Unsupervised Multiblock analyses

CODE

Florence Mehl

February 28, 2025

- Nutrimouse dataset
- Analysis of complete dataset
    - Question 1: based on lipids and genes data, do we observe clusters of samples ?
    - Question 2: how do both blocks contribute to each dimension?
    - Question 3: observe the samples distributions in the space of the common dimensions, what are the main sources of variation?
    - Question 4: which genes and lipids are responsible of the samples differences?
- Repeat analysis for wt samples only
- Repeat analysis for ppar samples only

Nutrimouse dataset

The data sets come from a nutrigenomic study in the mouse (Martin et al., 2007) in which the effects of five regimens with contrasted fatty acid compositions on liver lipids and hepatic gene expression in mice were considered.

Two sets of variables were acquired on forty mice: - genes: expressions of 120 genes measured in liver cells, selected (among about 30,000) as potentially relevant in the context of the nutrition study. These expressions come from a nylon macroarray with radioactive labelling - lipids: concentrations (in percentages) of 21 hepatic fatty acids measured by gas chromatography

Biological units (mice) were cross-classified according to two factors experimental design (4 replicates): - genotype: 2-levels factor, wild-type (WT) and PPARalpha -/- (PPAR) - diet: 5-levels factor. Oils used for experimental diets preparation were corn and colza oils (50/50) for a reference diet (REF), hydrogenated coconut oil for a saturated fatty acid diet (COC), sunflower oil for an Omega6 fatty acid-rich diet (SUN), linseed oil for an Omega3-rich diet (LIN) and corn/colza/enriched fish oils for the FISH diet (43/43/14)

HIDE

```
data("nutrimouse")
genes <- nutrimouse$gene
lipids <- nutrimouse$lipid
metadata <- data.frame(genotype = nutrimouse$genotype, diet = nutrimouse$diet)
metadata$sample_name <- paste0(rownames(metadata), "_", metadata$genotype, "_", metadata$diet)
rownames(genes) <- metadata$sample_name
rownames(lipids) <- metadata$sample_name
```

Analysis of complete dataset

Question 1: based on lipids and genes data, do we observe clusters of samples ?

Prepare dataset - concatenate genes and lipids dataframes - define the number of variables of both block

Run ComDim analysis - use ComDim() from MBAnalysis package

HIDE

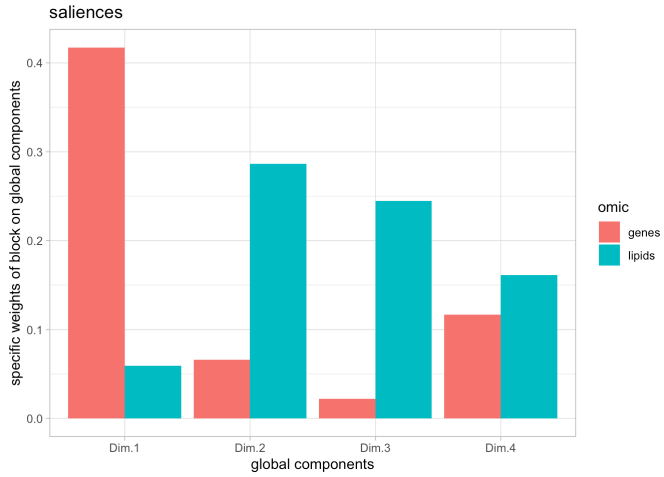
```
# prepare dataset
ComDim_data <- cbind.data.frame(genes, lipids)
n_group <- c(dim(genes)[[2]], dim(lipids)[[2]])

# run analysis
ComDim_res <- ComDim(X = ComDim_data,
                     block = n_group,
                     name.block = c("genes", "lipids"),
                     scale = T,
                     scale.block = T)
```

Question 2: how do both blocks contribute to each dimension?

- plot saliences

HIDE

```
# saliences

saliences <- ComDim_res$saliences
rownames(saliences) <- c("genes", "lipids")
saliences <- as.data.frame(t(saliences[,1:4]))
saliences$Dim <- rownames(saliences)
saliences <- melt(saliences)

ggplot(saliences, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components",
       y = "specific weights of block on global components",
       fill = "omic",
       title = "saliences")
```

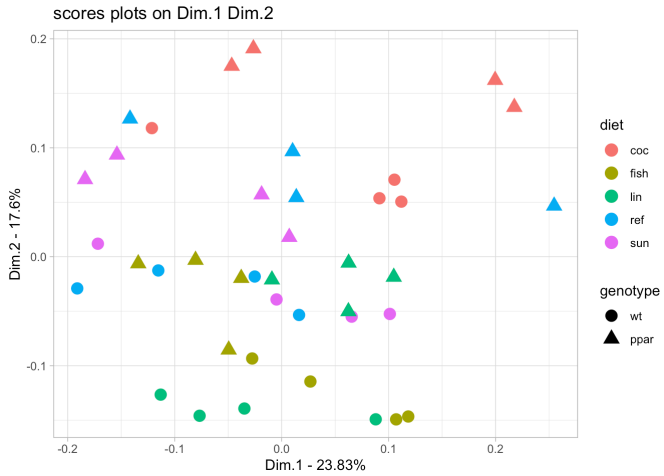


Question 3: observe the samples distributions in the space of the common dimensions, what are the main sources of variation?

- plot scores (Scor.g) on Dim.1 vs Dim.2 with percentages of explained variance on axes
- plot scores (Scor.g) on Dim.3 vs Dim.4 with percentages of explained variance on axes

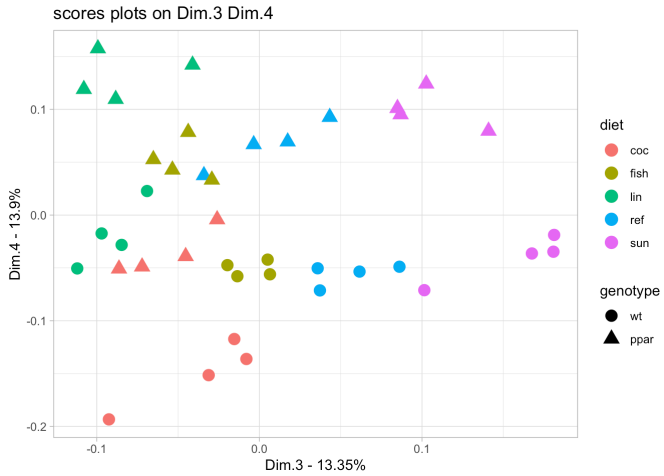
HIDE

```
scores <- data.frame(metadata, ComDim_res$Scor.g)

ggplot(scores, aes(x=Dim.1, y=Dim.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "scores plots on Dim.1 Dim.2") +
  theme_light()
```



HIDE

```
ggplot(scores, aes(x=Dim.3, y=Dim.4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.3 - ", ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "scores plots on Dim.3 Dim.4") +
theme_light()
```

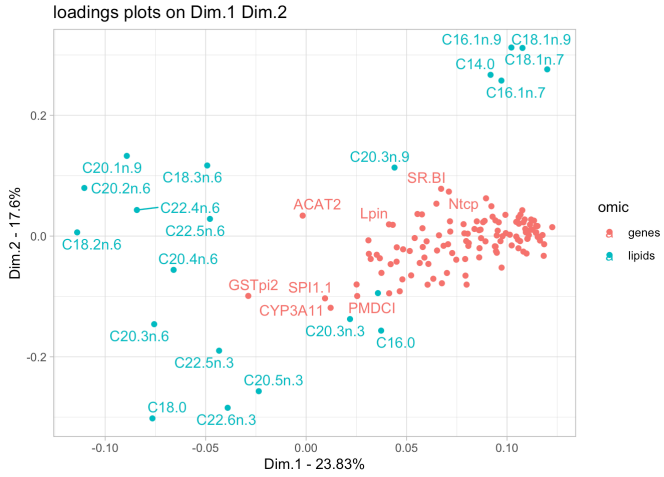


Question 4: which genes and lipids are responsible of the samples differences?

- plot loadings (Load.g) on Dim.1 vs Dim.2 with percentages of explained variance on axes
- plot loadings (Load.g) on Dim.3 vs Dim.4 with percentages of explained variance on axes

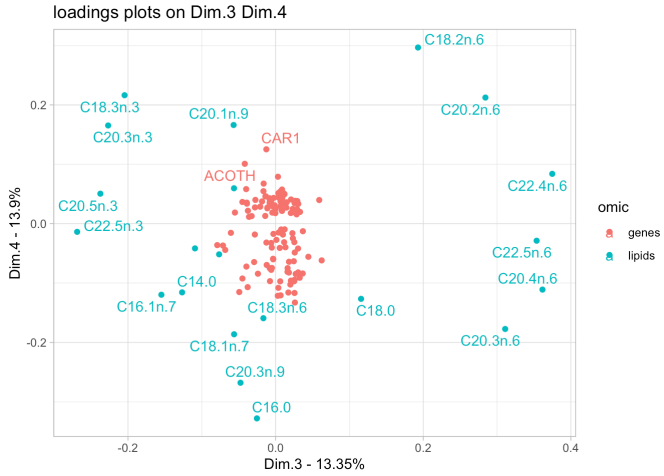
HIDE

```
loadings <- data.frame(ComDim_res$Load.g)
loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
loadings$variable <- rownames(loadings)

ggplot(loadings, aes(x=Dim.1, y=Dim.2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```



HIDE

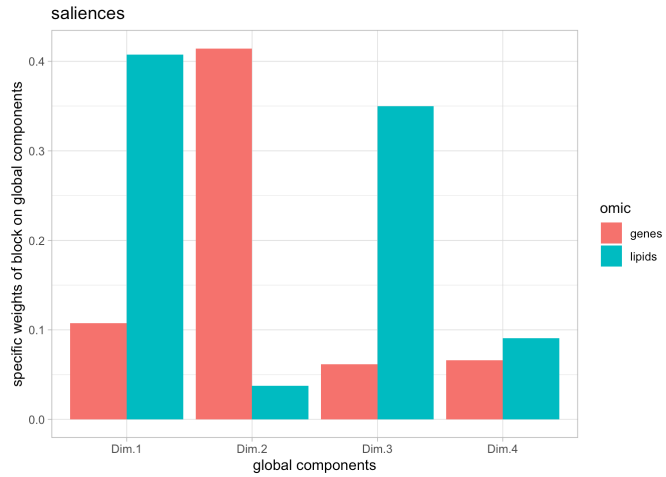
```
ggplot(loadings, aes(x=Dim.3, y=Dim.4, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.3 - ", ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "loadings plots on Dim.3 Dim.4") +
  theme_light()
```



Repeat analysis for wt samples only

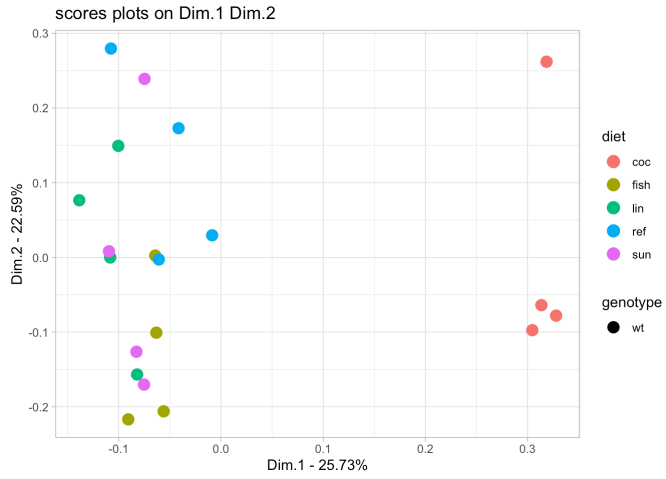
HIDE

```
# prepare dataset
wt_samples <- metadata$sample_name[metadata$genotype == "wt"]
wt_ComDim_data <- cbind.data.frame(genes[wt_samples,], lipids[wt_samples,])
n_group <- c(dim(genes)[[2]], dim(lipids)[[2]])

# run analysis
wt_ComDim_res <- ComDim(X = wt_ComDim_data,
                        block = n_group,
                        name.block = c("genes", "lipids"),
                        scale = T,
                        scale.block = T)

# saliences
wt_saliences <- wt_ComDim_res$saliences
rownames(wt_saliences) <- c("genes", "lipids")
wt_saliences <- as.data.frame(t(wt_saliences[,1:4]))
wt_saliences$Dim <- rownames(wt_saliences)
wt_saliences <- melt(wt_saliences)
ggplot(wt_saliences, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components",
       y = "specific weights of block on global components",
       fill = "omic",
       title = "saliences")
```



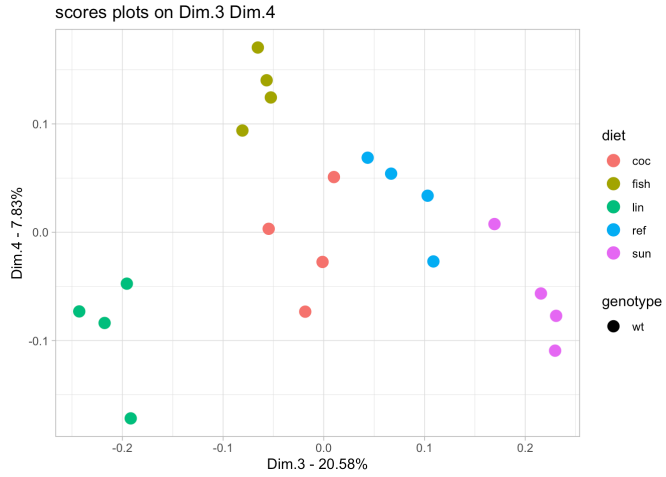
HIDE

```
# scores plots
wt_scores <- data.frame(metadata[metadata$genotype == "wt",], wt_ComDim_res$Scor.g)
ggplot(wt_scores, aes(x=Dim.1, y=Dim.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", wt_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", wt_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "scores plots on Dim.1 Dim.2") +
theme_light()
```



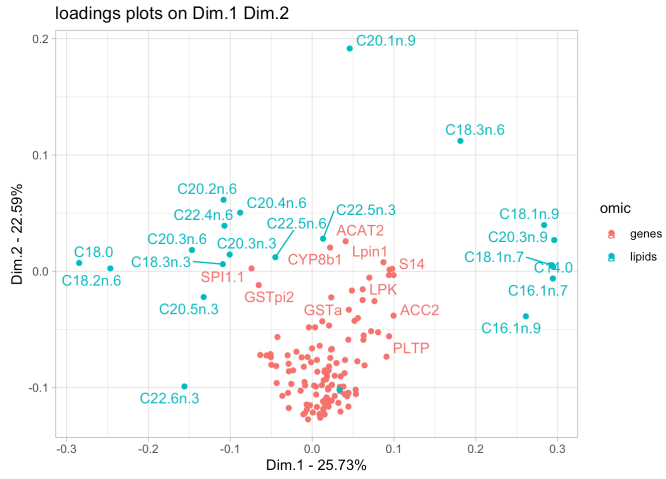
HIDE

```
ggplot(wt_scores, aes(x=Dim.3, y=Dim.4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.3 - ", wt_ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", wt_ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "scores plots on Dim.3 Dim.4") +
theme_light()
```



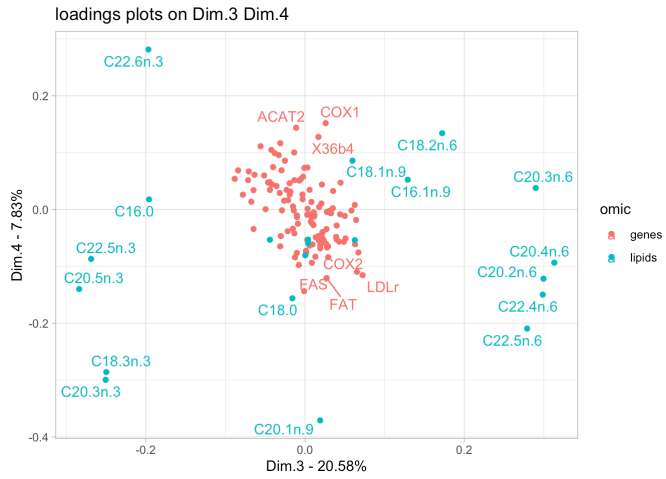
HIDE

```
# loadings plots
wt_loadings <- data.frame(wt_ComDim_res$Load.g)
wt_loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
wt_loadings$variable <- rownames(wt_loadings)
ggplot(wt_loadings, aes(x=Dim.1, y=Dim.2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", wt_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", wt_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```



HIDE

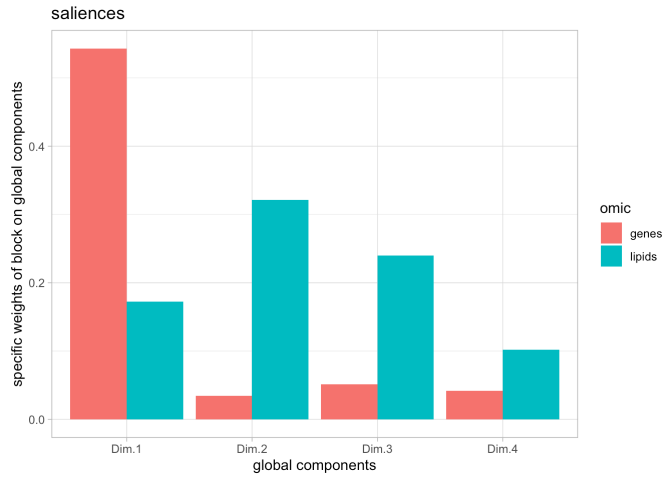
```
ggplot(wt_loadings, aes(x=Dim.3, y=Dim.4, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.3 - ", wt_ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", wt_ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "loadings plots on Dim.3 Dim.4") +
  theme_light()
```



Repeat analysis for ppar samples only

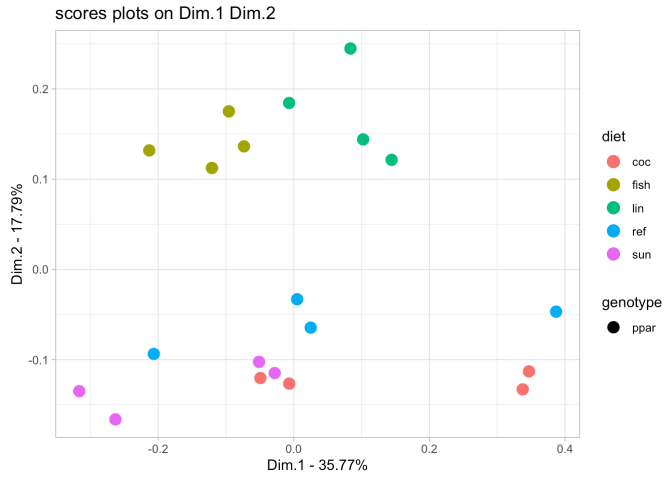
HIDE

```
# prepare dataset
ppar_samples <- metadata$sample_name[metadata$genotype == "ppar"]
ppar_ComDim_data <- cbind.data.frame(genes[ppar_samples,], lipids[ppar_samples,])
n_group <- c(dim(genes)[[2]], dim(lipids)[[2]])

# run analysis
ppar_ComDim_res <- ComDim(X = ppar_ComDim_data,
                    block = n_group,
                    name.block = c("genes", "lipids"),
                    scale = T,
                    scale.block = T)

# saliences
ppar_saliences <- ppar_ComDim_res$saliences
rownames(ppar_saliences) <- c("genes", "lipids")
ppar_saliences <- as.data.frame(t(ppar_saliences[,1:4]))
ppar_saliences$Dim <- rownames(ppar_saliences)
ppar_saliences <- melt(ppar_saliences)
ggplot(ppar_saliences, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components",
       y = "specific weights of block on global components",
       fill = "omic",
       title = "saliences")
```



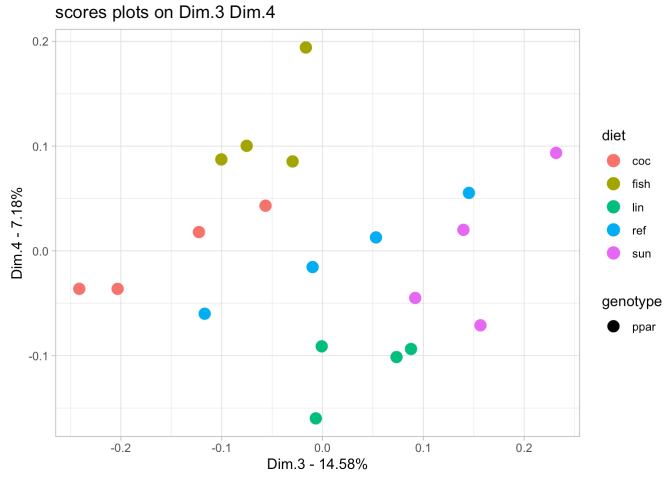
HIDE

```
# scores plots
ppar_scores <- data.frame(metadata[metadata$genotype == "ppar",], ppar_ComDim_res$Scor.g)
ggplot(ppar_scores, aes(x=Dim.1, y=Dim.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", ppar_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ppar_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "scores plots on Dim.1 Dim.2") +
  theme_light()
```



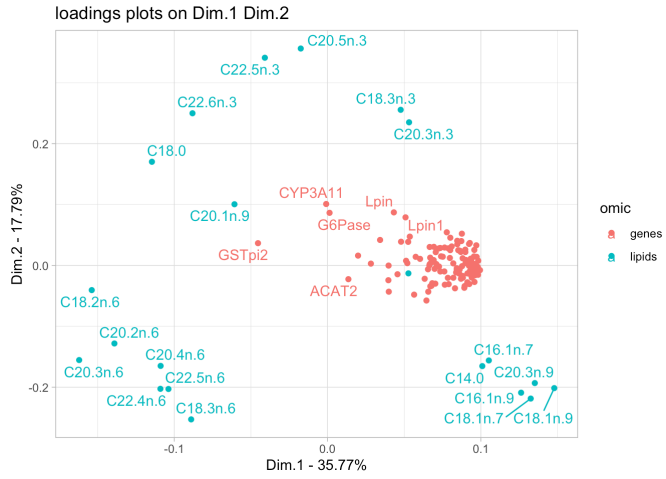
HIDE

```
ggplot(ppar_scores, aes(x=Dim.3, y=Dim.4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.3 - ", ppar_ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", ppar_ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "scores plots on Dim.3 Dim.4") +
  theme_light()
```



HIDE

```
# loadings plots
ppar_loadings <- data.frame(ppar_ComDim_res$Load.g)
ppar_loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
ppar_loadings$variable <- rownames(ppar_loadings)
ggplot(ppar_loadings, aes(x=Dim.1, y=Dim.2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", ppar_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ppar_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```



HIDE

```
ggplot(ppar_loadings, aes(x=Dim.3, y=Dim.4, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.3 - ", ppar_ComDim_res$cumexplained[3,"%explX"], "%"),
       y=paste0("Dim.4 - ", ppar_ComDim_res$cumexplained[4,"%explX"], "%"),
       title = "loadings plots on Dim.3 Dim.4") +
  theme_light()
```