10 ExpoNTA module

10.1 Application domain

ExpoNontarget module is designed to conduct the analysis of the features from the high-resolution mass spectrometry. It mainly aims to screen and annotate the significant features associated with the health outcomes.

10.2 Theory

10.3 Work pipeline

Users can easily get the modeling results and their visualization plots with high quality by following the detailed instructions in each step. Three methods are adopted to screen the important features, including "Stepwise" "LASSO", and "Random forest". At present, the annotation is only applied for the MS1 features acquired from high-resolution liquid chromatography–mass spectrometry.

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

# library(exnta)
# library(extidy)
library(tidyverse)

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

```
res = InitNTA()

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

res1$Expo$Data
```

```
## # A tibble: 150 x 206
##    SampleID SubjectID    Y1    Y2    C1    C2     X1     X2     X3     X4     X5
##    <chr>        <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 001_26_A         1  -101     0    35     7 8.12e3 3.62e4 6.12e4 1.56e5 4.38e4
##  2 001_33_B         1   -51     0    51   779 9.57e3 2.61e4 5.90e4 2.21e5 6.41e4
##  3 001_35_C         1   -37     1    30   613 7.95e3 3.06e4 1.09e5 1.87e5 5.31e4
##  4 003_25_A         3   -61     0    50   454 1.12e5 3.57e5 6.15e4 9.91e4 8.96e4
##  5 003_30_B         3   -28     1    39   766 1.16e5 4.60e5 9.56e4 3.85e5 9.88e4
##  6 003_32_C         3    -8     1    76   768 1.46e4 6.24e4 3.15e4 2.10e5 1.12e5
##  7 005_25_A         5   -63     0    27   418 2.69e4 7.84e4 5.33e4 1.80e5 9.93e4
##  8 005_29_B         5   -35     1    50   484 8.48e3 3.24e4 6.11e4 6.50e4 8.28e4
##  9 005_32_C         5   -14     1    16   647 7.92e3 3.92e4 7.10e4 1.75e5 8.92e4
## 10 006_25_A         6   -99     0    49   329 1.51e4 4.80e4 3.67e4 1.99e5 5.22e4
## # ... with 140 more rows, and 195 more variables: X6 <dbl>, X7 <dbl>, X8 <dbl>,
## #   X9 <dbl>, X10 <dbl>, X11 <dbl>, X12 <dbl>, X13 <dbl>, X14 <dbl>, X15 <dbl>,
## #   X16 <dbl>, X17 <dbl>, X18 <dbl>, X19 <dbl>, X20 <dbl>, X21 <dbl>,
## #   X22 <dbl>, X23 <dbl>, X24 <dbl>, X25 <dbl>, X26 <dbl>, X27 <dbl>,
## #   X28 <dbl>, X29 <dbl>, X30 <dbl>, X31 <dbl>, X32 <dbl>, X33 <dbl>,
## #   X34 <dbl>, X35 <dbl>, X36 <dbl>, X37 <dbl>, X38 <dbl>, X39 <dbl>,
## #   X40 <dbl>, X41 <dbl>, X42 <dbl>, X43 <dbl>, X44 <dbl>, X45 <dbl>, ...
```

```
res2 = DelNearZeroVar(OutPath = "default",
                      PID = res$PID)

res3 = DelMiss(PID = res$PID)

res4 = TransType(PID=res$PID,
                 OutPath = "default",
                 Vars="Y2",
                 To="factor")

res5 = TransScale(PID=res$PID,
                  OutPath = "default",
                  Group= F,
                  Vars="all.x",
                  Method="normal")
res5$Expo$Data
```

```
## # A tibble: 150 x 206
##    SampleID Subje~1    Y1 Y2       C1    C2     X1     X2     X3       X4     X5
##    <chr>      <dbl> <dbl> <fct> <dbl> <dbl>  <dbl>  <dbl>  <dbl>    <dbl>  <dbl>
##  1 001_26_A       1  -101 0        35     7 -0.655 -0.577 -0.312 -0.258   -1.28
##  2 001_33_B       1   -51 0        51   779 -0.643 -0.600 -0.320  0.114   -0.560
##  3 001_35_C       1   -37 1        30   613 -0.657 -0.590 -0.135 -0.0815  -0.949
##  4 003_25_A       3   -61 0        50   454  0.240  0.155 -0.311 -0.590    0.346
##  5 003_30_B       3   -28 1        39   766  0.275  0.390 -0.184  1.07     0.669
##  6 003_32_C       3    -8 1        76   768 -0.600 -0.517 -0.422  0.0549   1.13
##  7 005_25_A       5   -63 0        27   418 -0.493 -0.480 -0.341 -0.120    0.689
##  8 005_29_B       5   -35 1        50   484 -0.652 -0.586 -0.312 -0.787    0.103
##  9 005_32_C       5   -14 1        16   647 -0.657 -0.570 -0.275 -0.151    0.329
## 10 006_25_A       6   -99 0        49   329 -0.595 -0.550 -0.403 -0.00932 -0.982
## # ... with 140 more rows, 195 more variables: X6 <dbl>, X7 <dbl>, X8 <dbl>,
## #   X9 <dbl>, X10 <dbl>, X11 <dbl>, X12 <dbl>, X13 <dbl>, X14 <dbl>, X15 <dbl>,
## #   X16 <dbl>, X17 <dbl>, X18 <dbl>, X19 <dbl>, X20 <dbl>, X21 <dbl>,
## #   X22 <dbl>, X23 <dbl>, X24 <dbl>, X25 <dbl>, X26 <dbl>, X27 <dbl>,
## #   X28 <dbl>, X29 <dbl>, X30 <dbl>, X31 <dbl>, X32 <dbl>, X33 <dbl>,
## #   X34 <dbl>, X35 <dbl>, X36 <dbl>, X37 <dbl>, X38 <dbl>, X39 <dbl>,
## #   X40 <dbl>, X41 <dbl>, X42 <dbl>, X43 <dbl>, X44 <dbl>, X45 <dbl>, ...
```

```
res6 = FindCovaNta(PID=res$PID,
                   OutPath = "default",
                   VarsY = "Y1",
                   VarsC_Prior = "default",
                   VarsC_Fixed = "C2",
                   Method = "single.factor",
                   Thr = 0.1)
res6$Covariables
```

```
## [1] "C2"
```

```
res7 = NtaCros(PID=res$PID,
               OutPath = "default",
               VarsY = "Y1",
               VarsX = "all.x",
               VarsN = "single.factor",
               FdrCorrect = "T",
```

```
               SelMethod = "lasso",
               StepwizeThr = 0.1,
               RF_ImpThr = 0.9,
               IncCova = F,
               Family = "gaussian",
               RepMsr = F,
               Corstr = "ar1")
res7$Y1_single.factor
```

```
## # A tibble: 15 x 11
##    SerialNo Vars.dummy Importance FullN~1 Exact~2 beta.~3   ci_l   ci_h  p.value
##    <chr>    <chr>           <dbl> <chr>     <dbl>   <dbl>  <dbl>  <dbl>    <dbl>
##  1 X196     X196            185.  X132.0~    132.  -16.8  -21.2  -12.3  9.73e-12
##  2 X20      X20              82.0 X103.0~    103.  -12.9  -17.6   -8.10 4.29e- 7
##  3 X16      X16              78.3 X102.0~    102.   12.7    7.89  17.5  6.65e- 7
##  4 X163     X163             65.0 X130.0~    130.   11.9    7.06  16.7  3.46e- 6
##  5 X164     X164             56.8 X130.0~    130.   11.4    6.49  16.2  1.00e- 5
##  6 X135     X135             50.8 X128.0~    128.   10.9    6.04  15.8  2.25e- 5
##  7 X90      X90              49.4 X119.0~    119.   10.8    5.92  15.7  2.73e- 5
##  8 X173     X173             42.4 X130.0~    130.   10.3    5.33  15.2  7.33e- 5
##  9 X17      X17              41.5 X102.0~    102.   10.2    5.25  15.1  8.41e- 5
## 10 X104     X104             40.9 X121.0~    121.  -10.1  -15.1   -5.19 9.17e- 5
## 11 X56      X56              40.8 X114.0~    114.   10.1    5.19  15.1  9.24e- 5
## 12 X161     X161             36.2 X130.0~    130.    9.70   4.75  14.7  1.85e- 4
## 13 X98      X98              35.9 X120.0~    120.    9.68   4.72  14.6  1.93e- 4
## 14 X176     X176             34.9 X130.9~    131.    9.58   4.61  14.5  2.25e- 4
## 15 X191     X191             34.5 X131.1~    131.    9.54   4.57  14.5  2.41e- 4
## # ... with 2 more variables: std.error <dbl>, formula <chr>, and abbreviated
## #   variable names 1: FullName, 2: ExactMass, 3: beta.value
```

```
res8 = VizNtaCros(PID=res$PID,
                  OutPath = "default",
                  VarsY = "Y1",
                  VarsN = "single.factor",
                  Layout = "forest",
                  Brightness = "light",
                  Palette = "default1")
res8$Y1_single.factor_forest_light_default1
```

```
## [[1]]
## TableGrob (1 x 1) "arrange": 1 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
##
## attr(,"class")
## [1] "arrangelist" "list"
```

```
res9 = NtaCros(PID=res$PID,
               OutPath = "default",
               VarsY = "Y2",
               VarsX = "all.x",
               VarsN = "multiple.factor",
               FdrCorrect = "T",
               SelMethod = "lasso",
               StepwizeThr = 0.1,
```

```
               RF_ImpThr = 0.9,
               IncCova = F,
               Family = "binomial",
               RepMsr = F,
               Corstr = "ar1")

res10 = VizNtaCros(PID=res$PID,
                   OutPath = "default",
                   VarsY = "Y2",
                   VarsN = "multiple.factor",
                   Layout = "forest",
                   Brightness = "light",
                   Palette = "default1")
res10$Y2_multiple.factor_forest_light_default1
```

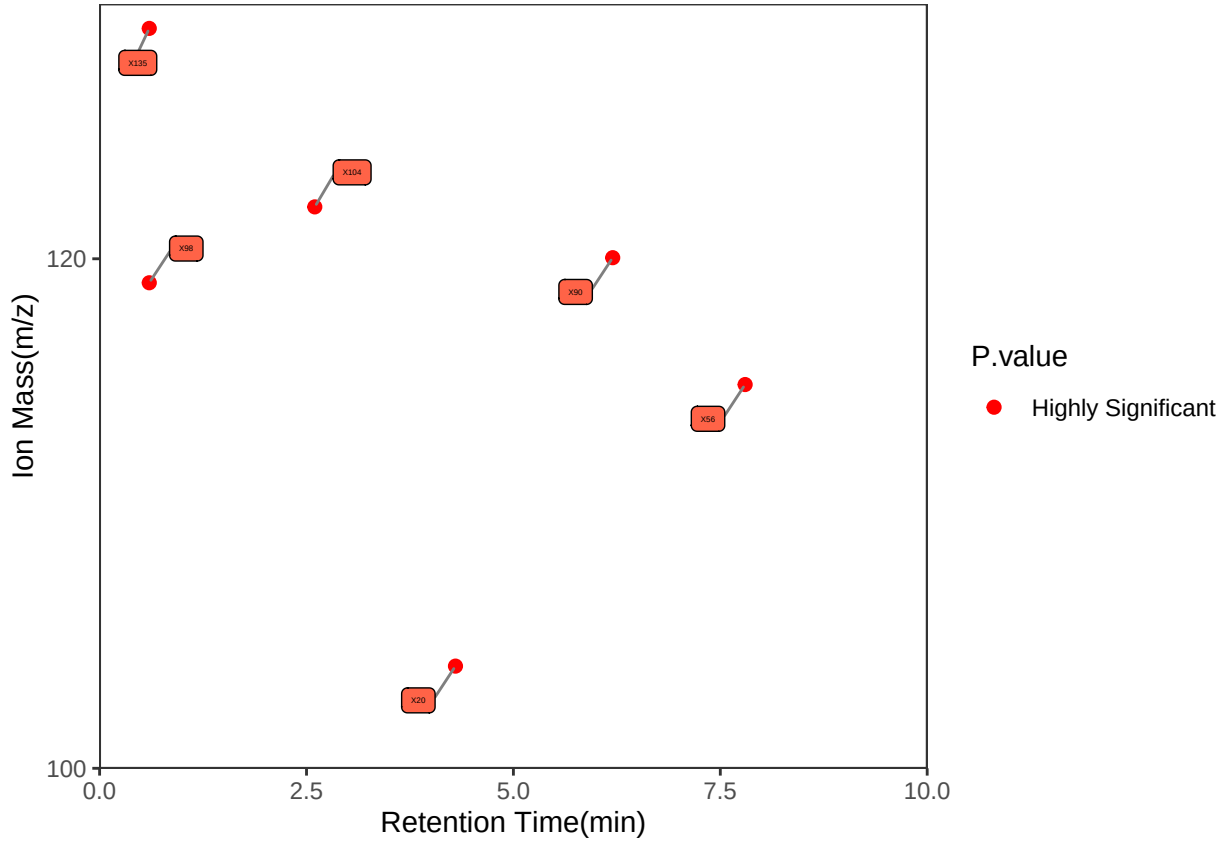
```
## [[1]]
## TableGrob (1 x 1) "arrange": 1 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
##
## attr(,"class")
## [1] "arrangelist" "list"
```

```
res11 = NtaAnno(PID=res$PID,
                OutPath = "default",
                VarsY = "Y1",
                VarsX = "default",
                VarsN = "single.factor",
                FdrCorrect = F,
                AdductPos = "M+H",
                AdductNeg = "M-H",
                Accuracy = 1)
res11$NtaAnno_Y1_single.factor_1
```

```
## # A tibble: 218 x 9
##    SerialNo    RT ExactMass Name             SMILES Monoi~1 IonMode Adduct Group
##    <chr>    <dbl>     <dbl> <chr>            <chr>    <dbl> <chr>   <chr>  <chr>
##  1 X98        0.6      120. 2-Methyl-2-nitr~ CC(C)~    119. positi~ M+H    pare~
##  2 X98        0.6      120. L-Threonine      C[C@@~    119. positi~ M+H    pare~
##  3 X98        0.6      120. 2-Nitro-1-butan~ CCC(C~    119. positi~ M+H    pare~
##  4 X98        0.6      120. 4-Amino-3-hydro~ NCC(O~    119. positi~ M+H    pare~
##  5 X98        0.6      120. N-Butylnitrite   CC(CC~    119. positi~ M+H    prec~
##  6 X98        0.6      120. Dibutyl sulfide  CC(CC~    119. positi~ M+H    prec~
##  7 X98        0.6      120. Isobutyl nitrite CC(C)~    119. positi~ M+H    prec~
##  8 X98        0.6      120. 2-Heptanol       CC(C)~    119. positi~ M+H    prec~
##  9 X98        0.6      120. Dichlorvos       CC(CC~    119. positi~ M+H    prec~
## 10 X98        0.6      120. Ethyl N-methylc~ CN=C(~    119. positi~ M+H    prec~
## # ... with 208 more rows, and abbreviated variable name 1: Monoisotopic_Mass
```

```
res12 = VizNtaAnno(PID=res$PID,
                   OutPath = "default",
                   VarsY = "Y1",
                   VarsN = "single.factor",
                   Accuracy = 1,
                   Brightness = "light",
```

```
                    Palette = "default1")
res12$VizNtaAnno_Y1_single.factor_1_light_default1
```

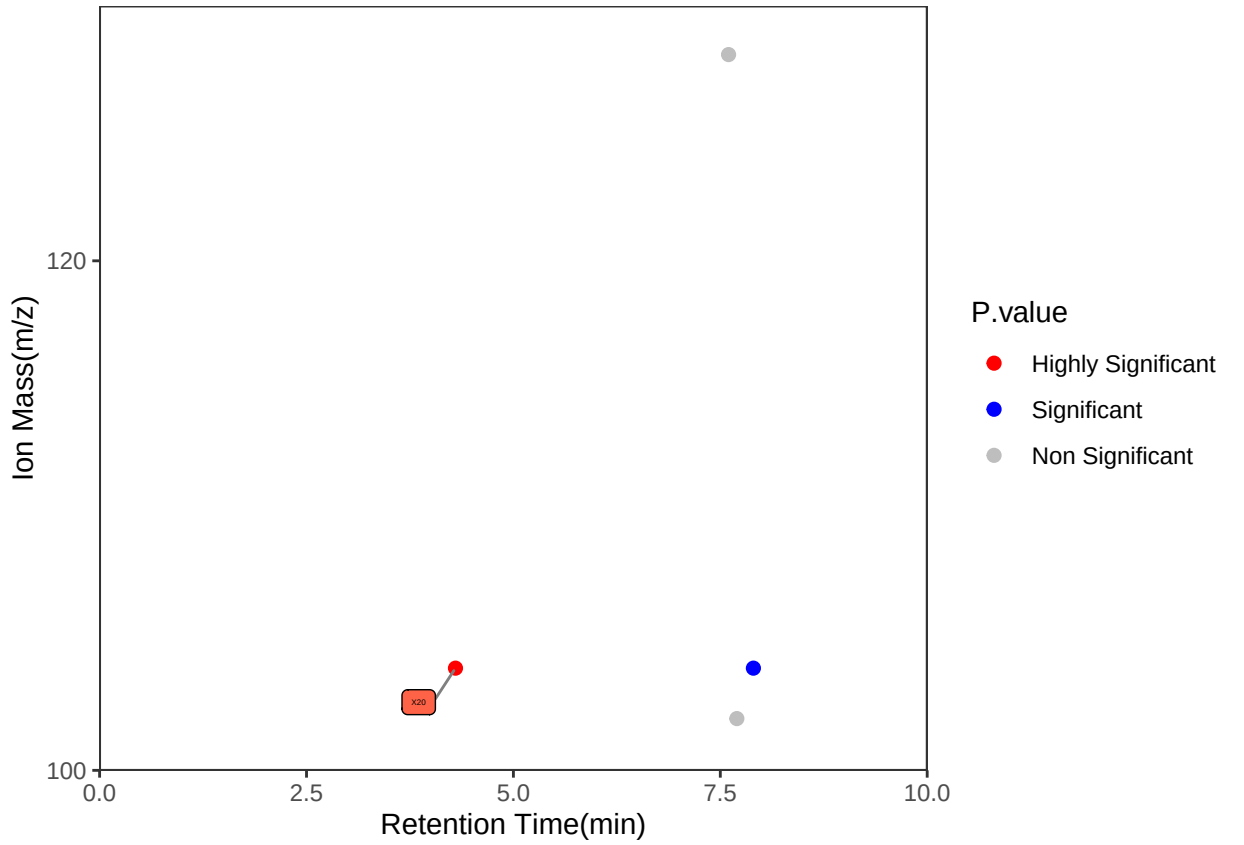



```
res13 = NtaAnno(PID=res$PID,
                OutPath = "default",
                VarsY = "Y2",
                VarsX = "default",
                VarsN = "multiple.factor",
                FdrCorrect = F,
                AdductPos = "M+H,M+2H,M+3H,M+3Na",
                AdductNeg = "M-H,M-2H,M+Cl,M-3H,2M-H",
                Accuracy = 1)
res13$NtaAnno_Y2_multiple.factor_1
```

```
## # A tibble: 43 x 9
##    SerialNo    RT ExactMass Name             SMILES Monoi~1 IonMode Adduct Group
##    <chr>    <dbl>     <dbl> <chr>            <chr>    <dbl> <chr>   <chr>  <chr>
##  1 X20        4.3      103. Propanedioic ac~ OC(=O~    104. negati~ M-H    pare~
##  2 X20        4.3      103. Hydroxypyruvic ~ OCC(=~    104. negati~ M-H    pare~
##  3 X196       4.8      132. alpha-Hydroxybe~ OC(C#~    133. negati~ M-H    pare~
##  4 X196       4.8      132. 4-Anisonitrile   COC1=~    133. negati~ M-H    pare~
##  5 X196       4.8      132. Benzoxazole, 2-~ CC1=N~    133. negati~ M-H    pare~
##  6 X196       4.8      132. 4-Methoxyphenyl~ c1cc(~    133. negati~ M-H    prec~
##  7 X196       4.8      132. (2-Methoxypheny~ c1ccc~    133. negati~ M-H    prec~
##  8 X196       4.8      132. Hydroxy(2-metho~ c1ccc~    133. negati~ M-H    prec~
##  9 X196       4.8      132. (3-Methoxypheny~ c1cc(~    133. negati~ M-H    prec~
```

```
## 10 X196       4.8      132. (3-Methyl-2-fur~ c1cc(~    133. negati~ M-H    prec~
## # ... with 33 more rows, and abbreviated variable name 1: Monoisotopic_Mass
```

```
res14 = VizNtaAnno(PID=res$PID,
                   OutPath = "default",
                   VarsY = "Y2",
                   VarsN = "multiple.factor",
                   Accuracy = 1,
                   Brightness = "light",
                   Palette = "default1")
res14$VizNtaAnno_Y2_multiple.factor_1_light_default1
```




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

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