

EDA of Human Cardiac Tissue-specific Proteome (Feb 23 2021)

CaseOLAP Scores - Full Text

Ashlyn Jew

4/15/2021

Load libraries

```
suppressMessages(library(tidyverse))
suppressMessages(library(ggplot2))
```

Load Data

```
# Human reference proteome - Mar 29 2021 with PMC full text
human_full <- read_csv("https://raw.githubusercontent.com/asjew/heart_caseolap_EDA/main/Data/Human%20reference%20proteome.csv")
```

```
##
## -- Column specification -----
## cols(
##   protein = col_character(),
##   IHD = col_double(),
##   CM = col_double(),
##   ARR = col_double(),
##   VD = col_double(),
##   CHD = col_double(),
##   CCD = col_double(),
##   VOO = col_double(),
##   OTH = col_double()
## )
```

```
head(human_full)
```

```
## # A tibble: 6 x 9
##   protein      IHD      CM      ARR      VD      CHD      CCD      VOO      OTH
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 p01137    0.0408  0.0355  0.0255  0.0243  0.0336  0.00479  0.0141  0.0293
## 2 a0a0c4dga6 0.0127  0.00687  0      0.0122  0      0      0.0131  0.00705
## 3 e5ri75     0      0.0130  0.0134  0      0      0.0144  0      0
## 4 f5h5q2    0.00672  0      0.00893  0.00746  0      0      0      0.00445
## 5 c9jhz9     0      0.00433  0      0.00746  0      0      0      0.00445
## 6 e7ew20    0.00847  0.0196  0.00708  0.0169  0      0.00762  0      0.0141
```

```
summary(human_full)
```

```
##      protein      IHD      CM      ARR
## Length:53145   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
## Class :character 1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
## Mode  :character Median :0.006716   Median :0.004334   Median :0.000000
##              Mean  :0.010833   Mean  :0.010081   Mean  :0.005698
##              3rd Qu.:0.016134   3rd Qu.:0.015539   3rd Qu.:0.007079
##              Max.  :0.207226   Max.  :0.204358   Max.  :0.138092
##      VD      CHD      CCD      VOO
## Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
## 1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
## Median :0.000000   Median :0.004614   Median :0.000000   Median :0.000000
## Mean   :0.005154   Mean   :0.007754   Mean   :0.003062   Mean   :0.00226
## 3rd Qu.:0.007458   3rd Qu.:0.011924   3rd Qu.:0.000000   3rd Qu.:0.000000
## Max.   :0.110973   Max.   :0.110199   Max.   :0.112764   Max.   :0.10759
##      OTH
## Min.   :0.000000
## 1st Qu.:0.000000
## Median :0.004447
## Mean   :0.008364
## 3rd Qu.:0.012485
## Max.   :0.117415
```

Exploratory Data Analysis

```
# Finding proteins that are strictly exclusive to VOO (all groups are 0 except VOO)
# None in this dataset
which(human_full$IHD == 0 & human_full$CM == 0 & human_full$ARR == 0 &
      human_full$VD == 0 & human_full$CHD == 0 & human_full$CCD == 0 & human_full$OTH == 0)

## integer(0)

# Get index of proteins that have caseolap score < 0.01 in all groups except VOO and where VOO > 0.01
VOO_01 <- which(human_full$IHD < 0.01 & human_full$CM < 0.01 & human_full$ARR < 0.01 &
                human_full$VD < 0.01 & human_full$CHD < 0.01 & human_full$CCD < 0.01 &
                human_full$OTH < 0.01 & human_full$VOO > 0.01)

human_full[VOO_01, ] # Displaying the proteins and caseolap scores with indices VOO_01

## # A tibble: 8 x 9
##   protein      IHD      CM      ARR      VD      CHD      CCD      VOO      OTH
##   <chr>      <dbl>    <dbl> <dbl>    <dbl>    <dbl> <dbl>    <dbl>    <dbl>
## 1 e9ppe6    0          0          0 0.00941 0          0 0.0101 0.00889
## 2 h0yet8    0          0          0 0.00941 0          0 0.0101 0.00889
## 3 h0ydi4    0          0          0 0.00941 0          0 0.0101 0.00889
## 4 h0yys3    0          0          0 0.00941 0          0 0.0101 0.00889
## 5 e9pmc9    0          0          0 0.00941 0          0 0.0101 0.00889
## 6 r4gmrv7  0.00672    0          0 0.00941 0          0 0.0101 0.00705
## 7 h0yep9    0          0          0 0.00941 0          0 0.0101 0.00889
## 8 q8iwr8    0          0.00687    0 0.00941 0.00461    0 0.0101 0.00889
```

```
# Proteins that have V00 caseolap score > 0.05
human_full[which(human_full$V00 > 0.05), ]
```

```
## # A tibble: 102 x 9
##   protein      IHD      CM      ARR      VD      CHD      CCD      V00      OTH
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 h7c2e4    0.207 0.0498 0.0497 0.0464 0.0327 0.0376 0.0512 0.0506
## 2 b7z509    0.0467 0.0562 0.0405 0.0537 0.0404 0.0362 0.0552 0.0679
## 3 f8wci6    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 4 j3kt25    0.0918 0.0751 0.0731 0.0557 0.0494 0.0464 0.0566 0.0617
## 5 b8zzc5    0.0467 0.0562 0.0405 0.0537 0.0404 0.0362 0.0552 0.0679
## 6 q92839    0.102 0.0943 0.0944 0.0684 0.0663 0.0733 0.0600 0.0846
## 7 q9nzs2    0.105 0.104 0.0729 0.0618 0.0481 0.0484 0.0517 0.0829
## 8 c9j3b7    0.101 0.0954 0.0978 0.0925 0.0890 0.0935 0.0880 0.0959
## 9 e7eqi0    0.107 0.0986 0.0976 0.0676 0.0816 0.0740 0.0578 0.0991
## 10 p09486   0.112 0.111 0.110 0.103 0.105 0.105 0.0988 0.108
## # ... with 92 more rows
```

```
# Proteins where V00 has highest caseolap score and also has a caseolap score > 0.05
human_full[which(human_full$IHD < human_full$V00 & human_full$CM < human_full$V00 &
  human_full$ARR < human_full$V00 & human_full$VD < human_full$V00 &
  human_full$CHD < human_full$V00 & human_full$CCD < human_full$V00 &
  human_full$OTH < human_full$V00 & human_full$V00 > 0.05), ]
```

```
## # A tibble: 12 x 9
##   protein      IHD      CM      ARR      VD      CHD      CCD      V00      OTH
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 f8wci6    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 2 a0a590ujl0 0.0439 0.0381 0.0389 0.0615 0.0420 0.0306 0.0766 0.0378
## 3 p02787    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 4 p08842    0.0439 0.0381 0.0389 0.0615 0.0420 0.0306 0.0766 0.0378
## 5 a0a590ujy9 0.0439 0.0381 0.0389 0.0615 0.0420 0.0306 0.0766 0.0378
## 6 o14656    0.0613 0.0562 0.0577 0.0865 0.0702 0.0545 0.104 0.0520
## 7 f8wek9    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 8 f8wc57    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 9 a0a590ujt4 0.0439 0.0381 0.0389 0.0615 0.0420 0.0306 0.0766 0.0378
## 10 c9jvg0    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 11 c9jb55    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
## 12 h7c5e8    0.0553 0.0432 0.0417 0.0561 0.0472 0.0302 0.0665 0.0423
```

EDA with Z-scores

```
# Convert caseolap scores into z-scores by column/group
human_standard <- data.frame("protein" = human_full$protein, apply(human_full[2:9], 2, scale))
summary(human_standard)
```

```
##   protein      IHD      CM      ARR
## Length:53145   Min.   :-0.8390   Min.   :-0.7848   Min.   :-0.5406
## Class :character 1st Qu.: -0.8390   1st Qu.: -0.7848   1st Qu.: -0.5406
```

```
## Mode :character      Median :-0.3189      Median :-0.4474      Median :-0.5406
##                      Mean   : 0.0000      Mean   : 0.0000      Mean   : 0.0000
##                      3rd Qu.: 0.4105      3rd Qu.: 0.4248      3rd Qu.: 0.1310
##                      Max.    :15.2092      Max.    :15.1228      Max.    :12.5589
##          VD          CHD          CCD          VOO
## Min.    :-0.5461      Min.    :-0.7279      Min.    :-0.3706      Min.    :-0.3236
## 1st Qu. :-0.5461      1st Qu. :-0.7279      1st Qu. :-0.3706      1st Qu. :-0.3236
## Median :-0.5461      Median :-0.2948      Median :-0.3706      Median :-0.3236
## Mean    : 0.0000      Mean    : 0.0000      Mean    : 0.0000      Mean    : 0.0000
## 3rd Qu. : 0.2440      3rd Qu. : 0.3915      3rd Qu. :-0.3706      3rd Qu. :-0.3236
## Max.    :11.2105      Max.    : 9.6169      Max.    :13.2787      Max.    :15.0828
##          OTH
## Min.    :-0.6996
## 1st Qu. :-0.6996
## Median :-0.3276
## Mean    : 0.0000
## 3rd Qu. : 0.3447
## Max.    : 9.1216
```

```
human_standard[which(human_standard$V00 > 10), ]$protein # proteins with V00 z-score > 10
```

```
## [1] "c9j3b7"      "p09486"      "o15178"      "a0a087x1b6" "a0a087wux9"
## [6] "p42768"      "a0a590ujl0" "p08842"      "e9peg3"      "a0a590ujy9"
## [11] "q04671"      "o14656"      "q16831"      "p10275"      "p28072"
## [16] "a0a590ujt4" "f5gzg9"      "p19526"      "o43934"      "q9nzc7"
```

```
human_standard[which(human_standard$IHD < human_standard$V00 &
  human_standard$CM < human_standard$V00 &
  human_standard$ARR < human_standard$V00 &
  human_standard$VD < human_standard$V00 &
  human_standard$CHD < human_standard$V00 &
  human_standard$CCD < human_standard$V00 &
  human_standard$OTH < human_standard$V00 &
  human_standard$V00 > 10), ]$protein
```

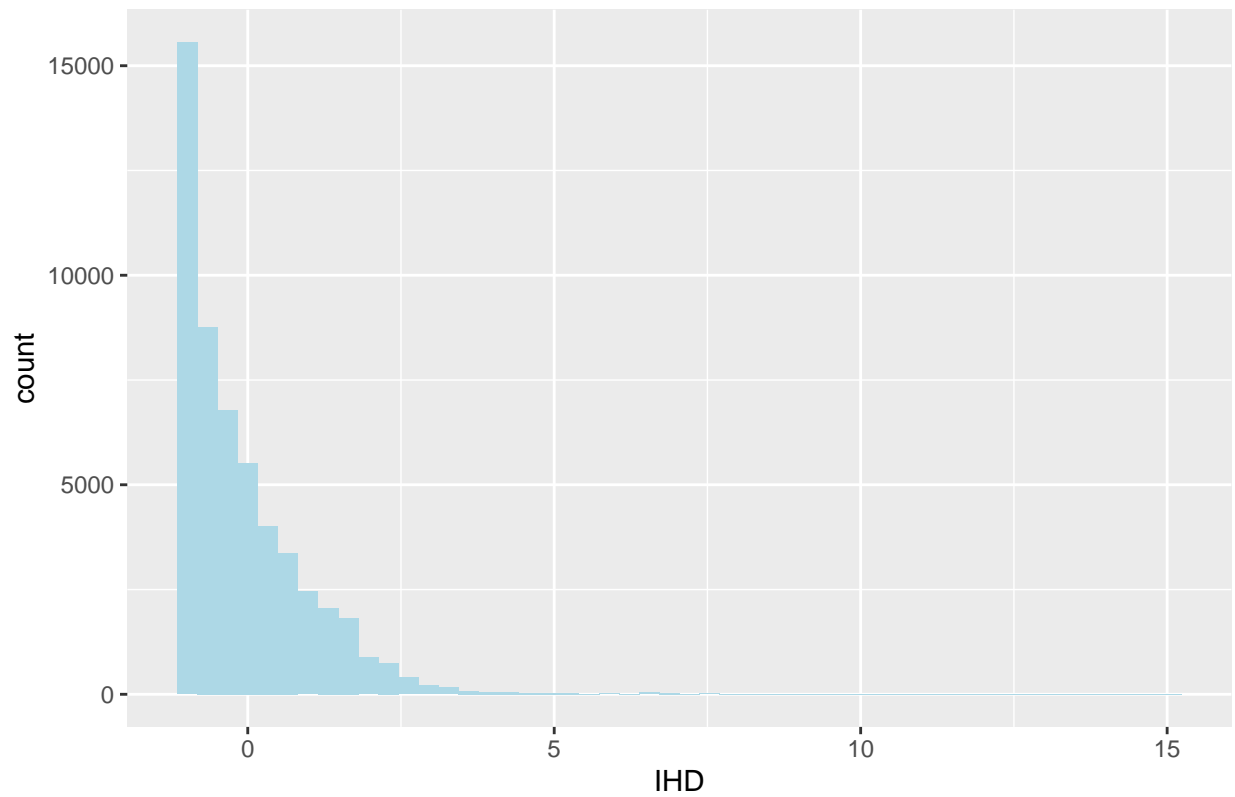
```
## [1] "c9j3b7"      "p09486"      "o15178"      "a0a087x1b6" "a0a087wux9"
## [6] "p42768"      "a0a590ujl0" "p08842"      "e9peg3"      "a0a590ujy9"
## [11] "q04671"      "o14656"      "q16831"      "p10275"      "p28072"
## [16] "a0a590ujt4" "f5gzg9"      "p19526"      "o43934"      "q9nzc7"
```

```
# proteins where V00 has highest z-score and V00 > 10 (same as proteins above)
```

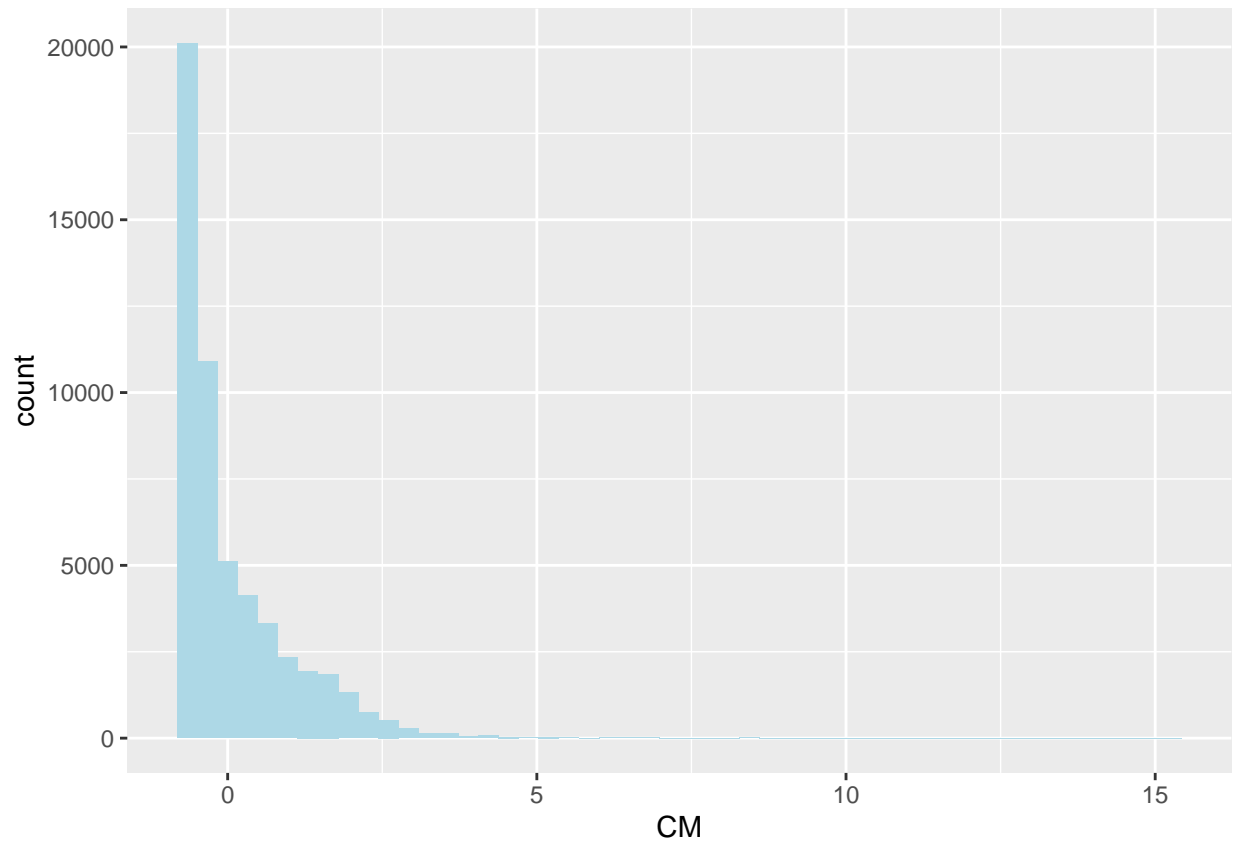
```
# Histogram of Z-score for each group
```

```
ggplot(human_standard, aes(x = IHD)) + geom_histogram(fill = "lightblue", bins = 50) + ggtitle("Distrib
```

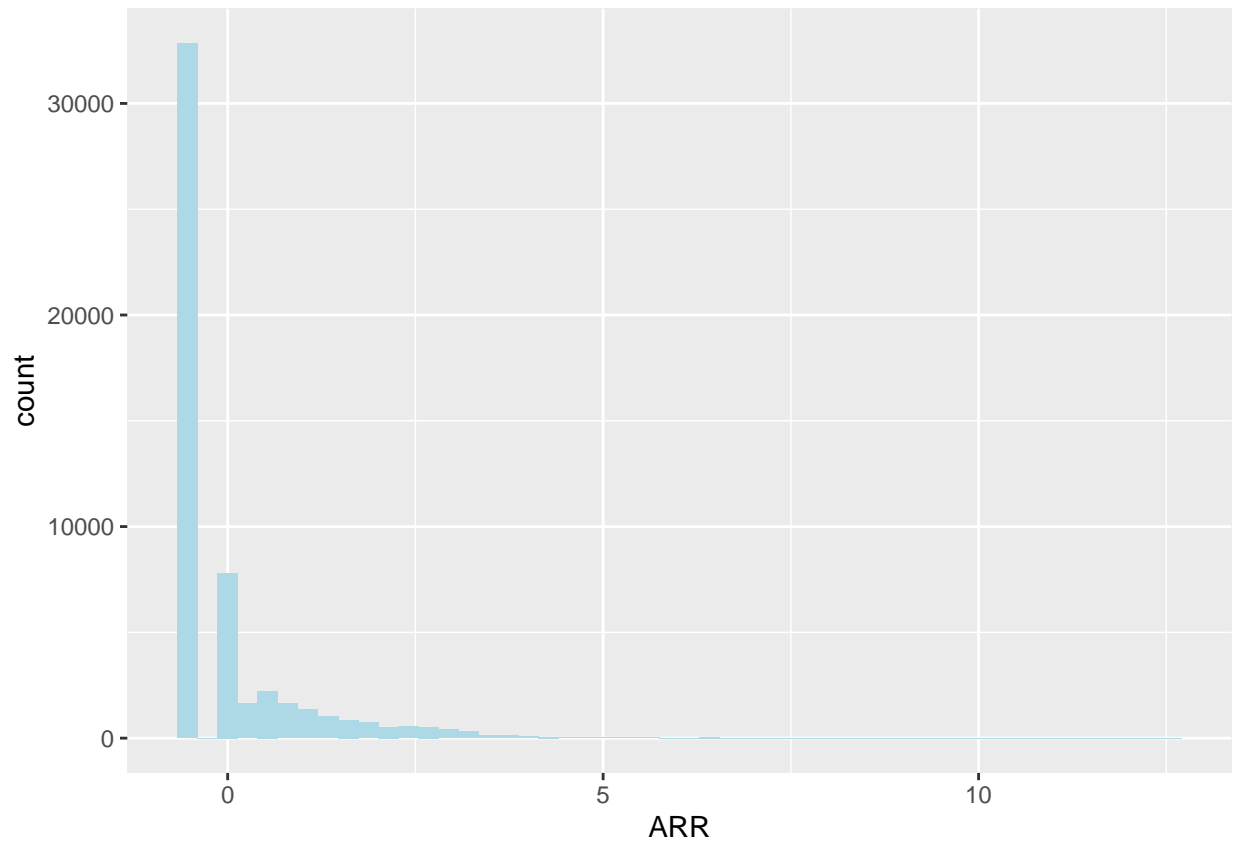
Distribution of CaseOLAP Scores for IHD Disease Group (Human)



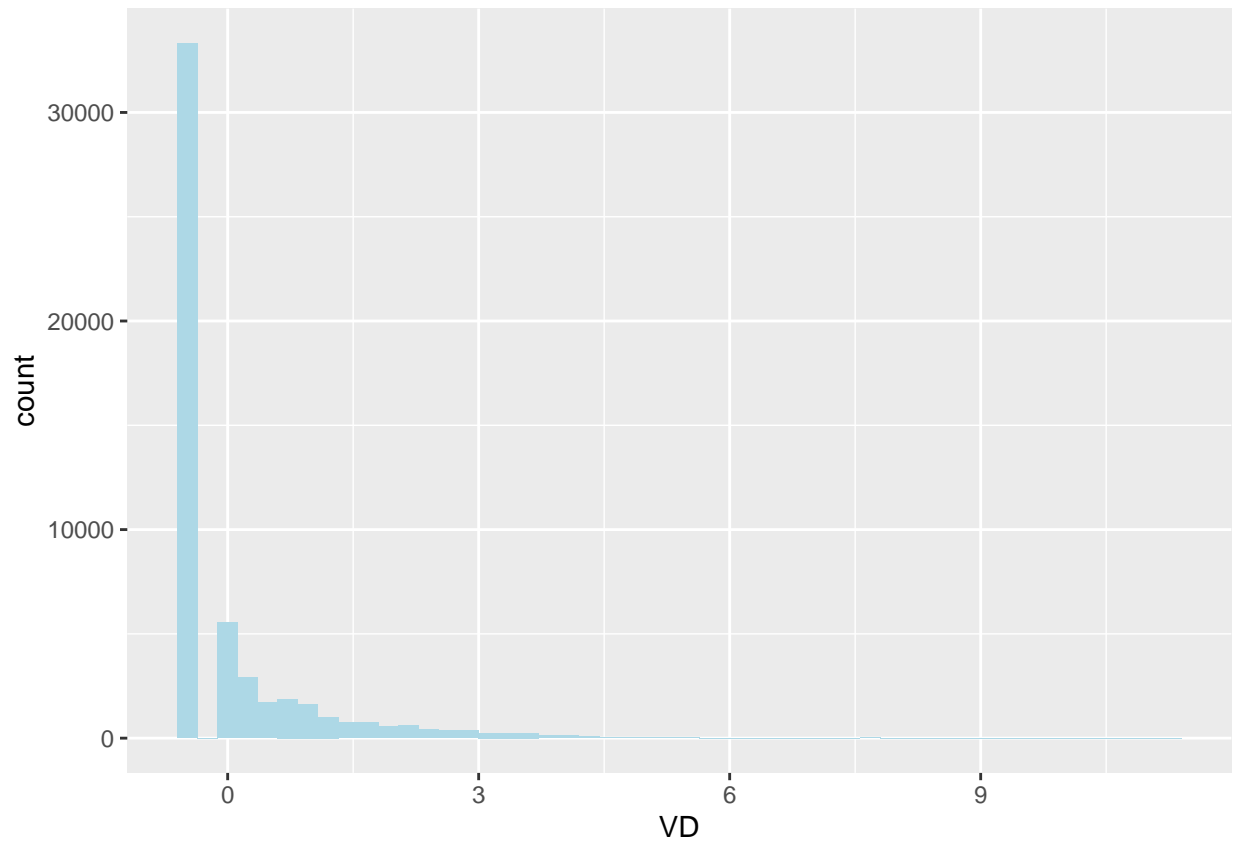
```
ggplot(human_standard, aes(x = CM)) + geom_histogram(fill = "lightblue", bins = 50)
```



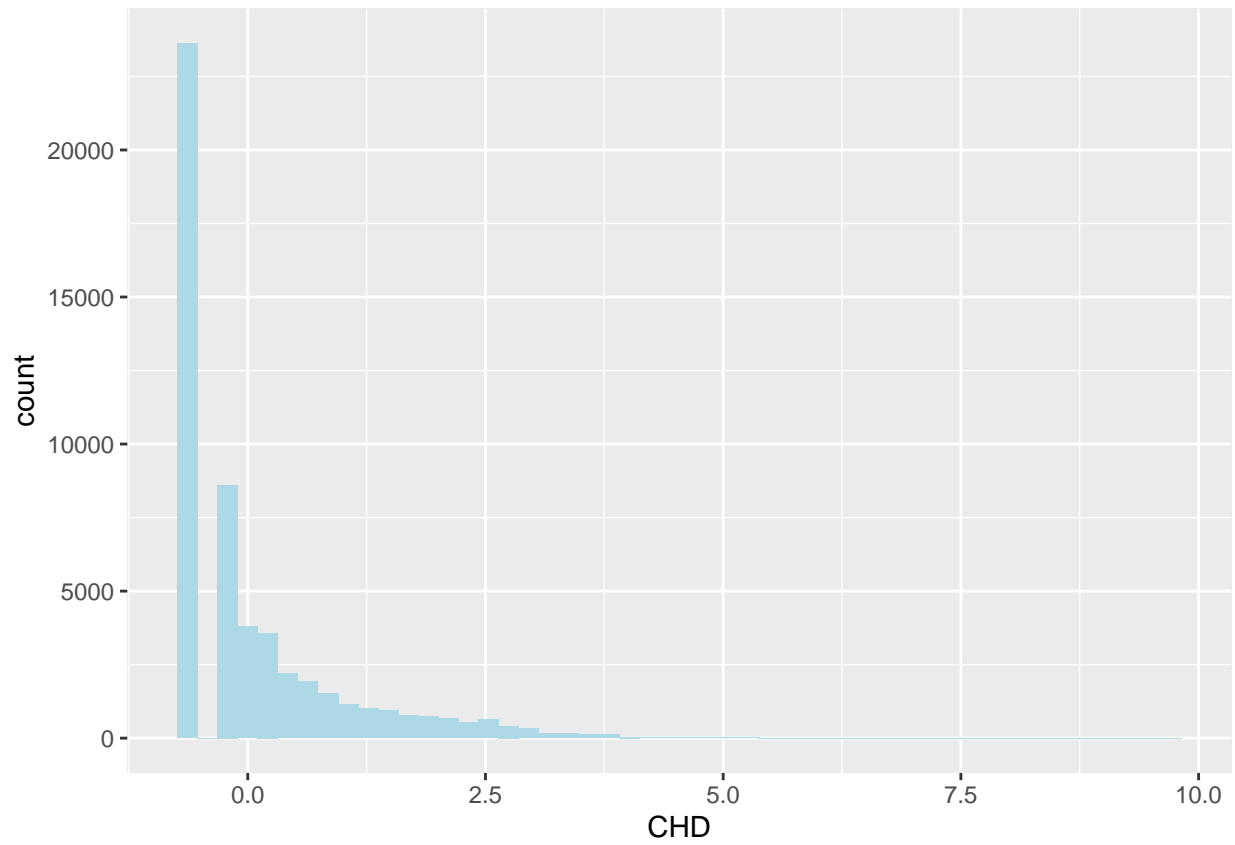
```
ggplot(human_standard, aes(x = ARR)) + geom_histogram(fill = "lightblue", bins = 50)
```



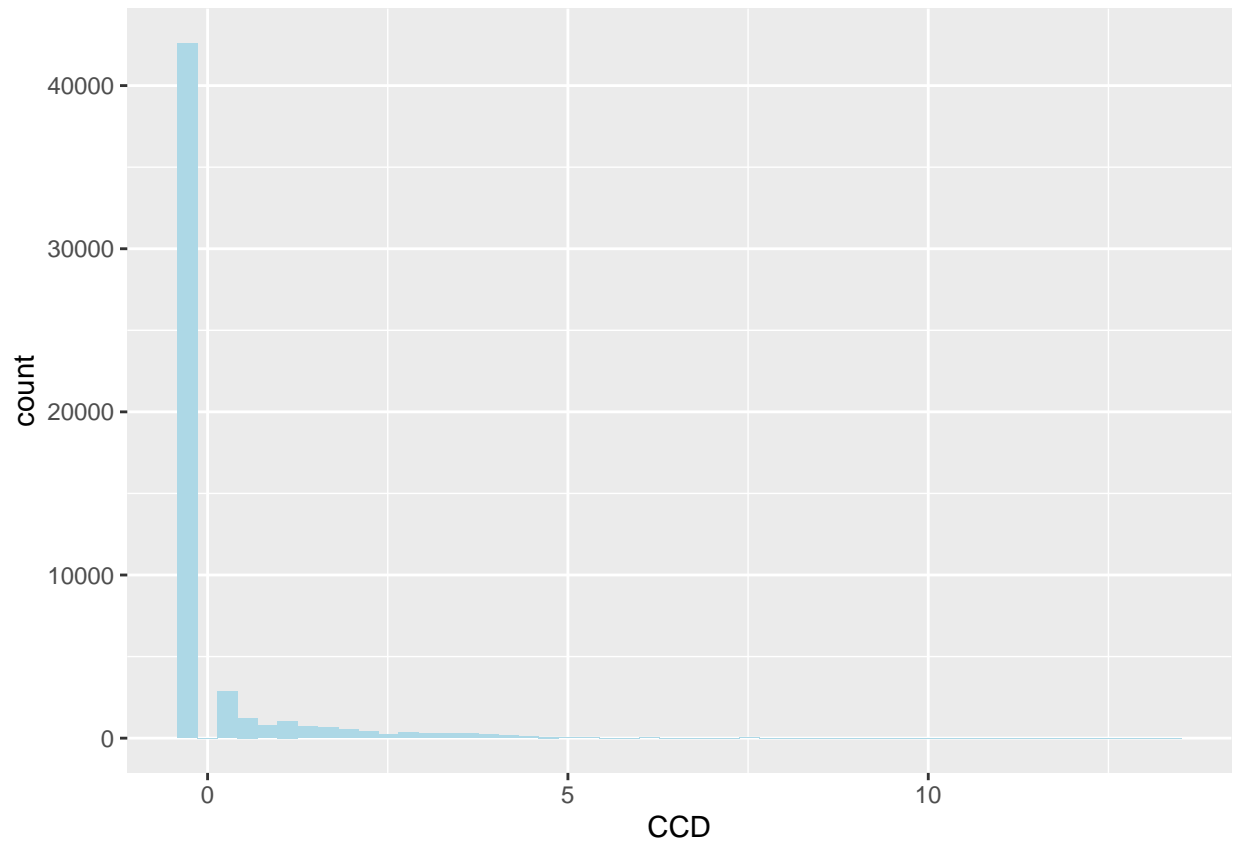
```
ggplot(human_standard, aes(x = VD)) + geom_histogram(fill = "lightblue", bins = 50)
```



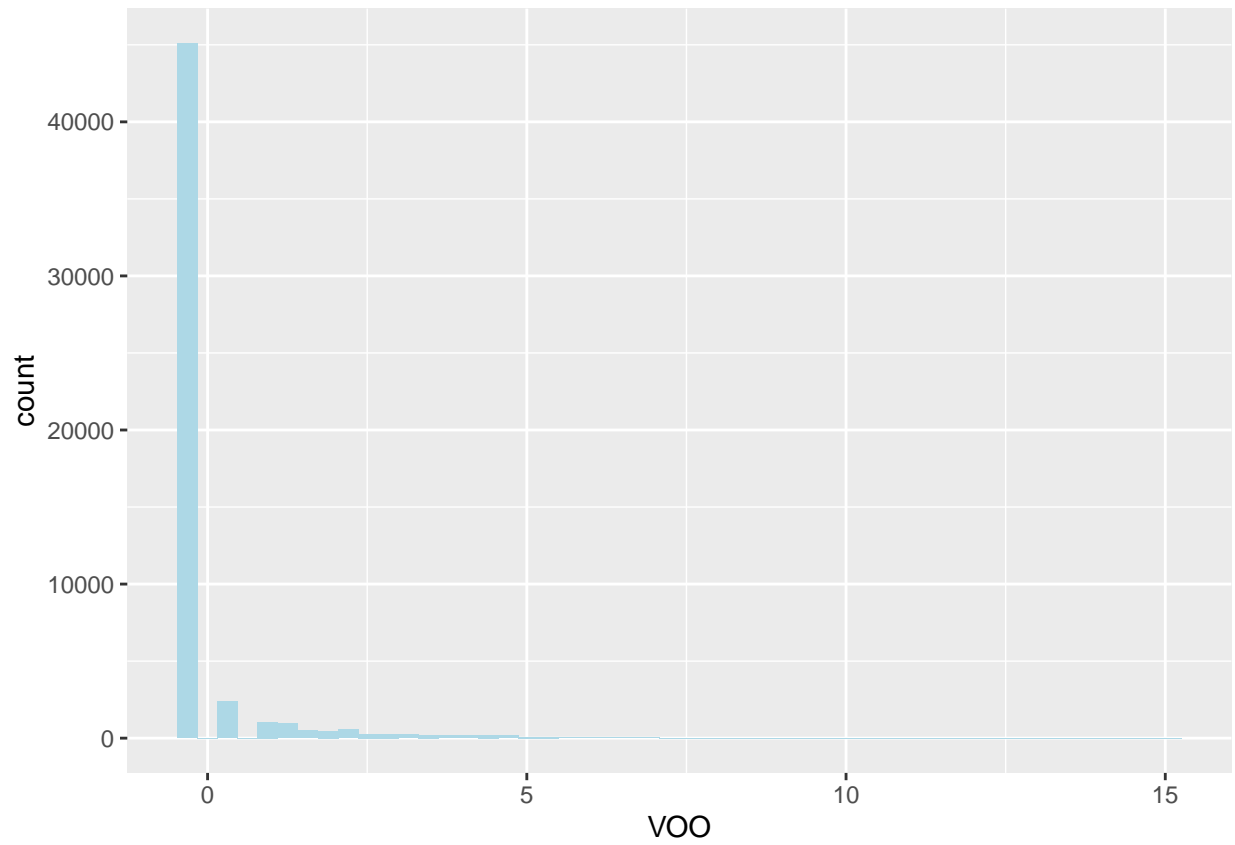
```
ggplot(human_standard, aes(x = CHD)) + geom_histogram(fill = "lightblue", bins = 50)
```

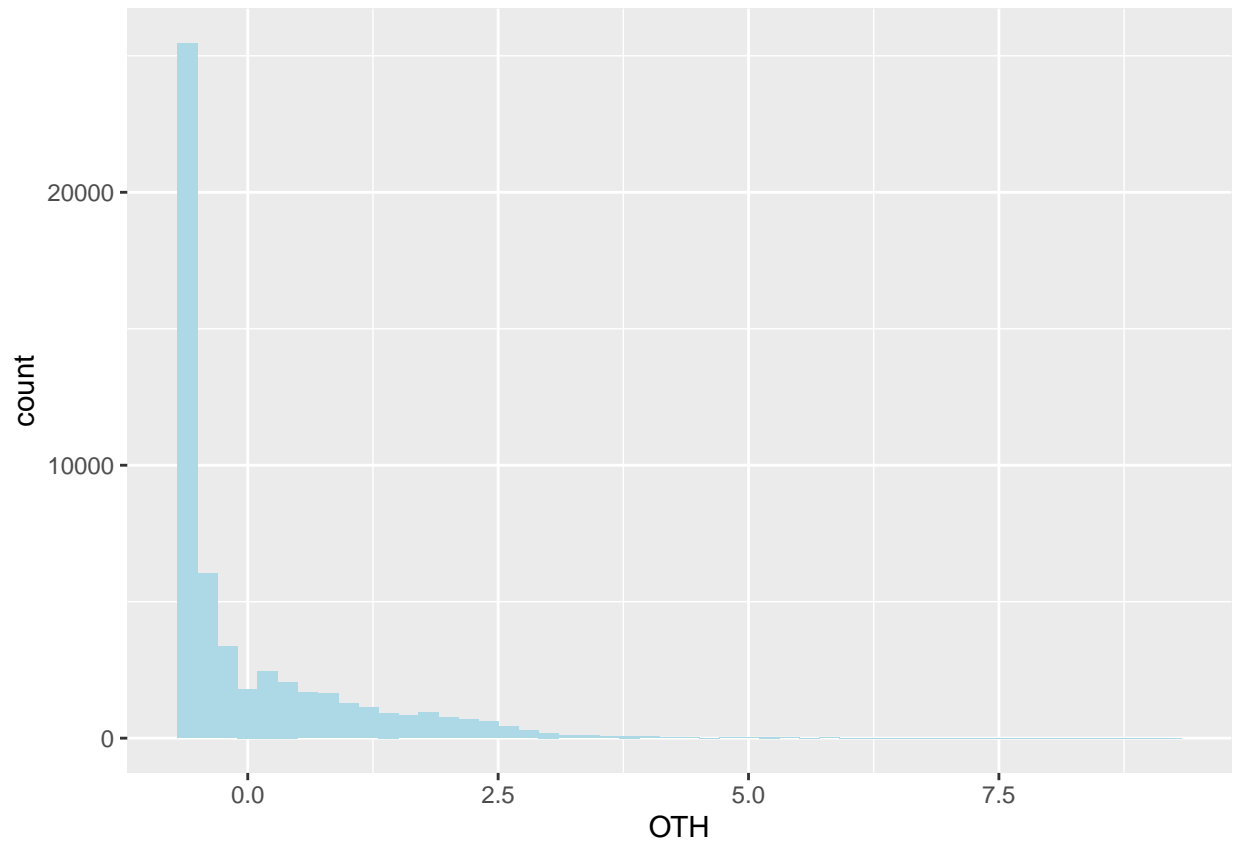
```
ggplot(human_standard, aes(x = CHD)) + geom_histogram(fill = "lightblue", bins = 50)
```



```
ggplot(human_standard, aes(x = V00)) + geom_histogram(fill = "lightblue", bins = 50)
```



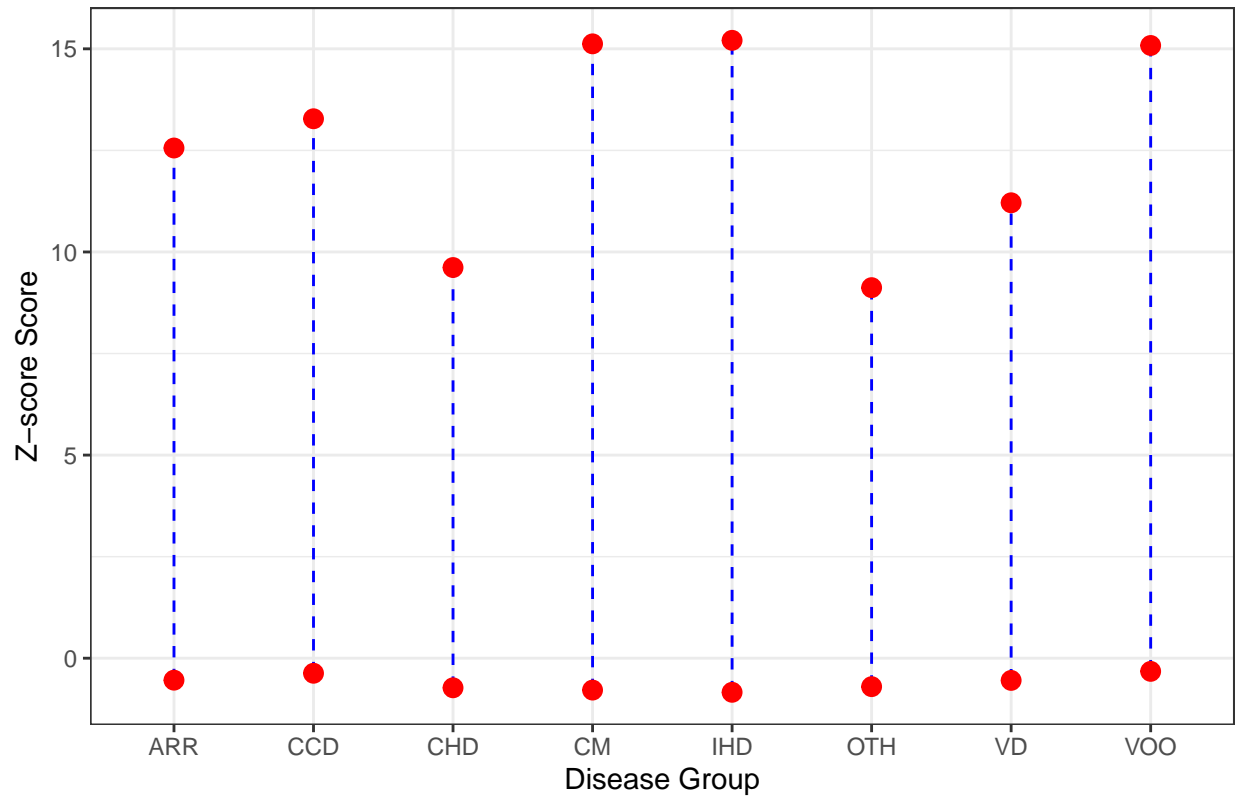
```
ggplot(human_standard, aes(x = OTH)) + geom_histogram(fill = "lightblue", bins = 50)
```



```
# Ranges of Z-score scores by group
human_standard_min <- sapply(human_standard[2:9], min)
human_standard_max <- sapply(human_standard[2:9], max)
human_standard_ranges <- data.frame(id=c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
                                     min=human_standard_min, max=human_standard_max)

ggplot(human_standard_ranges, aes(x=id))+
  geom_linerange(aes(ymin=min,ymax=max),linetype=2,color="blue")+
  geom_point(aes(y=min),size=3,color="red")+
  geom_point(aes(y=max),size=3,color="red")+
  theme_bw() + ggtitle("Ranges of Z-score Scores by Disease Group (Human)") + xlab("Disease Group") + ylab("Z-score Range")
```

Ranges of Z-score Scores by Disease Group (Human)



Finding top unique proteins for each group

Cut-off z-scores are roughly chosen and can be changed to get a smaller or larger list of proteins

IHD

```
human_standard[which(human_standard$V00 < human_standard$IHD &
  human_standard$CM < human_standard$IHD &
  human_standard$ARR < human_standard$IHD &
  human_standard$VD < human_standard$IHD &
  human_standard$CHD < human_standard$IHD &
  human_standard$CCD < human_standard$IHD &
  human_standard$OTH < human_standard$IHD &
  human_standard$IHD > 9), ]
```

| ## | protein | IHD | CM | ARR | VD | CHD | CCD | V00 |
|----------|---------|----------|----------|----------|----------|----------|----------|----------|
| ## 719 | h7c2e4 | 15.20919 | 3.095237 | 4.176024 | 4.371658 | 2.343464 | 4.179438 | 7.012858 |
| ## 5735 | p05154 | 14.68073 | 3.108552 | 3.847722 | 3.376770 | 2.015281 | 3.772437 | 5.516512 |
| ## 11679 | q9une0 | 10.97836 | 5.122472 | 5.041837 | 4.349208 | 2.816573 | 4.566056 | 6.446133 |
| ## 15709 | p27708 | 15.20896 | 3.095432 | 4.176436 | 4.372216 | 2.343395 | 4.179336 | 7.012695 |
| ## 17030 | o76075 | 15.20918 | 3.095205 | 4.175984 | 4.371617 | 2.343918 | 4.179399 | 7.012796 |
| ## 17234 | h7bzb3 | 15.20896 | 3.095432 | 4.176436 | 4.372216 | 2.343395 | 4.179336 | 7.012695 |

```
## 23785 q05682 15.19216 3.138532 4.189114 4.367555 2.341986 4.173143 7.001848
## 34447 h7c3z5 15.20896 3.095432 4.176436 4.372216 2.343395 4.179336 7.012695
## 52061 f8vdp4 15.20892 3.095422 4.176424 4.372203 2.343684 4.179325 7.012676
##      OTH
## 719   3.530471
## 5735  3.511839
## 11679 4.534558
## 15709 3.530377
## 17030 3.530435
## 17234 3.530377
## 23785 3.533901
## 34447 3.530377
## 52061 3.530366
```

proteins where IHD has highest z-score and is above 9

| Entry | Entry name | Gene names | Protein names |
|---------------------------------|--------------|-----------------------------|--|
| <input type="checkbox"/> H7C2E4 | H7C2E4_HUMAN | CAD | Aspartate carbamoyltransferase, EC 2.1.3.2 |
| <input type="checkbox"/> P05154 | IPSP_HUMAN | SERPINA5 PCI, PLANH3, PROCI | Plasma serine protease inhibitor (Acrosomal serine protease inhibitor) (Plasminogen activator inhibitor 3, PAI-3, PAI3) (Protein C inhibitor, PCI) (Serpin A5) |
| <input type="checkbox"/> Q9UNE0 | EDAR_HUMAN | EDAR DL | Tumor necrosis factor receptor superfamily member EDAR (Anhidrotic ectodysplasin receptor 1) (Downless homolog) (EDA-A1 receptor) (Ectodermal dysplasia receptor) (Ectodysplasin-A receptor) |
| <input type="checkbox"/> P27708 | PYR1_HUMAN | CAD | CAD protein [Includes: Glutamine-dependent carbamoyl-phosphate synthase, EC 6.3.5.5; Aspartate carbamoyltransferase, EC 2.1.3.2; Dihydroorotase, EC 3.5.2.3] |
| <input type="checkbox"/> O76075 | DFFB_HUMAN | DFFB CAD, DFF2, DFF40 | DNA fragmentation factor subunit beta, EC 3.-.-.- (Caspase-activated deoxyribonuclease, CAD, Caspase-activated DNase) (Caspase-activated nuclease, CPAN) (DNA fragmentation factor 40 kDa subunit, DFF-40) |
| <input type="checkbox"/> H7BZB3 | H7BZB3_HUMAN | CAD | CAD protein |
| <input type="checkbox"/> Q05682 | CALD1_HUMAN | CALD1 CAD, CDM | Caldesmon, CDM |
| <input type="checkbox"/> H7C3Z5 | H7C3Z5_HUMAN | CAD | CAD protein |
| <input type="checkbox"/> F8VPD4 | F8VPD4_HUMAN | CAD | Aspartate carbamoyltransferase, EC 2.1.3.2, EC 3.5.2.3, EC 6.3.5.5 (CAD protein) (Dihydroorotase) (Glutamine-dependent carbamoyl-phosphate synthase) |

Figure 1: UniProt entries for IHD

CM

```
human_standard[which(human_standard$IHD < human_standard$CM &
  human_standard$V00 < human_standard$CM &
  human_standard$ARR < human_standard$CM &
  human_standard$VD < human_standard$CM &
  human_standard$CHD < human_standard$CM &
  human_standard$CCD < human_standard$CM &
  human_standard$OTH < human_standard$CM &
  human_standard$CM > 8), ]
```

```
##      protein      IHD      CM      ARR      VD      CHD      CCD
## 1577   j3ksq5 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 8892   p08603 4.223978 15.122810 6.748300 3.463074 1.740468 5.080118
## 9535   j3krh5 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 10131  q9p121 6.250077 9.891019 6.985538 7.096340 5.093733 5.792587
```

```
## 12897      j3qsc7 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 22950      j3ks28 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 24430      p12821 6.709497 8.432963 5.296883 4.163121 2.378726 4.475182
## 28375      j3qrx8 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 43882      j3qlr4 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 48185      j3ktb8 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 50461      j3kth9 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
## 52365 a0a0a0msn4 6.689920 8.433404 5.299055 4.165123 2.367500 4.477689
##          V00      OTH
## 1577  5.461282 5.911577
## 8892  4.123195 6.193185
## 9535  5.461282 5.911577
## 10131 8.355606 7.740212
## 12897 5.461282 5.911577
## 22950 5.461282 5.911577
## 24430 5.455220 5.911980
## 28375 5.461282 5.911577
## 43882 5.461282 5.911577
## 48185 5.461282 5.911577
## 50461 5.461282 5.911577
## 52365 5.461282 5.911577
```

proteins where CM has highest z-score and above 8

| <input type="checkbox"/> | Entry | Entry name | Gene names | Protein names | |
|--------------------------|------------|------------------|-------------------------------|--|--|
| <input type="checkbox"/> | J3KSQ5 | J3KSQ5_HUMAN | ACE | Angiotensin-converting enzyme | |
| <input type="checkbox"/> | P08603 | CFAH_HUMAN | CFH HF, HF1, HF2 | Complement factor H (H factor 1) | |
| <input type="checkbox"/> | J3KRH5 | J3KRH5_HUMAN | ACE | Angiotensin-converting enzyme | |
| <input type="checkbox"/> | Q9P121 | NTRI_HUMAN | NTM IGLON2, NT, UNQ297/PRO337 | Neurotrimin, hNT (IgLON family member 2) | |
| <input type="checkbox"/> | J3QSC7 | J3QSC7_HUMAN | ACE | Angiotensin-converting enzyme | |
| <input type="checkbox"/> | J3KS28 | J3KS28_HUMAN | ACE | Angiotensin-converting enzyme, EC 3.4.-.- | |
| <input type="checkbox"/> | P12821 | ACE_HUMAN | ACE DCP, DCP1 | Angiotensin-converting enzyme, ACE, EC 3.2.1.-, EC 3.4.15.1 (Dipeptidyl carboxypeptidase I) (Kininase II) (CD antigen CD143) [Cleaved into: Angiotensin-converting enzyme, soluble form] | |
| <input type="checkbox"/> | J3QRX8 | J3QRX8_HUMAN | | Deleted. | |
| <input type="checkbox"/> | J3QLR4 | J3QLR4_HUMAN | ACE | Angiotensin-converting enzyme | |
| <input type="checkbox"/> | J3KTB8 | J3KTB8_HUMAN | ACE | Angiotensin-converting enzyme, EC 3.4.-.- | |
| <input type="checkbox"/> | J3KTH9 | J3KTH9_HUMAN | ACE | Angiotensin-converting enzyme | |
| <input type="checkbox"/> | A0A0A0MSN4 | A0A0A0MSN4_HUMAN | ACE | Angiotensin-converting enzyme, EC 3.4.-.- | |

Figure 2: UniProt entries for CM

ARR

```
human_standard[which(human_standard$IHD < human_standard$ARR &
                     human_standard$CM < human_standard$ARR &
```

```

human_standard$V00 < human_standard$ARR &
human_standard$VD < human_standard$ARR &
human_standard$CHD < human_standard$ARR &
human_standard$CCD < human_standard$ARR &
human_standard$OTH < human_standard$ARR &
human_standard$ARR > 9), ]

```

```

##      protein      IHD      CM      ARR      VD      CHD      CCD      V00
## 7992   q9ul49 2.657266 2.916266 12.558886 3.678617 2.330739 4.846638 4.011663
## 11723  e5rk80 7.621587 6.884133 9.632561 4.705597 2.892766 8.135945 6.506734
## 12731  q16473 3.520232 2.221996 9.694233 3.821210 2.246642 2.323893 4.002783
## 21661  o43593 7.621587 6.884133 9.632561 4.705597 2.892766 8.135945 6.506734
## 42437  p49366 3.616542 3.270556 10.563750 4.467255 4.766417 4.979301 4.554361
##      OTH
## 7992  2.500736
## 11723 5.396820
## 12731 2.032001
## 21661 5.396820
## 42437 3.149157

```

proteins where ARR has highest z-score and above 9

| <input type="checkbox"/> | Entry | Entry name | Gene names | Protein names |
|--------------------------|--------|--------------|------------------|--|
| <input type="checkbox"/> | Q9UL49 | TCFL5_HUMAN | TCFL5 CHA, E2BP1 | Transcription factor-like 5 protein (Cha transcription factor) (HPV-16 E2-binding protein 1, E2BP-1) |
| <input type="checkbox"/> | E5RK80 | E5RK80_HUMAN | HR | Lysine-specific demethylase hairless |
| <input type="checkbox"/> | Q16473 | TENXA_HUMAN | TNXA XA | Putative tenascin-XA, TN-XA |
| <input type="checkbox"/> | O43593 | HAIR_HUMAN | HR | Lysine-specific demethylase hairless, EC 1.14.11.65 ([histone H3]-dimethyl-L-lysine(9) demethylase hairless) |
| <input type="checkbox"/> | P49366 | DHYS_HUMAN | DHPS DS | Deoxyhypusine synthase, DHS, EC 2.5.1.46 |

Figure 3: UniProt entries for ARR

VD

```

human_standard[which(human_standard$IHD < human_standard$VD &
  human_standard$CM < human_standard$VD &
  human_standard$ARR < human_standard$VD &
  human_standard$V00 < human_standard$VD &
  human_standard$CHD < human_standard$VD &
  human_standard$CCD < human_standard$VD &
  human_standard$OTH < human_standard$VD &
  human_standard$VD > 8), ]

```

```

##      protein      IHD      CM      ARR      VD      CHD      CCD      V00
## 19024  p25116 4.65126 6.172324 5.373991 11.183498 6.112869 4.872527 8.671207
## 51911  q6e213 1.94605 2.162937 2.279152 8.399987 2.475388 2.115031 4.872195
##      OTH
## 19024 6.166133
## 51911 2.119385

```


proteins where VD has highest z-score and above 8

| Entry | Entry name | Gene names | Protein names |
|--------|-------------|------------------------------|---|
| P25116 | PAR1_HUMAN | F2R CF2R, PAR1, TR | Proteinase-activated receptor 1, PAR-1 (Coagulation factor II receptor) (Thrombin receptor) |
| Q6E213 | AWAT2_HUMAN | AWAT2 DC4, DGAT2L4, MFAT, WS | Acyl-CoA wax alcohol acyltransferase 2, EC 2.3.1.75 (11-cis-specific retinyl-ester synthase, 11-cis-RE-synthase) (Acyl-CoA retinol O-fatty-acyltransferase, ARAT, Retinol O-fatty-acyltransferase, EC 2.3.1.76) (Diacylglycerol O-acyltransferase 2-like protein 4) (Diacylglycerol O-acyltransferase candidate 4, hDC4) (Long-chain-alcohol O-fatty-acyltransferase 2) (Multifunctional O-acyltransferase) (Wax synthase, hWS) |

Figure 4: UniProt entries for VD

CHD

```
human_standard[which(human_standard$IHD < human_standard$CHD &
  human_standard$CM < human_standard$CHD &
  human_standard$ARR < human_standard$CHD &
  human_standard$VD < human_standard$CHD &
  human_standard$V00 < human_standard$CHD &
  human_standard$CCD < human_standard$CHD &
  human_standard$OTH < human_standard$CHD &
  human_standard$VD > 5), ]
```

```
##      protein      IHD      CM      ARR      VD      CHD      CCD      V00
## 8993  q8iwt0 2.477818 2.240472 2.962209 5.413161 8.447828 3.523173 7.904877
##      OTH
## 8993 2.795316
```

proteins where CHD has highest z-score and above 5

| Entry | Entry name | Gene names | Protein names |
|--------|------------|--------------|--|
| Q8IWT0 | ARCH_HUMAN | ZBTB80S ARCH | Protein archease (Protein ZBTB80S) (Zinc finger and BTB domain-containing opposite strand protein 8) |

Figure 5: UniProt entries for CHD

CCD

```
human_standard[which(human_standard$IHD < human_standard$CCD &
  human_standard$CM < human_standard$CCD &
  human_standard$ARR < human_standard$CCD &
  human_standard$VD < human_standard$CCD &
  human_standard$CHD < human_standard$CCD &
  human_standard$V00 < human_standard$CCD &
  human_standard$OTH < human_standard$CCD &
  human_standard$CCD > 10), ]
```

| ## | protein | IHD | CM | ARR | VD | CHD | CCD |
|----------|------------|----------|----------|-----------|----------|----------|----------|
| ## 223 | p42357 | 4.177873 | 4.640655 | 8.527665 | 5.119573 | 5.323045 | 12.05445 |
| ## 1760 | q96s42 | 1.695108 | 1.970701 | 7.093345 | 2.570365 | 3.404003 | 11.81856 |
| ## 6798 | h7c0e4 | 1.695108 | 1.970701 | 7.093345 | 2.570365 | 3.404003 | 11.81856 |
| ## 8283 | q8ng44 | 3.938835 | 5.019702 | 10.210906 | 5.263583 | 5.089687 | 12.96113 |
| ## 9231 | e9pg18 | 1.409237 | 2.188305 | 6.630038 | 2.282354 | 3.954432 | 10.72187 |
| ## 11561 | a3ey21 | 1.403318 | 2.185107 | 6.625567 | 2.289853 | 3.952090 | 10.71290 |
| ## 16928 | q12809 | 1.451550 | 1.723750 | 8.080234 | 1.846309 | 5.914939 | 12.77803 |
| ## 19259 | q96hu1 | 3.962873 | 3.884718 | 8.757496 | 3.950484 | 3.665883 | 10.29203 |
| ## 27097 | h9kvd2 | 1.409237 | 2.188305 | 6.630038 | 2.282354 | 3.954432 | 10.72187 |
| ## 35707 | q8ng42 | 3.938835 | 5.019702 | 10.210906 | 5.263583 | 5.089687 | 12.96113 |
| ## 38780 | k4dia1 | 1.409237 | 2.188305 | 6.630038 | 2.282354 | 3.954432 | 10.72187 |
| ## 39034 | e9phb6 | 1.409237 | 2.188305 | 6.630038 | 2.282354 | 3.954432 | 10.72187 |
| ## 42468 | a0a0a0mt39 | 1.409237 | 2.188305 | 6.630038 | 2.282354 | 3.954432 | 10.72187 |
| ## 44654 | q14524 | 1.403796 | 2.181298 | 6.642634 | 2.279847 | 3.964270 | 10.75395 |
| ## 45337 | q8wxs4 | 3.930572 | 5.010012 | 10.217844 | 5.265819 | 5.091058 | 12.97247 |
| ## 47772 | q15477 | 6.862250 | 7.131732 | 9.438035 | 6.730224 | 6.217419 | 10.02837 |
| ## 49418 | q8ng45 | 3.938835 | 5.019702 | 10.210906 | 5.263583 | 5.089687 | 12.96113 |
| ## 49754 | p19526 | 8.531811 | 8.582576 | 10.084570 | 8.708805 | 8.416007 | 11.93331 |
| ## | V00 | OTH | | | | | |
| ## 223 | 5.402255 | 5.152783 | | | | | |
| ## 1760 | 2.411202 | 2.133770 | | | | | |
| ## 6798 | 2.411202 | 2.133770 | | | | | |
| ## 8283 | 5.990651 | 5.086934 | | | | | |
| ## 9231 | 2.204544 | 2.974921 | | | | | |
| ## 11561 | 2.205546 | 2.973325 | | | | | |
| ## 16928 | 2.010776 | 2.256843 | | | | | |
| ## 19259 | 5.077978 | 3.867956 | | | | | |
| ## 27097 | 2.204544 | 2.974921 | | | | | |
| ## 35707 | 5.990651 | 5.086934 | | | | | |
| ## 38780 | 2.204544 | 2.974921 | | | | | |
| ## 39034 | 2.204544 | 2.974921 | | | | | |
| ## 42468 | 2.204544 | 2.974921 | | | | | |
| ## 44654 | 2.202302 | 2.972967 | | | | | |
| ## 45337 | 5.997930 | 5.080532 | | | | | |
| ## 47772 | 7.437046 | 6.646555 | | | | | |
| ## 49418 | 5.990651 | 5.086934 | | | | | |
| ## 49754 | 10.609381 | 8.836096 | | | | | |

proteins where CCD has highest z-score and above 10

V00

```
human_standard[which(human_standard$IHD < human_standard$V00 &
                    human_standard$CM < human_standard$V00 &
                    human_standard$ARR < human_standard$V00 &
                    human_standard$VD < human_standard$V00 &
                    human_standard$CHD < human_standard$V00 &
                    human_standard$CCD < human_standard$V00 &
                    human_standard$OTH < human_standard$V00 &
                    human_standard$V00 > 10), ]
```

| Entry | Entry name | Gene names | Protein names |
|------------|------------------|-------------------------------|---|
| P42357 | HUTH_HUMAN | HAL HIS | Histidine ammonia-lyase, Histidase, EC 4.3.1.3 |
| Q96542 | NODAL_HUMAN | NODAL | Nodal homolog |
| H7C0E4 | H7C0E4_HUMAN | NODAL | Nodal homolog |
| Q8NG44 | Q8NG44_HUMAN | PGR PR, hCG_39657 | Nuclear receptor subfamily 3 group C member 3 (Progesterone receptor) |
| E9PG18 | E9PG18_HUMAN | SCN5A | Sodium channel protein |
| A3EY21 | A3EY21_HUMAN | SCN5A | Cardiac sodium channel alpha subunit (Sodium channel protein type 5 subunit alpha) |
| Q12809 | KCNH2_HUMAN | KCNH2 ERG, ERG1, HERG | Potassium voltage-gated channel subfamily H member 2 (Eag homolog) (Ether-a-go-go-related gene potassium channel 1, ERG-1, Eag-related protein 1, Ether-a-go-go-related protein 1, H-ERG, hERG-1, hERG1) (Voltage-gated potassium channel subunit Kv11.1) |
| Q96HU1 | SGSM3_HUMAN | SGSM3 MAP, RABGAPL3, RUTBC3 | Small G protein signaling modulator 3 (Merlin-associated protein) (RUN and TBC1 domain-containing protein 3) (Rab-GTPase-activating protein-like protein, RabGAPLP) |
| H9KVD2 | H9KVD2_HUMAN | SCN5A | Sodium channel protein |
| Q8NG42 | Q8NG42_HUMAN | PGR PR, hCG_39657 | Nuclear receptor subfamily 3 group C member 3 (Progesterone receptor) |
| K4DIA1 | K4DIA1_HUMAN | SCN5A | Sodium channel protein |
| E9PHB6 | E9PHB6_HUMAN | SCN5A | Sodium channel protein |
| A0A0A0MT39 | A0A0A0MT39_HUMAN | SCN5A | Sodium channel protein |
| Q14524 | SCN5A_HUMAN | SCN5A | Sodium channel protein type 5 subunit alpha (Sodium channel protein cardiac muscle subunit alpha) (Sodium channel protein type V subunit alpha) (Voltage-gated sodium channel subunit alpha Nav1.5) (HH1) |
| Q8WXS4 | CCGL_HUMAN | TMEM37 PR | Voltage-dependent calcium channel gamma-like subunit (Neuronal voltage-gated calcium channel gamma-like subunit) (Transmembrane protein 37) |
| Q15477 | SKIV2_HUMAN | SKIV2L DDX13, SKI2W, SKIV2, W | Helicase SKI2W, Ski2, EC 3.6.4.- (Helicase-like protein, HLP) |
| Q8NG45 | Q8NG45_HUMAN | PGR PR, hCG_39657 | Nuclear receptor subfamily 3 group C member 3 (Progesterone receptor) |
| P19526 | FUT1_HUMAN | FUT1 H, HSC | Galactoside alpha-(1,2)-fucosyltransferase 1 (Alpha(1,2)FT 1) (Blood group H alpha 2-fucosyltransferase) (Fucosyltransferase 1) (GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase 1) (Type 1 galactoside alpha-(1,2)-fucosyltransferase FUT1, EC 2.4.1.69) (Type 2 galactoside alpha-(1,2)-fucosyltransferase FUT1, EC 2.4.1.344) |

Figure 6: UniProt entries for CCD

| ## | protein | IHD | CM | ARR | VD | CHD | CCD | |
|----|---------|------------|----------|----------|-----------|-----------|----------|-----------|
| ## | 3063 | c9j3b7 | 6.965302 | 6.641857 | 8.738036 | 9.254790 | 7.629177 | 10.947898 |
| ## | 5006 | p09486 | 7.814915 | 7.872700 | 9.890282 | 10.348461 | 9.123827 | 12.285759 |
| ## | 5877 | o15178 | 8.643149 | 8.908017 | 10.663996 | 10.596955 | 9.616940 | 13.278656 |
| ## | 6661 | a0a087x1b6 | 4.977847 | 6.024066 | 5.630716 | 11.208309 | 5.043599 | 5.269674 |
| ## | 10792 | a0a087wux9 | 4.977847 | 6.024066 | 5.630716 | 11.208309 | 5.043599 | 5.269674 |
| ## | 13447 | p42768 | 6.965302 | 6.641857 | 8.738036 | 9.254790 | 7.629177 | 10.947898 |
| ## | 15108 | a0a590ujl0 | 2.559614 | 2.182201 | 3.147555 | 5.969290 | 3.213214 | 3.335805 |
| ## | 18427 | p08842 | 2.559549 | 2.182144 | 3.147484 | 5.969165 | 3.217474 | 3.335734 |
| ## | 20344 | e9peg3 | 4.977847 | 6.024066 | 5.630716 | 11.208309 | 5.043599 | 5.269674 |
| ## | 20517 | a0a590ujy9 | 2.559614 | 2.182201 | 3.147555 | 5.969290 | 3.213214 | 3.335805 |
| ## | 24905 | q04671 | 8.231804 | 8.066166 | 10.062439 | 10.924957 | 8.722026 | 12.356011 |
| ## | 26564 | o14656 | 3.906485 | 3.591231 | 4.932987 | 8.620294 | 5.861550 | 6.229122 |
| ## | 27149 | q16831 | 7.027328 | 6.504890 | 8.550979 | 8.298618 | 6.353187 | 9.357606 |
| ## | 36256 | p10275 | 4.977663 | 6.023832 | 5.630504 | 11.210487 | 5.043400 | 5.269480 |
| ## | 38781 | p28072 | 7.460803 | 7.044626 | 8.885437 | 8.314015 | 7.943831 | 9.371391 |
| ## | 42793 | a0a590ujt4 | 2.559614 | 2.182201 | 3.147555 | 5.969290 | 3.213214 | 3.335805 |
| ## | 45683 | f5gzg9 | 4.977847 | 6.024066 | 5.630716 | 11.208309 | 5.043599 | 5.269674 |
| ## | 51633 | o43934 | 7.274412 | 7.321452 | 8.572355 | 8.072049 | 7.307056 | 8.940106 |
| ## | 52185 | q9nzc7 | 7.112592 | 6.932865 | 8.928574 | 9.204862 | 8.002171 | 10.727447 |
| ## | V00 | OTH | | | | | | |
| ## | 3063 | 12.27602 | 7.325668 | | | | | |
| ## | 5006 | 13.82318 | 8.338084 | | | | | |
| ## | 5877 | 13.30115 | 9.121595 | | | | | |
| ## | 6661 | 14.23471 | 5.771785 | | | | | |
| ## | 10792 | 14.23471 | 5.771785 | | | | | |
| ## | 13447 | 12.27602 | 7.325668 | | | | | |
| ## | 15108 | 10.64255 | 2.459249 | | | | | |
| ## | 18427 | 10.64234 | 2.459189 | | | | | |

```
## 20344 14.23471 5.771785
## 20517 10.64255 2.459249
## 24905 15.08283 8.576291
## 26564 14.52802 3.652153
## 27149 10.90972 6.381701
## 36256 14.23421 5.771562
## 38781 10.12635 7.120730
## 42793 10.64255 2.459249
## 45683 14.23471 5.771785
## 51633 10.17515 7.377648
## 52185 12.34425 7.329424
```

| <input type="checkbox"/> | Entry | Entry name | Gene names | Protein names |
|--------------------------|------------|------------------|------------------|--|
| <input type="checkbox"/> | C9J3B7 | C9J3B7_HUMAN | WAS | Wiskott-Aldrich syndrome protein |
| <input type="checkbox"/> | P09486 | SPRC_HUMAN | SPARC ON | SPARC (Basement-membrane protein 40, BM-40) (Osteonectin, ON) (Secreted protein acidic and rich in cysteine) |
| <input type="checkbox"/> | O15178 | TBXT_HUMAN | TBXT T | T-box transcription factor T (Brachyury protein) (Protein T) |
| <input type="checkbox"/> | A0A087X1B6 | A0A087X1B6_HUMAN | AR | Androgen receptor (Dihydrotestosterone receptor) (Nuclear receptor subfamily 3 group C member 4) |
| <input type="checkbox"/> | A0A087WUX9 | A0A087WUX9_HUMAN | AR | Androgen receptor |
| <input type="checkbox"/> | P42768 | WASP_HUMAN | WAS IMD2 | Wiskott-Aldrich syndrome protein, WASp |
| <input type="checkbox"/> | A0A590UJL0 | A0A590UJL0_HUMAN | STS | Steryl-sulfatase |
| <input type="checkbox"/> | P08842 | STS_HUMAN | STS ARSC1 | Steryl-sulfatase, EC 3.1.6.2 (Arylsulfatase C, ASC) (Estrone sulfatase) (Steroid sulfatase) (Steryl-sulfate sulfohydrolase) |
| <input type="checkbox"/> | E9PEG3 | E9PEG3_HUMAN | AR | Androgen receptor |
| <input type="checkbox"/> | A0A590UJY9 | A0A590UJY9_HUMAN | STS | Steryl-sulfatase |

Figure 7: UniProt entries for VOO

OTH

```
human_standard[which(human_standard$IHD < human_standard$OTH &
                      human_standard$CM < human_standard$OTH &
                      human_standard$ARR < human_standard$OTH &
                      human_standard$VD < human_standard$OTH &
                      human_standard$CHD < human_standard$OTH &
                      human_standard$CCD < human_standard$OTH &
                      human_standard$VOO < human_standard$OTH &
                      human_standard$OTH > 7), ]
```

```
##      protein      IHD      CM      ARR      VD      CHD      CCD      VOO
## 25902  p03950 4.081178 5.296124 4.070665 4.146777 2.871126 4.417061 5.367782
##      OTH
## 25902 7.288494
```

```
# proteins where OTH has highest z-score and above 7
```

| <input type="checkbox"/> | Entry | Entry name | Gene names | Protein names | |
|--------------------------|--------|------------|------------|---|--|
| <input type="checkbox"/> | P03950 | ANGI_HUMAN | ANG RNASE5 | Angiogenin, EC 3.1.27.- (Ribonuclease 5, RNase 5) | |

Figure 8: UniProt entries for OTH

UpSet Plots

```
library(UpSetR)
```

```
## Warning: package 'UpSetR' was built under R version 4.0.5
```

```
# http://gehlenborglab.org/research/projects/upsetr/

# Create dataframe with 8 columns corresponding to the 8 CVD groups
# 1 if protein is above threshold, 0 if not

# Threshold 0
human_binary <- ifelse(human_full[, 2:9] > 0, 1, 0)
human_binary <- data.frame("protein" = human_full$protein, human_binary)

# Threshold 0.01
human_binary_01 <- ifelse(human_full[, 2:9] > 0.01, 1, 0)
human_binary_01 <- data.frame("protein" = human_full$protein, human_binary_01)

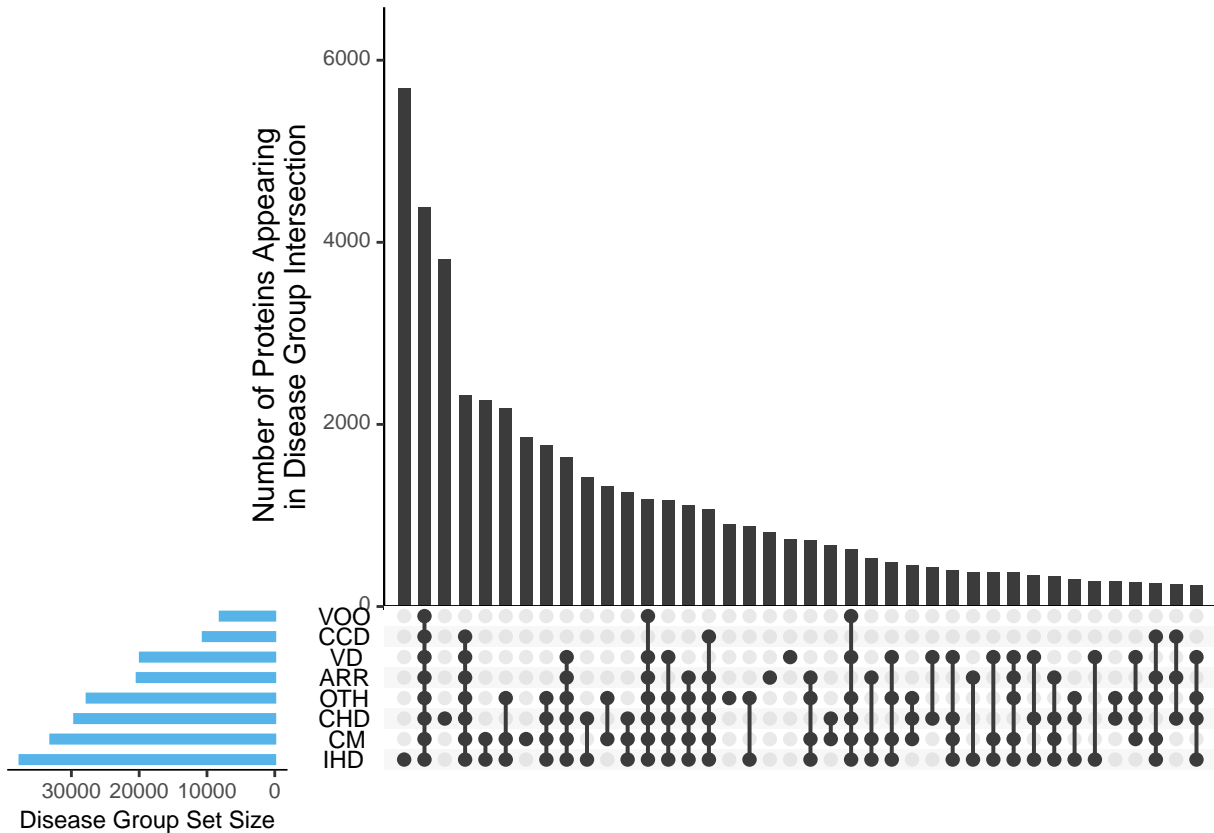
# Threshold 0.02
human_binary_02 <- ifelse(human_full[, 2:9] > 0.02, 1, 0)
human_binary_02 <- data.frame("protein" = human_full$protein, human_binary_02)

# Threshold 0.03
human_binary_03 <- ifelse(human_full[, 2:9] > 0.03, 1, 0)
human_binary_03 <- data.frame("protein" = human_full$protein, human_binary_03)

# Threshold 0.05
human_binary_05 <- ifelse(human_full[, 2:9] > 0.05, 1, 0)
human_binary_05 <- data.frame("protein" = human_full$protein, human_binary_05)
```

UpSet Plot with 50 Intersections

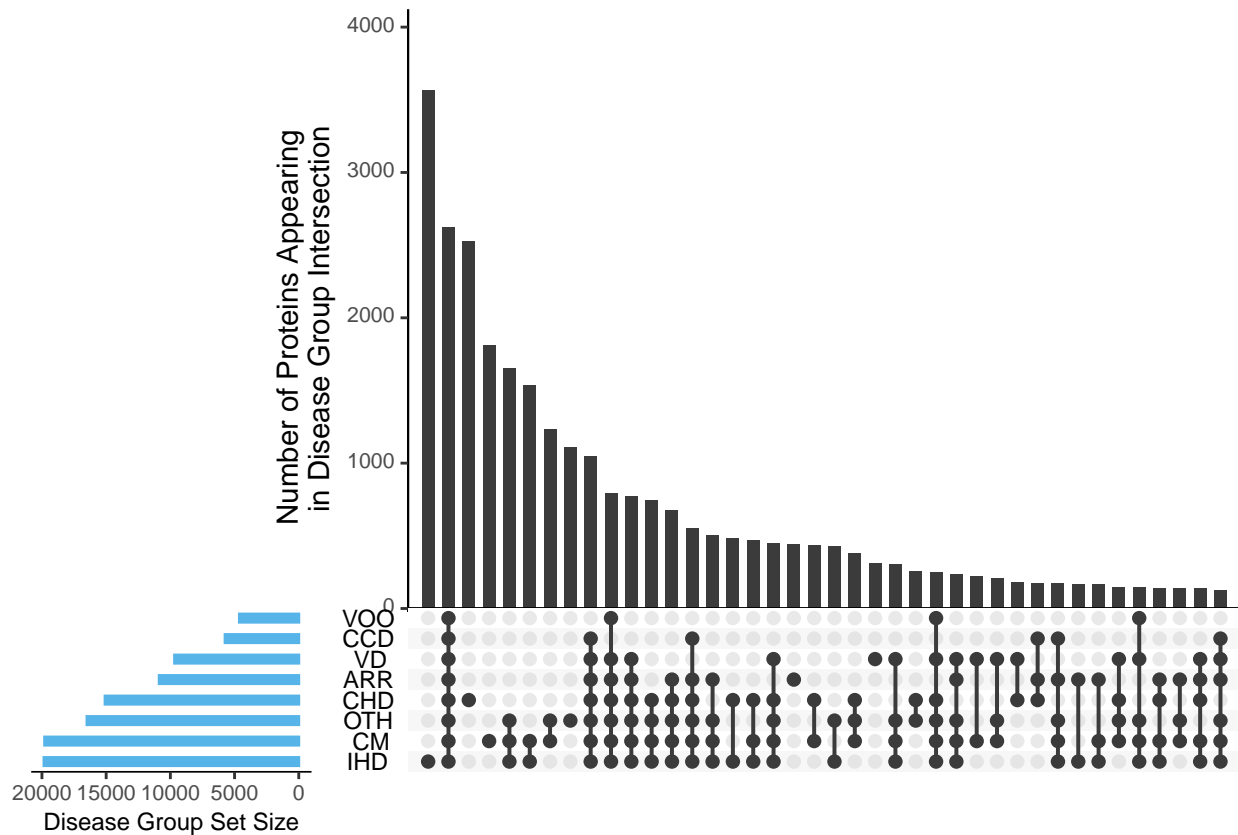
```
upset(human_binary,
      sets = c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
      sets.bar.color = "#56B4E9",
      number.angles = 5,
      # c(intersection size title, intersection size tick labels, set size title, set size tick labels,
      text.scale = c(1, 1, 1, 1, 1, 1.1),
      set_size.show = TRUE,
      empty.intersections = "on",
      nintersects = 50)
```

Threshold of 0.01 UpSet Plot

```
upset(human_binary_01,
      sets = c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
      sets.bar.color = "#56B4E9",
      number.angles = 5,
      # c(intersection size title, intersection size tick labels, set size title, set size tick labels,
      text.scale = c(1.3, 1.2, 1.1, 1.1, 1.25, 1.1),
      order.by = "freq",
      show.numbers = FALSE,
      mb.ratio = c(0.72, 0.28),
      empty.intersections = "on",
      mainbar.y.label = "Number of Proteins Appearing \n in Disease Group Intersection",
      sets.x.label = "Disease Group Set Size")
```

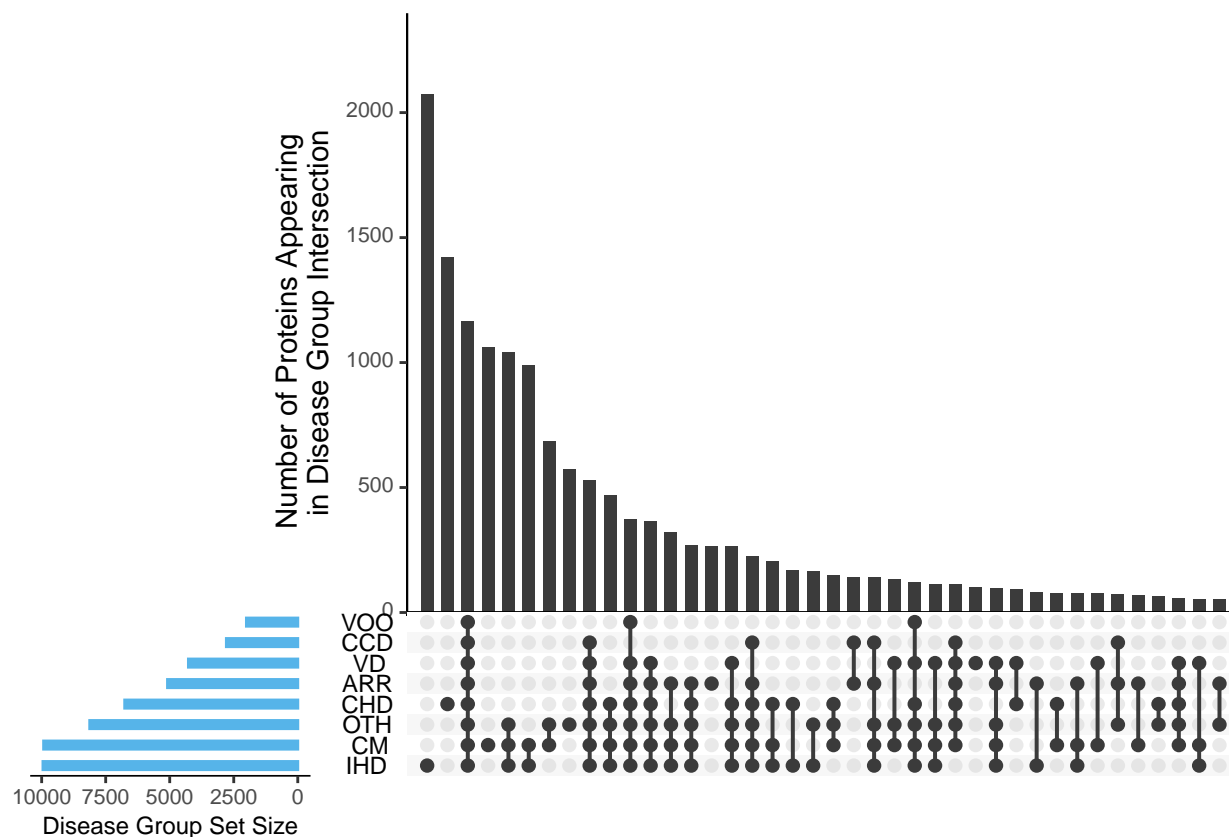
```
## Warning in Make_base_plot(Main_bar_plot = x$Main_bar, Matrix_plot = x$Matrix, :
## Plot might be out of range if ratio > 0.7 or < 0.3
```

Threshold of 0.02 UpSet Plot

```
upset(human_binary_02,
      sets = c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
      sets.bar.color = "#56B4E9",
      number.angles = 5,
      # c(intersection size title, intersection size tick labels, set size title, set size tick labels,
      text.scale = c(1.3, 1.2, 1.1, 1.1, 1.25, 1.1),
      order.by = "freq",
      show.numbers = FALSE,
      mb.ratio = c(0.72, 0.28),
      empty.intersections = "on",
      mainbar.y.label = "Number of Proteins Appearing \n in Disease Group Intersection",
      sets.x.label = "Disease Group Set Size")
```

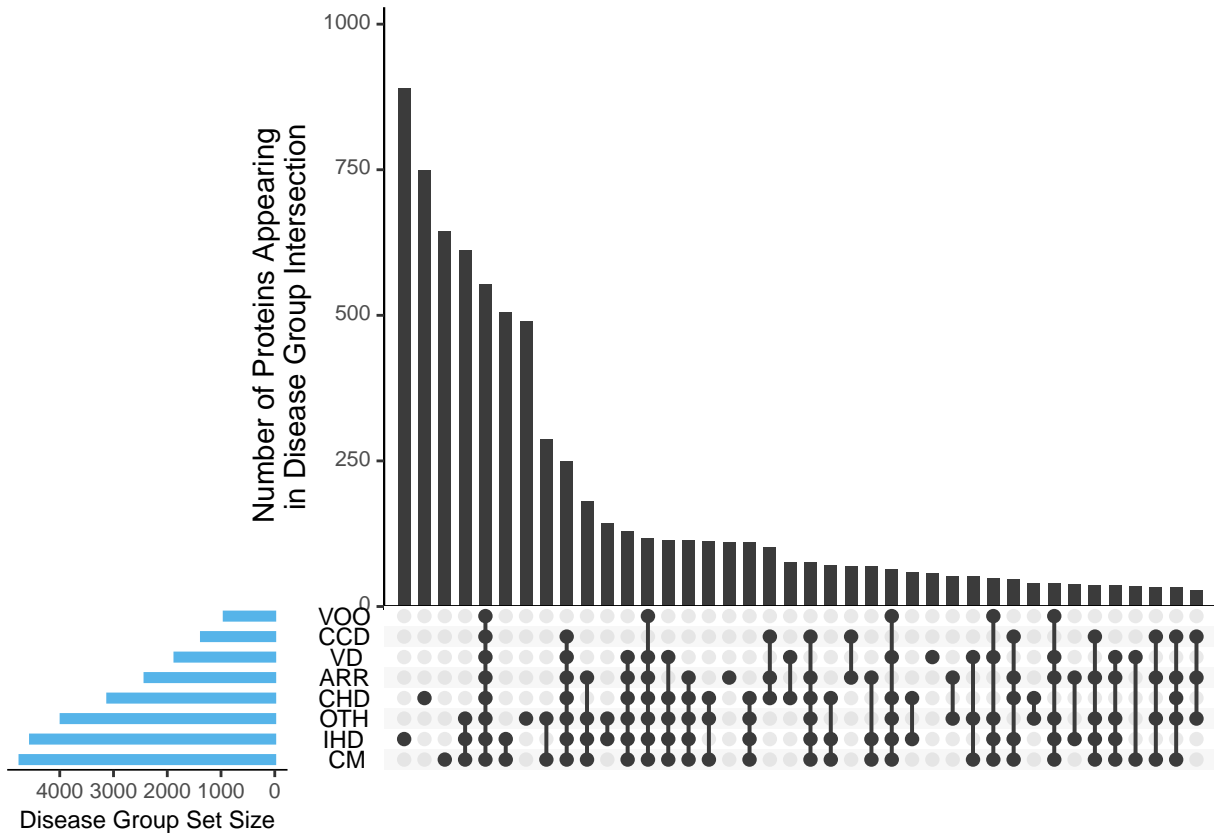
```
## Warning in Make_base_plot(Main_bar_plot = x$Main_bar, Matrix_plot = x$Matrix, :
## Plot might be out of range if ratio > 0.7 or < 0.3
```



Threshold of 0.03 UpSet Plot

```
upset(human_binary_03,
      sets = c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
      sets.bar.color = "#56B4E9",
      number.angles = 5,
      # c(intersection size title, intersection size tick labels, set size title, set size tick labels,
      text.scale = c(1.3, 1.2, 1.1, 1.1, 1.25, 1.1),
      order.by = "freq",
      show.numbers = FALSE,
      mb.ratio = c(0.72, 0.28),
      empty.intersections = "on",
      mainbar.y.label = "Number of Proteins Appearing \n in Disease Group Intersection",
      sets.x.label = "Disease Group Set Size")
```

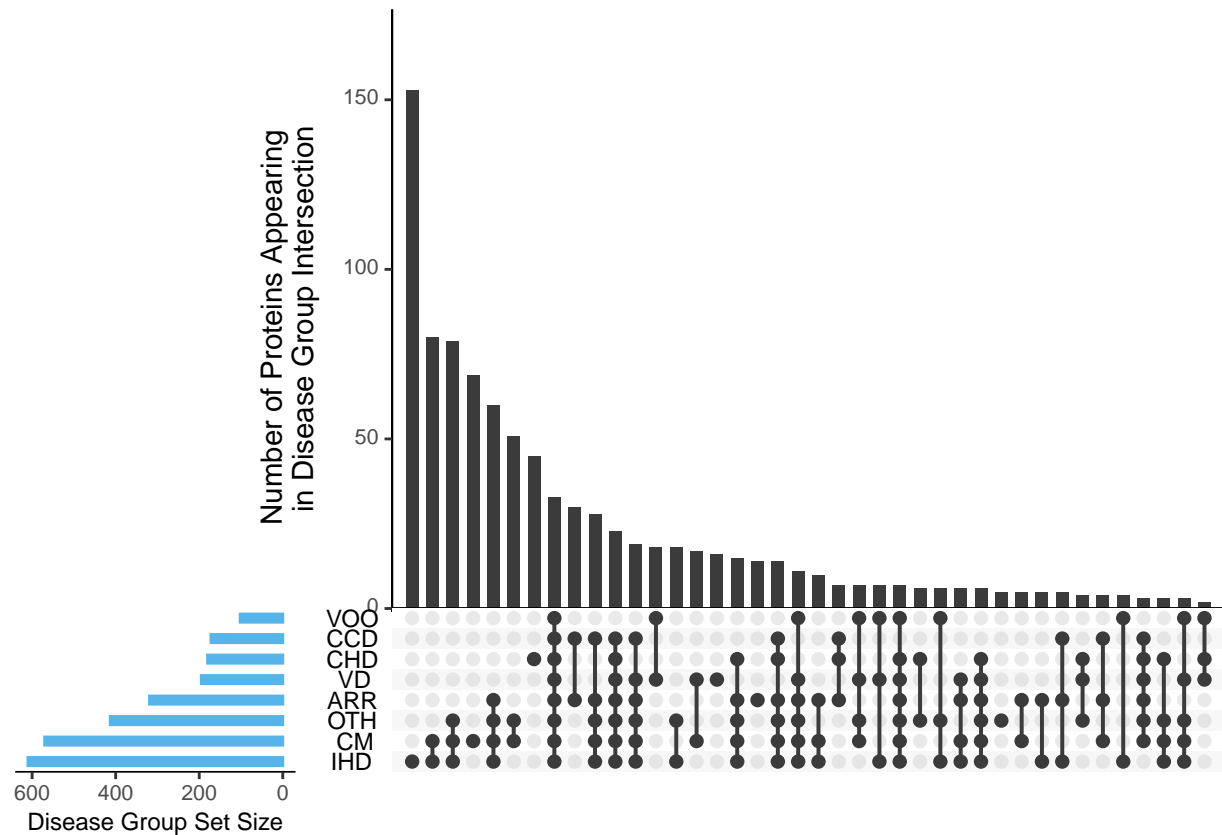
```
## Warning in Make_base_plot(Main_bar_plot = x$Main_bar, Matrix_plot = x$Matrix, :
## Plot might be out of range if ratio > 0.7 or < 0.3
```



Threshold of 0.05 UpSet Plot

```
upset(human_binary_05,
      sets = c("IHD", "CM", "ARR", "VD", "CHD", "CCD", "VOO", "OTH"),
      sets.bar.color = "#56B4E9",
      number.angles = 5,
      # c(intersection size title, intersection size tick labels, set size title, set size tick labels,
      text.scale = c(1.3, 1.2, 1.1, 1.1, 1.25, 1.1),
      order.by = "freq",
      show.numbers = FALSE,
      mb.ratio = c(0.72, 0.28),
      empty.intersections = "on",
      mainbar.y.label = "Number of Proteins Appearing \n in Disease Group Intersection",
      sets.x.label = "Disease Group Set Size")
```

```
## Warning in Make_base_plot(Main_bar_plot = x$Main_bar, Matrix_plot = x$Matrix, :
## Plot might be out of range if ratio > 0.7 or < 0.3
```



```
# library(devtools)
# install_github("jokergoo/ComplexHeatmap")
# Refer to https://jokergoo.github.io/ComplexHeatmap-reference/book/upset-plot.html#example-with-the-mo

# For Threshold 0 dataframe
# Make binary matrix
human_binary_mat <- ComplexHeatmap::make_comb_mat(human_binary)
# Get size for each intersection
human_binary_comb_size <- sort(ComplexHeatmap::comb_size(human_binary_mat), decreasing = TRUE)
# Get size for each category set
human_binary_set_size <- ComplexHeatmap::set_size(human_binary_mat)

# For Threshold 0.03 dataframe
# Make binary matrix
human_binary_mat_03 <- ComplexHeatmap::make_comb_mat(human_binary_03)
# Get size for each intersection
human_binary_comb_size_03 <- sort(ComplexHeatmap::comb_size(human_binary_mat_03), decreasing = TRUE)
# Get size for each category set
human_binary_set_size_03 <- ComplexHeatmap::set_size(human_binary_mat_03)

# Get intersections with intersection size > 200 in the Threshold 0 dataframe
human_binary_names_subset <- names(human_binary_comb_size[human_binary_comb_size > 200])
# Find the size of those intersections in the Threshold 0.03 dataframe
human_binary_comb_size_03[human_binary_names_subset]
```

```
## 10000000 11111111 00001000 11111101 11000000 11000001 01000000 11001001
```

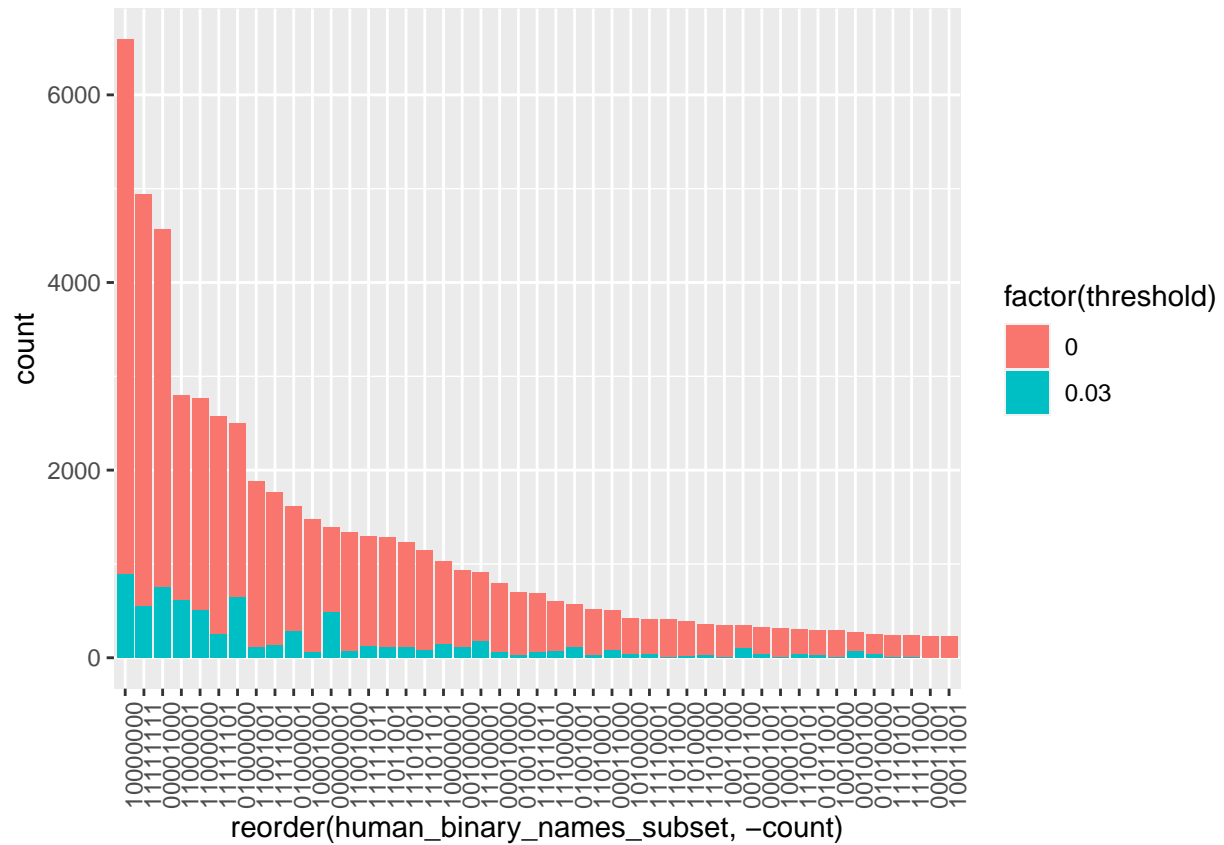
```
##      891      554      750      249      505      613      644      110
## 11111001 10001000 01000001 11001000 11111011 11011001 11101001 11101101
##      129      59      287      72      118      115      114      76
## 00000001 10000001 00100000 00010000 11100001 01001000 11011011 11100000
##      490      143      111      58      181      24      64      69
## 11010001 01001001 00011000 11011000 10100000 11010000 11110001 10011000
##      24      112      77      10      39      12      37      2
## 11101000 10001001 10010000 00001001 01011001 11100101 00101100      <NA>
##      28      10      7      40      26      37      102      NA
## 11110101 11111000      <NA> 01010000 00100100
##      7      7      NA      36      69
```

Stacked Barplot comparing threshold 0 and threshold 0.03

```
human_binary_subset <- data.frame(human_binary_names_subset,
                                  "count" = human_binary_comb_size[human_binary_names_subset],
                                  "threshold" = rep(0, 45))
human_binary_subset_03 <- data.frame(human_binary_names_subset,
                                     "count" = human_binary_comb_size_03[human_binary_names_subset],
                                     "threshold" = rep(0.03, 45))
human_binary_subset_double <- rbind(human_binary_subset, human_binary_subset_03)

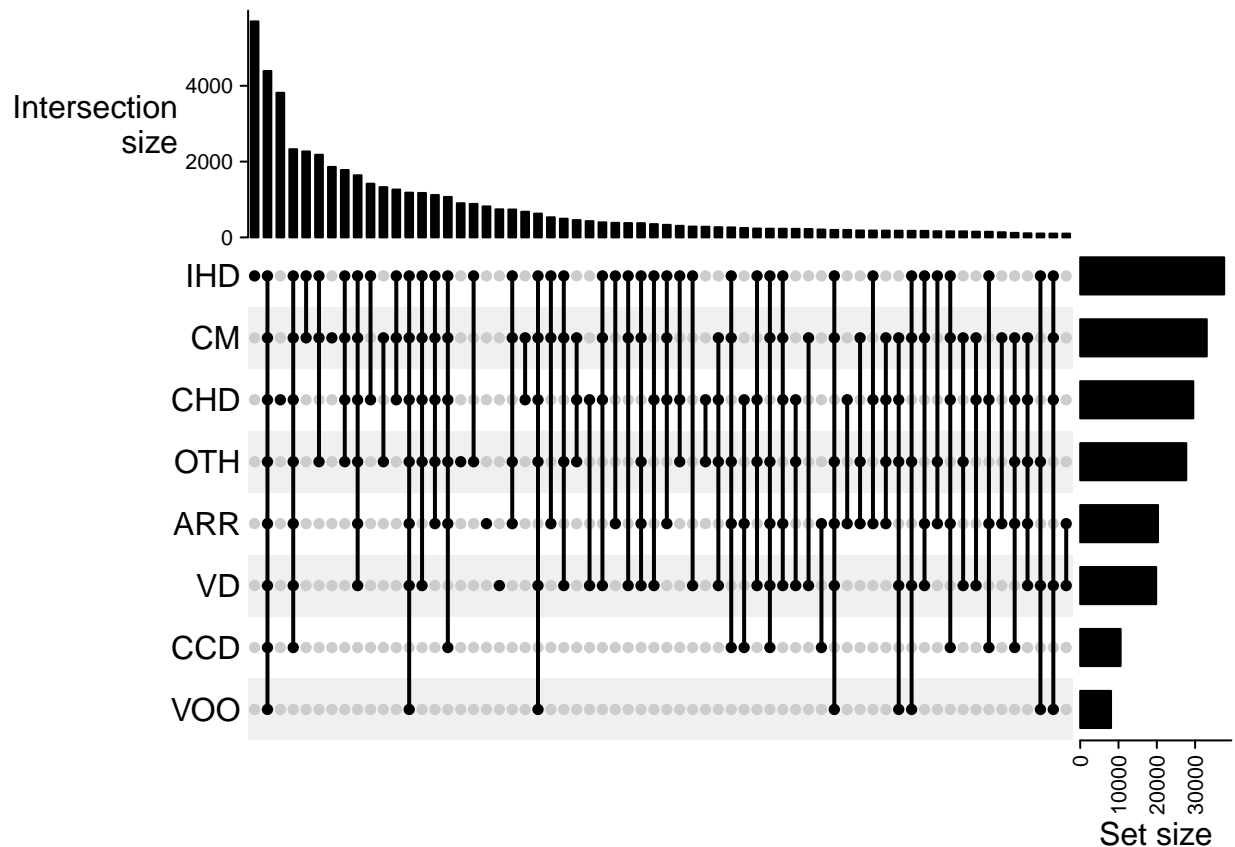
ggplot(human_binary_subset_double, aes(fill=factor(threshold),
                                              y=count, x=reorder(human_binary_names_subset, -count))) +
  geom_bar(position="stack", stat="identity") + theme(axis.text.x = element_text(angle = 90))

## Warning: Removed 2 rows containing missing values (position_stack).
```



```
# Exploring the ComplexHeatmap UpSet() function
human_binary_mat_subset <- human_binary_mat[ComplexHeatmap::comb_size(human_binary_mat) >= 100]

ComplexHeatmap::UpSet(human_binary_mat_subset,
  set_order = order(ComplexHeatmap::set_size(human_binary_mat_subset),
    decreasing = TRUE),
  comb_order = order(ComplexHeatmap::comb_size(human_binary_mat_subset),
    decreasing = TRUE),
  pt_size = unit(2, "mm"))
```



Stacked UpSet Plot comparing threshold 0 and threshold 0.03

```
# Setting up dataframe
human_binary_TF <- human_binary
human_binary_TF[2:9] <- human_binary[2:9] == 1
human_binary_TF["Threshold"] <- '0'

human_binary_03_TF <- human_binary_03
human_binary_03_TF[2:9] <- human_binary_03[2:9] == 1
human_binary_03_TF["Threshold"] <- '0.03'

human_binary_double_TF <- rbind(human_binary_TF, human_binary_03_TF)
CVD_groups <- colnames(human_binary)[2:9]

# Upset plot
library(ComplexUpset)

##
## Attaching package: 'ComplexUpset'

## The following object is masked from 'package:UpSetR':
##
##      upset
```

```

upset(
  human_binary_double_TF,
  CVD_groups,
  base_annotations = list(
    'Number of Proteins Appearing \n in Disease Group Intersection' = intersection_size(
      counts = FALSE,
      mapping = aes(fill = Threshold)
    ) + scale_fill_manual(values=c(
      '0'='#E41A1C', '0.03'='#377EB8'))
  ),
  width_ratio = 0.15,
  n_intersections = 40,
  max_size = 12000,
  set_sizes=(
    upset_set_size(geom=geom_bar(
      aes(x=group),
      width=0.8),
      position = 'right')
    + theme(axis.text.x=element_text(angle=90))
    + ylab('Disease Group\n Set Size')),
  guides='over',
  name=''
)

```

