

Gene identifiers and pathway annotation

Contents

Intro	1
Data	1
Pathway names	1

Intro

This markdown aims to measure the proportion of neural pathways for each gene

```
library(readr)
library(tidyr)
library(dplyr)
library(tidylog)
library(magrittr)
```

Data

```
data(gene_ids, entrez_to_pathway, package = "neurotransmission")
```

Pathway names

```
if (!file.exists("pathway_names.tsv")) {
  download.file("http://rest.kegg.jp/list/pathway/hsa", "pathway_names.tsv")
}
```

```
pathway_names <- read_tsv(
  "pathway_names.tsv",
  col_names = c("pathway_id", "pathway_name"),
  col_types = "cc"
)
```

```
selected_genes_pathways <- entrez_to_pathway %>% filter(entrez_id %in% gene_ids[["entrez_id"]])
```

```
## filter: removed 25,091 rows (79%), 6,610 rows remaining
```

```
unique_pathway_ids <- selected_genes_pathways %>% pull(pathway_id) %>% unique
```

```
pathway_names %<>% filter(pathway_id %in% unique_pathway_ids) %>% write_tsv("pathway_names.tsv")
```

```
## filter: removed one row (<1%), 230 rows remaining
```

```
pathway_names
```

```
## # A tibble: 230 x 2
##   pathway_id pathway_name
##   <chr>      <chr>
## 1 path:hsa00140 Steroid hormone biosynthesis - Homo sapiens (human)
## 2 path:hsa00190 Oxidative phosphorylation - Homo sapiens (human)
## 3 path:hsa00220 Arginine biosynthesis - Homo sapiens (human)
## 4 path:hsa00230 Purine metabolism - Homo sapiens (human)
## 5 path:hsa00250 Alanine, aspartate and glutamate metabolism - Homo sapien~
## 6 path:hsa00260 Glycine, serine and threonine metabolism - Homo sapiens (~
## 7 path:hsa00280 Valine, leucine and isoleucine degradation - Homo sapiens~
## 8 path:hsa00330 Arginine and proline metabolism - Homo sapiens (human)
## 9 path:hsa00340 Histidine metabolism - Homo sapiens (human)
## 10 path:hsa00350 Tyrosine metabolism - Homo sapiens (human)
## # ... with 220 more rows
```

```
pathway_classification <- read_tsv("pathway_classification.tsv", col_types = "cci")
```

```
pathway_neuroexclusivity <- selected_genes_pathways %>%
  inner_join(pathway_classification) %>%
  drop_na %>%
  group_by(entrez_id) %>%
  summarise(pathway_neuroexclusivity = sum(neural)/length(neural))
```

```
## Joining, by = "pathway_id"
```

```
## inner_join: added 2 columns (pathway_name, neural)
```

```
##           > rows only in x (    0)
```

```
##           > rows only in y (    0)
```

```
##           > matched rows      6,610
```

```
##           >                    =====
```

```
##           > rows total        6,610
```

```
## drop_na: removed 160 rows (2%), 6,450 rows remaining
```

```
## group_by: one grouping variable (entrez_id)
```

```
## summarise: now 382 rows and 2 columns, ungrouped
```

```
# exporting for package use
```

```
usethis::use_data(pathway_neuroexclusivity, overwrite = TRUE)
```

```
## <U+2714> Setting active project to 'C:/R/neurotransmission'
```

```
## <U+2714> Saving 'pathway_neuroexclusivity' to 'data/pathway_neuroexclusivity.rda'
```

```
pathway_neuroexclusivity
```

```
## # A tibble: 382 x 2
##   entrez_id pathway_neuroexclusivity
##   <chr>          <dbl>
## 1 10000          0.0667
## 2 100137049     0.273
## 3 10243         0.5
## 4 10312         0.1
## 5 10411         0.286
## 6 10488         0.182
## 7 10497         1
## 8 10681         0.529
## 9 107         0.184
## 10 107987478    0.25
## # ... with 372 more rows
```