Improving visualisation and interpretation of metabolome-wide association NMR metabolic profiling studies (MWAS): an application in a population-based cohort using untargeted 1H

Raphaele Castagne

Marc Chadeau-Hyam

## About the Tutorial

This tutorial will provide an overview of how to estimate the effective number of tests, generate the Manhattan plot (Figure 2) and regional association plot (Figure 3). We will use simulated data/results.

## Estimation of the effective number of tests

Full description of the Metabolome-Wide Significance Level (MWSL) and Effective Number of Tests (ENT) is given in the Methods section of the article. Briefly its calculation relies on N permutations of the outcome in the study population (N=3948). For each permutation (N=10000), we shuffled the outcome, we then calculate the p-value for each NMR data point (N=30590) using a linear regression model and record the p-value.

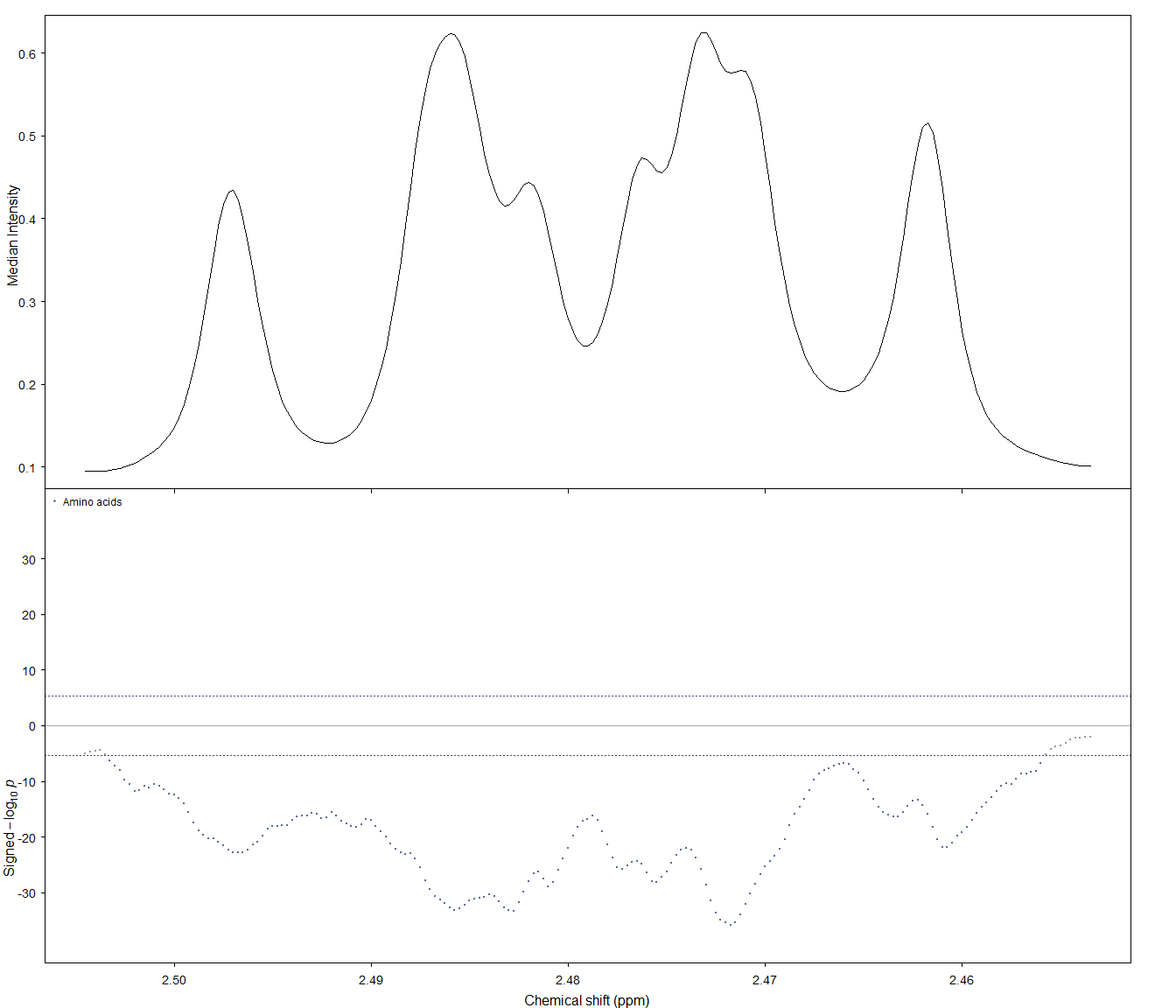
#Define parameters  
N.samples <- 3948   
N.permutation <- 1000 #10000  
N.data.point <- 150 #30590  
  
#Create an empty matrix to save the p-value   
p.perm <- matrix(NA, nrow=N.data.point, ncol=N.permutation)   
  
#Fill p.perm matrix with simulated p-value   
for (k in 1:N.permutation)  
{  
 p <- c()  
for (i in 1:N.data.point) {  
 x <- rnorm(N.samples)  
 p <- append(p, t.test(x)$p.value)  
  
}   
 p.perm[,k] <- p  
  
 }  
  
  
#Extract minimum p-value across N permutation  
min.p=apply(p.perm,2,min)  
#Order minimum p-value   
min.p.ord=sort(min.p)  
  
#Define qnalpha which denotes the n × alpha th smallest value for q, among its N realizations for alpha=0.05  
q.n.alpha=dim(p.perm)[2]\*0.05  
#Extract MWSL  
MWSL=min.p.ord[q.n.alpha]  
#And corresponding ENT for alpha = 0.05  
ENT = 0.05/ MWSL  
#Extract 95% confidence interval   
q.n.up = q.n.alpha+sqrt(q.n.alpha\*0.95)  
q.n.lw = q.n.alpha-sqrt(q.n.alpha\*0.95)  
 #MWSL  
MWSL.up = min.p.ord[q.n.up]  
MWSL.lw = min.p.ord[q.n.lw]  
 #ENT  
ENT.CI.up = 0.05/MWSL.up  
ENT.CI.lw = 0.05/MWSL.lw  
  
 #Return results  
res.ENT <- cbind(ENT, ENT.CI.up, ENT.CI.lw, MWSL, MWSL.lw, MWSL.up)  
print(res.ENT)

## ENT ENT.CI.up ENT.CI.lw MWSL MWSL.lw MWSL.up  
## [1,] 171.0646 148.7955 194.6655 0.0002922872 0.0002568509 0.0003360317

## Manhattan plot

This type of representation arose from the field of genome-wide association studies. It offers a global view of the spectral regions associated with the outcome of interest, and can be further informed by the annotation of spectral features. We report for each spectral variable the -log10 p-value multiplied by the sign of the corresponding regression coefficient.

rm(list=ls())  
#library needed  
library(RColorBrewer)  
  
#load results for the CPMG, glutamine, model 2  
results <- readRDS("C:/Users/Raphaële/Documents/CombiBio/CombBio\_2/NMR\_final\_JPR/Revised\_manuscript/R\_tutorials/Results\_glutamine\_CPMG\_mod2\_glucose\_log.rds")  
#define x and y  
results$x <- as.numeric(row.names(results))  
results$mean.spectra <- with(results, (mean.mesa.ph1+mean.mesa.ph2)/2)  
results$y <- with(results, sign(coef) \* -log10(pval))  
  
#define significant data points  
#FWER  
alpha = 0.05  
#ENT  
ent = 12554.26  
results$mod2.sig <- results$pval < alpha / ent  
  
#Set graphical parameters  
#point size  
results$cex <- 0.25  
results$cex[results$mod2.sig] <- 0.30  
  
#point color  
getPalette = colorRampPalette(brewer.pal(9, "Set1"))(11)   
results$col <- "#999999"  
results[results$mod2.sig,"col"] <- getPalette[2]  
  
#point shape  
results$pch <- 20  
  
#define window device  
par(mfrow = c(2, 1))  
par(oma=c(3,3,1,3))  
par(cex = 0.9)  
par(mar = c(0,0,0,0))  
par(tcl = -0.25)  
par(mgp = c(2, 0.6, 0))  
   
# Plot mean spectra on the top panel  
plot(results$x,results$mean.spectra,  
 xlim=rev(range(results$x)),  
 type="l",axes=F, ylim=range(results$mean.spectra))   
   
axis(1,labels = FALSE)  
axis(2,las=1)  
mtext("Median Intensity", 2, side=2, line=1.7, adj=0.8, cex=1)  
box(lwd=0.5)  
   
# Plot non-significant associations first  
ylim <- c(-1.1, 1.1) \*max(abs(results$y))  
main <-""  
with(subset(results, !mod2.sig),  
 plot(x, y,  
 pch=pch, cex=cex, col=col,  
 xlim=rev(range(x)), ylim=ylim,  
 xlab="Chemical shift (ppm)", ylab="",  
 main=main, axes=FALSE)  
)  
  
abline(h=c(-1, 1) \* -log10(alpha / ent), lty=3,col="darkslateblue")  
abline(h=0, col="darkgrey")  
   
# Plot axes  
axis(1)  
axis(2,seq(-30,30,10),las=1)  
mtext(expression(paste("Signed ", -log[10]~italic(p))), 2, side=2, line=1.7, adj=0.1, cex=1)  
   
with(subset(results, mod2.sig),   
 points(x,y,  
 pch=pch,  
 col= col,  
 cex=cex)  
 )   
box(lwd=0.5)   
   
# Add legend   
legend("topleft",  
 legend=c("Amino acids"),  
 pch=20,  
 col=getPalette[2],  
 bty="n", ncol=1, cex=0.8)   
   
# Add x-axis legend   
mtext("Chemical shift (ppm)",1,side=1,line=2)



## Regional association plot

Interpretation of MWAs can be facilitated by visualisation. Regional association plot aims to highlight the context of the association results for a given metabolite. It provides information about (i) the association between each NMR data point within a region and the outcome of interest (ii) the correlation within datapoint (iii) mean intensity and (iv) the stability of the results.

rm(list=ls())  
#library needed  
library(RColorBrewer)  
library(plotrix)  
   
#FWER  
alpha = 0.05  
#ENT  
ent = 12554.26  
   
#load results for the CPMG, glutamine, model 2  
results <- readRDS("C:/Users/Raphaële/Documents/CombiBio/CombBio\_2/NMR\_final\_JPR/Revised\_manuscript/R\_tutorials/Results\_glutamine\_CPMG\_mod2\_glucose\_log.rds")  
  
#define x and y  
results$x <- as.numeric(row.names(results))  
results$y <- with(results, -log10(pval))  
  
#graph parameters  
#sympbol  
results[sign(results$coef) < 0,"pch"] <- 25  
results[sign(results$coef) > 0,"pch"] <- 24  
#size  
results$cex <- 0.4  
results[results$cor.nmr.t>= 0 & results$cor.nmr.t<0.4,"cex"] <- 0.4  
results[results$cor.nmr.t>= 0.4 & results$cor.nmr.t<0.6,"cex"] <- 0.8  
results[results$cor.nmr.t>= 0.6 & results$cor.nmr.t<0.8,"cex"] <- 1  
results[results$cor.nmr.t>= 0.8 & results$cor.nmr.t<1,"cex"] <- 1.2  
results[which.min(results$pval),"cex"]=3  
#colour  
heat=heat.colors(10, alpha = 1)  
results[results$cor.nmr.t>= 0 & results$cor.nmr.t<0.1,"bg.pch"] <- heat[10]  
results[results$cor.nmr.t>= 0.1 & results$cor.nmr.t<0.2,"bg.pch"] <- heat[10]  
results[results$cor.nmr.t>= 0.2 & results$cor.nmr.t<0.3,"bg.pch"] <- heat[8]  
results[results$cor.nmr.t>= 0.3 & results$cor.nmr.t<0.4,"bg.pch"] <- heat[7]  
results[results$cor.nmr.t>=0.4 & results$cor.nmr.t<0.5,"bg.pch"] <- heat[6]  
results[results$cor.nmr.t>=0.5 & results$cor.nmr.t<0.6,"bg.pch"] <- heat[5]  
results[results$cor.nmr.t>=0.6 & results$cor.nmr.t<0.7,"bg.pch"] <- heat[4]  
results[results$cor.nmr.t>= 0.7 & results$cor.nmr.t<0.8,"bg.pch"] <- heat[3]  
results[results$cor.nmr.t>=0.8 & results$cor.nmr.t<0.9,"bg.pch"] <- heat[2]  
results[results$cor.nmr.t>=0.9 & results$cor.nmr.t<1,"bg.pch"] <- heat[1]  
results[which.min(results$pval), "bg.pch"] <- "grey30"  
  
#define window device  
op <- par(mfrow = c(2,1),  
 oma = c(5,5,0,3) + 0.1,  
 mar = c(0,0,1,3) + 0.1,xpd=TRUE)  
   
#plot mean corrected intensity samples in the 5th with low and high glucose (after adjustement for the FRS variable)  
with(results,   
 plot(x,mean.spectra.reg.lw, pch=1,type="l",  
 xlim=rev(range(as.numeric(x))),col="darkslateblue",  
 yaxt="n",ylim=range(c(mean.spectra.reg.lw, mean.spectra.reg.up)),   
 ylab="",xaxt="n",xlab="",bty="n", lty=1, cex=1.2) )  
  
  
with(results,   
 lines(x,mean.spectra.reg.up,pch=1,type="l",  
 xlim=rev(range(as.numeric(x))),col="darkgreen",  
 yaxt="n",ylim=range(c(mean.spectra.reg.lw, mean.spectra.reg.up)),   
 ylab="",xaxt="n",xlab="", lty=2, cex=1.2))  
#add vertical lines  
abline(v= round(seq(min(as.numeric(results$x)), max(as.numeric(results$x)), 0.005),2)[-1], lty=2, col="lightgrey" )  
#add legend  
legend("top",c("5% high glucose","5% low glucose"),  
 xpd = TRUE, inset = c(0.50,-0.1),  
 bty = "n", lty = c(2,1), col = c("darkgreen","darkslateblue"),  
 cex = 1, horiz=F,ncol=2, bg="white")  
#add axis on the right  
axis(side=4,las=2,cex.axis=1,cex.lab=1)   
mtext("Mean corrected intensity", side=4, line=4)  
   
#on the same plot, add NMR results  
par(new=TRUE)  
with(results,   
 plot(x,y,xlim=rev(range(as.numeric(x))),type="n",yaxt="n",  
 ylim=c(min(y)-1,max(y)+50),  
 ylab="",axes=FALSE,xlab="",main=""))  
with(results, points(x,y,pch=pch,cex=cex,type="p",col="black", bg=bg.pch))  
   
axis(side=2,las=1)   
mtext(expression(paste(-log[10]~italic(p))), side = 2, line=2.5)   
text(x=results[which.min(results$pval),"x"],  
 y=results[which.min(results$pval),"y"]+6,results[which.min(results$pval),"x"] )  
abline(h= -log10(alpha / ent), lty="dotted", lwd=1, col="black")  
   
#bottom panel  
#plot mean corrected intensity by MESA phase  
all.mean <- with(results,c(mean.mesa.ph1,mean.mesa.ph2 ))  
with(results,plot(x,mean.mesa.ph1,  
 xlim=rev(range(as.numeric(x))),  
 ylim=range(c(all.mean)),  
 type="l",  
 axes=FALSE,  
 col="black"))  
with(results, lines(x, mean.mesa.ph2, col="black", lty=2))  
  
axis(side=2,las=1)  
axis(1, round(seq(min(as.numeric(results$x)), max(as.numeric(results$x)), 0.005),2), las=1)  
mtext("Mean intensity", side=2, line=3)  
mtext("Chemical shift [ppm]", side=1, line=2)  
abline(v= round(seq(min(as.numeric(results$x)), max(as.numeric(results$x)), 0.005),2)[-1], lty=3, col="lightgrey" )  
legend("top",c("MESA phase 1","MESA phase 2"),  
 xpd = TRUE, inset = c(0.50,-0.1),  
 bty = "n", lty = c(1, 2), col = c("black","black"),  
 cex = 0.9, horiz=F,ncol=2)  
   
#Add prioritisation results  
par(new=TRUE)  
with(results, plot(x, r.pc,xlim=rev(range(as.numeric(x))),yaxt="n",ylab="",  
axes=FALSE,xlab="",main="", pch=21, cex=0.8, col="black", bg="mediumaquamarine"))  
axis(side=4,las=1)  
mtext("Proportion of replication", side=4, line=4)  
  
#add legend at the bottom plot  
par(fig = c(0, 1, 0, 1), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)  
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")  
  
legend.coord= legend("bottom", c("IM", "IBD", "1R", "2R"),  
 xpd = TRUE, horiz = TRUE, inset = c(0,0), bty = "n",  
 pch = c(4, 2, 15, 19), col = 1:4, cex = 0.9, plot=FALSE)  
gradient.rect(legend.coord$rect$left,  
 legend.coord$rect$top-(legend.coord$rect$w/12),  
 legend.coord$rect$left+(5\*legend.coord$rect$h),  
 legend.coord$rect$top,  
 col=rev(heat.colors(10)),gradient="x",border=NA)   
  
text(legend.coord$rect$left,  
 legend.coord$rect$top-0.08 , "0",cex=0.9)  
  
text(legend.coord$rect$left+(5\*legend.coord$rect$h),  
 legend.coord$rect$top-0.08 , "1",cex=0.9)  
  
text( legend.coord$rect$left + 0.25,  
 legend.coord$rect$top-0.08, "Correlation", cex=0.9)

