Exploring AMP-AD metadata and gene expression.

Gilderlanio Santana de Araújo

2023-03-13

Exploring metadata from AMP-AD.

```
mayo <- read.csv("original_data/RNAseq_Harmonization_Mayo_combined_metadata.csv")
mayo <- unique(mayo[, c("specimenID", "tissue", "diagnosis")])
mayo <- mayo[mayo$diagnosis %in% c("AD", "CN"),]
write.table(mayo, "sample_tissue_diag_data/AMPAD_Mayo_Samples.txt",
            row.names = F, sep = "\t", quote = F)
table(mayo$tissue, mayo$diagnosis)
```

```
##
##                   AD CN
##   cerebellum      84 78
##   temporal cortex 82 78
```

```
# COGDX
# 1 NCI: No cognitive impairment
# 2 MCI: Mild cognitive impairment, no other condition contributing to CI
# 3 MCI+: Mild cognitive impairment AND another condition contributing to CI
# 4 AD: Alzheimer's dementia, no other condition contributing to CI (NINCDS/ADRDA Probable AD)
# 5 AD+: Alzheimer's dementia AND other condition contributing to CI (NINCDS/ADRDA Possible AD)
# 6 Other dementia: Other primary cause of dementia, no clinical evidence of Alzheimer's dementia

# ACC : anterior cingulate cortex
# DLPFC : dorsolateral prefrontal cortex
# PCC : posterior cingulate cortex

rosmap <- read.csv("original_data/ROSMAP_Covariates_ages_censored.tsv", sep = "\t")
rosmap <- unique(rosmap[, c("specimenID", "tissue", "diagnosis")])
rosmap <- rosmap[rosmap$diagnosis %in% c("CT", "AD"),]
rosmap <- rosmap[rosmap$tissue != "",]
rosmap[rosmap$diagnosis == "CT", ]$diagnosis <- "CN"
rosmap <- na.omit(rosmap)
write.table(rosmap, "sample_tissue_diag_data/AMPAD_ROSMAP_Samples.txt",
            row.names = F, sep = "\t", quote = F)
table(rosmap$tissue, rosmap$diagnosis)
```

```
##
##          AD  CN
##   ACC   171  94
```

```
##   DLPFC 294 143
##   PCC   151  99
```

```
# CERAD criteria is defined as 1 = normal, 2 = definite AD, 3 = Probable AD, 4 = possible AD
# https://www.sciencedirect.com/science/article/pii/B978012804832000002X

msbb <- read.csv("original_data/RNAseq_Harmonization_MSBB_combined_metadata.csv")
msbb <- unique(msbb[, c("specimenID", "tissue", "CDR", "CERAD")])
msbb <- na.omit(msbb)
msbb$diagnosis <- "X"
msbb[msbb$CERAD == 1, ]$diagnosis <- "CN"
msbb[msbb$CERAD == 2, ]$diagnosis <- "AD"
msbb[msbb$CERAD == 3, ]$diagnosis <- "Prob AD"
msbb[msbb$CERAD == 4, ]$diagnosis <- "Poss AD"

write.table(msbb, "sample_tissue_diag_data/AMPAD_MSBB_Samples.txt",
            row.names = F, sep = "\t", quote = F)

table(msbb$tissue, msbb$diagnosis)
```

```
##
##                            AD  CN Poss AD Prob AD
##   frontal pole            132  90      43      41
##   inferior frontal gyrus  132  91      41      37
##   parahippocampal gyrus   146  83      36      46
##   prefrontal cortex         8   3       0       3
##   superior temporal gyrus 149  86      44      48
```

```
mayo.exp <- read.csv("original_data/Mayo_Filtered_counts_(greater_than_1cpm).tsv", sep = "\t", check.nam
rownames(mayo.exp) <- mayo.exp$feature

rosmap.exp <- read.csv("original_data/ROSMAP_Filtered_counts_(greater_than_1cpm).tsv", sep = "\t", check
rownames(rosmap.exp) <- rosmap.exp$feature

msbb.exp <- read.csv("original_data/MSBB_Filtered_counts_(greater_than_1cpm).tsv", sep = "\t", check.nam
rownames(msbb.exp) <- msbb.exp$feature

split.data <- function(project, samples, transcriptome, outdir) {
  diagnosis <- unique(samples$diagnosis)
  tissues <- unique(samples$tissue)
  for (diag in diagnosis) {
    if (diag %in% c("AD", "CN")) {
        for (tissue in tissues) {
          target.samples <- samples[samples$tissue == tissue & samples$diagnosis == diag, ]$specimenID
          target.samples <- intersect(colnames(transcriptome), target.samples)
          exp.data <- transcriptome[,target.samples]
          write.table(exp.data, paste(paste(outdir, project, sep="/"), diag, tissue, ".count", sep = "_
                row.names = T, sep = "\t", quote = F, append = F)
        }
    }
  }
}
```

```
split.data("AMPAD_Mayo", mayo, mayo.exp, "counts")
split.data("AMPAD_ROSMAP", rosmap, rosmap.exp, "counts")
split.data("AMPAD_MSBB", msbb, msbb.exp, "counts")
```