackagePaper/Danio_...
## 
## Building URLs to retrieve mapping of gene to organs from Bgee...
## URL successfully built (http://r13\_2.bgee.org/?page=dao&action=org.bgee.model.dao.api.expressionond)
## Submitting URL to Bgee webservice (can be long)
## Got results from Bgee webservice. Files are written in "/Users/jroux/Dropbox/RpackagePaper/Danio_...
## 
## Parsing the results.....
## 
## Adding BGEE:0 as unique root of all terms of the ontology.....
## 
## Done.

str(myTopAnatData)

## List of 4
## $ gene2anatomy      :List of 18715
##   ..$ ENSDARG000000000001: chr [1:3] "UBERON:0000468" "UBERON:0001997" "ZFA:0001093"
##   ..$ ENSDARG000000000002: chr [1:11] "UBERON:0000019" "UBERON:0000468" "UBERON:0000965" "UBERON:0002...
##   ..$ ENSDARG000000000018: chr [1:28] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000...
##   ..$ ENSDARG000000000019: chr [1:23] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000...
##   ..$ ENSDARG000000000068: chr [1:53] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000...
##   ..$ ENSDARG000000000069: chr [1:23] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000...
##   ..$ ENSDARG000000000102: chr [1:7] "UBERON:0000151" "UBERON:0000468" "UBERON:0001774" "UBERON:00023...
##   ..$ ENSDARG000000000103: chr [1:20] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000...
##   ..$ ENSDARG000000000142: chr [1:4] "UBERON:0000080" "UBERON:0000468" "UBERON:0000992" "UBERON:00015...
##   ..$ ENSDARG000000000151: chr [1:13] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000...
##   ..$ ENSDARG000000000161: chr "UBERON:0000955"
##   ..$ ENSDARG000000000175: chr [1:36] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000...
##   ..$ ENSDARG000000000183: chr [1:10] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000...
##   ..$ ENSDARG000000000189: chr [1:6] "UBERON:0000965" "UBERON:0001891" "UBERON:0001893" "UBERON:00018...
##   ..$ ENSDARG000000000212: chr [1:20] "UBERON:0000004" "UBERON:0000019" "UBERON:0000080" "UBERON:0000...
##   ..$ ENSDARG000000000229: chr [1:4] "UBERON:0000307" "UBERON:0000468" "UBERON:0000992" "UBERON:00070...
##   ..$ ENSDARG000000000241: chr [1:22] "UBERON:0000080" "UBERON:0000468" "UBERON:0000922" "UBERON:0000...
##   ..$ ENSDARG000000000324: chr "UBERON:0000473"

```

```

## ..$ ENSDARG00000000369: chr [1:2] "UBERON:0000080" "UBERON:0000468"
## ..$ ENSDARG00000000370: chr "UBERON:0000468"
## ..$ ENSDARG00000000380: chr [1:9] "UBERON:0000019" "UBERON:0000468" "UBERON:0000922" "UBERON:00009
## ..$ ENSDARG00000000384: chr [1:25] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:00009
## ..$ ENSDARG00000000394: chr [1:17] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:00009
## ..$ ENSDARG00000000423: chr [1:3] "UBERON:0000468" "UBERON:0000473" "UBERON:0000992"
## ..$ ENSDARG00000000442: chr [1:20] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000000460: chr "UBERON:0000948"
## ..$ ENSDARG00000000472: chr [1:24] "UBERON:0000468" "UBERON:0000955" "UBERON:0001675" "UBERON:0001
## ..$ ENSDARG00000000476: chr [1:16] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000489: chr [1:18] "UBERON:0000307" "UBERON:0000468" "UBERON:0000473" "UBERON:0000
## ..$ ENSDARG00000000503: chr [1:14] "UBERON:0000019" "UBERON:0000468" "UBERON:0000473" "UBERON:0000
## ..$ ENSDARG00000000516: chr [1:3] "UBERON:0000468" "UBERON:0000473" "UBERON:0004734"
## ..$ ENSDARG00000000529: chr [1:22] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000540: chr [1:20] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000551: chr [1:22] "UBERON:0000468" "UBERON:0000922" "UBERON:0000948" "UBERON:0000
## ..$ ENSDARG00000000563: chr [1:31] "UBERON:0000019" "UBERON:0000080" "UBERON:0000151" "UBERON:0000
## ..$ ENSDARG00000000567: chr [1:2] "UBERON:0000019" "UBERON:0000468"
## ..$ ENSDARG00000000568: chr [1:20] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000588: chr [1:2] "UBERON:0000468" "UBERON:0001997"
## ..$ ENSDARG00000000606: chr [1:2] "UBERON:0000019" "UBERON:0000468"
## ..$ ENSDARG00000000631: chr [1:21] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000000638: chr [1:3] "UBERON:0000019" "UBERON:0000468" "UBERON:0000966"
## ..$ ENSDARG00000000651: chr [1:14] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000
## ..$ ENSDARG00000000656: chr [1:12] "UBERON:0000019" "UBERON:0000080" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000666: chr [1:6] "UBERON:0000019" "UBERON:0000080" "UBERON:0000468" "UBERON:00004
## ..$ ENSDARG00000000690: chr [1:24] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000729: chr [1:19] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000000730: chr [1:4] "UBERON:0000468" "UBERON:0002107" "UBERON:0008897" "ZFA:0001093"
## ..$ ENSDARG00000000747: chr [1:3] "UBERON:0000019" "UBERON:0000468" "UBERON:0001555"
## ..$ ENSDARG00000000760: chr [1:20] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000767: chr [1:30] "UBERON:0000019" "UBERON:0000033" "UBERON:0000080" "UBERON:0000
## ..$ ENSDARG00000000779: chr [1:24] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000000796: chr [1:31] "UBERON:0000004" "UBERON:0000019" "UBERON:0000080" "UBERON:0000
## ..$ ENSDARG00000000804: chr [1:3] "UBERON:0000468" "UBERON:0001997" "ZFA:0001093"
## ..$ ENSDARG00000000830: chr [1:5] "UBERON:0000080" "UBERON:0000468" "UBERON:0000992" "UBERON:00021
## ..$ ENSDARG00000000837: chr [1:3] "UBERON:0000178" "UBERON:0000468" "UBERON:0001997"
## ..$ ENSDARG00000000853: chr [1:22] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000000857: chr [1:34] "UBERON:0000004" "UBERON:0000080" "UBERON:0000151" "UBERON:0000
## ..$ ENSDARG00000000861: chr [1:13] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000000887: chr [1:29] "UBERON:0000019" "UBERON:0000080" "UBERON:0000151" "UBERON:0000
## ..$ ENSDARG00000000906: chr [1:6] "UBERON:0000080" "UBERON:0000468" "UBERON:0001555" "UBERON:00019
## ..$ ENSDARG00000000935: chr [1:18] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000
## ..$ ENSDARG00000000943: chr [1:3] "UBERON:0000080" "UBERON:0000468" "ZFA:0001093"
## ..$ ENSDARG00000000966: chr [1:26] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000
## ..$ ENSDARG00000000972: chr [1:23] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG000000001014: chr [1:4] "UBERON:0000019" "UBERON:0000080" "UBERON:0000468" "UBERON:00088
## ..$ ENSDARG000000001015: chr [1:3] "UBERON:0000019" "UBERON:0000468" "UBERON:0000955"
## ..$ ENSDARG000000001057: chr [1:33] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG000000001127: chr [1:5] "CL:0005000" "UBERON:0000468" "UBERON:0000955" "UBERON:0001898"
## ..$ ENSDARG000000001129: chr [1:20] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG000000001162: chr [1:17] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000
## ..$ ENSDARG000000001169: chr [1:14] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG000000001210: chr [1:7] "UBERON:0000019" "UBERON:0000080" "UBERON:0000468" "UBERON:00004
```

```

## ..$ ENSDARG00000001220: chr [1:23] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000001234: chr [1:3] "UBERON:0000468" "UBERON:0000955" "UBERON:0000966"
## ..$ ENSDARG00000001241: chr [1:27] "UBERON:0000019" "UBERON:0000080" "UBERON:0000151" "UBERON:0000
## ..$ ENSDARG00000001244: chr [1:22] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000001249: chr [1:4] "UBERON:0000080" "UBERON:0000468" "UBERON:0000473" "UBERON:00009
## ..$ ENSDARG00000001285: chr [1:2] "UBERON:0000468" "UBERON:0001997"
## ..$ ENSDARG00000001303: chr [1:16] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000001313: chr [1:25] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000001354: chr "UBERON:0000468"
## ..$ ENSDARG00000001414: chr [1:7] "UBERON:0000468" "UBERON:0000473" "UBERON:0000922" "UBERON:00019
## ..$ ENSDARG00000001431: chr [1:28] "UBERON:0000019" "UBERON:0000080" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000001437: chr [1:8] "UBERON:0000033" "UBERON:0000468" "UBERON:0000966" "UBERON:00010
## ..$ ENSDARG00000001442: chr [1:13] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000001447: chr [1:3] "UBERON:0000307" "UBERON:0000468" "UBERON:0007010"
## ..$ ENSDARG00000001452: chr [1:13] "UBERON:0000080" "UBERON:0000468" "UBERON:0000473" "UBERON:0000
## ..$ ENSDARG00000001463: chr [1:4] "UBERON:0000019" "UBERON:0000468" "UBERON:0000966" "UBERON:00019
## ..$ ENSDARG00000001498: chr [1:4] "UBERON:0000080" "UBERON:0000468" "UBERON:0000992" "UBERON:00018
## ..$ ENSDARG00000001549: chr [1:13] "UBERON:0000080" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000001557: chr [1:25] "UBERON:0000019" "UBERON:0000307" "UBERON:0000468" "UBERON:0000
## ..$ ENSDARG00000001558: chr [1:22] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000001559: chr "ZFA:0001093"
## ..$ ENSDARG00000001572: chr "UBERON:0000080"
## ..$ ENSDARG00000001578: chr [1:28] "UBERON:0000019" "UBERON:0000080" "UBERON:0000307" "UBERON:0000
## ..$ ENSDARG00000001621: chr "UBERON:0000468"
## ..$ ENSDARG00000001634: chr [1:13] "UBERON:0000019" "UBERON:0000080" "UBERON:0000151" "UBERON:0000
## ..$ ENSDARG00000001637: chr [1:16] "UBERON:0000307" "UBERON:0000468" "UBERON:0000922" "UBERON:0000
## ..$ ENSDARG00000001656: chr [1:13] "UBERON:0000019" "UBERON:0000468" "UBERON:0000473" "UBERON:0000
## ... [list output truncated]
## -- attr(*, "dim")= int 18715
## -- attr(*, "dimnames")=List of 1
## ... : chr [1:18715] "ENSDARG00000000001" "ENSDARG00000000002" "ENSDARG00000000018" "ENSDARG0000
## $ organ.relationships:List of 12587
## ..$ AEO:0000013 : chr "UBERON:0000479"
## ..$ AEO:0000127 : chr "UBERON:0005423"
## ..$ AEO:0000129 : chr "UBERON:0005423"
## ..$ AEO:0000131 : chr "UBERON:0005423"
## ..$ AEO:0000153 : chr "AEO:0000187"
## ..$ AEO:0000162 : chr "UBERON:0006984"
## ..$ AEO:0000173 : chr [1:2] "UBERON:0002416" "UBERON:0000020"
## ..$ AEO:0000180 : chr "UBERON:0002418"
## ..$ AEO:0000185 : chr [1:2] "UBERON:0004288" "UBERON:0007475"
## ..$ AEO:0000187 : chr "UBERON:0000481"
## ..$ AEO:0000188 : chr [1:2] "UBERON:0000463" "UBERON:0000476"
## ..$ AEO:0000189 : chr "AEO:0000187"
## ..$ AEO:0000196 : chr "UBERON:0002548"
## ..$ AEO:0000210 : chr "UBERON:0000480"
## ..$ AEO:0000211 : chr [1:2] "UBERON:0000477" "UBERON:0005423"
## ..$ AEO:0000215 : chr "UBERON:0010064"
## ..$ AEO:0000217 : chr "UBERON:0005423"
## ..$ AEO:0001003 : chr "UBERON:0000476"
## ..$ AEO:0001006 : chr "UBERON:0000463"
## ..$ AEO:0001007 : chr [1:2] "UBERON:0000483" "UBERON:0005423"
## ..$ AEO:0001009 : chr "UBERON:0004022"
## ..$ AEO:0001010 : chr "UBERON:0004022"

```

```

## ..$ AEO:0001013 : chr [1:2] "UBERON:0010314" "UBERON:0002020"
## ..$ AEO:0001014 : chr "UBERON:0005423"
## ..$ AEO:0001019 : chr "UBERON:0005423"
## ..$ AEO:0001020 : chr "UBERON:0000479"
## ..$ AEO:0001021 : chr "UBERON:0000479"
## ..$ CARO:0000049 : chr "UBERON:0010895"
## ..$ CARO:0000050 : chr "UBERON:0010895"
## ..$ CARO:0000064 : chr "CL:0000003"
## ..$ CL:0000000 : chr "GO:0005575"
## ..$ CL:0000001 : chr "CL:0000010"
## ..$ CL:0000003 : chr [1:3] "GO:0005623" "CL:0000000" "UBERON:0000061"
## ..$ CL:0000005 : chr [1:2] "UBERON:0010313" "CL:0000057"
## ..$ CL:0000006 : chr [1:2] "CL:0000197" "CL:0000101"
## ..$ CL:0000007 : chr "CL:0002321"
## ..$ CL:0000008 : chr "CL:0000333"
## ..$ CL:0000010 : chr "CL:0000578"
## ..$ CL:0000011 : chr [1:2] "UBERON:0003083" "CL:0000333"
## ..$ CL:0000014 : chr [1:2] "CL:0000039" "CL:0000034"
## ..$ CL:0000015 : chr "CL:0000586"
## ..$ CL:0000016 : chr [1:2] "CL:0000014" "CL:0000015"
## ..$ CL:0000017 : chr [1:2] "UBERON:0000473" "CL:0000015"
## ..$ CL:0000018 : chr [1:2] "CL:0000015" "CL:0000413"
## ..$ CL:0000019 : chr [1:3] "CL:0000064" "CL:0000219" "CL:0000408"
## ..$ CL:0000020 : chr [1:2] "UBERON:0000473" "CL:0000015"
## ..$ CL:0000021 : chr "CL:0000586"
## ..$ CL:0000022 : chr [1:2] "CL:0000014" "CL:0000021"
## ..$ CL:0000023 : chr [1:2] "UBERON:0001305" "CL:0000021"
## ..$ CL:0000024 : chr "CL:0000021"
## ..$ CL:0000025 : chr "CL:0000675"
## ..$ CL:0000026 : chr [1:2] "CL:0000039" "CL:0000412"
## ..$ CL:0000027 : chr [1:2] "UBERON:0010313" "CL:0000192"
## ..$ CL:0000028 : chr "CL:2000029"
## ..$ CL:0000029 : chr [1:2] "UBERON:0010313" "CL:0000540"
## ..$ CL:0000030 : chr [1:2] "CL:0000055" "UBERON:0004121"
## ..$ CL:0000031 : chr [1:2] "CL:0000047" "CL:0000055"
## ..$ CL:0000032 : chr [1:2] "CL:0002321" "UBERON:0004121"
## ..$ CL:0000033 : chr "CL:0000151"
## ..$ CL:0000034 : chr "CL:0000003"
## ..$ CL:0000035 : chr "CL:0000723"
## ..$ CL:0000036 : chr "CL:0000035"
## ..$ CL:0000037 : chr [1:3] "UBERON:0002390" "CL:0000048" "CL:0008001"
## ..$ CL:0000038 : chr [1:2] "CL:0000764" "CL:0000839"
## ..$ CL:0000039 : chr "CL:0000548"
## ..$ CL:0000040 : chr [1:3] "CL:0000763" "CL:0002194" "UBERON:0004120"
## ..$ CL:0000041 : chr [1:2] "CL:0000771" "CL:0002242"
## ..$ CL:0000042 : chr [1:2] "CL:0000835" "CL:0000834"
## ..$ CL:0000043 : chr [1:3] "CL:0000767" "CL:0002274" "CL:0002242"
## ..$ CL:0000047 : chr [1:3] "CL:0000048" "CL:0002319" "UBERON:0001017"
## ..$ CL:0000048 : chr "CL:0000723"
## ..$ CL:0000049 : chr [1:2] "UBERON:0004120" "CL:0002032"
## ..$ CL:0000050 : chr [1:3] "CL:0000839" "CL:0000763" "CL:0002032"
## ..$ CL:0000051 : chr [1:2] "UBERON:0004120" "CL:0002032"
## ..$ CL:0000052 : chr "CL:0000723"
## ..$ CL:0000054 : chr "CL:0000151"

```

```

##   ..$ CL:0000055    : chr "CL:0000003"
##   ..$ CL:0000056    : chr "CL:0000680"
##   ..$ CL:0000057    : chr [1:2] "UBERON:0004120" "CL:0002320"
##   ..$ CL:0000058    : chr [1:2] "CL:0000667" "CL:0000055"
##   ..$ CL:0000059    : chr [1:3] "CL:0000151" "CL:0000146" "UBERON:0005176"
##   ..$ CL:0000060    : chr [1:4] "UBERON:0001754" "CL:0000055" "CL:0002159" "UBERON:0005176"
##   ..$ CL:0000061    : chr [1:4] "UBERON:0000078" "CL:0000062" "CL:0000151" "UBERON:0005176"
##   ..$ CL:0000062    : chr [1:4] "UBERON:0010313" "CL:0000055" "CL:0002320" "UBERON:0004120"
##   ..$ CL:0000064    : chr "CL:0000003"
##   ..$ CL:0000065    : chr [1:3] "UBERON:0001016" "CL:0000067" "CL:0000710"
##   ..$ CL:0000066    : chr [1:2] "UBERON:0000483" "CL:0000548"
##   ..$ CL:0000067    : chr [1:2] "CL:0000075" "CL:0000064"
##   ..$ CL:0000068    : chr [1:3] "CL:0000066" "CL:0002371" "UBERON:0000058"
##   ..$ CL:0000069    : chr "CL:0000068"
##   ..$ CL:0000071    : chr [1:3] "CL:0000076" "CL:0002139" "UBERON:0004638"
##   ..$ CL:0000072    : chr [1:2] "CL:0000068" "UBERON:0004120"
##   ..$ CL:0000073    : chr [1:2] "CL:0000066" "CL:0002371"
##   ..$ CL:0000075    : chr [1:2] "CL:0000066" "CL:0002371"
##   ..$ CL:0000076    : chr "CL:0000066"
##   ..$ CL:0000077    : chr [1:3] "CL:0000213" "CL:0000076" "UBERON:0004120"
##   ..$ CL:0000078    : chr [1:2] "CL:0000076" "CL:0002371"
##   ..$ CL:0000079    : chr [1:2] "CL:0000066" "CL:0002371"
##   ..$ CL:0000080    : chr [1:2] "CL:0000003" "UBERON:0000179"
##   ... [list output truncated]
## $ organ.names      :'data.frame': 12588 obs. of  2 variables:
##   ..$ ID : chr [1:12588] "AEO:0001009" "AEO:0001010" "AEO:0001013" "CL:0000005" ...
##   ..$ NAME: chr [1:12588] "proliferating neuroepithelium" "differentiating neuroepithelium" "neurona...
## $ bgee.object      :Reference class 'Bgee' [package "BgeeDB"] with 13 fields
##   ..$ species       : chr "Danio_rerio"
##   ..$ speciesName   : chr "Danio_rerio"
##   ..$ speciesId     : num 7955
##   ..$ dataType      : chr [1:4] "rna_seq" "affymetrix" "est" "in_situ"
##   ..$ pathToData    : chr "/Users/jroux/Dropbox/RpackagePaper/Danio_rerio_Bgee_13_2"
##   ..$ release       : chr "13_2"
##   ..$ annotationUrl: chr(0)
##   ..$ experimentUrl: chr(0)
##   ..$ allExperimentsUrl: chr(0)
##   ..$ topAnatUrl    : chr "http://r13_2.bgee.org/"
##   ..$ sendStats     : logi TRUE
##   ..$ quantitativeData: logi FALSE
##   ..$ apiKey        : chr "696a8ee5f27f212d71349c131d413fb0de7947e688c95fcdd001689789c7333dbdf6fa...
##   ...and 18 methods, of which 4 are possibly relevant:
##   ...  format_data, get_annotation, get_data, initialize

```

```

library(biomaRt)

# zebrafish data in Ensembl 85 (stable link)
ensembl <- useMart("ENSEMBL_MART_ENSEMBL",
                     dataset="drerio_gene_ensembl",
                     host="jul2016.archive.ensembl.org")

# get the mapping of Ensembl genes to phenotypes
genesToPhenotypes <- getBM(filters=c("phenotype_source"),
                           value=c("ZFIN")),

```

```

    attributes=c("ensembl_gene_id","phenotype_description"),
    mart=ensembl)

# select phenotypes related to pectoral fin
myPhenotypes = grep("pectoral fin", unique(genesToPhenotypes$phenotype_description), value=T)

# select the genes annotated to select phenotypes
myGenes <- unique(genesToPhenotypes$ensembl_gene_id[genesToPhenotypes$phenotype_description
    %in% myPhenotypes])

# prepare the gene list vector
geneList <- factor(as.integer(unique(genesToPhenotypes$ensembl_gene_id) %in% myGenes))
names(geneList) <- unique(genesToPhenotypes$ensembl_gene_id)
summary(geneList)

##      0      1
## 2986 150

# prepare the topAnat object based on topGO
myTopAnatObject <- topAnat(myTopAnatData, geneList)

##  

## Checking topAnatData object.....  

##  

## Checking gene list.....  

##  

## WARNING: Some genes in your gene list have no expression data in Bgee, and will not be included in the analysis.  

##  

## Building 'most specific' Terms..... ( 942 Terms found. )  

##  

## Build DAG topology..... ( 1707 terms and 3442 relations. )  

##  

## Annotating nodes (Can be long)..... ( 2578 genes annotated to the nodes. )

myTopAnatObject

##  

## ----- topGOdata object -----  

##  

## Description:  

## -  

##  

## Ontology:  

## -  

##  

## 3136 available genes (all genes from the array):  

## - symbol: ENSDARG00000063924 ENSDARG00000028663 ENSDARG00000088634 ENSDARG00000063197 ENSDARG00000058996  

## - 150 significant genes.  

##  

## 2578 feasible genes (genes that can be used in the analysis):  

## - symbol: ENSDARG00000063924 ENSDARG00000028663 ENSDARG00000063197 ENSDARG00000058996 ENSDARG00000058996  

## - 128 significant genes.

```

```

##  

## GO graph (nodes with at least 10 genes):  

##   - a graph with directed edges  

##   - number of nodes = 891  

##   - number of edges = 1798  

##  

## ----- topGOdata object -----  

results <- runTest(myTopAnatObject, algorithm='weight', statistic='fisher')  

##  

##      -- Weight Algorithm --  

##  

##      The algorithm is scoring 819 nontrivial nodes  

##      parameters:  

##          test statistic: fisher : ratio  

##  

##      Level 28: 1 nodes to be scored.  

##  

##      Level 27: 1 nodes to be scored.  

##  

##      Level 26: 2 nodes to be scored.  

##  

##      Level 25: 3 nodes to be scored.  

##  

##      Level 24: 3 nodes to be scored.  

##  

##      Level 23: 2 nodes to be scored.  

##  

##      Level 22: 10 nodes to be scored.  

##  

##      Level 21: 20 nodes to be scored.  

##  

##      Level 20: 19 nodes to be scored.  

##  

##      Level 19: 35 nodes to be scored.  

##  

##      Level 18: 34 nodes to be scored.

```

```
##  
##    Level 17: 53 nodes to be scored.  
  
##  
##    Level 16: 76 nodes to be scored.  
  
##  
##    Level 15: 87 nodes to be scored.  
  
##  
##    Level 14: 71 nodes to be scored.  
  
##  
##    Level 13: 80 nodes to be scored.  
  
##  
##    Level 12: 92 nodes to be scored.  
  
##  
##    Level 11: 74 nodes to be scored.  
  
##  
##    Level 10: 56 nodes to be scored.  
  
##  
##    Level 9:   34 nodes to be scored.  
  
##  
##    Level 8:   19 nodes to be scored.  
  
##  
##    Level 7:   11 nodes to be scored.  
  
##  
##    Level 6:   9 nodes to be scored.  
  
##  
##    Level 5:   15 nodes to be scored.  
  
##  
##    Level 4:   3 nodes to be scored.  
  
##  
##    Level 3:   6 nodes to be scored.  
  
##  
##    Level 2:   2 nodes to be scored.  
  
##  
##    Level 1:   1 nodes to be scored.
```

```
results
```

```
##  
## Description:  
## Ontology:  
## 'weight' algorithm with the 'fisher : ratio' test  
## 891 GO terms scored: 53 terms with p < 0.01  
## Annotation data:  
##     Annotated genes: 2578  
##     Significant genes: 128  
##     Min. no. of genes annotated to a GO: 10  
##     Nontrivial nodes: 819
```

```
# retrieve anatomical structures enriched at a 1% FDR threshold, sorted by decreasing fold enrichment  
tableOver <- makeTable(myTopAnatData, myTopAnatObject, results, cutoff=0.01)
```

```
##  
## Building the results table for the 22 significant terms at FDR threshold of 0.01...  
## Ordering results by pValue column in increasing order...  
## Done
```

```
head(tableOver)
```

	organId		organName		
##	UBERON:0004357	UBERON:0004357	paired limb/fin bud		
##	UBERON:0000151	UBERON:0000151	pectoral fin		
##	UBERON:2000040	UBERON:2000040	median fin fold		
##	UBERON:0003051	UBERON:0003051	ear vesicle		
##	UBERON:0005729	UBERON:0005729	pectoral appendage field		
##	UBERON:0007390	UBERON:0007390	pectoral appendage cartilage tissue		
##		annotated significant	expected	foldEnrichment	pValue
##	UBERON:0004357	144	41	7.15	5.734266 1.622480e-22
##	UBERON:0000151	420	70	20.85	3.357314 1.037552e-18
##	UBERON:2000040	51	18	2.53	7.114625 7.171001e-12
##	UBERON:0003051	304	41	15.09	2.717031 3.135769e-10
##	UBERON:0005729	16	10	0.79	12.658228 4.004917e-10
##	UBERON:0007390	17	9	0.84	10.714286 2.411891e-08
##		FDR			
##	UBERON:0004357	1.445630e-19			
##	UBERON:0000151	4.622296e-16			
##	UBERON:2000040	2.129787e-09			
##	UBERON:0003051	6.984926e-08			
##	UBERON:0005729	7.136762e-08			
##	UBERON:0007390	3.581659e-06			

```
sessionInfo()
```

```
## R version 3.3.1 (2016-06-21)  
## Platform: x86_64-apple-darwin13.4.0 (64-bit)  
## Running under: OS X 10.12.1 (Sierra)  
##
```

```

## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] tcltk      stats4     parallel   stats      graphics   grDevices utils
## [8] datasets   methods    base
##
## other attached packages:
## [1] biomaRt_2.30.0      edgeR_3.16.1        limma_3.30.0
## [4] Mfuzz_2.34.0        DynDoc_1.52.0       widgetTools_1.52.0
## [7] e1071_1.6-7         BgeeDB_2.0.0       tidyR_0.6.0
## [10] topGO_2.26.0        SparseM_1.72       GO.db_3.4.0
## [13] AnnotationDbi_1.36.0 IRanges_2.8.0       S4Vectors_0.12.0
## [16] Biobase_2.34.0      graph_1.52.0       BiocGenerics_0.20.0
## [19] BiocInstaller_1.24.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.7          formatR_1.4        tkWidgets_1.52.0
## [4] bitops_1.0-6         class_7.3-14       tools_3.3.1
## [7] digest_0.6.10        RSQLite_1.0.0       evaluate_0.10
## [10] tibble_1.2           lattice_0.20-34    DBI_0.5-1
## [13] yaml_2.1.13          stringr_1.1.0       dplyr_0.5.0
## [16] knitr_1.14           locfit_1.5-9.1     grid_3.3.1
## [19] data.table_1.9.6     R6_2.2.0           XML_3.98-1.4
## [22] rmarkdown_1.1         magrittr_1.5        htmltools_0.3.5
## [25] matrixStats_0.51.0   assertthat_0.1     stringi_1.1.2
## [28] RCurl_1.95-4.8       chron_2.3-47

```