

reactome.db: How to use the reactome.db package

Willem Ligtenberg

October 12, 2011

1 Introduction

The *reactome.db* package contains mappings from Gene Ontology (GO) identifiers and Entrez Gene identifiers to Reactome identifiers of pathways. Reactome is an open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. The core unit of the Reactome data model is the reaction. Entities (nucleic acids, proteins, complexes and small molecules) participating in reactions form a network of biological interactions and are grouped into pathways. Examples of biological pathways in Reactome include signaling, innate and acquired immune function, transcriptional regulation, translation, apoptosis and classical intermediary metabolism. The *reactome.db* package also contains a mapping from Reactome identifier to the name of the pathway. It can be used for gene set enrichment analysis or just to look up the biological context of a given gene. The Bioconductor version contains information from the Reactome database that was available at the release of the new Bioconductor version. Open Analytics also has their own repository, where we update the package as soon as Reactome releases a new version. This repository can be found at: <http://repos.openanalytics.eu/>

2 Examples

2.1 Basic information

The *reactome.db* package provides an interface to an SQLite database, which contains a subset of the Reactome database. The command `ls` can be used to get an overview of the methods that the package contains.

```
R> library("reactome.db")
```

The same basic set of objects is provided with the db packages:

```
R> ls("package:reactome.db")

 [1] "reactome"                "reactomeEXTID2PATHID"
 [3] "reactomeGO2REACTOMEID"  "reactomeMAPCOUNTS"
 [5] "reactomePATHID2EXTID"   "reactomePATHID2NAME"
 [7] "reactomePATHNAME2ID"    "reactomeREACTOMEID2GO"
 [9] "reactome_dbInfo"        "reactome_dbconn"
[11] "reactome_dbfile"        "reactome_dbschema"
```

Exercise 1

Start an R session and use the `library` function to load the `reactome.db` software package.

It is possible to call the package name as a function to get some QC information about it.

```
R> qcdata = capture.output(reactome())
R> head(qcdata, 20)

 [1] "Quality control information for reactome:"
 [2] ""
 [3] ""
 [4] "This package has the following mappings:"
 [5] ""
 [6] "reactomeEXTID2PATHID has 28363 mapped keys (of 28363 keys)"
 [7] "reactomeGO2REACTOMEID has 3217 mapped keys (of 3217 keys)"
 [8] "reactomePATHID2EXTID has 8320 mapped keys (of 8320 keys)"
 [9] "reactomePATHID2NAME has 13778 mapped keys (of 13778 keys)"
[10] "reactomePATHNAME2ID has 13876 mapped keys (of 13876 keys)"
[11] "reactomeREACTOMEID2GO has 47575 mapped keys (of 47575 keys)"
[12] ""
[13] ""
[14] "Additional Information about this package:"
[15] ""
[16] "DB schema: REACTOME_DB"
[17] "DB schema version: 36"
```

Alternatively, you can get similar information on how many items are in each of the provided maps by looking at the `MAPCOUNTS`:

```
R> reactomeMAPCOUNTS
```

To demonstrate the *environment* API, we'll start with a random sample of Reactome IDs.

```
R> all_reactomeIds <- ls(reactomePATHID2EXTID)
R> length(all_reactomeIds)
```

```
[1] 8320
```

```
R> set.seed(Oxa1beef)
R> reactomeIds <- sample(all_reactomeIds, 3)
R> reactomeIds
```

```
[1] "1614213" "111464" "74158"
```

The usual ways of accessing annotation data are also available.

```
R> reactomePATHID2EXTID[[reactomeIds[1]]]
```

```
[1] "852955" "854999" "855320" "854157" "856317" "852810" "854322"
[8] "853301" "852013" "852451" "854791" "854385" "854399" "851415"
[15] "855925" "855155" "853098" "852888" "856201" "854369" "853807"
[22] "852497" "854875" "851723" "853191" "854993" "851745" "852766"
[29] "855276" "853840" "856918" "854663" "856169" "850691" "855981"
[36] "851906" "851450" "856136" "852071"
```

```
R> reactomePATHID2EXTID$"1163762"
```

```
NULL
```

```
R> pathwayNames <- unlist(mget(reactomeIds, reactomePATHID2NAME))
R> pathwayNames
```

```
1614213
"Saccharomyces cerevisiae: RNA Polymerase II Pre-transcription Events"
111464
"Homo sapiens: SMAC-mediated dissociation of IAP:caspase complexes "
74158
"Homo sapiens: RNA Polymerase III Transcription"
```

For all mappings the *reactome.db* package provides, there is a manual page provided which describes the data in the mapping and where it came from.

```
R> ?reactomeEXTID2PATHID
```

2.2 Getting pathways for a specific species

To get all pathways for a specific species one can use the following code:

```
R> pathways <- toTable(reactomePATHNAME2ID)
R> pathwaysSelectedSpecies <- pathways[grepl("Homo sapiens: ", iconv(pathways$path_name, "UTF-8", "UTF-8"))]
```

Exercise 2

For the reactome IDs in 'reactomeIds' above, use the annotation mappings to find the GO annotations.

Exercise 3

How many Reactome IDs do not have a GO mapping in the reactome.db package? Find a Reactome ID that has a GO mapping. Now look at the GO mappings for this Reactome ID in table form.

The version number of R and packages loaded for generating the vignette were:

```
R version 2.14.0 alpha (2011-10-10 r57211)
Platform: x86_64-unknown-linux-gnu (64-bit)
```

locale:

```
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8      LC_COLLATE=C
[5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=C               LC_NAME=C
[9] LC_ADDRESS=C            LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods
[7] base
```

other attached packages:

```
[1] reactome.db_1.0.38  RSQLite_0.10.0      DBI_0.2-5
[4] AnnotationDbi_1.15.36 Biobase_2.13.10
```

loaded via a namespace (and not attached):

```
[1] IRanges_1.11.29 tools_2.14.0
```