UVIC UNIVERSITAT DE VIC

Master of Science in Omics Data Analysis

Master Thesis

Long-term health risks of hyperinsulinaemic androgen excess appear in adolescence with serum markers of oxidative stress and impaired HDL maturation through oxidation of methionine residues in apolipoprotein A1

By

Sara Samino Gené

Supervisor: Oscar Yanes Torrado

Department of Systems Biology
University of Vic – Central University of Catalonia
15/09/14

OUTLINE

BACKGROUND

This Project has been carried out within the Metabolomics Platform. The Metabolomics Platform is a joint research facility created in 2008 by URV (Universitat Rovira i Virgili) and CIBERDEM (Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas). The main goal of the Metabolomics Platform is to offer metabolomic services to the biomedical and clinical research groups from CIBERDEM and URV. This project is part of an on going internal CIBERDEM project in colaboration with Dra. Lourdes Ibañez from Hospital Sant Joan de Déu (Barcelona).

PRESENTATION

I started my phD in Metabolomics Platform in 2009. In my thesis, I have participated in several metabolomics experiments involving different analytical platforms (NMR, LC-MS and GC-MS). The project here described have been the last work presented in my doctoral thesis. I have contributed intensively in this work, technically and scientifically.

In this project, I have developed an untargeted metabolomic workflow from sample preparation, data processing for metabolite identification (including statistical analysis) to biological interpretation. I have been involved in all of this steps. The main goal of this project was to implement a multiplatform approach based NMR and LC/MS to provide new insights in disease in a cohort of young lean hyperinsulinaemic androgen excess patients (For more details see ARTICLE).

Untargeted metabolomics have the aim to simultaneously measure as many metabolites as possible from biological samples without bias. This metabolomic approach is typically hypothesis generating and provides a global comparative overview of metabolites abundances between two or more sample groups (i.e., experimental conditions), for example, healthy vs. disease or, WT (wild type) vs. KO (knock-out). The main bottleneck of an untargeted approach is the identification of metabolites on the basis of MS and/or NMR peaks from exceedingly complex datasets. Despite this challenge, this approach has the potential to involve previously unrecognized metabolites in seemingly known human pathological conditions. For this reason, this project has been focused on an untargeted metabolomic approach.

MY CONTRIBUTION

The aim of this project was to determine the underlying mechanism by which hyperinsulinemic androgen exces is associated with long-term health risks in adulthood such as anovulatory infertility and long-term health risks such as type 2 diabetes, metabolic syndrome and possibly cardiovascular disease. In this project we have been used NMR and MS-based metabolomics to compare the lipoprotein profile and the serum metabolome of non-obese adolescents with HIAE, with those in age- and weight-matched control girls.

I have contributed to the experimental design; NMR, LC-MS and MALDI-TOF MS sample preparation; LC-MS data adquisition; NMR, LC-MS and MALDI-TOF MS data analysis; biological interpretation and integration of all results. Moreover, I have written the whole final manuscript, excluding NMR and MADLI-TOF MS technical methods which were written by other authors. The entire study was supervised by Oscar Yanes. The final manuscript was supervised by Oscar Yanes and Lourdes Ibañez.

UNTARGETED METABOLOMICS EXPERIMENT

In general, in untargeted metabolomic studies, numerous factors must be carefully considered. These factors include: sample preparation, implementation of appropriate MS and NMR analytical tools for sample analysis, data analysis, and last but not least, biological interpretation. Hence, a well-defined workflow characterizes untargeted metabolomics. All of this sections have been performed in this project. Sample preparation have been described in detail in (ARTICLE: MATERIALS AND METHODS). However, I have considered to expand the data analysis of this project. Once, data has adquired for LC-MS the first step is to convert the raw data into a standard format such as mzData using ProteoWizard. Then, a package based on R (xcms) has been used for data processing. The XCMS package provide methods for peak picking, non-linear retention time alignment, visualization, and relative quantitation. XCMS is implemented as a package within the R programming environment. XCMS implements three main steps:

1. *Peak detection*: the algorithm identifies peaks in each of the samples;

2. *Non-linear retention time alignment*: matches peaks with similar retention times across multiple samples, and use the groups of matched peaks for time alignment;

3. *Fill in any missing peaks:* the peaks that have failed to be detected in step 1, are filled in properly from raw data.

As a result, the output matrix of XCMS includes the intensity under the peak for all features detected, for every sample. In this process, thousands of so called metabolite features are routinely generated. A feature is defined as a peak corresponding to an individual ion with a unique mass-to-charge ratio and a unique retention time. It is worth emphasizing that one feature is not equivalent to one metabolite. Generally, more than one feature belongs to the same metabolite. In LC-MS data, isotopic distributions, potential adducts or in-source fragmentation generates more than one feature for the same metabolite. Next, the statistical analysis will reveal which features are significantly different between sample groups. Finally, the last step is metabolite identification.

Identification of metabolites by LC-MS:

To determine the identity of a feature of interest, the accurate mass of the compound is first searched in metabolite databases such as Metlin or the Human Metabolome Database (HMDB).

In this project, more than 50000 features were obtained after xcms processing from samples analyzed by LC-MS using HILIC chromatography. All of this features were filtered following different criterias (see the script section) and after filtering for statistical analysis (Yuen-Welch's test) and fold change, I obtained 250 features that were statistically significant between control and disease. This 250 features were searched in Metlin database and only 80 of the initial 250 features returned a hit in the database. However, a database match represents only a putative assignment that must be confirmed by comparing the retention time and MS/MS data of a model compound to that from the feature of interest in the research sample. Currently, MS/MS data for features selected from the profiling results are obtained from additional experiments and matching of MS/MS fragmentation patterns is performed manually by inspection. Then, after running samples again with the aim to perform the MS/MS fragmentation experiments only 11 metabolites were identified.

SCRIPT

XCMS package:

```
library(xcms)
```

```
# Perform peak detection on data in centroid mode using centWave
algorithm
xset<-xcmsSet(method="centWave",ppm=30,peakwidth=c(10,60))
# Match peaks across samples: place in to groups peaks representing
the same metabolite #across samples
xset<-group(xset)
# Make use of group information to identify and correct retention time
missalignments from run #to run using Obi-warp
xset2<-retcor(xset,method="obiwarp",profStep=1)
#Make a second pass of grouping algorithm to a more accurate group
definition
xset2<-group(xset2, mzwid=0.025,minfrac=0.5,bw =5)
#Finally fill in missing peaks
xset3<-fillPeaks(xset2)
```

DATA PROCESSING (Filtering data)

###Intensity criteria

The accuracy of a metabolite identi_cation depends on the detection of all, most, or informative fragment ions generated during the MS/MS experiments. Lower-intensity parent ions obtained in profling mode lead to poor MS/MS spectra quality hampering this way a proper identi_cation. Therefore, it is worth to perform MS/MS experiments for those parent ions presenting a minimum intensity threshold. This threshold depends on the experimental conditions. However, according to our experience, a reasonable threshold in our example would be 10000 counts.

```
# Get intensities from xcmsSet object
X1 <- groupval(xset3, value = "maxo")
# Define experimental groups
class <- as.factor(xset3@phenoData$class)
# Compute the mean intensities for each group
meanintensities <- apply(X1, 1, function(x) tapply(x, class, mean))
# Getting index of those features with mean intensity above 10000
counts in at least one of the #groups
idx_intensity <- which(apply(meanintensities, 2, function(x) any(x
>10000)) == TRUE)
```

###Handle analytical variation:

Most common sources of analytical variation in LC-MS experiments are due to sample preparation and instrumental drifts caused by chromatographic columns and MS detectors. The ideal method to examine analytical variation is to analyze quality control (QC) samples, which will provide robust quality assurance of each detected mzRT feature. To this end, QC samples should be prepared by pooling aliquots of each individual sample entering the study and analyze them periodically through-out the sample work list.

Being replicates of the same pooled samples QC samples are expected to not contain biological variation. Hence variation observed in QC samples reects analytical variation. the performance of the analytical platform can be calculated individually for each detected mzRT feature computing the variation of each mzRT feature around their mean in QC pooled samples (CV_{QC}).

This leads to an estimation of the analytical variation. Additionally, we can also calculate the variation of mzRT features around their mean in the samples entering the experiment (CV_S). This variation would enclose both, analytical and biological variation.

Consequently, those mzRT features where $CV_{QC} > CV_S$ contain more analytical variation than other sources of variation and they should be conveniently removed from further analysis. This criteria helps to focus on mzRT features holding the lowest proportion of analytical variation

```
# 1-. Define (CV) function
```

```
co.var <- function(x) (100 * sd(x)/mean(x))
# 2-. Define QC and Sample samples classes
cl1 <- rep(c("Sample", "QC"), times = c(38, 8))
# 3-. Compute CV for QC and Samples
CV <- t(apply(X1, 1, function(x) tapply(x, cl1, co.var)))
# 4-. Determine the percenteage of features where CV(Samples)>CV(QC)
idx_qc <- which(CV[, "Sample"] > CV[, "QC"])
```

###Combine both criteria

Combine both QC and Intensity criteria, filter the original dataset accordingly and get rid out of QC samples.

```
# Combine both intensity and QC criteria
Ii <- colnames(D)[idx_intensity]
Iqc <- colnames(D)[idx_qc]
Ib <- intersect(Ii, Iqc)
# Create a new dataset without QC samples and with mZRT features
meeting both criteria
D2 <- subset(D, class != "QC", select = Ib)
```

###Statistical analysis

Provided the limitation in the number of MS/MS con_rmation experiments, additional criteria to those described above are necessary to further reduce the number of initial mzRT features to an amenable number. In our pipeline once QC and intensity criteria has been applied, the resulting dataset is usually specified via hypothesis testing to select mzRT features showing statistical signifcance.

```
# Yuen function

library(PairedData)

Yuen_sara<-function (x,y)
{
  resultyuen<-list(1:ncol(x))
  pvalue<-matrix(nrow=ncol(x),ncol=1)
  colnames(pvalue)<-"pvalue"
  rownames(pvalue)<-colnames(x)

  for (i in 1:ncol(x)){

    resultyuen[[i]]<-yuen.t.test(x[,i],y[,i], tr=0.2, paired=FALSE)
    pvalue[i,1]<-resultyuen[[i]]$p.value

  }
  return(pvalue)
}

##Create two matrices

CTR<-D2[1:14,]
PCOS<-D2[15:26,]

pvalue_ctr_pcos <- Yuen_sara(CTR,PCOS)


#Create the fold change function
fc.test<-function(D2,classvector){
```

```
means<-apply(D2, 2, function(x) tapply(x, classvector, mean))
means <- t(means)
case <- means[,"case"];control <- means[,"control"]
logFC <- log2(case/control)
FC <- case/control;
FC2 <- -control/case
FC[FC<1] <- FC2[FC<1]
fc.res <- cbind(FC, logFC)
return(fc.res)
}

# Create a new group factor that does not include QC samples
gr <- as.factor(xset3@phenoData$class[xset3@phenoData$class != "QC"])

#Calculate FC for "DISEASE-CTR" groups
#Define control/case groups
gr2 <- as.character(gr)
gr2[gr2=="CTR"] <- "control"
gr2[gr2=="DISEASE"] <- "case"
#Calculate FC using fc.test function
fc.res <- fc.test(D2,gr2)
D3<- data.frame(pvalue_ctr_pcos, fc.res)

#Find those mzRT features matching statistical and fold change
criteria
D3$threshold = as.factor(abs(D3$FC) > 1.5 & D3$pvalue.ctr_pcos <
0.01);

#Calculate the number of features that match this criteria
table(D3$threshold)["TRUE"]
## TRUE
## 250

#Select those features
idxD3red <- rownames(D3)[which(D3$threshold == "TRUE")]
D3red <- D3[idxD3red,]
```

#Putative identification

Once initial xcmSet has been properly filtered, a good way to organize the resulting dataset is to rank significantly varied mZRT features according to their FC value. The final result would be a list of relevant features and their corresponding exact mass to query them to databases. From the initially filtered mZRT features, just those matching exact mass to database are retained for further MS/MS monitoring. It is important to collect all necessary information for these MS/MS spectra adcquisition.

Retention time, exact mass, and group mean intensities are mandatory.Another important poind is to visually inspect these relevant mZRT features. Although XCMS have some functionalities to display diferential data, it does not allow an iterative environment to explore it. Functions in XCMS to visualize data require to go back to mzXML files. Therefore, we should always operate in the directory where mzXML data is located. At this point, vendor software might offer better capabilities for displaying purposes and more exibility to visually inspect main results.

For NMR, raw data are typically processed by means of Fourier Transform (FT) and baseline removal. Baseline distortions affect not only the statistical analysis but also the quantification of metabolites. Multi integration of Regions of Interest (ROIs) is the

method used in this project. In this approach, the analysis is not directly based on peaks or spectal binned regions but on regions which are a priori known to contain compound resonances or ROIs. The area under ROIs representative of metabolites can be considered as a surrogate of the relative abundance of these metabolites in the the biological sample. The key issue here is to properly assign ROIs resonances to metabolite structures. This is not a trivial issue. NMR peak structure assignemnt in complex data matrices such as in cell cultures requires a skilled NMR specialist, acquisition of 2D NMR spectra and libraries of pure compounds to compare with. This assignment is actually the bottle neck in the whole NMR-based metabolic analysis. Noteworthy, it is eventually impossible to assign all resonances arising from complex matrices to chemical structures. Therefore, we will much probably end-up with some unknown resonances. Despite this daughting scenario, in this project, I have used a commercial software, AMIX, which incorporate libraries of pure standar compounds resulting in great help for the resonance assignment process.

Analyzing Multi Integration Results

```
# Yuen function

library(PairedData)

Yuen_sara<-function (x,y)
{
  resultyuen<-list(1:ncol(x))
  pvalue<-matrix(nrow=ncol(x),ncol=1)
  colnames(pvalue)<-"pvalue"
  rownames(pvalue)<-colnames(x)

  for (i in 1:ncol(x)){

    resultyuen[[i]]<-yuen.t.test(x[,i],y[,i], tr=0.2, paired=FALSE)
    pvalue[i,1]<-resultyuen[[i]]$p.value

  }
  return(pvalue)
}

##Create two matrices

data<-read.table("multiintegration.txt",sep="\t", header=T)
CTR<-data[1:14,]
PCOS<-data[15:26,]

##Apply the function

pvalue_ctr_pcos <- Yuen_sara(CTR,PCOS)
```

Lipoprotein profile and biochemical parameters

The technical issues of this parameters have been performed for other colleges which have contributed in the elaboration of this article. However, the statistical analysis of all of them and the figures and tables in the paper related to this issues have been performed by me. Statistical analysis have been performed in R.

SCRIPT

```
data1 <- read.table("Bioquimica_ctr_pcos_pio.txt", header=T, sep="\t")
data2<- read.table("Lipos_roger.txt", header=T, sep="\t")
data <- data1 [,2:13]
rownames(data) <- data1[,1]
lipos <- data2 [,2:13]
rownames(lipos) <- data2[,1]
rm(data1,data2)
##divido la matriz
CTR <- data[1:14,]
PCOS <- data[15:26,]
PIO <- data[27:32,]
##LIPOS
CTRL <- lipos[1:14,]
PCOSL <- lipos[15:26,]
PIOL <- lipos[27:32,]

##yuen test
Yuen_sara<-function (x,y)
{
  resultyuen<-list(1:ncol(x))
  pvalue<-matrix(nrow=ncol(x),ncol=1)
  colnames(pvalue)<-"pvalue"
  rownames(pvalue)<-colnames(x)
    for (i in 1:ncol(x)){
        resultyuen[[i]]<-yuen.t.test(x[,i],y[,i], tr=0.2,
paired=FALSE)
    pvalue[i,1]<-resultyuen[[i]]$p.value
      }
  return(pvalue)
}
library(PairedData)
pvalue_ctr_pcos <- Yuen_sara(CTR,PCOS)
pvalue_ctr_pio <- Yuen_sara(CTR,PIO)
pvalue_pcos_pio <- Yuen_sara(PCOS,PIO)
pvalue_ctr_pcos_LIPOS <- Yuen_sara(CTRL,PCOSL)
pvalue_ctr_LIPOS <- Yuen_sara(CTRL,PIOL)
pvalue_pcos_LIPOS <- Yuen_sara(PCOSL,PIOL)
```

Once, all data have been generated, biological interpretation have to be performed. The use of Cytoscape software for metabolic pathway analysis and searches in the literature were performed exhaustively by me.

I found that 9 metabolites found in the untargeted metabolomic experiment belong to the same pathway, in concrete to the gamma glutamyl cycle. The gamma-glutamyl cycle plays a key role in the synthesis and degradation of GSH. Most metabolites are increased in hyperinsulinemic androgen excess women which I associated with an increased demand of GSH due to increased oxidative stress (which has been extensively described in the literature). However

methionine and methionine sulfoxide were not involved in this pathway. Interestingly, lower levels of methionine in serum samples of hyperinsulinemic androgen excess women were associated with greater levels of methionine sulfoxide in these patients, indicating increased oxidation of this aminoacid. By other hand, I found a paper publish in 2008 demonstrates that methionine oxidation in apo-AI impairs reverse cholesterol transport by LCAT. Specifically oxidation of Met-148 residue in apo-AI changes the 3D structure of the protein imparing the activity of LCAT in HDL lipoproteins affecting the maturation of nascent HDL particles. In addition, I observe a very good correlation between large HDL and methionine sulfoxide levels. Increased concentration of methionine sulfuxide in serum of PCOS is associated with lower levels of large HDL particles. Therfore, I hypothetized that women with hyperinsulinemia androgen excess present greater oxidation of methionine residues in apo-AI relative to healthy women, and the increased levels of free methionine sulfoxide in hyperinsulinemic androgen excess patients may result from the turnover and degradation of these apo-AI proteins. With the aim to verify this hypothesis I proposed to perform a quantitative analysis of oxidized methionine/methionine ratio in apo AI. After protein precipitation serum samples were separed by SDS electrophoresis, and ApoA1 gel bands were digested and measured by MALDI-TOF MS (See results in appendice).

All plots performed in this project have been done by me. I have used ggplot2 library. Scripts code:

FIGURE 1:

```
####PASAR EL PLOT DE ROGER DE MATLAB A R####

setwd("C:/Documents and Settings/Sara/Escritorio/Ibañez_02/Final
Data/Julio2014/Figures/LIPOS ROGER/R")

methyl <- read.table("Total_metyl.txt", header=TRUE, sep="\t")

library(ggplot2)

cols <- c("Experimental Spectra"="black", "Fitted Spectra"="red",
"Background"="grey", "VLDL"="#E69F00","Large LDL"= "#56B4E9", "Small
LDL"="#009E73", "Large HDL"="#F0E442", "Medium HDL"="#0072B2", "Small
HDL"="#D55E00")

m <- ggplot(methyl, aes(x=x.axis, y=Total.line)) +
geom_line(aes(x=x.axis, y=Total.line, colour="Experimental Spectra"),
size=1) + xlim(1,0.7)

m <- m + geom_line(aes(x=x.axis, y=Total.subfractions, colour="Fitted
Spectra"), size=1)

m <- m + xlab("ppm")

m <- m + theme(axis.title.x= element_text(color="#999999",
face="bold",size=16))

m <- m + ylab("Intensity")

m <- m + theme(axis.title.y= element_text(color="#999999",
face="bold",size=16))

m <- m + geom_line(aes(x=x.axis, y=Background, colour="Background"),
size=0.5 )

m <- m + geom_line(aes(x=x.axis, y=VLDL, colour="VLDL"), size=1)
```

```
m <- m + geom_line(aes(x=x.axis, y=large.LDL, colour="Large LDL"),
size=1)

m <- m + geom_line(aes(x=x.axis, y=small.LDL, colour="Small LDL"),
size=1)

m <- m + geom_line(aes(x=x.axis, y=large.HDL, colour="Large HDL"),
size=1)

m <- m + geom_line(aes(x=x.axis, y=medium.HDL, colour="Medium HDL"),
size=1)

m <- m + geom_line(aes(x=x.axis, y=small.HDL, colour="Small HDL"),
size=1)

m <- m + scale_colour_manual( values = c("Experimental
Spectra"="black", "Fitted Spectra"="red", "Background"="grey",
"VLDL"="#E69F00","Large LDL"= "#56B4E9", "Small LDL"="#009E73", "Large
HDL"="#F0E442", "Medium HDL"="#0072B2", "Small HDL"="#D55E00"),
breaks=c("Experimental Spectra", "Fitted Spectra", "Background",
"VLDL","Large LDL", "Small LDL", "Large HDL", "Medium HDL", "Small
HDL"))

m <- m + ggtitle("Lipoprotein Deconvolution") + theme(plot.title =
element_text(size=18))

m
```

FIGURE 2 and FIGURE 3:

```
###PARA PINTAR LOS PLOTS INDIVIDUALES DE CADA VARIABLES

##tengo mi matriz data, quiero que sea un data.frame y que incorpore
#dos columnas, una con un factor que me defina las clases y otra
#con un vector que me defina los grupos(para que me pinte por colores)

colores <- rep(c("CTR","PCOS","PIOFLUMET"), times=c(14,12,6))
class <- as.factor(colores)
data <- as.data.frame(data)

###defino un dataframe al que le añado las dos columnas con las
variables
##creadas

pintar <-cbind(class,colores,data)

##quiero pintar para todos los metabolitos primero solo de CTR vs PCOS

metabolites <- c("Methionine","Methionine Sulfoxide", "5-
oxoproline","Taurine", "Glu-Cys", "Glu-Glu", "Glutamate", "GSH",
"GSH/GSSG", "Glu-Taurine", "Glycine")

ctrvspcos <- pintar[1:26,]

library(ggplot2)

###Para los gráficos del gamma-glutamyl cycle

ylabel <- rep(c("Intensity","Arbitrary Units"), times=c(10,1))
metabolites <- colnames(pintar)
```

```
cont <- 1


for (i in 12:22) {

  name<-paste(metabolites[i],".svg")

  pintar1 <- ctrvspcos[,1:2]
  pintar1 <- cbind(pintar1, ctrvspcos[,i])
  #colnames(pintar1)[3] <- colnames(pintar)[i]
  #metabolite <- colnames(pintar1)[3]

  b <- ggplot(pintar1, aes(x=class, y = log10(ctrvspcos[,i]), color =
colores)) + geom_point(size = 4, position = position_jitter(w = 0.1))

  b <- b + geom_segment(x=0.75,
y=mean(log10(pintar1[1:14,3]),trim=0.2), xend=1.25,
yend=mean(log10(pintar1[1:14,3]),trim=0.2, colour="black"), size=1,
colour="black")

  b <- b <- b + geom_segment(x=1.75,
y=mean(log10(pintar1[15:26,3]),trim=0.2), xend=2.25,
yend=mean(log10(pintar1[15:26,3]),trim=0.2, colour="black"), size=1,
colour="black")

  #b <-  b + geom_segment(x=2.75, y=mean(pintar1[27:32,3],trim=0.2),
xend=3.25, yend=mean(pintar1[27:32,3],trim=0.2, colour="black"),
size=1, colour="black")

  b <- b + ggtitle(label.title[i]) + theme(plot.title =
element_text(size=18, face="bold"))

  b <- b + theme(axis.title.x = element_blank()) +
ylab(paste("log10(",ylabel[cont], ")"))

  b <- b + theme(axis.text.x  = element_text(size=16)) +
theme(axis.text.x= element_text(face="bold"))

  b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(face="bold"))

  b <- b + theme(legend.position="none")

  b <- b + coord_fixed(ratio=3/0.75)

  b
  ggsave(name, dpi=300)
  cont <- cont+1
}
```

FIGURE 4:

```
####plots de proteomica

setwd("X:/YANES LAB/USUARIS
(YanesLab)/Sara/Plataforma/Usuaris/Ibañez_02/Final
Data/Julio2014/Figures/Proteomica_Lipos")

pintar <- DATA[1:18,]
```

```
colores <- rep(c("CTR","HIAE","PIOFLUMET"), times=c(8,10,5))
class <- as.factor(colores)
col <- rep(c("#F8766D", "#619CFF", "#00BA38"), times=c(8,10,5))

pintar <- cbind(class,colores,col, DATA)

ctrvspcos <- pintar[1:18,]

library(ggplot2)

###para el gráfico de la Met-148

name<-paste("Met-148",".svg")

pintar1 <- ctrvspcos[,1:4]
pintar1 <- pintar1[,-3]

b <- ggplot(pintar1, aes(x=class, y = ctrvspcos[,4], color = colores))
+ geom_point(size =4, position = position_jitter(w = 0.1),
colour=col1)

b <- b + geom_segment(x=0.75, y=mean(pintar1[1:8,3],trim=0.2),
xend=1.25, yend=mean(pintar1[1:8,3],trim=0.2), colour="black", size=1,
colour="black")

b <- b <- b + geom_segment(x=1.75, y=mean(pintar1[9:18,3],trim=0.2),
xend=2.25, yend=mean(pintar1[9:18,3],trim=0.2), colour="black",size=1,
colour="black")

#b <-  b + geom_segment(x=2.75, y=mean(pintar1[27:32,3],trim=0.2),
xend=3.25, yend=mean(pintar1[27:32,3],trim=0.2, colour="black"),
size=1, colour="black")

b <- b + ggtitle("MetOx-148/Met-148") + theme(plot.title =
element_text(size=18))

b <- b + theme(axis.title.x = element_blank()) + ylab("Ratio MetOx-
148/Met-148 in ApoAI")

b <- b + theme(axis.text.x  = element_text(size=16, color="#999999"))
+ theme(axis.text.x= element_text(face="bold"))

b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))

b <- b + theme(legend.position="none")

b <- b + coord_fixed(ratio=2/2)

b
ggsave(name, dpi=300)


###ahora vamos a hacer scatter plot de metox vs meth sufloxide in
serum
###creamos el dataframe

pintar1 <- pintar[,c(4,6)]
ctrvspcos <- pintar1[1:18,]

###modelo lineal
```

```
source("C:/ScriptsR/Functions R/pbcor.r")
source("C:/ScriptsR/Functions R/pbos.r")
library(MASS)

model <- pbcor(ctrvspcos[,2],ctrvspcos[,1], beta=0.2)
robust <-rlm(Ratio.Met.148~Methionine.Sulfoxide, data=ctrvspcos)

p <- ggplot(ctrvspcos, aes(x=Methionine.Sulfoxide, y=Ratio.Met.148,
color=colores[1:18])) + geom_point(shape=19, size=4, colour=col[1:18])

p <- p + geom_abline(intercept = robust[[1]][1], slope =
robust[[1]][2], colour="black", size=1)

p <- p + ylab("Ratio MetOx-148/Met-148 in ApoAI") + theme(axis.title.y
= element_text(size=16)) + theme(axis.title.y=
element_text(color="#999999", face="bold"))

p <- p + xlab("Methionine Sulfoxide in Serum \n (Intensity Value)") +
theme(axis.title.x  = element_text(size=16)) + theme(axis.title.x=
element_text(color="#999999", face="bold"))

p <- p +  theme(axis.text.x= element_text("Intensity Value"))

p

name <- paste("correlation_metox_metsulfo",".svg")
ggsave(name, dpi=300)


###ahora vamos a hacer scatter plot de large HDL vs meth sufloxide in
serum
###creamos el dataframe

pintar1 <- cbind(pintar[,2],pintar[,6], pintar[,19])
colnames(pintar1) <- c("colores", "Methionine.Sulfoxide", "Large.HDL")
ctrvspcos <- pintar1[1:18,]
ctrvspcos <- as.data.frame(ctrvspcos)
ctrvspcos[,1] <- rep(c("CTR","HIAE"), times=c(8,10))
ctrvspcos[,1] <- as.factor(ctrvspcos[,1])


###modelo lineal

source("C:/ScriptsR/Functions R/pbcor.r")
source("C:/ScriptsR/Functions R/pbos.r")

library(MASS)
library(ggplot2)

model <- pbcor(ctrvspcos[,2],ctrvspcos[,3], beta=0.2)
robust <-rlm(Large.HDL~Methionine.Sulfoxide, data=ctrvspcos)

p <- ggplot(ctrvspcos, aes(x=Methionine.Sulfoxide, y=Large.HDL,
color=colores)) + geom_point(shape=19, size=4,colour=col[1:18])

p <- p + geom_abline(intercept = robust[[1]][1], slope =
robust[[1]][2], colour="black", size=1)
```

```
p <- p + ylab("Large HDL \n (% Relative area)") + theme(axis.title.y
= element_text(size=16)) + theme(axis.title.y=
element_text(color="#999999", face="bold"))

p <- p + xlab("Methionine Sulfoxide in Serum \n (Intensity Value)") +
theme(axis.title.x  = element_text(size=16)) + theme(axis.title.x=
element_text(color="#999999", face="bold"))

p

name <- paste("correlation_large_hdl_metsulfo",".svg")
ggsave(name, dpi=300)
```

FIGURE 5:

```
###Plots del tratamiento

setwd("X:/YANES LAB/USUARIS
(YanesLab)/Sara/Plataforma/Usuaris/Ibañez_02/Final
Data/Julio2014/Figures/Treatment")

###funcion para saber los codigos de los colores por default de ggplot

ggplotColours <- function(n=6, h=c(0, 360) +15){
  if ((diff(h)%%360) < 1) h[2] <- h[2] - 360/n
  hcl(h = (seq(h[1], h[2], length = n)), c = 100, l = 65)
}

codigos <- ggplotColours(n=3)


###voy a pintar la ratio de MetOx
library(ggplot2)

pintar1 <- pintar[,1:3]
col <- rep(c("#F8766D", "#619CFF", "#00BA38"), times=c(8,10,5))
pintar1 <- cbind(pintar1,col)

b <- ggplot(pintar1, aes(x=class, y = pintar1[,3], color = colores)) +
geom_point(size = 4, position = position_jitter(w = 0.1), colour=col)

b <- b + geom_segment(x=0.75, y=mean(pintar1[1:8,3],trim=0.2),
xend=1.25, yend=mean(pintar1[1:8,3],trim=0.2, colour="black"), size=1,
colour="black")

b <- b <- b + geom_segment(x=1.75, y=mean(pintar1[9:18,3],trim=0.2),
xend=2.25, yend=mean(pintar1[9:18,3],trim=0.2, colour="black"),
size=1, colour="black")

b <-  b + geom_segment(x=2.75, y=mean(pintar1[19:23,3],trim=0.2),
xend=3.25, yend=mean(pintar1[19:23,3],trim=0.2, colour="black"),
size=1, colour="black")

b <- b + ggtitle("MetOx-148/Met-148") + theme(plot.title =
element_text(size=18))

b <- b + theme(axis.title.x = element_blank()) + ylab("Ratio MetOx-
148/Met-148 in ApoAI")
```

```
b <- b + theme(axis.text.x  = element_text(size=16, color="#999999"))
+ theme(axis.text.x= element_text(face="bold"))

b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))

b <- b + theme(legend.position="none")

b <- b + coord_fixed(ratio=2/2)

b

ggsave(paste("MetOx",".svg"), dpi=300)

###hago el pca

pca <- prcomp(DATA, scale=TRUE)
class <- pintar1$class
scores <- data.frame(pca$x[, 1:2], class,col)


r <- ggplot(data = scores, aes(x = PC1, y = PC2, colour = colores)) +
geom_point(size =6, colour=col)

r <- r + geom_hline(aes(yintercept = 0), linetype="dashed") +
geom_vline(aes(yintercept = 0), linetype="dashed")

r <- r + xlab("PC1 (46.5%)") + ylab("PC2 (14.8%)")

r <- r + theme(axis.title.x  = element_text(size=16, color="#999999"))
+ theme(axis.title.x= element_text(face="bold"))

r <- r + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))

r

ggsave(paste("PCA",".svg"), dpi=300)

###LOADIGNS DEL PCA

load <- pca$rotation[,1]
loadings <- cbind(colnames(DATA),load)
loadings <- as.data.frame(loadings)
loadings <- cbind(loadings,load)
loadings <- loadings[-2]


loadings <- loadings[-1]

ylabel <- c("Ratio Met-148","Methionine","Methionine Sulfoxide","5-
oxoproline","Taurine",
            "Glu-Cys","Glu-Glu","Glutamate","GSH","GSH/GSSG","Glu-
Taurine","Glycine",
            "VLDL","Large LDL","Small LDL","Large HDL", "Medium HDL",
"Small HDL")

loadings <- cbind(ylabel,loadings)
colnames(loadings) <- c("metabolites","PC1")
metab <- loadings$metabolites
```

```
l <- ggplot(loadings, aes(x=metabolites, y=PC1)) +
geom_bar(data=subset(loadings, PC1>=0),stat="identity") +
  geom_bar(data=subset(loadings, PC1<0),stat="identity")

l <- l + coord_flip()

l <- l + scale_x_discrete(limits=metab)

l <- l + theme(axis.title.y = element_blank())

l

ggsave(paste("loadings",".svg"), dpi=300)

###para pintar methionine y methionine sulfoxide
###tengo q cargar el dataset que esta en la carpeta de gamma
glutamyl...

colores <- rep(c("CTR","HIAE","PIOFLUMET"), times=c(14,12,6))
class <- as.factor(colores)
col <- rep(c("#F8766D", "#619CFF", "#00BA38"), times=c(14,12,6))

pintar <- cbind(class,colores,col, data)
library(ggplot2)

pintar1 <- pintar[,1:2]
pintar1 <- cbind(pintar1,pintar[,13]) ###Methionine
colnames(pintar1[4]) <- "Methionine"

###pinto

b <- ggplot(pintar1, aes(x=class, y = log10(pintar1[,3]), color =
colores)) + geom_point(size = 4, position = position_jitter(w = 0.1),
colour=col)

b <- b + geom_segment(x=0.75, y=mean(log10(pintar1[1:14,3]),trim=0.2),
xend=1.25, yend=mean(log10(pintar1[1:14,3]),trim=0.2, colour="black"),
size=1, colour="black")

b <- b <- b + geom_segment(x=1.75,
y=mean(log10(pintar1[15:26,3]),trim=0.2), xend=2.25,
yend=mean(log10(pintar1[15:26,3]),trim=0.2, colour="black"), size=1,
colour="black")

b <-  b + geom_segment(x=2.75,
y=mean(log10(pintar1[27:32,3]),trim=0.2), xend=3.25,
yend=mean(log10(pintar1[27:32,3]),trim=0.2, colour="black"), size=1,
colour="black")

b <- b + ggtitle("Methionine") + theme(plot.title =
element_text(size=18))

b <- b + theme(axis.title.x = element_blank()) +
ylab("log10(Intensity)")

b <- b + theme(axis.text.x  = element_text(size=16, color="#999999"))
+ theme(axis.text.x= element_text(face="bold"))

b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))
```

```
b <- b + theme(legend.position="none")

b <- b + coord_fixed(ratio=2/1.5)

b

ggsave(paste("Methionine",".svg"), dpi=300)

###Methionine sulfoxide
pintar1 <- pintar[,1:2]
pintar1 <- cbind(pintar1,pintar[,14]) ###Methionine Sulfoxide
colnames(pintar1[3]) <- "Methionine.Sulfoxide"

###pinto

b <- ggplot(pintar1, aes(x=class, y = log10(pintar1[,3]), color =
colores)) + geom_point(size = 4, position = position_jitter(w = 0.1),
colour=col)

b <- b + geom_segment(x=0.75, y=mean(log10(pintar1[1:14,3]),trim=0.2),
xend=1.25, yend=mean(log10(pintar1[1:14,3]),trim=0.2, colour="black"),
size=1, colour="black")

b <- b <- b + geom_segment(x=1.75,
y=mean(log10(pintar1[15:26,3]),trim=0.2), xend=2.25,
yend=mean(log10(pintar1[15:26,3]),trim=0.2, colour="black"), size=1,
colour="black")

b <-  b + geom_segment(x=2.75,
y=mean(log10(pintar1[27:32,3]),trim=0.2), xend=3.25,
yend=mean(log10(pintar1[27:32,3]),trim=0.2, colour="black"), size=1,
colour="black")

b <- b + ggtitle("Methionine Sulfoxide") + theme(plot.title =
element_text(size=18))

b <- b + theme(axis.title.x = element_blank()) +
ylab("log10(Intensity)")

b <- b + theme(axis.text.x  = element_text(size=16, color="#999999"))
+ theme(axis.text.x= element_text(face="bold"))

b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))

b <- b + theme(legend.position="none")

b <- b + coord_fixed(ratio=2/1.5)

b

ggsave(paste("MethionineSulfoxide",".svg"), dpi=300)


###large HDL

pintar1 <- pintar[,1:2]
pintar1 <- cbind(pintar1,pintar[,10]) ###Large HDL
colnames(pintar1[3]) <- "Large.HDL"
###pinto
```

```
b <- ggplot(pintar1, aes(x=class, y = pintar1[,3], color = colores)) +
geom_point(size = 4, position = position_jitter(w = 0.1), colour=col)

b <- b + geom_segment(x=0.75, y=mean(pintar1[1:14,3],trim=0.2),
xend=1.25, yend=mean(pintar1[1:14,3],trim=0.2, colour="black"),
size=1, colour="black")

b <- b <- b + geom_segment(x=1.75, y=mean(pintar1[15:26,3],trim=0.2),
xend=2.25, yend=mean(pintar1[15:26,3],trim=0.2, colour="black"),
size=1, colour="black")

b <-  b + geom_segment(x=2.75, y=mean(pintar1[27:32,3],trim=0.2),
xend=3.25, yend=mean(pintar1[27:32,3],trim=0.2, colour="black"),
size=1, colour="black")

b <- b + ggtitle("Large HDL") + theme(plot.title =
element_text(size=18))

b <- b + theme(axis.title.x = element_blank()) + ylab("% Relative
Area")

b <- b + theme(axis.text.x  = element_text(size=16, color="#999999"))
+ theme(axis.text.x= element_text(face="bold"))

b <- b + theme(axis.title.y  = element_text(size=16)) +
theme(axis.title.y= element_text(color="#999999", face="bold"))

b <- b + theme(legend.position="none")


b

ggsave(paste("largeHDL",".svg"), dpi=300)
```

ARTICLE

Long-term health risks of hyperinsulinaemic androgen excess appear in adolescence with serum markers of oxidative stress and impaired HDL maturation through oxidation of methionine residues in apolipoprotein A1

Sara Samino[1,2], Maria Vinaixa[1,2,3], Marta Diaz[2,5], Antoni Beltran[1,2], Miguel A. Rodríguez[1,2], Roger Mallol[1,2,3], Mercedes Heras[2,4], Anna Cabre[2,4], Lorena Garcia[1], Nuria Canela[1], Xavier Correig[1,2,3], Lluís Masana[2,4], Lourdes Ibañez[2,5*], Oscar Yanes[1,2,3*]

1. Centre for Omic Sciences (COS), Rovira i Virgili University, IISPV, Avinguda Universitat 3, 43204 Reus, Spain; 2. Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), C/ Monforte de Lemos 3-5, 28029 Madrid, Spain; 3. Department of Electronic Engineering, Rovira i Virgili University, Avinguda Països Catalans 26, 43007 Tarragona, Spain; 4.Research Unit on Lipids and Atherosclerosis, Sant Joan University Hospital, Universitat Rovira i Virgili, IISPV, Carrer Sant Llorenç 21, 43201 Reus, Spain; 5.Endocrinology Unit, Hospital Sant Joan de Déu, University of Barcelona, Passeig de Sant Joan de Déu 2, 08950 Esplugues, Barcelona, Spain.

*To whom correspondence should be addressed.

Oscar Yanes, PhD.
Centre for Omic Sciences
Rovira i Virgili University
Avinguda Universitat, 3.
43204 Reus (Spain)
phone: +34 977776617
email: oscar.yanes@urv.cat

Lourdes Ibañez, PhD-MD.
Hospital Sant Joan de Déu
University of Barcelona
Passeig de Sant Joan de Déu, 2.
08950 Esplugues, Barcelona (Spain).
Phone: +34 932804000; 4424
Email: Libanez@hsjdbcn.org

ABSTRACT

Hyperinsulinaemic androgen excess (HIAE) is the phenotypic core for polycystic ovary syndrome (PCOS). HIAE in prepubertal and pubertal girls usually precedes a broader PCOS phenotype in adulthood that is associated with anovulatory infertility and long-term health risks such as type 2 diabetes (T2D), metabolic syndrome and possibly cardiovascular disease (CVD). With the aim of determining the underlying mechanism by which HIAE is associated with these long-term health risks in adulthood, here we use NMR and MS-based metabolomics to compare the lipoprotein profile and the serum metabolome of non-obese adolescents with HIAE, with those in age- and weight-matched control girls. We demonstrate that elevated levels of methionine sulfoxide in HIAE serum are strongly correlated with the degree of oxidation of methionine residues in apolipoprotein-A1. As a result, there is an impaired maturation of HDL reflected in a decline of large HDL particles in girls with HIAE. Remarkably, these metabolic disturbances were partially restored after 18 months of treatment with a combination of insulin sensitizers plus an anti-androgen (pioglitazone, metformin and flutamide). We postulate that hyperandrogenism and hyperinsulinemia are responsible for the oxidative damage in lipoproteins. Taken together, these metabolite changes might also constitute biomarkers of pre-diabetes and metabolic syndrome.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is an evolving concept that, depending on the diagnostic criteria applied and population studied, affects 8-21% of women of reproductive age worldwide (Azziz et al., 2004; Franks, 1995; March et al., 2010). The phenotypic core of PCOS as we know it today, however, consists of women with hyperinsulinemic androgen excess (HIAE) (Barbieri RL et al., 1988; Chang RJ et al., 1983; Dunaif, 1997; Ibanez et al., 2014). HIAE is a hallmark present in both obese and non-obese adolescent girls that continues into adult women with a PCOS-like phenotype (Apter D et al., 1995; Lewy VD et al., 2001). The mechanism by which HIAE is manifested at the early stages of a broader PCOS phenotype is currently under debate (Corbett S and Morin-Papunen L, 2013; de Zegher F and Ibáñez L, 2009; Tsilchorozidou et al., 2004). The “prenatal androgen excess” hypothesis (de Zegher F and Ibáñez L, 2009; Tsilchorozidou et al., 2004) appears to have been refuted in favour

of the “adipose tissue hypertrophy” hypothesis (de Zegher F et al., 2009; Ibanez et al., 2014). The latter represents a mismatch between early adipogenesis and later fat mass, which establishes an individual setpoint determined by a wide range of environmental and genetic factors beyond which further lipid storage is rapidly accompanied by lipotoxicity and metabolic complications, including insulin resistance and androgen excess (Tsilchorozidou et al., 2004). Irrespective of whether hyperandrogenism results from the hyperinsulinaemia of insulin resistance, or vice versa, HIAE in prepubertal and pubertal girls precedes a broader PCOS phenotype in adulthood that is associated with anovulatory infertility and long-term health risks such as metabolic syndrome, type 2 diabetes, (Gambineri A et al., 2012) and possibly cardiovascular disease (Fauser BC, 2012; Talbott EO et al., 2000).

Consequently, adolescence becomes a key period for the study of early biomarkers of long-term health risks, as well as for identifying symptomatic girls who might benefit from early therapeutic interventions, thereby decreasing subsequent metabolic abnormalities related to HIAE (Ibanez et al., 2014). In this regard, a low-dose combination of pioglitazone (an insulin-sensitizer), flutamide (an androgen-receptor blocker) and metformin (an insulin-sensitizer) has been found to normalize the endocrine-metabolic profile of adolescent girls with HIAE, as judged by markers of insulin sensitivity, visceral adiposity, arterial health, low-grade inflammation and menstrual regularity (Ibanez et al., 2011; Ibanez et al., 2010). This combination of insulin sensitizers plus an anti-androgen demonstrates the potential of therapeutic approaches based on a good knowledge of the pathological process of HIAE, in contrast to symptom-directed treatments –for example, with oral contraceptives- that seek at improving the clinical signs (hirsutism and/or acne) and at inducing regular cycles yet persistently anovulatory (Dronavalli and Ehrmann, 2007; Kobaly et al., 2014; Legro RS, 2013).

In this context, here we aimed at identifying as yet undefined metabolic alterations implicated in the early stages of the PCOS phenotype that may facilitate early intervention and thus potentially prevent later metabolic complications.

To achieve this goal, we used mass spectrometry (MS) and nuclear magnetic resonance (NMR)-based metabolomics. Metabolomics enables the characterization of metabolites, the chemical entities that are transformed during metabolism and

provide a functional readout of cellular biochemistry (Patti et al., 2012b). Global metabolite profiling studies are revealing new discoveries linking cellular pathways to biological mechanism, shaping our understanding of cell biology, physiology and medicine (Panopoulos et al., 2012; Patti et al., 2012a; Tomita and Kami, 2012).

In the present study we used serum samples of young and non-obese adolescents with HIAE, and age-, weight- and ethnicity-matched healthy controls. The results revealed that markers for the long-term health risks of HIAE, such as a type 2 diabetes and cardiovascular disease may be already detected early in adolescence with oxidative stress markers impacting HDL functionality through oxidation of apolipoprotein A1. We also demonstrate that a low-dose combination of PioFluMet during 18 months partially restores the levels of oxidative markers to the levels found in healthy girls.

MATERIALS AND METHODS

Study population. The study population consisted of 12 young, non-obese adolescents (age, 16.3±0.4 yr; BMI, 22.8±0.5 Kg/m^2) diagnosed with HIAE and 14 age-, weight- and ethnicity-matched healthy controls. Adolescents with HIAE were recruited at the Endocrinology Unit of the Sant Joan de Déu-Barcelona Children's Hospital, Barcelona (Spain), among those randomized into a clinical study comparing the effects of low-dose combination of PioFluMet with those of a frequently prescribed oral contraceptive [cyproteroneacetate 2 mg + 35 mcg ethynilestradiol for for 21 of 28 d, and placebo for 7 of 28 d] (Ibáñez L et al., 2013). The girls were chosen among those having enough serum sample left to allow for the required assessments. Controls were recruited among age-matched student mates with no clinical signs of androgen excess and normal menstrual cycles.

Inclusion criteria were: 1) hyperinsulinemia, defined as fasting-insulinemia above 15 U/ml and/or hyperinsulinemia on a standard 2-h oral glucose tolerance test, defined as peak insulin levels >150 U/mL and/or mean serum insulin >84 µU/mL; and 2) the presence of both clinical and biochemical androgen excess, as defined by the following: hirsutism score above 8 (Ferriman-Gallwey), amenorrhea (no menses for 3 months) or oligomenorrhea (menstrual cycles longer than 45 d); and high circulating

levels of and/or total testosterone (in the follicular phase (days 3–7) or after 2 months of amenorrhea.

Exclusion criteria were: evidence of anemia, thyroid dysfunction, bleeding disorder, Cushing syndrome, or hyperprolactinemia; glucose intolerance; diabetes mellitus; late-onset adrenal hyperplasia; abnormal electrolytes; abnormal screening of liver or kidney function; use of medication affecting gonadal or adrenal function, or carbohydrate or lipid metabolism. Pregnancy risk was a particular exclusion criterion that was not only taken into account when the study began, but was also maintained throughout the study in the PioFluMet subgroup.

Ethics. This clinical study was registered as ISRCTN12871246 and conducted in Sant Joan de Déu University Hospital (Barcelona, Spain), after approval by the Institutional Review Board of Sant Joan de Déu University Hospital, and after written informed consent by each patient.

Global metabolomics profiling. Untargeted metabolomic analyses on serum samples of HIAE and control girls were performed using two analytical platforms: ^{1}H-NMR and LC-ESI-QTOF; each serum sample was split into two aliquots and run in parallel using the two analytical platforms. For the NMR measurement 250 µL of serum were mixed with 250 µL of phosphate buffer (0.75 mM Na2HPO4 adjusted at pH 7.4, and 20% D2O to provide the field frequency lock). The final solution was transferred to a 5 mm NMR tube and kept refrigerated at 4ºC in the autosampler until the analysis. ^{1}H-NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz using a 5 mm CPTCI triple resonance (^{1}H, ^{13}C, ^{31}P). Three different ^{1}H-NMR pulse experiments were performed for each sample: 1) Nuclear Overhauser Effect Spectroscopy (NOESY)-presaturation sequence to suppress the residual water peak; and 2) Carr-Purcell-Meiboom-Gill sequence (CPMG, spin-spin T2 relaxation filter) with a total time filter of 410 ms to attenuate the signals of serum macro-molecules to a residual level; 20 ppm spectral width and a total of 64 transients collected into 64 k data points.

The second aliquot was used for LC-MS analysis. 30 µL of serum sample was mixed with 120 µL of cold ACN/H2O (1:1) with 1% meta-phosphoric acid (MPA) and 0.1% formic acid (previously filtered). Samples were vortexed vigorously for 30 seconds and stored at –20ºC for 2 hours to enable protein precipitation. Subsequently, samples

were centrifuged 15 minutes at 4ºC and 15000 rpm and the supernetant was transferred to a LC-MS vial. Samples were injected in a UHPLC system (1290 Agilent) coupled to a quadrupole time of flight (QTOF) mass spectrometer (6550 Agilent Technologies) operated in positive electrospray ionization (ESI+) mode. Metabolites were separated using either C18-RP (ACQUITY UPLC HSS T3 1.8 µL, Waters) or HILIC (ACQUITY UPLC BEH 1.7 µL, Waters) chromatography at a flow rate of 0.4 mL/min. The solvent system in C18-RP was A= 0.1% formic in water, and B= 0.1% formic in acetonitrile. The linear gradient elution started at 100% A (time 0–3 min) and finished at 100% B (20–21 min). The solvent system in HILIC was A= 50mM NH4OcA in water, and B= ACN. The linear gradient elution started at 95% B (time 0–2 min) and finished at 55% B (6 min). The injection volume was 2 µL. ESI conditions were gas temperature, 225 °C; drying gas, 13 L min^{-1}; nebulizer, 20 psig; fragmentor, 125 V; and skimmer, 65 V. The instrument was set to acquire over the *m*/*z* range 80–1200 with an acquisition rate of 4 spectra/s. Quality control samples (QC) consisting of pooled serum samples of all patients were used. QC samples were injected before the first study sample and then periodically after five-study samples. Furthermore, real samples were randomized to reduce systematic error associated with instrumental drift. MS/MS was performed in targeted mode, and the instrument was set to acquire over the *m*/*z* range 50–1000, with a default iso width (the width half-maximum of the quadrupole mass bandpass used during MS/MS precursor isolation) of 4 *m*/*z*. The collision energy was fixed at 20 V.

Targeted metabolomics: Relevant metabolites were measured again using an UHPLC system (1290 Agilent) coupled to a triple quadrupole (QqQ) MS (6490 Agilent Technologies) operated in multiple reaction monitoring (MRM) and positive electrospray ionization (ESI+) mode. MRM transitions were: methionine (150→ 56, 61), methionine sulfoxide (166→ 56, 74), Taurine (126→ 41, 85), glutamate (148→56, 84), cysteine-glycine (179→ 59, 116), glutathione (308→ 76, 162), glutathione oxidized (613→ 355, 484), glutamate-glutamate (277→ 84, 130), glutamate-cysteine (251→ 84, 122), 5-oxoproline (130→ 56, 84).

Characterization of lipoprotein classes: ^{1}H NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz (14.1 T). We used the double stimulated echo (DSTE) pulse program with bipolar

gradient pulses and a gradient pulse strength of 25% of the maximum strength of 53.5 Gauss cm^{-1} in order to completely attenuate signals from low molecular weight metabolites. The relaxation delay was 2 s, the free induction decays (FIDs) were collected into 64K complex data points and 32 scans were acquired on each sample. The methyl signal was line-shape fitted using eight 1D Lorentzian functions using a modification of a previously reported protocol (Mallol et al., 2013). According to the NMR-derived lipoprotein sizes previously described, functions 2 to 8 were associated with 1 VLDL, 2 LDL, and 4 HDL lipoprotein subclasses, respectively. For simplification, functions 7 to 8 were grouped to obtain the small HDL subclass.

Data analysis and statistical methods. The acquired CPMG NMR spectra were phased, baseline corrected and referenced to the chemical shift of the α-glucose anomeric proton doublet at 5.23 ppm. Pure compound references in BBioref AMIX (Bruker), HMDB and Chenomx databases were used for metabolite identification. After baseline correction, intensities of each 1H-NMR regions identified in the CPMG 1D-NMR spectra were integrated for each sample entering the study using the AMIX 3.8 software package (Bruker, GmBH).

LC-MS (RP-C18 and HILIC ESI+ mode) data were processed using the XCMS software (Smith et al., 2006) (version 1.38.0) to detect and align features. A feature is defined as a molecular entity with a unique m/z and a specific retention time. XCMS analysis of these data provided a matrix containing the retention time, m/z value, and integrated peak area of each feature for every serum sample. The tab-separated text files containing LC-MS data were imported into R software where QC samples were used to filter analytical variation as previously reported (Vinaixa et al., 2012). Then univariate statistical analysis was performed using robust statistics (Yuen-Welch's t-test). Differentially regulated metabolites (fold>1.5) that passed our statistical criteria (p-value<0.01) were characterized by LC-qTOF MS/MS and identified using Metlin database or pure standards purchased on our lab.

SDS-PAGE and trypsin digestion of apolipoprotein A1. Protein precipitation was carried out adding 10% of pure trichloroacetic acid (TCA) to 10 µL of serum. Samples were vortexed vigorously and incubated on ice for 1 hour. Samples were then centrifuged at 4ºC and 14.000 rpm for 15 minutes and the supernatants were discarded. 800 µL of cold acetone (-20ºC) were added to the pellet and proteins were

suspended and incubated overnight (-20ºC). Samples were centrifuged and the supernatant discarded again. This step was repeated and the pellet was air-dried. Pellet was resuspended with 100 µL of urea (7M), thiourea (2M) and CHAPS (4%) buffer. 10 µL of this solution were added to 30 µL of Laemmli Buffer (4x). 40 µL of such solution was applied to a home-made 12% acrylamide/bis-acrylamide SDS-PAGE gel.

Proteins were Coomassie stained. The band of interest corresponding to Apo-A1 (MW: 28,1 kDa) was manually excised from 1D SDS-PAGE gels, destained and washed with 25 mM AmBic for 15 min followed by a wash with acetonitrile for 15 min. These washes were twice repeated and samples were finally dehydrated with 100% acetonitrile and dried in a Speed-Vack concentrator. Apo-A1 was cysteine carbamidomethylated by placing the dried gel at 56 ºC for 1 h in a reducing solution containing 10 mM DTT and 50mM ammonium bicarbonate. Alkylation of the cysteines was achieved by incubation of the gel for 30 min in the dark with 55 mM iodoacetamide in 25 mM ammonium bicarbonate buffer. Gel pieces were alternately washed with 25 mM AmBic and 25 mM AmBic with acetonitrile, and finally dehydrated with 100% acetonitrile and dried under vacuum. Gel pieces were incubated with 12.5 ng/µl sequencing grade trypsin (Roche Molecular Biochemicals) in 25 mM AmBic overnight at 37°C. After digestion, the supernatants were separated. Peptides were extracted from the gel pieces into 50% ACN, 0.1% trifluoroacetic acid. For each extraction, samples were incubated for 10 min in an orbital shaker. All extracts were pooled and the volume reduced using a vacuum concentrator. In order to obtain a suitable sample for mass spectrometry analysis the pellet was resuspended in 25µL of 0.1% TFA/water, desalted and concentrated using C18 ZipTips (Millipore). Tryptic peptides were sequentially eluted with 5 µL of 70% acetonitrile with 0.1% TFA in water.

MALDI-TOF MS analysis of Apo-A1. Samples were spotted on the MALDI plate following the dried-droplet method. Briefly, 1 µl of the reconstituted in-gel digest sample was spotted on a BigAnchorChip target plate (BrukerDaltonics), followed by 1 µl of matrix (10 mg/ml α-cyano-4-hydroxycinnamic acid matrix (BrukerDaltonics) in 50% ACN, 0.1% trifluoroacetic). Sample and matrix mixture was dried at room temperature. Mass spectra were obtained on an UltrafleXtreme (BrukerDaltonics, Bremen, Germany) matrix-assisted laser desorption ionization–tandem time of flight

(MALDI-TOF/TOF) mass spectrometer. Mass spectra were recorded in positive ionization reflectron mode in the mass range of 700–3500 Da. Operating conditions were as follows: ion source 1 = 25.00 kV, ion source 2 = 24.40 kV, lens voltage = 8.50 kV, reflector voltage = 26.45 kV, optimized pulsed ion extraction time = 130 ns, matrix suppression = 500 Da. 1500 single-shot spectra were accumulated by recording 50-shot spectra at 10 random positions using fixed laser attenuation. Mass spectra were externally calibrated using a standard peptide mixture (Bruker); calibration was considered good when a value below 1 ppm was obtained.

Peptide mass fingerprinting of Apo-A1 and quantization of the ratio MetOx-148/Met-148. ProteinScape software (Bruker) supported by the Mascot search engine (Matrix Science) was used with the following parameters: SWISS-PROT non-redundant database filtered by *homo sapiens* taxonomy, two missed cleavage permission, 50-ppm measurement tolerance.

Carbamidomethylation of cysteines was set as a fixed modification and methionine oxidation was set as a variable modification. Positive identifications were accepted with a Mascot score higher than that corresponding to a *P* value of 0.05. The quantification of the ratio MetOx-148/Met-148 was performed using the intensity of the peptide K.LSPLGEEMS.D (Flex Analysis, Bruker).

RESULTS

Endocrine and metabolic alterations associated with HIAE in young, non-obese girls.

Endocrine and metabolic parameters were compared between a group of 12 young, non-obese girls with HIAE and 14 age-, weight- and ethnicity-matched healthy controls. Table 1 shows mean values and the standard error of the mean for each variable. As expected by definition, endocrine alterations in adolescent girls with HIAE relative to healthy controls included increased significant ($p \leq 0.05$) serum levels of insulin, testosterone, dehydroepiandrosterone sulfate (DHEAS) and leptin. On the other hand, metabolic values such as the fasting glucose to insulin ratio, total cholesterol were increased in HIAE girls and HDL-cholesterol was decreased in HIAE girls. Levels of superoxide dismutase (SOD), a key antioxidant enzyme protecting the cells, were also decreased in PCOS. Two-hour glucose levels on an oral glucose

tolerance test (oGTT), triglycerides and LDL-cholesterol levels did not show differences from healthy control adolescents.

Thus, hyperinsulinemia with normal fasting glucose levels in HIAE girls may reflect insulin resistance, as suggested by the increased ratio of glucose to insulin. However, we could disclose no evidences of impaired glucose tolerance, hyperglycemia or hypertriglyceridemia at this early age.

Table 1. **Physiological and biochemical parameters**.

	CTR	PCOS	p-value
Age (yr)	17.2 ± 0.4	16.3 ± 0.4	0.15
BW SDS	0.3 ± 0.1	-0.2 ± 0.4	0.58
Wt (kg)	58.8 ± 1.8	58.2 ± 1.2	0.92
Ht (cm)	163.9 ± 1.3	160.0 ± 1.5	0.06
BMI (Kg/m^2)	21.8 ± 0.6	22.8 ± 0.5	0.23
BMI SDS	0.2 ± 0.2	0.5 ± 0.2	0.16
WBC (cell/mm3)	7.3 ± 0.3	7.8 ± 0.5	0.26
Neutrophils (x1000/mm3)	4.1 ± 0.3	4.4 ± 0.5	0.63
Lymphocytes (x1000/mm3)	2.2 ± 0.1	2.5 ± 0.22	0.41
N/L (ratio)	1.9 ± 0.2	2.0 ± 0.4	0.51
AST (µL/L)	16.6 ± 0.7	16.8 ± 1.6	0.59
ALT (µL/L)	13.6 ± 1.0	13.2 ± 1.1	0.83
oGTT (mg/dL)	89.1 ± 1.5	85.4 ± 2.0	0.14
Insulin (µU/mL)	3.5 ± 0.6	10.3 ± 1.6	**0.01**
G/I ratio	32.8 ±3.6	11.2 ± 1.9	**0.0004**
Total Cholesterol	143.9 ± 5.9	145.9 ± 6.8	0.75
HDL-cholesterol	52.6 ± 2.3	51.9 ± 3.3	0.77
LDL-cholesterol	80.5 ± 5.4	78.7 ± 4.5	0.88
Triglycerides	53.4 ± 3.6	76.8 ± 16.5	0.57
Testosterone (ng/dL)	32 ± 2.4	64.2 ± 10.2	**0.05**
DHEAS (µg/dL)	222.1 ± 27.8	280.8 ± 31.5	**0.03**
Leptin (ng/mL)	13.9 ± 2.3	20.9 ± 2.7	**0.05**
usCRP (mg/L)	0.7 ± 0.2	1.1 ± 0.2	0.14
SOD (U/mL)	6.1 ± 0.3	5.4 ± 0.2	**0.03**

Data are represented as mean±standard error of the mean (SEM). P-values are calculated from a robust Yuen-Welch's t-test. BW SDS: Birth weight standard desviation, Wt: weight, Ht: height, BMI: body mass index, BMI SDS: body mass index standard desviation, WBC: white blood cells, N/L: Neutrophils/Lymphocytes, AST: aspartate transaminase, ALT: alanine aminotransferase, oGTT: oral glucose tolerance test, G/I ratio: Glucose/insuline ratio, DHEAS: dehydroepiandrosterone sulfate, usCRP: ultrasensitive-C-Reactive Protein, SOD: superoxid dismutase.

Alteration of the VLDL, LDL and HDL serum profile in non-obese adolescents with HIAE

Dyslipidemia is a characteristic lipid abnormality in PCOS and it is an important risk factor linked to metabolic syndrome, T2D and CVD. Here we studied the lipoprotein profile of adolescent girls with HIAE beyond standard measurement of cholesterol content of lipoproteins, by using an advanced lipoprotein analysis based on nuclear magnetic resonance (NMR) spectroscopy (Mallol et al., 2013). ^{1}H-NMR allowed us to characterize the size and relative abundance of lipoprotein particles in serum. Briefly, depending on the size of the lipoprotein particle, the methyl moieties of the lipids in lipoproteins resonate at slightly different frequencies, the smaller particles resonating at lower frequencies (Figure 1a). Our NMR-derived lipoprotein subclasses were defined as VLDL, large LDL, small LDL, large HDL, medium HDL, and small HDL.

Relative levels of VLDL, small LDL and large LDL were significantly increased in HIAE relative to control girls. In contrast, the relative abundance of large, medium and small HDL subclasses were decreased in HIAE, with the greatest decline though associated with large HDL (Figure 1b). Table 2 shows the differences in serum lipoproteins between adolescent girls with HIAE and healthy controls. Thus, the NMR-based characterization of lipoprotein subclasses revealed a dyslipidemic profile in girls with HIAE, which is similar to the one observed in the metabolic syndrome and its associated pathologies; T2D and CVD (Dokras, 2013).

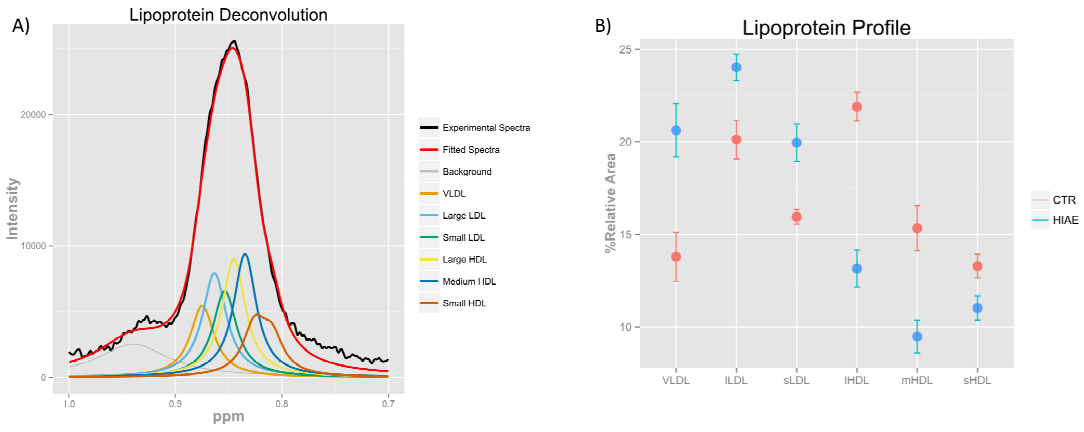

Figure 1. Lipoprotein profile measured by NMR spectroscopy. (A) Bipolar LED pulse sequence ^{1}H NMR spectra of a HIAE serum showing the fitting of the methyl band using the seven Lorentzian functions derived from our previously described methodology (B) Row-wise normalized areas showed as mean±sem. VLDL: very low-density lipoprotein, lLDL: large low-density lipoprotein, sLDL: small low-

density lipoprotein, lHDL: large high-density lipoprotein, mHDL: medium high-density lipoprotein, sHDL: small high-density lipoprotein

Table 2. NMR lipoprotein profile in serum samples.

	% variation	p-value
VLDL	50	0.012
Large LDL	20	0.0264
Small LDL	31	0.0245
Large HDL	-63	0.00001
Medium HDL	-52	0.0058
Small HDL	-20	0.02

Percentage of variation of each lipoprotein subclass in HIAE with respect to control girls, and p-values (Yuen-Welch's t-test).

Metabolomics reveals altered glutathione biosynthesis via the γ-glutamyl cycle, and elevated levels of methione sulfoxide in HIAE.

The relative abundance of serum metabolites in girls with HIAE and healthy controls were compared using an untargeted metabolomic approach based on reverse phase (RP)-C18 and HILIC chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry (LC ESI-QTOF-MS) in positive ionization mode in combination with ^{1}H-NMR metabolite profiling (Supplementary Table 1). Our MS-based platform enabled us to observe greater than 38.000 metabolite features defined as molecular entities with a unique mass/charge and retention time value, after the analytical variability had been corrected. Only the integrated areas of those metabolite features above 5,000 spectral counts in at least one of the two groups were considered for quantification (see the Methods section for further details). We next structurally identified metabolites based on accurate mass and MS/MS data that differed between HIAE and healthy controls. We identified 7 and 7 metabolites using RP-C18 and HILIC, respectively, that showed a greater than 50% of variation with a p-value < 0.01 (Yuen-Welch's t-test) (Supplementary Table 2). We found that levels of γ-glutamyl dipeptides were higher in serum of adolescent girls with HIAE compared to healthy controls. γ-glutamyl dipeptides are involved in the γ-glutamyl cycle that transports amino acids into cells and plays a key role for the synthesis and degradation of glutathione, which is a key reducing agent that protects from oxidative damage. Furthermore, the NMR analysis revealed lower levels of glycine in serum samples of

HIAE girls. Glycine can be added to the C-terminal of γ-glutamylcysteine (Glu-Cys) via the enzyme glutathione synthetase to produce reduced glutathione. Finally, lower levels of methionine in serum samples of HIAE were associated with greater levels of methionine sulfoxide in these patients, indicating increased oxidation of this aminoacid.

To confirm that the biosynthesis of glutathione via the γ-glutamyl cycle was really altered in girls with HIAE, we quantified reduced glutathione and the ratio of reduced to oxidized glutathione (GSH/GSSG) using triple-quadrupole mass spectrometry (QqQ MS) and multiple reaction monitoring (MRM). At the same time, we requantified most significant metabolites described in Supplementary Table 1 using pure standards (when commercially available), and quantified the additional γ-glutamyl dipeptide Glu-Cys not detected by our untargeted metabolomic approach (Table 3).

Table 3.Targeted metabolomics.

	% variation	p-value
Methionine	-344	0.00005
Methionine Sulfoxide	40	0.0018
5-oxoproline	39	0.005
Taurine	28	0.015
Glu-Cys	37	0.0002
Glu-Glu	35	0.027
Glutamate	33	0.004
GSH	34	0.036
GSH/GSSG	29	0.023
Glu-Taurine*	73	0.0000003
Glu-Gly*	98	0.00015
Glycine	-20	0.031

*Percentage of variation and p-values (Yuen-Welch's t-test) of metabolites. Negative and positive values indicate lower and higher levels, respectively, in girls with HIAE relative to healthy controls. Glu: glutamate, cys: cysteine, GSH: glutathione, Gly: glycine, GSH/GSSG: glutathione/oxidated glutathione. *Glu-taurine and *Glu-Gly could not be quantified by LC-QqQ in MRM mode due to the lack of pure standards and reported values are from LC-qTOF MS. Gly's value is from NMR.*

Glutamyl dipeptides (Glu-Glu, Glu-Tau, Glu-Cys and Glu-Gly), amino acids (glutamate, 5-oxoproline and taurine), GSH and the GSH/GSSG ratio were significantly elevated in HIAE. Therefore, the targeted analysis validated our untargeted metabolomic results and proved the activation of the γ-glutamyl cycle in HIAE to synthesize reducing agents in the form of GSH (Figure 2).

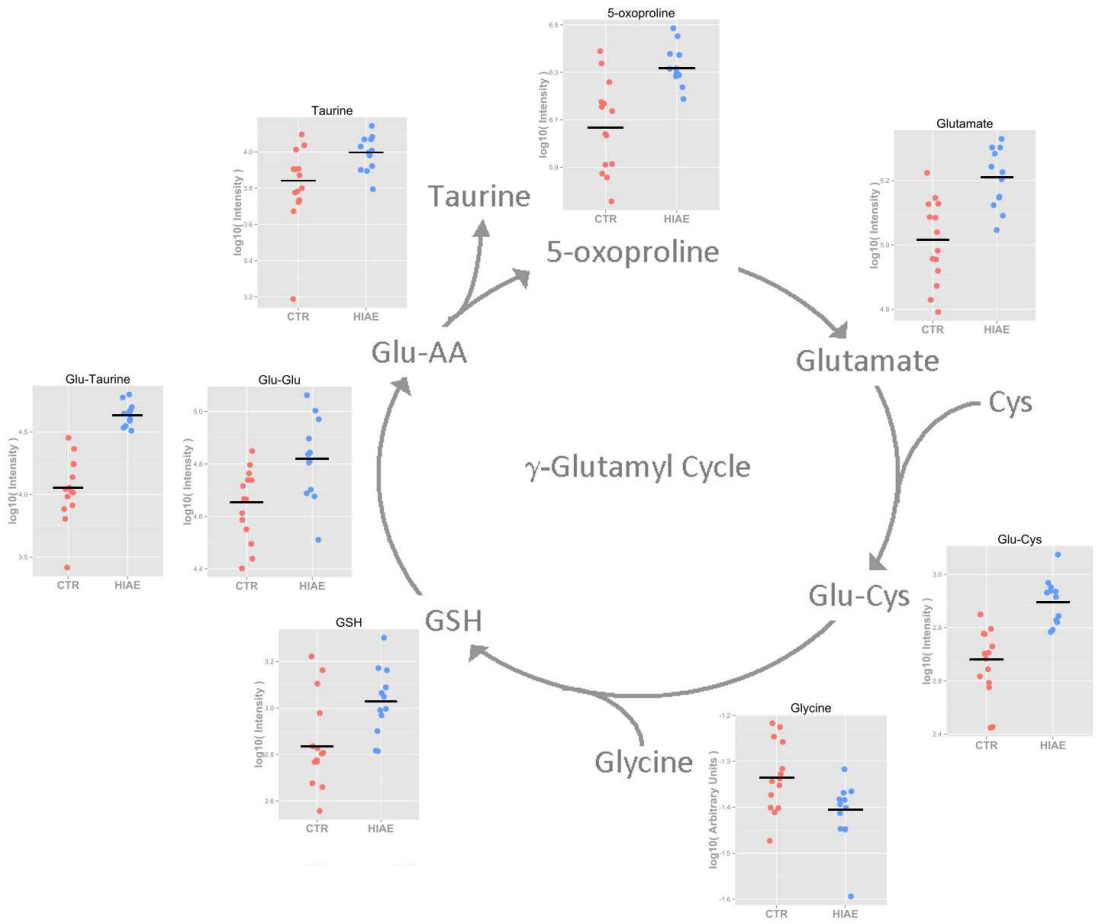

Figure 2. Identified and quantified metabolites involved in the γ-glutamyl cycle. The scatter plots show the relative abundance (i.e., intensity) of individual metabolites in controls and HIAE samples and trimmed mean (control in red and HIAE in blue).

Moreover, the trend that we observed between methionine and methionine sulfoxide (MetOx) was confirmed, that is, a drop in methione levels (Figure 3A) comes with a significant increase in MetOx (Figure 3B) in adolescent girls with HIAE.

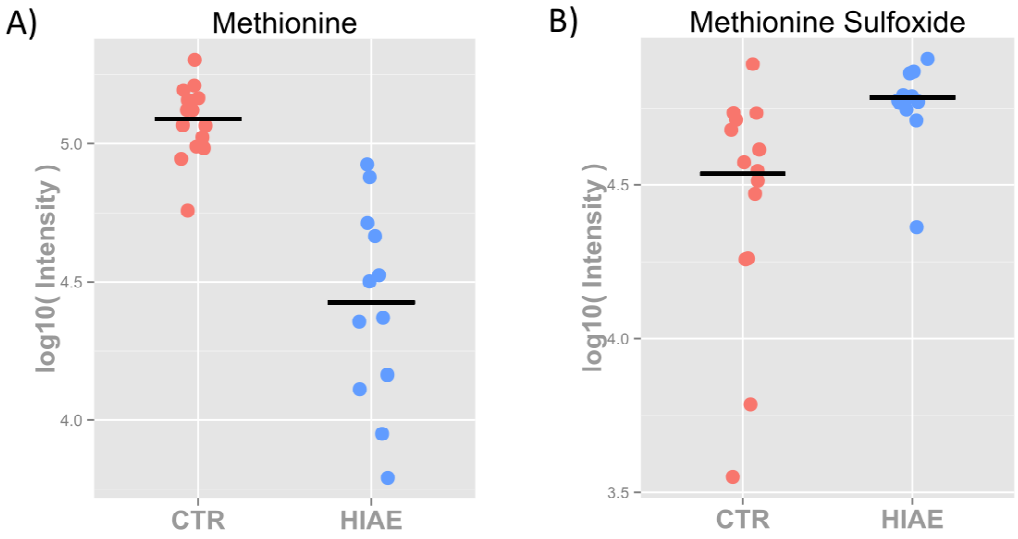

Figure 3.Targeted analysis of methionine and methionine sulfoxide. The scatter plots show the relative abundance of methionine (A) and methionine sulfoxide (B) in controls and HIAE samples and trimmed mean (control in red and HIAE in blue).

Methionine residues in proteins can be readily oxidized by reactive oxygen species to MetOx (Stadtman et al., 2003). In the context of our study, oxidation of methione residues in apolipoprotein A1 (Apo-A1) has been associated with impaired reverse cholesterol transport by HDL, and consequently, impaired maturation of HDL particles (Shao et al., 2008). This could partly explain the lower percentage of HDL particles, and more specifically, of large HDL particles in girls with HIAE relative to healthy controls (Figure 1). Since Apo-A1 is the major protein component of HDL and one of the most abundant proteins in human serum (Fisher et al., 2012), we hypothesized that girls with HIAE show greater oxidation of methionine residues in Apo-A1 relative to healthy girls.

Quantitative analysis of the MetOx/Met ratio in Apo-A1 by MALDI-TOF MS

To test the hypothesis that girls with HIAE suffer from increased oxidation of methionine residues in Apo-A1, we measured the ratio of MetOx/Met in Apo-A1 using SDS-PAGE and MALDI-TOF MS. In particular, we focused on a single methionine residue of Apo-A1, Met-148, the oxidation of which has been associated with loss of LCAT activity, a critical early step in reverse cholesterol transport (Sorci-Thomas and Thomas, 2002).

Briefly, we separated serum proteins of 10 non-obese adolescent girls with HIAE and 6 healthy controls using 1D SDS-PAGE, and the bands of interest corresponding to the molecular weight of Apo-A1 (28.1 kDa) were manually excised from the gel electrophoresis. After in-gel digestion with trypsin, peptides were recovered for MALDI-TOF MS analysis (see the Methods sections for details). Apo-A1 sequence coverage was 61% on average and MetOx-148/Met-148 was calculated from the ratio of peak intensity of peptides m/z 1047.51 (sequence K.LSPLGEEMR.D) and m/z 1411.67 of Apo-A1. We recognize that MetOx could partly be generated during sample preparation and analysis as an artifact. However, since we measured the ratio MetOx-148/Met-148 of Apo-A1 separately for each sample, this should not alter the relative differences between HIAE and control group.

The ratio MetOx-148/Met-148 in Apo-A1 was significantly increased in girls with HIAE compared with healthy controls (Figure 4a). Furthermore, we found a positive and statistically significant correlation between the ratio MetOx-148/Met-148 in Apo-A1 and free methionine sulfoxide in serum (Figure 4b).

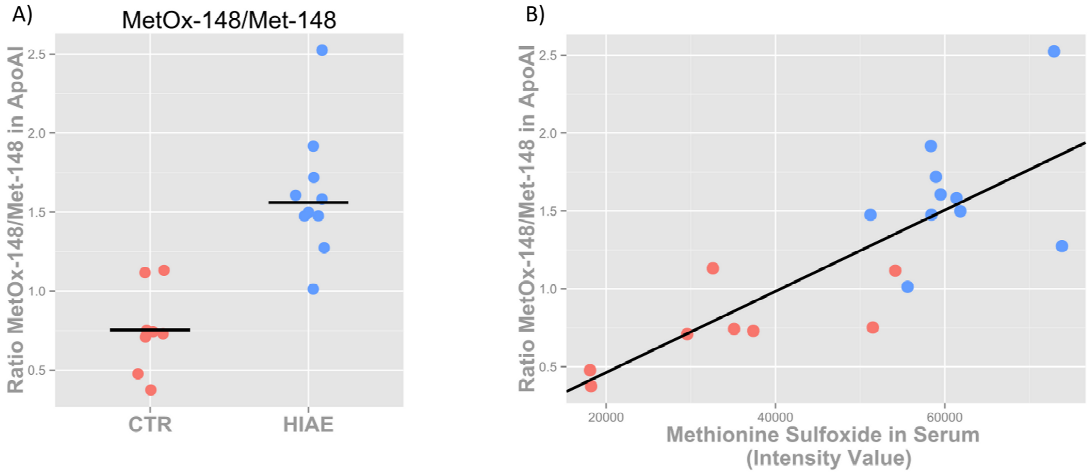

Figure 4. **Apo-AI measurement**. (A) Ratio of MetOx-148/Met-148 in apo-AI calculated from the intensity of the peptide K.LSPLGEEMR.D in Apo-A1. (B) Statistically significant correlation (p=6.95E-05 and r=0.8) between free methionine sulfoxide in serum and the oxidation state of apo-AI on the basis of the ratio of MetOx-148/Met-148 by MALDI-TOF MS. Control in red and HIAE in blue.

Therefore, levels of methionine sulfoxide in serum are directly linked to the degree of oxidation of the Met-148 residue in Apo-A1, which could reflect turnover and proteolytic degradation of Apo-A1 proteins.

Finally, there is also a negative correlation between the number of large (i.e., mature) HDL particles and methionine sulfoxide in serum (Figure 5), which reinforces our hypothesis that levels of methionine sulfoxide in serum reflect HDL oxidation, and indirectly, impaired maturation of HDL particles.

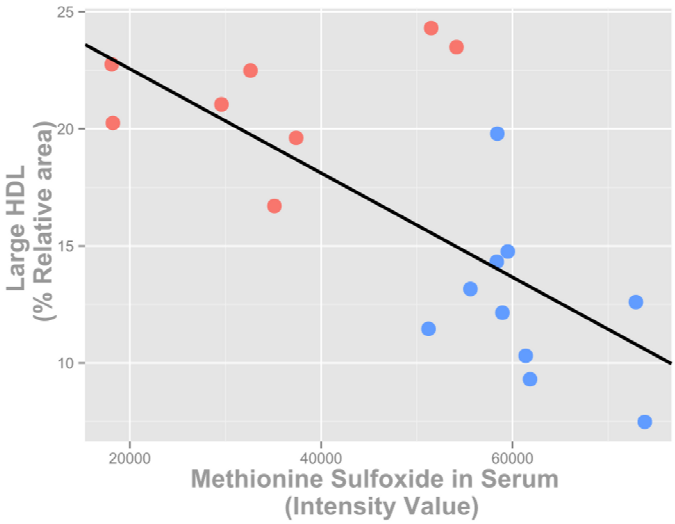

Figure 5. Correlation results. Statistically significant negative correlation (p=1.13E-03 and r=-0.7) between free levels of methionine sulfoxide in serum and the percentage of large HDL particles. Control in red and HIAE in blue.

Metabolic changes after 18 months of PioFluMet polytherapy in non-obese girls with HIAE

Ibañez and colleagues demonstrated that a low-dose combination of PioFluMet proved to be more beneficial than oral contraceptives in regulating endocrine-metabolic parameters, decreasing inflammation and visceral fat, and in improving markers of cardiovascular health (Ibanez et al., 2007).

Here, we also measured the novel metabolic markers described above in girls with HIAE, after 18 months of PioFluMet polytherapy. Specifically, metabolites involved in the biosynthesis of GSH via the γ-glutamyl cycle, the size and relative abundance of lipoprotein particles, and the degree of oxidation of Apo-A1 by means of the MetOx-148/Met-148 ratio were monitored using NMR and MS in the serum of 6 patients after 18 months of PioFluMet treatment.

The abundance of individual markers was scaled to unit variance and projected using an unsupervised principal component analysis (PCA) (Figure 6A). The scores plot reveals two clusters along PC1 (~47% of the variance) corresponding to adolescent girls

with HIAE on the one hand, and healthy controls and HIAE girls treated with PioFluMet on the other. This distribution indicates that the metabolic state of non-obese girls with HIAE after 18 months of PioFluMet treatment more closely resemble the state of healthy girls, suggesting an overall improvement of the metabolic conditions of HIAE. To interpret the pattern displayed in the scores plot, Figure 6B shows a loading bar plot of the PCA using each individual metabolic marker measured. The relative abundance of large HDL particles and levels of free methionine in serum are the two largest contributing variables to positive values in PC1. After the treatment with PioFluMet, the levels of methionine and large HDL particles recovered to the levels found in healthy girls (Figure 6c and 6d). Similarly, the levels of methionine sulfoxide and the oxidation of Apo-A1 in the Met-148 residue decreased after the treatment, reaching similar levels seen in healthy girls (Figure 6e and 6f).

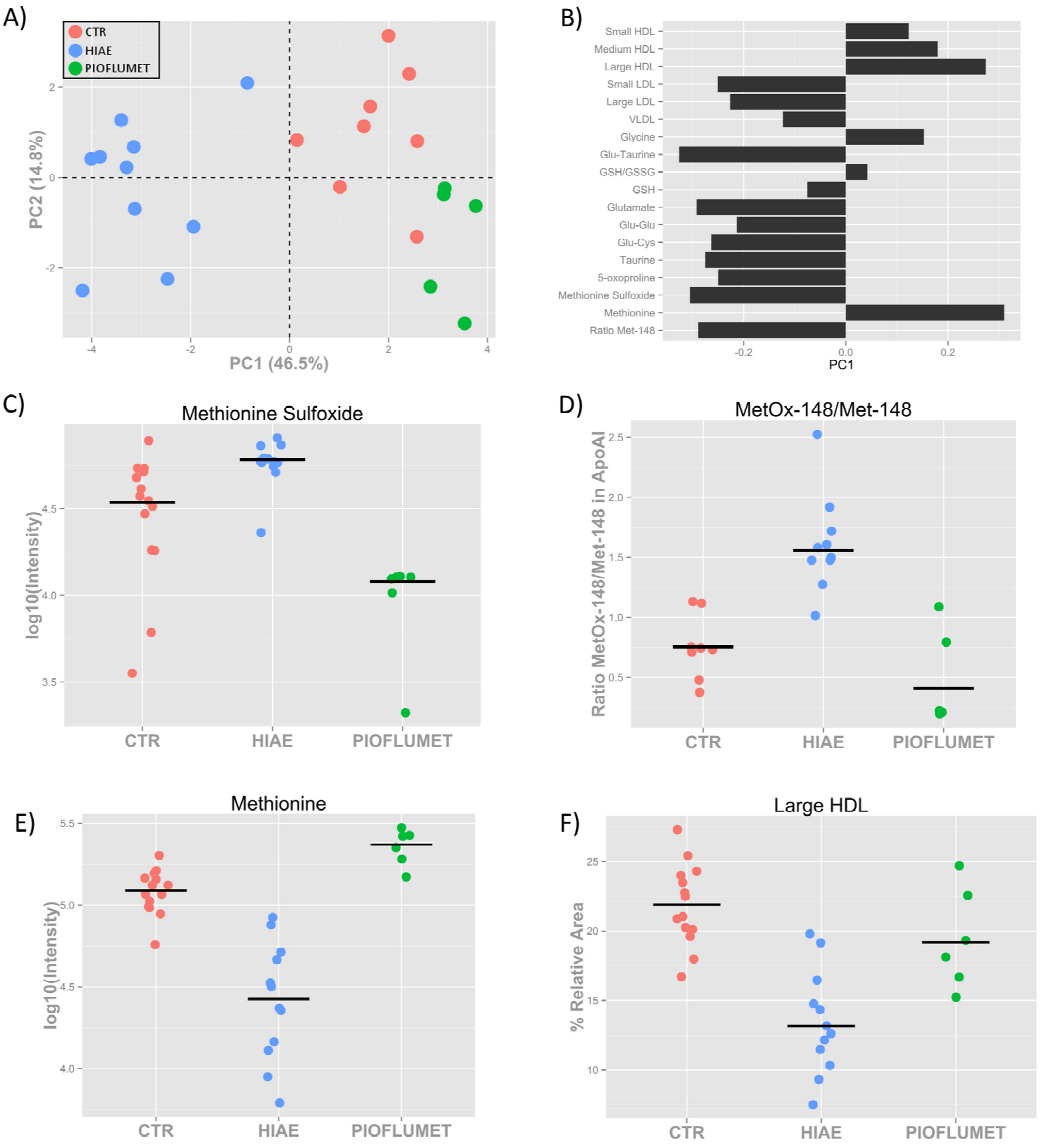

Figure 6. Metabolic changes after 18 months of PioFluMet polytherapy. (A) PC1/PC2 scatter scores plot and (B) PC1 loading bar plot of PCA showing all the metabolites measured in HIAE patients after the treatment. (C) Relative intensity of free methionine in serum. (D) Percentage of large HDL particles in serum (E) Relative intensity of free methionine sulfoxide in serum. (F) Ratio of MetOx-148/Met-148 in apo-AI calculated from the intensity of the pepide K.LSPLGEEMR.D.

DISCUSSION

A better insight into hyperinsulinemic androgen excess (HIAE) in non-obese adolescents with a mild PCOS phenotype may contribute to the elucidation of early origins of PCOS and the causes of its association with an increased incidence of pre-diabetic states (Apridonidze et al., 2005) and possibly cardiovascular events (Sukalich and Guzick, 2003). Ultimately, novel insights should sharpen the perspective of early

PCOS prevention (Ibanez et al., 2014).

The PCOS phenotype has been associated with an imbalance between pro-oxidant and anti-oxidant mechanisms (Blair SA et al., 2013; Macut D et al., 2013), however, the underlying causes of this association remains unclear (Murri M et al., 2013). In our study the pro-oxidant events in adolescent girls with HIAE include low levels of the enzyme SOD, in accordance to previous studies (Seleem et al., 2014), increased oxidation of methionine residues in Apo-AI and accumulation of the oxidative marker methionine sulfoxide in HIAE serum. Interestingly, girls with HIAE appear to activate a compensatory and anti-oxidant mechanism that aims to regulate their redox status by synthesizing glutathione (GSH) through the γ-glutamyl cycle (Zhang et al., 2005). This is reflected in greater levels of the anti-oxidant GSH and the GSH/GSSG ratio in HIAE. Redox imbalances have been also described in atherosclerotic diseases (Stocker and Keaney, 2005), diabetes (Martitim et al., 2003; Sundaram et al., 1996) and metabolic syndrome (Roberts and Sindhu, 2009).

Oxidative modifications are considered an initial step in lipoprotein conversion into more atherogenic particles (Kaysen and Eiserich, 2004; Shao et al., 2008), particularly in LDL particles (Jialal and Devaraj, 1996). In contrast, HDL cholesterol is generally associated with lower risk of cardiovascular disease (Valkenburg et al., 2008), metabolic syndrome and type 2 diabetes (von Eckardstein A and Widmann C, 2014). Concretely, HDL cholesterol is generally associated with atheroprotective properties, which include mediation of reverse cholesterol transport (Rosenson et al., 2013). The functional status of HDL is closely linked to its primary protein component, Apo-AI, an abundant apolipoprotein whose plasma concentrations are inversely correlated with the incidence of coronary artery disease (Borja et al., 2013). Notwithstanding, when Rajkhowa et al. measured the concentration of apo-AI in HDL particles in control and PCOS women, no change in the content of apoA-I was observed (Rajkhowa et al., 1997).

Alternatively, oxidation of methionine residues in apo-AI has been shown to impair reverse cholesterol transport by HDL (Shao et al., 2008). Specifically, the oxidation of Met-148 in apo-AI impairs apo-AI's ability to activate lecithin cholesterol acyltransferase (LCAT) (Shao et al., 2008). LCAT is the enzyme responsible for transforming nascent HDL into spherical HDL particles containing a central

hydrophobic core of cholesteryl esters and an outer layer composed of apo-AI and phospholipids. Our NMR lipoprotein study shows lower percentage of large (i.e., mature) HDL particles in HIAE relative to control girls. This circumstance, together with the increased oxidation of Met-148 residues in apo-AI of HIAE girls suggests impaired HDL function and, consequently, reduced ability to form mature HDL particles.

It has recently been shown that hyperandrogenemia and glucose ingestion induce oxidative stress and that women with androgen excess fail to suppress the release of inflamation markers after glucose ingestion (González F et al., 2012; González F et al., 2006; González et al., 2014). In the current period of caloric abundance and chronically positive energy balance for most adolescents (Corbett SJ et al., 2009), we postulate that hyperandrogenism in combination with excessive carbohydrate intake in a likely scenario of adipose tissue hypertrophy, may cause oxidative stress and oxidation of lipoprotein particles, resulting ultimately in impaired lipoprotein function and dyslipidemia. In our view, therefore, HIAE leads to alterations in lipoprotein metabolism, and both events precede other metabolic complications associated with the increased risk for metabolic syndrome (Anderson et al., 2014; Sharpless, 2003) and diabetes, including hypertriglyceridemia, impaired fasting glycemia and impaired glucose tolerance (Figure 7).

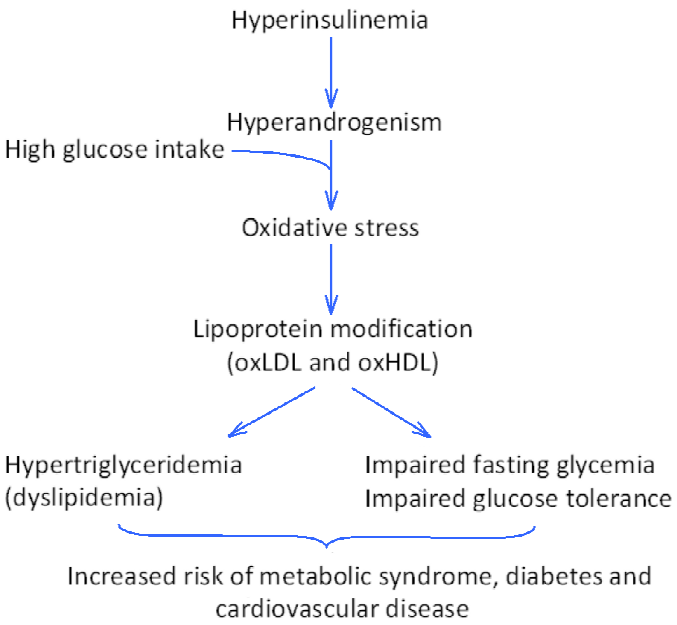

Figure 7. Schematic representation of the underlying pathway by which HIAE is associated with long-term health risks, namely metabolic syndrome, diabetes and possibly, cardiovascular disease.

It is also tempting to speculate that the triad of methionine sulfoxide in serum, concentration of HDL subclasses (i.e., small, medium, large) and the ratio of methionine oxidation in apo-AI may represent novel biomarkers of pre-diabetes and metabolic syndrome. Further work is needed to study this combination of biomarkers in the general population.

Finally, early treatment with low-dose PioFluMet proves again its efficacy on androgen excess and hyperinsulinemia (Ibáñez L et al., 2013), while improving the lipoprotein profile and most oxidative stress markers of non-obese adolescents with HIAE previously unknown. These observations extend further the benefits of therapies leading to a more physiological condition in adolescents with HIAE, questioning the rationale for applying symptom-directed therapies that might potentially impact on later co-morbidities (Beaber et al., 2014).

REFERENCES

Anderson, S.G., Dunn, W.B., Banerjee, M., Brown, M., Broadhurst, D.I., Goodacre, R., Cooper, G.J.S., Kell, D.B. and mail, J.K.C. (2014) Evidence That Multiple Defects in Lipid Regulation Occur before Hyperglycemia during the Prodrome of Type-2 Diabetes *PLoS ONE*, **9**, e103217.

Apridonidze, T., Essah, P.A., Iuorno, M.J. and Nestler, J.E. (2005) Prevalence and Characteristics of the Metabolic Syndrome in Women with Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology & Metabolism*, **90**, 1929-1935.

Apter D, Bützow T, Laughlin GA and SS, Y. (1995) Metabolic features of polycystic ovary syndrome are found in adolescent girls with hyperandrogenism. . *Journal of Clinical Endocrinology & Metabolism*, **80**, 2966-2973.

Azziz, R., Woods, K.S., Reyna, R., Key, T.J., Knochenhauer, E.S. and Yildiz, B.O. (2004) The Prevalence and Features of the Polycystic Ovary Syndrome in an Unselected Population. *The Journal of Clinical Endocrinology & Metabolism*, **89**, 2745-2749.

Barbieri RL, Smith S and KJ, R. (1988) The role of hyperinsulinemia in the pathogenesis of ovarian hyperandrogenism. *Fertility and Sterility*, **50**, 197–212.

Beaber, E.F., Buist, D.S.M., Barlow, W.E., Malone, K.E., Reed, S.D. and Li, C.I. (2014) Recent Oral Contraceptive Use by Formulation and Breast Cancer Risk among Women 20 to 49 Years of Age. *Cancer Research*, **74**, 4078–4089.

Blair SA, Kyaw-Tun T, Young IS, Phelan NA, Gibney J and J., M. (2013) Oxidative stress and inflammation in lean and obese subjects with polycystic ovary syndrome. *The Journal of reproductive medicine*, **58**, 107-114.

Borja, M.S., Zhao, L., Hammerson, B., Tang, C., Yang, R., Carson, N., Fernando, G., Liu, X., Budamagunta, M.S., Genest, J., Shearer, G.C., Duclos, F. and Oda, M.N. (2013) HDL-apoA-I Exchange: Rapid Detection and Association with Atherosclerosis. *PLoS ONE*, **8**, e71541.

Corbett S and Morin-Papunen L. (2013) The Polycystic Ovary Syndrome and recent human evolution. Molecular and cellular endocrinology. *Molecular and Cellular Endocrinology*, **373**, 39-50.

Corbett SJ, McMichael AJ and Prentice AM. (2009) Type 2 diabetes, cardiovascular disease, and the evolutionary paradox of the polycystic ovary syndrome: a fertility first hypothesis. . *American journal of human biology : the official journal of the Human Biology Council*, **21**, 587-598.

Chang RJ, Nakamura RM, Judd HL and SA, K. (1983) Insulin resistance in nonobese patients with polycystic ovarian disease. *Journal of Clinical Endocrinology & Metabolism*, **57**, 356–359.

de Zegher F and Ibáñez L. (2009) Early Origins of polycystic ovary syndrome: hypotheses may change without notice. *Journal of Clinical Endocrinology & Metabolism*, **94**, 3682-3685.

de Zegher F, Lopez-Bermejo A and L, I. (2009) Adipose tissue expandability and the early origins of PCOS. . *Trends in endocrinology and metabolism: TEM*, **20**, 418-423.

Dokras, A. (2013) Cardiovascular disease risk in women with PCOS. *Steroids*, **78**, 773-776.

Dronavalli, S. and Ehrmann, D.A. (2007) Pharmacologic therapy of polycystic ovary syndrome. *Clinical Obstetrics and Gynecology*, **50**, 244-254.

Dunaif, A. (1997) Insulin resistance and the polycystic ovary syndrome: Mechanism and implications for pathogenesis. *Endocrine Reviews*, **18**, 774-800.

Fauser BC, T.B., Rebar RW, Legro RS, Balen AH, Lobo R, Carmina E, Chang J, Yildiz BO, Laven JS, Boivin J, Petraglia F, Wijeyeratne CN, Norman RJ, Dunaif A, Franks S, Wild RA, Dumesic D, Barnhart K,. (2012) Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertility and Sterility*, **97**, 28-38.e25.

Fisher, E., Feig, J., Hewing, B., Hazen, S. and Smith, J. (2012) High-Density Lipoprotein Function, Dysfunction, and Reverse Cholesterol Transport. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **32**, 2813-2820.

Franks, S. (1995) Polycystic Ovary Syndrome. *New England Journal of Medicine*, **333**, 853-861.

Gambineri A, Patton L, Altieri P, Pagotto U, Pizzi C, Manzoli L and R., P. (2012) Polycystic ovary syndrome is a risk factor for type 2 diabetes: results from a long-term prospective study. *Diabetes*, **61**, 2369-2374.

González F, Nair KS, Daniels JK, Basal E, Schimke JM and HE., B. (2012) Hyperandrogenism sensitizes leukocytes to hyperglycemia to promote oxidative stress in lean reproductive-age women. *Journal of Endocrinology Metabolism*, **97**, 2836-2843.

González F, Rote NS, Minium J and JP., K. (2006) Reactive oxygen species-induced oxidative stress in the development of insulin resistance and hyperandrogenism in polycystic ovary syndrome. *Journal of Clinical Endocrinology & Metabolism*, **91**, 226-240.

González, F., Sia, C.L., Shepard, M.K., Rote, N.S. and Minium, J. (2014) The altered mononuclear cell-derived cytokine response to glucose ingestion is not regulated by excess adiposity in polycystic ovary syndrome. *Journal of Clinical Endocrinology & Metabolism*, **epub**, 2014-2046.

Ibanez, L., Diaz, M., Sebastiani, G., Sanchez-Infantes, D., Salvador, C., Lopez-Bermejo, A. and de Zegher, F. (2011) Treatment of Androgen Excess in Adolescent Girls: Ethinylestradiol-Cyproteroneacetate Versus Low-Dose Pioglitazone-Flutamide-Metformin. *Journal of Clinical Endocrinology & Metabolism*, **96**, 3361-3366.

Ibanez, L., Lopez-Bermejo, A., del Rio, L., Enriquez, G., Valls, C. and de Zegher, F. (2007) Combined low-dose pioglitazone, flutamide, and metformin for women with androgen excess. *Journal of Clinical Endocrinology & Metabolism*, **92**, 1710-1714.

Ibanez, L., Lopez-Bermejo, A., Diaz, M., Enriquez, G., Del Rio, L. and De Zegher, F. (2010) Low-dose pioglitazone, flutamide, metformin plus an estro-progestagen for non-obese young women with polycystic ovary syndrome: increasing efficacy and persistent safety over 30 months. *Gynecological Endocrinology*, **26**, 869-873.

Ibanez, L., Ong, K.K., Lopez-Bermejo, A., Dunger, D.B. and de Zegher, F. (2014) Hyperinsulinaemic androgen excess in adolescent girls. *Nat Rev Endocrinol*, **p499 doi:10.1038/nrendo.2014.58**.

Ibáñez L, Díaz M, Sebastiani G, Marcos MV, López-Bermejo A and de Zegher F. (2013) Oral contraception vs insulin sensitization for 18 months in nonobese adolescents with androgen excess: posttreatment differences in C-reactive protein, intima-media thickness, visceral adiposity, insulin sensitivity, and menstrual regularity. *The Journal of clinical endocrinology and metabolism.* , **98**, E902-907.

Jialal, I. and Devaraj, S. (1996) The role of oxidized low density lipoprotein in atherogenesis. *Journal of Nutrition*, **126**, 1053S-1057S.

Kaysen, G. and Eiserich, J. (2004) The role of oxidative stress-altered lipoprotein structure and function and microinflammation on cardiovascular risk in patients with minor renal dysfunction. *Journal American Society Nephrology*, **15**, 538-548.

Kobaly, K., Vellanki, P., Sisk, R.K., Armstrong, L., Lee, J.Y., Lee, J., Hayes, M.G., Urbanek, M., Legro, R.S. and Dunaif, A. (2014) Parent-of-Origin Effects on Glucose Homeostasis in Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology & Metabolism*, **0**, jc.2013-4338.

Legro RS, A.S., Ehrmann DA, Hoeger KM, Murad MH, Pasquali R, Welt CK; Endocrine Society,. (2013) Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology & Metabolism*, **98**, 4565-4592.

Lewy VD, Danadian K, Witchel SF and S., A. (2001) Early metabolic abnormalities in adolescent girls with polycystic ovarian syndrome. *The Journal of pediatrics*, **138**, 38-44.

Macut D, Bjekić-Macut J and A., S.-R. (2013) Dyslipidemia and oxidative stress in PCOS. *Frontiers of hormone research.* , **40**, 51-63.

Mallol, R., Rodriguez, M.A., Brezmes, J., Masana, L. and Correig, X. (2013) Human serum/plasma lipoprotein analysis by NMR: Application to the study of diabetic dyslipidemia. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **70**, 1-24.

March, W.A., Moore, V.M., Willson, K.J., Phillips, D.I.W., Norman, R.J. and Davie, M.J. (2010) The prevalence of polycystic ovary syndrome in a community sample

assessed under contrasting diagnostic criteria. *Human Reproduction*, **25**, 544-551.

Martitim, A., Sanders, R. and Watkins, J. (2003) Diabetes, oxidative stress, and antioxidants: a review. *Journal of Biochemical Molecular Toxicology*, **17**, 24-38.

Murri M, Luque-Ramírez M, Insenser M, Ojeda-Ojeda M and HF, E.-M. (2013) Circulating markers of oxidative stress and polycystic ovary syndrome (PCOS): a systematic review and meta-analysis. . *Human reproduction update*, **19**, 268-288.

Panopoulos, A.D., Yanes, O., Ruiz, S., Kida, Y.S., Diep, D., Tautenhahn, R., Herrerias, A., Batchelder, E.M., Plongthongkum, N., Lutz, M., Berggren, W.T., Zhang, K., Evans, R.M., Siuzdak, G. and Belmonte, J.C.I. (2012) The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Research*, **22**, 168-177.

Patti, G.J., Yanes, O., Shriver, L.P., Courade, J.-P., Tautenhahn, R., Manchester, M. and Siuzdak, G. (2012a) Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nature Chemical Biology*, **8**, 232-234.

Patti, G.J., Yanes, O. and Siuzdak, G. (2012b) Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, **13**, 263-269.

Rajkhowa, M., Neary, R., Kumpatla, P., Game, F., Jones, P., Obhrai, M. and Clayton, R. (1997) Altered Composition of High Density Lipoproteins in Women with the Polycystic Ovary Syndrome. *Journal of Clinical Endocrinology and Metabolism*, **82**, 3389-3394.

Roberts, C. and Sindhu, K. (2009) Oxidative stress and metabolic syndrome. *Life Sci.*, **84**, 705-712.

Rosenson, R.S., Jr, H.B.B., Ansell, B., Barter, P., Chapman, M.J., Heinecke, J.W., Kontush, A., Tall, A.R. and Webb, N.R. (2013) Translation of High-Density Lipoprotein Function Into Clinical Practice

Current Prospects and Future Challenges. *Circulation*, **128**, 1256-1267.

Seleem, A.K., Refaeey, A.A.E., Shaalan, D., Sherbiny, Y. and Badawy, A. (2014) Superoxide dismutase in polycystic ovary syndrome patients undergoing intracytoplasmic sperm injection. *Journal of Assisted Reproduction and Genetics*, **31**, 499-504.

Shao, B., Cavigiolio, G., Brot, N., Oda, M. and Heinecke, J. (2008) Methionine oxidation impairs reverse cholesterol transport by apoprotein A-1. *PNAS*, **105**, 12224-12229.

Sharpless, J.L. (2003) Polycystic Ovary Syndrome and the Metabolic Syndrome. *Clinical Diabetes*, **21**, 154-161.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, **78**, 779-787.

Sorci-Thomas, M. and Thomas, M. (2002) The effects of altered apolipoprotein A-I structure on plasma HDL concentration. *Trends in Cardiovascular Medicine*, **12**, 121-128.

Stadtman, E.R., Moskovitz, J. and Levine, R.L. (2003) Oxidation of methionine residues of proteins: Biological consequences. *Antioxidants & Redox Signaling*, **5**, 577-582.

Stocker, R. and Keaney, J. (2005) New insights on oxidative stress in the artery wall. *Journal Thromb Haemost.* , **3**, 1825-1834.

Sukalich, S. and Guzick, D. (2003) Cardiovascular health in women with polycystic ovary syndrome. *Seminars in Reproductive Medicine*, **21**, 309-315.

Sundaram, R., Bhaskar, A., Vijayalingam, S., Viswanathan, M., Moahn, R. and Shanmugasundaram, K. (1996) Antioxidant status and lipid peroxidation in type II diabetes mellitus with and without complications. *Clin Sci*, **90**, 255-260.

Talbott EO, Guzick DS, Sutton-Tyrrell K, McHugh-Pemu KP, Zborowski JV, Remsberg KE and Kuller LH. (2000) Evidence for association between polycystic ovary syndrome and premature carotid atherosclerosis in middle-aged women. . *Arteriosclerosis, thrombosis, and vascular biology*, **20**, 2414-2421.

Tomita, M. and Kami, K. (2012) Systems Biology, Metabolomics, and Cancer Metabolism. *Science*, **336**, 990-991.

Tsilchorozidou, T., Overton, C. and Conway, G. (2004) The pathophysiology of polycystic ovary syndrome. *Clinical Endocrinology* **60**, 1-17.

Valkenburg, O., Steegers-Theunissen, R., Smedts, H., Dallinga-Thie, G., Fauser, B., Westerveld, E. and Laven, J. (2008) A more atherogenic serum lipoprotein profile is present in women with polycystic ovary syndrome: a case-control study. *Journal Clinical Endocrinology Metabolism*, **93**, 470-476.

Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J. and Yanes, O. (2012) A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites*, **2**, 775-795.

von Eckardstein A and Widmann C. (2014) High-density lipoprotein, beta cells, and diabetes. *Cardiovascular Research*, **103**, 385-394.

Zhang, H., Forman, H. and Choi, J. (2005) γ-Glutamyl Transpeptidase in Glutathione Biosynthesis. *Methods in Enzimology*, **401**, 468-483.

SUPPLEMENTARY MATERIAL

	% variation	p-value
Lactate	25	0.0092
Acetoacetate	-20	0.1493
Acetates	-10	0.2945
Alanine	-3	0.6404
Valine	11	0.0792
Isoleucine+Valine	-1	0.9184
Leucine+Isoleucine	-23	0.1711
Glucose	-9	0.0656
Tyrosine	-19	0.1009
Hystidine	-19	0.1951
Glutamine	-56	0.0004
Free Glycerol	-64	0.043
Lysine	-47	0.0530
Glutamate	16	0.0717
Creatine	3	0.4382
Citrate	-239	0.0187
Glycine	-20	0.031

Supplementary Table 1. Metabolites identified by NMR. P-values were obtained from a Yuen-Welch's t-test.

	% variation	p-value	UPLC Column
Taurine	48	5.2E-03	HILIC
Glutamate	94	4E-03	HILIC
Methionine Sulfoxide	297	2.4E-05	HILIC
Choline	57	1.9E-03	HILIC
Methionine	-99	1E-04	HILIC
5-oxoproline	76	2.8E-03	HILIC
Glutamine	-31	1.5E-04	RP-C18
Glu-Gly	98	1.5E-04	RP-C18
Glu-Glu	82	3.1E-03	RP-C18
Val-Glu	122	7.9E-04	RP-C18
Glu-Taurine	250	3E-07	RP-C18
PC (16:1)	667/428	3.7E-05	HILIC & RP-C18
PC (10:2)	700	5E-04	RP-C18

Supplementary Table 2. Metabolites identified by RP-C18 and HILIC that showed a greater than 50% of variation with a p-value < 0.01 (Yuen-Welch's t-test)