Coor_PCA

```
source_rmd = function(file, ...) {
  tmp_file = tempfile(fileext=".R")
  on.exit(unlink(tmp_file), add = TRUE)
  knitr::purl(file, output=tmp_file)
  source(file = tmp_file, ...)
}
```

```
options(knitr.duplicate.label = "allow")
source_rmd("rawdata_normalization.rmd")
```

Correlation analysis heatmap

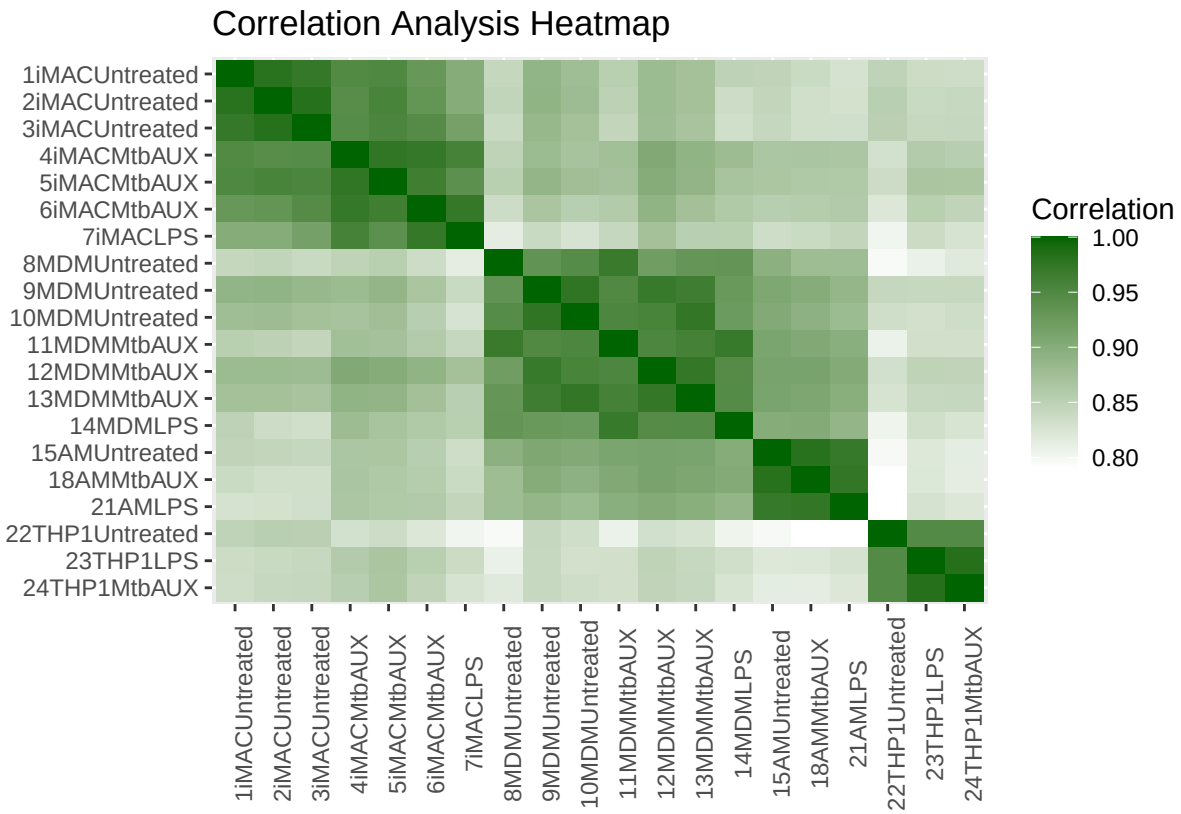
```
cols2 <- norm_exp_matrix_am_rm[,1:20]

cc2 = cor(cols2, method = "spearman")
cc_df = as.data.frame(cc2)
cc_df$samples = row.names(cc_df)

ccm = melt(cc_df, id = "samples")

ccm$samples <- factor(ccm$samples,levels=unique(ccm$samples))

ggplot(ccm, mapping = aes(x= variable, y = samples))+
   geom_tile(aes(fill = value)) +
   scale_fill_gradient(low = "white", high = "darkgreen") +
    scale_y_discrete(limits = rev(levels(ccm$samples))) +
    theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Correlation Analysis Heatmap", x = "", y = "", fill = "Correlation")
```



PCA plot

```
norm_pca <- inner_join(as.data.frame(z_transformed_avg_norm_exp_am_rm) %>% rownames_to_column("Gene_ID")

norm_pca <- inner_join(avg_norm_exp_as_df_am_rm, gene.info[c(1,11)])

norm_pca[1] <- NULL


pca_matrix <-
   distinct(norm_pca, gene_source, .keep_all = TRUE) %>%
   column_to_rownames("gene_source") %>%
   as.matrix() %>%
   t()
```

```
# sample_pca <- prcomp(pca_matrix, scale. = TRUE)
sample_pca <- prcomp(pca_matrix)


pc_eigenvalues <- sample_pca$sdev ^ 2

# create a "tibble" manually with
# a variable indicating the PC number
# and a variable with the variances
pc_eigenvalues <- tibble(PC = factor(1:length(pc_eigenvalues)),
                         variance = pc_eigenvalues) %>%
```

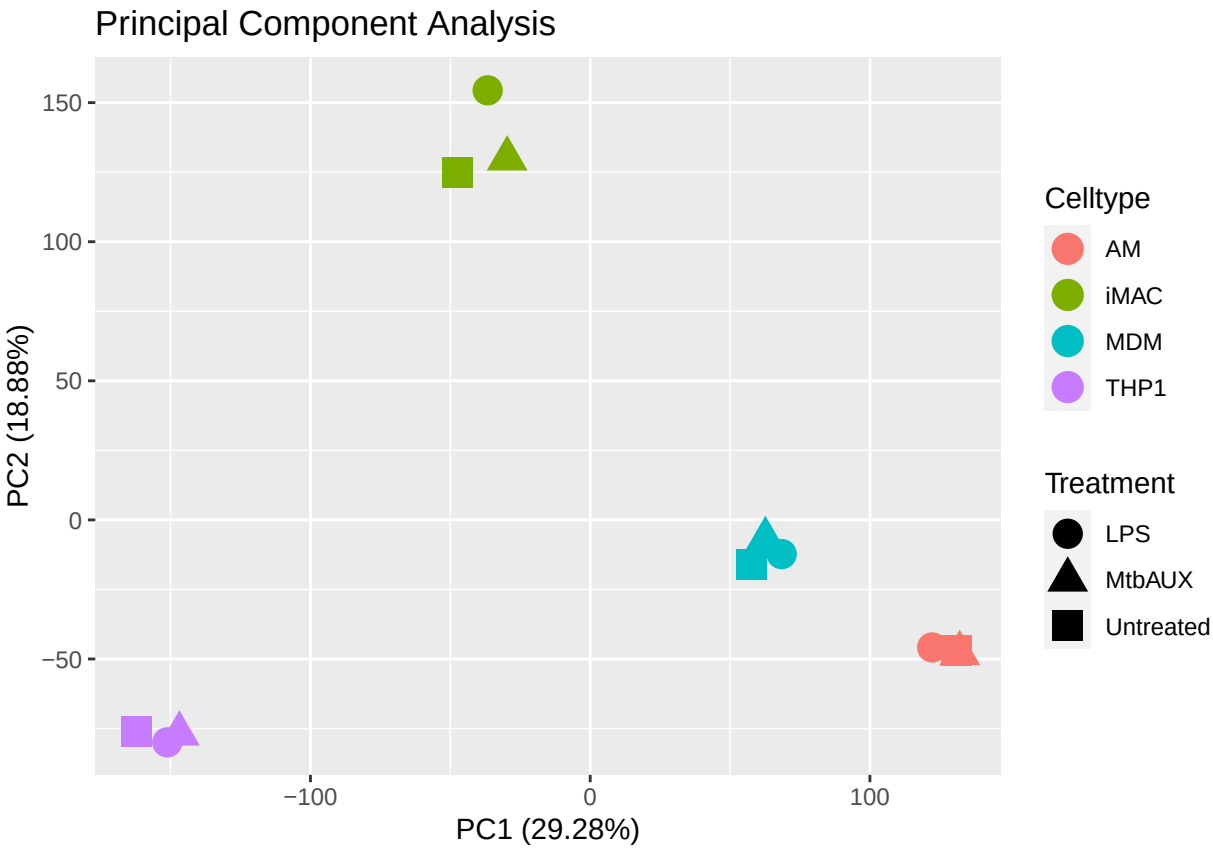
```
   # add a new column with the percent variance
   mutate(pct = variance / sum(variance) * 100) %>%
   # add another column with the cumulative variance explained
   mutate(pct_cum = cumsum(pct))



pc_scores <- sample_pca$x

pc_scores <- pc_scores %>%
   # convert to a tibble retaining the sample names as a new column
   as_tibble(rownames = "sample")
```

```
ggplot(
   pc_scores %>%
      mutate(
         "celltype" = rep(sub(
            "iMACs", "iMAC", unique(sample.info[[2]])
         ), each = 3),
         "treatment" = rep(unique(sample.info[[3]]), times =4)
      ) ,
   mapping = aes(PC1, PC2, color = celltype, shape = treatment)
) +
   geom_point(size = 5) +
   labs(
      color = "Celltype",
      shape = "Treatment",
      x = "PC1 (29.28%)",
      y = "PC2 (18.88%)",
      title = "Principal Component Analysis"
   )
```


Principal Component Analysis
150
100
50
0
−50
PC2 (18.88%)
−100
0
100
PC1 (29.28%)
Celltype
AM
iMAC
MDM
THP1
Treatment
LPS
MtbAUX
Untreated