

# Unsupervised Multiblock analyses

CODE ▾

Florence Mehl

February 21, 2023

- Concatenated PCA
- Statis
- MFA
- MCOA
- ComDim analysis of all samples
- ComDim analysis of wt samples
- ComDim analysis of ppar samples

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:

gene: 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;

lipid: 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 PPAR $\alpha$  -/- (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
```

## Concatenated PCA

HIDE

```

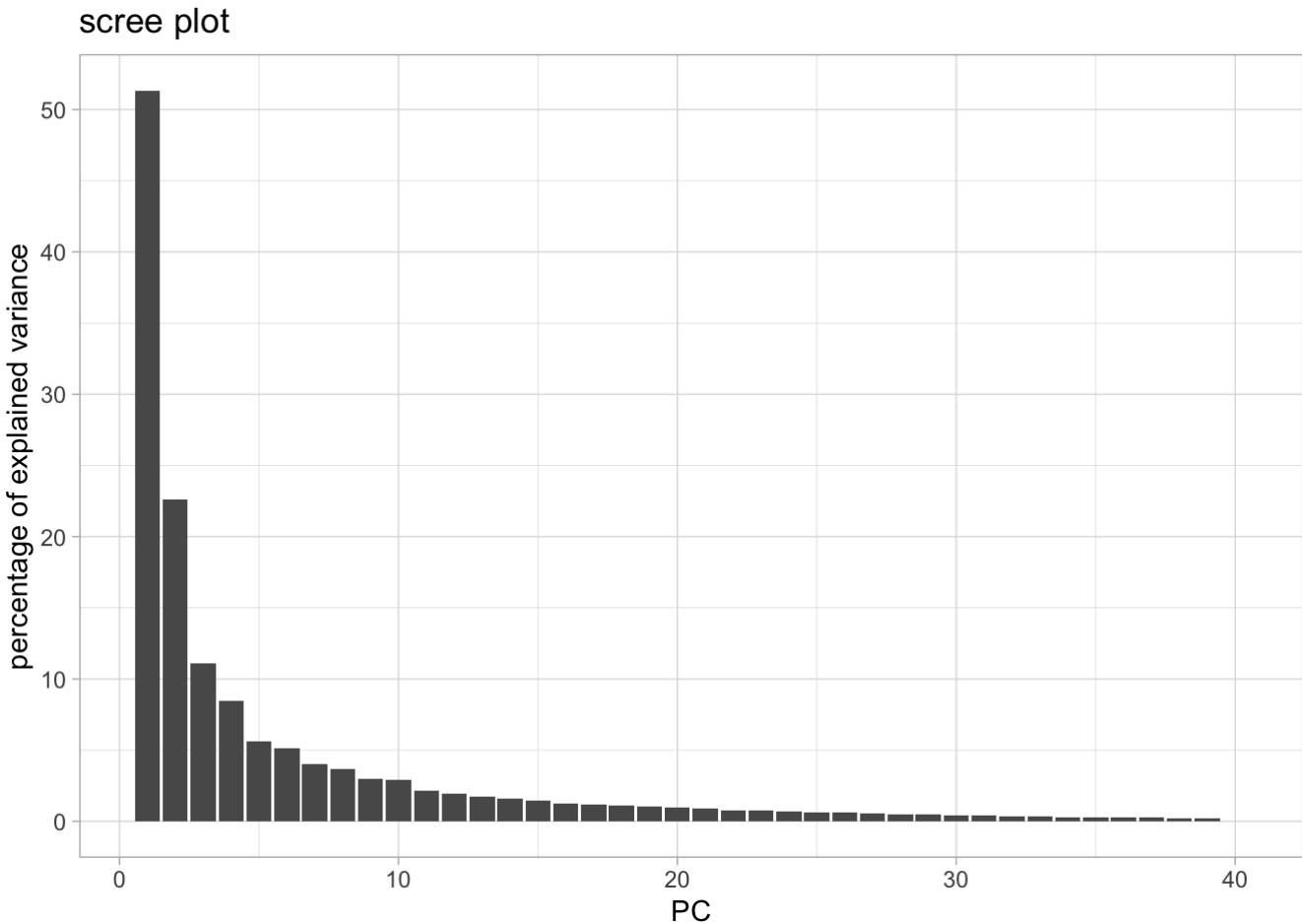
# prepare dataset
concatPCA_data <- cbind.data.frame(genes, lipids)

# run analysis
concatPCA_res <- prcomp(concatPCA_data, center=TRUE, scale.=TRUE)

# plot of variances: scree plot
concatPCA_variances <- data.frame("percentage_explX" = concatPCA_res$sdev^2)
concatPCA_variances$PC <- as.numeric(rownames(concatPCA_variances))

ggplot(concatPCA_variances, aes(x=PC, y=percentage_explX)) +
  geom_bar(stat = "identity") +
  labs(x = "PC", y = "percentage of explained variance", title = "scree plot") +
  theme_light()

```



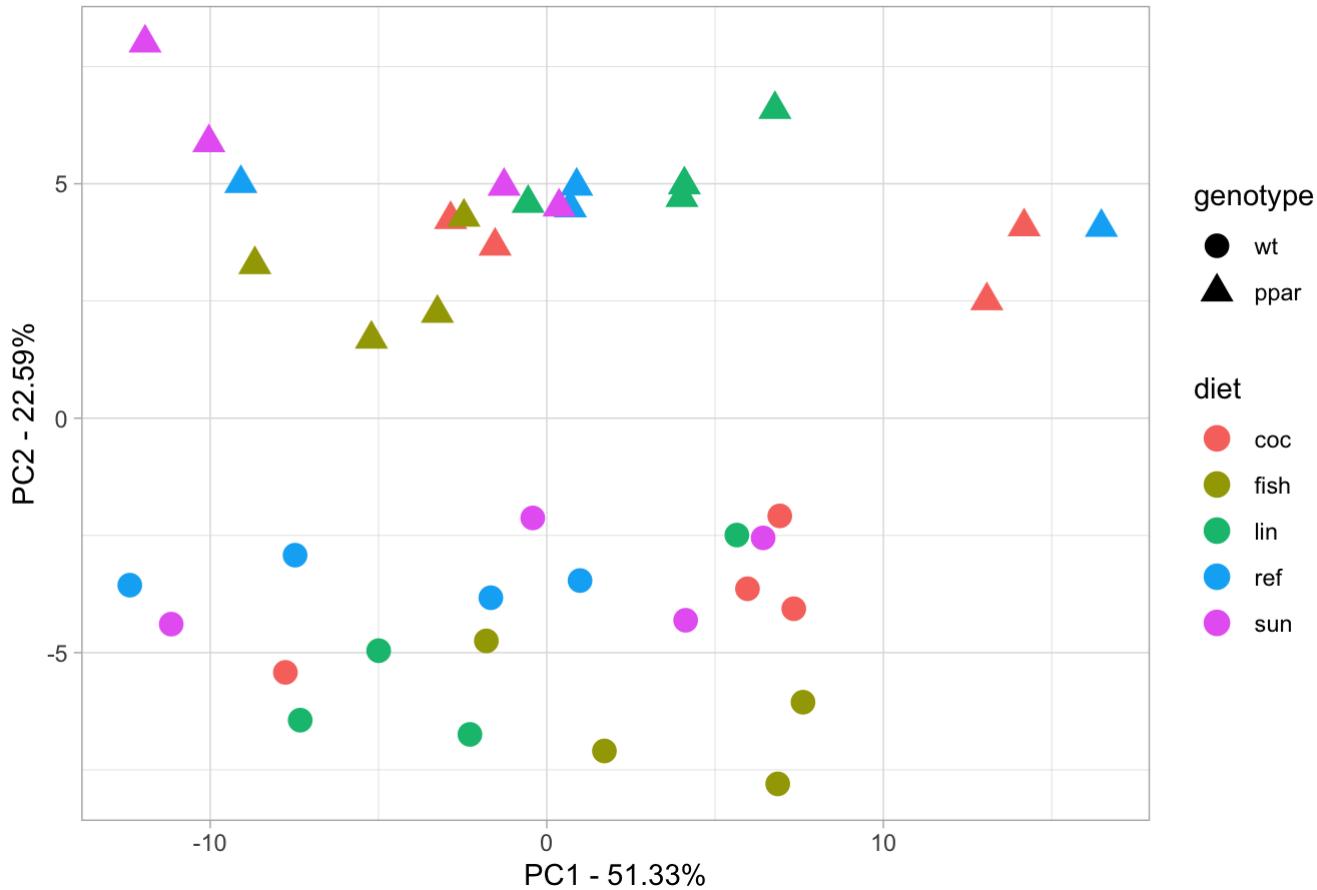
HIDE

```
# scores plots

concatPCA_scores <- data.frame(metadata, concatPCA_res$x)

ggplot(concatPCA_scores, aes(x=PC1, y=PC2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("PC1 - ", round(concatPCA_variances$percentage_explX[1], digits = 2), "%"),
       y=paste0("PC2 - ", round(concatPCA_variances$percentage_explX[2], digits = 2), "%"),
       title= "scores plot PC1 PC2") +
  theme_light()
```

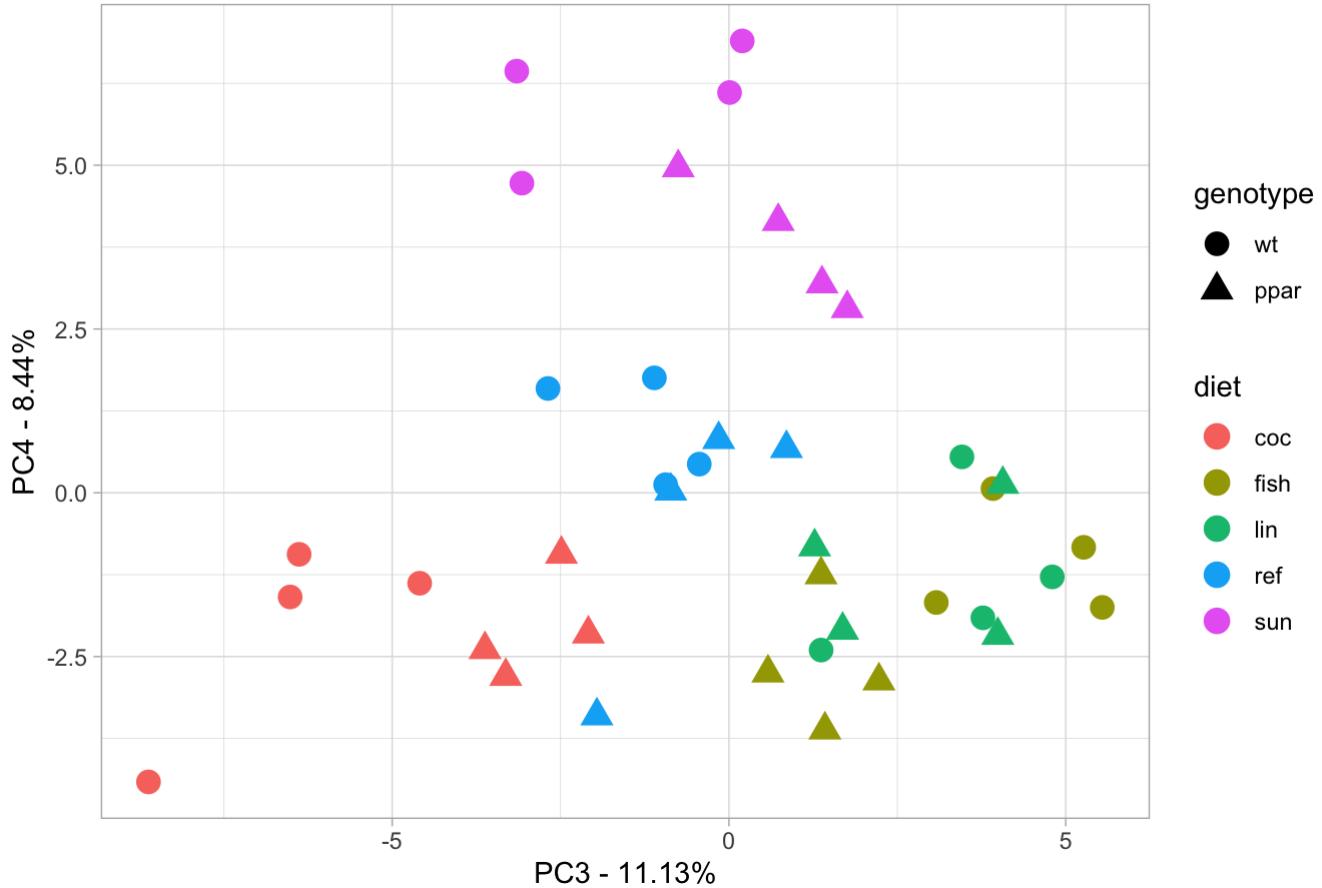
scores plot PC1 PC2



HIDE

```
ggplot(concatPCA_scores, aes(x=PC3, y=PC4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("PC3 - ", round(concatPCA_variances$percentage_explX[3], digits = 2), "%"),
       y=paste0("PC4 - ", round(concatPCA_variances$percentage_explX[4], digits = 2), "%"),
       title= "scores plot PC3 PC4") +
  theme_light()
```

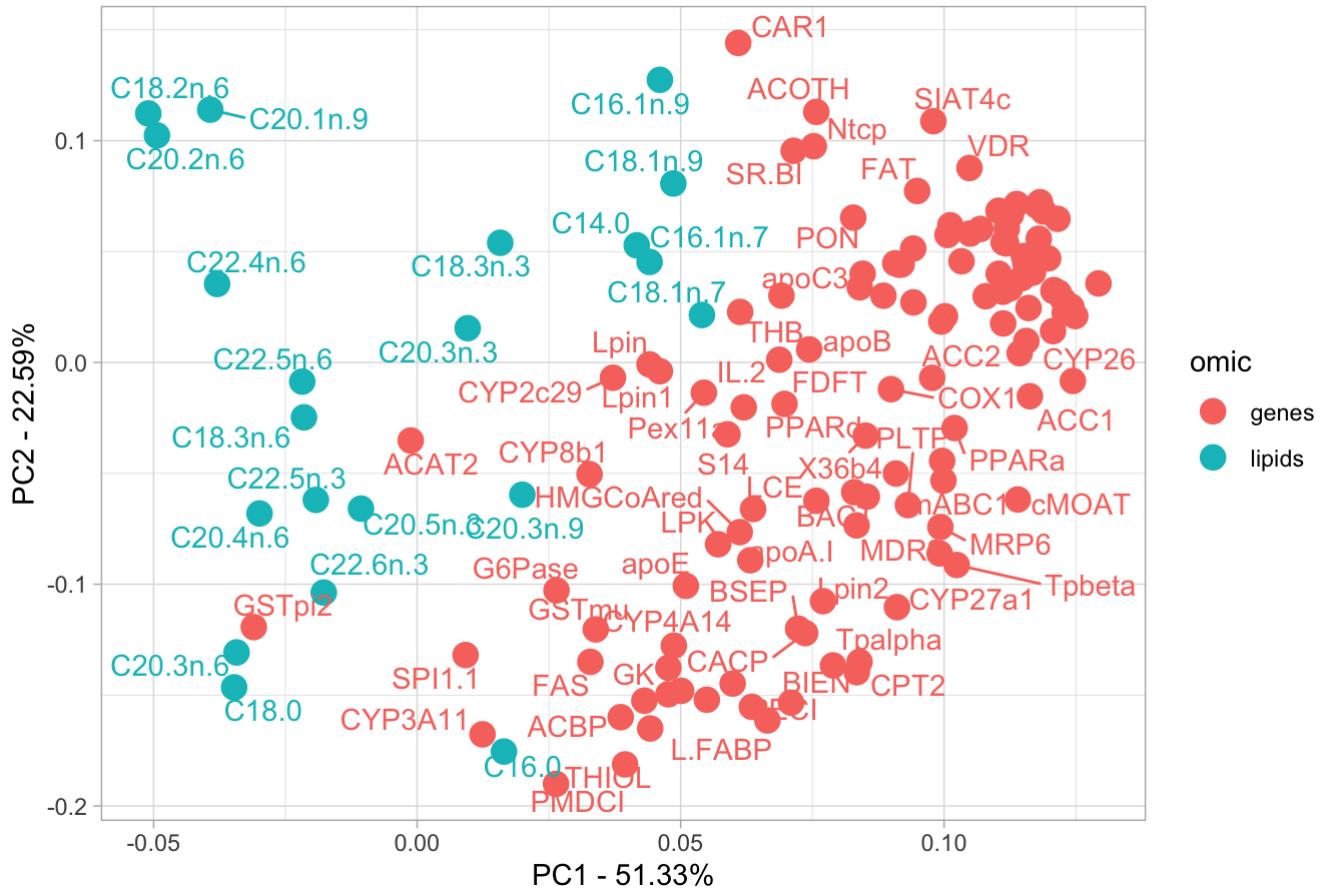
scores plot PC3 PC4



HIDE

```
# loadings plots
concatPCA_loadings <- data.frame(concatPCA_res$rotation)
concatPCA_loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
concatPCA_loadings$variable <- rownames(concatPCA_loadings)
ggplot(concatPCA_loadings, aes(x=PC1, y=PC2, col=omic, label=variable)) +
  geom_point(size=4) +
  labs(x=paste0("PC1 - ", round(concatPCA_variances$percentage_explX[1], digits = 2), "%"),
       y=paste0("PC2 - ", round(concatPCA_variances$percentage_explX[2], digits = 2), "%"),
       title= "loadings plot PC1 PC2") +
  geom_text_repel() +
  theme_light()
```

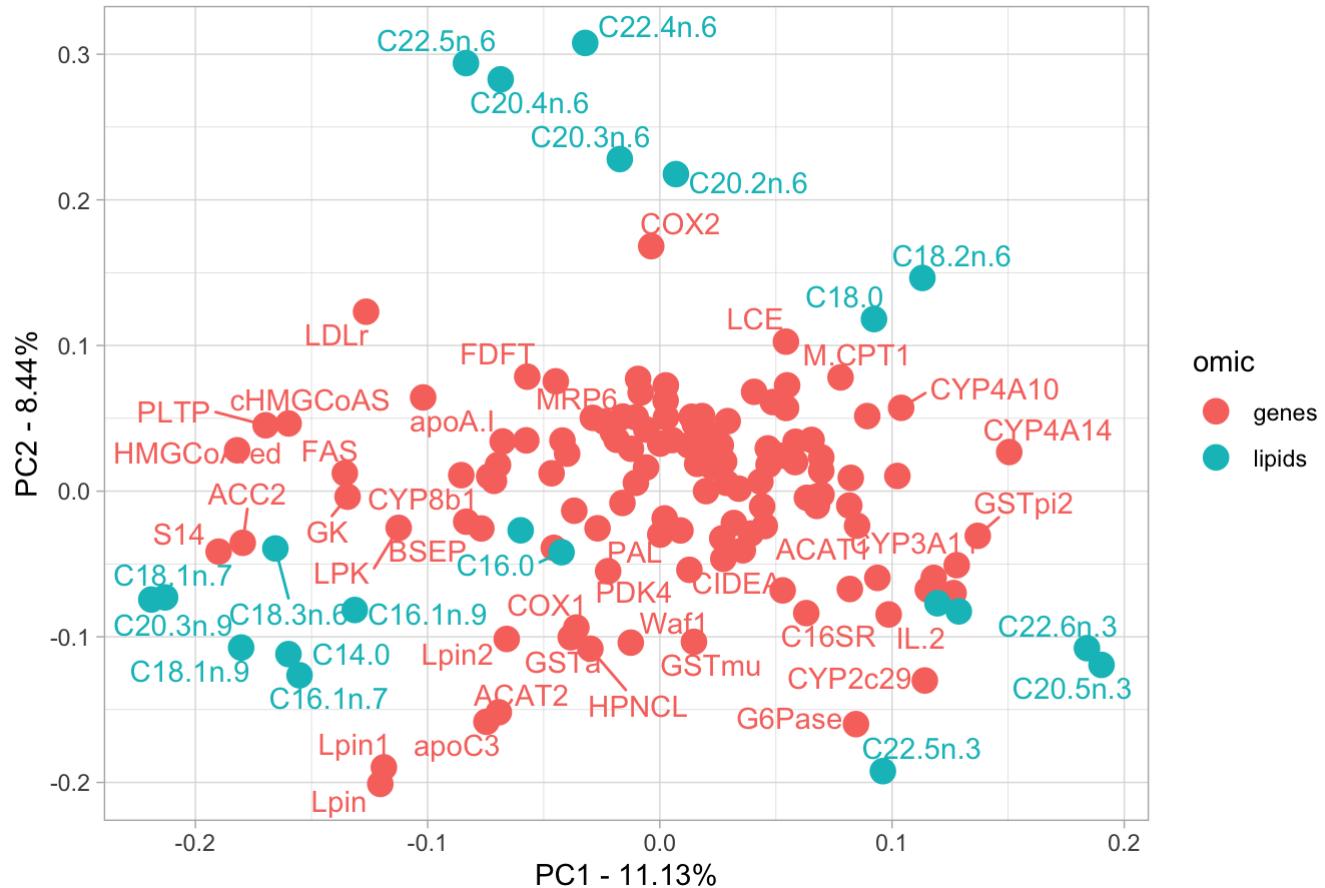
loadings plot PC1 PC2



HIDE

```
ggplot(concatPCA_loadings, aes(x=PC3, y=PC4, col=omic, label=variable)) +
  geom_point(size=4) +
  labs(x=paste0("PC1 - ", round(concatPCA_variances$percentage_explX[3], digits = 2), "%"),
       y=paste0("PC2 - ", round(concatPCA_variances$percentage_explX[4], digits = 2), "%"),
       title= "loadings plot PC3 PC4") +
  geom_text_repel() +
  theme_light()
```

loadings plot PC3 PC4



# Statis

HIDE

```

# statis_ktab <- ktab.list.df(statis_data, rownames = metadata$sample_name) # ????????????? r
ien compris
#
# # fonction statis ds package multiblock, dans le dataset candy les blocks sont des blocks
d'observations -> d'abord pca dans chaque block puis les résultats des pca sont pris comme
blocks? avec ems données ça joue seulement si je transpose les matrices mais les scores son
t les variables et les loadings les échantillons...
# statis <- function(X, ncomp = 3, scannf = FALSE, tol = 1e-07, ...){
#   X_frame <- as.data.frame(do.call(rbind, X))
#   X_factor <- factor(unlist(lapply(1:length(X), function(x)rep(x,nrow(X[[x]])))))
#   kta <- ktab.within(withinpca(X_frame, X_factor, scannf=scannf, nf=ncomp))
#   ret <- ade4:::statis(kta, scannf=scannf, nf=ncomp, tol=tol)
#   scores <- as.matrix(ret$C.Co); loadings <- as.matrix(ret$C.li)
#   blockScores <- list(); j <- 0
#   for(i in 1:length(X)){
#     blockScores[[i]] <- scores[j+(1:nrow(X[[i]])),,drop=FALSE]; j <- j+nrow(X[[i]])
#   }
#   names(blockScores) <- names(X)
#   colnames(scores) <- colnames(loadings) <- paste0('Comp ', 1:ncomp)
#   blockScores <- colnamesList(blockScores, paste0('Comp ', 1:ncomp))
#   info <- list(method = "STATIS",
#                 scores = "Concatenated scores", loadings = "Loadings",
#                 blockScores = "Block-wise scores", blockLoadings = "Not used")
#   obj <- list(scores = scores, loadings = loadings, blockScores = blockScores,
#               info = info, statis = ret, call = match.call())
#   obj$data <- list(X = X)
#   class(obj) <- c("multiblock", "list")
#   return(obj)
#   # Has plot and print in ade4
#   # Clustatis in https://cran.r-project.org/web/packages/ClustBlock/ClustBlock.pdf
# }

# prepare dataset
statis_data <- list(genes = t(genes), lipids = t(lipids))

# run analysis
statis_res <- multiblock:::statis(statis_data, ncomp = 4, scannf = F)

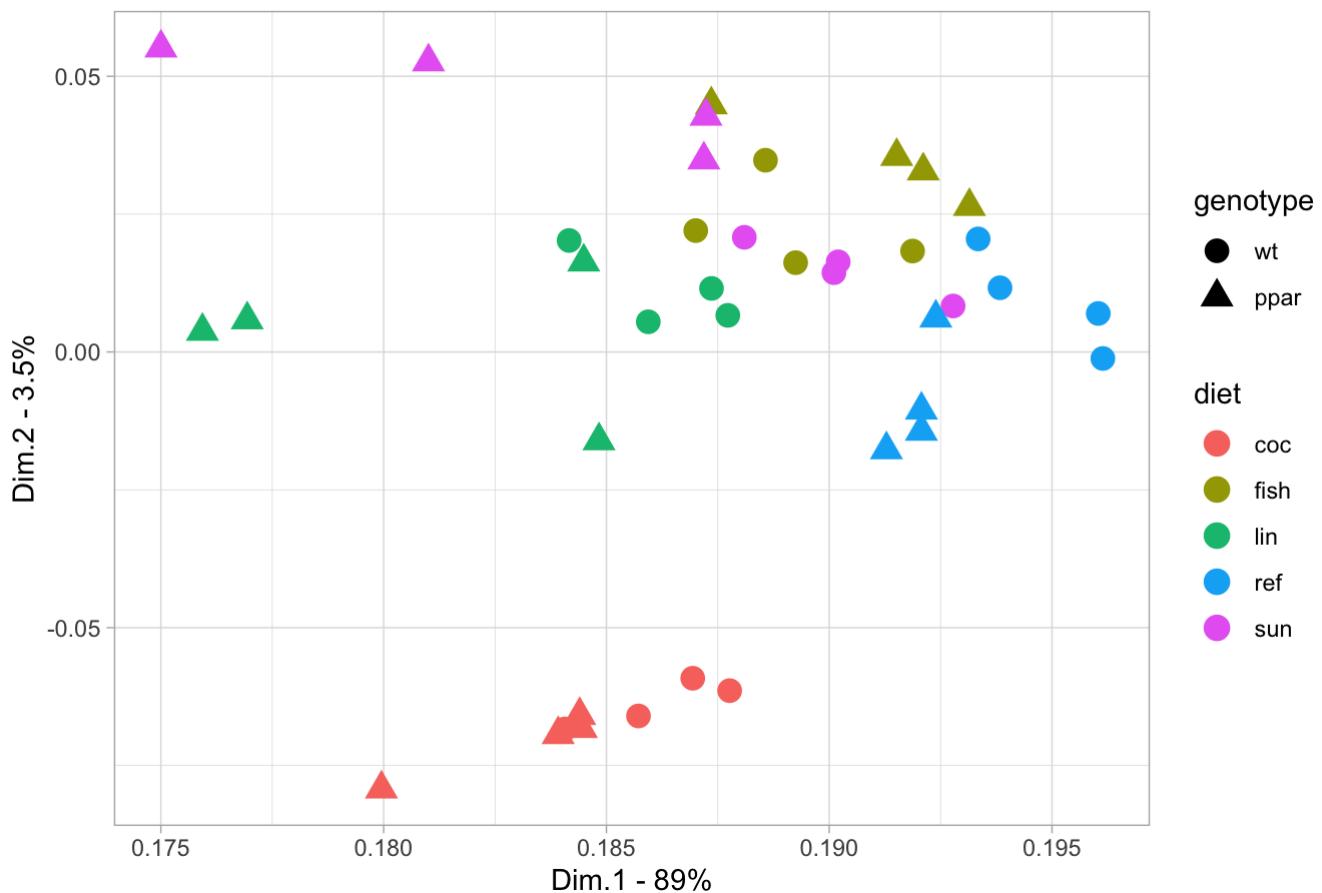
statis_res_explVar <- statis_res$statis$C.eig/sum(statis_res$statis$C.eig)

# scores plots
scores <- data.frame(metadata, statis_res$loadings)
ggplot(scores, aes(x=Comp.1, y=Comp.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", round(statis_res_explVar[1], digits=3)*100, "%"),
       y=paste0("Dim.2 - ", round(statis_res_explVar[2], digits=3)*100, "%"),

```

```
title = "scores plots on Dim.1 Dim.2") +
theme_light()
```

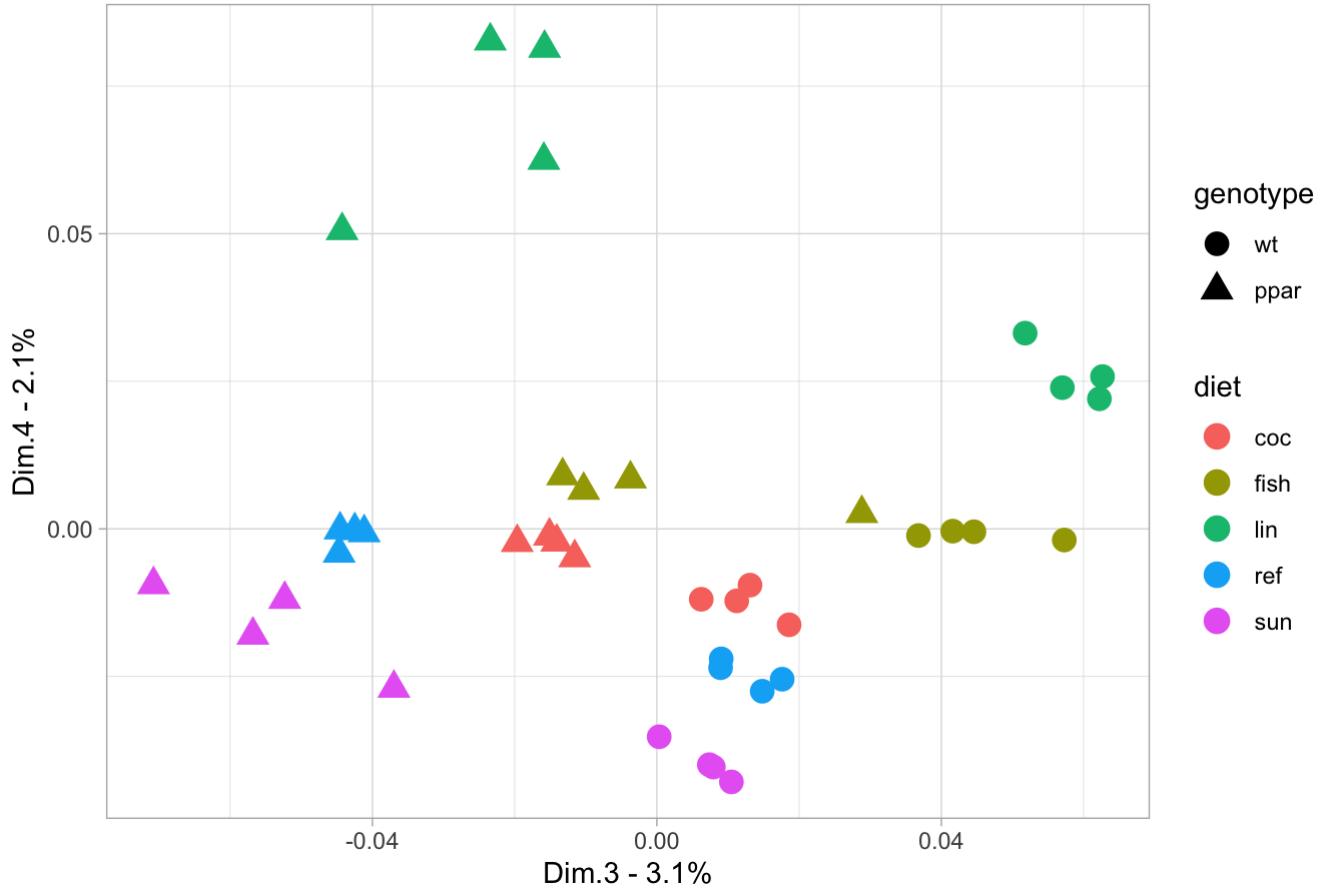
scores plots on Dim.1 Dim.2



HIDE

```
ggplot(scores, aes(x=Comp.3, y=Comp.4, col=diet, shape = genotype)) +
geom_point(size=4) +
labs(x=paste0("Dim.3 - ", round(statis_res_explVar[3], digits=3)*100, "%"),
y=paste0("Dim.4 - ", round(statis_res_explVar[4], digits=3)*100, "%"),
title = "scores plots on Dim.3 Dim.4") +
theme_light()
```

### scores plots on Dim.3 Dim.4



## MFA

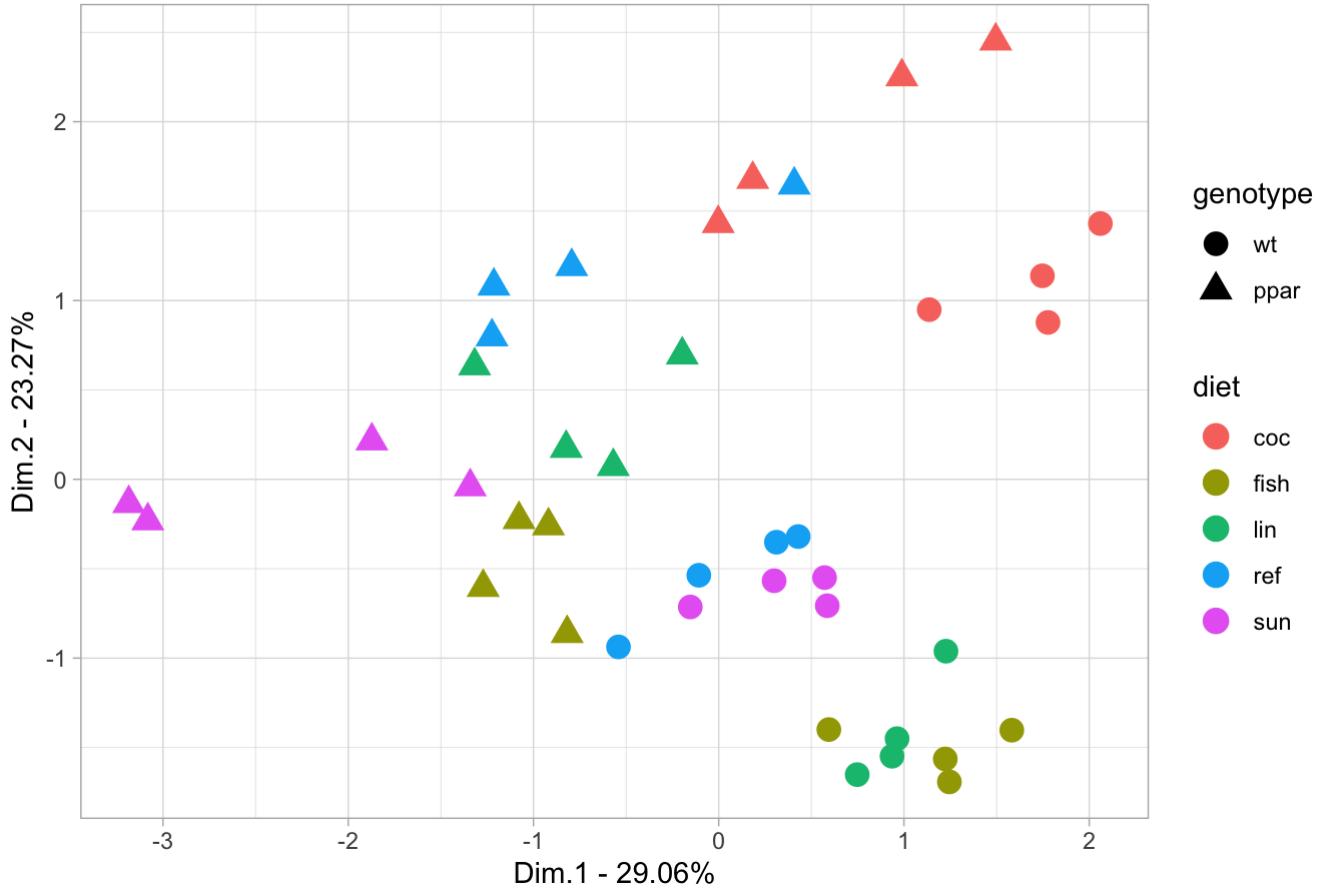
HIDE

```
# prepare dataset
mfa_data <- list(genes = genes, lipids = lipids)

# run analysis
mfa_res <- multiblock::mfa(mfa_data)

# scores plots
scores <- data.frame(metadata, mfa_res$scores)
ggplot(scores, aes(x=Comp.1, y=Comp.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", round(mfa_res$MFA$eig[1, "percentage of variance"], digits=2),
  "%"),
  y=paste0("Dim.2 - ", round(mfa_res$MFA$eig[2, "percentage of variance"], digits=2),
  "%"),
  title = "scores plots on Dim.1 Dim.2") +
  theme_light()
```

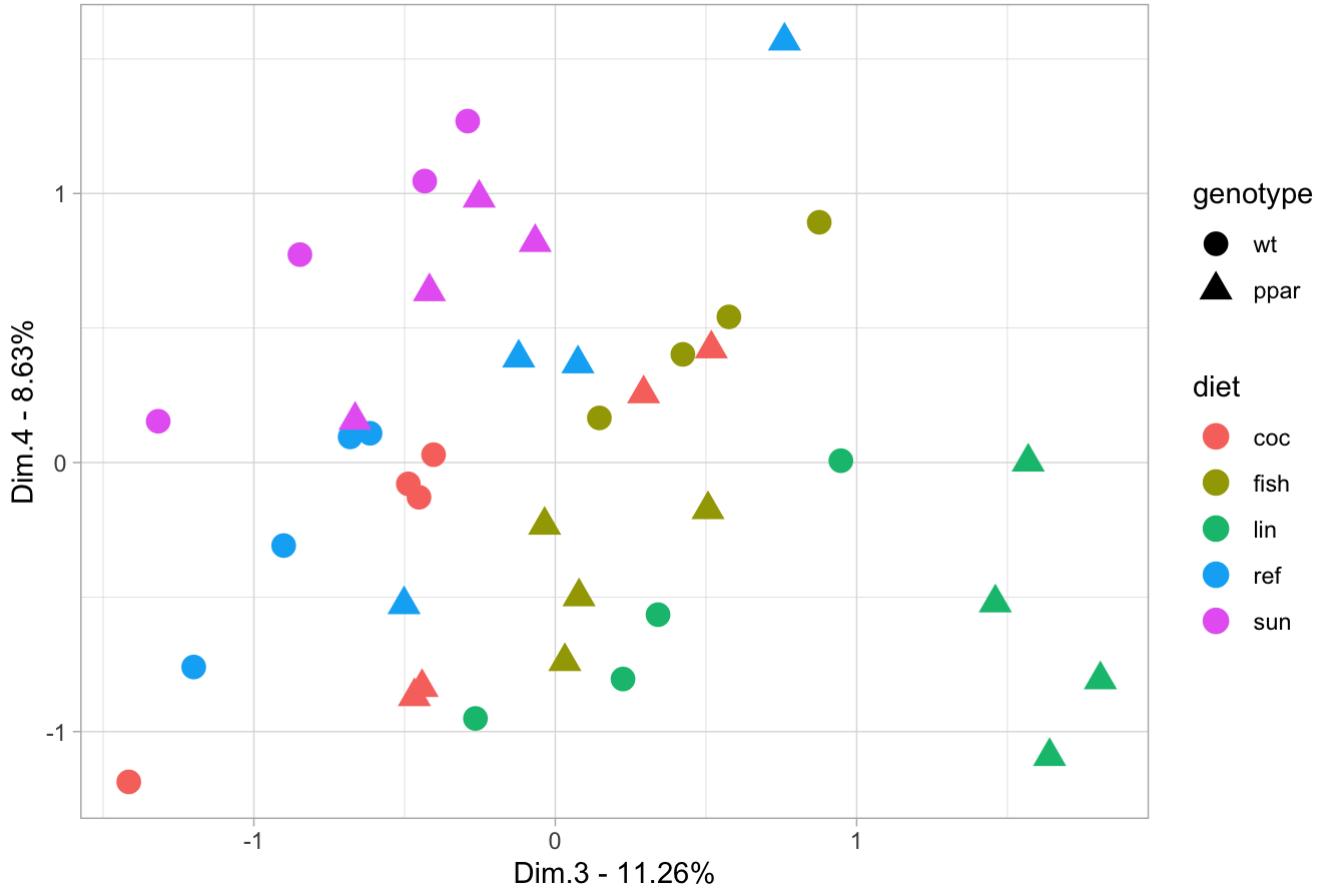
### scores plots on Dim.1 Dim.2



HIDE

```
ggplot(scores, aes(x=Comp.3, y=Comp.4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.3 - ", round(mfa_res$MFA$eig[3, "percentage of variance"], digits=2),
  "%"),
  y=paste0("Dim.4 - ", round(mfa_res$MFA$eig[4, "percentage of variance"], digits=2),
  "%"),
  title = "scores plots on Dim.3 Dim.4") +
  theme_light()
```

### scores plots on Dim.3 Dim.4



## MCOA

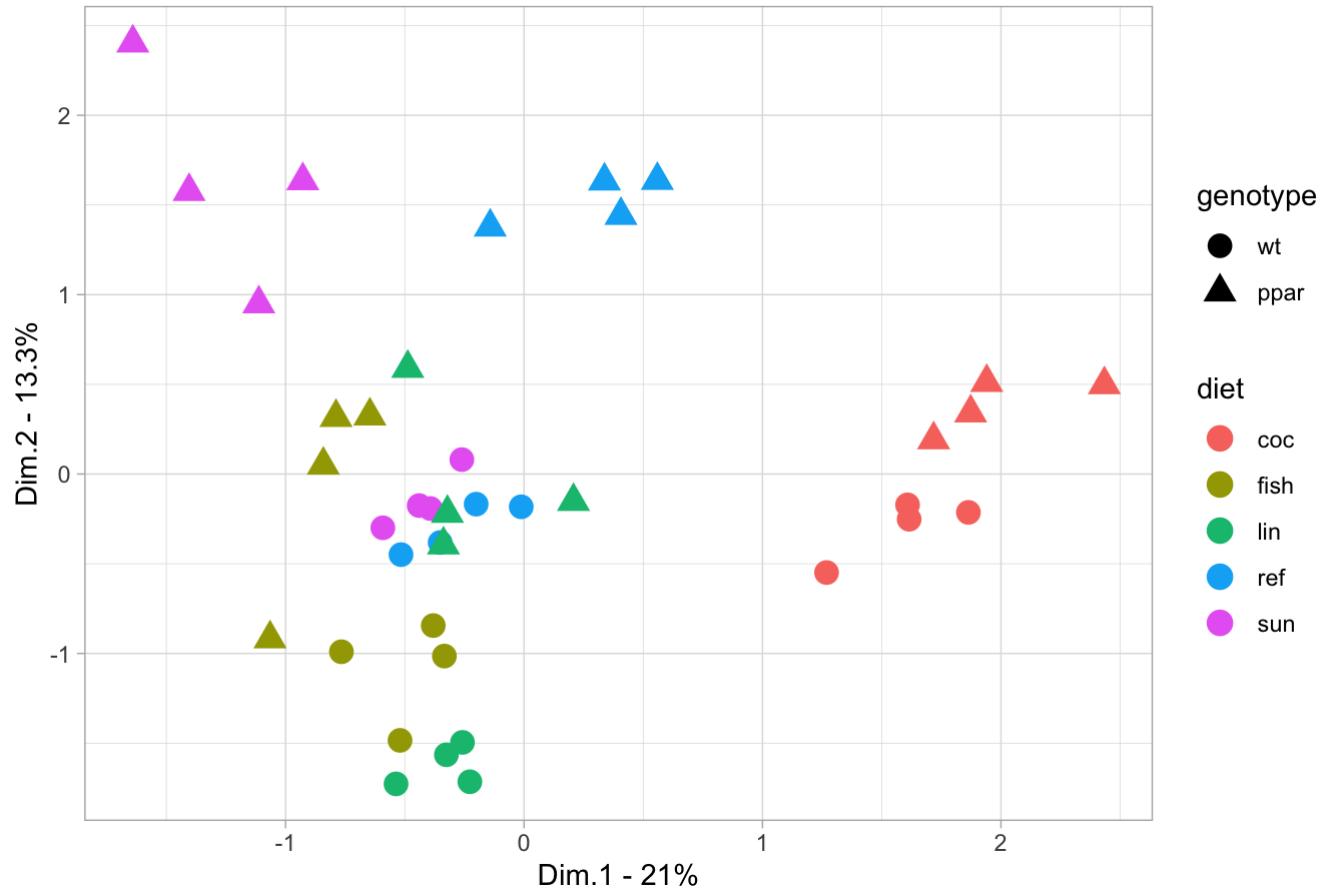
HIDE

```
# prepare dataset
mcoa_data <- list(genes = genes, lipids = lipids)

# run analysis
mcoa_res <- multiblock::mcoa(mcoa_data, ncomp = 4)

# scores plots
scores <- data.frame(metadata, mcoa_res$scores)
ggplot(scores, aes(x=Comp.1, y=Comp.2, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.1 - ", round(mcoa_res$rgcca$AVE$AVE_outer_model[1], digits = 3)*100,
            "%"),
       y=paste0("Dim.2 - ", round(mcoa_res$rgcca$AVE$AVE_outer_model[2], digits = 3)*100,
            "%"),
       title = "scores plots on Dim.1 Dim.2") +
  theme_light()
```

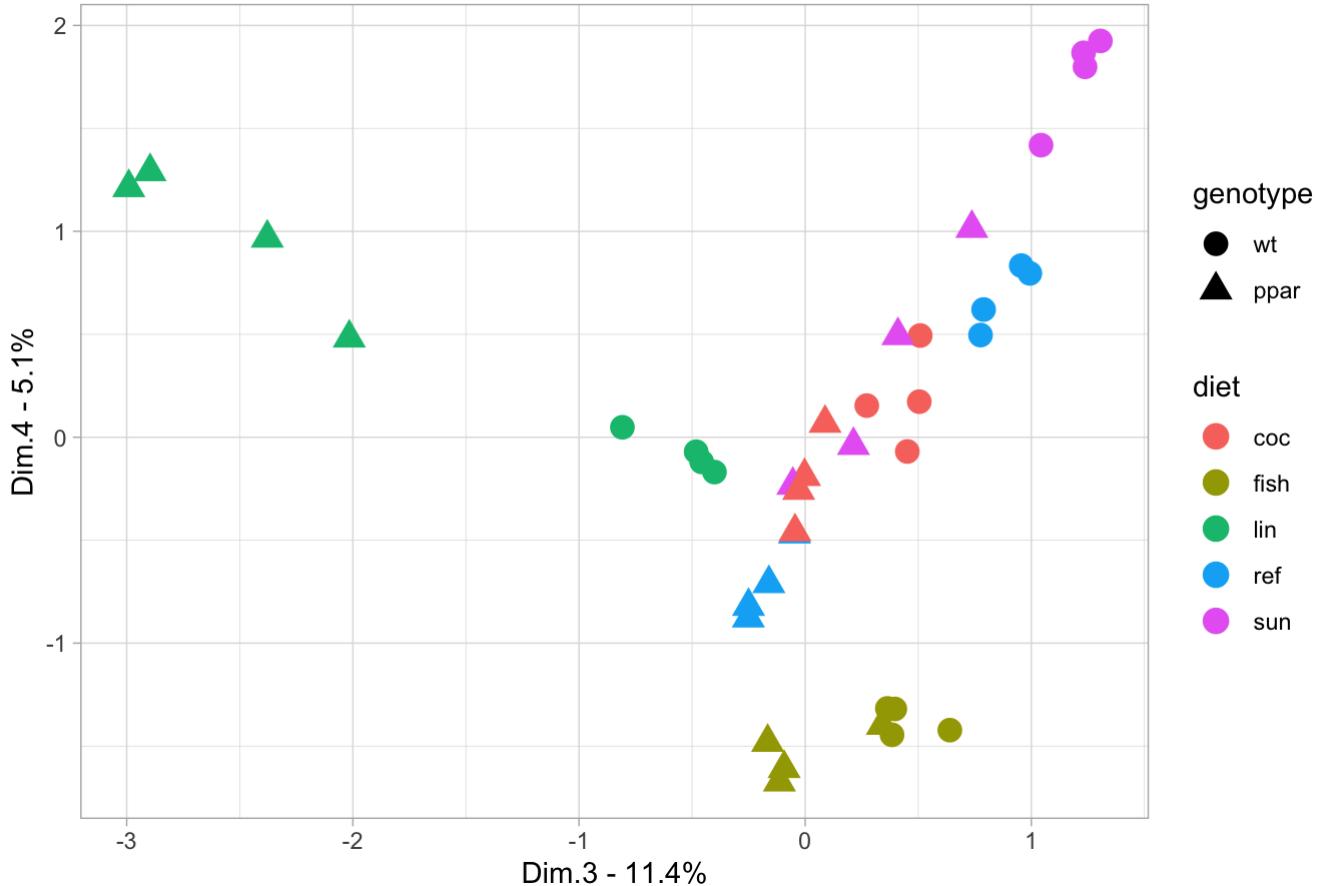
## scores plots on Dim.1 Dim.2



HIDE

```
ggplot(scores, aes(x=Comp.3, y=Comp.4, col=diet, shape = genotype)) +
  geom_point(size=4) +
  labs(x=paste0("Dim.3 - ", round(mcoa_res$rgcca$AVE$AVE_outer_model[3], digits = 3)*100,
  "%"),
       y=paste0("Dim.4 - ", round(mcoa_res$rgcca$AVE$AVE_outer_model[4], digits = 3)*100,
  "%"),
       title = "scores plots on Dim.3 Dim.4") +
  theme_light()
```

scores plots on Dim.3 Dim.4



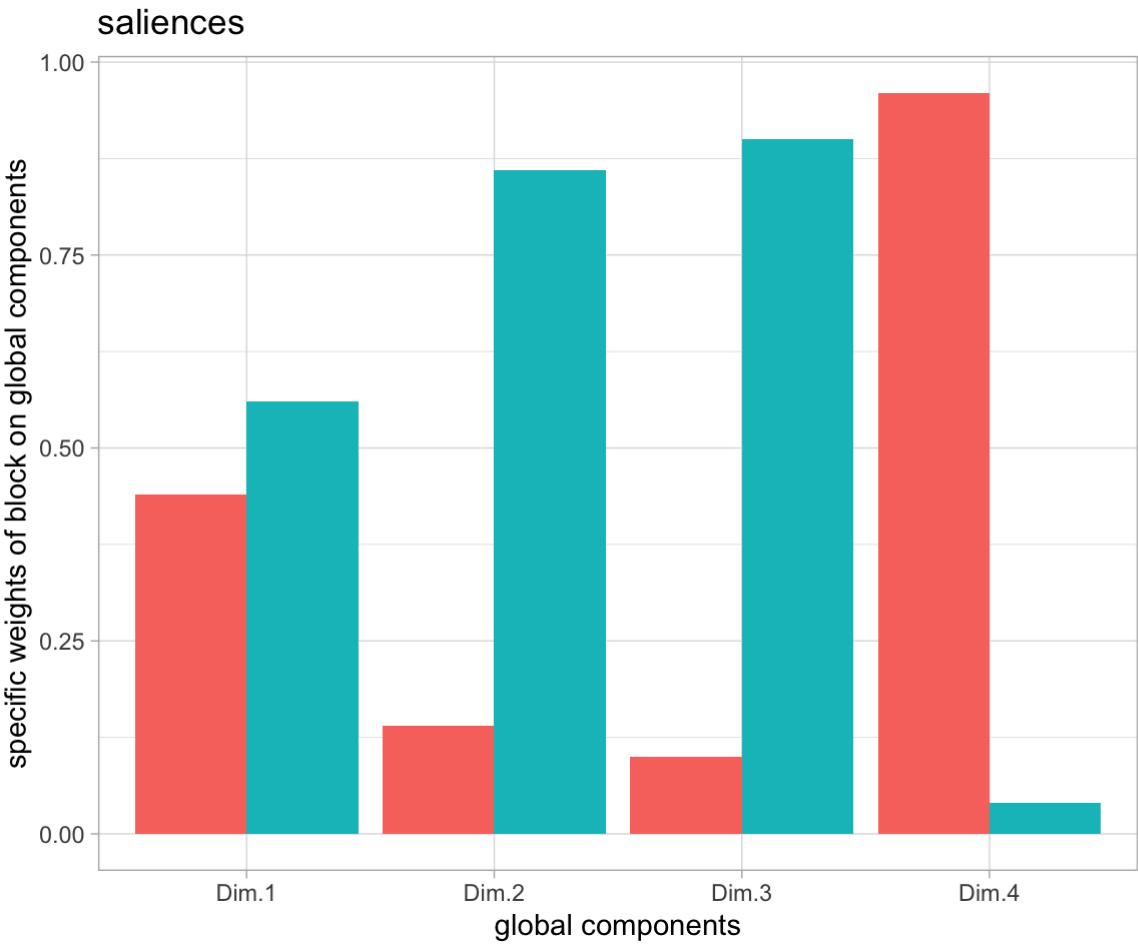
## ComDim analysis of all samples

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, group = n_group, option = "uniform", plotgraph = F)

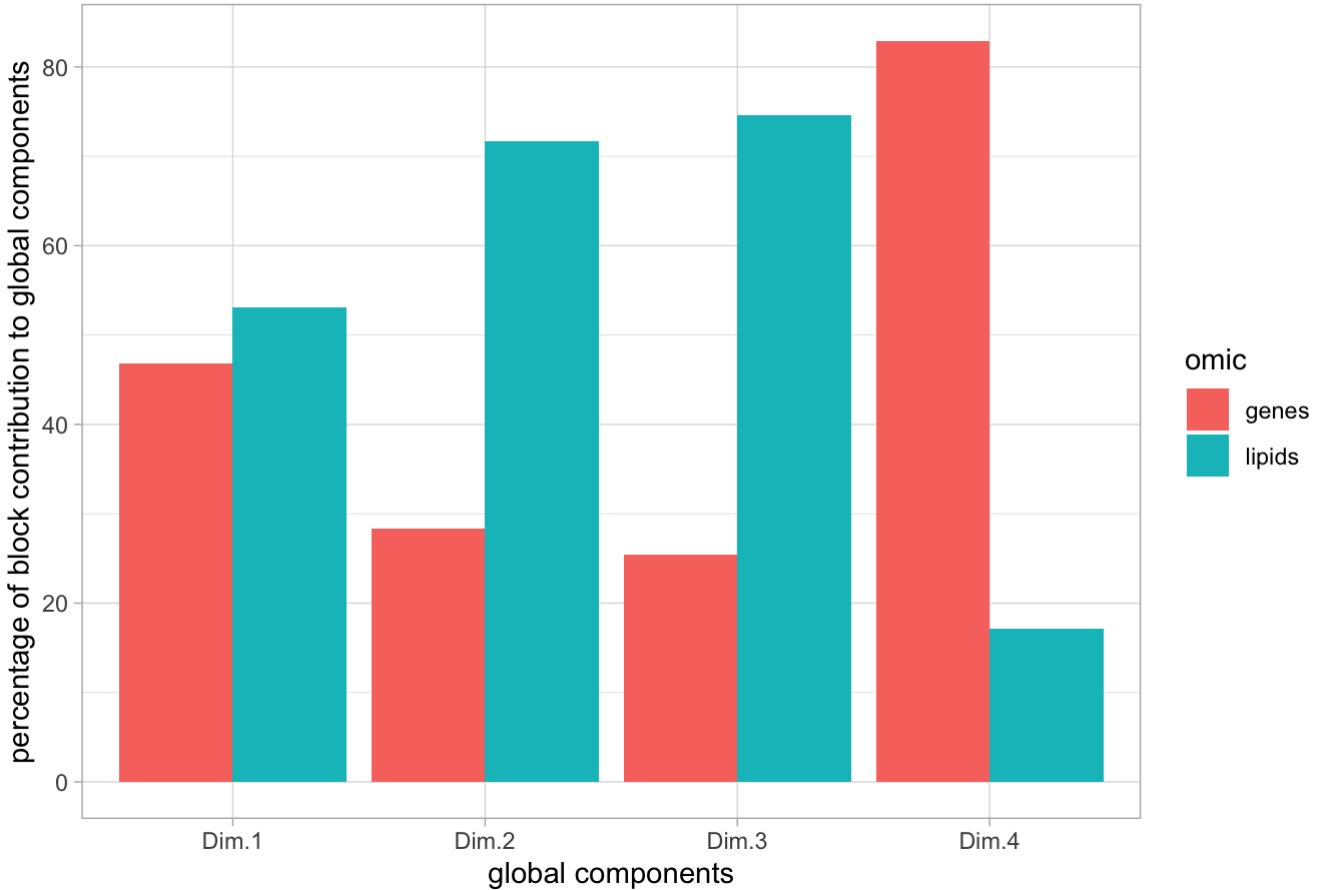
# saliences
saliences <- ComDim_res$saliences
rownames(saliences) <- c("genes", "lipids")
saliences <- as.data.frame(t(saliences[,1:4]))
saliences$Dim <- rownames(saliences)
saliences <- melt(saliences)
ggplot(saliences, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge())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



HIDE

```
# block contributions
contributions <- ComDim_res$contrib
rownames(contributions) <- c("genes", "lipids")
contributions <- as.data.frame(t(contributions[,1:4]))
contributions$Dim <- rownames(contributions)
contributions <- melt(contributions)
ggplot(contributions, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components", y = "percentage of block contribution to global component",
       fill = "omic",
       title = "blocks contributions")
```

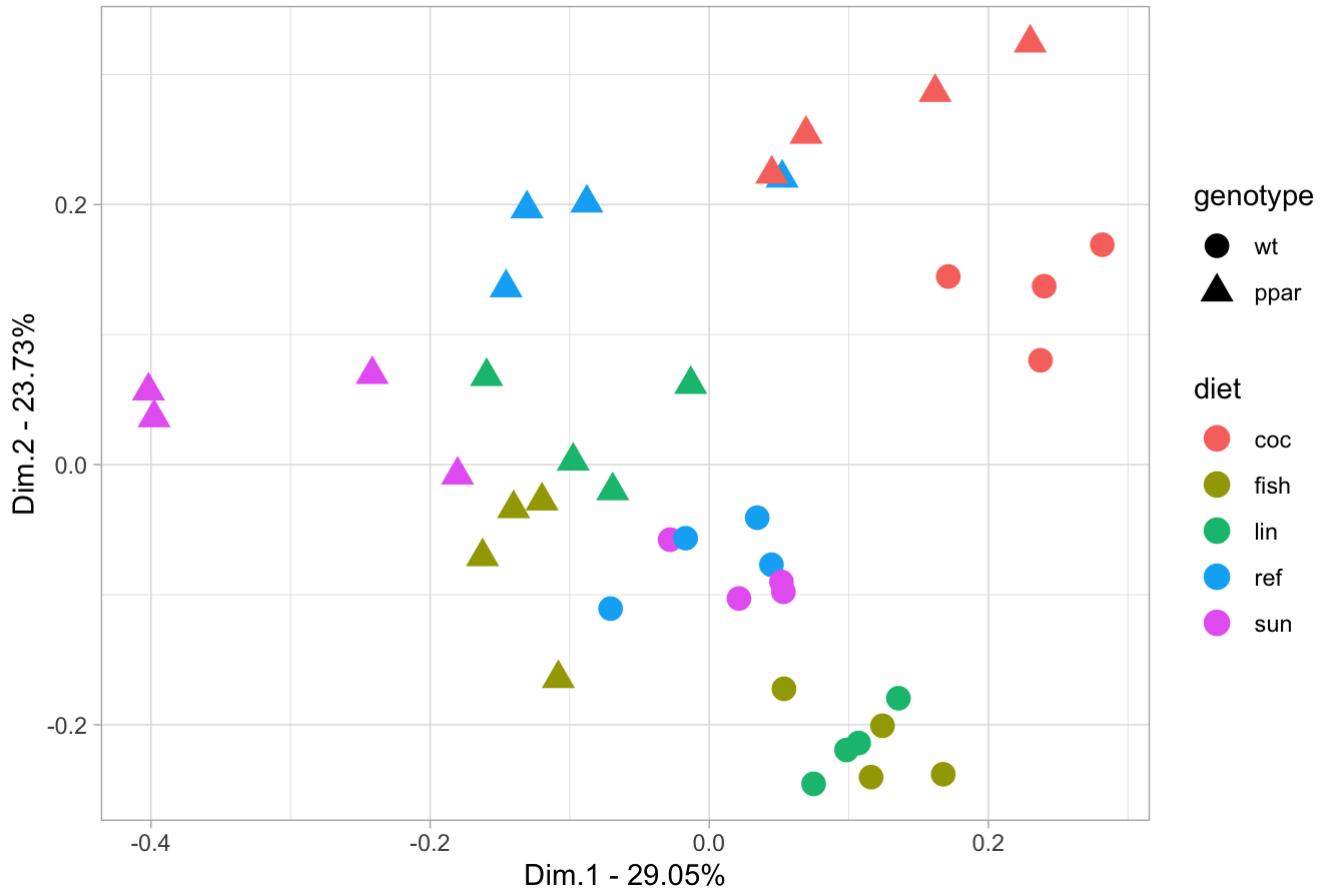
## blocks contributions



HIDE

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

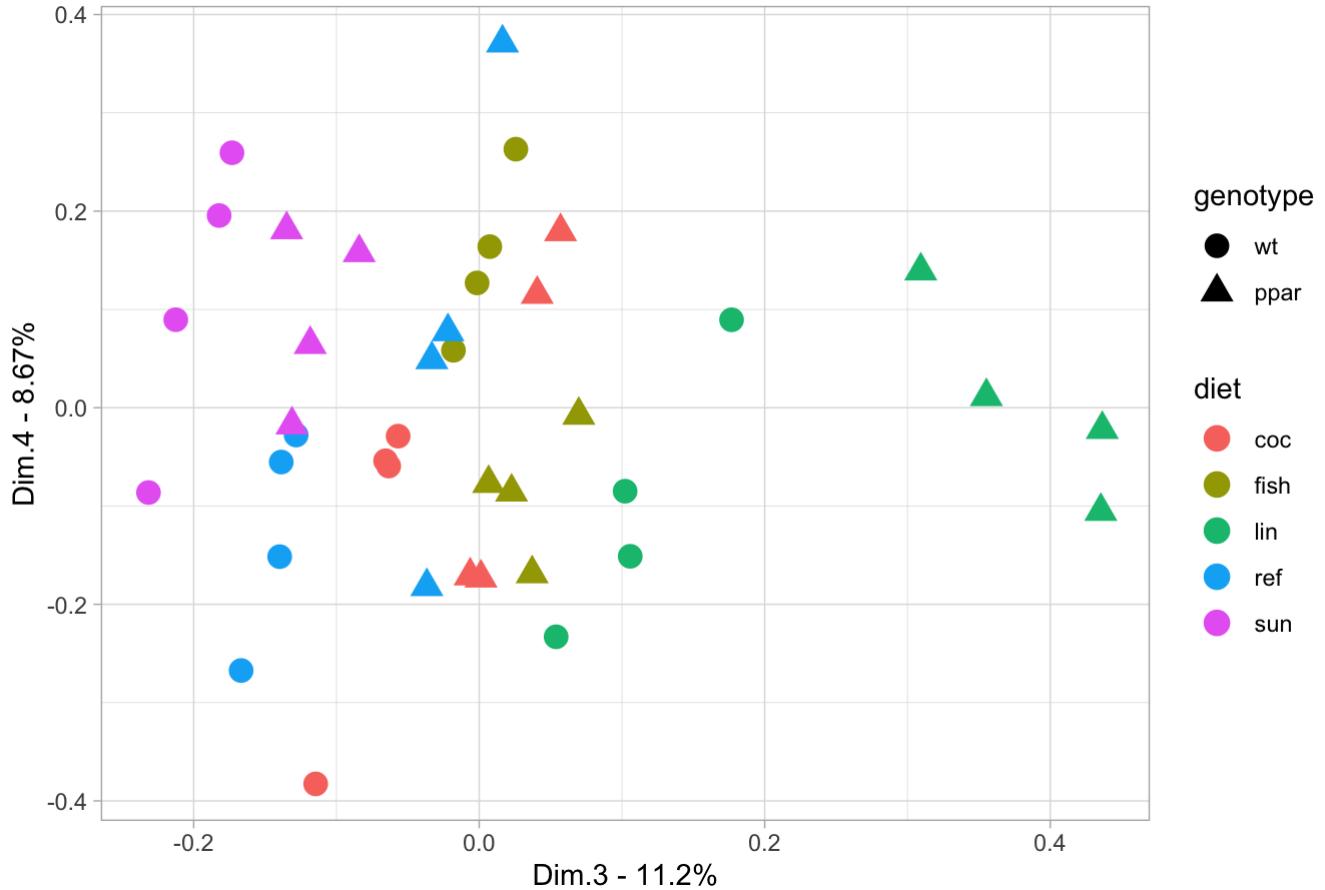
### scores plots on Dim.1 Dim.2



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()
```

### scores plots on Dim.3 Dim.4

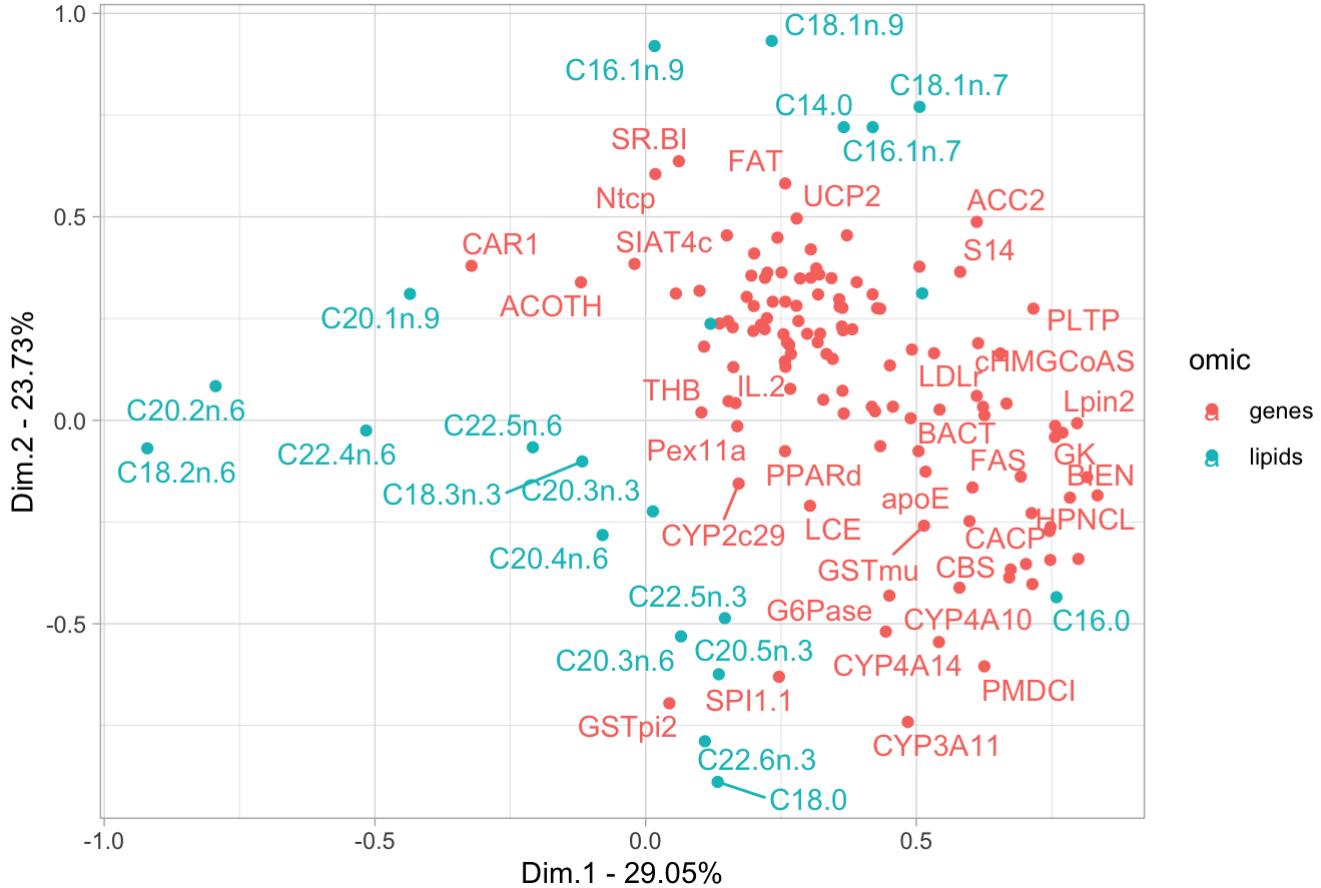


HIDE

```
# loadings plots

loadings <- data.frame(ComDim_res$globalcor)
loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
loadings$variable <- rownames(loadings)
ggplot(loadings, aes(x=X1, y=X2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```

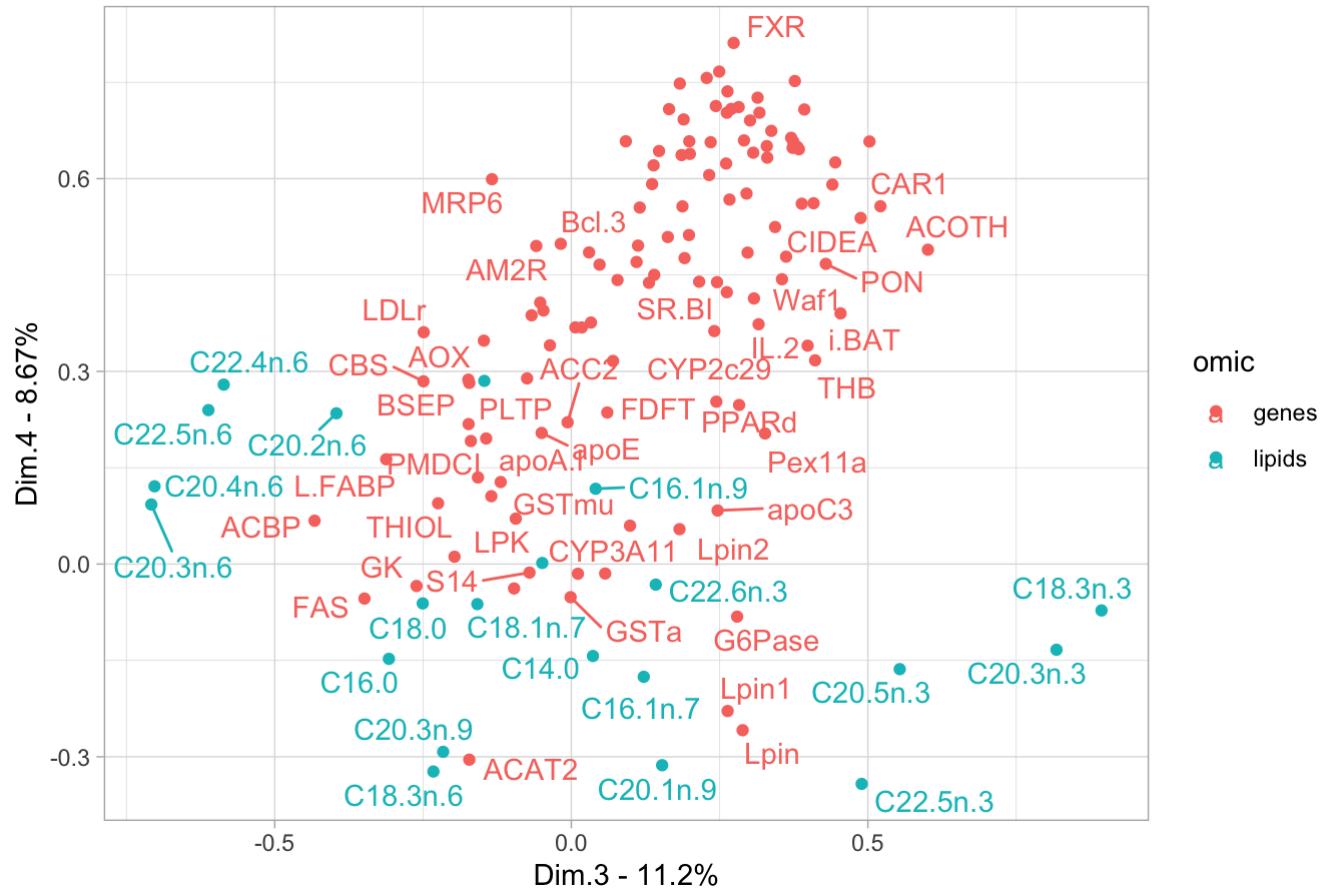
### loadings plots on Dim.1 Dim.2



HIDE

```
ggplot(loadings, aes(x=X3, y=X4, 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()
```

### loadings plots on Dim.3 Dim.4



## ComDim analysis of wt samples

HIDE

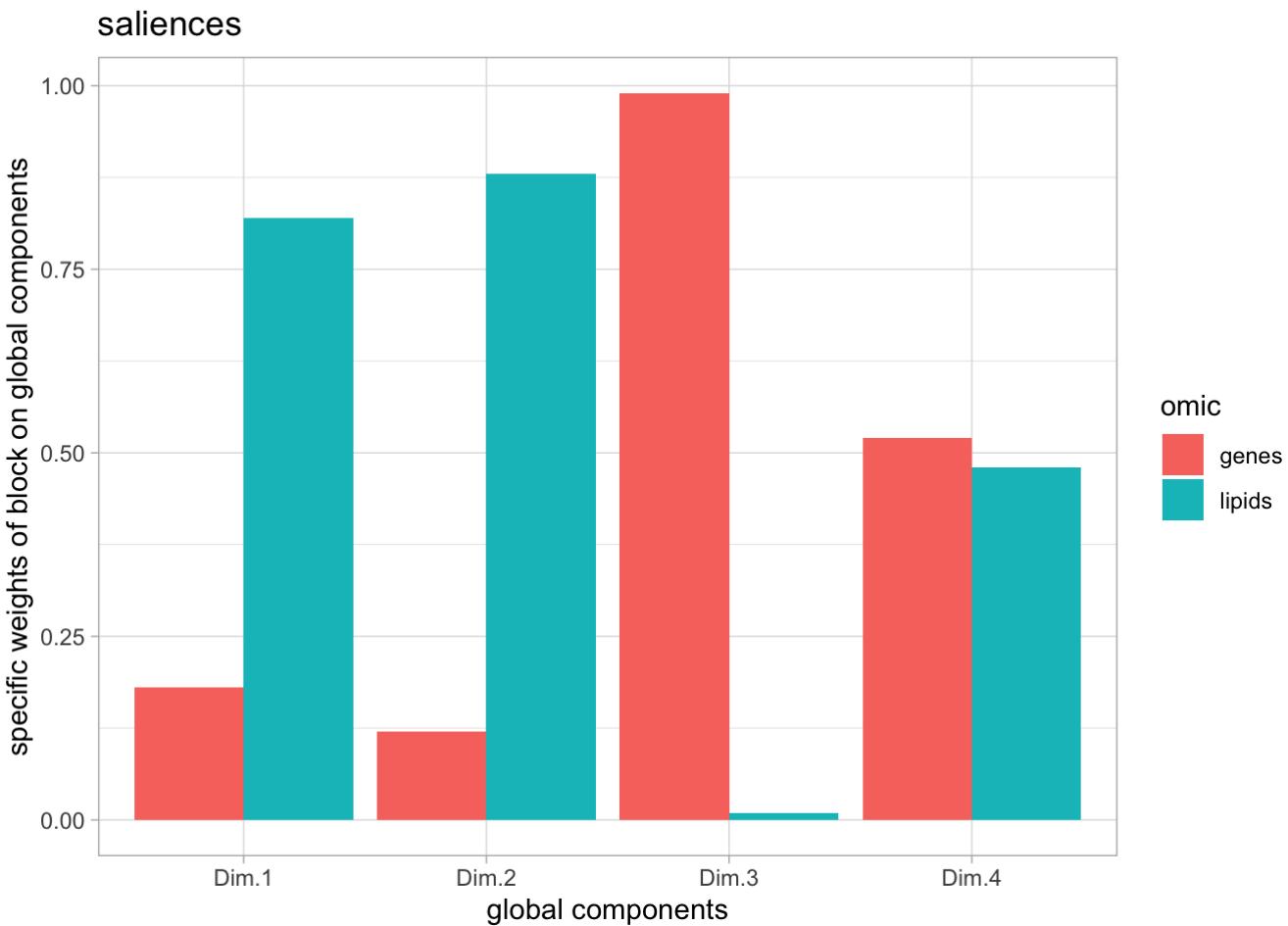
```

# 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, group = n_group, option = "uniform", plotgraph = F)

# 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")

```



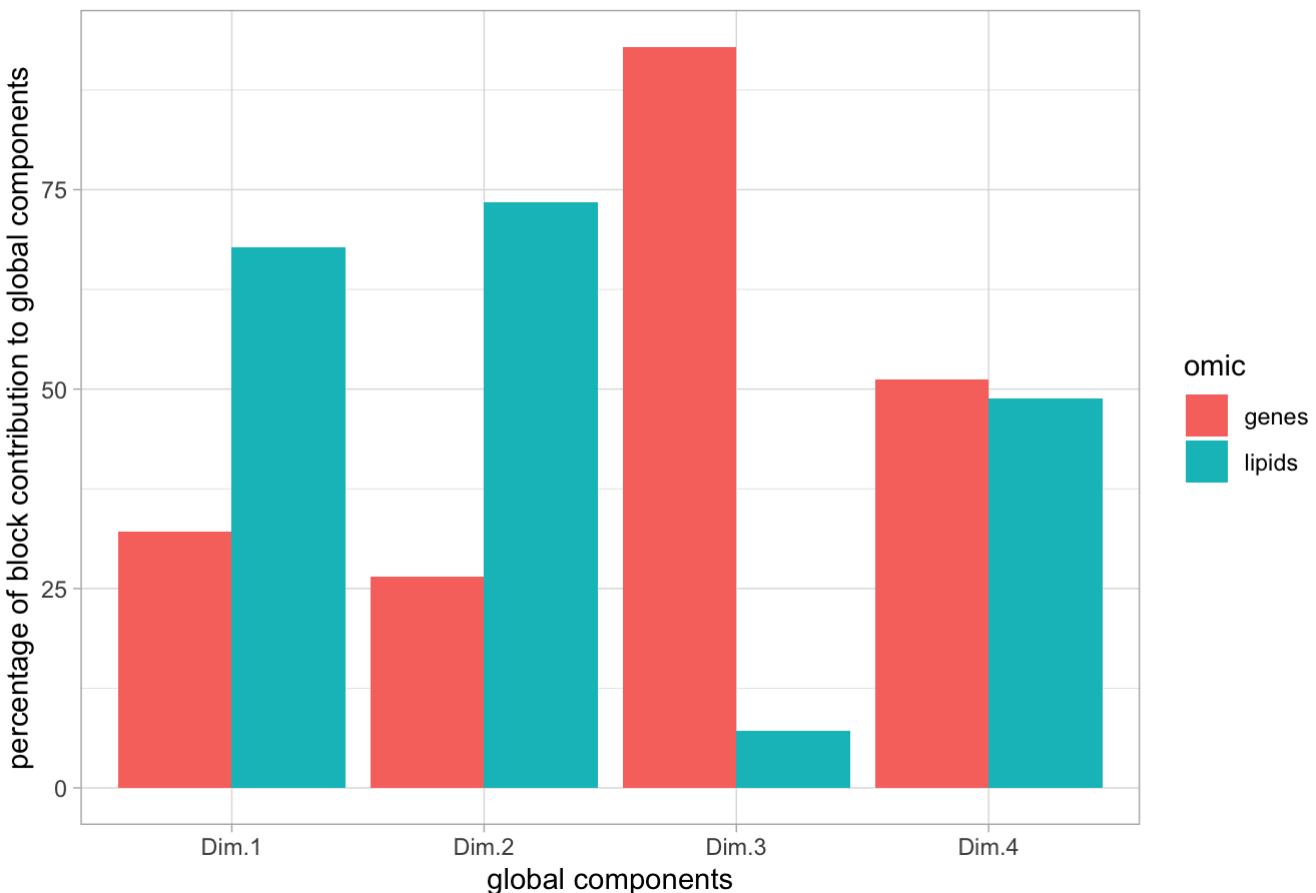
HIDE

```

# block contributions
wt_contributions <- wt_ComDim_res$contrib
rownames(wt_contributions) <- c("genes", "lipids")
wt_contributions <- as.data.frame(t(wt_contributions[,1:4]))
wt_contributions$Dim <- rownames(wt_contributions)
wt_contributions <- melt(wt_contributions)
ggplot(wt_contributions, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components", y = "percentage of block contribution to global component
s", fill = "omic",
       title = "blocks contributions")

```

blocks contributions



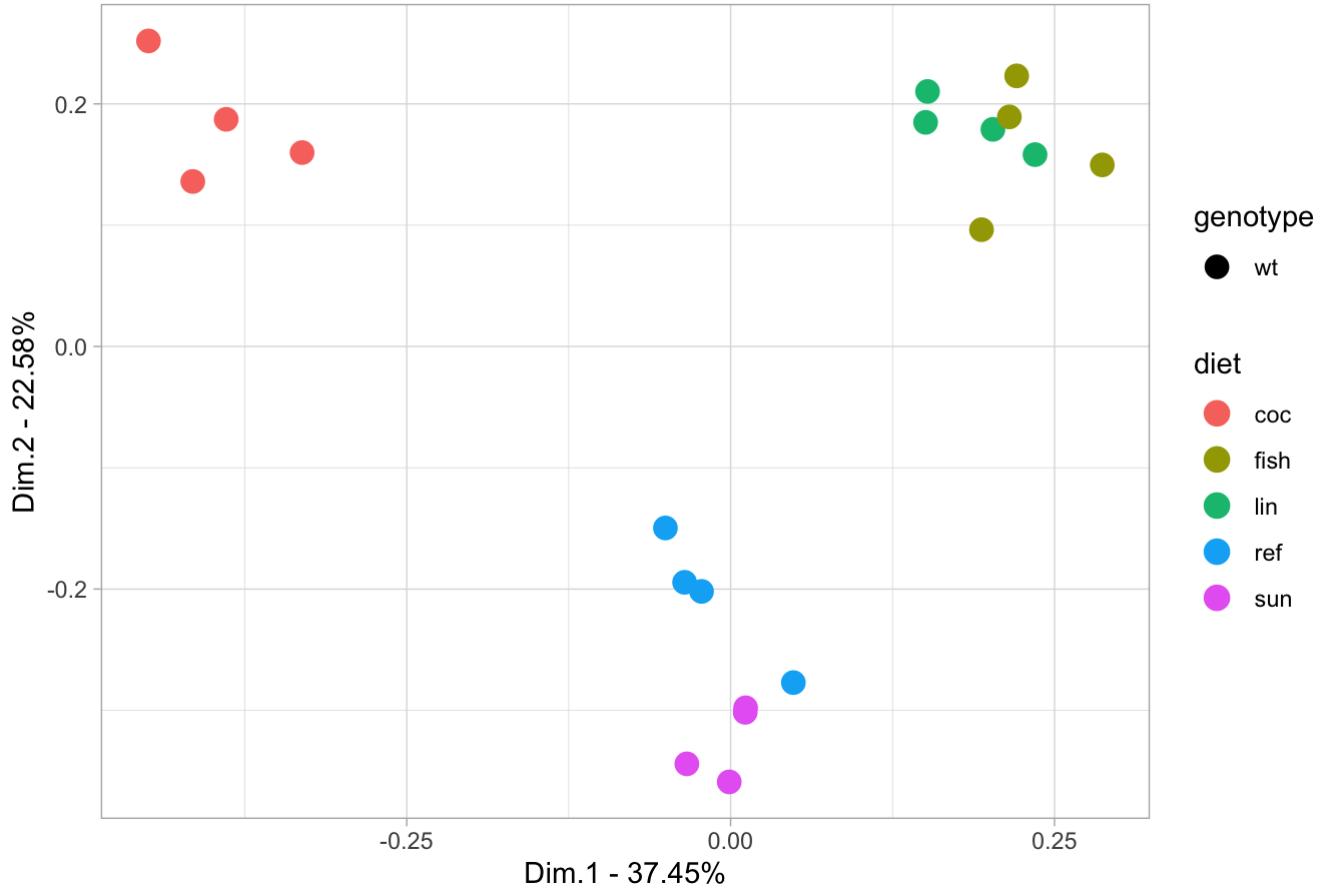
HIDE

```

# scores plots
wt_scores <- data.frame(metadata[metadata$genotype == "wt",], wt_ComDim_res$T)
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()

```

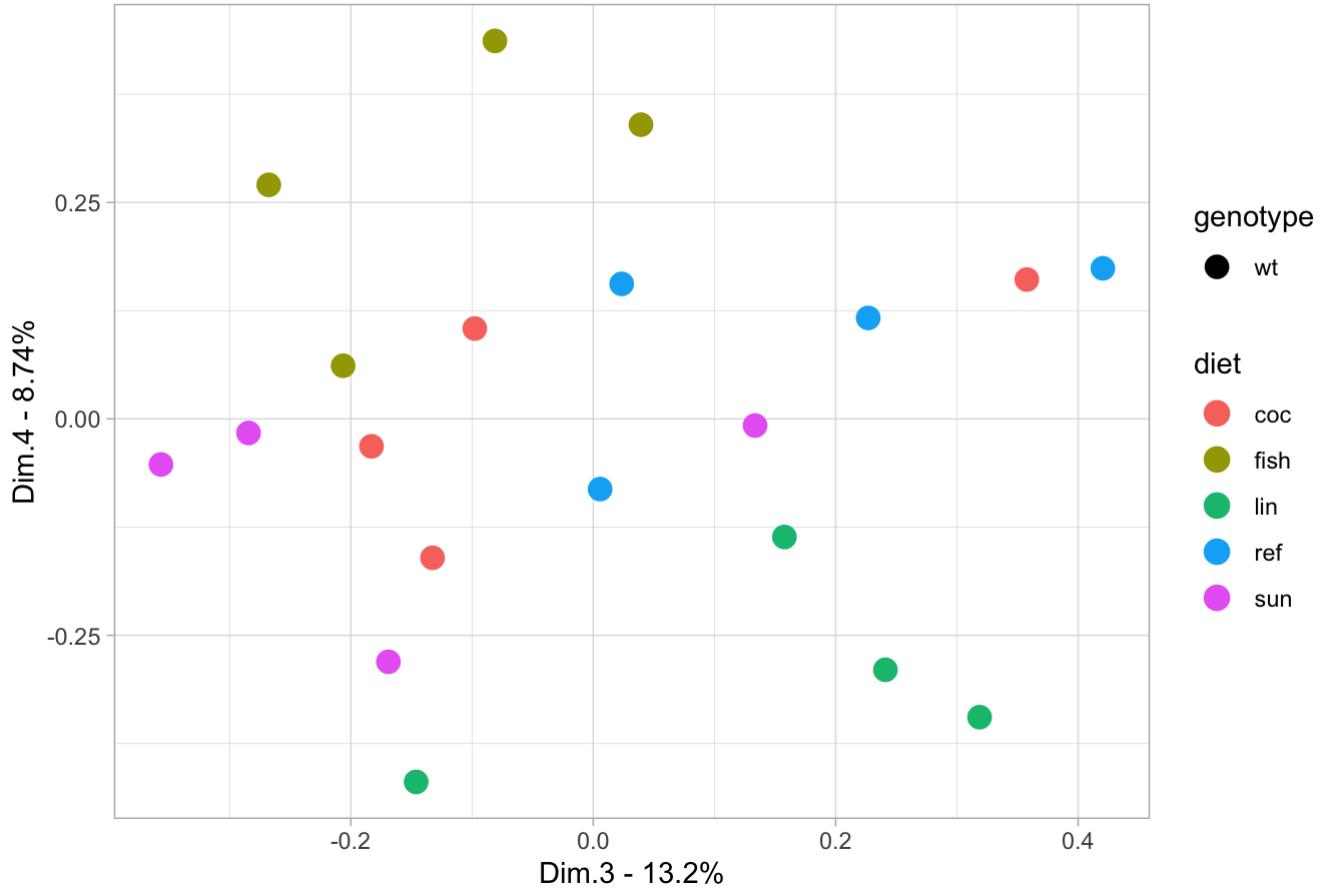
## scores plots on Dim.1 Dim.2



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()
```

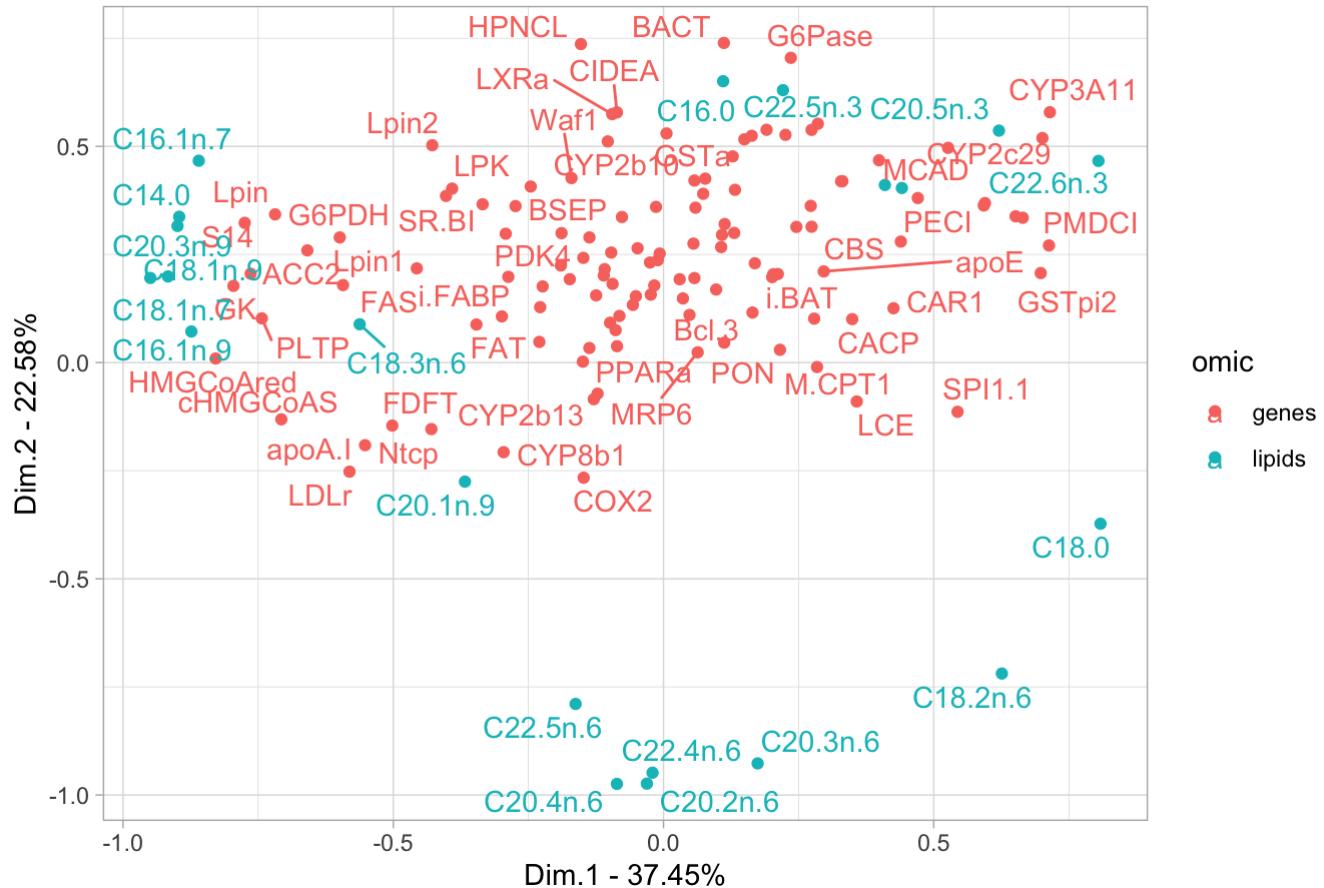
### scores plots on Dim.3 Dim.4



HIDE

```
# loadings plots
wt_loadings <- data.frame(wt_ComDim_res$globalcor)
wt_loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
wt_loadings$variable <- rownames(wt_loadings)
ggplot(wt_loadings, aes(x=X1, y=X2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", wt_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", wt_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```

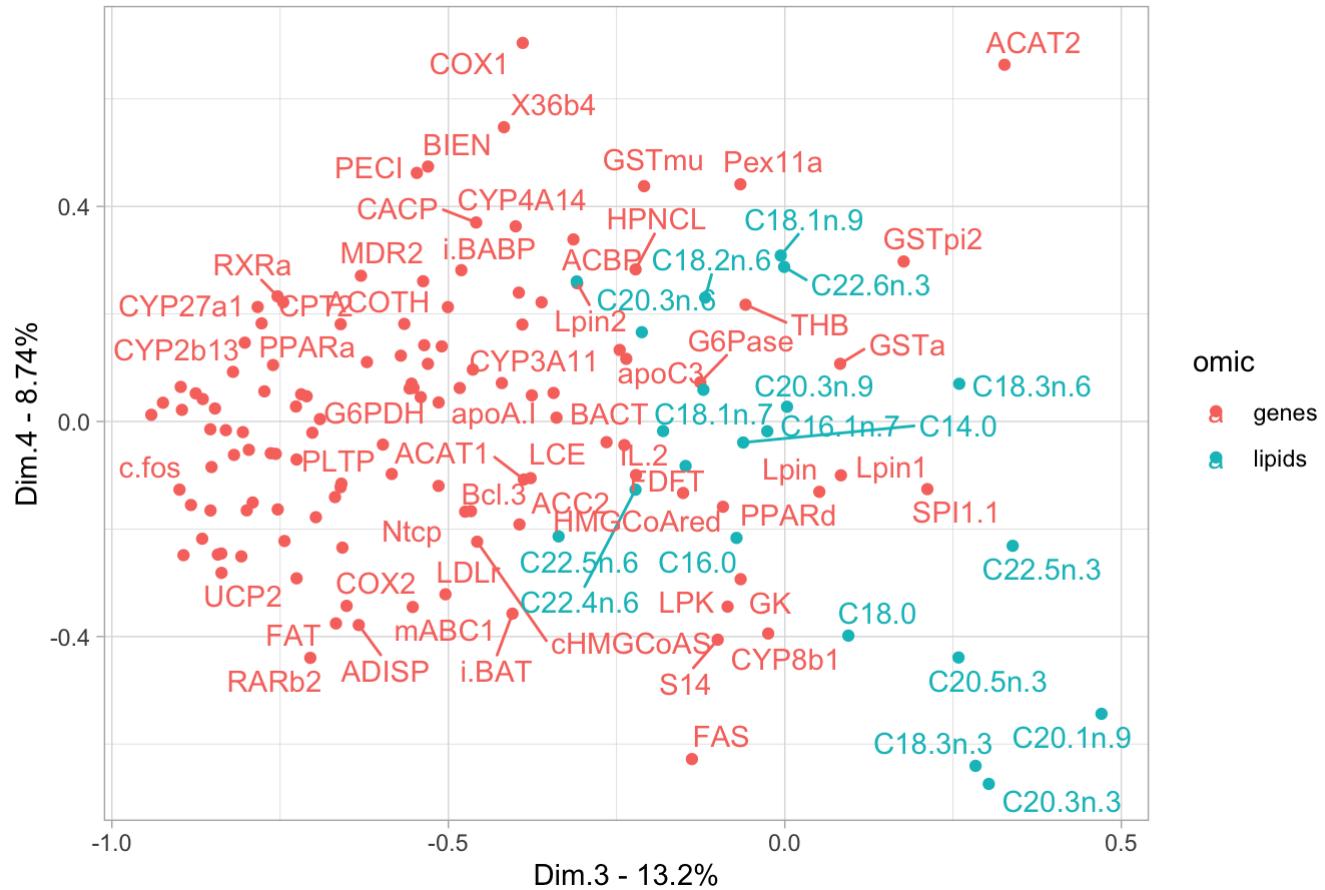
### loadings plots on Dim.1 Dim.2



HIDE

```
ggplot(wt_loadings, aes(x=X3, y=X4, 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()
```

loadings plots on Dim.3 Dim.4



## ComDim analysis of ppar samples

HIDE

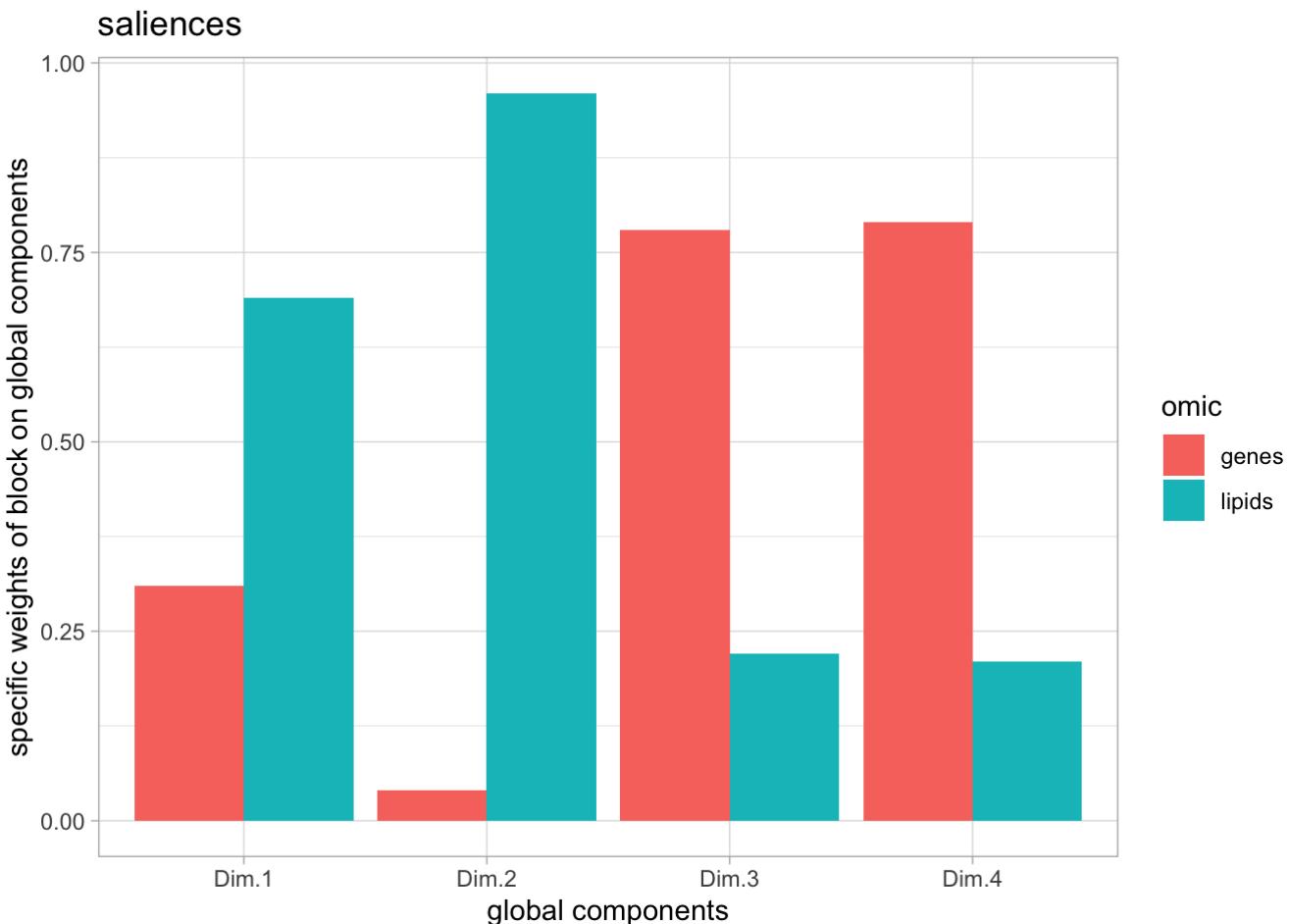
```

# 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, group = n_group, option = "uniform", plotgraph = F)

# 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

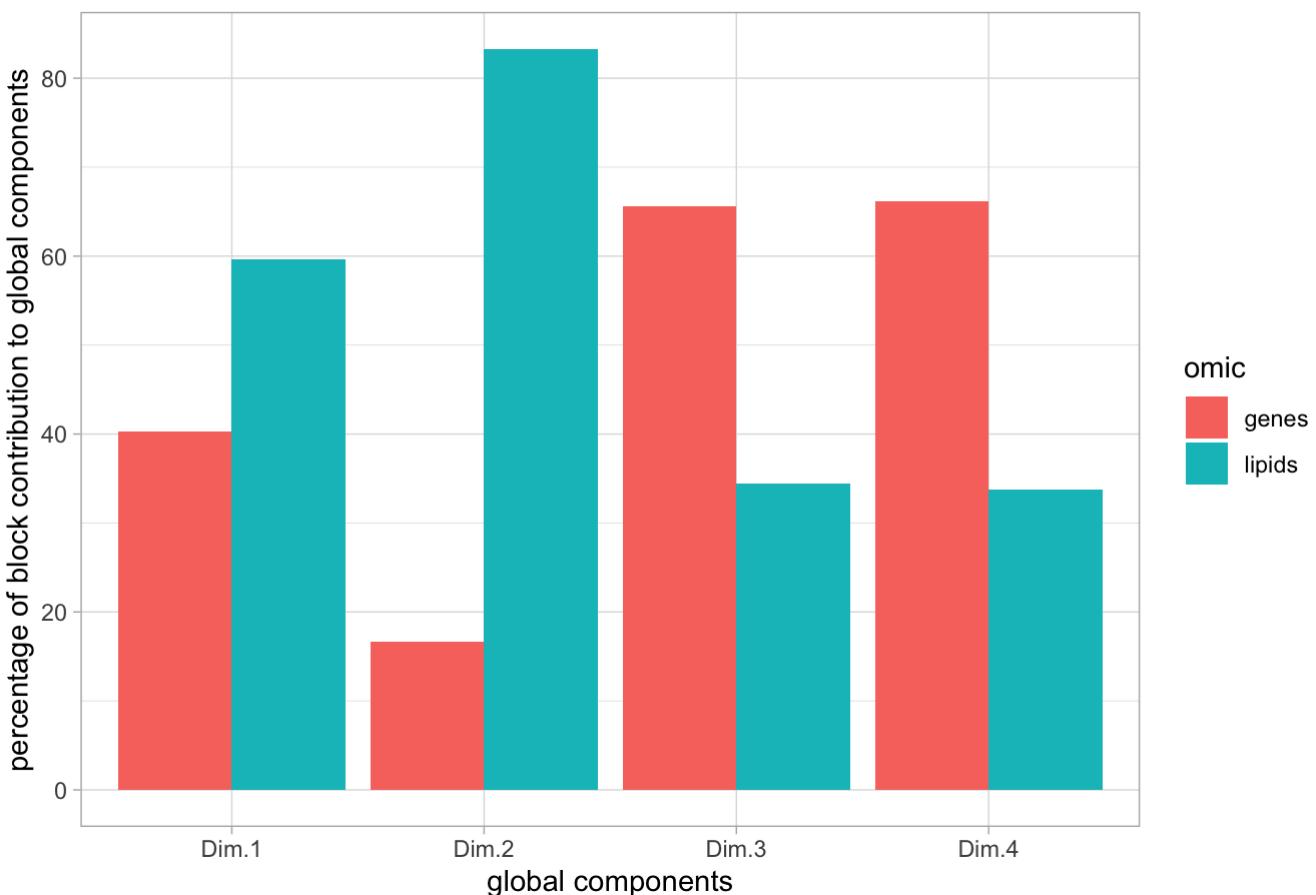
```

# block contributions

ppar_contributions <- ppar_ComDim_res$contrib
rownames(ppar_contributions) <- c("genes", "lipids")
ppar_contributions <- as.data.frame(t(ppar_contributions[,1:4]))
ppar_contributions$Dim <- rownames(ppar_contributions)
ppar_contributions <- melt(ppar_contributions)
ggplot(ppar_contributions, aes(x=Dim, y=value, fill=variable)) +
  geom_bar(stat = "identity", position=position_dodge()) +
  theme_light() +
  labs(x = "global components", y = "percentage of block contribution to global component
s", fill = "omic",
       title = "blocks contributions")

```

blocks contributions



HIDE

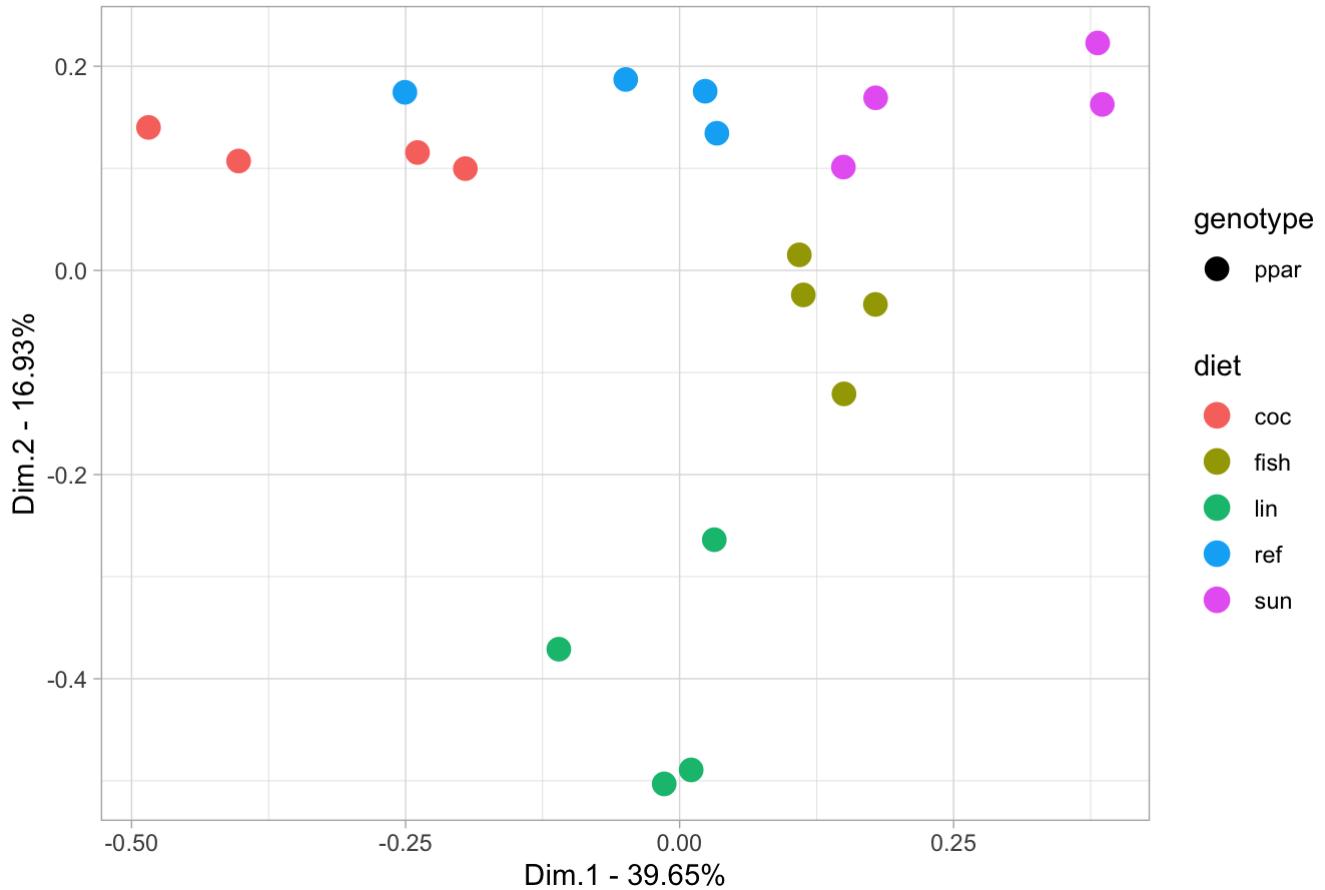
```

# scores plots

ppar_scores <- data.frame(metadata[metadata$genotype == "ppar",], ppar_ComDim_res$T)
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()

```

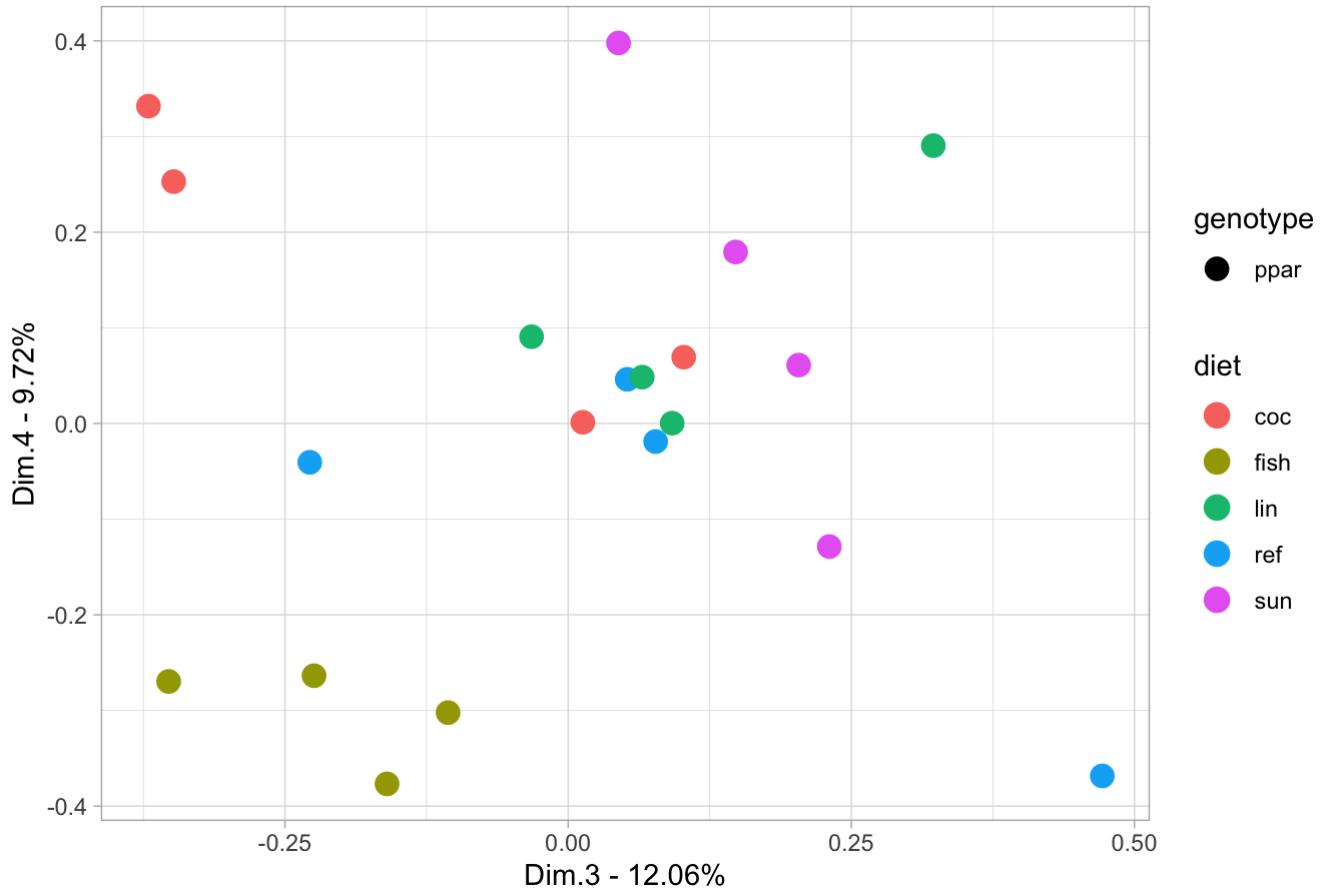
### scores plots on Dim.1 Dim.2



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()
```

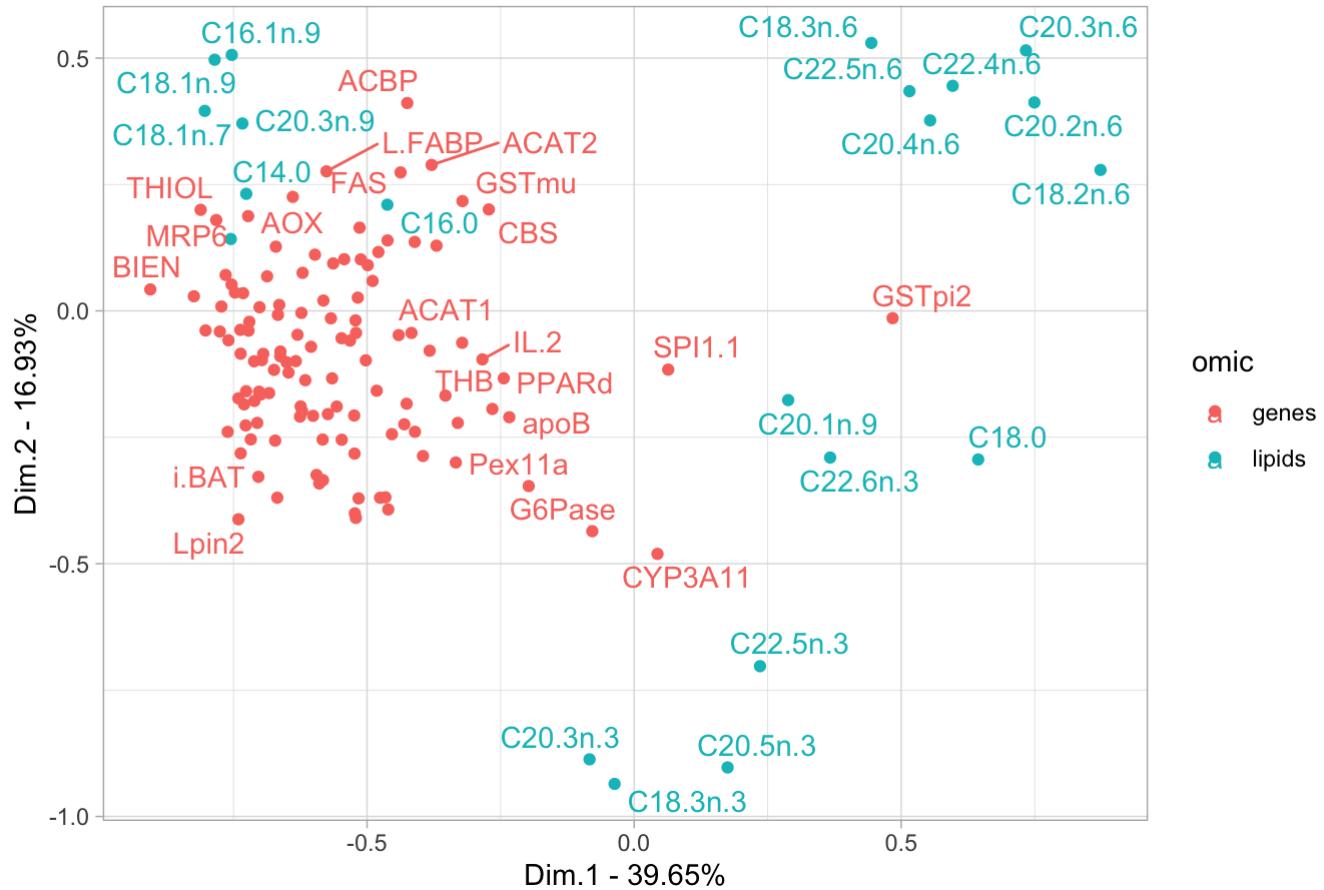
### scores plots on Dim.3 Dim.4



HIDE

```
# loadings plots
ppar_loadings <- data.frame(ppar_ComDim_res$globalcor)
ppar_loadings$omic <- c(rep("genes", dim(genes)[[2]]), rep("lipids", dim(lipids)[[2]]))
ppar_loadings$variable <- rownames(ppar_loadings)
ggplot(ppar_loadings, aes(x=X1, y=X2, col=omic, label=variable)) +
  geom_point() +
  geom_text_repel() +
  labs(x=paste0("Dim.1 - ", ppar_ComDim_res$cumexplained[1,"%explX"], "%"),
       y=paste0("Dim.2 - ", ppar_ComDim_res$cumexplained[2,"%explX"], "%"),
       title = "loadings plots on Dim.1 Dim.2") +
  theme_light()
```

### loadings plots on Dim.1 Dim.2



HIDE

```
ggplot(ppar_loadings, aes(x=X3, y=X4, 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()
```

loadings plots on Dim.3 Dim.4

