2 ExpoDb module

2.1 Application domain

ExpoDB module is designed as a convenient tool to explore the data, as well as facilitating to find the biological relationship between exposure and diseases from the perspective of bioinformatics. This module adopts the most frequently-used and authoritative databases, e.g., T3DB, CTD, ToxCast, StringDB, STITCH, KEGG, and GO.

2.2 Theory

2.3 Work pipeline

Users can easily get the modeling results by following the detailed instructions in each step. It mainly has five functions, including searching the related information (Dictionary), converting the IDs between different databases (Convert), exploring the nexus between exposure, protein, phenotype, and disease (Nexus), annotating the non-target features from high-resolution mass spectrometry (Annotation), and learning the nomenclature of "ExposomeX" platform (Abbreviation).

```
# The following two packages should be installed in advance
# devtools::install_github("ExposomeX/exdb", force = TRUE)
# devtools::install_github("ExposomeX/extidy", force = TRUE)

#library(exdb)
# library(extidy)

# devtools::install_github("ExposomeX/exposomex", force = TRUE)
library(exposomex)
```

```
res = InitDb()

res1 = LoadDb(PID=res$PID,
              UseExample="example#1")

res2 = ExpoAbbr(PID=res$PID,
                OutPath = 'default',
                Keys = "default")
res2
```

```
## # A tibble: 12 x 3
##    Keywords     FullName        Group
##    <chr>        <chr>           <chr>
##  1 CrossSection Cross sectional EpiDesign
##  2 Cros         Cross sectional EpiDesign
##  3 CaseControl  Case control    EpiDesign
##  4 CaseCtr      Case control    EpiDesign
##  5 Cohort       Cohort          EpiDesign
##  6 Panel        Panel           EpiDesign
##  7 Longitudinal Longitudinal    EpiDesign
##  8 Longi        Longitudinal    EpiDesign
##  9 CaseFollow   Case followup   EpiDesign
## 10 TimeSeries   Time series     EpiDesign
## 11 TimeSeri     Time series     EpiDesign
## 12 GO           Gene ontology   <NA>
```

```
res3 = ExpoDict(PID=res$PID,
                OutPath = 'default',
                Class = "GO",
                Keys = "default")
res3
```

```
## # A tibble: 7 x 6
##   go_id      Term                              Ontology Definit~1 Synonym Secon~2
##   <chr>      <chr>                             <chr>    <chr>     <chr>   <lgl>
## 1 GO:0000001 mitochondrion inheritance         BP       The dist~ mitoch~ NA
## 2 GO:0004857 enzyme inhibitor activity         MF       Binds to~ metall~ NA
## 3 GO:0004857 enzyme inhibitor activity         MF       Binds to~ GO:004~ NA
## 4 GO:0004866 endopeptidase inhibitor activity MF       Stops, p~ alpha-~ NA
## 5 GO:0004866 endopeptidase inhibitor activity MF       Stops, p~ endopr~ NA
## 6 GO:0004866 endopeptidase inhibitor activity MF       Stops, p~ protei~ NA
## 7 GO:0030414 peptidase inhibitor activity      MF       Stops, p~ protea~ NA
## # ... with abbreviated variable names 1: Definition, 2: Secondary
```

```
res4 = ExpoConv(PID=res$PID,
                OutPath = 'default',
                From = "chemical",
                To = "cas.rn",
                Keys = "default")
res4
```

```
## # A tibble: 18 x 3
##    chemical cas.rn       EXC
##    <chr>    <chr>        <chr>
##  1 zinc     1162648-93-2 EX:C00212
##  2 zinc     24359-56-6   EX:C00212
##  3 zinc     25016-79-9   EX:C00212
##  4 zinc     23713-49-7   EX:C00212
##  5 zinc     7440-66-6    EX:C00212
##  6 zinc     7646-85-7    EX:C00212
##  7 zinc     9025-42-7    EX:C00212
##  8 water    1310-73-2    EX:C46309
##  9 water    39388-36-8   EX:C46309
## 10 water    13670-17-2   EX:C46309
## 11 water    14314-42-2   EX:C46309
## 12 water    558440-22-5  EX:C46309
## 13 water    7732-18-5    EX:C46309
## 14 water    67747-09-5   EX:C46309
## 15 water    7789-20-0    EX:C46309
## 16 water    14280-30-9   EX:C46309
## 17 water    3352-57-6    EX:C46309
## 18 water    60426-60-0   EX:C46309
```

```
res5 = ExpoNexus(PID=res$PID,
                 OutPath = 'default',
                 ClassA = "chemical",
                 ClassB = "protein",
                 KeysA = "default",
                 KeysB = "default")
res5
```

```
## # A tibble: 556 x 8
```

```
##    EXC       inchikey                cas.rn EXP   ENSP  Unipr~1 datab~2 remarks
##    <chr>     <chr>                   <chr>  <chr> <chr> <chr>   <chr>   <chr>
##  1 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ ENSP~ <NA>    ctd     affect~
##  2 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ ENSP~ Q15057  ctd     affect~
##  3 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ ENSP~ F8WAU0  ctd     affect~
##  4 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ ENSP~ C9J8L1  ctd     affect~
##  5 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ <NA>  <NA>    ctd     affect~
##  6 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ <NA>  Q15057  ctd     affect~
##  7 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ <NA>  F8WAU0  ctd     affect~
##  8 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 11118~ EX:P~ <NA>  C9J8L1  ctd     affect~
##  9 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 15502~ EX:P~ ENSP~ <NA>    ctd     affect~
## 10 EX:C07343 OWTFKEBRIAXSMO-UHFFFAOY~ 15502~ EX:P~ ENSP~ Q15057  ctd     affect~
## # ... with 546 more rows, and abbreviated variable names 1: Uniprot_KB,
## #   2: database
```

```
res6 = ExpoAnno(PID=res$PID,
                OutPath = 'default',
                MassToCharge = "default",
                AdductPos = "all",
                AdductNeg = "all",
                Accuracy = 5)
res6
```

```
## # A tibble: 41 x 9
##    IonMass Bias_PPM IonMode  Adduct Name            MassC~1 Monoi~2 SMILES Group
##      <dbl>    <dbl> <chr>    <chr>  <chr>             <dbl>   <dbl> <chr>  <chr>
##  1     200    0.364 positive M+2Na  1,4-Bis(4-chlo~    354.    354. ClC1=~ pare~
##  2     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  3     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  4     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  5     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  6     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  7     200    0.127 positive M+2Na  (S)-1-(4-Bromo~    354.    354. CN(C)~ prec~
##  8     200    3.99  positive M+2Na  2-Chloro-.beta~    354.    354. c1ccc~ prec~
##  9     200    3.99  positive M+2Na  2-Chloro-.beta~    354.    354. c1ccc~ prec~
## 10     200    3.99  positive M+2Na  2-Chloro-.beta~    354.    354. c1ccc~ prec~
## # ... with 31 more rows, and abbreviated variable names 1: MassConv,
## #   2: Monoisotopic_Mass
```

```
FuncExit(PID = res$PID)
```

```
## [1] "Success to exit. Thanks for using ExposomeX platform!"
```