Integration of Omics, and Clinical Data of the hormonal and inflammatory response at the basal level in response to the different macronutrients of the diet

Edmond Géraud Aguilar

Master: Bioinformatics and Biostatistics
Work Area: Biostatistics
Supervisor: Susana Barcelo Cerdá
Co-Supervisor: Francisco García García

21/01/2021

Free Documentation License (GNU FDL)

C) Copyright

Title of the thesis:	*Integration of Omics, and Clinical Data of the hormonal and inflammatory response at the basal level in response to the different macronutrients of the diet*
Author name:	*Edmond Géraud Aguilar*
Consultor name	*Susana Barceló Cerdá*
PRA name	*Carles Ventura Royo*
Fecha de entrega (mm/aaaa):	*01/2021*
Program:	*Bioinformatics and Biostatistics*
Work area:	*Biostatistics*
Language of the thesis	*English*
Key words	*data integration, omics, Polycystic ovarian syndrome (PCOS)*

Resumen del trabajo:

El síndrome de ovario poliquístico (PCOS) es un síndrome multigénico, metabólico y con fuerte asociación a factores externos como la dieta. Es un síndrome muy común, con una prevalencia hasta del 20% en mujeres en edad reproductiva. El objetivo del trabajo es entender mejor dicho síndrome puesto que su etiología es poco conocida, mediante el entendimiento de los mecanismos moleculares a nivel basal, de la respuesta hormonal, metabólica e inflamatoria frente a los diferentes macronutrientes de la dieta. Para poder abordar el problema, se realizó un análisis estadístico integrativo de cuatro conjuntos de datos: clínicos, marcadores de permeabilidad intestinal, metaboloma y metagenoma. Con el fin de integrar dichos bloques, primero se procedió a un análisis descriptivo univariante con el fin de entender cada variable y reconocer la estructura de cada bloque. Después se procedió con técnicas de reducción de la dimensionalidad, las que incluyen el clásico análisis de componentes principales (PCA), su variante con el algoritmo NIPALS en el caso de más variables que observaciones, y técnicas de escalado multidimensional métrico y no métrico. Finalmente se llevó a cabo el análisis integrativo. Se probaron tres métodos: regresión por mínimos cuadrados multivariante (PLS2), su versión ortogonal (O2PLS), y el método final, una versión generalizada de PLS2 con la habilidad de análisis discriminante con selección de variables, llamado DIABLO (Data Integration Analysis for Biomarker discovery using a Latent Components). En conjunto, las diversas fuentes de datos, separaron mejor los sujetos entre sexo, obesidad y mujeres con SOP, que individualmente. En general, los niveles basales de las diferentes biomoléculas medidas son más altos que a nivel postprandial. Los esteroides sexuales diferencian ambos sexos, y las mujeres obesas con el síndrome, tienen niveles más altos de estos esteroides. Las mujeres no obesas se caracterizan por tener niveles bajos de macromoléculas relacionadas con la respuesta inflamatoria y un desequilibrio en la microbiota intestinal. Por otro lado, las mujeres obesas con el síndrome PCOS se caracterizan por tener mayor concentración de productos metabólicos en sangre que el resto de sujetos.

Abstract (in English, 250 words or less):

Polycystic ovary syndrome (PCOS) is a multigenic, metabolic syndrome with a strong association with external factors such as diet. It is a very common syndrome, with a prevalence up to 20% in women of reproductive age. The aim of the work is to better understand this syndrome since it is little known its etiology in all areas. The main objective is to understand the molecular mechanisms at the basal level, of the hormonal, metabolic and inflammatory response to the different macronutrients in the diet. In order to address the problem, a joint integrative statistical analysis of four datasets was performed: clinical, Intestinal Permeability markers, metabolome, and metagenome. In order to integrate these blocks, a univariate descriptive analysis was first carried out in order to understand each variable and recognize the structure of each block. Then we proceeded with dimensionality reduction techniques, which include the classic principal component analysis (PCA), its variant with the NIPALS algorithm in the case of more variables than observations, and metric and non-metric multidimensional scaling techniques. Finally, the integrative analysis was carried out. Three methods were tested: multivariate least squares regression (PLS2), its orthogonal version (O2-PLS), and the final method, a generalized version of PLS2 with the ability of discriminant analysis and variable selection, called DIABLO (Data Integration Analysis for Biomarker discovery using a Latent Components). The several blocks taken together, differentiate better the subjects between sex, obesity, and women with PCOS, than individually. In general, the basal levels of the different biomolecules measured are higher than at the postprandial level. Sex steroids differentiate both sexes, and the obese women with the syndrome have higher levels of these steroids. Non-obese women are characterized by low levels of macromolecules related to the inflammatory response and an imbalance in the intestinal microbiota. On the other hand, obese women with the syndrome are characterized by having higher concentration of metabolic products in the blood than the rest of the subjects.

Table of Contents

Table of Figures

Index of Tables

1 Introduction

1.1 Context and justification of the project

Polycystic ovarian syndrome (PCOS), is a metabolic and endocrine disorder, very common in women of reproductive age. Its prevalence ranges between 6% and 20%. This range of prevalence depends on the diagnosis criteria. The etiology of the syndrome remains unknown, however, mounting evidence suggests that PCOS is a multigenic disorder with strong environmental and epigenetic influences, including lifestyle and diet. PCOS is associated with abdominal adiposity, insulin resistance, obesity, metabolic disorders, and cardiovascular risk factors. The pathophysiology of the syndrome increases the risk for comorbidities such as endometrial cancer, diabetes mellitus II and psychiatric disorders like anxiety, depression, use of recreational drugs, low self-esteem among others (1–3).

For patients with this syndrome, scientists and clinicians do not have so much clear what is about. Moreover, it is of little interest to pharmaceutical companies and health authorities. This is reflected by the scarce commercial studies in comparison with diabetes mellitus II, despite both disorders showing similar worldwide prevalence. This lack of interest may be due to PCOS being the most poorly understood medical disorder and its long-term consequences for health for patients and their relatives. One of the main reasons is the inadequacy of its naming (3).

Taking into account the previous considerations, one of the main objective of this project is to better understand the syndrome, by physicians, scientists, patients, and the general public.

1.2 Motivation

Science has evolved through the centuries by leaps and bounds, through the implementation of the various tools of these tools, includingMathematics, an exact science, and thanks to it, researchers have been able to implement probability and statistics strategies. Therefore, by implementing probabilistic and statistical methods in this research on PCOS, It will be able to spread the word about a little-known disease such as PCOS in all sectors of society. Indeed it is a really interesting and exciting topic to be investigated. Learning about such a health problem, which affects a large number of women around the world, and given the fact that is little known is of real interest.

As we know, any health problem has its basis on the molecular blocks of life, that is, genetics, epigenetics, metabolomics, i.e the omics in general, and as we discussed, environmental factors. Due to the simple fact that the human is a biological system, undoubtedly, molecular bases that are made must be related to its environment. Thus, it is more logical to think about addressing health problems that affect us as a whole and not in an isolated way. In other words, to address a health problem like PCOS, particularly in this work, the main objective is to develop and implement a probabilistic and statistical methodology to integrate the metagenome, metabolome, and clinical biomarkers to be able to understand better not only the system under study, but also to be able to export this methodology to other biological areas, and why not to other sciences. Furthermore, due to the nature of the project, which is projected into the field of personalized medicine, if the results are good enough, the physicians could exploit the results to obtain a set of reduced clinical and omics variables to diagnose PCOS and pursue a treatment for the patient.

1.3 Problems to address

This work is under the project called *Hormonal, metabolic, inflammatory, and oxidative response to the different macronutrients of the diet: influence of sex steroids.* Within the framework of this project, several responses have already been analyzed separately. Data and results are already available at each of these levels of biological information. Now, what we want to carry out within this work, is to integrate these data globally, allowing us to understand the relationships between the different elements evaluated so far for a better understanding of the described pathology.

We have metabolic, metagenomic, Intestinal Permeability, and clinical records (4–11). Thus the main questions we want to address are the following:

- Is there any relationship between the gut microbiome, Intestinal Permeability, and metabolome? At the basal level? (**Main question)**

- Is this relationship different as a function of sex, PCOS, and/or obesity? The subjects are grouped by sex, PCOS, and obesity. Is there an interaction between the different groups? In other words, is everything related to sex hormones?

1.4 Main objectives of the thesis

The main goal of this work is to improve the comprehension of the roles of the different macronutrients of the diet in women with PCOS, through the integration of data coming from the different biological and clinical responses.

The fulfillment of the mentioned above goal will be carried out through the following specific objectives:

1. Descriptive statistical analysis of the data belonging to each of the several biological levels quantified: metabolomics, markers of gut permeability, microbiome, and clinical records. The univariate and bivariate descriptive analysis will provide us a better understanding of the variables that participate in the study as well as their main relationships.

2. Integrative statistical analysis to include the information generated in each of the biological and clinical dimensions described, providing results that improve the

characterization of this disease in a systems biology context (12–14).

1.5 Specific tasks of the thesis.

To state the objectives in a more specific way, we should start with the following *hypothesis*:

There is a relationship between the diet, gut microbiome, and Intestinal Permeability markers, and metabolites. This relationship may be influenced by sex hormones. For this reason, it should be different between women, men, and women with PCOS, furthermore, this relationship should be modulated by obesity.

Thus, to pursue this hypothesis, the following tasks have been specified:

1. Load and make a pretreatment of the data, i.e to know the structure of each data set.
2. Perform a univariate descriptive statistical analysis both numerically and graphically.
3. Perform a multivariate descriptive analysis both numerically and graphically.
4. Review of the current literature for statistical integration analysis.
5. Carry out a statistical integrative description analysis, to find out the possible correlations between the different data types, in an integrated manner to understand the pathophysiological mechanisms and pathways involved in the syndrome.
6. Selection of the most appropriate statistical workflow to integrate several types of data. Specifically, metagenomic, metabolomics, and Clinical Data.
7. To integrate data from the microbiome, metabolome, and gut permeability markers to observe if there are relationships and if these relationships are different between women, men, and women with PCOS.

1.6 Metagenomics, metabolomics, and the exposome.

Biological systems rely on the transmission of information from nucleic acids to metabolites to shape function and phenotype. This is known as the omic’s cascade. As we discussed above, this lens approach gives a more powerful technique than single omics (12,13).

Though there have been several advances in science, there is a lack of understanding of microbial function in disease causality. Metagenomics is the study of genetic material retrieved directly from environmental samples, including the gut, soil and water. It is host specific. Altered bacterial abundance in the gut has been associated with complex diseases such as symptomatic atherosclerosis, diabetes type II, obesity and colorectal cancer. Moreover, dietary compounds may modulate bacterial abundance within the gut and interact with microbiome composition to alter host metabolism. For this reason, an important tool to understand microbial functions is metabolomics. Metabolomics is a field that comprises the measurement of hundreds of thousands of small molecules in a system. While other omics such as genomics, transcriptomics provide valuable information regarding phenotypes and molecular mechanisms, metabolomics provides a real-time view of dynamic changes. The metabolome is involved in many processes affected by exogenous influences, such as the response to oxidative stress, inflammatory response, energy metabolism, among others. This functional knowledge is essential to cover this gap between microbial communities and the disease. However, we cannot fully understand a system under study, such as women with PCOS, without the linking of the microbiome, the metabolome, and the exposome. The exposome is exogenous factors, such as pollutants, hygiene products, lifestyle, and diet. Diet composition alters the human microbiome, and diet-induced changes can occur within a day. Furthermore, this change can alter the bacterial expression of the genes (14).

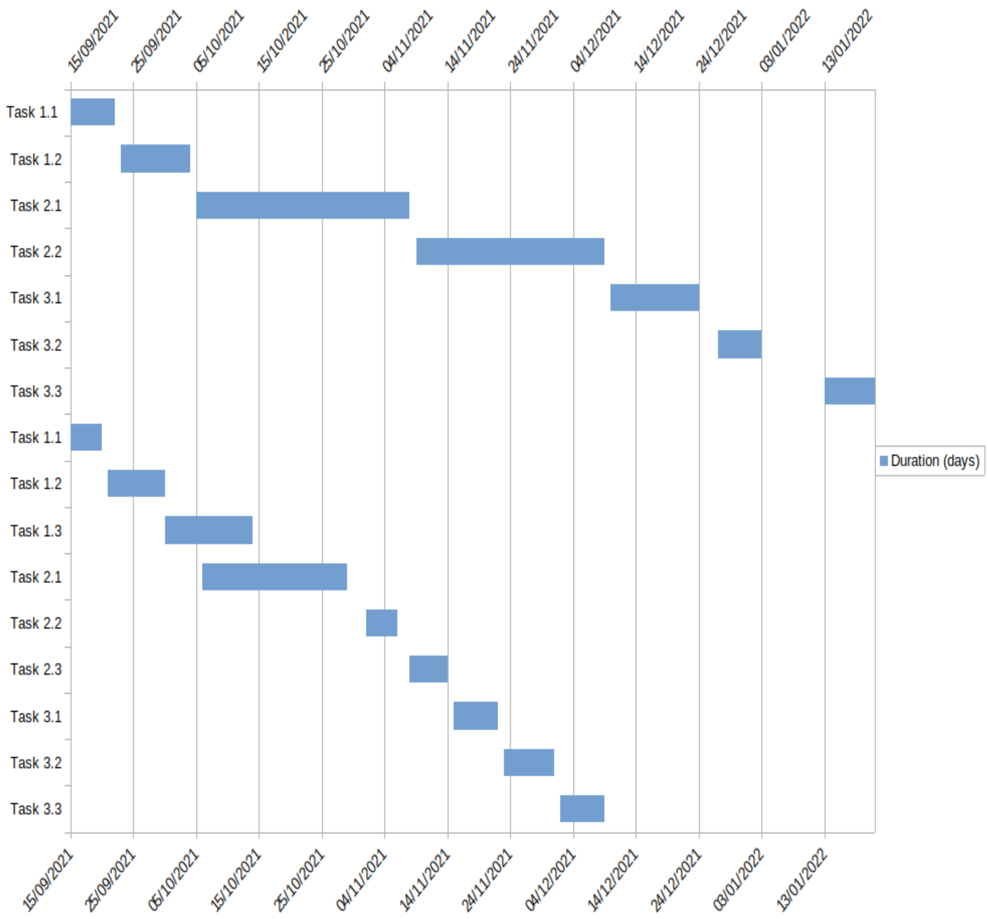
Multi-omics integration can be conceptual, statistical, and/or model-based. The conceptual approach combines the insights from a single omic to form a more comprehensive understanding of the system under study. Statistical integration focuses on statistical relationships across omics and model-based integration attempts to build a predefined system from layers to try to understand molecular organization and function 12).

Mainly there are three strategies: (a) independent, (b) conditional, (c), and joint modeling. In the independent modeling approach, also known as late integration, both models omics and non-omics are built independently. Then both models are combined in a final model. Regarding the accuracy, this metric is compared with the non-omics model. In this strategy, we cannot capture the correlation structure of different nature, and the predicted power is overestimated. The conditional modeling, firstly, a clinical model with non-omics data is defined, and later on, is added the omics variable build on the non-omics. Finally, we have the joint modeling approach, in which both data types are jointly modeled in a supervised or unsupervised way. The last strategy is the suggested one in the literature (15) and is the one used in this project.

Integrative statistical analysis arises from several challenges. For example, non-omic data is heterogeneous, i.e, there is a lack of uniformity. Data must be preprocessed, i.e transformed and normalized. And of course, we have to take into account the curse of dimensionality. Furthermore, when doing the integrative analysis, there is an ascertainment bias. This bias may induce a correlation between omics and non-omics data. And we have to address if all variable blocks contribute equally, and take into account the presence of subphenotypes, that may affect the performance of the model (12,13,15).

Descriptive, inferential, and integrative analysis on this work, is performed on R statistical software (16). Though there are several packages in R regarding the integration of omics, such as mixOmics (17), moCluster (18), mCIA(19), STATegRa (20), and STATegraEMS (21) among others, here we are applying several functions from different packages according to the questions we want to address.

Thus, for understanding PCOS, it is necessary to integrate several types of data including, metagenomic, metabolomic, Intestinal Permeability markers, and Clinical Data. This integration facilitates testing the hypothesis that we have stated above. Taking into consideration the literature review, a statistical integrative joint analysis should be conducted in the context of this work.



1. Figure: Gantt Diagram of the Project

1.7 Planning of the Master Thesis

A project planning was implemented before the statistical analysis. This planning is a must to establish a timeline of tasks and deadlines for the different parts of the master's thesis. The distinct tasks are divided into two main categories, the academic and the specific of the project according to general and specific objectives detailed in the previous paragraphs. The planning of the project can be seen on table 1, and the respective Gantt diagram on figure 1.

Tasks related to the project:

1. Study papers related to the project.
2. Study omics integration and its challenges: State of art.
3. Pre-treatment of the data.
4. Perform univariate descriptive statistical analysis.
5. Perform multivariate descriptive analysis.
6. Study integrative statistical analysis.
7. Integration of two data-sets.
8. Integration of all data-sets.
9. Results of the basal response.

Academic tasks:

1.
   1. Write evaluation report 1.
   2. Write evaluation report 2.
2.
   1. Write development report phase 1.
   2. Write development report phase 2.
3.
   1. Finish memory.
   2. Prepare a presentation.
   3. Study presentation.

PHASES	Tasks	Begin date	End date	Duration (days)
Phase I	Task 1.1	15/09/2021	22/09/2021	7
	Task 1.2	23/09/2021	04/10/2021	11
Phase II	Task 2.1	05/10/2021	08/11/2021	34
	Task 2.2	09/11/2021	09/12/2021	30
Phase III	Task 3.1	10/12/2021	24/12/2021	14
	Task 3.2	27/12/2021	03/01/2022	7
	Task 3.3	13/01/2022	21/01/2022	8
Phase I	Task 1.1	15/09/2021	20/09/2021	5
	Task 1.2	21/09/2021	30/09/2021	9
	Task 1.3	30/09/2021	14/10/2021	14
Phase II	Task 2.1	06/10/2021	29/10/2021	23
	Task 2.2	01/11/2021	06/11/2021	5
	Task 2.3	08/11/2021	14/11/2021	6
Phase III	Task 3.1	15/11/2021	22/11/2021	7
	Task 3.2	23/11/2021	01/12/2021	8
	Task 3.3	02/12/2021	09/12/2021	7

1. Table: Planning of the Project

1.8 Milestones and Risk Analysis

We were not able to complete all the initial tasks, but we have accomplished the main objectives. In summary, we have completely analyzed descriptively all variables between the blocks that comprise the project. Both in a univariate and multivariate way. The most consuming task was the descriptive univariate statistics, due to the huge amount of variables, and the transformations needed and this is the main reason why we did not complete all tasks specified initially on the project. Due to the core of the project is the integration of several blocks, and for the reasons mentioned, we could not perform inference.

The main risks, apart from the time that was consumed by the univariate descriptive analysis, were to address the multivariate analysis when more variables than observations were present. This was the case on three of four data sets. Only the

Intestinal Permeability markers (IP) data had more observations than variables. Furthermore, it was not expected at the beginning of the study, the fact that the microbiome comprises two data sets, the genus, and the phyla (because this data we got it later). Moreover, this data set was analyzed in several ways. In other words, we had to analyze 2 times the microbiome data, with the absolute abundance and the relative abundance of both taxa and the transformed data. Moreover, not only the descriptive statistics were necessary, but also, we had to compute other multivariate statistical methods, such as PCoA and non-metric multidimensional scaling.

Regarding the integrative analysis, it was computed on the expected time, but only at the basal level, though with delayed days due to the univariate analysis. Though good results were obtained, better results could be obtained, if the input variables were computed according to the best practices mentioned in the DIABLO paper. Which are basically to compute a cross-validation technique on the variables to obtain the most important ones. However, on a personal computer, this procedure could lead to several hours, even days. So, we must have a computation cluster for this purpose.

Regarding the computational requirements, we had access on the last day of the milestone, but if we have time, once the master's thesis is written, then we could try to obtain better results with the cross-validation technique, but maybe this could be done in a future project. But we are going to explain the prospects of the project in another section.

1.9 Summary of obtained results

From the univariate statistical analysis of the data sets where there were information at both times, i.e, basal and postprandial, we concluded that, across all subjects, the levels of almost all macromolecules, were lower before the ingestion of the macronutrients. The best method for the joint integration of several blocks was the multiblock generalized sparse canonical correlation analysis, also known as ***Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)*** method. More information is better. For example, the Intestinal Permeability marker data alone could not differentiate between sex, however, if we introduce the information of the Clinical Data, we can, distinct between obese and non-obese subjects is clearer and between sexes. This improvement is clearly seen on the interaction between the metabolome and Clinical Data.

The sex steroids were crutial for the distinction between males and females, and between females, the testosterone is higher on non obese PCOS women as well as the other steroids, in fact is one of the variables more importants on these patients. In general non obese subjects have higher levels of this hormones.
These women are characterized by low levels of macromolecules related to the inflammatory response and an imbalance in the intestinal microbiota. On the other hand, obese women with the syndrome are characterized by having higher metabolic products in the blood than the rest of the subjects, but also with a disorder on the gut microbiota, but less pronuncied.

The maini correlation between all blocks is positive. There are features that obviously correlates nevatively. And the main interaction are between de Clinical Data and the rest of the blocks, followed by the microbiome-metabolome. Six genera interacts with the sex steroids, and some of them are characteristic of the women with the syndrome.

1.10 Summary of chapters

The first chapter, which is the introduction, explains the context of the study, the main goals, problems to address, the specific objectives, and a review of the literature of the syndrome, as well as the different methods to perform the statistical integration analysis.

Materials and methods describe the blocks of data and the distribution of the subjects. Also includes a description of the multivariate techniques and joint integration methods used. Also, the materials and section show the final strategy to obtain the final results, as well as the tools used

The third chapter describes the results for each block, the univariate, bivariate, multivariate descriptive analysis. And the last section of this chapter reports the results of the techniques used and the chosen one for the joint integrative analysis.

Finally, a discussion of the results is presented, followed by the conclusions and finally the prospects, if the project could be continued in the future

The appendices contain the descriptive tables for each data set, as well as, the majority of the bivariate analysis plots. Descriptive tables for each data set, as well as, the majority of the bivariate analysis plots. It has to be metnioned that the meaning of the variables are in the descriptive tables, that is why there is no glossary.

2 Materials and Methods

2.1 Materials

2.1.1 Data

The study comprises 53 subjects, divided into 6 groups, moreover, have the data for all the 53 subjects for the Clinical, Intestinal Permeability markers (IP), and Metabolome data, for both basal and postprandial levels. However, we only have 46 subjects for the microbiome at the basal level and not at the postprandial time, as we can see on table 2

Clinical IP Metabolome	No Obese	Obese	Total	Microbiome	No Obese	Obese	Total
Females	9	8	17	Females	8	8	16
PCOS	9	8	17	PCOS	7	8	15
Males	10	9	19	Males	8	7	15
Total	28	25	53	Total	23	23	46

2. Table: Number of subjects distributed by groups

The Clinical Data which is described on table 3, comprises 58 variables plus the *SEX*, *GROUP*, and *OBESE* categorical variables. The *GROUP* variable consists of *female*, *male,* and *PCOS* levels, thus for the analysis, the sex variable is not taken into account but rather the *GROUP* variable is considered, because *GROUP* already contain the sex. The clinical variables consist of age, and body mass index (BMI), Waist Circumference (WC), and waist to hip ratio (WHR). Apart from the Clinical Data, there are 15 biochemical biomarkers: free testosterone, total testosterone, free estradiol, total estradiol, the ratio between the free sex steroids, sex hormone-binding globulin (SHBG), the homeostatic model assessment of insulin resistance (HOMA-IR), the insulin sensitivy index (ISI), cholesterol, High Density Lipoproteins (HDL), Low Density Lipoproteins (LDL,) triglycerides (TG), insulin, and glucose. Sex steroids, the ratio between them, the (HOMA-IR), and the SHBG are only calculated at basal levels. The other variables are calculated also at the postprandial level. ISI is measured at the postprandial level, however, it is computed between two-time points and is a single value. HOMA-IR is measured before the meal ingestion each day. The other variables: Insulin, Glucose, HDL, LDL, TG, and cholesterol are measured at the basal time and two-time points after the macronutrient load: 0-60-120 min for the protein and glucose and 0-120-240 for the lipids. The postprandial measurement of lipids is distinct because it takes more time to process them in the body.

Clinical Data set	Clinical and Biochemical markers	Resistance Insulin Variables	Other
Basal Variables	6x3+6	1x3+1	1
Postprandial Variables	6x3+6	1	0
Clinical Variables	4	0	0

3. Table: Clinical Data set. 6X3 +6 At basal and postprandial time there are 6 variables, plus the mean across all days of the macronutrient ingestion days. 1X3 corresponds to HOMA-IR plus the mean value of this feature in addition to the SHBG. There are 3 clinical variables: age, BMI, WC, and WHR, At postprandial time, the ISI is measured

The Intestinal Permeability (IP) biomarkers data set, consists of 5 variables: the cluster differentiation 14 (CD14), Lipopolysaccharide-binding protein (LBP), Glucagon-like peptide 2 (GLP2), Zonulin, and Succinate. The last variable is not an Intestinal Permeability marker, but rather a metabolite. The data includes the mean basal values for all the variables, plus the measurements of the three days at the basal level, apart from the three days at the postprandial stage. Thus in total, there are 35 variables, 5 of which are the mean value at the basal level. Here the postprandial levels were measured only at two time points, 0-60 min for glucose and proteins and 0-120 min for lipids.

Data	Basal	Postprandial	Subjects	Total Variables	Time Points	Final Dimension
Intestinal Permeability	Yes	Yes	53	5+5x3+5x3	2	53x35
Metabolome	Yes	Yes	53	37+37x3+37x3	3	53x259
Microbiome (phyla)	Yes	No	46	22	1	46x17
Microbiome (genera)	Yes	No	46	282	1	46x213

4. Table: Intestinal Permability markers (IP) 5x3+5x3+5; tthere are 5 IP one of which is succinate. 5X3 means that there are 3 measurements for each day at basal and posptrandial times, and +3 are the mean basal values. However only on this dataset the measurements on each day are computed on 2 time-points, 0-60 for glucose and proteins, and 0-120 for lipids. The metabolome data set, comprises actually of 37 metabolites, measured three times for each day, plus the mean basal values. The microbiome dataset consists of the phylum and the genera. We only have the values at basal time only once.

The microbiome data is different in the sense that we have only 46 subjects of the 53. Furthermore, the abundance of the microbiome at the phyla and genera taxa levels have been analyzed. The unidentified microbiome was excluded from the analysis. Without the unidentified operational taxonomic units (OTUs) from 282 genera, we ended up with 213, and from 22 phylum the final nomber of phylum OTUs were 17. It as to be mentioned, that we filtered both taxas at 0.01% of the abundance, leading up to 9 phylum OTUs and 89 genus.

All data was already collected and pre-processed, thus the blocks were already in a matrix form. This information is summarised on table 4. Moreover, All data was already collected and pre-processed, thus the blocks were already in a matrix form.

The description of all variables from all blocks, and the numerical descriptive statistics are in the appendices on tables 11-14.

2.1.2 Materials

The data was processed with a personal computer with 16 GB of RAM. The analysis was performed with R statistical software (16) version 4.1.2. The platform used was X86 64-pc-Linux.gnu (64 bit), running on the operative system Ubuntu 20.04.3 LTS Focal release (22)

To read the data, which is in Excel (23) format, the package *xlsx* was used, and especially the microbiome data process *phyloseq* was used (24,25). Regarding the univariate descriptive statistical analysis, no special package was used except from *ggplot2 (26)* for plotting and *base*. To perform multivariate analysis, *chemometrics*, *sparsepca* and, *pls* were used to compute NIPALS algorithm, sparse PCA, and PLS-regression (27–29). For the multidimensional scaling techniques, the package *vegan* was used (30). For the integration process, *pls*, *OmicsPLS* and, *mixOmics*, packages was used (17,29,31)

2.2 Methods

To corroborate the study and check whether the groups, i.e, the number of control females, males, and PCOS women, and obesity, actually not are related between them, a chi-squared test was performed, on the 53 subjects and the 46 subjects respectively.

In general, firstly univariate descriptive methods were used to observe the main features of the variables within the groups of interest. Density plots and histograms were used to assess the normality of the variables. Regarding the bivariate analysis Pearson correlation coefficient was computed, as well as, ordinarly least squares (OLS), to observe the relationship between the variables, regardless of the normality of the variables and the assumptions for OLS

Secondly, methods for reduction analysis were assessed to observe if each block could separate the groups of interest.

Thirdly, several methods were used for investigating the integration between the blocks of data. The integration of metabolome and microbiome was first assessed to choose the best integrative method. Then all blocks at a basal level were integrated.

2.2.1 Multivariate analysis methods

Two principal techinques were used for the multivariate analysis. Principal Component Analysis (PCA), and Multidimensional Scaling (MDS) techinques. These two techiques are dimensional reduction methods. The first one is linear whilst the second is non linear. When using PCA, if the block is of high dimensionalty, i.e more variables than observations $n \leq p$, the NIPALS algorithm is used, if not, the classical method is computed. Regarding MDS techinques, were used on the microbiome data set, to compute the beta diversity, which is essentially, a similarity matrix between subjects. If this matrix is euclidean, a Principal Coordinate Analysis is computed, if not, a non-metric MDS techinque is used, such as isometric regression or monotic regression. All these methods are describe here briefly. Another technique used is Partial Least Squares -Discrimination Analysis. This method was used for observe the microbiome as whole, also is a multivariate tenchinque, however, it is supervised, insted of the other techniques used that are unsupervised.

Principal Component Analysis (PCA), is a multivariate technique with the central aim of reducing the dimensionality of a multivariate data set while accounting for as much of the original variation as possible present in the data set by maximizing the covariance matrix. This aim is achieved by transforming to a new set of variables, the principal components, that are linear combinations of the original variables, which are uncorrelated and are ordered so that the first few of them account for most of the variation in all the original variables. This method is obtained by the maximization of the covariance matrix.

It can be done by computing the eigen decomposition, or by Singular Value Descomposition (SVD). The last method is the one used by the *prcomp* function in **R.** The data is mean centered, and if the features are in the same range normally is not scaled, if not it should be scaled. Then the covariance matrix is computed, and by SVD, one can obtain the eigen vectors that corresponds to the loadings, and the singular values squared to the variance of each component in a decrasing order. The score matrix is computed by the multplication of the loadigns by the data set.

In the case, when more variables than observations is present, one can compute the covariance matrix of the transpose data, and then the same procedure is applied, but now, what it should be the loadings matrix is the score matrix (32,33). In other words, (34) the PCA algorithm calculates the first PC along the first eigenvector by minimizing the projection error and then iteratively projects all the points to a subspace orthogonal to the last PC and repeats the process on the projected points. An alternative formulation is that the principal component vectors are given by the eigenvectors of the non-singular portion of the covariance matrix *C* given by:

$$C = \frac{1}{(n-1)^2} X^T C_n X$$

Where C_n is the $n \times n$ centering matrix. The loading vectors, denoted by $L1, \cdots, L_n$, are given in terms of the eigenvectors, $e_1, \cdots, e_n$ and the eigenvalues, $\lambda_1, \cdots, \lambda_n$ of C as follows:

$$L_i = \sqrt{\lambda_i}\, e_i,\ i = 1 \cdots, n$$

The Non-linear iterative partial least squares (NIPALS) algorithm is the first method for solving the problem of Partial Least Squares regression problem. For very-high-dimensional datasets, such as those generated in the omics sciences, for example, what happens on the metabolome, or in the Clinical Data, is a good option. The algorithm can be used to find the first few (or all) principal components with the decomposition:

$$X = TP^T$$

Where the columns of T are called scores and the columns of P are the loadings. The algorithm begins by initializing $h = 1$ and $X_h = X$, then proceeds through the following basic steps:

1. Choose t_h as any column of X_h
2. Compute loadigns as: $p_h = X'_h t_h / t'_h t_h$
3. Let $p_h = p_h / \sqrt{p'_h p_h}$
4. Compute scores $t_h = X_h p_h / p'_h p_h$

This algorithm repeats the (3) and (4) steps until convergence for the h^{th} principal component (35). This algorithm is present on the *chemometrics* **R** package (27).

The Multidimensional scaling (**MDS**) techinques are methods for visualizing the level of similarity of individual cases of a dataset. MDS is used to translate the information about the pairwise distances, similarities or disimilarities among a set of n individuals into a configuration of n points mapped into an abstract Cartesian space. Is a method of dimensionality reduction such as PCA, but it is non linear. There are two main approaches to solve this class of problems.

One one hand, there is the Classical multidimensional scaling It is also known as

Principal Coordinates Analysis (PCoA), It takes an input matrix giving dissimilarities between pairs of items and outputs a coordinate matrix whose configuration minimizes a loss function called *strain*. For example, given the Euclidean adistances d_{ij}between various individuals indexed by i and j one should be able to find the coordinates (x_i, x_j) of the subjects such that $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. If the euclidean distance is exact, the solution exists.

On the other hand, there is the **non-metric multidimensional scaling (N-MDS)**. In contrast to metric MDS, non-metric MDS finds arelationship between the dissimilarities in the item-item matrix and the Euclidean distances between items, and the location of each item in the low-dimensional space. The relationship is typically found using isotonic regression: let x denote the vector of proximities $f(x)$ a monotic transformation of x, and d the point distances; then coordinates have to be found, that minimize the so-called stress

$$Stress = \sqrt{\frac{\sum(f(x) - d)^2}{\sum d^2}}$$

A few variants of this cost function exist. MDS programs automatically minimize stress in order to obtain the MDS solution. The core of a non-metric MDS algorithm is a twofold optimization process. First the optimal monotonic transformation of the proximities has to be found. Secondly, the points of a configuration have to be optimally arranged, so that their distances match the scaled proximities as closely as possible. The basic steps in a non-metric MDS algorithm are:

1. Find a random configuration of points, e. g. by sampling from a normal distribution.
2. Calculate the distances d between the points.
3. Find the optimal monotonic transformation of the proximities, in order to obtain optimally scaled data.
4. Minimize the stress between the optimally scaled data and the distances by finding a new configuration of points.
5. Compare the stress to some criterion. If the stress is small enough then exit the algorithm else return to 2.

Another approach is by using a monotic regression with the Kruskal-s original monotone regression to minimize the stress:

$$Stress^2 = \frac{\sum (d - \hat{d})^2}{\sum (d - d_{null})^2}$$

Where d are distances among points in ordination configuration, $\hat{d}$ are the fitted ordination distances, and d_{null} are the ordination distances under null model. These techinque of MDS are in the package *vegan* of **R** (35–37).

Although **Partial Least Squares** was not originally designed for classification and discrimination problems, it has often been used for that purpose. **Partial Least-Squares Discriminant Analysis (PLS-DA)** is a multivariate dimensionality-reduction tool that has been popular in the field of chemometrics for well over two decades. Both chemometrics and omics data sets are characterized by large volume, large number of features, noise and missing data. PLS-DA can be thought of as a “supervised” version of Principal Component Analysis (PCA) in the sense that it achieves dimensionality reduction but with full awareness of the class labels. Besides its use for dimensionality-reduction, it can be adapted to be used for feature selection, within its sparse variant, as well as for classification.

The objective of dimensionality-reduction methods such as PCA and PLS-DA is to arrive at a linear transformation that converts the data to a lower dimensional space with as small an error as possible. PLS-DA In its standard variant the components are required to be orthogonal to each other. The first PC of PLS-DA can be formulated as the eigenvectors of the non-singular portion of *C*, given by:

$$C = \frac{1}{(n-1)^2} X^T C_n y y^T C_b X,$$

Where C_n is the $n \times n$ centering matrix. And y is the class label vector. The iterative process computes the loading vectors, $a_1, \cdots, a_d$, which which give the importance of each feature in that component. In iteration *h*, it has the following objective:

$$\max_{a_h, b_h} \; cov(X_h a_h, y_h b_h)$$

Where b_h is the loading for the label vector y_h, $X_1 = X$, and X_h and y_h are the residual matrices after transforming with the previous $h - 1$ components.

2.2.2 Methods on Statistical Integrative analysis

The PLS2 algorithm was developed within the framework of the NIPALS algorithm. The method and the algorithm used are implemented in several commercial software for multivariate data analysis.

Given the matrix $Y$of the responses and the matrix $X$of the predictors, the aim of PLS2 is to calculate the matrix of the regression coefficients B that produces the linear regression model:

$$Y = XB + F_A$$

where F_A is the matrix of residuals. Since no assumptions about the statistical distribution of the residuals are made, the term ‘model’ indicate a particular matrix decomposition.

PLS2 regression is performed decomposing the matrix of the predictors and that of the responses using a suitable orthogonal score matrix $T = [t_1, ..., t_A]$ as :

$$X = TP^t + E_A$$

and:

$$Y = TQ^T + F_A$$

under the constraint :

$$T = XW^*$$

where W^* is a suitable matrix to be calculated. Moreover, $P = X^tT(T^tT)^{-1}$ and $Q = Y^tT(T^tT)^{-1}$ are the loading matrices of the X and Yblocks, and E_A the matrix of the residuals of the X-block. PLS2 linearly transforms the space of the predictors to obtain a new set of orthogonal variables (called latent variables whose values are the scores t_i), which can be used to model the response in the least-squares sense. In this way, the limits of ordinary least squares (OLS) regression due to redundancy and multicollinearity are overcome. The matrix of the regtression coefficients resutls : $B = W^*Q^t$. The core of PLS2 is to choose a strategy to calculate W^* that can be done with the NIPALS algorithm (17,35,39). This algorithm is provided in several packages, but we used *mixOmics.*

The **Orthogonal 2- Partial Least Squares (O2-PLS) i**s an extension of the PLS2 method, when a systematic Y–orthogonal variation in X exists. The basic idea of O2-PLS is (1) to use multiple linear regression to estimate the pure constituent profiles and (2) to divide the systematic part in $X$and Yinto two parts, one which is related to both X and Y (covarying) and one that is not (orthogonal). For each matrix the latter is computed in a way that makes it orthogonal to the other matrix, i.e. linearly independent. Thus the O2-PLS model can be written as a factor analysis model where some factors (T) are common to both X and Y

$$X\ model:\ \ X = TW^T + T_{Y-ortho}P^T_{Y-ortho} + E \quad \text{(Equation 1 O2-PLS)}$$

$$Y\ model:\ \ Y = UC^T + U_{X-ortho}P^T_{X-ortho} + F \quad \text{(Equation 2 O2-PLS)}$$

$$Inner\ relation:\ \ U = T + H \quad \text{(Equation 3 O2-PLS)}$$

$$Prediction\ of\ Y: \hat{Y} = TC^T \quad \text{(Equation 4 O2-PLS)}$$

The inner relation for X and Y is given in Equation 1 , where H is the residual matrix. The predictive equation of Y is given in Equation (2). Here T and U are the score matrices for X and Y respectively. W and C are the joint component loading matrices E [N,K] and F [N,M] are the residual matrices. $P_{Y-ortho}$ is the loading matrix, and $T_{Y-ortho}$is the corresponding Y-orthogonal score matrix. $P_{X-ortho}$ is the loading matrix , and $U_{X-ortho}$ is the corresponding X-orthogonal score matrix. This method is available on the package *OmicsPLS (31,40)*

Finally, the ***Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO),*** a multi-omics integrative method that seeks for common information across different data types through the selection of a subset of molecular features, while discriminating between multiple phenotypic groups. DIABLO extends the sparse Generalized Canonical Correlation Analysis (sGCCA) (41) to a classification or supervised framework. sGCCA is a multivariate dimension reduction technique that uses singular value decomposition and selects co-expressed (correlated) variables from several omics datasets. sGCCA maximizes the covariance between linear combinations of variables (latent component scores) and projects the data into the smaller dimensional subspace spanned by the components. The selection of the correlated molecules across omics levels is performed internally with l_1 penalization on the variable coefficient vector defining the linear combinations. Since all latent components are scaled in the algorithm, sGCCA maximizes the correlation between components. However, we will retain the term ‘covariance’ instead of ‘correlation’ throughout this section to present the general sGCCA framework.

The sGCCA works as follows: denote Q data sets centered and scaled: $X^{(1)}(N \times P_1), X^{(2)}(N \times P_2), \cdots, X^{(Q)}(N \times P_Q)$ measuring the expression levels of $P_1, \cdots, P_Q$ omics variables on the same N samples. The method sovles the optimization function for each dimension $h = 1, ..., H$:

$$\max_{a_h^{(1)},\cdots a_h^{(Q)}} \sum_{i,j=1,i\neq j}^{Q} c_{i,j}(X_h^{(i)}a_h^{(i)}, X_h^{(j)}a_h^{(j)})$$

$$s.t.||a_h^{(q)}||_2 = 1 \ and \ ||a_h^{(q)}||_1 \leq \lambda^{(q)} \forall 1 \leq q \leq Q$$

where $a_h^{(q)}$ he variable coefficient or loading vector on dimension h associated to the residual matrix $X_h^{(q)}$of the dataset $X^{(q)}$. C is a (Q × Q) design matrix that specifies whether datasets should be connected. Elements in C can be set to zeros when datasets are not connected and ones where datasets are fully connected. Moreover, $\lambda^{(q)}$ s a non-negative parameter that controls the amount of shrinkage and thus the number of non-zero coefficients in $a_h^{(Q)}$. Similar to the LASSO (42) and other $l_1$penalized multivariate models the penalization enables the selection of a subset of variables with non-zero coefficients that define each component score $t_h^{(q)} = X_h^{(q)}a_h^{(q)}$. The result is the identification of variables that are highly correlated between and within omics datasets. To extend sGCCA for a classification framework, they substitute one omics dataset, $X^{(q)}$ on the loss function, with a dummy indicator matrix Y (N × G) to indicate the class membership of each sample, where G is the number of phenotype groups. For easier use of DIABLO, we replaced the l_1 penalty parameter $\lambda^{(q)}$ by the number of variables to select in each dataset and each component, as there is a direct correspondence between both parameters (43).

2.2.3 Methods on each data set

Before computing the univariate analysis, research on the biochemical markers and their relationship with Clinical Data was made. This was done, apart from being familiarized with the variables, to observe recorded relationships between them, to corroborate the results obtained. As mentioned, boxplots, ordinary least squares regression, and Pearson correlation were computed to observe the main differences between groups and the relationship between them and each variable.

When computing the multivariate analysis, the NIPALS algorithm (35) as the best method for differentiating between the groups of interest. The data was mean-centered. For the Clinical Data, due to the fact that sex steroids levels for men are out of range in comparison with the other variables, the resulting analysis the main distinction was on sex, however, if the data is scaled, the main distinction is between obese and non-obese subjects, for this reason, for this data set, the data was scaled and not scaled in order to compute PCA. This method was used because of the high dimensionality of both

Clinical and Metabolome data sets.

For the multivariate analysis on the IP, a classical PCA was computed with the mean-centered and scaled data. Mainly, because we have more observations than variables.

Methods on the metabolome

On this data set, the approach was different, in the sense that metabolites play an important role on very different pathways. Thus, we computed a Pearson correlation matrix to observe the relationship between them. So, no research on each metabolite was performed. Rather for each metabolite, only boxplots were computed. For the multivariate analysis, like on the Clinical Data, the NIPALS algorithm was used (35).

Methods on the microbiome

Since this data set comprises two sub-datasets, the phyla, and the genera, and as occurs with the metabolome, each microorganism plays a different role and interacts with different molecules on the system, the strategy for the analysis of this data set was different.

Alpha diversity is the mean species diversity in sites or habitats at a local scale, on the other side ***beta*** diversity is the ratio between regional and local species diversity. These terms were introduced by R. H. Whittaker (44). BBoth alpha and beta are diversity indices, which are quantitative measures that reflects how many different types, are in a data set, and that can simultaneously take into account the phylogenetic relations among the individuals (45)

The first equation is the Shannon index for the alpha diversity, this equation is closely related to the Shannon entropy, initially proposed by Claude Shannon in 1948. The idea is that, the more letters there are, and the closer their proportional abundances in the string of interest, the more difficult it is to correctly predict which letter will be the next one in the string, where is the proportion of individuals belonging to the *i*th species in the dataset of interest. Then the Shannon entropy quantifies the uncertainty in predicting the species identity of an individual that is taken at random from the dataset.et.

$$H = -\sum p_i ln\ p_i \qquad \text{(Equation 1)}$$

Regarding the Chao1 index, equation 2, estimate richness, i.e, the measurement of Operational Taxonomic Units (OTU), expected in samples given all the bacterial species that were identified in the samples. It is an estimator on the abundance of species richness. Where and are the count of singletons and doubletons respectively. A singleton is defined as an observation with exactly one individual in the sample, whereas a doubleton is defined as an observation with exactly two individuals in the sample. And is the number of observed species.

$$S_{chao1} = S_{obs} + \frac{F_1(F_1 - 1)}{2(F_2 - 1)} \qquad \text{(Equation 2)}$$

The Sørensen-Dice coefficient is a statistic used to gauge the similarity of two samples. The Sørensen index was intended to be applied to discrete data. Given two sets, it is defined as:

$$S(A,B) = 2\frac{|A \cap B|}{|A| + |B|} \qquad \text{(Equation 3)}$$

Where $|A|$ and $|B|$ are the cardinalities of the two sets (i.e. the number of elements in each set). The Sørensen index equals twice the number of elements common to both sets divided by the sum of the number of elements in each set

Finally, the Jaccard index is a statistic as well as the Sørensen used to specify the diversity of sample sets, it has a similar formulation to the Sørensen similarity, except that is the intersection of both sets divided by the cardinality of their union.

$$J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad \text{(Equation 4)}$$

Both Jaccard and the Sørensen similarity indices, compare members for two sets to see which members are shared and which are distinct. It's a measure of similarity for the two sets of data. The higher the percentage, the more similar the two populations (44–46).

For each sub-block of the microbiome, firstly, the Shannon and Chao-1 index, which tells us about the alpha diversity of the microbiome was computed, these indices are univariate, so to compare groups descriptively, only boxplots were made. Afterward, the Jaccard and Sørensen indexes were computed. Those indices allow us to clarify the beta diversity of the genera and the phyla. However, the beta diversity comes in similarity matrices. Thus to observe the possible differences between the subjects, Multidimensional Scaling (MDS) techniques were assessed.

To compute a metric-MDS, i.e, to apply principal coordinate analysis (PCoA), the similarity matrix has to meet the conditions for a euclidean distance, if not a non-metric-MDS is computed, wither an isotonic regression or a monotonic regression.

Since we are dealing with an abundance of microorganisms, an approach for assessing the differences between groups in a multivariate way, Canonical Correlation Analysis was computed and a comparison with PLS-DA was made. For the last method, log-center-ratio transformation is needed.

Afterward, bar plots for each phylum and genera were computed. Concerning the results obtained with the above methods, boxplots were made to observe the possible differences between subjects.

For an integration statistical analysis, all variables from all blocks of data have to be mean-centered and scaled to the unit variance to be comparable.

Firstly two blocks were integrated, the metabolome and the microbiome. To perform such analysis, only the basal variables were taken into account. To the proper integration, the microbiome data has to be transformed either to the relative abundance or the log-center-ratio.

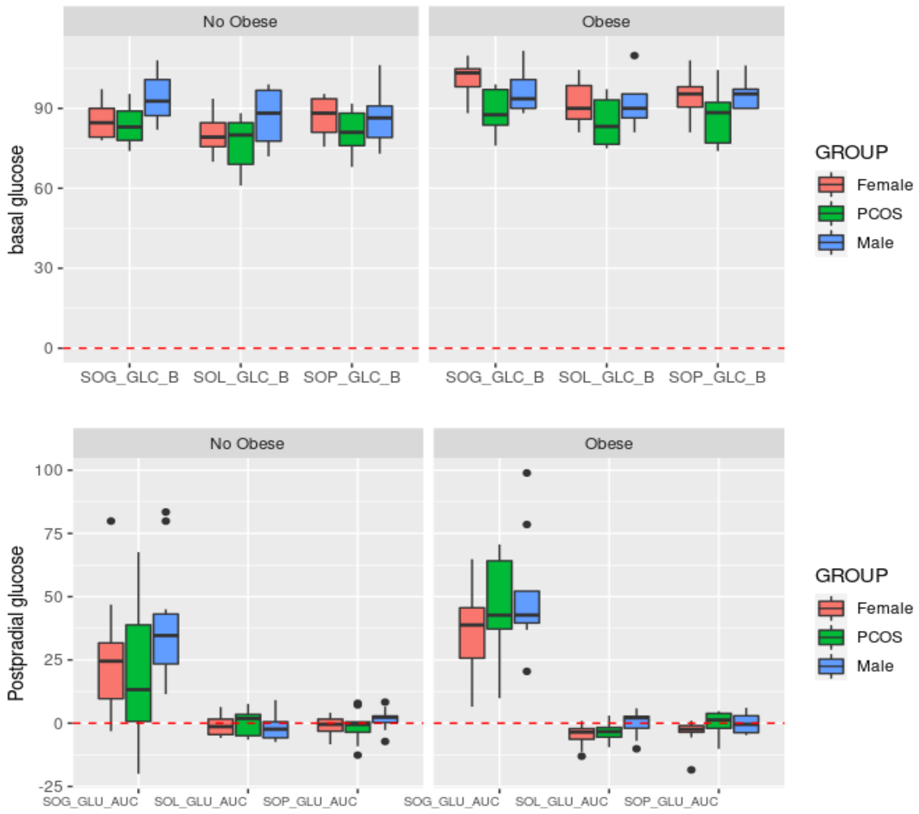
After that, three main statistical procedures were computed: Partial Least Squares (2PLS), Orthogonal 2O-PLS (40,47), and sparse Generalized Canonical Correlation analysis extended for a classification framework (43).

2.3 Final Strategy

The final strategy for the analysis of each sub-group and the total basal integration is conceptual and a statistical integrative joint analysis. Following the steps below:

1. Understand each block of data, i.e how it is structured.
2. Univariate and bivariate descriptive analysis using boxplots, histograms, density plots and Pearson correlation analysis.
3. Multivariate descriptive analysis. The following PCA methods are needed:
    1. Clinical Data: mean-centered, scaled, and not scaled NIPALS PCA.
    2. IP: classic PCA mean-centered and scaled to unit variance.
    3. Metabolome: NIPALS PCA
    4. Microbiome: MDS metric and non-metric PLS-DA
        1. The data has to be log-center-ratio transformed
4. Joint integrative statistical analysis.
    1. Sparse Generalized Canonical Correlation Analysis (sGCCA).

The



2. Figure: Basal and postprandial levels of glucose. Each trio of boxplots corresponds to the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Pink, green and blue boxplots represents females, PCOS women and men integrative method requires a design matrix that takes into consideration the correlation between blocks, the optimal number of components, and the optimal number of variables to compute.

1. The design matrix is as follows:

$$M = \begin{pmatrix} 0 & \rho_1 & \cdots & \rho_b \\ \rho_1 & 0 & \cdots & \rho_b \\ \vdots & & \ddots & \\ \rho_b & \cdots & & 0 \end{pmatrix} \qquad \text{(Equation 5)}$$

Where b is the number of data setts and each correlation coefficient ρ can be placed based on prior knowledge of the biological system, or data-driven. In the context of this study, the last method was used. For obtaining the relationship between two pairs of blocks, a simple PLS with one component was computed. With the obtained scores for the dependent and independent variables, a Pearson correlation coefficient was computed.

1. Based on the previous knowledge of the correlation between each block, the optimal number of components is assessed via the Cross-Validation technique (48).
2. Finally, the optimal number of variables has to be computed. This is a very consuming task. For example, if we consider a block with 9 variables randomly chosen, with a 9 grid span, the method will compute 81 models. This takes approximately 1.5 hours if we consider all nodes in a 16 RAM personal computer. Thus it is unfeasible for integrating all the blocks with all the variables, it will take days to compute. For this reason, this last step is not performed.

3 Results

3.1 Statistical analysis for each block.

3.1.1 Clinical Data

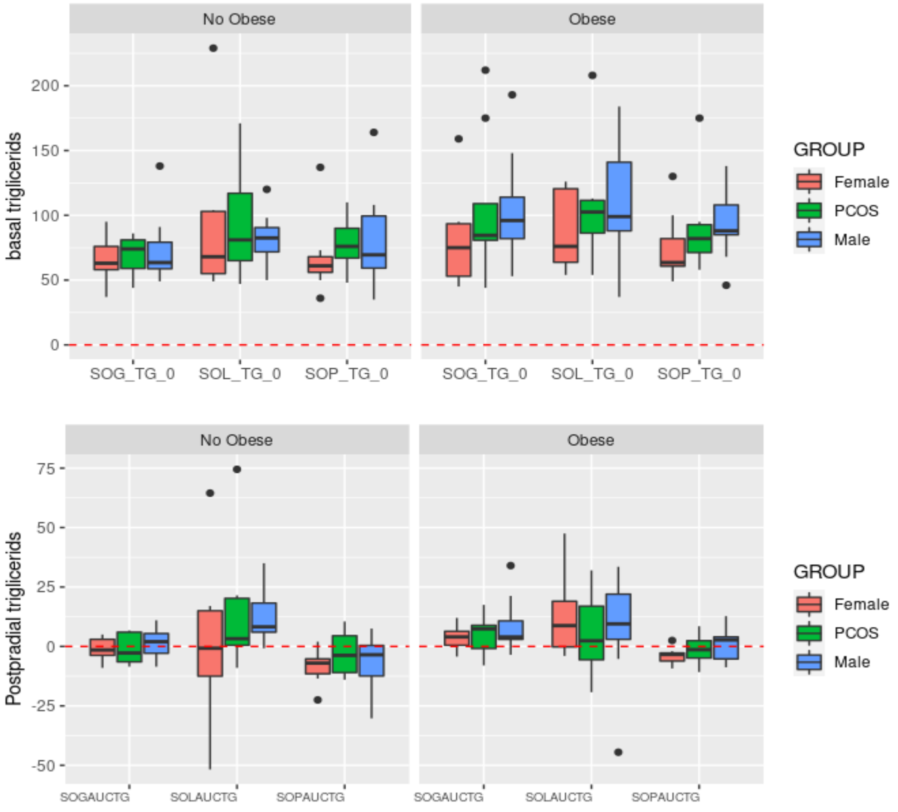
Univariate Analysis Results

At the Basal time women with PCOS showed higher values of testosterone but they did not present a higher waist to hip ratio measurement or waist circumference in comparison with non-hyperandrogenic women, as is described in the literature (10)

The SHBG is a protein that binds with sex hormones. This means that when SHBG increases, the free sex steroids diminish because they are bound to the globulin, but the total sex steroids augment. Therefore for healthy subjects and non-obese, we expect to have a higher concentration of E2 in women and lower in men. The contrary happens with testosterone, figure 4. On obese subjects, the SHBG, have lower levels than non-obese. In obese women, tend to have slightly levels of total testosterone, but it is barely apreciable than non-obese women. Obese men have lower levels of total testosterone than non-obese men. Free testosterone has the same behavior in women.

But in men, free testosterone is higher in obese than non-obese. PCOS women have similar values of SHBG to control women and higher concentration of testosterone levels.

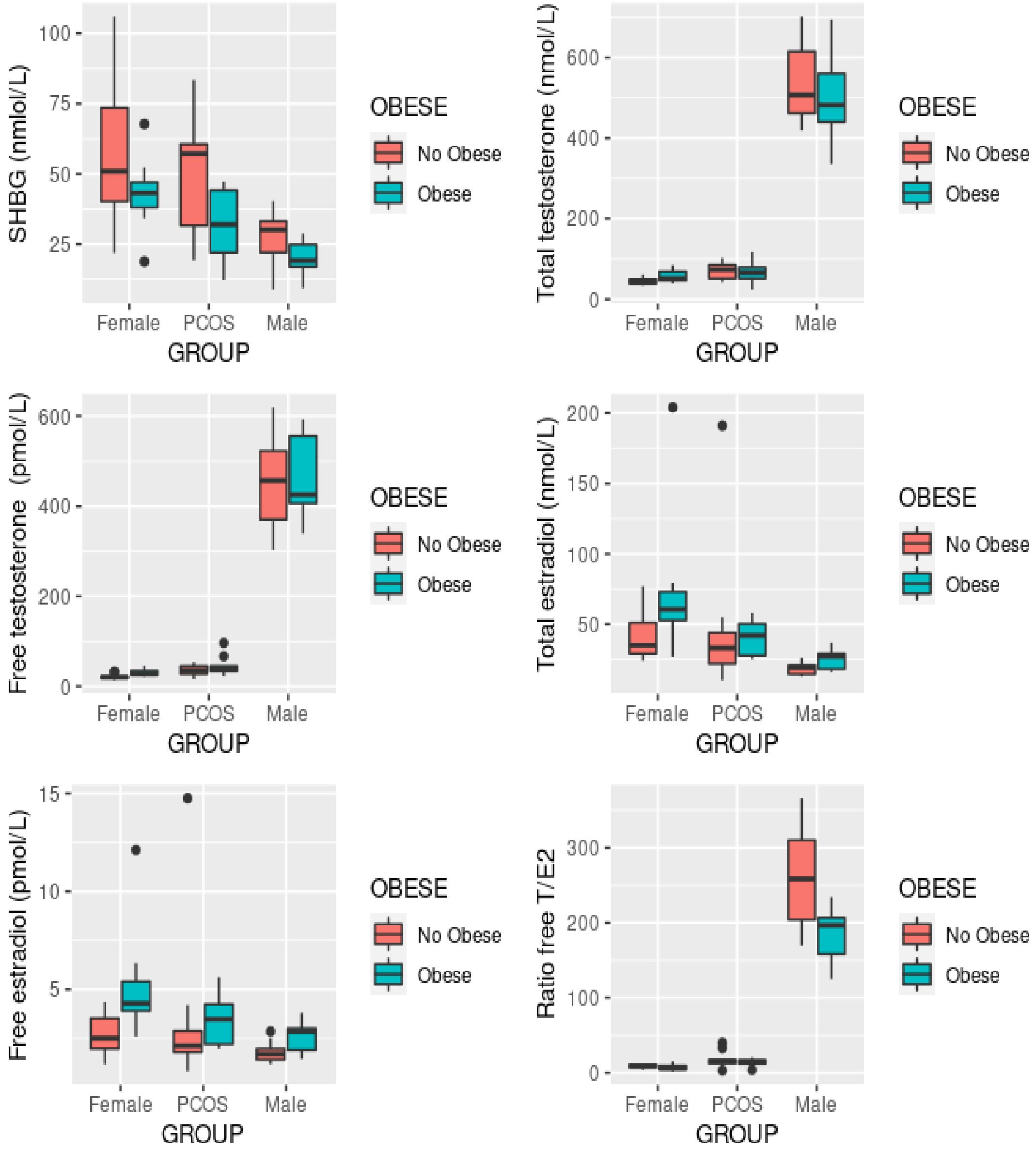
The SHBG is higher in women independently if they have PCOS or not. Regarding E2, the concentration is higher in obese subjects than non-obese subjects. And higher in women irrespective of they have PCOS than men. E2 has a lower concentration in PCOS women than control women.



3.Figure: Basal and postprandial levels of triglycerids. Trellis plot. Each trio boxplots are arragange in the following way: Firstly the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Blue boxplots belongs to males, green to PCOS women and pink to control women.

The ratio of both free sex steroids is higher in non-obese subjects and higher in men. But it is not distinguishable between obese and non-obese PCOS women (8,10). These results are summarized on figure 4

For the variables hsCRP (C-reactive Protein), HOMA-IR, insulin sensitivity index (ISI),



4. Figure: Clinical variables 2: Blue boxplots are obese subjects, and pink not obese. GROUP is the categorical variable that includes the sex levels and PCOS.

and lipids, a similar procedure was addressed. HsCRP is higher on obese subjects, but this interaction is only observed on PCOS women as well as the fasting concentration of insulin according to (49,50) however on (9) PCOS women this interaction is not observed. Fasting glucose, figure 2 was higher between obese females and control females, but no differences were found among the other subjects. HOMA-IR shows higher levels in obese subjects but no distinction between groups was observed (9). This results are observed on figure 8.

Triglycerides were higher on obese subjects, being higher on males and PCOS women. A similar pattern was found on cholesterol levels and LDL cholesterol. An opposite pattern was observed on HDL cholesterol according to (5), figure 3.

Postprandial levels:

Insulin postprandial levels are higher than basal levels on the first day. Being higher on

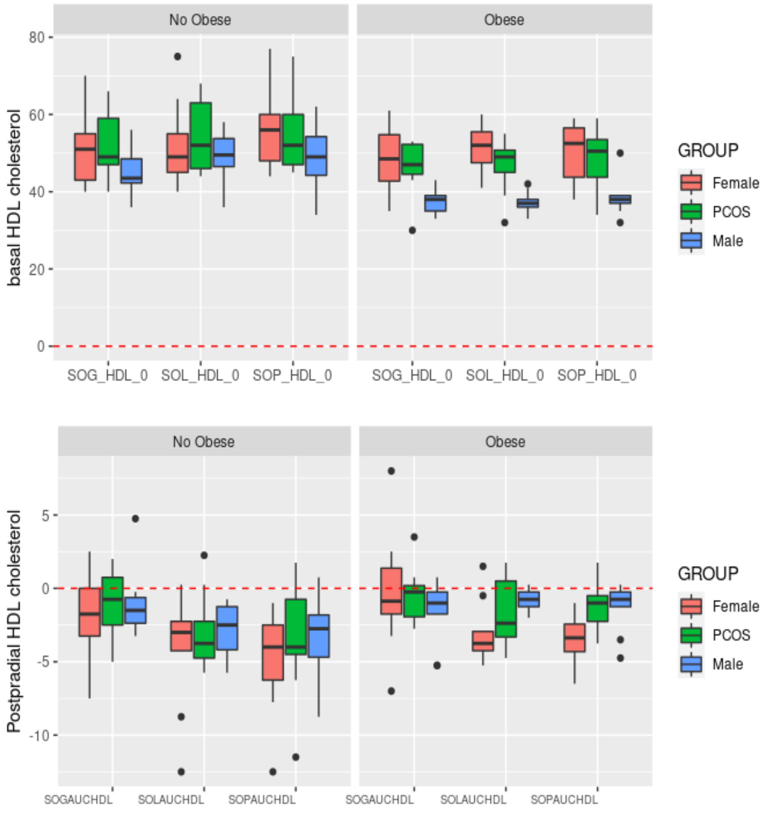
PCOS women. The protein ingestion day showed similar values at the basal time, but on the second day, values were lower. Obese PCOS subjects present a higher concentration of glucose on the first day. But all subjects present lower values than at the basal stage, but still lower, these observations agreed with (10)**.**

Triglycerides show higher values on the lipid load day, with no distinction between groups. Still lower values than at the basal time. On the contrary levels of cholesterol were lower at postprandial levels, but higher in obese PCOS women. The same pattern was observed on LDL cholesterol. And similar pattern but without distinction between groups was observed in HDL cholesterol (9), figures 5-6

Bivariate Analysis Results

Regarding BMI, and WC the data shows a positive correlation with SHBG in healthy men, obese women, and non-obese PCOS women. On the other hand, a negative correlation is found in healthy women, obese PCOS women, and obese men, this resultos can be observed on figures Error: Reference source not found-69 (see apendices). WHR correlated with SHGB follows a similar pattern between subjects but not on healthy men, in which this correlation becomes negative, due to the men and women body shape relationships (8,10).

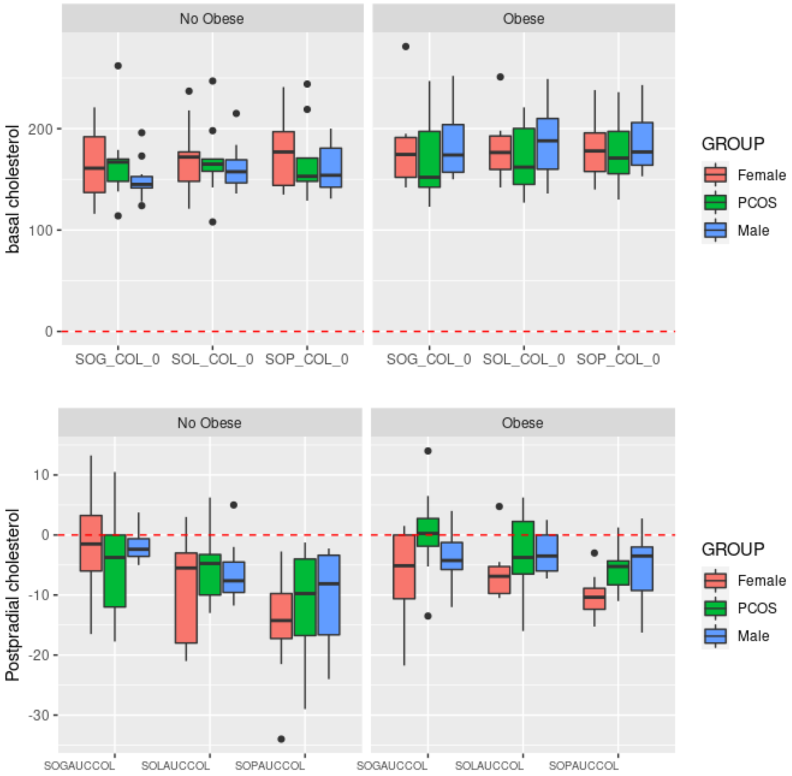
As expected, and in concordance with the literature, SHBG increases with age. Regarding the correlation with BMI, on healthy control women, this correlation is negative, as well as on obese PCOS and obese males. The other subjects show a positive correlation. A similar pattern is observed on WC and WHR, figures Error: Reference source not found-69 .



5. Figure: Basal and postprandial levels of HDL cholesterol. Each trio boxplots are arragange in the following way: Firstly the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Blue boxplots belongs to males, green to PCOS women and pink to control women.

A correlation matrix between HDL, LDL cholesterol, triglycerides, HOMA-IR, hscRP BMI, and age was computed. This correlation was performed on the whole subjects, to observe the general trend, figure 22.

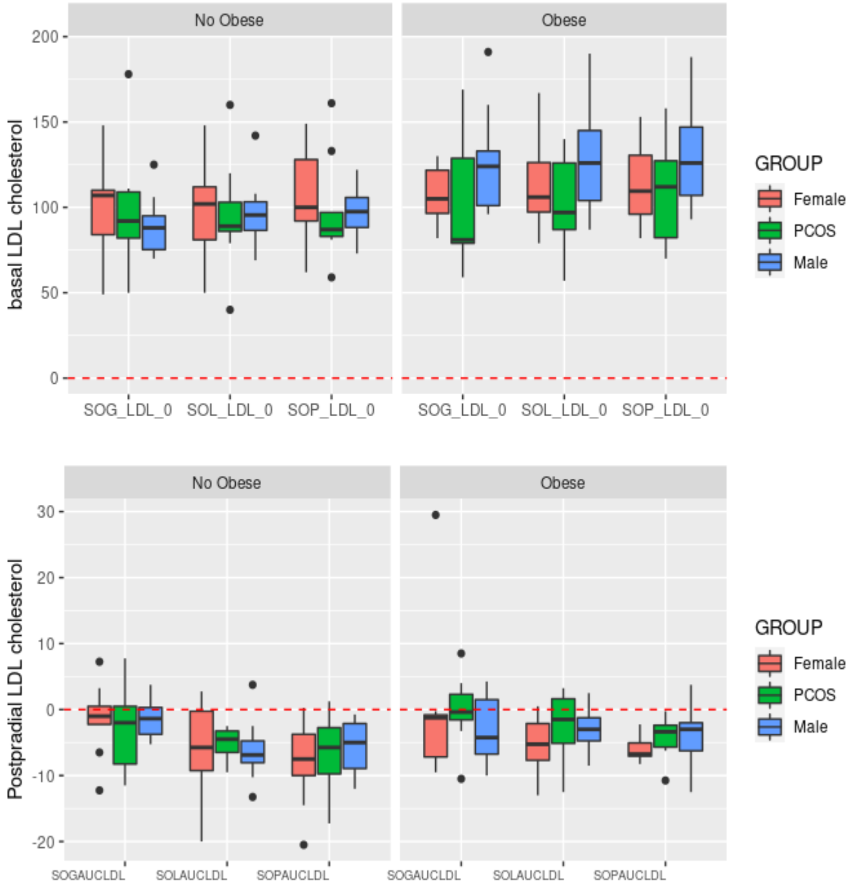
In general, age correlates possibly with BMI, hs-CRP, cholesterol, HDL, and LDL. There is no association between age, HOMA-IR, and triglycerides. BMI correlates possibly with all variables except on HDL, which is negatively correlated. hsCRP correlates positively with BMI, HOMA-IR, and TG. negatively with HDL, and no association with cholesterol and LDL cholesterol- HOMA-IR correlates positively with all variables except con HDL which is negatively correlated. TG and age are not correlated, but all variables except on HDL this correlation is positive.



6. Figure: Basal and posptrandial levels of cholesterol. each trio boxplots are arragange in the following way: Firstly the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Blue boxplots belongs to males, green to PCOS women and pink to control women.

Cholesterol correlates positively with all remaining variables. HDL has a negative correlation with all variables, and no association is found between cholesterol and LDL. All these associations agreed with the literature (51–53).

A positive correlation between insulin, glucose, and HOMA-IR levels was found and a negative correlation was found regarding ISI accordingly with the definition of these two variables. Differences were found when the correlation was performed between groups, figure 60.Regarding hsCRP and BMI, only on obese males, this correlation was negative. hs-CRP and HOMA-IR, no relationship exists between healthy control women, obese women, and obese PCOS women. A negative correlation is found in non-obese PCOS women and a positive correlation is observed between men. HsCRP and TG, show a negative correlation for women, except on obese PCOS women and males. HsCRP and cholesterol the correlation is negative in women and positive in males.



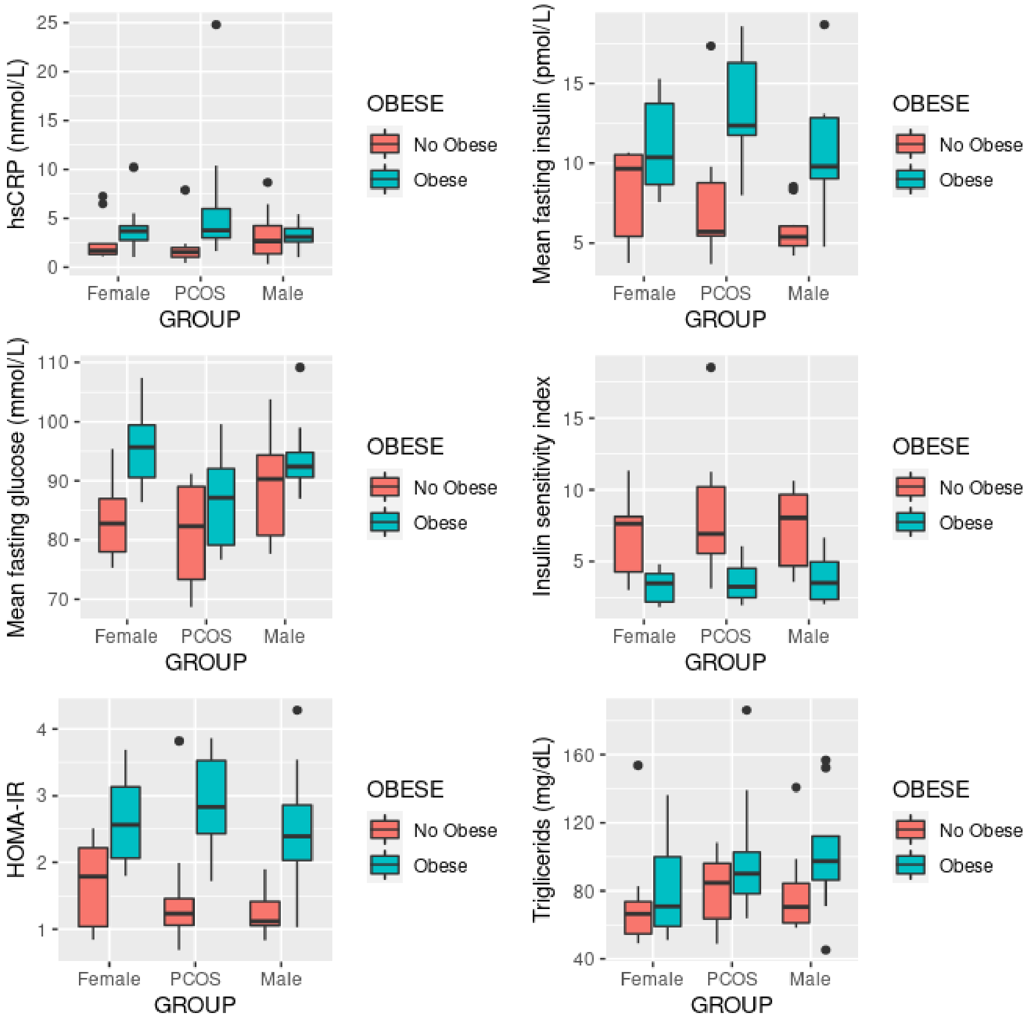
7. Figure: Basal and postprandial levels of LDL cholesterol. Trellis plot. Each trio boxplots are arragange in the following way: Firstly the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Blue boxplots belongs to males, green to PCOS women and pink to control women.

HDL cholesterol and hsCRP show a positive correlation between non-obese PCOS females and obese control females. A negative correlation is observed on the other subjects. LDL cholesterol and hscRP correlation are negative on females and positive on males. These results can be observed on the apendices, figures 61-68

Multivariate Analysis of the Clinical Data

When the data set is not scaled previous to the PCA, the first component explains 84%

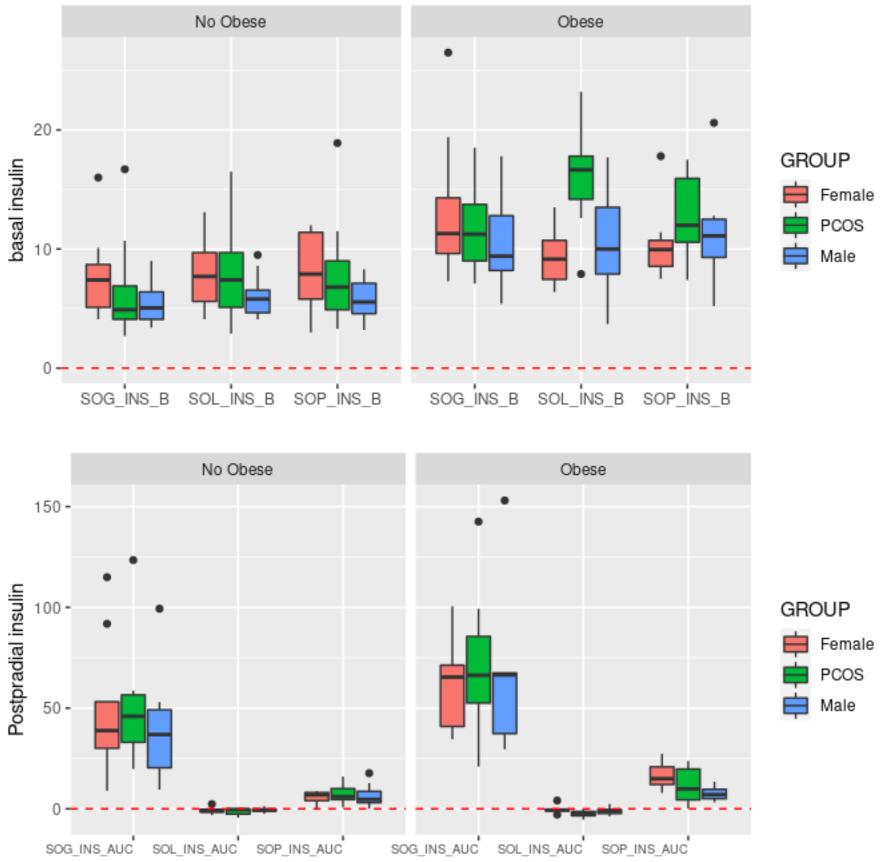
of the variability and the second component explains 7.2 % of the data set, figure 11. The first component differentiates well between males and females and PCOS women.



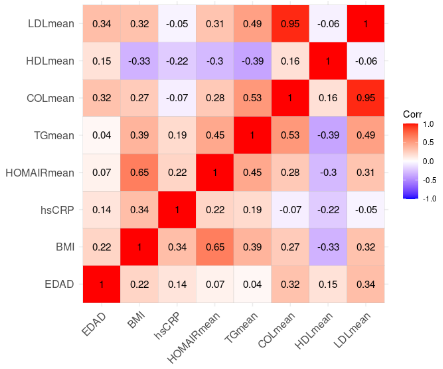
8. Figure: Clinical variables 1: Mean fasting insulin, glucose, triglycerids, the homeostatic model assessment (HOMA-IR) are the mean values across all days. Blue boxplots are obese subjects, and pink not obese. GROUP is the categorical variable that includes the sex levels and PCOS.

This is due mainly to the free and total levels of testosterone and the ratio between free levels of testosterone and estradiol. Moreover, the second principal component divides the subjects into obese and non-obese, especially males, however, it is not possible to observe this distinction between females. Notwithstanding, when the data set is scaled to unit variance, the first component explains only 25% of the variability.

The subjects are separated by obesity, but the distinction between sex is barely appreciable. The main variables that distinguish between sex, if the data is scaled, are the ratio between free levels of testosterone-estradiol and the total testosterone. Regarding the distinction between obese and non-obese subjects, the variables with the most impact are cholesterol, low density lipoprotein cholesterol (LDL), triglycerides (TG), wasit to hip ratio (WHR), and the high sensitivity c-reactive protein (hsCRP). Differentiation between lean and obese subjects is due to HDL, SHBG, and Glucose.



9. Figure: Basal and postprandial levle of insulin ach trio boxplots are arragange in the following way: Firstly the oral glucose load, secondly the lipid challenge, and finally the protein oral ingestion day. Blue boxplots belongs to males, green to PCOS women and pink to control women.



	EDAD	BMI	hsCRP	HOMAIRmean	TGmean	COLmean	HDLmean	LDLmean
LDLmean	0.34	0.32	-0.05	0.31	0.49	0.95	-0.06	1
HDLmean	0.15	-0.33	-0.22	-0.3	-0.39	0.16	1	-0.06
COLmean	0.32	0.27	-0.07	0.28	0.53	1	0.16	0.95
TGmean	0.04	0.39	0.19	0.45	1	0.53	-0.39	0.49
HOMAIRmean	0.07	0.65	0.22	1	0.45	0.28	-0.3	0.31
hsCRP	0.14	0.34	1	0.22	0.19	-0.07	-0.22	-0.05
BMI	0.22	1	0.34	0.65	0.39	0.27	-0.33	0.32
EDAD	1	0.22	0.14	0.07	0.04	0.32	0.15	0.34

10. Figure: Correlation matrix plot between the mean basal values across all days of LDL, HDL, cholesterol, HOMA-IR, hsCRP, BMI, and age

3.1.2 Intenstinal

3.1.3 Permeability markers data set

Univariate Descriptive Analysis
All variables on this data set show higher values at the basal levels than at the postprandial time. The cluster differenciation 14, (CD14) ,figure 12, at basal levels, is higher on non-obese subjects. Especially on healthy control women. No distinction between the other groups is found regarding non-obese subjects. Obese PCOS women show increased levels of this IP, but lower than healthy control women. At postprandial levels, similar values are found across all subjects. But on the second and third days, on non-obese subjects males and PCOS women show higher values. Obese males on the first day show the highes values at postprandial levels.

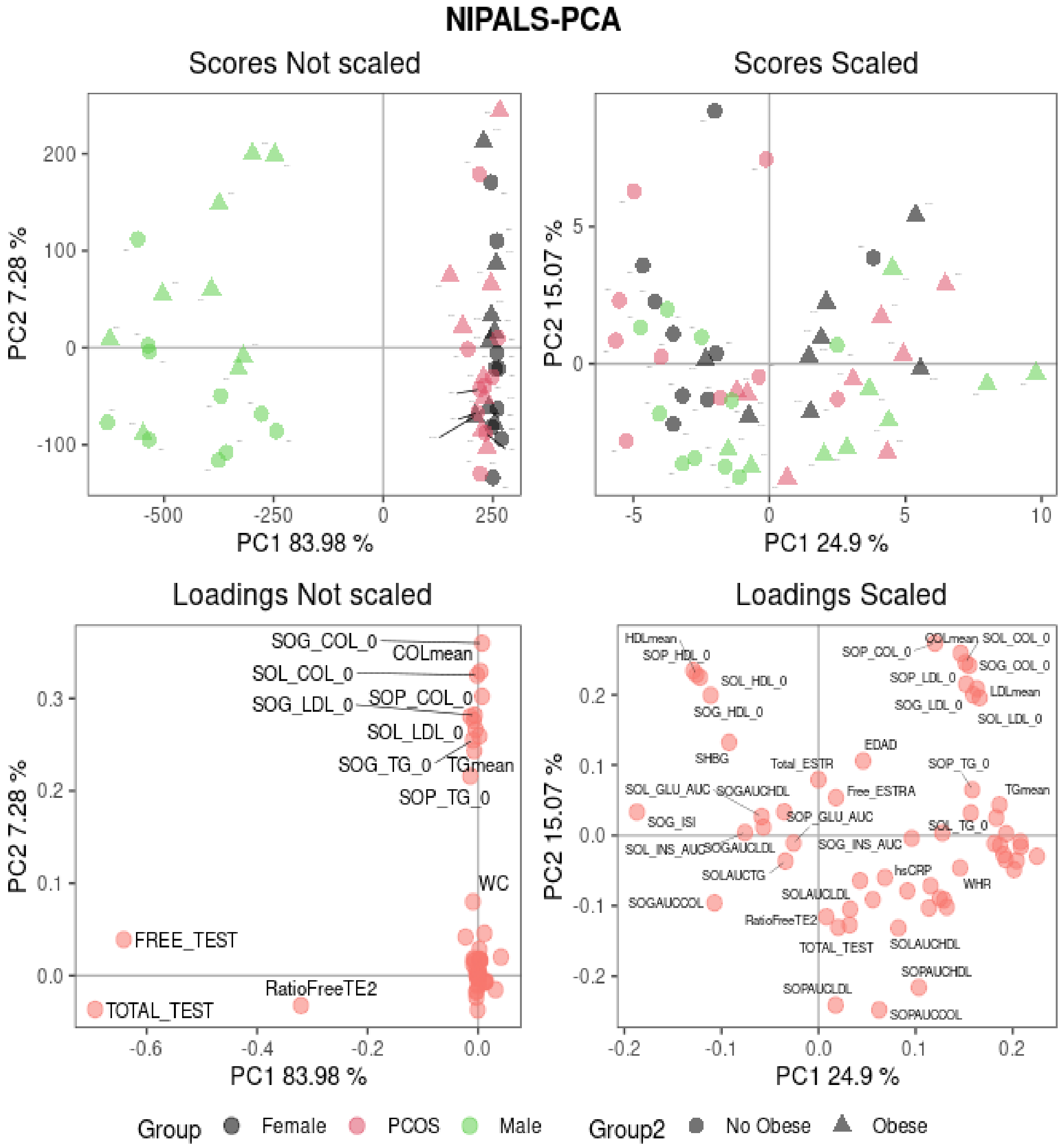
LBP mean basal values are higher on obese subjects, figure 23, specifically on PCOS obese women. And no distinction between groups is observed at postprandial levels.

GLP2 values are higher at basal levels on obese PCOS women and obese males. Between control women, no distinction is observed. At postprandial levels, higher values are observed on the second and third days. With no distinction between non-obese subjects. Obese control subjects show higher values than obese PCOS women, figure 15.

Zonulin mean basal levels show no distinction between groups. At postprandial levels, on the second day, between obese subjects, control females have the highest values followed by PCOS and males, fugure 13. Succinate levels are higher on obese subjects, but remarkably higher on obese women with PCOS. On the postprandial time, the first day, levels of this metabolite are higher, being higher on nonobese males and obese PCOS women. On the other days, no distinction is found between groups. These results are shown on figure 14

Bivariate analysis

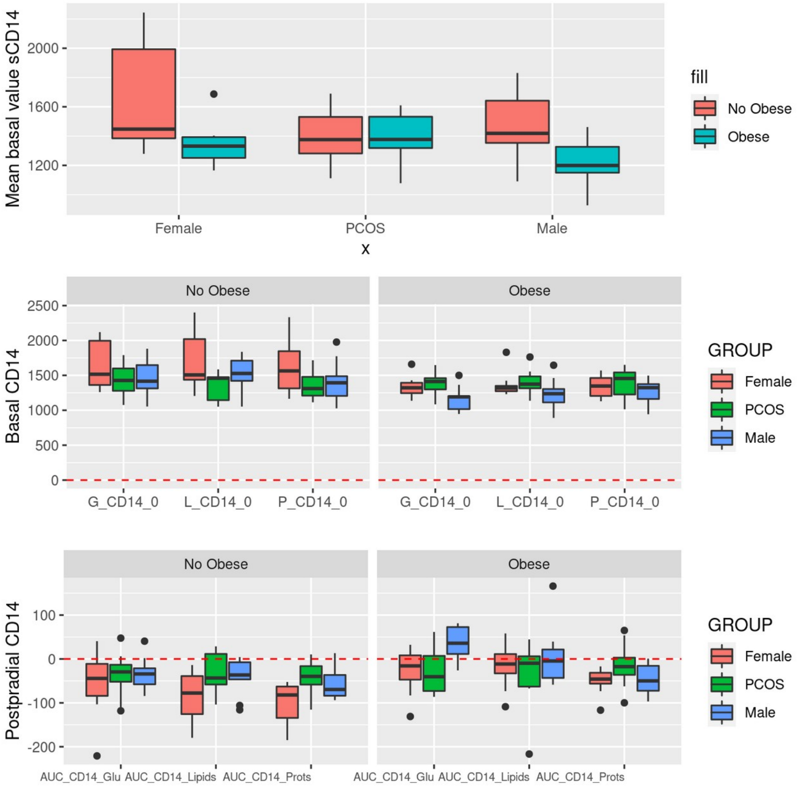
A correlation-scatter matrix plot, which can be observed on figure 16, was performed to observe the relationships among the mean basal values of the variables, BMI included, as is an indicator of diabetes II and inflammation (54).
BMI, correlates positively with all variables, except on sCD14. But CD14 correlates positively with LBP and Zonulin, and negatively with GLP2 and succinate. The correlation of LBP with GLP2, zonulin and succinate is positive. There is no relationship between GLP2 and zonulin, but a positive correlation exists between succinate. Finally, succinate and zonulin present a barely positive correlation. All these relationships are described also in (55–58).



11. Figure: PCA with the NIPALS method. Green objects are male, black objects are females, and the pink ones are PCOS women. Circles are lean subjects whilst triangles are obese subjects

Multivariate Analysis Results

Classical PCA, figure 17, was computed to observe the possible differences between groups of patients. The first principal component accounts for almost 20% of the total variation, whilst the second principal component explains 16%. Obese and non-obese subjects are separated by the first component, meanwhile, the second component differentiates the sex. Zonulin, lipopolysaccharide binding protein (LBP) and succinate differentiate obese subjects and the cluster differentiation 14 (CD14) at basal time differentiate lean subjects. The postprandial levels of Glucagon Like Peptide 2 (GLP2) on the lipid and glucose ingestion days, succinate, Zonulin and LBP, on all postprandial days, makes the distinction on lean subjects The Body Mass Index (BMI) makes the distinction on obese subjects. This marker though is not a IP is included on this data set because on obese subjects, this measurement may lead to a chronic inflammation situation. On the plot, 0 means the measurements of the metabolites at the basal stage, and the letters P, L and G correspond to proteins, lipids and glucose ingestion days.



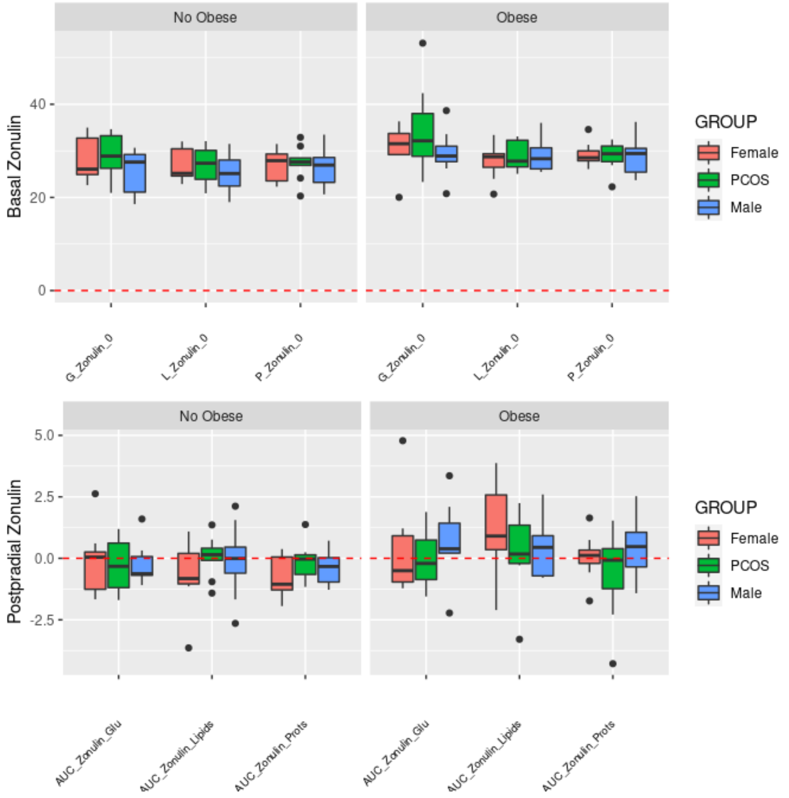
12. Figure: Basal and posptrandial levels of the cluster differenciation 14 (CD14)

3.1.4 Metabolome

Univariate statistical analysis

Mean basal values of ***alanine*** are higher in obese subjects than non-obese, but this relationship is only observable in men and women with PCOS. Separately on each day, one can visualize that, women with PCOS have higher concentrations of this metabolite especially in the obese ones. Basal levels are higher than a postprandial stage. On the lipid load day, the subjects present a lower concentration of Alanine in comparison with the other days. No differences are observed between subjects. Higher values are present on the first day in comparison with the second. On the last day, the protein load day, subjects present higher levels of Alanine than the other days, figure 19

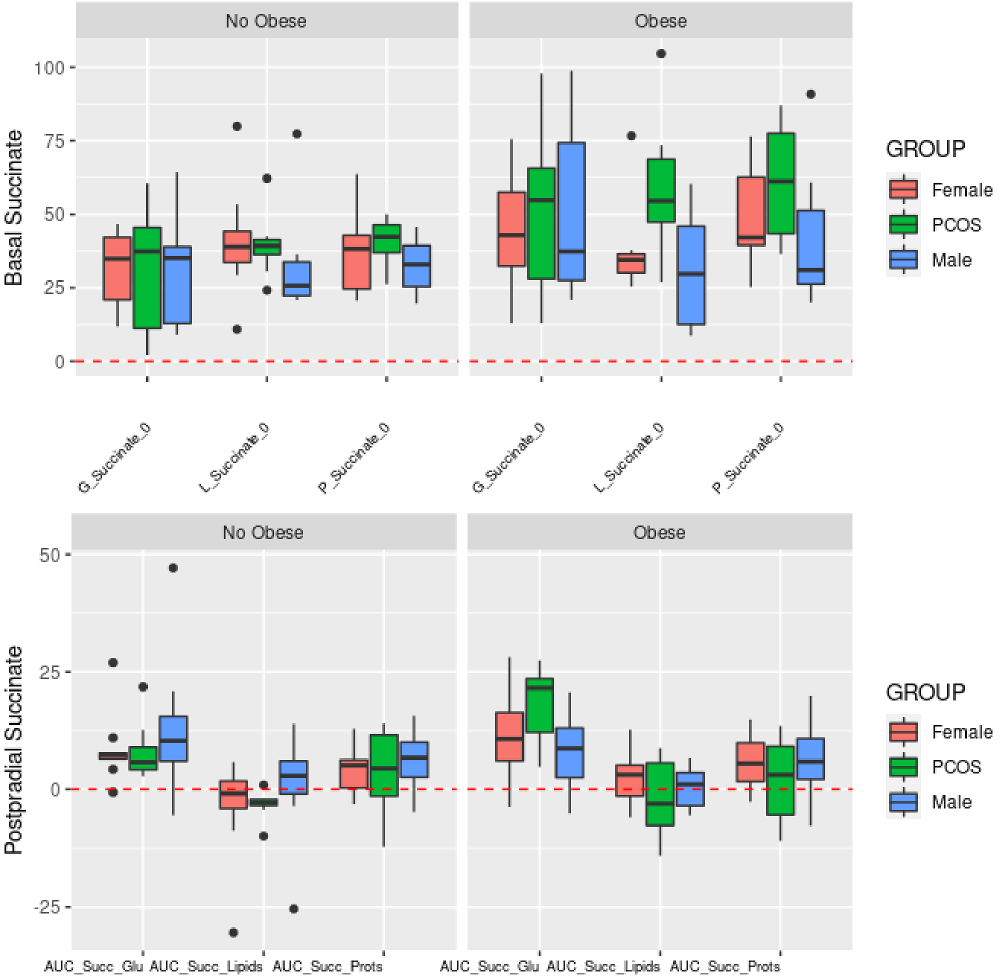
The mean basal values of ***acetate***, figure 23 are higher in non-obese females in comparison with obese females. The contrary happens with men. Reversely, healthy women present higher values than PCOS non-obese patients. However, obese PCOS women have higher values than control obese women.. The PCOS women have higher than men if they are non-obeses and the contrary in obeses individuals. Every day, obese men have higher values than obese PCOS women and these higher than obese females. This last is pointing out the relationship between basal acetate values and obesity.Basal values are higher in comparison with postprandial levels. There are almost no differences across subjects among all postprandial days.



13. Figure: Basal and posptrandial levels of zonulin. Pink boxplots are women, the green ones are females with the syndrome, and the blue boxplots are control males

The ***pyroglutamic acid ,*** levels, *figure 18*, are higher in non-obese subjects, with the lowest values on lean PCOS women.Postprandial levels are lower than basal levels. On the second day, the lipid oral day, these values increase being comparable with basal levels and are higher in both obese and non-obese women than men.

Lysine_1 levels at basal time, are higher in obese subjects, but this difference is observable only between women. PCOS women have higher values than control women. Postprandial levels are lower than basal levels, except on the protein load day, with obese PCOS women having higher values than the rest of the obese subjects. Non-obese men and non-obese PCOS women have higher values than control women. And obese PCOS women present higher values than the other obese subjects. In women, the ***Lysine_1*** is associated with PCOS, especially in the obese ones both at basal and postpandrial levels. Figure 21



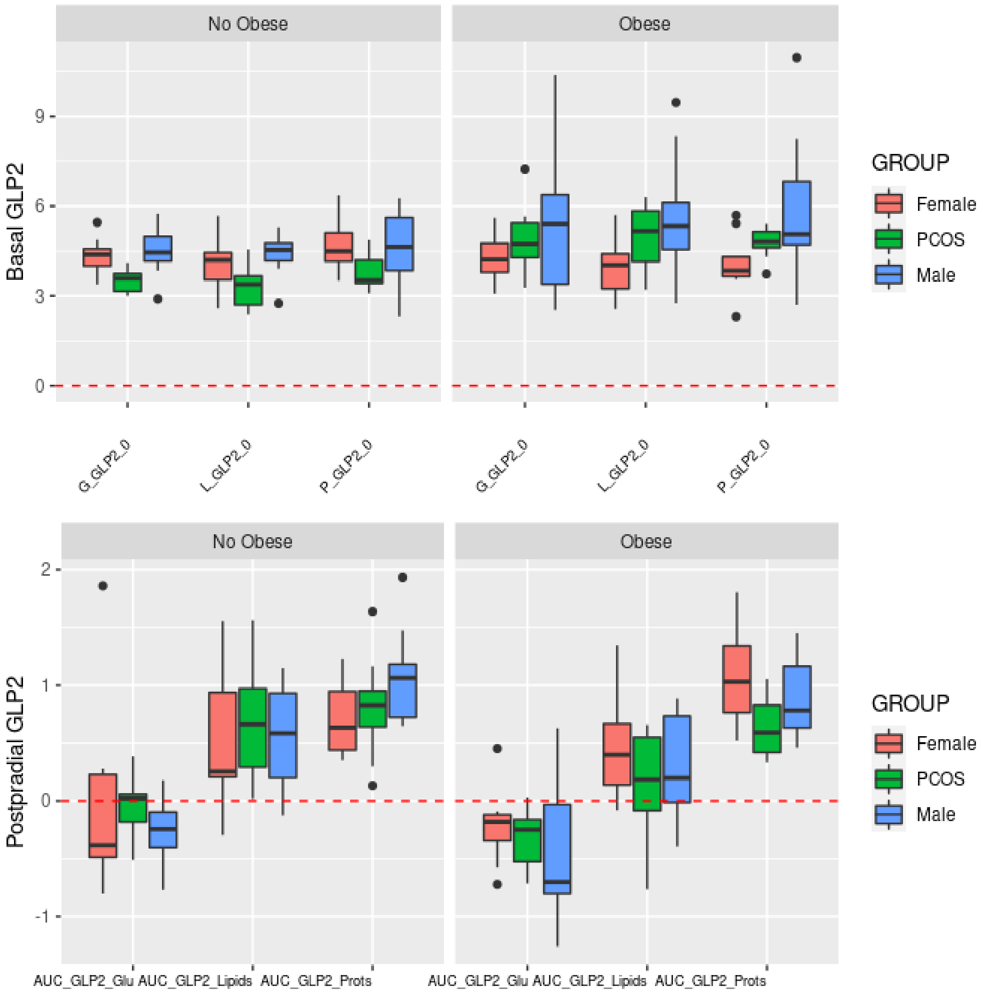
14. Figure: Basal and posptrandial levels of succinate. Pink boxplots are women, the green ones are females with the syndrome, and the blue boxplots are control males

Basal levels of ***formate***, figure 20, have the highest value on healthy control women when they are not obeses. However, between obeses PCOS women have the highest values. Postprandial levels are comparable with basal levels. They were lower on the first and second days in non-obese individuals. Obese control women show higher values at the postprandial stage.

The postprandial levels in general are lower than the basal levels, leading to negative values, this is due to the fact that are arbitrary units that came from the technology used, which is nuclear magnetic resonance.

Bivariate Analysis

Each metabolite is associated with many pathways, and the fact that most of them are amino acids, we performed a correlation matrix plot. From this plot, we infer that the correlations, whether positive or negative, may be due to the physical/chemical properties as it shown on figure 24



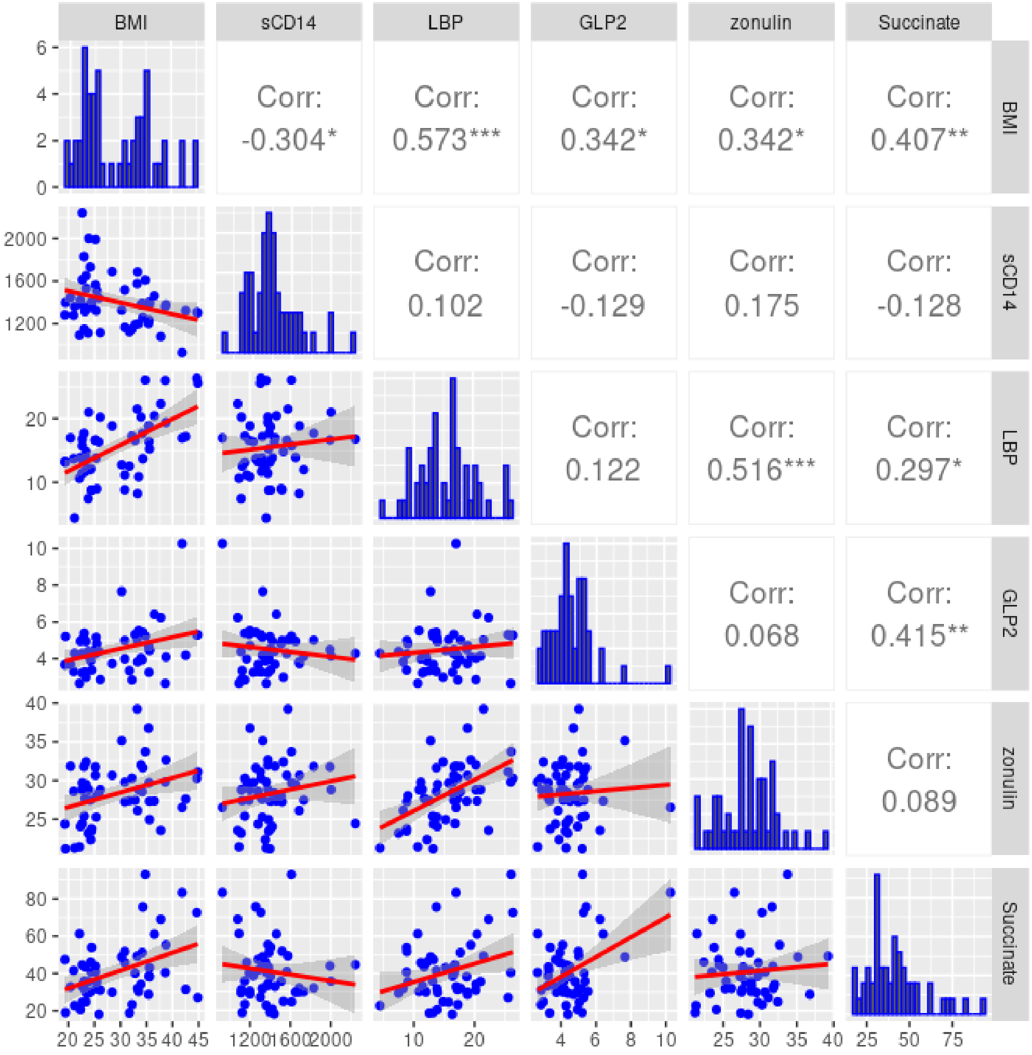
15. Figure: Basal and posptrandial levels of GLP2 Pink boxplots are women, the green ones are females with the syndrome, and the blue boxplots are control males

For instance, Oxoisocaproic, Leucine, Isoleucine and Valine are positively correlated. These metabolites have in common that Oxoisocaproic acid is derived from leucine, isoleucine, and valine. And leucine, isoleucine and valine are non-polar.

It has to be mentioned that, the correlation matrix was computed on the mean basal values across all days. And no correlation between groups were made.

Multivariate Statistical Analysis (NIPALS-PCA)

We have computed PCA via the NIPALS algorithm, figure 22, centered and scaled. The differentiation between all groups is very clear. The first principal component accounts for 20% of the variance and makes the distinction between sex and lean control women and obese women with the syndrome. Whilst the second component explains 5.7% and differentiates between obese PCOS women and control obese females.PCOS women, and especially obese PCOS women, appear to have a very similar concentration of metabolites to men. The distinction is clear between lean PCOS women and lean women. Though there are subjects that do not belong to a specific cluster. The pyroglutamic acid, formate, and acetate segregate the group of lean control women. Meanwhile, acetone, alanine, and lysine differentiate PCOS obese women and males. So far, is the best block that differentiate between each group. Even better that the Clinical Data.



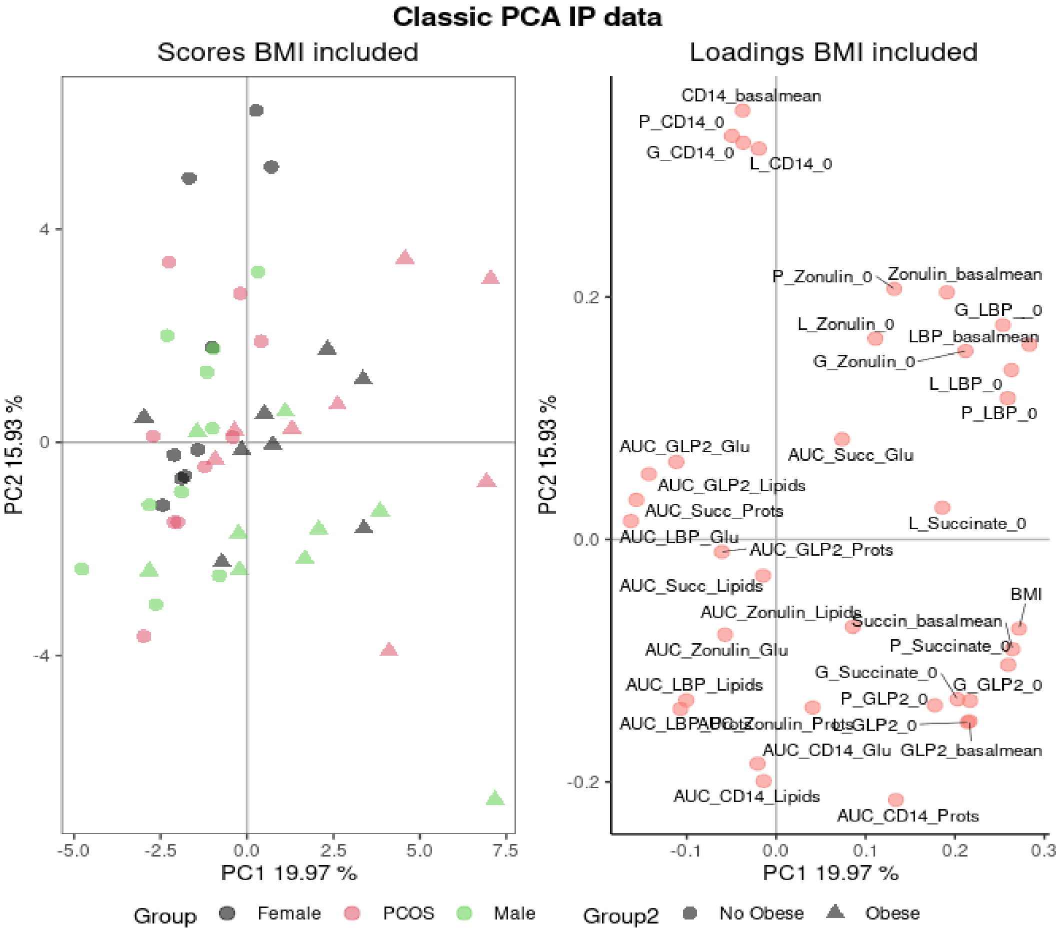
16. Figure: Scatter OLS matrix. The correlation between BMI, sCD14, LBP, GLP2, zonulin and succinate.

3.1.5 Microbiome

Data from the gut was extracted from the faces of the subjects. As previously mentioned, we do not have information on 5 subjects. Thus we performed a chi-square test on the subjects to check whether this change may affect the data. The null hypothesis of no relationship is accepted.

The approach for the descriptive analysis was different from the previous datasets. From the raw reads of the absolute abundance, to perform the multivariate descriptive analysis, apart from the MDS techniques used, CCA and PLS-DA were computed. With

CCA we could not obtain a good differentiation between groups, thus we transformed the data into a log-center ratio, and then the PLS-DA technique was used.



17. Figure: Classical PCA was computed due to the fact that on the IP data more observations than variables were present. Green dots are males, the pink ones are PCOS females, and the black dots are control women. Triangles represent obese subjects, and circles corresponds to lean subjects

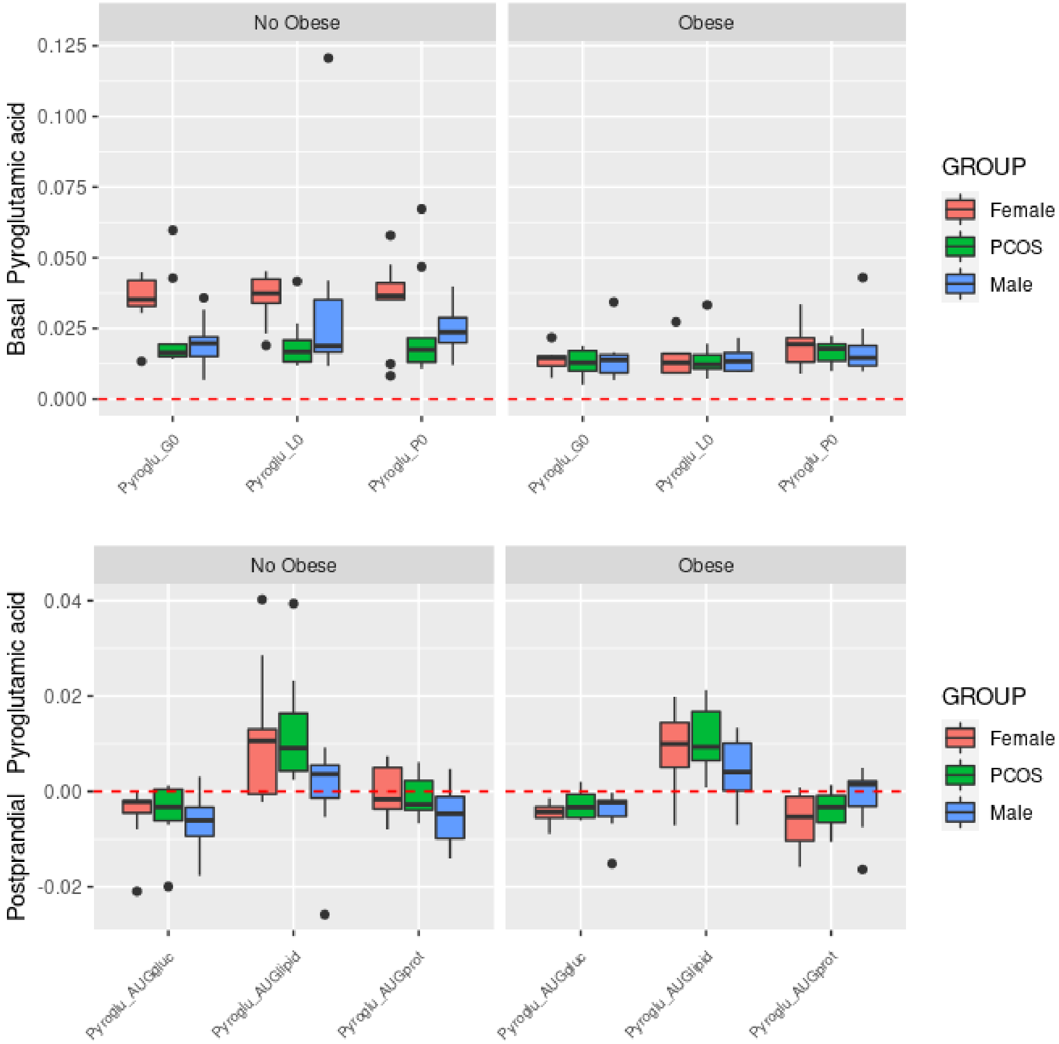
Phylum

We had at the beginning 24 phyla. However, there were cases where certain phyla it is present in one sample. Iby removing those phyla, ee ended up with 17 phyla. Moreover, to perform the PLS-DA analysis we filtered the data up to a minimal abundance of 0.01% ending up to 9 operational taxonimic units (OTU).

First of all, we worked with the raw read data. We made use of the package *phyloseq* to observe the most abundant phylum.

The data tells us that the most abundant phylum is Firmicutes, farly followed by Bacteroidetes and Actinobacteria. One can notice on figure 28 an increase of several phyla on PCOS women, specifically on obese subjects.

Thus secondly we computed the Shannon and Chao 1 index, figure 27, which are indices of ***alpha*** diversity. Both indices show a higher diversity of lean control women followed by lean PCOS women. However in men, it is the contrary, being higher in obeses .



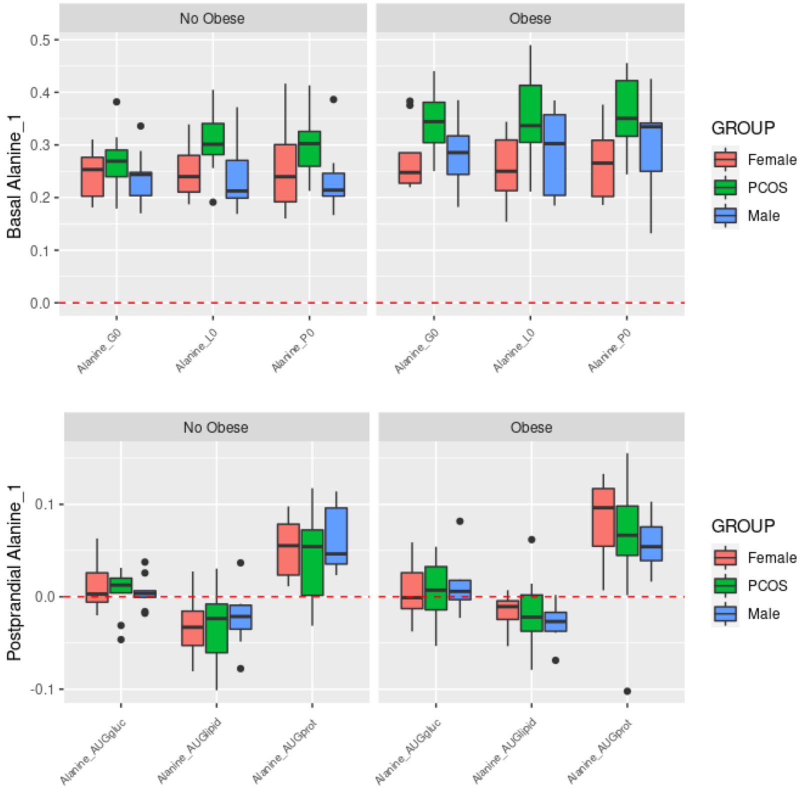
18. Figure: Basal and Postprandial levels of pyroglutamatic acid

Beta diversity, on the other hand, is a similarity index, thus we need to compute a Multidimensional Scaling technique. If the resulting matrix is euclidean or not, then we proceed with a metric or non-metric MDS respectively

For both indices, the similarity matrix did not follow a euclidean distribution, thus, we opted for non-metric MDS techniques, figure 25. We made use of the isotonic regression, for the Jaccard index and a monotic regression for the Sørensen index. These engines were chosen because of the lower stress obtained.

From the Jaccard index, one can observe that control women are separated from PCOS women and males. However, no distinction between the *OBESE* group is observed.

Regarding the Sørensen index, several subjects are overlapped, thus no conclusions can be drawn**.**



19. Figure: Basal and postprandial levels of alanine 1. Pink, green and blue boxplots corresponds to control females, females with the syndrome, and control males. The first trio of boxplots belongs to the oral challenge day of glucose, the second to the protein ingestion day, and the thith trio to the portein day.

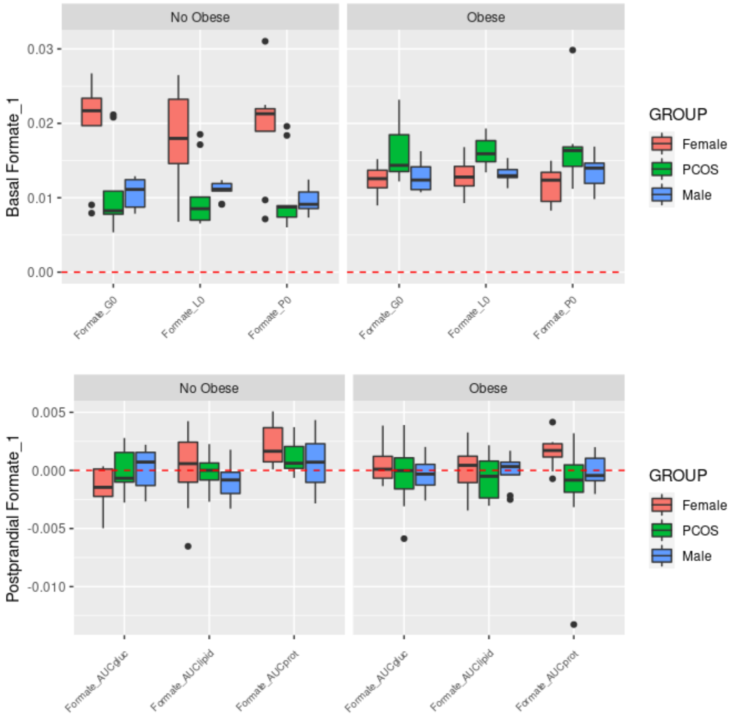
Univariate Descriptive Analysis

Actinobacteria and Candidus Saccharibacteria have an especially more abundance on obeses, figure 30 moreover, Actinobacteria, has the lowest abundance on lean PCOS women. Moreover, Verrucromicrobia is more likely to be present on non obese PCOS women. Firmicutes, which is the phylum more present, its abundance is similar on males, however, on obese females with PCOS, has similar values in comparison with men, being higher than lean women with the syndrome and control females. Bacteriodetes which is the second most abundance phylum on the subjects, has similar values across all sobjects, except on obese PCOS women and obese males, with similar values. Finally Proteobacteria, has similar values across all subjects.

Multivariate Analysis results

As mentioned, the best results were obtained with PLS-DA analysis, figure with two components. The optimal number of components was obtained via the CV (Cross Validation) technique. Males and PCOS females are separated by the first component which accounts for 27% of the variability. Being control females placed between these two groups.and the second component that explains 16% differentiate between obese subjects. Furthermore, the phyla that contribute more to this differentiation are

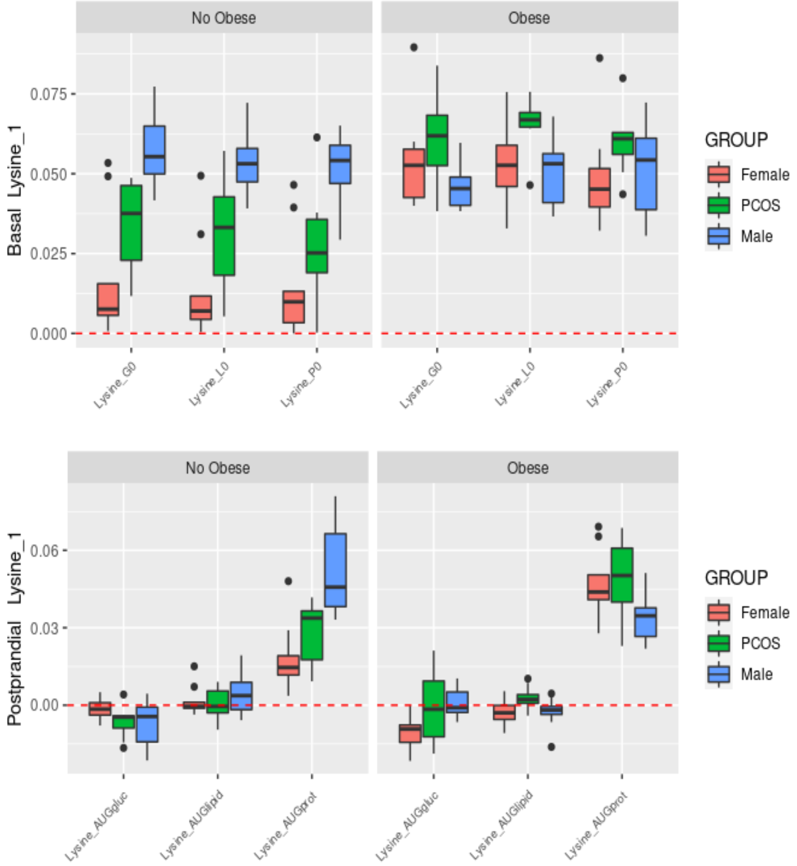
Bacteroides, Firmicutes, Candidatus Saccharibacteria, Actinobacteria and Proteobacteria for males, and Verrucomicrobia and Synergistetes for PCOS females.



20. Figure: Basal and postprandial levels of formate.Pink, green and blue boxplots corresponds to control females, females with the syndrome, and control males. The first trio of boxplots belongs to the oral challenge day of glucose, the second to the protein ingestion day, and the thith trio to the portein day.

Genera

The gut microbiome comprises 282 genera. However, as before, not all genera were correctly classified, thus we ended up with 213 genera. As we proceed with the phyla, we used the 213 genera for alpha and beta diversity. When computing PLS-DA, the log-center-ratio transformation was done, with 0.01% of minimum abundance across all samples. So we ended up with 83 genera.



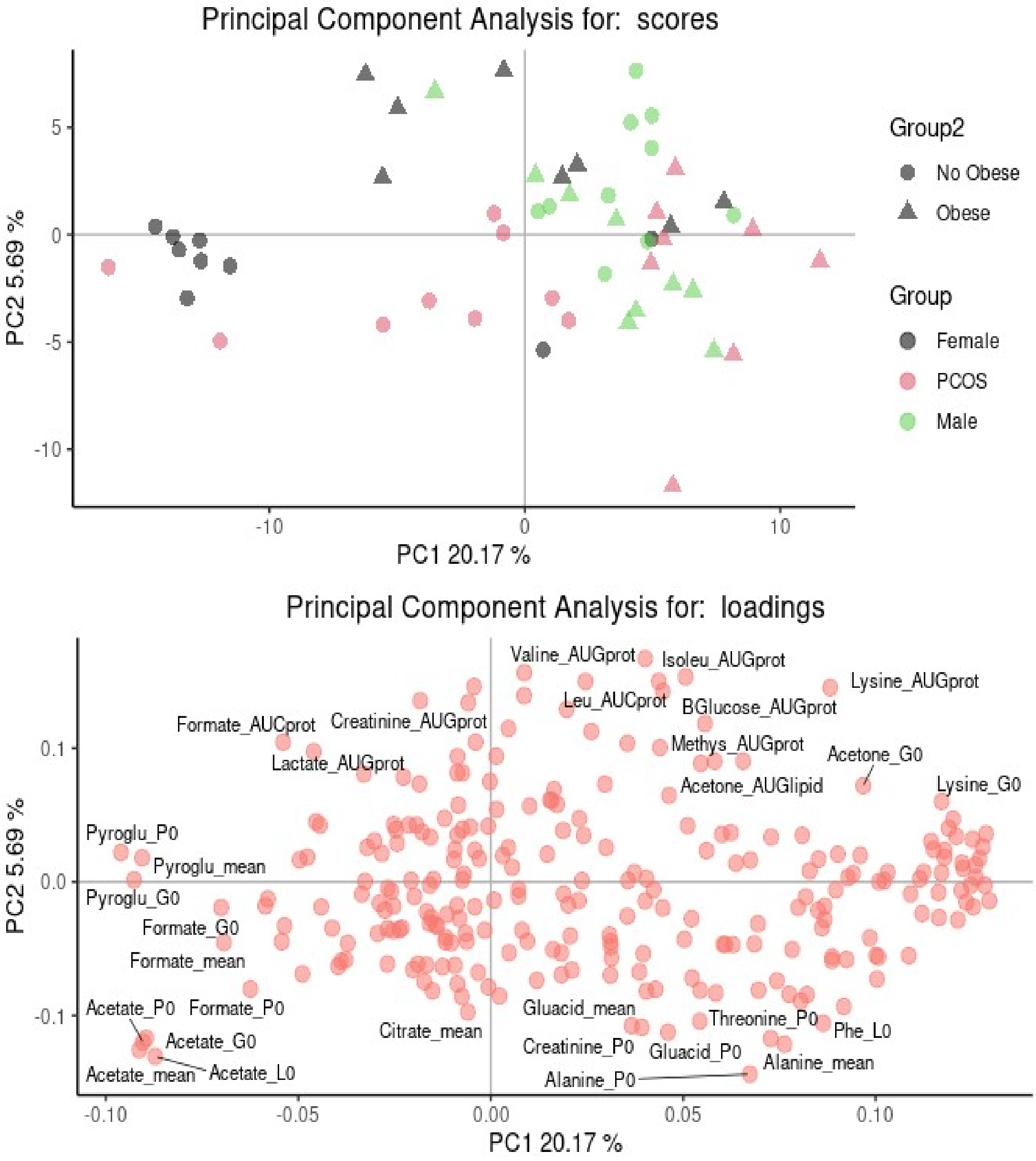
21. Figure: Basal and postprandial levels of Lysine

Multivariate Descriptive Analysis

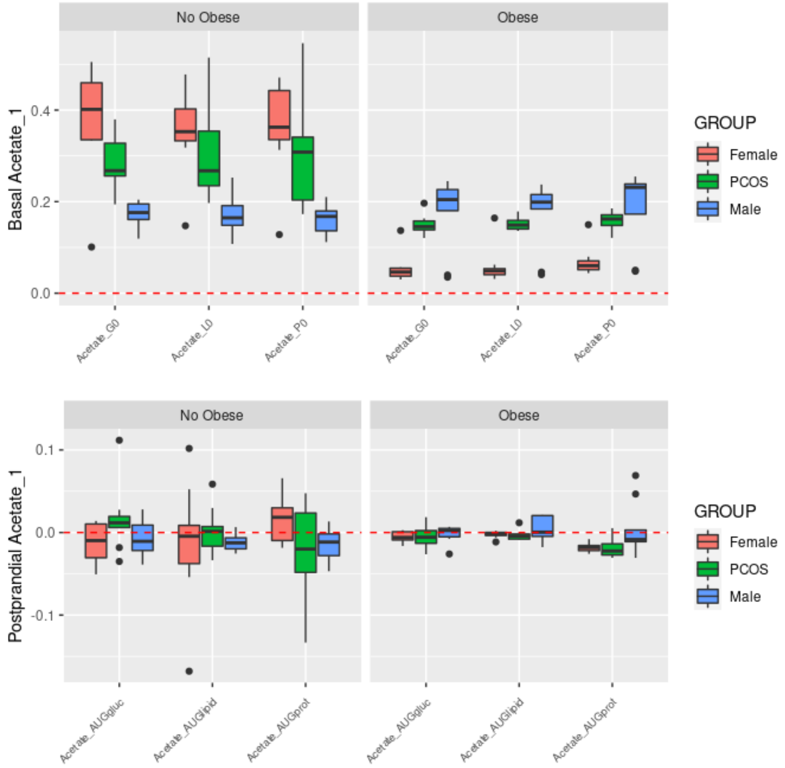
PLS-DA was computed with the log-center-ratio transformation. The first component accounted for 12.64% of the variance. This component differentiates between obese PCOS women and lean PCOS women at both extremes of the plot. Whilst the second component explains only 7% and differentiates between sexes . No further distinction among groups is observed, figure 37.

Regarding the variables, Fusicatenilbacter, Dorea are associated with PCOS obese women and Oscillibacter, and Desulfovibrio with PCOS lean women contributing That are the genera that most contribute to this dissemination.

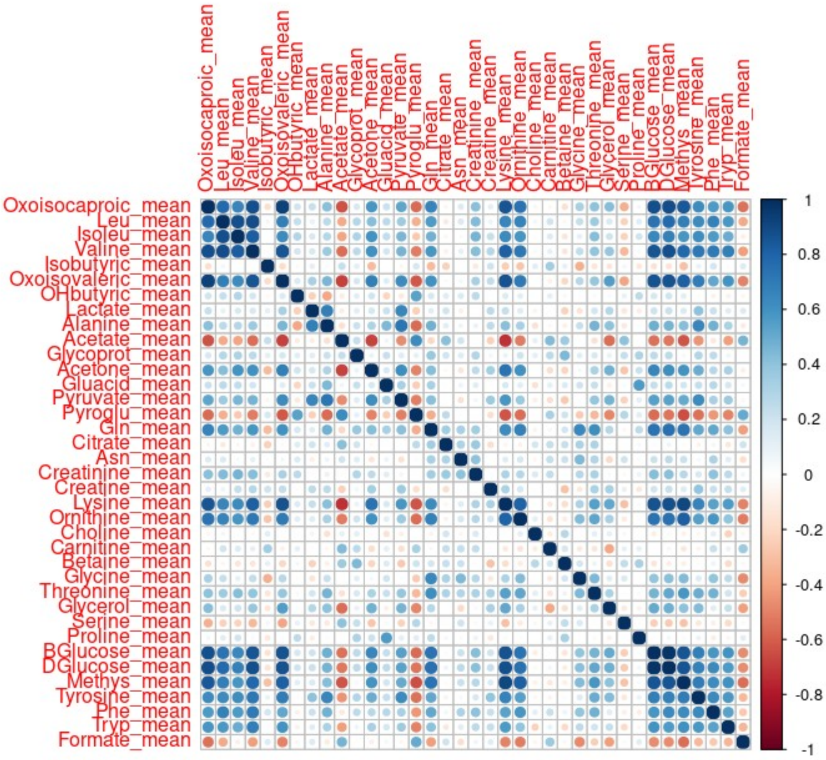
The ***Alpha*** diversity, with the Chao1 index, indicates that there is more abundance of the several genera on obese males and similar abundance between non-obese subjects. Surprisingly, obese women have the lowest abundance. The pattern in males,for the Shannon index, is similar to the Chao1 index. PCOS obese women, have a lower abundance than the rest of the subjects .



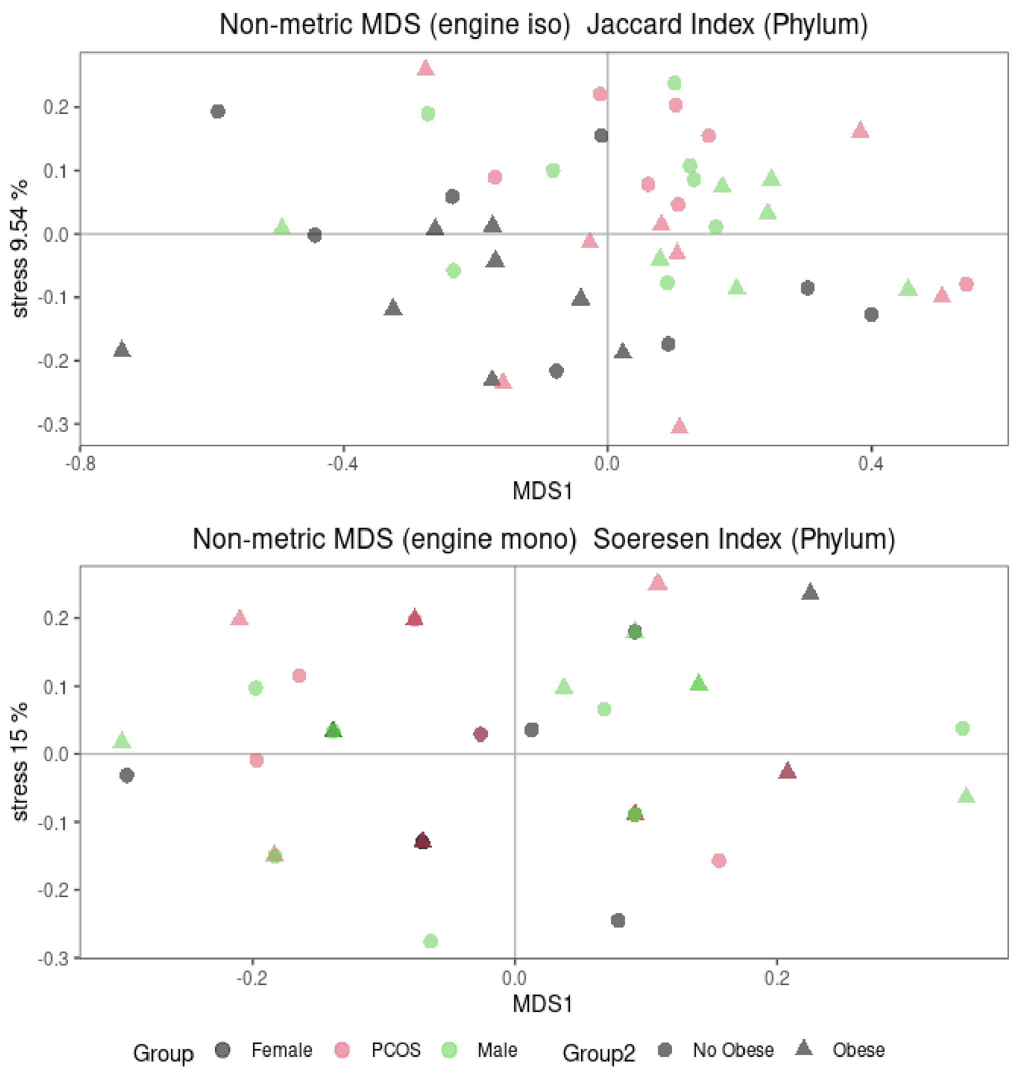
22. Figure: NIPALS-PCA analysis for the metabolome. Black, green and pink dots corresponds to control females, males, and women with PCOS. Triangles to obeses, and, circles to non obeses subjects



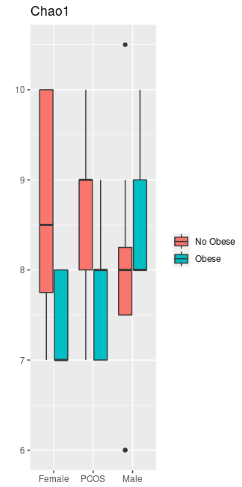
23. Figure: Basal and posptrandial levels of acetate



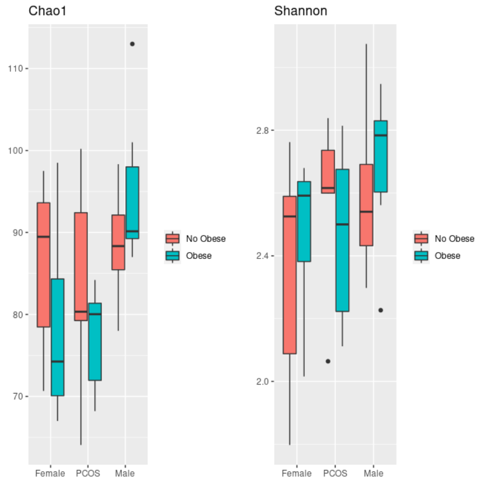
24. Figure: Correlation matrix plot. Mean basal values across all days of each macronutrient load.



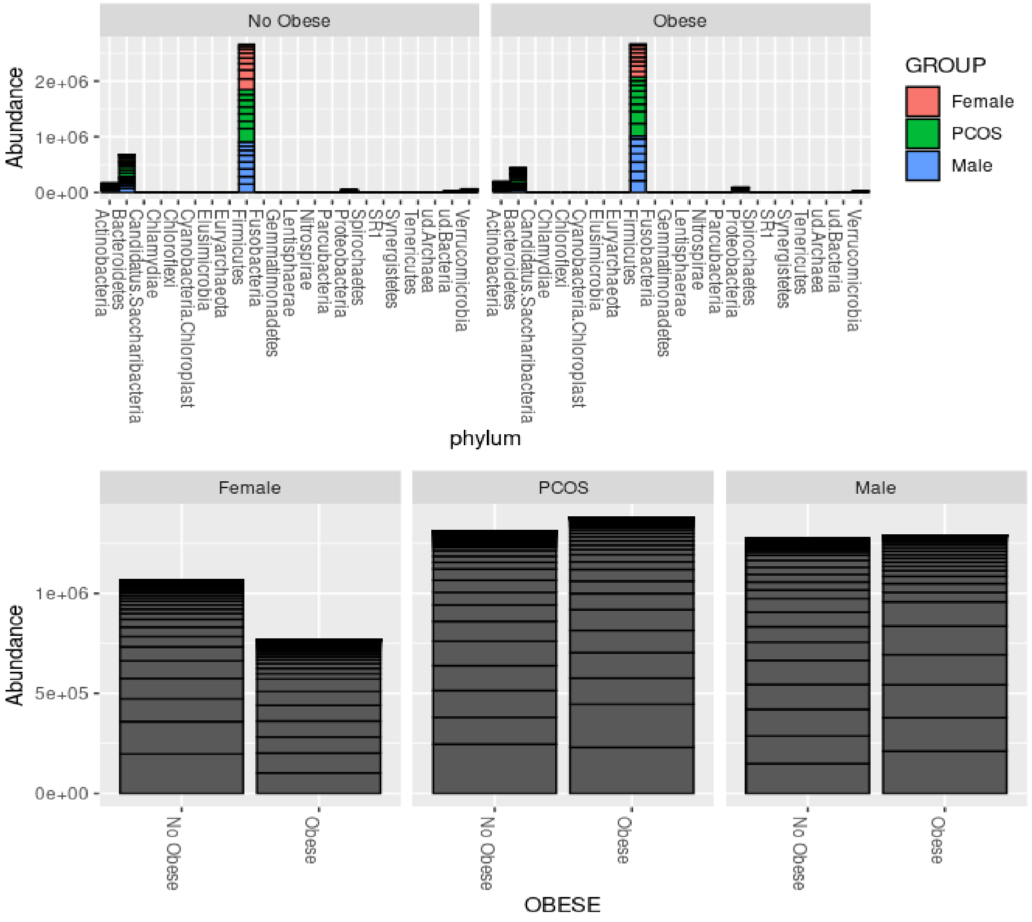
25. Figure: Jaccard similarities with isotonic regression techinque, and Sørensen similarities with monotic regression. Beta diverisity on the Phylum. Pink and blue boxplots corresponds to non obese and obese subjects respectively



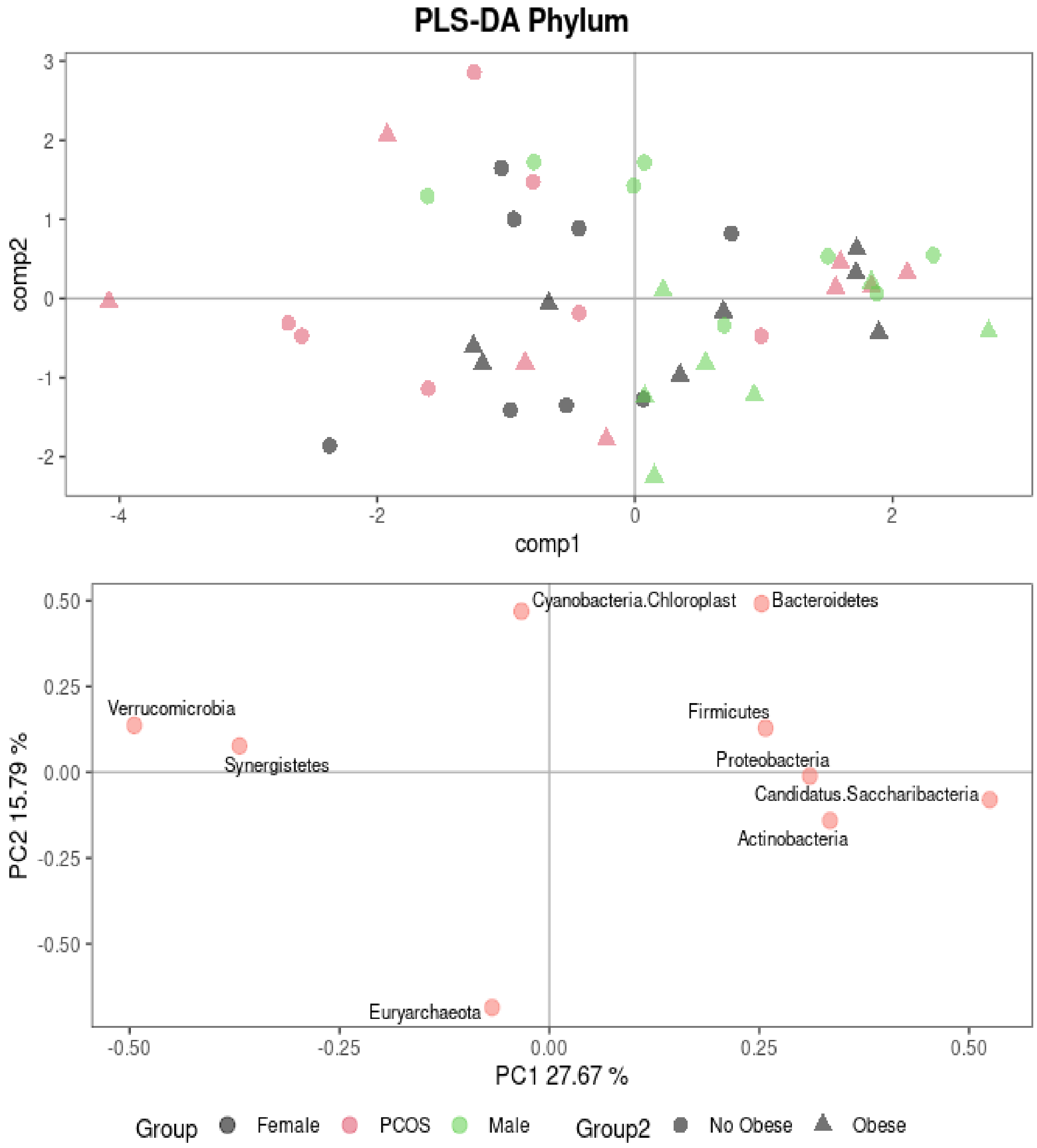
27. Figure: Chao1 and Shannon indices for the alpha diverisity on the Phylum. Pink and blue boxplots corresponds to non obese and obese subjects respectively



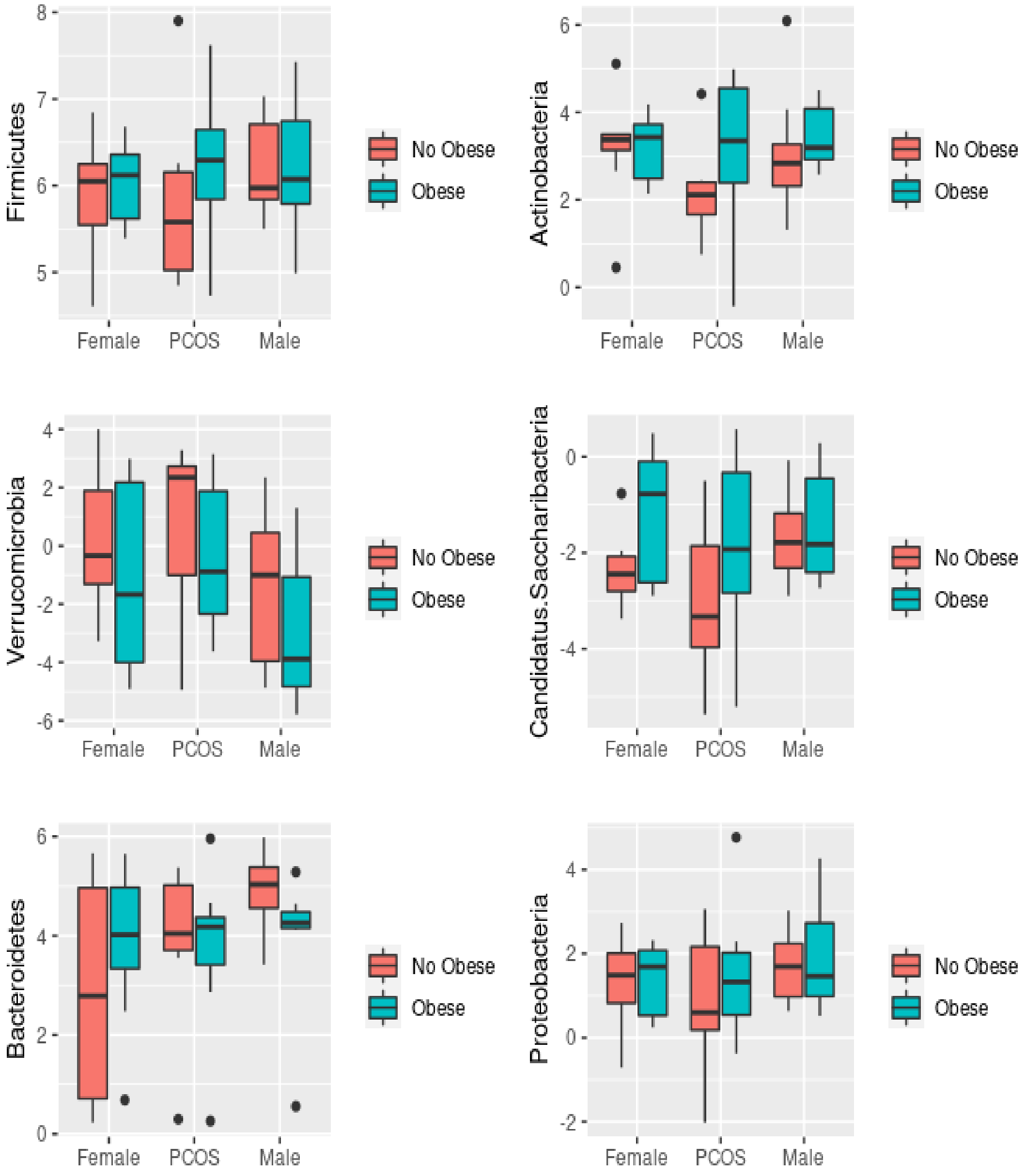
26. Figure: Chao1 and Shannon indices for the alpha diverisity on the Genus. Pink and blue boxplots corresponds to non obese and obese subjects respectively



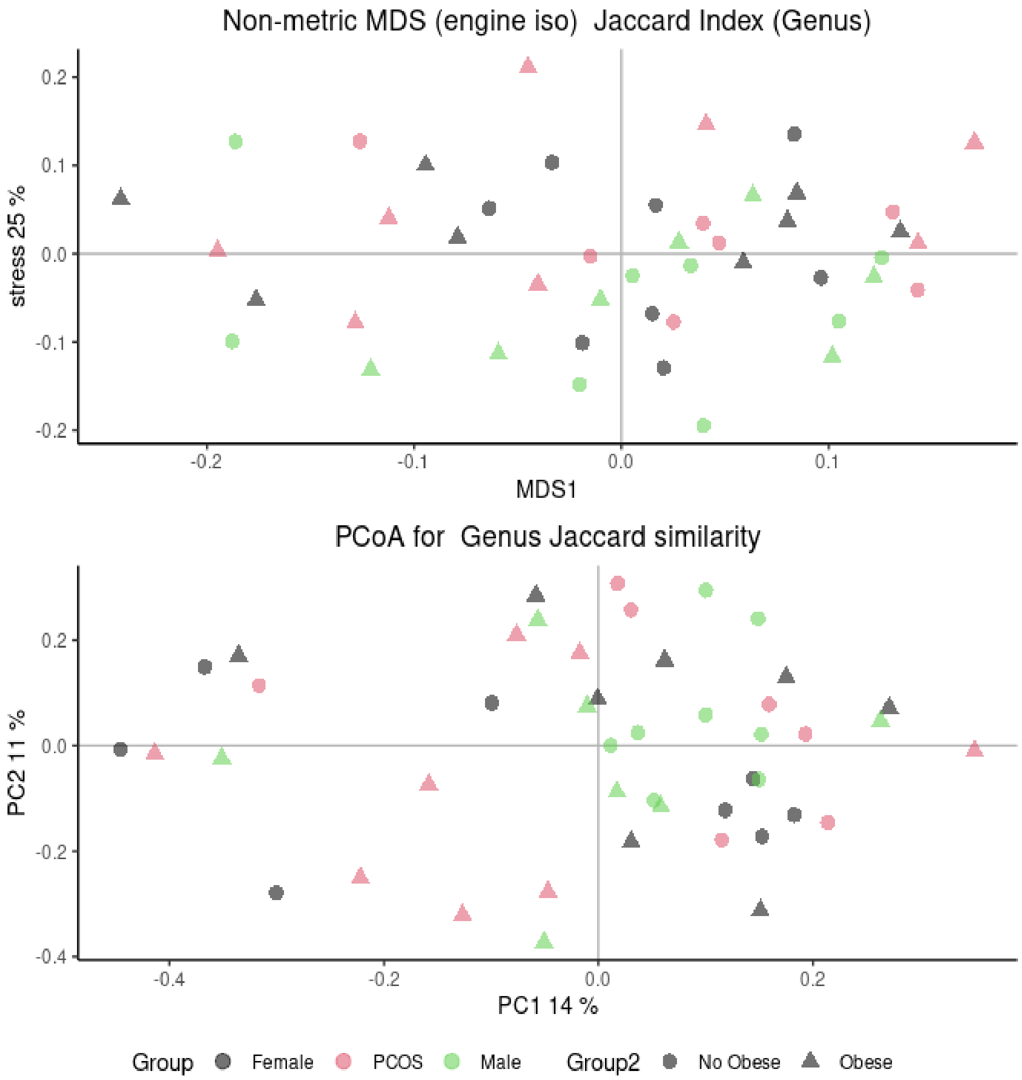
28. Figure: Barplot of the absolute abundance of the phylum



29. Figure: PLS-DA phylum. Black, green and pink dots are women, men and females with PCOS. Triangles are obeses and circles are non obeses



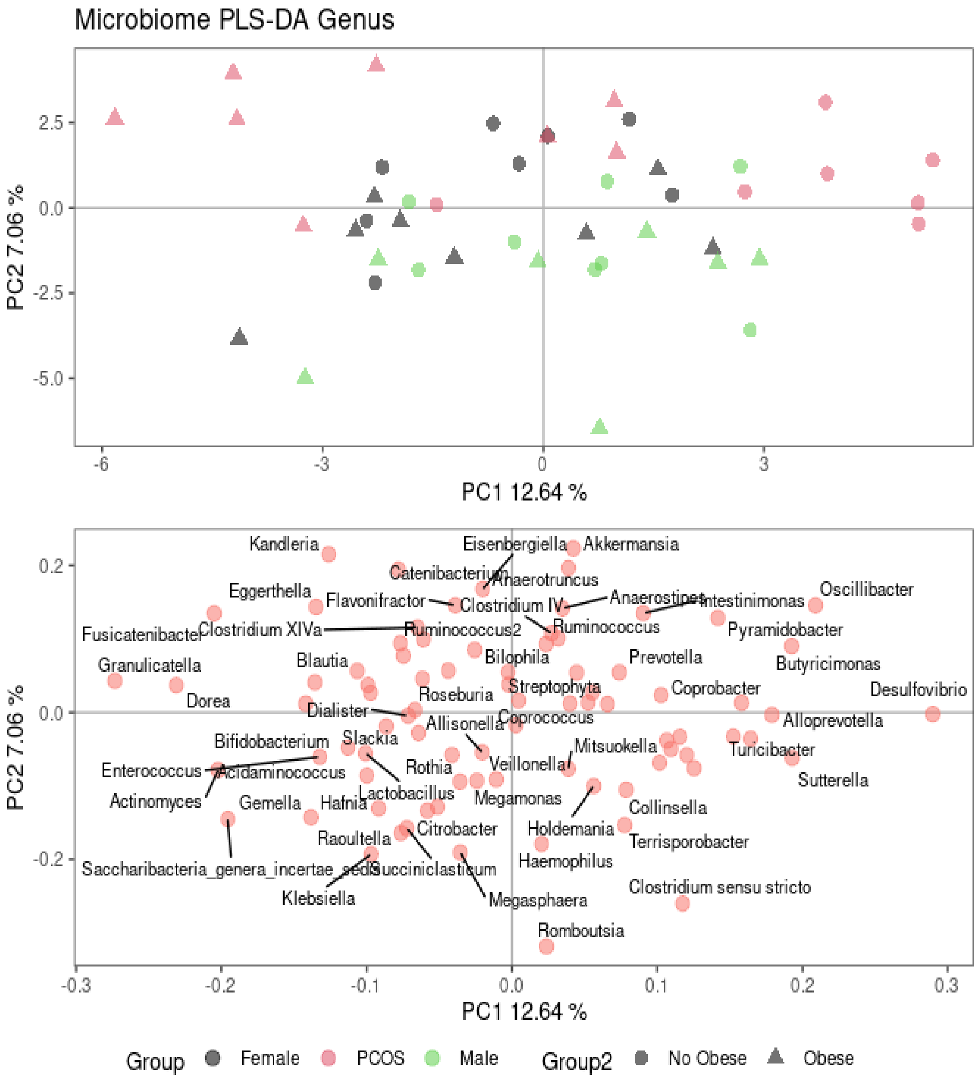
30. Figure: Descriprive values of the log-center-transformation of the phylum abundance. Pink and blue boxplots corresponds to non obese and obese subjects respectively



31. Figure: Metric and non metric MDS for the genus. Pink and blue boxplots corresponds to non obese and obese subjects respectively

3.2 Statistical Integrative Analysis

As mentioned in the methods section, three main joint statistical integrative analyses were computed: PLS2, O2PLS, and sparse Generalized Canonical Correlation Analysis. The last method is a sparse generalization of PLS2-DA with the possibility of integrating several blocks, which is called horizontal integration. Although the algorithm can make the variable selection, it is very time-consuming in terms of computational power, for this reason, this property of the algorithm is not taken into account.



32. Figure: PLS-DA analysis on the mean-centered, scaled, and log-center-ratio transformed of the genus. Pink and blue boxplots corresponds to non obese and obese subjects respectively

To understand better the methods for joint integration, firstly, the interaction between the microbiome and the metabolome was assessed. For all three methods, models with absolute, relative, and log-centered-ratio transformed data were computed. The best results were obtained with the transformed data and filtered by 0.01% of abundance. Moreover, we computed the same models for the phyla and genera in a separate way.

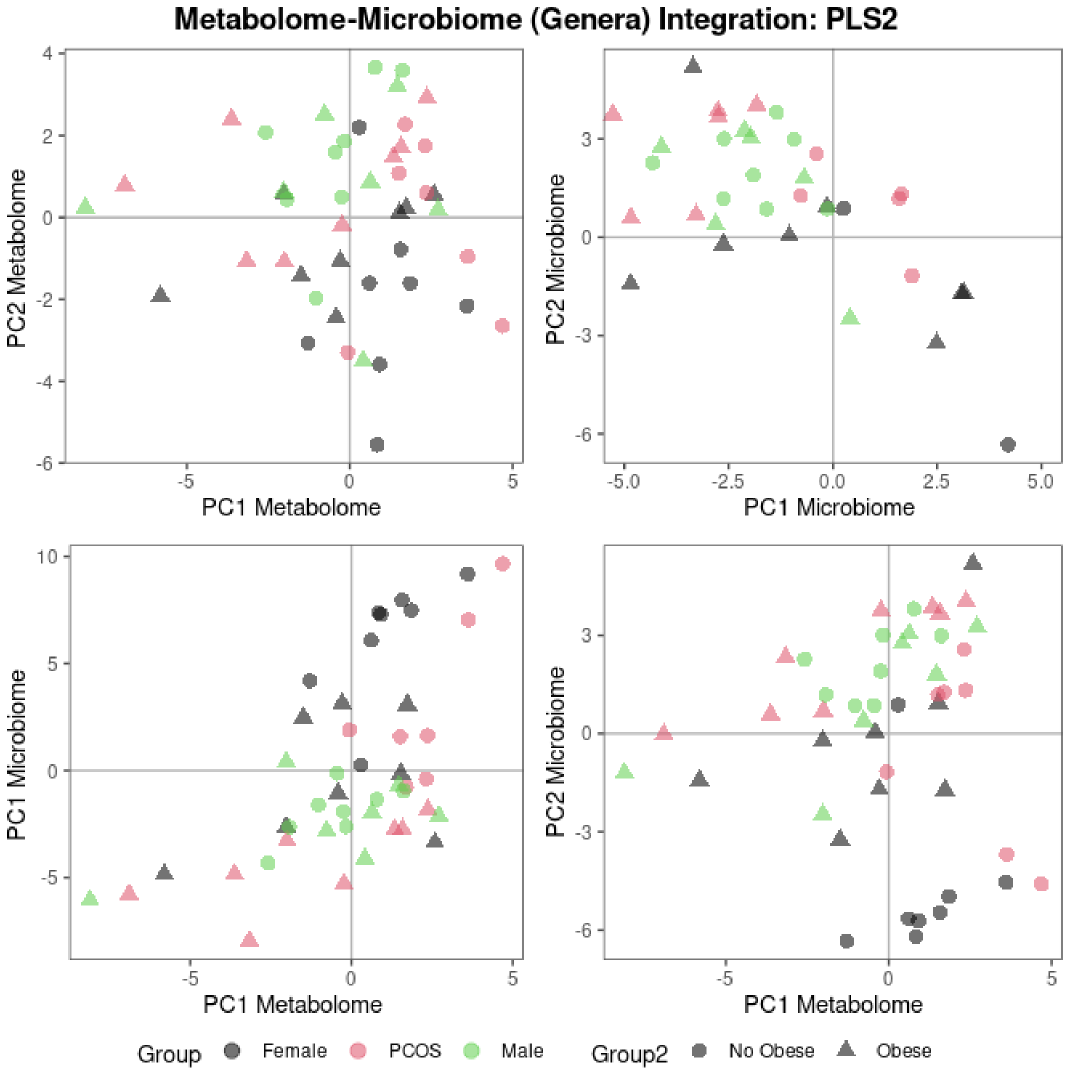
3.2.1 Metabolome-Microbiome Integration

Results from PLS2

The best results obtained with this method were obtained with the scaled metabolome data, and the transformation mentioned above with the genera data, figure 33.

On the first plot, the metabolome block, we can observe on one hand, a distinction between obese and non-obese subjects discriminated by the first component. On other han, the second component separes sex and PCOS from healthy vomen.

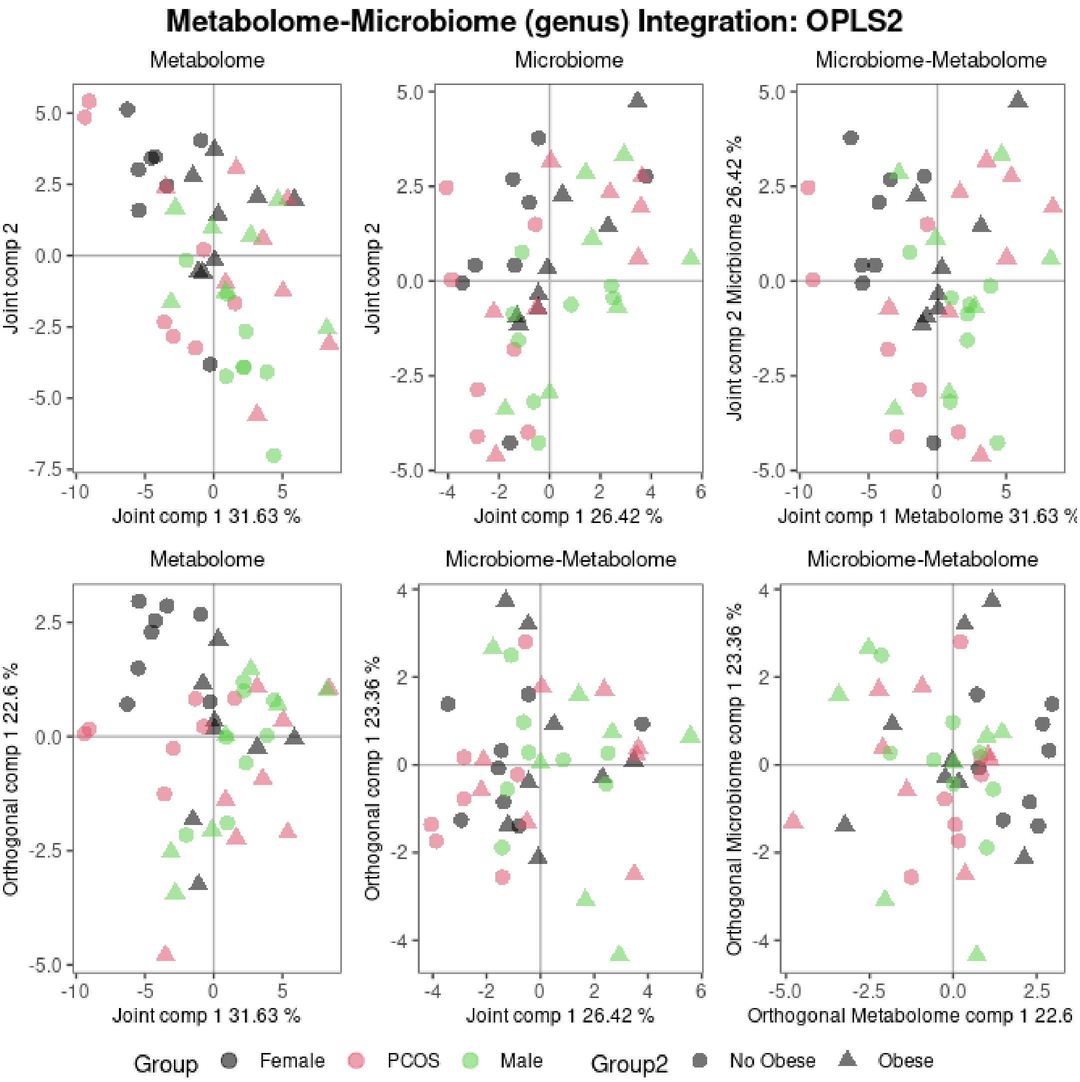
If we take a look at the last plot which belongs to the first principal component of the metabolome and the second component of the microbiome,. It can be seen, the similarity between males and PCOS women, that though slightly differentiated from each other, are close to each other in both components.



33. Figure: PLS2 integration of the meabolome-microbiome blocks. Black, green and pink dots corresponds to control female, men, and PCOS women.

Results from O2-PLS

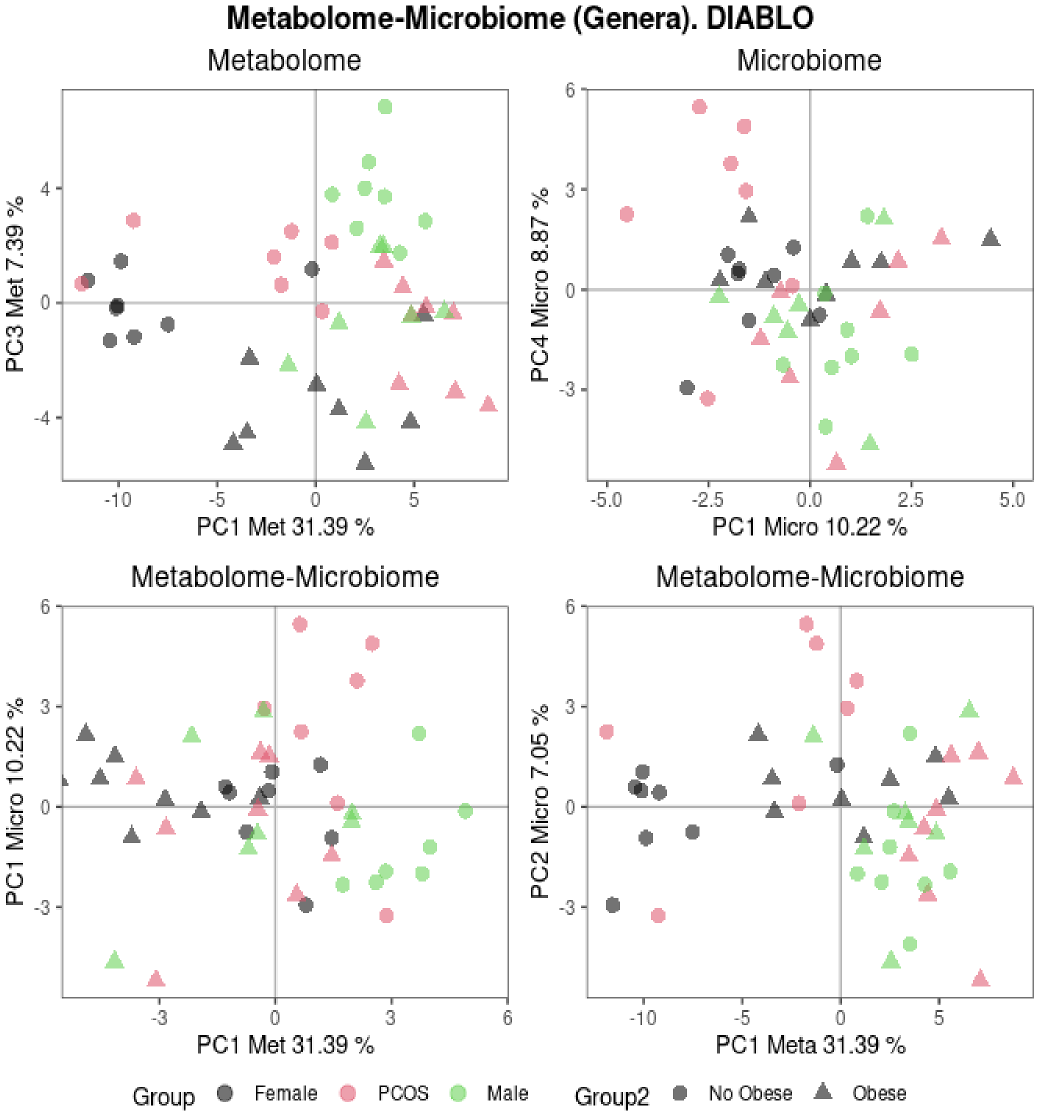
O2-PLS estimates the pure constituents profiles on the X block as well as models the Y orthogonal variation on the X block, the X block orthogonal variation in Y, and the joint $X-Y$ covariation. Furthermore, it is predictive in both directions. In this analysis the important biplot is the third one given that points out the joint modeling between Microbiome and Metabolome. There, it is possible to see on figure 34 that the metabolome separates between sex but with the PCOS women bewteen both sexes as seen througallthe this investigation and the microbiome separates the non-obese from the obese subjects. One can observe similar distinction between sexes and PCOS women on the joint components of the metabolome and the microbiome. Moreover the separation between lean control women is very clear on the first component and orthogonal component of the metabolome. But futhrer conclusions cannot be done.



34. Figure: O2-PLS integration of the metabolome-microbiome (phylum) blocks. Black, green and pink dots corresponds to control female, men, and PCOS women.

Results from DIABLO

There is a possible and logical explanation regarding why the above methods are not convincing: regardless of the technical details, firstly PLS2 assumes a relationship of independence-dependence between the two blocks. On the results obtained, the best differentiation was obtained when the metabolome was considered as the independent block. Suggesting that the metabolome induces the abundance of the types of microorganisms. Similar conclusions can be drawn with O2PLS. The main improvement of O2PLS is that it takes into account the joint and orthogonal variation between the blocks and it is predictive in both ways.



35. Figure: DIABLO: Metabolome-Microbiome (genus) mean-centered and scaled to unit variance. Black, green and pink dots corresponds to control females, males, and PCOS women. Triangles to obese subjects and circles to non obese

However, on the last method, the dependence between blocks still exists, leading to the same concussions. On the other hand, it is well documented that the metabolome and the microbiome are codependent, for example in (59). To pursue these conclusions more work has to be done. Secondly, from what it was investigated so far, these methods are only for the integration of two blocks of data. Since the aim of the project is the integration of all data sets a better approach is horizontal or *N*-integration.

The sparse Generalized Canonical Correlation Analysis, which is a generalization of PLS2-DA with multiple independent blocks, is a supervised method. Firstly and to investigate the performance of this technique, the blocks to be analyzed are the microbiome and the metabolome at basal levels. The dependent variable is the interaction between groups of interest, that's it sex, PCOS women, and obese subjects. As commented on methods, we used the CV technique to obtain the optimal number of components for each block. However, all variables in the blocks were used, thus the sparse property was not used. On the other hand, one has to take into account the correlation between the metabolome and the microbiome. This was assessed by computing a simple PLS2 with one component for each block. Then a Pearson correlation coefficient was calculated between the scores.

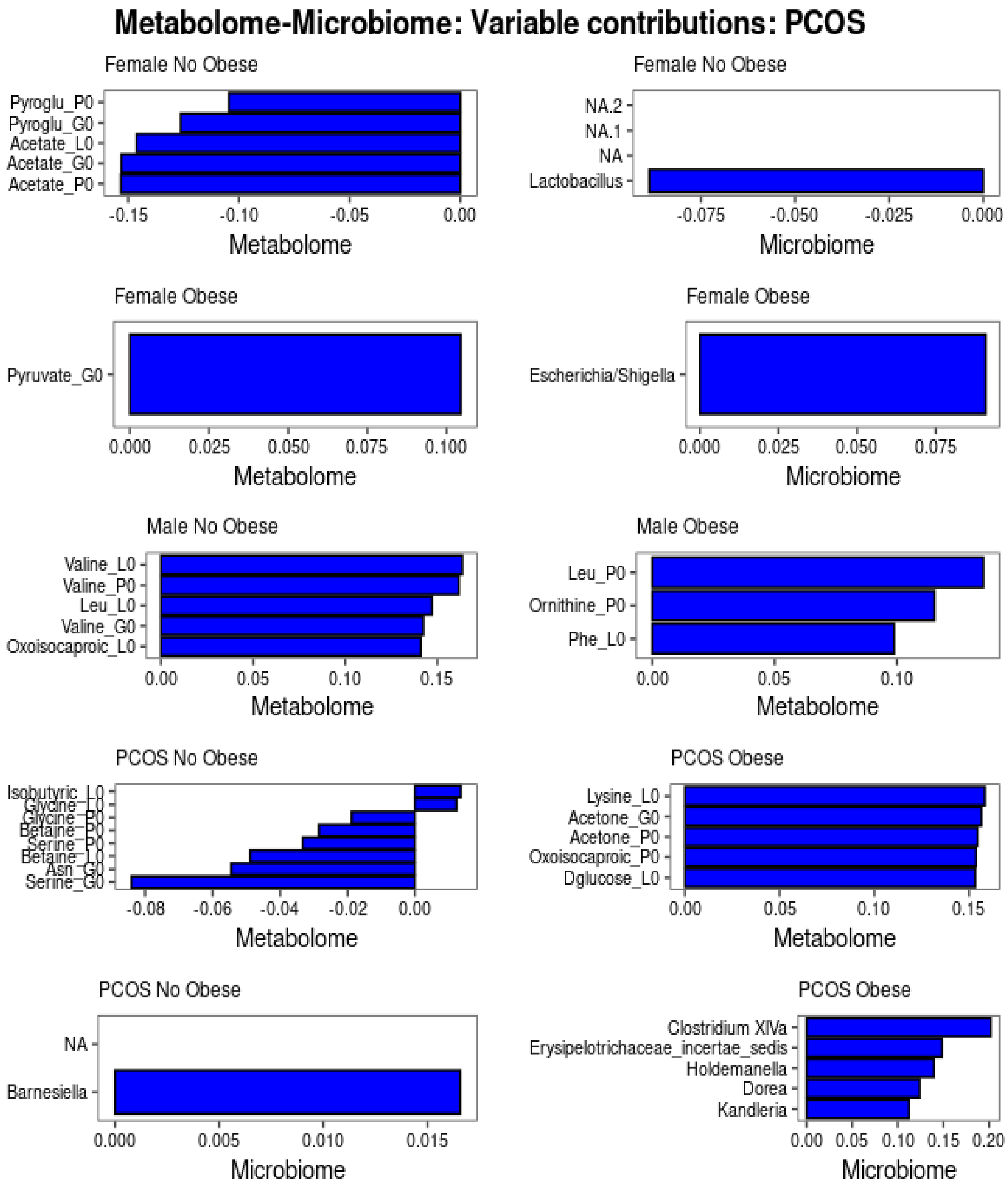
It has to be noted that the best results from the total integration were obtained with the genus data set. Although the results from the integration with the phyla were good, under the published results from the univariate analysis of each data set, it was decided that the total integration was computed only with the genus.

The optimal number of components is 5. Sparse PLS2-DA tries to maximize the correlation between the blocks. For each component the correlation is maximized as observed on the following table 5, doing successfully on the second to fifth component since the original design matrix accounted for a correlation of 0.6

Regarding the variation explained by each component is shown in table 6. The first components for both data sets account for the maximum variability within the blocks. However, the third component for the metabolome follows the second maximal variation, 7.38%, and the fourth component of the microbiome explains 8.87% of the variance.

Principal Components	Correlation Between Blocks
PC1	0.6
PC2	0.77
PC3	0.67
PC4	0.7
PC5	0.65

5. Table: Correlation between the microbiome and metabolome on each principal component



36. Figure: Variable contribution for women with PCOS. Only the variables from the first component are shown.

The metabolome block, figure35, on the first component, differentiates clearly between lean control women and the rest of the subjects, however, two lean PCOS women are on the same cluster as the control women. This dimension also differentiates between lean PCOS women and males. The third component of the metabolome which explains 7.39 % of the variation, separates between the obese and non-obese subjects.The latter reveals that there are differences regarding metabolome between groups. Again, it is shown how PCOS women have more similarities with men than the women control, especially if these last are lean.

Variance Explained	Metabolome	Microbiome
Comp 1	31.38	10.22
Comp 2	6.67	7.05
Comp 3	7.38	7.04
Comp 4	3.7	8.87
Comp 5	4.46	4.77

6. Table: Variance explained by each component

Regarding the microbiome dataset, lean PCOS women are differentiated by the second component, but no further conclusions can be made.

The contribution of each variable for each block is represented in bar plots, figure36. This contribution belongs to the loading weight of the feature. It can be positive or negative indicating if the relationship between one group and the feature, either metabolite or microorganism es positive or negatively correlated

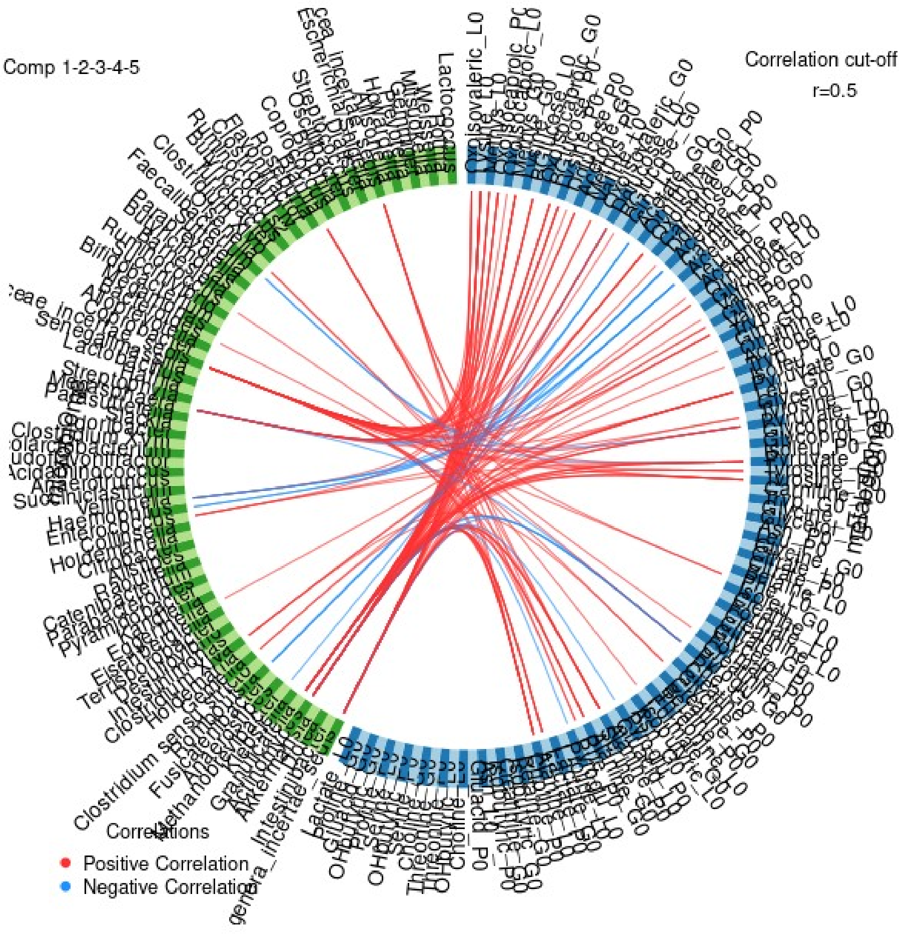
On control, healthy females Lactobacillus, Pyroglutamate, and Acetate have a negative contribution on these subjects in the sense that these variables make the distinction between the other patients. Regarding the control of obese females, the features Pyruvate and Escherichia contribute positively.

For the control healthy males, the metabolites, Valine, Leucine, and Oxoisocaproic contribute positively, whilst no microorganism contributes to the distinction between the other subjects.

Regarding the group under study, PCOS, lean women have the metabolites Isobutyric acid and Glycine at basal level lipid load day that contribute positively while Serine is the amino acid which contributes the most, followed by Asn, Betaine and Glycine, the last at basal level protein load day but negatively together with a positive association in the genera Barnesciella. In what concerns to obese females with the syndrome, the metabolites associated positively are the following: Lysine, Acetone, Oxosiocaproic acid, and D-Glucose, and the genera, Clostridium XIVa together with Erysipelotrichaceae Incertae Cedis, Holdemanella, Dorea and Kandleria.

Interaction between the metabolome and the microbiome.

From the figure 37 one can observe that the main relationship between both blocks is negative. For example, Saccharibacteria genera incertae sedis is negatively correlated with oxoisovaleric acid, lysine, oxoisocaproic and methylsisteine and alsw with alanine. On the other hand, Veillonella, is possitevly correlated only with acetate, as well as Haemoplhillus, Romboustia, and Methanobrevibacter. On the other hand, Blautia , Olsenella, Rominococcus, Lachnospiracea_incertae_sedis, Megasphaera, Enterococcus, Eggerthella, Clostridium XlVa are negatively correlated with several metabolites such as acetone, ornithine, glutamine, valine, pyruvate, tyrosine, carnitine, phenilalanine, formate, alanine and serine.



37. Figure: Circos plot of the correlation between the microbiome-metabolome blocks. Red possitive correlation, blue negative correlation

3.2.2 Total Integration at basal level

The biomarkers discovered by this method are found in two ways. Firstly, the contribution of each feature, i.e the loadings, the most important ones are present. Secondly heatmap with the correlation between subjects, and the correlation between features is computed. The last method provides more insightful results according to the ***mixOmics*** package of Bioconductor (17)

The total integration of all blocks at the basal level was assessed in the same way, in the sense that the microbiome data was log-center-ratio transformed.

Although the best results were obtained with the log-center-ratio transformation of the genus, it has been mentioned that the correlation between the microbiome and the metabolome was higher when accounting for the absolute abundance of the data (80%). However, as we will see briefly, the distinction between the groups of interest and the variance explained on the microbiome was higher (data not shown).

One of the highest initial correlations regarding the design matrix surprisingly was between the microbiome and the Clinical Data. In general, the correlations between the Clinical Data and all data sets are high, as one can see on the follwing table 7

Design	Clinical Data	IP	Metabolome	Microbiome
Clinical Data	0	0.78	0.72	0.72
IP	0.78	0	0.64	0.67
Metabolome	0.72	0.64	0	0.59
Microbiome	0.72	0.67	0.59	0

7. Table: Design Matrix for the total integration at basal level

The number of optimal components is 2, and The variance explained by each component on each data set is the one shown below. The metabolome explains almost all the data on the first component, accounting for 30% of the variance, and the second component only 6%

Variance Explained	PC1 (%)	PC2 (%)
Clinical Data	28	17
IP	28	21
Metabolome	30	6
Microbiome	12	8

8. Table: Variance explained for each component

Regarding the performance of the algorithm, the clinical-metabolome correlation is maximized at the second component as well as the interaction metabolome-microbiome The other components show a lower correlation than the design matrix, table 9.

Maximized Correlation	PC1	PC2
Clinical-IP	0.74	0.51
Clinical-Metabolome	0.7	0.88
Clinical-Microbiome	0.7	0.61
Metabolome-IP	0.55	0.55
Metabolome-Microbiome	0.51	0.59
Microbiome-IP	0.49	0.66

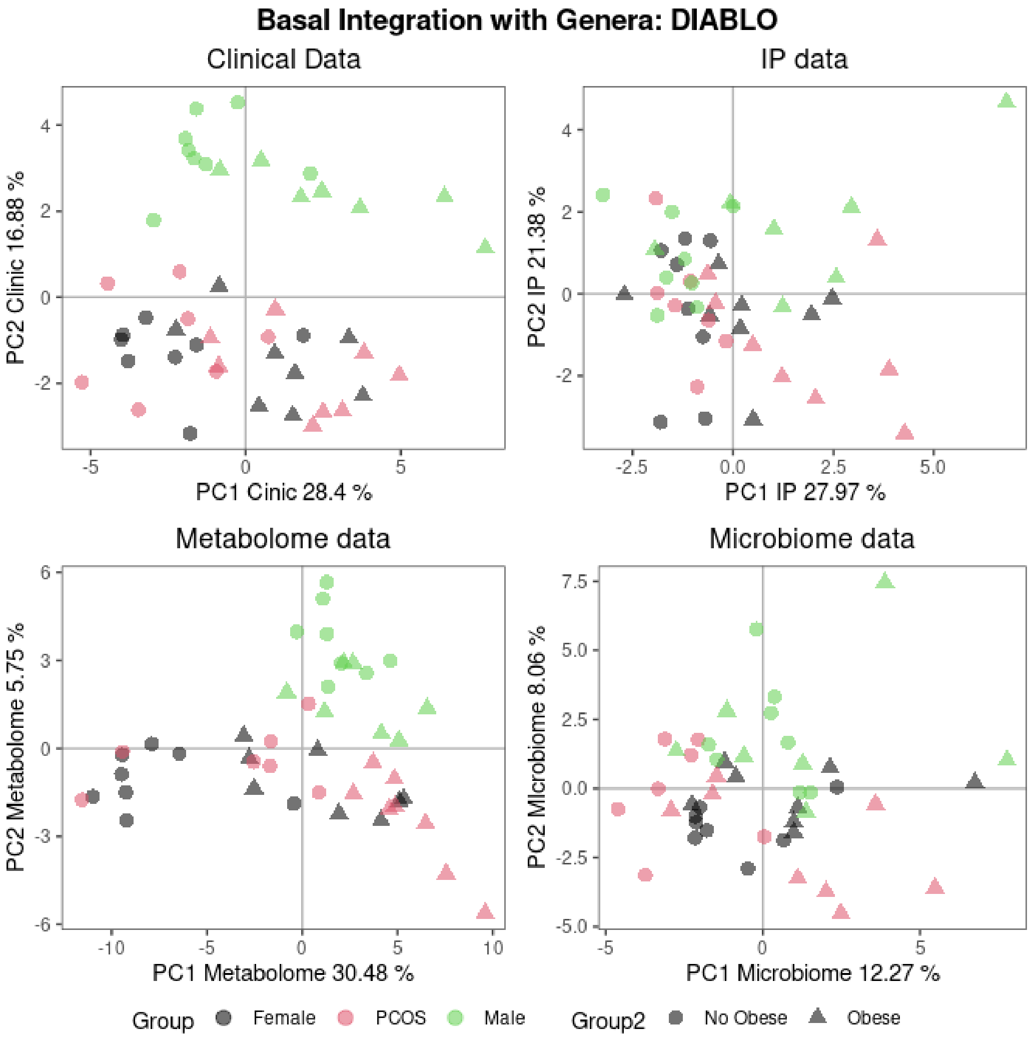
9. Table: Maximized correlation between blocks of data on each component

If the biplot PC2 Clinical Data and PC1 Intestinal Permeability data is observed, figure 40 the separation between females and males is very clear. This is due to the second component of the Clinical Data, which accounts for 17% of the variability in that block. The first IP component accounts for 28% of the variation, and it distinguishes between obese and non-obese subjects. But no distinction can be made between PCOS women and control females

The metabolome block is, so far, the best data set that distinguishes between all groups, figure 38. The first component accounts for 30 % of the variation. It is clear the distinction between controlled healthy women and the rest of the subjects. PCOS

women are in the middle of the plot together with obese control females. Moreover, PCOS is differentiated from the rest, except for some overlap with control obese females. Whilst the first between obese and non-obese patients,the second component which accounts
for almost 6% of the variation, distinguishes between sex.

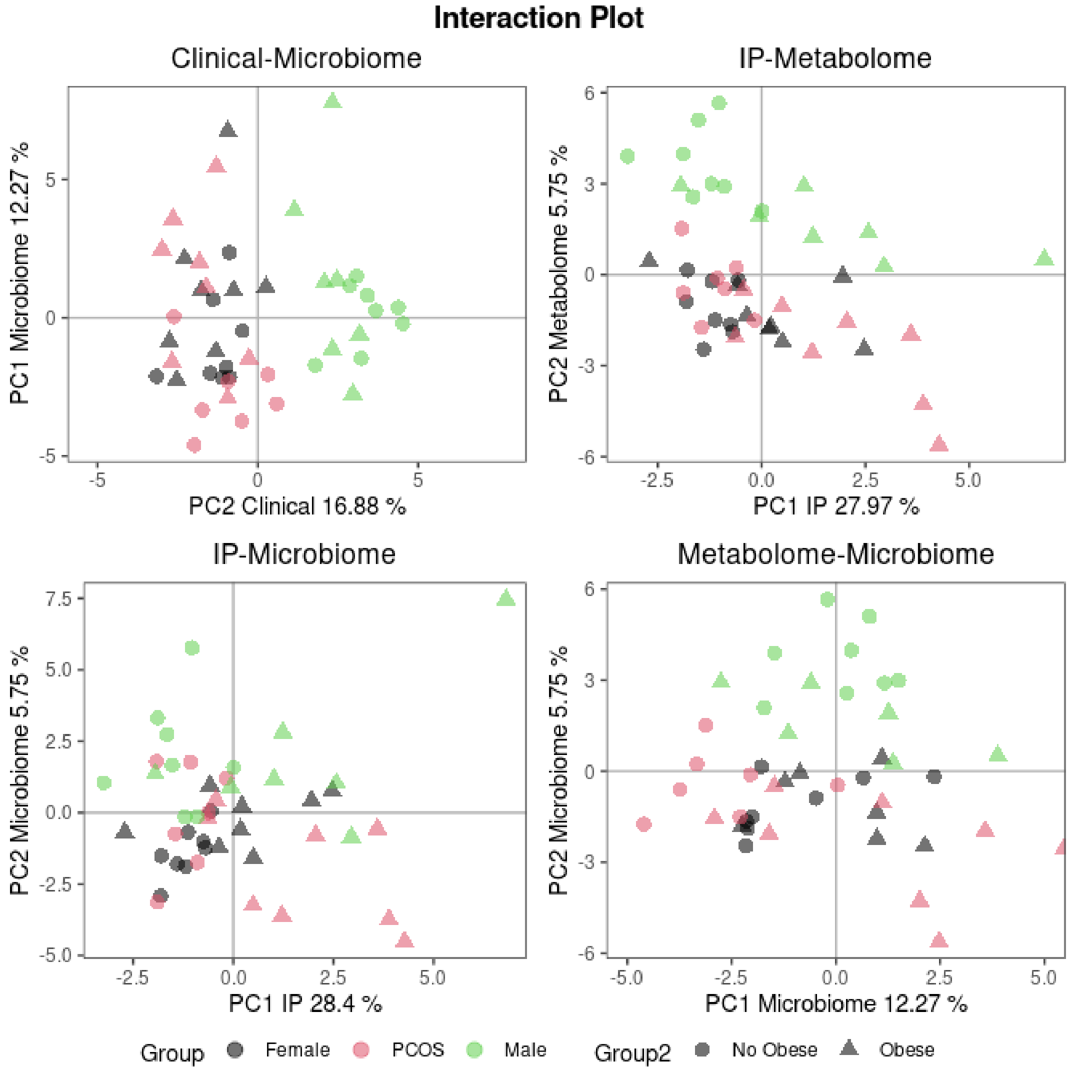
Finally, the microbiome data set, the distinction of obese and non-obese subjects is mainly due to the first component that accounts for 12% of the variation. The separation between lean and obese PCOS females is very clear, meanwhile, the control subjects, both males and females, though overlapped, are separated by the second component that explains 6% of the variation. Again it can be see, how PCOS women take features from both sex, being intermediate between control healthy women and men.



38. Figure: DLABO; total basal integration. Clinical, IP, Metabolome, and Microbiome (genus) blocks are integrated. Black, green, and pink dots corresponds to control females, males and PCOS women. Triangles belongs to obese subjects and circles to non obese subjects.

When it is taken into consideration the maximized correlations between the blocks, figures 39-40, microbiome-metabolome, and clinical-metabolome, both on the second component, and we plot the corresponding biplots, we observe that more information and more robust conclusions can be drawn.

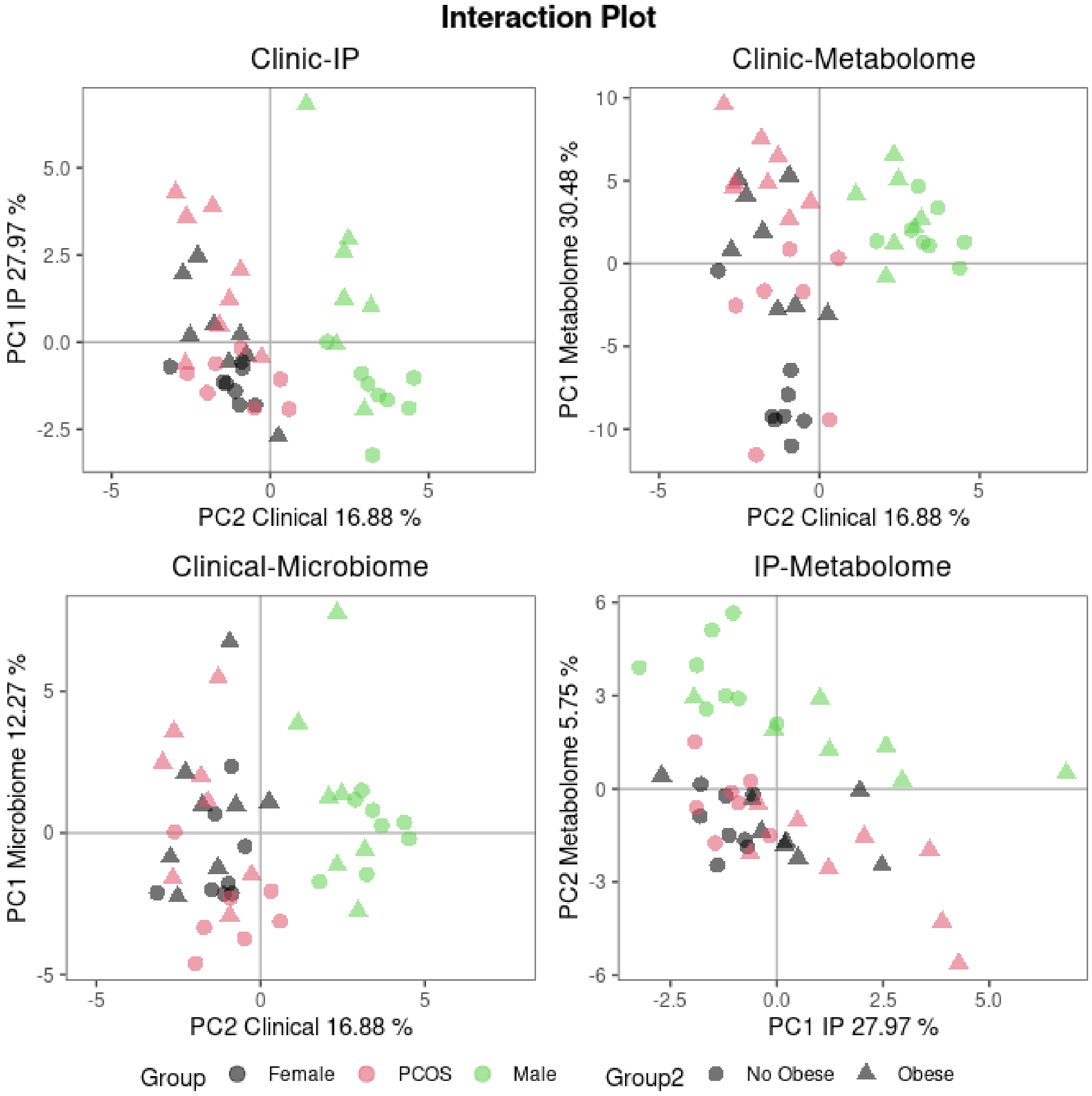
This is specifically true on the clinic-metabolome data. The second component of the clinic data differentiates between sex and the second component distinguishes between obese subjects. Though the distinction between obese subjects is not very clear in males, between females is very obvious. Lean PCOS women are clustered together, however, there are two patients that are in the cluster of the healthy control women. This should be investigated further by physicians. Furthermore, there is a lean woman which is in the cluster of obese females. PCOS lean women and obese control women appear to be on the same cluster. The distinction between PCOS obese females and non-obese also is very clear, with some overlap with control obese females.



39. Figure: Basal total integration with Diablo. Interaction plots of the clinical-microbiome, IP-metabolome, IP-microbiome, Metabolome-Microbiome. Black, green, and Pink dots corresponds to control females, males and PCOS women. Triangles belongs to obese subjects and circles to non obese subjects.

Regarding the Microbiome-Metabolome interaction, the distinction between sexes is very clear, mainly to the second component of the microbiome data set.

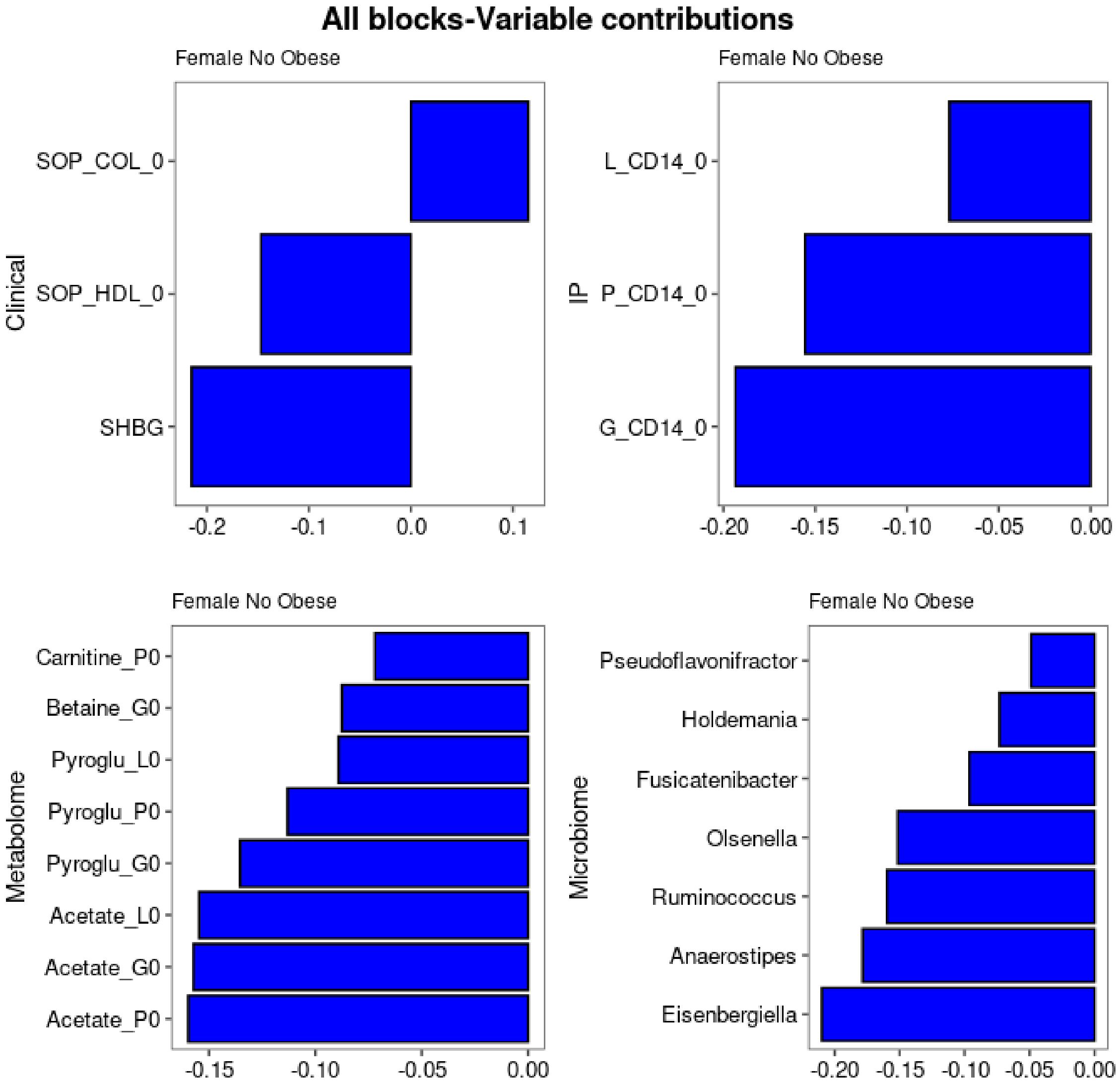
The interaction between all blocks and IP the distinction is clear between obese and non-obeses subjects.



40. Figure: Basal total integration with Diablo. Interaction plots of the Clinical-IP, Clinic-Metabolome, Clinical-Microbiome, and IP-Metabolome are computed. Black, green, and Pink dots corresponds to control females, males and PCOS women. Triangles belongs to obese subjects and circles to non obese subjects.

Contribution of features to the principal components

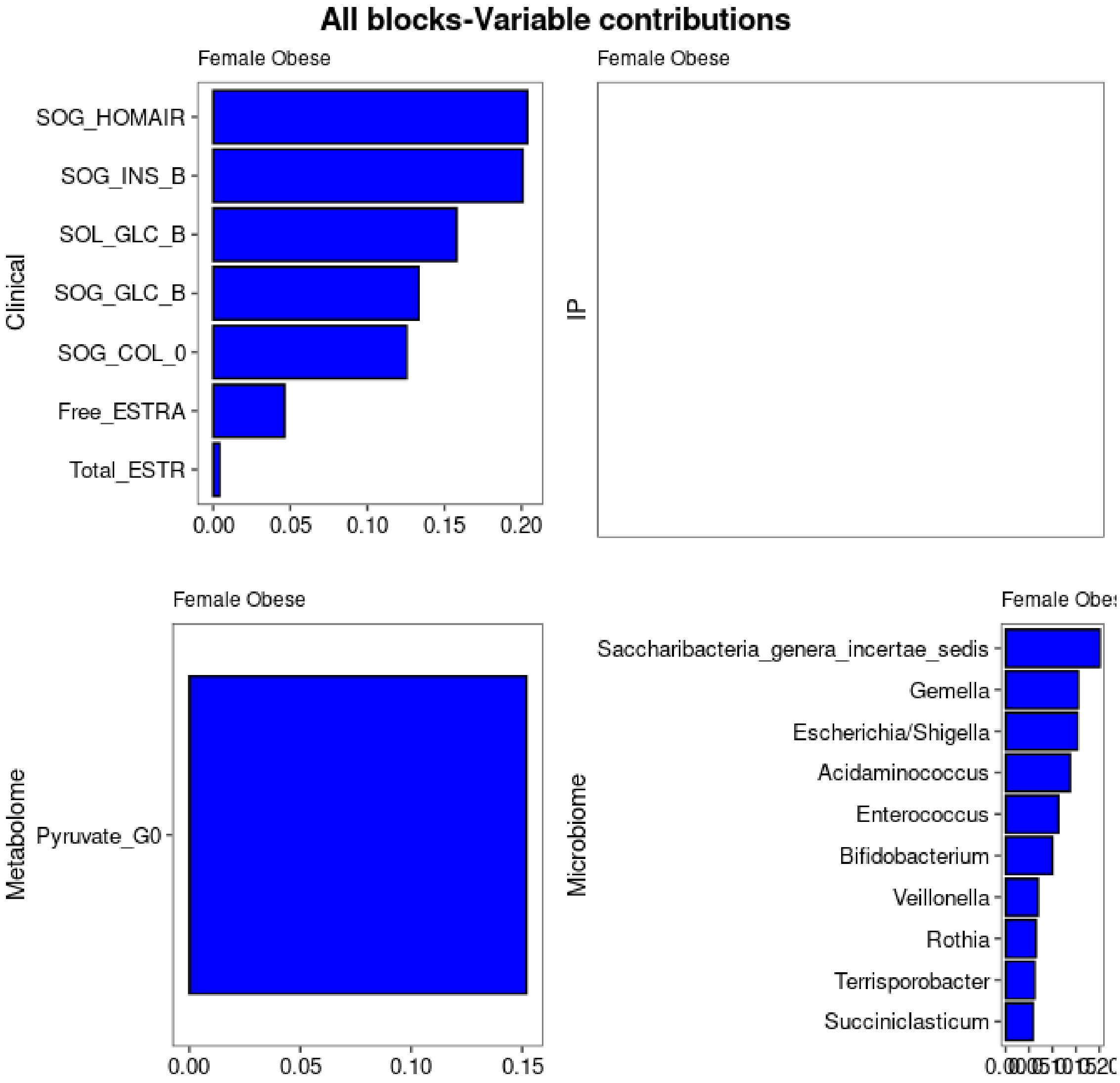
In general, the distinction between sexes is due to sex steroids. On healthy control females, the clinical, the metabolome and the microbiome contribute to the distinction between them and other subjects. Specifically, on the protein oral challenge, cholesterol, HDL-cholesterol, and the SHBG, contribute negatively to the plot to distinguish between them and the other patients. No permeability markers distinguish lean females. The acetate, when the protein is administered, makes the main distinction together with Anaerstipes and Eisenbergiella correlated negatively, and in a positive way with Slackia. The contribution plots can be observed on figures 41-46



41. Figure: Variable contributions for the control healthy females across all blocks

Regarding the obese control females subjects, the glucose at the basal level when this macronutrient is administrated together with the lipid oral challenge day administration, make the distinction with the cholesterol. On the other hand, HOMA-IR when the glucose is provided to the patient, makes the distinction in a negative way on the plot, as well as the insulin, free estradiol, and total estradiol, being the last one the most important clinical variable. No IP markers have been found, neither of the metabolome to characterize these individuals. However, Veillonella, Acidaminocococcus, Rothia contribute in a positive way and Bifidobacterium, Escherichia/Shigella in a negative way.

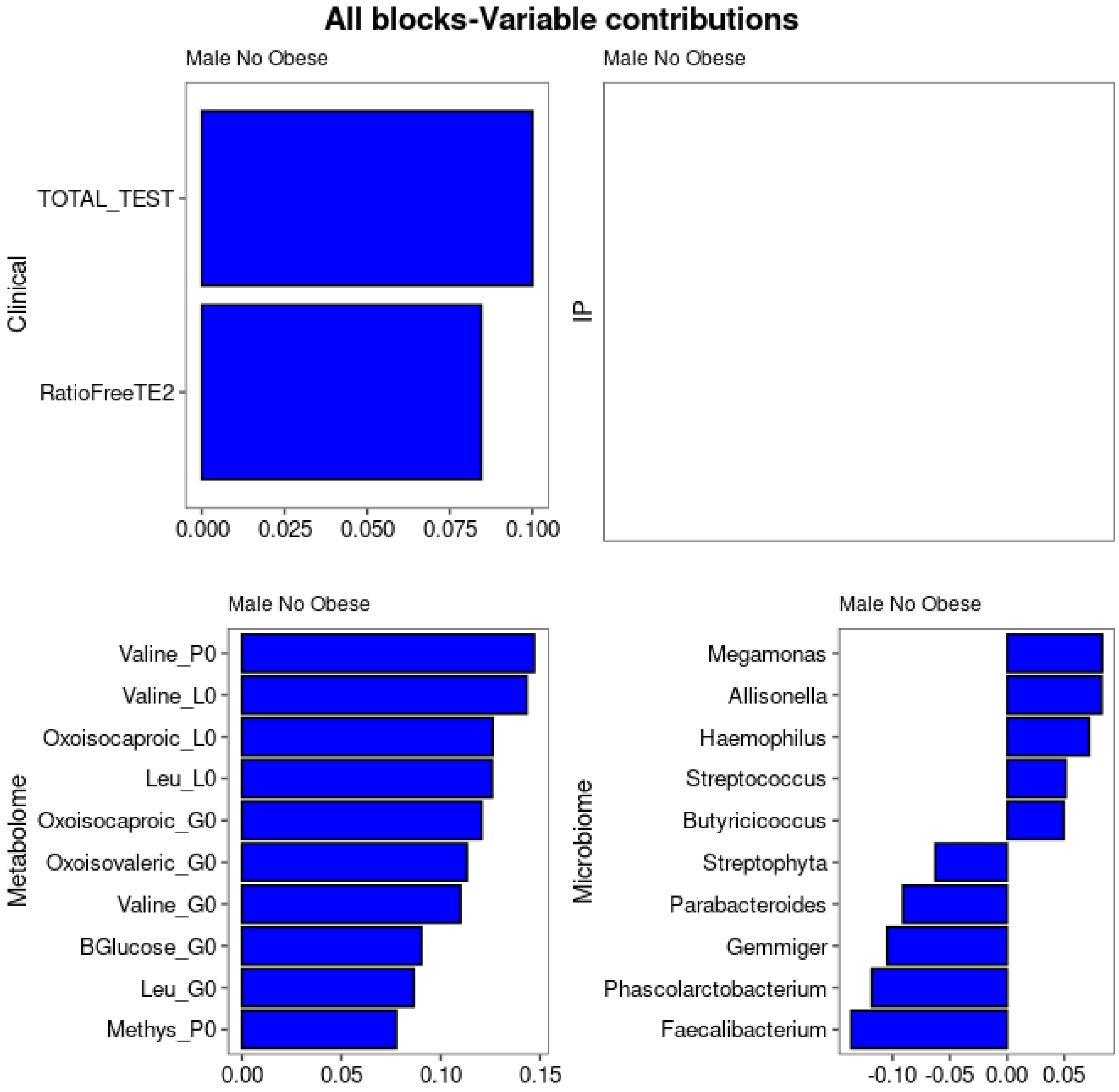
In what concerns healthy males, the most relevant clinical biomarkers for males are the ratio-free testosterone-estradiol, together with total testosterone. No IP markers characterize these subjects. Regarding the metabolome, only the Leucine and the Phenylalanine, contribute positively, being more important than the sex steroids, apart from acid oxisocaproic. The genera that characterize lean males is Megamonas.



42. Figure: Variable contributions for control obese women across all blocks

On the other hand obese males, free testosterone is the most important clinical biomarker, followed by glucose, WHR, triglycerides, with and in a negative contribution way the LDL, cholesterol, WC, and HOMA-IR. The IP markers are the glucagon-like protein 2 and Zonulin. The metabolites that differentiate these subjects are phenylalanine, ornithine, and leucine. The genera that make this distinction are Clostridium Sensu Stricto, Klebsiella, Corpococcus, Collinesella, and Hafnia.

Finally, the group of interest, the females with the syndrome, in particular, the clinical variables that contribute the most are: HDL cholesterol but in a negative way. No IP biomarkers characterize these subjects. The only metabolite that contributes to these subjects is glycine. Anaerotruncus and Lactococcus are the genera positively correlated with lean PCOS women.

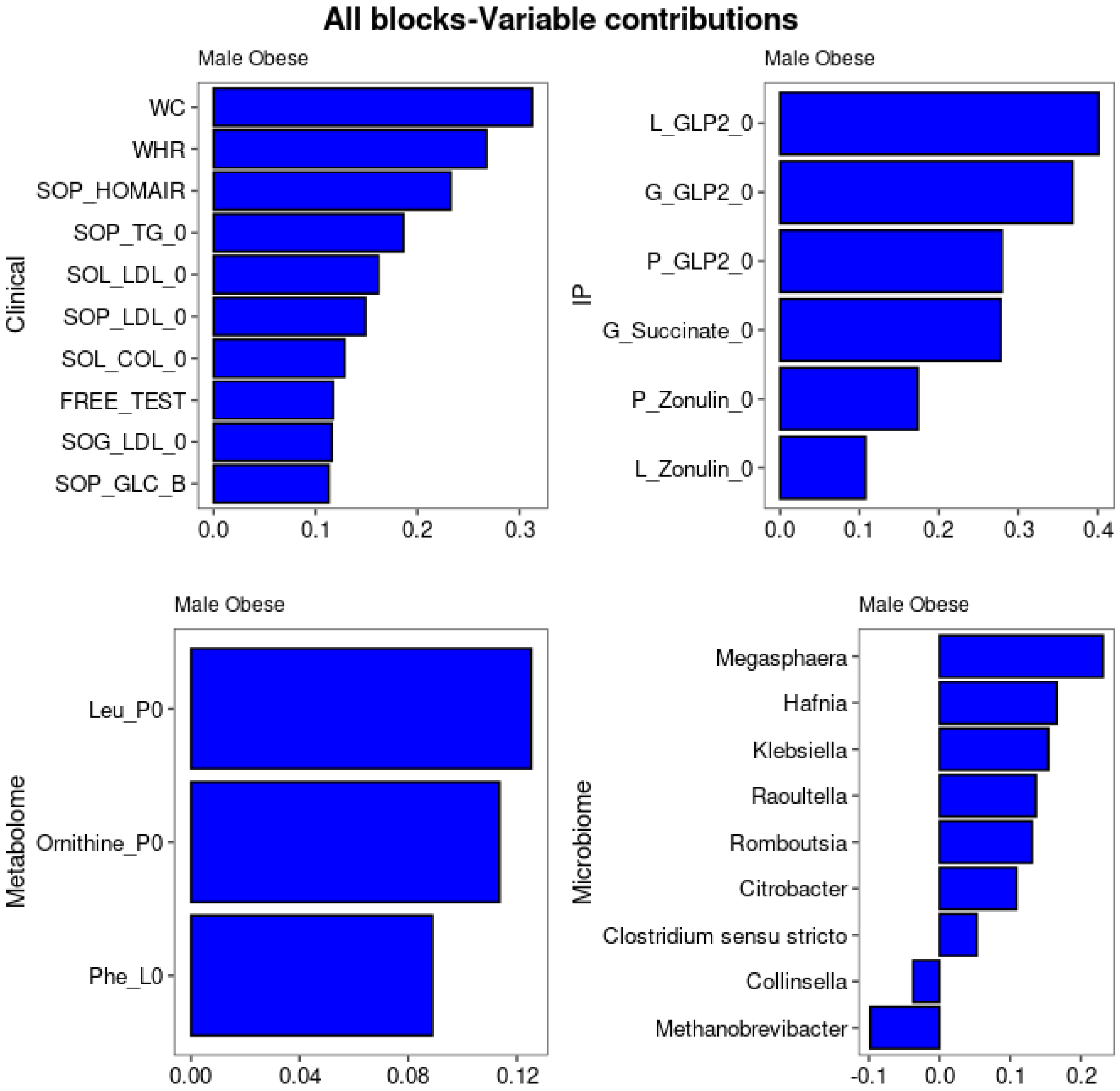


43. Figure: Variable contributions for healthy males

For what concerns obese PCOS females, the clinical biomarkers that characterize these patients are BMI, Insulin, hs-CRP, Trigicerics, Insulin, HOMA-IR, and surprisingly age. Regarding IP markers, the LBL is the only one as well as succinate but, this is a metabolite, other metabolites implicated in obese PCOS females are glutamine, tyrosine, creatinine, alanine, glycoprotein, creatinine, citrate, and pyruvate. The bacterias that characterize these females are Eggerthella, Catenibacterium, Kandleria, Actinomyces, and Granulicatella.

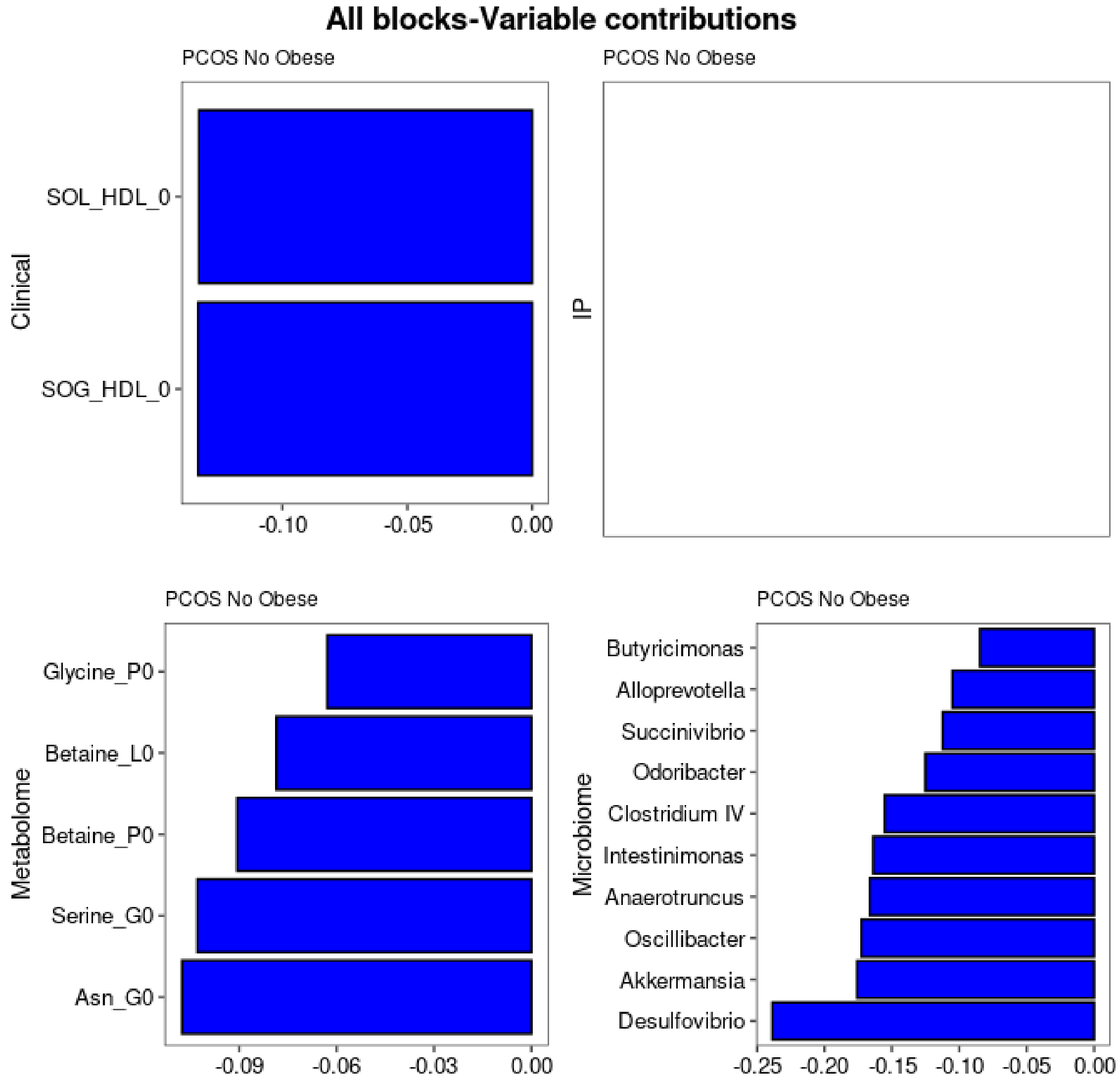
Correlation between subjects and features

A heatmap between all blocks and subjects, taking into account the most important variables was done by performing a euclidean distance between subjects, and a correlation between variables figure 47. The barplots shown in,, where computed due to visualization purposes by takin the mean of the variables, thus the results are not perse correlation, but a mixture of distances, correlations and the mean of the variables, figures 48-51



44. Figure: Variable contributions for obese males across all blocks.

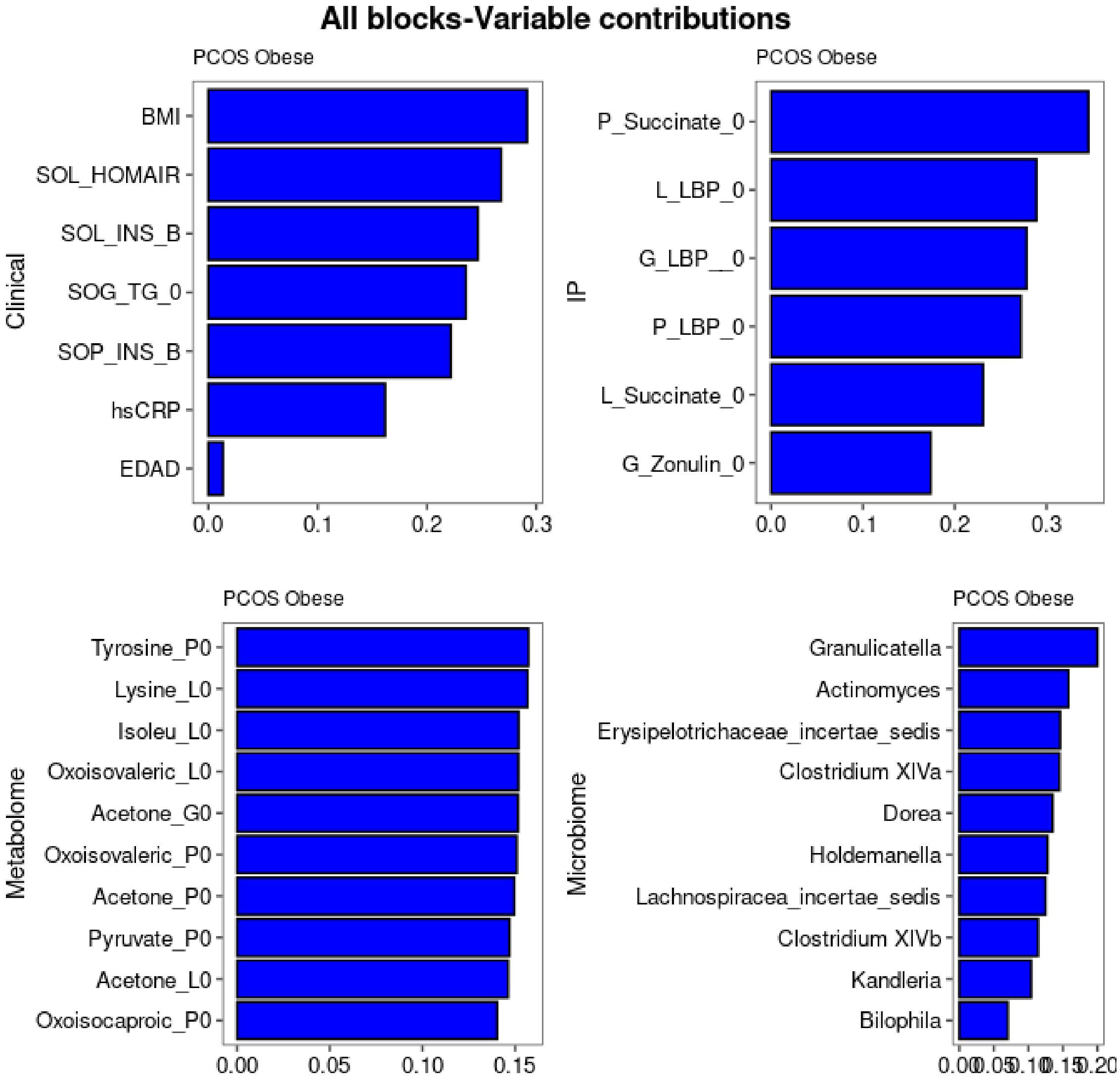
Total, free testosterone and the ratio between free testosterone and estradiol, in males, correlate positively, and negatively the tesotsin PCOS females, and obese females, however, this association is not present in control non-obese healthy women. Regarding body shape measurements, as expected, on lean subjects, BMI, WHR, and WC correlate negatively with them contrary to what happens on obese subjects. HOMA-IR correlates positively with obese PCOS women and negatively with obese males. Acetone on the glucose ingestion day, has a positive correlation with obese females and a negative association with lean women. On lean PCOS women, the genera Butycirimonas, Oscillibacter, Turicibacter, Intestinimonas, Ingtestinibacer, Desulfovibrio, Coprobacter, Alloprevotella correlates positively, however, Sachibacteria genera incertae sedis, Fusicatenibacter, Dorea, Actinomyces, Klebsiella, Granulicatella have a negative correlation. Surprisingly, fewer bacterias are found on obese women with the conditiion studied, specifically, Romboustia, and Clostridium senso stricto correlates negatively, but, Kandleria, Holdemanella, Granulicatella, Clostridium XIVa, and Catenibacterium correlates positevley.



45. Figure: Variable contributions for non obese females with PCOS

In general, terms, what represent the best obese PCOS women are the metabolites, and the most important ones, those that are represented on the barplot all of them correlate positively. On the other hand, in lean PCOS women, serine and glycine on the glucose and protein oral challenge days respectively correlates positively as well as betaine on the protein day, and acetate during all days. However, creatine when the glucose is administrated, acetone across all days, formate on the lipid day, lysine, isoleucine, oxoisovalerc, and valine on all days, correlates negatively. In what respects to IP markers, on obese women with the syndrome, LBP, zonulin and succinate, correlates positively, whilst on lean women with PCOS, only GLP2 is present on the most important features of this data set, and correlates negatively on the glucose and lipid oral ingestion days.

Interestingly, the variable that is more important on lean PCOS women is the abundance of several genera, whilst, on obese PCOS women are the metabolites. Apart from the testosterone sexual steroids negatively correlated with obese PCOS women and the fact that two IP markers are present on obese PCOS women which correlates positively, but there are only one, the GLP2, lean PCOS women. That correlates negatively. Moreover, age correlates, surprisingly in a positive way on obese PCOS women.

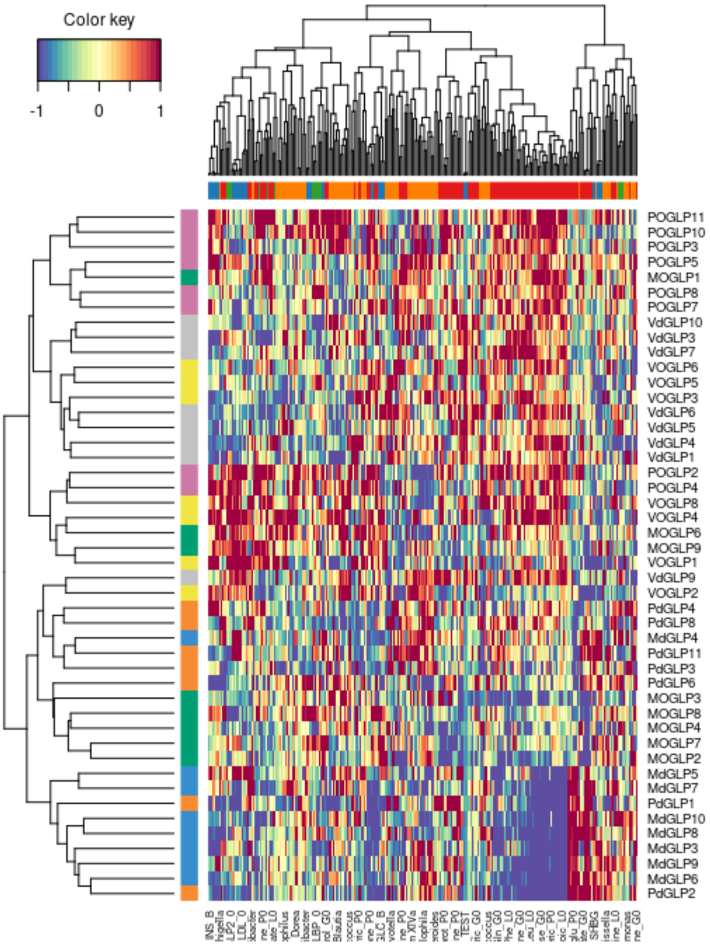


46. Figure: Variable contributions for obese women with PCOS.

In lean control women, the most important block is the metabolites, and almost all of them have a positive correlation, except the acetate in contrast with non obese PCOS women. Moreover, no sexual steroids are listed as the most important ones in this group. Regarding obese women, almost all metabolites. have a negative correlation.

Regarding lean PCOS is the only group with the most presence of bacteria, only obese males, have 1 genera: Rombustia, whilst PCOS lean women have 15 genera, above mentioned and 7 on obese PCOS women.

The group under study, PCOS, lean women have the metabolites Isobutyric acid and Glycine at basal level lipid load day that contribute positively while Serine is the amino acid which contributes the most, followed by aspargine (Asn),



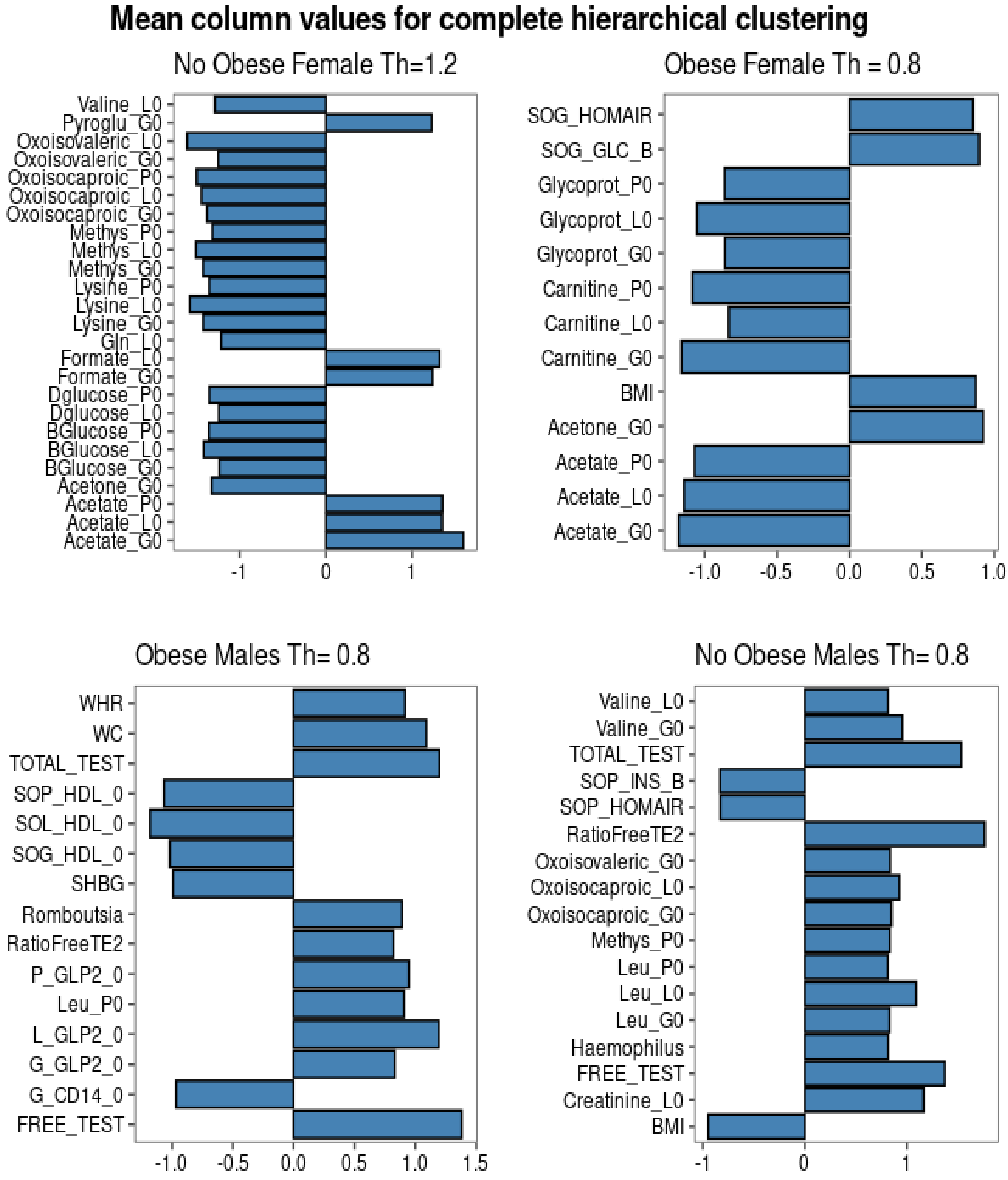
47. Figure: Heatmap DIABLO method. Euclidean between subjects is computed whilst correlation between variables is calculated.

Betaine and Glycine, the last at basal level protein load day but negatively together with a positive association in the genera Barnesciella. In what concerns to obese females with the syndrome, the metabolites associated positively are the following: Lysine, Acetone, Oxosiocaproic acid, and D-Glucose, and the genera, Clostridium XIVa together with Erysipelotrichaceae Incertae Cedis, Holdemanella, Dorea and Kandleria.

The main correlation between all blocks is positive, as seen on figure, where this correlation is colored in red, whilst the negative correlation are shown in blue. This figure show the correlation between variables of each blocks.

The main interaction is between the Clinical Data and the metabolome, however, the microbiome has a strong interaction with the metabolome and the Clinical Data.

Specifically, the SHGB, WC, BMI, insylin and glucose, interacts, with the osoisovaleric acid, lysine, methylistine, osoisocaproic, beta glucose, D glucose, and acetone.The sex steroids interacts with CD14 interacts negatively with CD14 negatively. HDL negatively with acetate. Also it seems that GLP2 and sex steroids are related to the microbiome. It is obvious that from the plot no further conclusions can be drawn, however, if we set an absolute threshold of 0.4, the main correlation can be obtained, as one can observe on the table 10.



48. Figure: Mean correlations between control subjects and features. Th stands for threshold.

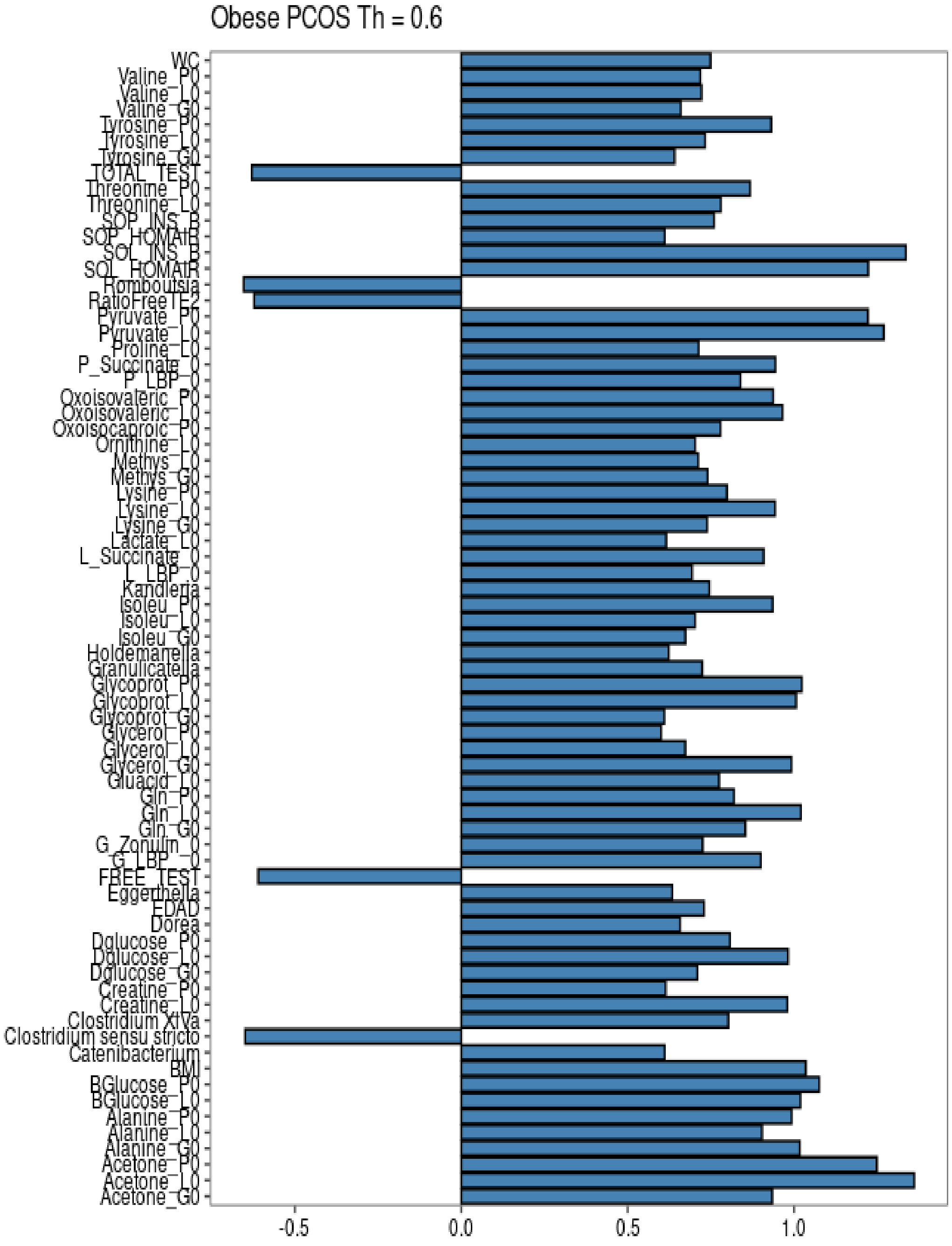
The sex steroids, specificaly, the total testosterone and the free testosterone, apart from the ratio of the free estradiol with the free testosterone, one can conclude, in general terms, withoug taking into accounts the gorups of interest, that, the SHBG correlates negatively with this features, regarding with the IP markers, CD14, LBO and GLP2 correlates as well in a negative way. On the metabolome block, the correlation with the sex steroids, the ones that are above or below of the threshold 0.4, are Leucine on the lipid ingestion day and creatinine on the same day, is positive. Whilst, the lactate on the on the glucose day correlates negatively. The last block, the microbiome, there are several genera thah correlates negatively, Blautia, Clostridium IV, Eggerthella and Eisenbergiella. The bacteria that correlate positively are the Clostrium sensu stricto, Hemophilus, Klebsiella and Romboustia. Moreover, the interlink in the Clinical Data, apart form the SHGB, there is a nevative correlation between the sex steroids and the high density lipoprotein.

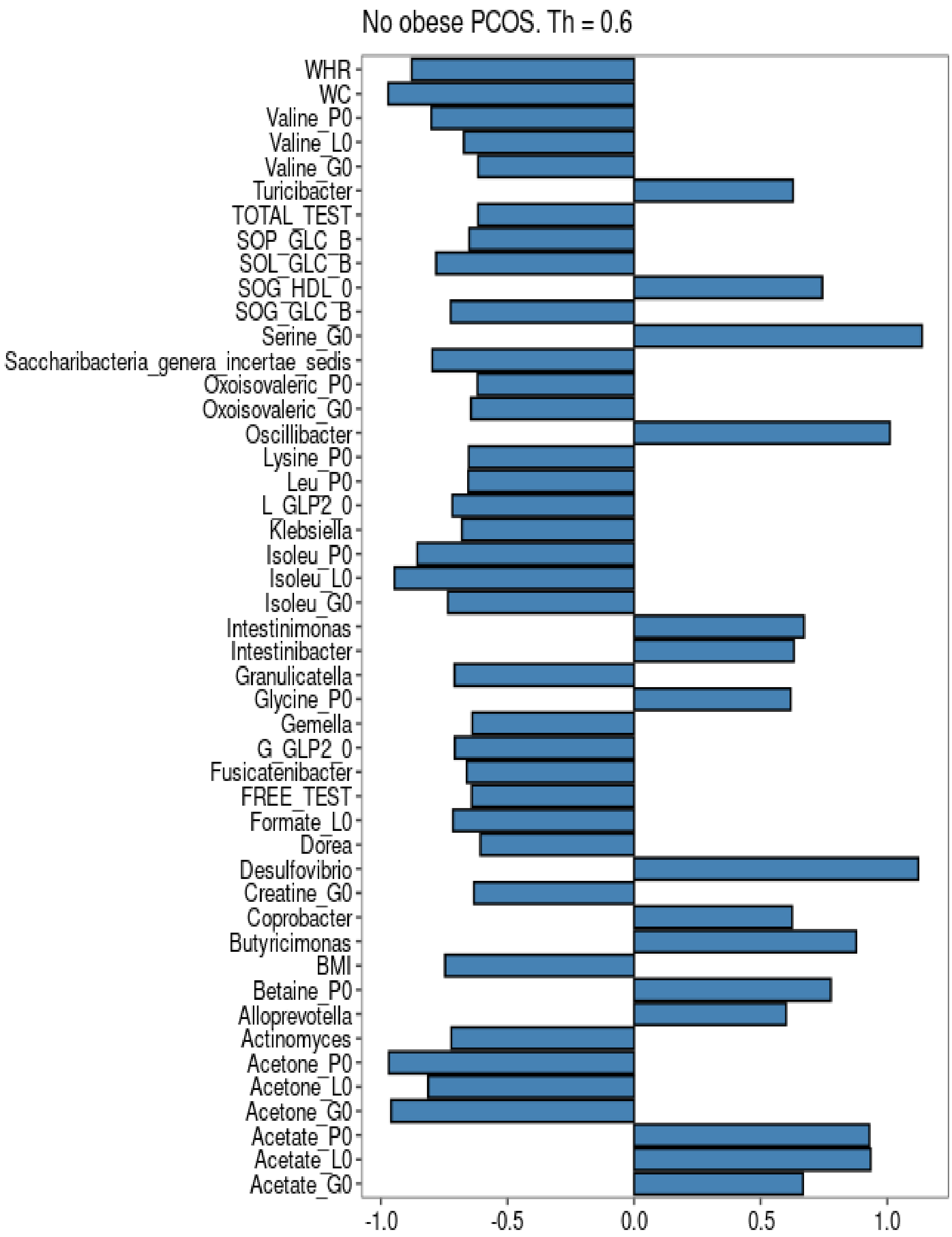
It has to be mentioned that we could obtain the main correlation between all variables at this threshold, but since the main goal of the project, is to know if all is related woth the sex steroids, then only the correlation between these hormones are taken into account.

10. Table: Results from CircosPlot, with an absolute threshold of 0.4

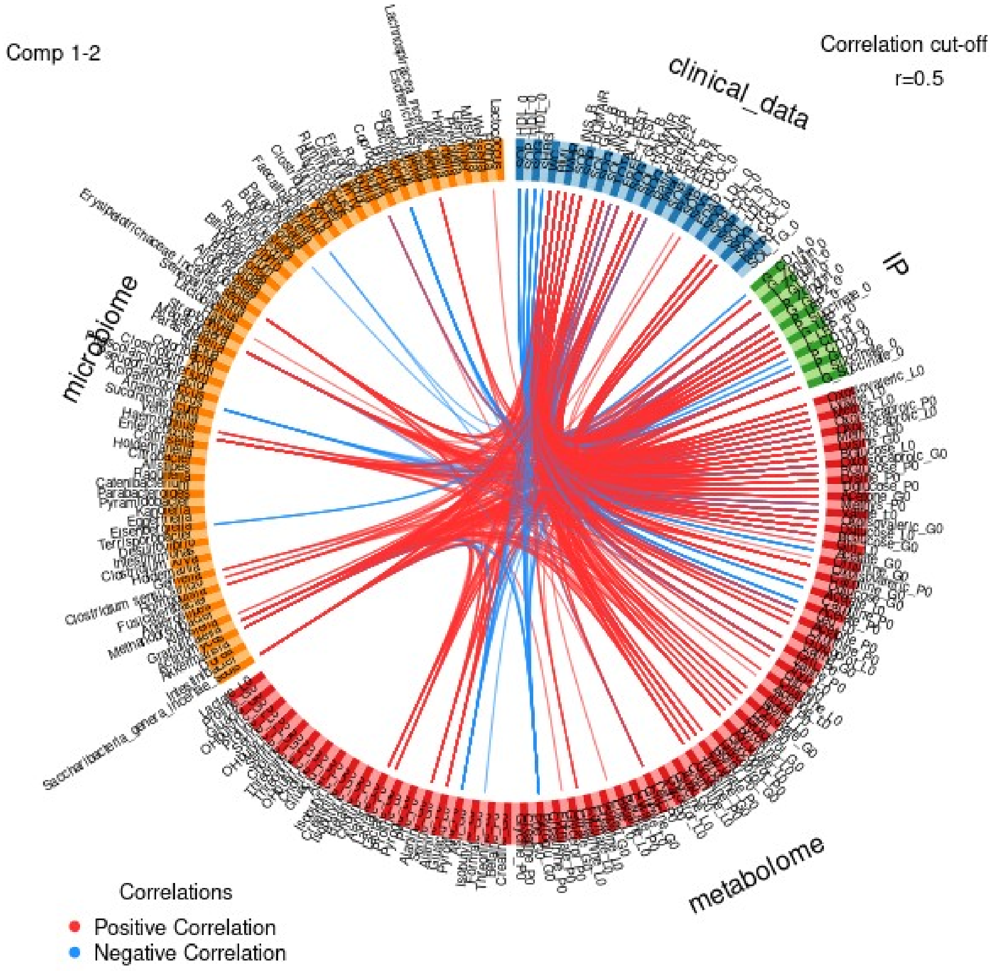
	FREE_TEST	RatioFreeTE2	TOTAL_TEST
SHBG	-0.53	-0.49	-0.52
SOG_HDL_0	-0.54	-0.5	-0.53
SOL_HDL_0	-0.54	-0.5	-0.53
SOP_HDL_0	-0.6	-0.55	-0.58
G_CD14_0	-0.58	-0.55	-0.58
L_CD14_0	-0.52	-0.5	-0.52
P_CD14_0	-0.6	-0.58	-0.6
G_LBP__0	-0.42	-0.49	-0.46
G_GLP2_0	0.51	0.44	0.48
L_GLP2_0	0.56	0.49	0.53
P_GLP2_0	0.59	0.52	0.57
Leu_L0	0.55	0.47	0.52
Lactate_G0	-0.43	-0.46	-0.45
Creatinine_L0	0.57	0.54	0.56
Blautia	-0.46	-0.5	-0.49
Clostridium IV	-0.45	-0.42	-0.45
Clostridium sensu stricto	0.54	0.52	0.54
Eggerthella	-0.43	-0.47	-0.46
Eisenbergiella	-0.55	-0.52	-0.55
Haemophilus	0.53	0.49	0.52
Klebsiella	0.5	0.48	0.5
Romboutsia	0.52	0.48	0.51

49. Figure: Mean correlation between the main features and obese women with PCOS.





50. Figure: Mean correlation between non obese women with the syndrome and the main features



51. Figure: Main correlations between features of each block. With a cutoff of 0.55. Red connections are positive correlation. The blue ones are negatives.

4 Discussion

Polycystic ovarian syndrome (PCOS), is a very common disease with 20% of prevalence. This syndrome implies a dysfunction in several areas of the body. The causes of PCOS may come from several sources, and it is still unknown the etiology. However, recent studies suggest that the syndrome is a multigenic disorder with strong influence of environmental factors. Some of the representations of this condition are abdominal adiposity, insulin resistance, obesity, metabolic disorders, and cardiovascular risks. Furthermore, it can be associated with other comorbidities such as endometrial cancer, diabetes mellitus II and psychiatric disorders (1–3).

In the context of this project, the aim was to have a better knowledge of the syndrome, since it is a disease that most of the population has a poor understanding of its mechanisms , even physicians not specialized in the subject.

Specifically, the project aimed to understand the syndrome by a multi omic-exomic joint integrative analysis of four blocks of data: clinical and biochemical markers, intestinal permeability markers, which explain the pathways of the inflammation response, the metabolome which is the product of the expressed genes, and the gut microbiome, since the bacterias in de body interact with the products of the genes.

To understand the features of each block, a univariate analysis was performed, only in the descriptive sense, after that, a multivariate descriptive analysis with several techniques was computed, depending on the number of features and observations.The NIPALS algorithm was used when more features were considered than observation, this holds for all data-sets except on the intestinal permeability markers. Finally, to understand the behavior of the three methods studied, PLS2, O2PLS and DIABLO firstly, the microbiome-metabolome integration was performed because of two reasons: the first one is of biological interest because it is described in the literature that both metabolites and the species on the gut interact with each other and change the abundance of the microorganism as a function of the metabolome and vice-versa. The second one is implied in the mentioned sentence because these two blocks are co-dependent, we have to figure out which is more important for the impact of the other block. Finally, because of this co-dependence and the fact that the first two methods used, i.e, PLS2 and O2PLS, only can integrate two blocks of data, we opted to use the generalized Sparse Canonical Correlation Analysis, also known as sparse block PLS2-DA, which is a specific design of sGCCA, also called DIABLO.

In which concerns the univariate descriptive analysis, though the results, in general, were following the literature, only in a descriptive way, some of the results were not in accordance with the reference bibliography. This could be since maybe the inference made on the same data on the published results, made the approach of the inference analysis with parametric approaches, whilst most of the data does not follow a normal distribution nor assumptions such as the homogeneity of variance is accomplished. However they transformed the data in such a way that accomplishes the normality and for that reason, they used parametric methods, but on other references, their conclusions were in discrepancy. Thus this should be further investigated.

Regarding the microbiome, we found differences when we transformed via log-center-ratio on the phylum. But on the other hand, the references used as *gold standard* did not observe these results.

For the microbiome data, the beta diversity was assessed via non-metric multidimensional scaling when the similarity matrix was not euclidean. Furthermore, the results from the beta diversity we could not compare with reference literature because there, this type of diversity is shown with bar plots, but actually, the beta diversity comes in form of a similarity matrix.

It has to be noted that the assumptions for all the principal component analysis methods used were not computed. So these results should be further investigated in this sense. However, they are robusts to the central limit theorem, because the methods are the sum of several features

For the joint integrative analysis, we opted for a joint horizontal supervised analysis already mentioned. However, for computational reasons, we could not take the profit from the sparse methodology. On the other hand, other methods may provide better results, such as the one in the *multiblock* package in **R.**

Another interesting approach, since the study by itself, has a design of repeated measures in the sense that we have measurements at both times basal and postprandial, vertical integration could be used such as the multivariate integration method, which allows the integration of independent studies measured on the same *P* predictors. In other words, a joint statistical analysis can also be made by variables after-post measurement, instead of doing it by subjects.

Regarding the clinical data, the feature that differentiates between males and females was obviuosly the sex steroids, and the Sex Hormone Binding Protein. Testosterone was higher in men but lower in obese men, but still higher than women. On females, the levels of testosterone were higher on lean females with the syndrome but a lower concentration of this hormone is found on obese females with this condition. These particular features were crucial for the differentiation between groups on the integrative analysis. The other hormone, the estradiol, was lower in patients with the syndrome in comparison with control women. Between women with PCOS, this hormone presents higher concentration levels on obese subjects. Regarding obese females with the syndrome, the levels of basal insulin, the measurements of HOMA-IR and hs-CRP were higher than in the other subjects. In what concerns obesity the main biomarkers that distinguish between subjects were BMI, waist circumference, and waist to hip ratio.

On the intestinal permeability markers, when the glucose was administrated to the patients, the levels of zonulin and LBP were higher on obese females with the disorder, as well as the succinate.

In the second-last block, the levels of alanine were higher on females with PCOS especially on obese women with the disorder at basal levels. The other metabolic feature that differentiates between females with the syndrome and those without it is the acetate. This metabolite was lower than control females but higher than the males. And the pyroglutamate, the females with this condition have the lowest values regarding males and control females.

Last but not least important, the microbiome data set comprises two sub-datasets, the phylum and the genus of the microorganism. Briefly, the alpha diversity for lean women with PCOS is higher than the rest of the females and similar value in comparison with males, and lower in obese subjects. Regarding the beta diversity, employing MDS and PLS-DA, we observed that Firmicutes and Actinobacteria were higher on obese females whilst Verrucomicrobia was higher on lean women with PCOS and Bacteriodetes have the lower presence among obese females with this condition. In what concerns with the genus, the alpha diversity was higher in obese PCOS females than control females, with similar values in comparison with men. On the contrary, using the dimensional reduction techniques mentioned above, one can conclude that Fusicatenibacter and Dorea are one of the least abundant genera on lean females with PCOS, but Dorea was especially high abundant in obese females with the syndrome. Oscillobacter and Desulovibrio are the most abundant in lean PCOS females.

Regarding the interaction, of the several blocks, is mainly a positive correlation, though there are features that correlates negatively. Moreover, and surprisingly, six genera correlates positively or negatively with testosterone, and some of them are characteristic of women with PCOS.

Although in the context of this study, the joint integrative analysis at the postprandial level was not performed, in general almost all the features studied, without taking into account the body shape measurements, and those variables that were not measured on both times, the plasma levels were higher at the basal time and lower at postprandial time. As commented, it should be interesting to pursue this study to observe what happens before and after the meal ingestion with a vertical integration.

On all blocks, especially on the microbiome, with more information provided with on the other data sets, when the joint analysis is computed, the distinction between groups is clearer. Moreover, when performing interaction score plots, i.e the joint scores of different blocks, for example between the microbiome and the clinical data, or the metabolome-clinical integration, due to the sex steroids, and other clinical variables, the distinction between all groups is more evident.

The obtained results may suggest that the metabolites, clinical biomarkers, body shape measurements, and microbiome are closely related to each other. For example, in lean females with the syndrome, more microorganisms are related to this disease rather than obese females with PCOS. On the contrary, in obese females with the disorder, almost all metabolites are in higher abundance on the body in comparison with the lean PCOS women. Furthermore, the sex steroids could play a more precise role in obese subjects given the results from the joint analysis, mainly because these hormones are present as the ones that differentiate this group, and the contrary happens with lean females with the condition, lower concentration of metabolites are present. As mentioned the same happens with microorganisms in lean PCOS women. The intestinal permeability biomarkers, which essentially are related to the inflammatory response, are more present in obese females with PCOS and low levels are found on lean PCOS females. Therefore, an important finding of this work is that the behaviour of women with PCOS is different if they are lean or obese.

5 Conclusion

This study is centered on the Polycystic Ovarian Syndrome, however, the methods for the joint integrative analysis were the core of this project. Regarding the technical details, with more information provided from different datasets, the more robust are the results rather than investigating each block individually. Foremost, the multivariate descriptive analysis is also a powerful tool for analyzing each data set, when high-dimensional data is provided.

The are several methods for integration of multiple blocks, we choose a joint integrative approach. The utility of these kind of methods, in general, all integrative methods, one can get more insights into the system under study. But firstly one has to be familiar with each block for the final integrative analysis. Although a univariate, bivariate and multivariate analysis proved us information of each variable, with the whole integration, we could obtain more robust conclusions.

To obtain more decisive results, with the method employed, it should be advisable to compute the sparse version of DIABLO. Other methods were not investigated.

Regarding the biological aspects of all subjects, it was observed that at the basal time, the levels of almost all blood features, such as metabolites, intestinal permeability biomarkers, and the clinical data set, were higher than at the postprandial time.

In what concerns PCOS, sex steroids may play a more significant role on obese women than non-obese. The most relevant metabolites are in more concentration in the organism on obese females with the syndrome than lean women. Though obese females present microorganisms that differentiate between non-obese, this microbiota was not so important in comparison with lean PCOS women. A negatice correlation of the levels of intestinal permeability markers are found on non-obese females with the syndrome, which is related to the inflammatory response, whilst a positive correlation is found on obese females with the syndrome.

Regarding the interaction between the blocks, a positive correlation between all data sets is predominant, however, there are features that correlates in negative way. The clinical data interacts stringly with all blocks, followed by the metabolome and the microbiome interaction.

In response to the research question, it was analyzed the correlation between features, and we found that surpinsingly, there are eight genera that correlates in a negative or positive way with the testosterone, and some of them are characteristic of the lean PCOS women.

It has to be mentioned that this project is mainly focused on the biostatistics aspects of the analysis of the syndrome. Specialized scientists on the subject, should review these results to obtain a biological and medical conclusion.

6 Future aspects.

To carry on this project in the future, several aspects should be considered.

1. Apart from the descriptive univariate and multivariate analysis, and inferential analysis should be performed, compare with the published results. Specifically the univariate inferential statistics.
2. Transform the data as needed.
3. Regarding the microbiome, evaluate the PLS-DA for the phylum and the genus, instead of MDS.
4. Get the relationship between the phyla and genus.
5. Compute the results of DIABLO in a computer cluster.
6. Evaluate other methods of integration such as multi-block analysis (60)

7 References

1. Bulsara J, Patel P, Soni A, Acharya S. A review: Brief insight into Polycystic Ovarian syndrome. Endocr Metab Sci. 2021 Jun 30;3:100085.

2. El Hayek S, Bitar L, Hamdar LH, Mirza FG, Daoud G. Poly Cystic Ovarian Syndrome: An Updated Overview. Front Physiol. 2016 Apr 5;7:124.

3. Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. Nat Rev Endocrinol. 2018 May;14(5):270–84.

4. Insenser M, Moncayo S, Martínez-García MÁ, Fernández-Durán E, Samino S, Álvarez-Blasco F, et al. 2D Diffusion-Ordered 1H-NMR Spectroscopy Lipidomic Profiling after Oral Single Macronutrient Loads: Influence of Obesity, Sex, and Female Androgen Excess. Mol Nutr Food Res. 2020;64(9):1900928.

5. Montes-Nieto R, Insenser M, Murri M, Fernández-Durán E, Ojeda-Ojeda M, Martínez-García MÁ, et al. Plasma thiobarbituric acid reactive substances (TBARS) in young adults: Obesity increases fasting levels only in men whereas glucose ingestion, and not protein or lipid intake, increases postprandial concentrations regardless of sex and obesity. Mol Nutr Food Res. 2017;61(11):1700425.

6. Insenser M, Murri M, del Campo R, Martínez-García MÁ, Fernández-Durán E, Escobar-Morreale HF. Gut Microbiota and the Polycystic Ovary Syndrome: Influence of Sex, Sex Hormones, and Obesity. J Clin Endocrinol Metab. 2018 Jul 1;103(7):2552–62.

7. Martínez-García MÁ, Moncayo S, Insenser M, Álvarez-Blasco F, Luque-Ramírez M, Escobar-Morreale HF. Metabolic Cytokines at Fasting and During Macronutrient Challenges: Influence of Obesity, Female Androgen Excess and Sex. Nutrients. 2019 Oct 24;11(11):2566.

8. Martínez-García MÁ, Moncayo S, Insenser M, Montes-Nieto R, Fernández-Durán E, Álvarez-Blasco F, et al. Postprandial inflammatory responses after oral glucose, lipid and protein challenges: Influence of obesity, sex and polycystic ovary syndrome. Clin Nutr. 2020 Mar;39(3):876–85.

9. Martínez-García MÁ, Ojeda-Ojeda M, Rodríguez-Martín E, Insenser M, Moncayo S, Álvarez-Blasco F, et al. TLR2 and TLR4 Surface and Gene Expression in White Blood Cells after Fasting and Oral Glucose, Lipid and Protein Challenges: Influence of Obesity and Sex Hormones. Biomolecules. 2020 Jan 9;10(1):111.

10. Martínez-García MÁ, Moncayo S, Insenser M, Álvarez-Blasco F, Luque-Ramírez M, Escobar-Morreale HF. Postprandial responses of circulating energy

homeostasis mediators to single macronutrient challenges: influence of obesity and sex hormones. Food Funct. 2021;12(3):1051–62.

11. Moncayo S, Insenser M, Martínez-García MÁ, Fuertes-Martín R, Amigó-Grau N, Álvarez-Blasco F, et al. Acute-phase glycoprotein profile responses to different oral macronutrient challenges: Influence of sex, functional hyperandrogenism and obesity. Clin Nutr. 2021 Mar;40(3):1241–6.

12. Santiago-Rodriguez TM, Hollister EB. Multi ‘omic data integration: A review of concepts, considerations, and approaches. Semin Perinatol. 2021 Oct 1;45(6):151456.

13. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinforma Biol Insights. 2020 Jan 31;14:1177932219899051.

14. Shaffer M, Armstrong AJ, Phelan VV, Reisdorph N, Lozupone CA. Microbiome and Metabolome Data Integration Provides Insight Into Health and Disease. Transl Res J Lab Clin Med. 2017 Nov;189:51–64.

15. López de Maturana E, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the Integration of Omics and Non-Omics Data. Genes. 2019 Mar 20;10(3):238.

16. R: The R Project for Statistical Computing [Internet]. [cited 2021 Sep 28]. Available from: https://www.r-project.org/

17. mixOmics | | Omics Data Integration Project [Internet]. [cited 2021 Sep 28]. Available from: http://mixomics.org/

18. Meng C, Helm D, Frejno M, Kuster B. moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets. J Proteome Res. 2016 Mar 4;15(3):755–65.

19. Meng C, Kuster B, Culhane AC, Gholami AM. A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics. 2014 May 29;15(1):162.

20. Gomez-Cabrero D, Tarazona S, Ferreirós-Vidal I, Ramirez RN, Company C, Schmidt A, et al. STATegra, a comprehensive multi-omics dataset of B-cell differentiation in mouse. Sci Data. 2019 Oct 31;6(1):256.

21. STATegra EMS v1.0 Documentation [Internet]. [cited 2021 Sep 28]. Available from: https://stategraems.readthedocs.io/en/latest/

22. Enterprise Open Source and Linux [Internet]. Ubuntu. [cited 2021 Dec 8]. Available from: https://ubuntu.com/

23. Microsoft Excel Spreadsheet Software | Microsoft 365 [Internet]. [cited 2021 Dec 18]. Available from: https://www.microsoft.com/en-us/microsoft-365/excel

24. xlsx.pdf [Internet]. [cited 2021 Dec 18]. Available from: https://cran.r-project.org/web/packages/xlsx/xlsx.pdf

25. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. PLoS ONE. 2013 Apr 22;8(4):e61217.

26. Create Elegant Data Visualisations Using the Grammar of Graphics [Internet]. [cited 2021 Dec 18]. Available from: https://ggplot2.tidyverse.org/

27. chemometrics.pdf [Internet]. [cited 2021 Dec 18]. Available from: https://cran.r-project.org/web/packages/chemometrics/chemometrics.pdf

28. Erichson NB, Zheng P, Manohar K, Brunton SL, Kutz JN, Aravkin AY. Sparse Principal Component Analysis via Variable Projection. SIAM J Appl Math. 2020 Jan;80(2):977–1002.

29. pls.pdf [Internet]. [cited 2021 Dec 18]. Available from: https://cran.r-project.org/web/packages/pls/pls.pdf

30. vegan.pdf [Internet]. [cited 2021 Dec 18]. Available from: https://cran.r-project.org/web/packages/vegan/vegan.pdf

31. OmicsPLS.pdf [Internet]. [cited 2021 Dec 18]. Available from: https://cran.r-project.org/web/packages/OmicsPLS/OmicsPLS.pdf

32. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Transact A Math Phys Eng Sci. 2016 Apr 13;374(2065):20150202.

33. Bishop - 2006 - Pattern recognition and machine learning.pdf [Internet]. [cited 2022 Jan 4]. Available from: http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf

34. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? BMC Bioinformatics. 2020 Dec 9;21(Suppl 1):2.

35. Preda C, Saporta G. THE NIPALS ALGORITHM FOR MISSING FUNCTIONAL DATA. :12.

36. Everitt BS, Dunn G. Multidimensional Scaling. In: Applied Multivariate Data Analysis [Internet]. John Wiley & Sons, Ltd; 2001 [cited 2022 Jan 4]. p. 93–124. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118887486.ch5

37. Cuadras - NUEVOS MÉTODOS DE ANÁLISIS MULTIVARIANTE.pdf [Internet]. [cited 2022 Jan 4]. Available from: http://www.est.uc3m.es/esp/nueva_docencia/getafe/estadistica/analisis_multivariante/doc_generica/archivos/metodos.pdf

38. Peña D. Análisis de Datos Multivariantes. 2002 Jan 1;

39. Stocchero M, Locci E, d’Aloja E, Nioi M, Baraldi E, Giordano G. PLS2 in

Metabolomics. Metabolites. 2019 Mar 15;9(3):51.

40. Trygg J, Wold S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. J Chemom. 2003;17(1):53–64.

41. Tenenhaus A, Philippe C, Guillemot V, Le Cao K-A, Grill J, Frouin V. Variable selection for generalized canonical correlation analysis. Biostat Oxf Engl. 2014 Jul;15(3):569–83.

42. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B Methodol. 1996;58(1):267–88.

43. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019 Sep 1;35(17):3055–62.

44. Whittaker RH. Vegetation of the Siskiyou Mountains, Oregon and California. Ecol Monogr. 1960;30(3):279–338.

45. Tucker CM, Cadotte MW, Carvalho SB, Davies TJ, Ferrier S, Fritz SA, et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biol Rev Camb Philos Soc. 2017 May;92(2):698–715.

46. Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945;26(3):297–302.

47. Bouhaddani S el, Houwing-Duistermaat J, Salo P, Perola M, Jongbloed G, Uh H-W. Evaluation of O2PLS in Omics data integration. BMC Bioinformatics. 2016 Jan 20;17(2):S11.

48. Brownlee J. A Gentle Introduction to k-fold Cross-Validation [Internet]. Machine Learning Mastery. 2018 [cited 2021 Dec 31]. Available from: https://machinelearningmastery.com/k-fold-cross-validation/

49. Legro RS, Finegood D, Dunaif A. A Fasting Glucose to Insulin Ratio Is a Useful Measure of Insulin Sensitivity in Women with Polycystic Ovary Syndrome1. J Clin Endocrinol Metab. 1998 Aug 1;83(8):2694–8.

50. Sumithra NUC, Lakshmi RL, Leela Menon N, Subhakumari KN, Sheejamol VS. Evaluation of Oxidative Stress and hsCRP in Polycystic Ovarian Syndrome in a Tertiary Care Hospital. Indian J Clin Biochem. 2015 Apr;30(2):161–6.

51. Cholesterol Levels: By age, LDL, HDL, and More [Internet]. Healthline. 2021 [cited 2021 Nov 1]. Available from: https://www.healthline.com/health/high-cholesterol/levels-by-age

52. Gayoso-Diz P, Otero-Gonzalez A, Rodriguez-Alvarez MX, Gude F, Cadarso-Suarez C, García F, et al. Insulin resistance index (HOMA-IR) levels in a general adult population: Curves percentile by gender and age. The EPIRCE study. Diabetes Res Clin Pract. 2011 Oct 1;94(1):146–55.

53. Dayal D, Jain H, Attri SV, Bharti B, Bhalla AK. Relationship of High Sensitivity

C-Reactive Protein Levels to Anthropometric and other Metabolic Parameters in Indian Children with Simple Overweight and Obesity. J Clin Diagn Res JCDR. 2014 Aug;8(8):PC05–8.

54. Esser N, Legrand-Poels S, Piette J, Scheen AJ, Paquot N. Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes. Diabetes Res Clin Pract. 2014 Aug 1;105(2):141–50.

55. Reiner AP, Lange EM, Jenny NS, Chaves PHM, Ellis J, Li J, et al. Soluble CD14. Arterioscler Thromb Vasc Biol. 2013 Jan 1;33(1):158–64.

56. Liu X, Lu L, Yao P, Ma Y, Wang F, Jin Q, et al. Lipopolysaccharide binding protein, obesity status and incidence of metabolic syndrome: a prospective study among middle-aged and older Chinese. Diabetologia. 2014 Sep 1;57(9):1834–41.

57. Martin TR, Rubenfeld GD, Ruzinski JT, Goodman RB, Steinberg KP, Leturcq DJ, et al. Relationship between soluble CD14, lipopolysaccharide binding protein, and the alveolar inflammatory response in patients with acute respiratory distress syndrome. Am J Respir Crit Care Med. 1997 Mar 1;155(3):937–44.

58. Tajik N, Frech M, Schulz O, Schälter F, Lucas S, Azizov V, et al. Targeting zonulin and intestinal epithelial barrier function to prevent onset of arthritis. Nat Commun. 2020 Apr 24;11(1):1995.

59. Nguyen QP, Karagas MR, Madan JC, Dade E, Palys TJ, Morrison HG, et al. Associations between the gut microbiome and metabolome in early life. BMC Microbiol. 2021 Aug 28;21(1):238.

60. SMILDE AGEN, TORMOD T.. LILAND KH. MULTIBLOCK DATA ANALYSIS: understanding complex data in the natural sciences. S.l.: WILEY-BLACKWELL; 2022.

8 Appendices

8.1 Descriptive table of the Clinical Data

Variable	Control Females				PCOS				Control Males				
	No Obese		Obese		No Obese		Obese		No Obese		Obese		
	9		8		9		8		10		9		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Age (years)	26.33	4.61	27.25	6.48	23.89	8.34	29.88	5.14	24	4.37	24.89	3.79	EDAD
BMI	23.41	1.9	35.88	4.23	24.06	2.5	37.03	4.77	22.92	1.77	34.41	3.44	BMI
Waist circumference (cm)	76.39	8.68	100.38	17.42	72	7.14	104.56	10.53	81.2	4.96	109.78	12.67	WC
Waist to hip ratio	0.75	0.08	0.83	0.12	0.73	0.05	0.85	0.06	0.83	0.04	0.9	0.05	WHR
SHBG (nmlol/L)	55.6	25.46	42.99	14.05	50.07	20.96	32.04	12.92	26.81	10.18	20.42	6.11	SHBG
Total testosterone (nmol/L)	44.79	8.4	57.16	15.48	70.69	20.43	68.44	29.72	533.72	96.06	498.53	102.98	TOTAL_TEST
Free testosterone (pmol/L)	21.37	6.61	30.93	8	36	11.8	45.33	24.36	449.95	104.4	464.15	94.1	FREE_TEST
Total estradiol (nmol/L)	40.44	17.22	75.12	54.52	49.56	54.67	40.62	13.32	18.5	4.22	25.56	7.14	Total_ESTR
Free estradiol (pmol/L)	2.72	1.11	5.26	2.99	3.58	4.3	3.45	1.36	1.79	0.55	2.6	0.75	Free_ESTRA
Ratio free T/E2	8.6	2.64	7.57	4.67	17.45	11.98	14.01	5.45	263.4	71.37	186.39	38.26	RatioFreeTE2
hsCRP (nmol/L)	2.78	2.35	4.03	2.89	2.09	2.26	6.82	7.73	3.28	2.63	3.22	1.29	hsCRP
Mean fasting insulin (pmol/L)	7.95	2.88	11.09	3.03	7.53	4.14	13.54	3.59	5.78	1.53	10.72	3.94	InsulinaBasal_mean
Mean fasting glucose (mmol/L)	84.04	6.6	95.76	7.05	80.6	8.34	86.89	8.55	89.17	9.07	94.2	6.51	GlucosaBasal_mean
Insulin sensitivity index	6.58	2.85	3.28	1.18	8.12	4.68	3.53	1.44	7.32	2.77	3.82	1.61	SOG_ISI
HOMA-IR	1.65	0.63	2.63	0.68	1.52	0.94	2.88	0.73	1.27	0.37	2.52	0.96	HOMAIRmean
Triglycerides (mg/dL)	74.11	31.78	81	29.82	79.44	22.16	100.96	41.45	78.83	25.3	101.56	35.59	TGmean
Cholesterol (mg/dL)	171.93	35.97	181.13	36.06	167.89	37.87	171.75	35.69	157.73	22.27	187.26	33.58	COLmean
HDL cholesterol (mg/dL)	53.59	10.11	49.79	7.49	54.19	8.89	47.21	7.49	47.43	7.19	37.63	3.28	HDLmean
LDL cholesterol (mg/dL)	103.07	30.19	111.46	22.25	97.41	31.95	104	31.07	94.13	16.93	128.85	30.69	LDLmean
Basal insulin: Glucose day (pmol/L)	7.69	3.74	13.31	6.49	6.63	4.48	11.86	3.92	5.5	1.84	10.49	3.77	SOG_INS_B
Basal insulin: Lipid day (pmol/L)	8.07	3.22	9.44	2.5	7.82	4.13	15.99	4.51	6.04	1.81	10.38	4.24	SOL_INS_B
Basal insulin: Protein day (pmol/L)	8.09	3.2	10.51	3.21	8.12	4.76	12.76	3.54	5.81	1.69	11.3	4.15	SOP_INS_B
Basal glucose: Glucose day (mmol/L)	85.16	6.74	100.75	7.67	84.16	7.68	88.9	8.18	94.3	9.33	96.2	7.9	SOG_GLC_B
Basal glucose: Lipid day (mmol/L)	80.8	7.57	92.03	8.58	76.53	9.82	84.95	9.52	86.82	10.45	91.2	8.59	SOL_GLC_B
Basal glucose: Protein day (mmol/L)	86.16	7.85	94.5	8.33	81.11	8.48	86.83	10.46	86.4	9.65	95.19	5.26	SOP_GLC_B
HOMA-IR: Glucose day	1.64	0.87	3.32	1.65	1.4	0.99	2.62	0.94	1.28	0.46	2.49	0.9	SOG_HOMAIR
HOMA-IR: Lipid day	1.61	0.7	2.12	0.5	1.5	0.88	3.32	0.85	1.3	0.44	2.38	1.05	SOL_HOMAIR
HOMA-IR: Protein day	1.71	0.69	2.45	0.79	1.66	1.09	2.71	0.71	1.23	0.37	2.68	1.04	SOP_HOMAIR
Insulin AUC: Glucose day (pmol/L)	48.5	34.16	61.61	22.13	49.3	30.88	72.56	36.4	39.23	25.86	62.91	37.36	SOG_INS_AUC
Insulin AUC: Lipid day (pmol/L)	-0.69	1.56	-0.34	2.01	-1.49	1.61	-2.6	1.74	-0.65	1.13	-1.09	2.01	SOL_INS_AUC
Insulin AUC: Protein day (pmol/L)	5.87	2.95	16.56	6.74	7.07	4.71	11.65	8.86	6.45	5.41	7.57	3.37	SOP_INS_AUC
Glucose AUC: Glucose day (mmol/L)	25.56	25.46	36.56	20.84	17.11	30.48	45.82	20.02	39.72	24.09	50.66	23.77	SOG_GLU_AUC
Glucose AUC: Lipid day (mmol/L)	-1.01	3.97	-4.73	4.99	0.6	5.12	-3.46	4.05	-1.73	5.01	-0.03	5.31	SOL_GLU_AUC
Glucose AUC: Protein day (mmol/L)	-0.81	3.75	-4.01	6.18	-1.35	6.7	-0.14	5.27	1.62	4.38	0.08	3.98	SOP_GLU_AUC
Basal triglycerides: Glucose day (mg/dL)	66.33	16.22	80.5	38.17	67.33	16.02	106	56.61	73.2	25.94	104.22	42.99	SOG_TG_0
Basal triglycerides: Lipid day (mg/dL)	89.11	56.41	87.12	30.49	93.78	39.7	107	45.61	82.5	19.15	108	43.51	SOL_TG_0
Basal triglycerides: Protein day (mg/dL)	66.89	28.35	75.38	26.77	77.22	19.36	89.88	37	80.8	36.71	92.44	27.04	SOP_TG_0
Triglycerides AUC: Glucose day (mg/dL)	-1.11	4.58	3.59	5.06	-1	6.36	4.59	8.58	1.23	6.39	8.69	11.95	SOGAUCTG
Triglycerides AUC: Lipid day (mg/dL)	2.75	31.41	12.53	17.41	13.22	25.16	5.16	16.48	12.68	12.49	8	23.33	SOLAUCTG
Triglycerides AUC: Protein day (mg/dL)	-7.83	7.38	-3.94	3.57	-2.75	9.38	-1.09	7.01	-7.22	13.01	1.39	7.56	SOPAUCTG
Basal cholesterol: Glucose day (mg/dL)	166	36.15	181.88	45.07	166.33	40.99	169	41.27	149.6	21.15	185.89	35.09	SOG_COL_0
Basal cholesterol: Lipid day (mg/dL)	171.11	36.73	181.75	33.33	168.78	38.02	170.38	33.25	162.2	23.22	190.22	37.17	SOL_COL_0
Basal cholesterol: Protein day (mg/dL)	178.67	36.82	179.75	32.29	168.56	37.84	175.88	34.34	161.4	24.84	185.67	29.8	SOP_COL_0
Cholesterol AUC: Glucose day (mg/dL)	-1.69	10.39	-6.81	8.48	-3.86	9.32	0.38	8.02	-1.93	2.69	-3.28	4.88	SOGAUCCOL
Cholesterol AUC: Lipid day (mg/dL)	-8.17	8.61	-6.19	4.96	-5.36	6.12	-3.09	7.05	-6.4	5.02	-3.11	3.54	SOLAUCCOL
Cholesterol AUC: Protein day (mg/dL)	-14.67	9.59	-10.06	3.77	-10.94	9.41	-5.72	3.79	-10.5	8.12	-5.14	6.03	SOPAUCCOL
Basal HDL: Glucose day (mg/dL)	51.11	9.98	48.12	9.33	52.78	8.77	46.25	7.57	45.3	6.29	37.33	3.08	SOG_HDL_0
Basal HDL: Lipid day (mg/dL)	52.78	10.93	51.12	6.58	54.44	9.26	46.75	7.61	49	7.2	37.22	2.95	SOL_HDL_0
Basal HDL: Protein day (mg/dL)	56.89	10.51	50.12	8.17	55.33	10.17	48.62	7.95	48	9.2	38.33	4.9	SOP_HDL_0
HDL AUC: Glucose day (mg/dL)	-1.64	3.24	-0.22	4.37	-0.92	2.37	-0.41	2.02	-1.02	2.25	-1.64	2.17	SOGAUCHDL
HDL AUC: Lipid day (mg/dL)	-4.31	3.94	-3.06	2.34	-2.89	2.65	-1.75	2.41	-2.85	1.79	-0.72	0.79	SOLAUCHDL
HDL AUC: Protein day (mg/dL)	-4.94	3.54	-3.47	1.82	-3.61	3.88	-1.12	1.69	-3.42	2.94	-1.31	1.69	SOPAUCHDL
Basal LDL: Glucose day (mg/dL)	101.22	32.32	107.38	17.26	99.67	34.73	101.12	37.65	89.2	17	127.33	31.41	SOG_LDL_0
Basal LDL: Lipid day (mg/dL)	100.11	29.2	112.75	27.64	95.11	32.36	101.88	28.34	96.4	20.17	130.78	32.43	SOL_LDL_0
Basal LDL: Protein day (mg/dL)	107.89	30.04	114.25	25.13	97.44	30.78	109	29.84	96.8	15.35	128.44	29.78	SOP_LDL_0
LDL AUC: Glucose day (mg/dL)	-1.44	5.56	0.12	12.44	-2.72	6.56	-0.16	5.52	-1.18	3.09	-3.39	5.04	SOGAUCLDL
LDL AUC: Lipid day (mg/dL)	-5.53	6.9	-5.53	4.53	-5.08	2.36	-2.47	5.52	-6.22	4.59	-3.22	3.14	SOLAUCLDL
LDL AUC: Protein day (mg/dL)	-8.06	6.43	-5.94	2.2	-6.78	5.71	-4.12	3.4	-5.68	4.04	-4.08	4.88	SOPAUCLDL

11. Table: Descriptive numerical table of the Clinical Data

8.5 Descriptive table of Intestinal Permeability makers data set

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	**SD**	Mean	SD	**Mean**	**SD**	Mean	SD	**Mean**	**SD**	**Mean**	**SD**
Mean basal value sCD14	1622.53	356.55	1349.8	158.86	1376.26	196.62	1395.71	172.3	1461.57	242.92	1218.02	157.48
Basal glucose load sCD14	1616.46	349.8	1333.78	168.48	1423.12	233.21	1392.3	176.97	1463.6	265.09	1162.24	183.34
Basal lipid load sCD14	1655.84	405.99	1371.43	194.75	1353.05	200.78	1408.38	189.84	1506.82	261.62	1236.1	229.28
Basal protein load sCD14	1595.27	367.41	1344.18	165.22	1352.62	194.04	1386.45	225.45	1414.27	286.78	1255.71	179.92
Postprandial AUC glucose sCD14	-56.2	77.48	-28.39	55.17	-36.21	52.44	-29.73	52.23	-32.8	36.27	33.71	39.57
Postprandial AUC lipid sCD14	-84.47	57.17	-16.63	53.27	-28.4	44.51	-36.66	82.26	-40.16	41.69	3.87	69.53
Postprandial AUC protein sCD14	-98.09	49.26	-51.3	31.31	-43.9	37.85	-14.94	54.48	-57.13	34.04	-45.16	35.67
Mean basal value LBP	14.51	3.92	17.18	6.32	13.67	4.35	20.23	5.57	12.92	3.69	15.72	3.14
Basal glucose load LBP	15.1	4.87	16.53	5.85	15.38	6.27	22.1	8.37	13.35	4.46	16.43	5.82
Basal lipid load LBP	13.06	3.06	16.76	7.12	12.26	4.84	17.86	4.15	12.3	3.23	14.66	2.91
Basal protein load LBP	15.36	5.13	18.24	6.81	13.37	3.61	20.74	5.64	13.1	4.05	16.07	3.75
Postprandial AUC glucose LBP	-0.13	0.81	-0.93	0.74	-0.3	1.44	-1.25	2.24	-0.31	0.55	-0.87	1.31
Postprandial AUC lipid LBP	-0.25	0.65	-0.12	1.16	-0.21	0.66	-0.54	0.89	0.2	0.82	0.34	0.82
Postprandial AUC protein LBP	-0.43	0.72	-0.79	1.09	-0.37	0.5	-0.42	1.33	-0.17	0.25	-0.34	0.5
Mean basal value GLP2	4.34	0.6	4.11	0.95	3.53	0.53	4.88	0.75	4.49	0.82	5.61	2.28
Basal glucose load GLP2	4.34	0.62	4.3	0.91	3.52	0.42	4.91	1.21	4.46	0.78	5.33	2.37
Basal lipid load GLP2	4.02	0.9	4	1.11	3.29	0.76	4.97	1.11	4.37	0.69	5.7	2.1
Basal protein load GLP2	4.66	0.81	4.04	1.07	3.78	0.62	4.76	0.53	4.64	1.22	5.81	2.53
Postprandial AUC glucose GLP2	-0.05	0.81	-0.21	0.35	-0.05	0.29	-0.32	0.26	-0.24	0.28	-0.45	0.6
Postprandial AUC lipid GLP2	0.53	0.63	0.46	0.47	0.68	0.5	0.15	0.47	0.56	0.45	0.28	0.45
Postprandial AUC protein GLP2	0.71	0.31	1.07	0.42	0.81	0.45	0.64	0.27	1.07	0.41	0.87	0.35
Mean basal value zonulin	27.36	3.21	29.26	1.85	27.82	3.34	30.79	4.78	25.85	4.05	29.41	4.01
Basal glucose load zonulin	28.28	4.55	30.67	5.03	28.97	4.55	34.59	9.44	25.64	4.74	29.41	4.92
Basal lipid load zonulin	26.82	3.47	27.83	3.95	27.1	3.83	28.91	3.21	25.13	3.98	29.41	4.05
Basal protein load zonulin	26.98	3.26	29.26	2.64	27.4	3.64	28.88	3.31	26.77	4.45	29.4	4.51
Postprandial AUC glucose zonulin	-0.11	1.32	0.33	2.01	-0.24	1.15	0.02	1.22	-0.26	0.78	0.68	1.54
Postprandial AUC lipid zonulin	-0.67	1.34	1.19	1.89	0.06	0.83	0.2	1.67	-0.08	1.4	0.49	1.17
Postprandial AUC protein zonulin	-0.78	0.89	0.05	0.97	-0.13	0.75	-0.65	1.86	-0.4	0.68	0.44	1.15
Mean basal value Succinate	36.36	7.53	43.64	14.93	37.14	9.72	57.11	20.53	31.97	13.24	41.98	20.37
Basal glucose load Succinate	31.96	12.97	44	23.65	31.76	21.29	51.41	28.77	31.58	19.36	51.06	28.23
Basal lipid load Succinate	40.77	18.71	37.89	16.26	39.35	10.34	59.32	23.25	31.77	16.95	33.03	19.76
Basal protein load Succinate	36.34	13.6	49.03	17.92	40.32	8.18	60.59	21.22	32.57	8.63	41.85	22.91
Postprandial AUC glucose Succinate	8.67	7.56	11.64	10.13	7.93	6.02	18.34	8.17	12.75	14.23	7.5	8.05
Postprandial AUC lipid Succinate	-3.95	10.76	2.42	6.23	-3.23	2.89	-1.78	8.37	1.02	10.6	0.23	4.25
Postprandial AUC protein Succinate	4.05	4.81	5.81	5.85	3.86	8.69	2.27	9.28	6.27	6.2	6.35	8.37

12. Table: Numerical descritpive table of the Intenstinal Permeability markers.

8.6 Descriptive table of the metabolome data

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean value (a.u) 2-Oxoisocaproic acid_2	0.0204	0.0183	0.0428	0.0101	0.0358	0.0144	0.0593	0.0087	0.064	0.0076	0.0597	0.0068
Mean value (a.u) Leucine_1	0.1216	0.025	0.1338	0.0349	0.1377	0.02	0.1703	0.0141	0.1981	0.0173	0.1878	0.0226
Mean value (a.u) Isoleucine_1	0.0358	0.0065	0.0402	0.0121	0.0336	0.0048	0.0555	0.0052	0.0534	0.008	0.0519	0.0074
Mean value (a.u) L-Valine_1	0.1063	0.0332	0.1409	0.035	0.1234	0.0315	0.1915	0.0136	0.1979	0.0205	0.183	0.0213
Mean value (a.u) Isobutyric acid_1	0.026	0.0087	0.0175	0.0043	0.0224	0.007	0.0192	0.0044	0.0212	0.0032	0.021	0.007
Mean value (a.u) 2-oxoisovaleric acid_1	0.0067	0.0071	0.0169	0.0043	0.0121	0.0067	0.024	0.0028	0.0212	0.002	0.0207	0.0024
Mean value (a.u) 3-Hydroxybutyric acid_1	0.0547	0.0263	0.0488	0.0212	0.0414	0.0206	0.0601	0.0204	0.1003	0.0787	0.054	0.0305
Mean value (a.u) Lactate_1	1.07	0.2641	1.1345	0.4346	0.9779	0.1789	1.3403	0.4347	0.8914	0.2428	1.1284	0.3531
Mean value (a.u) Alanine_1	0.2493	0.0477	0.2649	0.0575	0.2931	0.0507	0.3525	0.0545	0.238	0.0557	0.2902	0.0622
Mean value (a.u) Acetate_1	0.3652	0.0997	0.062	0.0364	0.3065	0.0976	0.1531	0.0158	0.1682	0.0269	0.1771	0.0766
Mean value (a.u) N-Ac-Glycoprotein_signals_1	0.0576	0.0111	0.0412	0.0082	0.0554	0.0062	0.0645	0.0151	0.0573	0.0129	0.0513	0.0104
Mean value (a.u) Acetone_1	0.0105	0.0032	0.023	0.003	0.0117	0.0023	0.0307	0.0053	0.0187	0.0026	0.0223	0.0041
Mean value (a.u) L-Glutamic acid_1	0.0454	0.0035	0.0436	0.0167	0.0453	0.0087	0.0506	0.0062	0.0457	0.0076	0.0524	0.0099
Mean value (a.u) Pyruvate_1	0.0068	0.0032	0.0209	0.0106	0.0125	0.0057	0.0305	0.0156	0.0137	0.0063	0.018	0.0055
Mean value (a.u) Pyroglutamic acid_1	0.035	0.0103	0.0157	0.0036	0.023	0.0149	0.0148	0.0043	0.0254	0.0127	0.0154	0.005
Mean value (a.u) L-Glutamine_1	0.1226	0.0266	0.133	0.0252	0.1514	0.0143	0.1683	0.0136	0.1568	0.0118	0.1516	0.024
Mean value (a.u) Citrate_2	0.0249	0.0053	0.0202	0.007	0.0252	0.0048	0.0267	0.0046	0.0242	0.0032	0.0194	0.0022
Mean value (a.u) L-Asparagine_4	0.0071	0.0006	0.0063	0.0016	0.0081	0.0016	0.0073	0.0014	0.008	0.0014	0.007	0.0011
Mean value (a.u) Creatinine_1	0.0463	0.0062	0.0399	0.0073	0.046	0.0082	0.0659	0.0281	0.0627	0.0066	0.0523	0.0093
Mean value (a.u) Creatine_1	0.0292	0.0106	0.0325	0.0107	0.0236	0.0113	0.0386	0.0072	0.0214	0.008	0.0279	0.0119
Mean value (a.u) Lysine_1	0.015	0.0171	0.0523	0.0131	0.03	0.0151	0.0619	0.0085	0.054	0.0088	0.0499	0.0088
Mean value (a.u) Ornithine_1	0.0116	0.0053	0.0193	0.0043	0.0158	0.006	0.0232	0.003	0.0215	0.0054	0.0224	0.0061
Mean value (a.u) Choline_1	0.029	0.0049	0.0264	0.0062	0.0306	0.009	0.0317	0.0042	0.0302	0.0092	0.0256	0.0085
Mean value (a.u) L-Carnitine_1	0.0934	0.0183	0.0615	0.0107	0.078	0.0206	0.0846	0.0108	0.0954	0.0093	0.081	0.0188
Mean value (a.u) Betaine_1	0.081	0.0265	0.057	0.0129	0.083	0.0247	0.0676	0.0193	0.0773	0.0114	0.0629	0.0213
Mean value (a.u) Glycine_1	0.0781	0.0236	0.0836	0.0226	0.1061	0.0195	0.0936	0.019	0.097	0.01	0.094	0.0257
Mean value (a.u) L-Threonine_1	0.0202	0.0074	0.0234	0.0107	0.0239	0.0094	0.0302	0.0051	0.0237	0.0054	0.0209	0.0066
Mean value (a.u) Glycerol_2	0.0144	0.0097	0.0272	0.007	0.0217	0.0083	0.0302	0.0076	0.0188	0.0039	0.0189	0.0069
Mean value (a.u) L-Serine_1	0.0266	0.0046	0.0238	0.008	0.0293	0.004	0.0231	0.0059	0.0225	0.0047	0.0226	0.0055
Mean value (a.u) L-Proline_3	0.0246	0.0041	0.0226	0.0102	0.0233	0.0055	0.028	0.0026	0.0263	0.0033	0.0236	0.0041
Mean value (a.u) beta-Glucose_1	0.1744	0.0393	0.2288	0.0399	0.2202	0.0442	0.2874	0.0186	0.2669	0.0238	0.2649	0.0312
Mean value (a.u) D-Glucose_2	0.2627	0.0663	0.3342	0.0608	0.3166	0.0606	0.4178	0.029	0.4015	0.0241	0.3862	0.0508
Mean value (a.u) 1-Methylhistidine_1	0.0053	0.0054	0.0135	0.0032	0.0099	0.0055	0.0159	0.0027	0.0157	0.0028	0.015	0.004
Mean value (a.u) Tyrosine_1	0.0274	0.0062	0.0325	0.0088	0.0317	0.0067	0.0402	0.0058	0.0336	0.0065	0.0358	0.0051
Mean value (a.u) L-Phenylalanine_2	0.0306	0.0039	0.0293	0.0058	0.0322	0.0043	0.0359	0.0021	0.0352	0.0048	0.0359	0.0054
Mean value (a.u) L-Tryptophan_4	0.0123	0.0028	0.0155	0.0025	0.0147	0.0026	0.0158	0.0021	0.0167	0.0027	0.0168	0.0018
Mean value (a.u) Formate_1	0.0191	0.006	0.0124	0.0018	0.0104	0.0051	0.0163	0.0023	0.0105	0.0009	0.0131	0.0014

13. Table: Descriptive nuemerical table of the metabolome data set

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Basal levels at glucose oral load day: 2-Oxoisocaproic acid_2 (a.u)	0.0216	0.0188	0.0418	0.01	0.0345	0.0125	0.0555	0.0131	0.0635	0.0097	0.0601	0.0077
Postprandial levels at lipids oral load day: 2-Oxoisocaproic acid_2 (a.u)	-0.0029	0.0036	-0.0027	0.0064	-0.0024	0.0067	-0.007	0.0113	-0.0057	0.0043	-0.0085	0.0065
Basal levels at glucose proteins oral load day: 2-Oxoisocaproic acid_2 (a.u)	0.0194	0.016	0.0449	0.0139	0.0368	0.0194	0.0596	0.0104	0.0653	0.0138	0.0595	0.0109
Postprandial levels at glucose oral load day: 2-Oxoisocaproic acid_2 (a.u)	-0.0021	0.0052	-0.0044	0.0105	-0.0053	0.008	-0.0036	0.0093	-0.0014	0.0099	-0.0012	0.0112
Basal levels at lipids oral load day: 2-Oxoisocaproic acid_2 (a.u)	0.0203	0.0216	0.042	0.012	0.036	0.0136	0.0629	0.0109	0.0632	0.0101	0.0596	0.0089
Postprandial levels at proteins oral load day: 2-Oxoisocaproic acid_2 (a.u)	-0.0035	0.0058	0.0136	0.0129	-0.0004	0.0066	0.0068	0.0153	0.0098	0.016	0.0091	0.0089
Basal levels at glucose oral load day: Leucine_1 (a.u)	0.1194	0.022	0.1422	0.0606	0.142	0.0273	0.1569	0.0177	0.1885	0.0267	0.1729	0.028
Postprandial levels at lipids oral load day: Leucine_1 (a.u)	-0.0205	0.0124	-0.0305	0.0217	-0.0311	0.0111	-0.0191	0.019	-0.0366	0.0116	-0.0264	0.0135
Basal levels at glucose proteins oral load day: Leucine_1 (a.u)	0.1199	0.0378	0.1305	0.0271	0.1358	0.0234	0.1776	0.0162	0.208	0.024	0.1922	0.0241
Postprandial levels at glucose oral load day: Leucine_1 (a.u)	-0.0082	0.022	-0.0076	0.0157	-0.0107	0.0132	-0.0062	0.0058	-0.0226	0.0156	-0.0214	0.0172
Basal levels at lipids oral load day: Leucine_1 (a.u)	0.1255	0.0281	0.1289	0.0299	0.1354	0.0266	0.1766	0.026	0.1983	0.0282	0.1985	0.0397
Postprandial levels at proteins oral load day: Leucine_1 (a.u)	0.1195	0.0381	0.1554	0.0428	0.0993	0.0618	0.1415	0.0573	0.1859	0.0738	0.1205	0.0331
Basal levels at glucose oral load day: Isoleucine_1 (a.u)	0.0344	0.0071	0.0406	0.0228	0.033	0.0093	0.052	0.0074	0.0514	0.008	0.0446	0.0096
Postprandial levels at lipids oral load day: Isoleucine_1 (a.u)	-0.0088	0.0037	-0.0124	0.0086	-0.0073	0.0044	-0.0094	0.0096	-0.0128	0.0058	-0.0089	0.0076
Basal levels at glucose proteins oral load day: Isoleucine_1 (a.u)	0.0374	0.0095	0.0407	0.0116	0.031	0.0068	0.0542	0.01	0.0538	0.01	0.0533	0.0111
Postprandial levels at glucose oral load day: Isoleucine_1 (a.u)	-0.0049	0.0046	-0.0055	0.0082	-0.0005	0.0049	-0.0021	0.0067	-0.0036	0.0041	-0.0063	0.0094
Basal levels at lipids oral load day: Isoleucine_1 (a.u)	0.0357	0.0096	0.0391	0.0077	0.0367	0.0084	0.0603	0.0081	0.055	0.0125	0.0577	0.0119
Postprandial levels at proteins oral load day: Isoleucine_1 (a.u)	0.0432	0.0097	0.0562	0.0121	0.0381	0.0176	0.0471	0.0261	0.0687	0.0216	0.0453	0.013
Basal levels at glucose oral load day: L-Valine_1 (a.u)	0.1095	0.0438	0.1424	0.0484	0.1286	0.0349	0.1893	0.0304	0.1998	0.0317	0.1712	0.0119
Postprandial levels at lipids oral load day: L-Valine_1 (a.u)	-0.022	0.0117	-0.0212	0.0132	-0.019	0.0153	-0.0213	0.0219	-0.0306	0.0204	-0.0172	0.0127
Basal levels at glucose proteins oral load day: L-Valine_1 (a.u)	0.0957	0.03	0.1406	0.0347	0.1199	0.0359	0.1927	0.0213	0.1979	0.0277	0.1879	0.0331
Postprandial levels at glucose oral load day: L-Valine_1 (a.u)	-0.0073	0.0049	-0.0066	0.013	-0.0086	0.0078	-0.0044	0.0109	-0.0127	0.0152	-0.0188	0.0222
Basal levels at lipids oral load day: L-Valine_1 (a.u)	0.1136	0.0351	0.1398	0.033	0.1215	0.0332	0.1927	0.0232	0.1962	0.0224	0.1896	0.0312
Postprandial levels at proteins oral load day: L-Valine_1 (a.u)	0.0898	0.0261	0.1155	0.0259	0.0693	0.0282	0.0999	0.0452	0.1315	0.0381	0.0856	0.0211
Basal levels at glucose oral load day: Isobutyric acid_1 (a.u)	0.0285	0.0134	0.0179	0.0058	0.024	0.0104	0.0185	0.0088	0.022	0.0074	0.0222	0.006
Postprandial levels at lipids oral load day: Isobutyric acid_1 (a.u)	-0.0057	0.0097	-0.0015	0.0043	-0.0012	0.0052	-0.0015	0.0032	-0.0009	0.0058	-0.0012	0.0048
Basal levels at glucose proteins oral load day: Isobutyric acid_1 (a.u)	0.021	0.008	0.0162	0.0049	0.0207	0.0092	0.0195	0.0066	0.0212	0.0039	0.0216	0.0075
Postprandial levels at glucose oral load day: Isobutyric acid_1 (a.u)	-0.0023	0.0069	-0.0007	0.0046	-0.0041	0.0089	-0.0044	0.0042	-0.0032	0.0024	-0.0035	0.0062
Basal levels at lipids oral load day: Isobutyric acid_1 (a.u)	0.0282	0.0143	0.0183	0.0031	0.0227	0.008	0.0195	0.0048	0.0206	0.0068	0.0195	0.0101
Postprandial levels at proteins oral load day: Isobutyric acid_1 (a.u)	0.0069	0.0117	0.0063	0.0042	0.0067	0.011	0.0064	0.0069	0.0105	0.0077	0.0094	0.0051
Basal levels at glucose oral load day: 2-oxoisovaleric acid_1 (a.u)	0.0077	0.0093	0.0162	0.0051	0.0115	0.0066	0.0219	0.0042	0.0236	0.0044	0.021	0.0034
Postprandial levels at lipids oral load day: 2-oxoisovaleric acid_1 (a.u)	-0.0013	0.0033	0.0006	0.0031	0.0006	0.0035	0.0014	0.0037	-0.0019	0.0035	0.0005	0.0038
Basal levels at glucose proteins oral load day: 2-oxoisovaleric acid_1 (a.u)	0.0041	0.0059	0.0172	0.0051	0.0126	0.0082	0.0258	0.0046	0.0207	0.005	0.0208	0.0033
Postprandial levels at glucose oral load day: 2-oxoisovaleric acid_1 (a.u)	0.0011	0.001	0.0003	0.0035	-0.0026	0.0053	-0.0009	0.0035	0.0012	0.0049	0.0002	0.0023
Basal levels at lipids oral load day: 2-oxoisovaleric acid_1 (a.u)	0.0084	0.008	0.0171	0.0045	0.0121	0.0074	0.0243	0.0058	0.0192	0.0043	0.0204	0.0041
Postprandial levels at proteins oral load day: 2-oxoisovaleric acid_1 (a.u)	-0.0015	0.0034	-0.0001	0.0024	-0.0025	0.0041	-0.0011	0.0058	0.0013	0.0033	-0.0002	0.0027
Basal levels at glucose oral load day: 3-Hydroxybutyric acid_1 (a.u)	0.0578	0.0448	0.0444	0.0208	0.0494	0.0239	0.0459	0.0181	0.0775	0.0456	0.0527	0.0465

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Postprandial levels at lipids oral load day: 3-Hydroxybutyric acid_1 (a.u)	-0.0336	0.0329	-0.0215	0.0116	-0.0241	0.0172	-0.0167	0.0128	-0.0418	0.0337	-0.0245	0.0338
Basal levels at glucose proteins oral load day: 3-Hydroxybutyric acid_1 (a.u)	0.0454	0.0205	0.0437	0.0294	0.0292	0.0118	0.0669	0.051	0.1278	0.1949	0.0446	0.0209
Postprandial levels at glucose oral load day: 3-Hydroxybutyric acid_1 (a.u)	0.0756	0.0693	0.0574	0.0402	0.0768	0.0661	0.0616	0.0466	0.0227	0.0581	0.026	0.0301
Basal levels at lipids oral load day: 3-Hydroxybutyric acid_1 (a.u)	0.0608	0.0543	0.0583	0.0333	0.0455	0.0414	0.0675	0.0194	0.0956	0.0541	0.0644	0.0573
Postprandial levels at proteins oral load day: 3-Hydroxybutyric acid_1 (a.u)	-0.0122	0.0314	-0.0159	0.0147	-0.0035	0.0257	-0.0148	0.0133	-0.0213	0.0358	-0.0154	0.0312
Basal levels at glucose oral load day: Lactate_1 (a.u)	1.1376	0.4693	1.1283	0.4463	1.0027	0.311	1.259	0.4811	0.8826	0.2112	1.0779	0.2999
Postprandial levels at lipids oral load day: Lactate_1 (a.u)	0.2059	0.2705	0.2565	0.2802	0.2856	0.1823	0.2964	0.2377	0.2752	0.2112	0.172	0.1334
Basal levels at glucose proteins oral load day: Lactate_1 (a.u)	1.1333	0.6387	1.0738	0.6222	0.9956	0.201	1.4677	0.6231	0.9666	0.4722	1.0979	0.4172
Postprandial levels at glucose oral load day: Lactate_1 (a.u)	-0.2396	0.4802	-0.1407	0.326	-0.0544	0.2198	-0.2554	0.2542	-0.0916	0.3599	-0.2321	0.1755
Basal levels at lipids oral load day: Lactate_1 (a.u)	0.939	0.1639	1.2016	0.4422	0.9355	0.3125	1.2942	0.474	0.8248	0.1733	1.2094	0.4606
Postprandial levels at proteins oral load day: Lactate_1 (a.u)	0.1209	0.234	-0.0425	0.2126	0.0115	0.2083	-0.0511	0.3269	0.0645	0.1087	-0.1341	0.2079
Basal levels at glucose oral load day: Alanine_1 (a.u)	0.2441	0.0478	0.2727	0.0672	0.2682	0.0589	0.343	0.0606	0.2386	0.0482	0.2793	0.0663
Postprandial levels at lipids oral load day: Alanine_1 (a.u)	0.0119	0.0278	0.0052	0.0315	0.0041	0.0258	0.0076	0.0368	0.005	0.0167	0.0108	0.0303
Basal levels at glucose proteins oral load day: Alanine_1 (a.u)	0.2497	0.0529	0.2551	0.0636	0.3088	0.0678	0.3528	0.0923	0.2432	0.0689	0.2885	0.0814
Postprandial levels at glucose oral load day: Alanine_1 (a.u)	-0.0324	0.0317	-0.0165	0.0203	-0.0365	0.045	-0.0158	0.0422	-0.0228	0.0299	-0.0281	0.0197
Basal levels at lipids oral load day: Alanine_1 (a.u)	0.2541	0.0805	0.267	0.0713	0.3022	0.0591	0.3618	0.0743	0.2322	0.0613	0.3027	0.0871
Postprandial levels at proteins oral load day: Alanine_1 (a.u)	0.0548	0.0322	0.0833	0.0473	0.0427	0.051	0.0555	0.0772	0.0625	0.0352	0.0572	0.0268
Basal levels at glucose oral load day: Acetate_1 (a.u)	0.3778	0.1204	0.0553	0.0344	0.2869	0.0604	0.1499	0.0231	0.1712	0.0289	0.1749	0.0806
Postprandial levels at lipids oral load day: Acetate_1 (a.u)	-0.0108	0.024	-0.0058	0.0073	0.0162	0.0408	-0.0052	0.0137	-0.0071	0.0222	-0.0019	0.0103
Basal levels at glucose proteins oral load day: Acetate_1 (a.u)	0.3548	0.0936	0.0606	0.0429	0.3107	0.115	0.1519	0.0156	0.1704	0.0403	0.1717	0.0744
Postprandial levels at glucose oral load day: Acetate_1 (a.u)	-0.0151	0.0747	-0.0025	0.0044	0.0017	0.0286	-0.0031	0.0068	-0.0121	0.0102	0.004	0.0139
Basal levels at lipids oral load day: Acetate_1 (a.u)	0.3629	0.1047	0.07	0.0342	0.3219	0.1326	0.1575	0.022	0.1628	0.033	0.1848	0.0817
Postprandial levels at proteins oral load day: Acetate_1 (a.u)	0.0148	0.0267	-0.018	0.006	-0.0242	0.0568	-0.0187	0.0123	-0.0142	0.0191	0.0038	0.033
Basal levels at glucose oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	0.0562	0.0111	0.0385	0.0088	0.0541	0.0083	0.0603	0.0237	0.0564	0.0147	0.0482	0.0091
Postprandial levels at lipids oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	-0.0023	0.0041	-0.0009	0.0037	-0.0006	0.0048	0.0045	0.0108	-0.0018	0.003	0.0012	0.0027
Basal levels at glucose proteins oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	0.0583	0.0119	0.0413	0.007	0.0552	0.0048	0.066	0.0111	0.0583	0.0129	0.0525	0.0116
Postprandial levels at glucose oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	-0.0053	0.0062	0.0004	0.0064	-0.0025	0.0021	-0.0013	0.0047	-0.0017	0.0041	-0.0028	0.0045
Basal levels at lipids oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	0.0584	0.0109	0.0437	0.0106	0.0567	0.0095	0.0676	0.0133	0.057	0.0123	0.0533	0.011
Postprandial levels at proteins oral load day: N-Ac-Glycoprotein_signals_1 (a.u)	-0.0027	0.0037	-0.0021	0.0019	-0.0056	0.0056	-0.0026	0.0051	-0.0028	0.0025	-0.0027	0.0033
Basal levels at glucose oral load day: Acetone_1 (a.u)	0.0115	0.0087	0.0274	0.0053	0.0122	0.0029	0.0275	0.005	0.0202	0.0044	0.0214	0.0047
Postprandial levels at lipids oral load day: Acetone_1 (a.u)	-0.0012	0.0046	-0.0051	0.0027	0.0002	0.0015	0.0001	0.0044	-0.0016	0.0031	-0.0023	0.0051
Basal levels at glucose proteins oral load day: Acetone_1 (a.u)	0.0103	0.002	0.0199	0.0016	0.0124	0.002	0.0334	0.0113	0.0167	0.004	0.0223	0.0067
Postprandial levels at glucose oral load day: Acetone_1 (a.u)	0.0004	0.0051	0.0043	0.0044	-0.0016	0.0026	0.0042	0.0055	0.0062	0.0043	0.0037	0.0051
Basal levels at lipids oral load day: Acetone_1 (a.u)	0.0096	0.0032	0.0216	0.0054	0.0108	0.0029	0.0315	0.0081	0.0193	0.0049	0.0229	0.0073
Postprandial levels at proteins oral load day: Acetone_1 (a.u)	8.00E-04	3.00E-03	0.0016	0.0032	0.0013	0.0029	0.0017	0.0066	0.0042	0.0065	0.0012	0.006
Basal levels at glucose oral load day: L-Glutamic acid_1 (a.u)	0.0486	0.0061	0.0477	0.0196	0.0461	0.0083	0.0446	0.01	0.0446	0.0108	0.0498	0.0076
Postprandial levels at lipids oral load day: L-Glutamic acid_1 (a.u)	-0.0034	0.0057	-0.0052	0.0044	-0.0034	0.004	0.0006	0.0054	-0.0019	0.0046	0.0006	0.0043

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Basal levels at glucose proteins oral load day: L-Glutamic acid_1 (a.u)	0.0428	0.0043	0.041	0.0147	0.0439	0.0115	0.0551	0.0042	0.0462	0.0067	0.0534	0.0126
Postprandial levels at glucose oral load day: L-Glutamic acid_1 (a.u)	-0.0032	0.0038	-0.0013	0.0034	-0.0046	0.0052	-0.0069	0.0042	-0.0027	0.0043	-0.0054	0.0065
Basal levels at lipids oral load day: L-Glutamic acid_1 (a.u)	0.0447	0.0064	0.042	0.017	0.0456	0.0088	0.0521	0.0106	0.0461	0.0089	0.0543	0.01
Postprandial levels at proteins oral load day: L-Glutamic acid_1 (a.u)	0.0172	0.0072	0.0189	0.0056	0.0105	0.0057	0.016	0.014	0.0201	0.0079	0.014	0.0076
Basal levels at glucose oral load day: Pyruvate_1 (a.u)	0.0061	0.0023	0.025	0.0129	0.0118	0.006	0.0249	0.0177	0.0147	0.0083	0.0215	0.008
Postprandial levels at lipids oral load day: Pyruvate_1 (a.u)	0.0033	0.0092	0.0068	0.0076	0.0011	0.0042	0.0162	0.0121	0.0094	0.0056	0.0067	0.0045
Basal levels at glucose proteins oral load day: Pyruvate_1 (a.u)	0.0065	0.0043	0.0182	0.0107	0.0135	0.0068	0.0367	0.0198	0.0157	0.0098	0.015	0.0078
Postprandial levels at glucose oral load day: Pyruvate_1 (a.u)	-0.0008	0.0029	-0.0008	0.0075	-0.0028	0.0038	-0.0045	0.0052	-0.0001	0.0072	-0.0001	0.0041
Basal levels at lipids oral load day: Pyruvate_1 (a.u)	0.0076	0.0055	0.0196	0.0125	0.0121	0.0062	0.0298	0.0137	0.0109	0.0052	0.0172	0.0069
Postprandial levels at proteins oral load day: Pyruvate_1 (a.u)	0.0064	0.0028	0.0099	0.0063	0.0027	0.004	0.008	0.0131	0.0102	0.0059	0.0083	0.0058
Basal levels at glucose oral load day: Pyroglutamic acid_1 (a.u)	0.0348	0.0094	0.0139	0.0042	0.0241	0.0161	0.0128	0.0048	0.0197	0.0089	0.0142	0.0084
Postprandial levels at lipids oral load day: Pyroglutamic acid_1 (a.u)	-0.0048	0.0065	-0.0046	0.0024	-0.0042	0.0066	-0.0027	0.0031	-0.007	0.0065	-0.0042	0.0046
Basal levels at glucose proteins oral load day: Pyroglutamic acid_1 (a.u)	0.0354	0.009	0.0141	0.0061	0.0199	0.0094	0.015	0.0082	0.0324	0.0326	0.0139	0.0045
Postprandial levels at glucose oral load day: Pyroglutamic acid_1 (a.u)	0.0117	0.0144	0.0089	0.0092	0.0129	0.0119	0.0108	0.0076	0.0003	0.0102	0.0044	0.0069
Basal levels at lipids oral load day: Pyroglutamic acid_1 (a.u)	0.0351	0.0157	0.0191	0.008	0.025	0.0192	0.0166	0.0046	0.0242	0.0081	0.018	0.0104
Postprandial levels at proteins oral load day: Pyroglutamic acid_1 (a.u)	-0.0001	0.0053	-0.0061	0.006	-0.0012	0.0043	-0.004	0.0042	-0.0051	0.0063	-0.0017	0.0066
Basal levels at glucose oral load day: L-Glutamine_1 (a.u)	0.122	0.0249	0.128	0.019	0.1526	0.0196	0.1617	0.0238	0.151	0.0155	0.144	0.023
Postprandial levels at lipids oral load day: L-Glutamine_1 (a.u)	-0.0012	0.0159	-0.0095	0.0093	-0.0057	0.0106	0.0019	0.0197	-0.0063	0.0125	-0.0044	0.0099
Basal levels at glucose proteins oral load day: L-Glutamine_1 (a.u)	0.1207	0.0283	0.1277	0.0201	0.1482	0.0208	0.1704	0.0191	0.1591	0.0141	0.1558	0.0245
Postprandial levels at glucose oral load day: L-Glutamine_1 (a.u)	0.0017	0.0097	0.0102	0.0097	0.0044	0.0129	0.0093	0.0115	0.0017	0.0117	-0.0024	0.0084
Basal levels at lipids oral load day: L-Glutamine_1 (a.u)	0.1253	0.0339	0.1431	0.0398	0.1535	0.0171	0.1728	0.0133	0.1604	0.0136	0.1549	0.0285
Postprandial levels at proteins oral load day: L-Glutamine_1 (a.u)	0.0149	0.0114	0.0159	0.017	0.0171	0.0172	0.0162	0.0181	0.0194	0.0129	0.0125	0.0054
Basal levels at glucose oral load day: Citrate_2 (a.u)	0.0257	0.0062	0.0195	0.008	0.0249	0.0076	0.0265	0.0047	0.0246	0.0065	0.0195	0.0041
Postprandial levels at lipids oral load day: Citrate_2 (a.u)	0.0047	0.0056	0.0004	0.0044	0.0041	0.0029	0.0022	0.0068	0.0022	0.0044	0.0025	0.0038
Basal levels at glucose proteins oral load day: Citrate_2 (a.u)	0.0247	0.0079	0.0189	0.0055	0.0249	0.0064	0.0262	0.006	0.0234	0.0045	0.0187	0.0032
Postprandial levels at glucose oral load day: Citrate_2 (a.u)	0.0037	0.0051	0.0035	0.0044	0.0057	0.0067	0.0064	0.0053	0.0034	0.0049	0.0006	0.0024
Basal levels at lipids oral load day: Citrate_2 (a.u)	0.0242	0.0056	0.0224	0.0086	0.0258	0.0069	0.0276	0.0073	0.0242	0.0046	0.02	0.0043
Postprandial levels at proteins oral load day: Citrate_2 (a.u)	0.0008	0.0041	-0.0014	0.0037	0.0006	0.0044	-2.00E-04	7.00E-03	-0.0008	0.0032	-0.0001	0.0024
Basal levels at glucose oral load day: L-Asparagine_4 (a.u)	0.007	0.0013	0.0059	0.0017	0.0078	0.0017	0.0064	0.003	0.0071	0.002	0.0059	0.0022
Postprandial levels at lipids oral load day: L-Asparagine_4 (a.u)	0.00E+00	9.00E-04	-0.0002	0.0011	-0.0009	0.0015	0.0007	0.0023	0	0.0012	0.0005	0.0014
Basal levels at glucose proteins oral load day: L-Asparagine_4 (a.u)	0.0072	0.0021	0.0072	0.0022	0.0085	0.0025	0.0074	0.0017	0.0086	0.002	0.0076	0.0012
Postprandial levels at glucose oral load day: L-Asparagine_4 (a.u)	0.0002	0.0022	-0.0008	0.0015	-2.00E-04	2.00E-03	0.0003	0.0017	-0.001	0.0015	-0.0006	0.0018
Basal levels at lipids oral load day: L-Asparagine_4 (a.u)	7.00E-03	9.00E-04	0.0056	0.0022	0.008	0.0027	0.008	0.0031	0.0081	0.0014	0.0072	0.0012
Postprandial levels at proteins oral load day: L-Asparagine_4 (a.u)	0.0023	0.0013	0.0034	0.0009	0.002	0.003	0.0027	0.0027	0.0026	0.0018	0.002	0.0012
Basal levels at glucose oral load day: Creatinine_1 (a.u)	0.0467	0.0062	0.0385	0.0066	0.0427	0.0103	0.0735	0.0749	0.0626	0.0077	0.0496	0.0096
Postprandial levels at lipids oral load day: Creatinine_1 (a.u)	-0.0033	0.0046	-0.0018	0.0066	0.0028	0.0039	-0.0158	0.0518	-0.0027	0.0075	0.0004	0.0073
Basal levels at glucose proteins oral load day: Creatinine_1 (a.u)	0.0445	0.0072	0.0425	0.0085	0.0468	0.0103	0.0523	0.0067	0.061	0.0098	0.0515	0.0083

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Postprandial levels at glucose oral load day: Creatinine_1 (a.u)	-0.0018	0.0071	-0.0019	0.0037	-0.004	0.0046	-0.0019	0.0049	-0.0009	0.0089	-0.0027	0.0076
Basal levels at lipids oral load day: Creatinine_1 (a.u)	0.0476	0.0102	0.039	0.0096	0.0484	0.0128	0.0719	0.0456	0.0644	0.0066	0.0555	0.0133
Postprandial levels at proteins oral load day: Creatinine_1 (a.u)	-0.0014	0.0053	0.0021	0.0059	-0.0035	0.0101	-0.0173	0.0361	-0.0044	0.0053	-0.0011	0.0112
Basal levels at glucose oral load day: Creatine_1 (a.u)	0.0311	0.0131	0.0345	0.0172	0.021	0.0115	0.0362	0.0136	0.0246	0.0099	0.0264	0.0104
Postprandial levels at lipids oral load day: Creatine_1 (a.u)	-0.0025	0.005	-0.0014	0.0073	0.0029	0.0054	0.0041	0.01	-0.0006	0.0041	0.0003	0.0044
Basal levels at glucose proteins oral load day: Creatine_1 (a.u)	0.0265	0.012	0.0306	0.0091	0.0222	0.0104	0.0394	0.0079	0.0178	0.0079	0.0259	0.0142
Postprandial levels at glucose oral load day: Creatine_1 (a.u)	-0.0046	0.0061	-0.0015	0.0035	-0.0018	0.0035	-0.0044	0.0058	0.0016	0.0038	-0.0006	0.0056
Basal levels at lipids oral load day: Creatine_1 (a.u)	0.0298	0.0112	0.0321	0.0082	0.0278	0.0147	0.0403	0.0157	0.022	0.0086	0.0313	0.017
Postprandial levels at proteins oral load day: Creatine_1 (a.u)	0.009	0.0079	0.0147	0.0088	0.0072	0.0056	0.0137	0.0072	0.0126	0.0061	0.008	0.0066
Basal levels at glucose oral load day: Lysine_1 (a.u)	0.017	0.0199	0.0545	0.0161	0.033	0.0149	0.06	0.015	0.0571	0.0118	0.0463	0.007
Postprandial levels at lipids oral load day: Lysine_1 (a.u)	-0.0014	0.0045	-0.0107	0.0066	-0.0064	0.0062	-0.001	0.014	-0.007	0.0092	0.0011	0.0057
Basal levels at glucose proteins oral load day: Lysine_1 (a.u)	0.0135	0.0161	0.0529	0.0129	0.0308	0.0176	0.0656	0.0086	0.0532	0.0092	0.0516	0.0102
Postprandial levels at glucose oral load day: Lysine_1 (a.u)	0.0019	0.0058	-0.0025	0.0053	0.0005	0.0062	0.0029	0.0046	0.0041	0.0081	-0.0026	0.0061
Basal levels at lipids oral load day: Lysine_1 (a.u)	0.0146	0.0168	0.0493	0.0168	0.0263	0.0176	0.06	0.0106	0.052	0.0103	0.0519	0.0159
Postprandial levels at proteins oral load day: Lysine_1 (a.u)	0.0185	0.0131	0.0466	0.0141	0.027	0.0126	0.0483	0.016	0.0516	0.0177	0.0338	0.0091
Basal levels at glucose oral load day: Ornithine_1 (a.u)	0.0112	0.0074	0.0186	0.0052	0.0176	0.0058	0.0216	0.0067	0.0208	0.0073	0.0198	0.005
Postprandial levels at lipids oral load day: Ornithine_1 (a.u)	-0.0012	0.004	-0.0017	0.0038	-0.0016	0.0028	-0.0007	0.0074	-0.0026	0.0053	0.0007	0.0025
Basal levels at glucose proteins oral load day: Ornithine_1 (a.u)	0.0124	0.0043	0.02	0.0047	0.0145	0.0088	0.0239	0.0071	0.0214	0.0076	0.0226	0.0059
Postprandial levels at glucose oral load day: Ornithine_1 (a.u)	-0.0004	0.0038	-0.0008	0.0028	6.00E-04	5.00E-03	-3.00E-04	4.00E-03	0.0005	0.0053	-0.003	0.0035
Basal levels at lipids oral load day: Ornithine_1 (a.u)	0.0111	0.0053	0.0194	0.0064	0.0156	0.0076	0.024	0.0051	0.0223	0.0054	0.0249	0.0109
Postprandial levels at proteins oral load day: Ornithine_1 (a.u)	0.001	0.0043	0.0042	0.0043	0.0019	0.0043	0.002	0.004	0.0048	0.0043	0.0007	0.0077
Basal levels at glucose oral load day: Choline_1 (a.u)	0.0316	0.006	0.0241	0.0069	0.0313	0.0129	0.0282	0.0054	0.0279	0.0141	0.0266	0.0109
Postprandial levels at lipids oral load day: Choline_1 (a.u)	0	0.0039	0.0001	0.0048	-0.0009	0.0064	0.0062	0.0086	0.0016	0.0088	0.0011	0.0115
Basal levels at glucose proteins oral load day: Choline_1 (a.u)	0.0286	0.0102	0.0258	0.0094	0.029	0.0095	0.0321	0.0072	0.0297	0.0126	0.0212	0.01
Postprandial levels at glucose oral load day: Choline_1 (a.u)	0.0019	0.0088	0.0024	0.0065	0.0025	0.0084	0.0023	0.0091	-0.0022	0.0068	0.0065	0.0125
Basal levels at lipids oral load day: Choline_1 (a.u)	0.027	0.011	0.0297	0.009	0.0316	0.011	0.035	0.01	0.0328	0.0143	0.029	0.0116
Postprandial levels at proteins oral load day: Choline_1 (a.u)	0.0019	0.0078	-0.0028	0.0102	4.00E-04	6.00E-03	-0.0011	0.0091	-0.0028	0.0138	-0.0016	0.0074
Basal levels at glucose oral load day: L-Carnitine_1 (a.u)	0.0962	0.0228	0.0582	0.0117	0.0772	0.0282	0.0813	0.0157	0.0991	0.0125	0.0843	0.0243
Postprandial levels at lipids oral load day: L-Carnitine_1 (a.u)	-0.0039	0.0098	-0.0001	0.0087	0.0064	0.0135	0.0055	0.0141	0.0039	0.0114	-0.0046	0.0101
Basal levels at glucose proteins oral load day: L-Carnitine_1 (a.u)	0.0881	0.022	0.0622	0.0128	0.0711	0.0229	0.0823	0.014	0.0907	0.0141	0.0748	0.0256
Postprandial levels at glucose oral load day: L-Carnitine_1 (a.u)	-0.0057	0.0168	0.006	0.005	0.0074	0.0168	0.0034	0.0107	0.0026	0.012	0.007	0.0207
Basal levels at lipids oral load day: L-Carnitine_1 (a.u)	0.096	0.0253	0.0641	0.0137	0.0853	0.0321	0.09	0.0156	0.0961	0.0149	0.0841	0.0204
Postprandial levels at proteins oral load day: L-Carnitine_1 (a.u)	-0.0022	0.0179	0.0025	0.0061	-0.0081	0.0208	0.0038	0.0113	-0.0006	0.0112	-0.0034	0.0131
Basal levels at glucose oral load day: Betaine_1 (a.u)	0.087	0.0289	0.0573	0.0096	0.0753	0.0291	0.0662	0.0281	0.0777	0.0209	0.0573	0.0167
Postprandial levels at lipids oral load day: Betaine_1 (a.u)	0.0017	0.0089	-0.0021	0.0113	0.0024	0.014	0.0043	0.0133	-0.0058	0.011	-0.0046	0.008
Basal levels at glucose proteins oral load day: Betaine_1 (a.u)	0.0796	0.0283	0.0585	0.0144	0.0849	0.0355	0.0711	0.0278	0.0715	0.016	0.0536	0.0188
Postprandial levels at glucose oral load day: Betaine_1 (a.u)	0.0007	0.0139	0.003	0.0067	-0.0083	0.0131	-0.0018	0.0153	-0.0002	0.0107	0.0014	0.01

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Basal levels at lipids oral load day: Betaine_1 (a.u)	0.0767	0.0323	0.0553	0.0203	0.0887	0.0285	0.0653	0.0248	0.0828	0.0109	0.078	0.0441
Postprandial levels at proteins oral load day: Betaine_1 (a.u)	0.0064	0.0128	0.0121	0.0185	-0.0039	0.0084	0.0001	0.0094	0.007	0.0066	-0.0008	0.0124
Basal levels at glucose oral load day: Glycine_1 (a.u)	0.0738	0.0203	0.0831	0.0186	0.1045	0.0246	0.0921	0.0194	0.0967	0.0141	0.089	0.0281
Postprandial levels at lipids oral load day: Glycine_1 (a.u)	-0.0069	0.0081	-0.007	0.0068	-0.0036	0.0032	-0.0021	0.0142	-0.0068	0.0077	-0.0025	0.011
Basal levels at glucose proteins oral load day: Glycine_1 (a.u)	0.0786	0.0261	0.0821	0.0245	0.1039	0.0253	0.0928	0.0196	0.1003	0.0142	0.0956	0.0249
Postprandial levels at glucose oral load day: Glycine_1 (a.u)	-0.0082	0.0089	-0.002	0.0096	-0.0007	0.0174	-0.0051	0.0079	-0.0013	0.0104	-0.0065	0.0067
Basal levels at lipids oral load day: Glycine_1 (a.u)	0.0819	0.0301	0.0859	0.0303	0.1099	0.0175	0.0958	0.0291	0.094	0.0101	0.0974	0.0283
Postprandial levels at proteins oral load day: Glycine_1 (a.u)	0.0015	0.004	0.0029	0.0066	0.0028	0.0116	0.0042	0.0187	0.0049	0.0074	0.0005	0.0038
Basal levels at glucose oral load day: L-Threonine_1 (a.u)	0.0206	0.011	0.0238	0.0159	0.0253	0.0123	0.0292	0.0062	0.0245	0.0096	0.0206	0.0079
Postprandial levels at lipids oral load day: L-Threonine_1 (a.u)	-0.0045	0.0077	-0.0024	0.0039	-0.003	0.0062	-0.0031	0.0093	-0.005	0.0041	-0.0031	0.0039
Basal levels at glucose proteins oral load day: L-Threonine_1 (a.u)	0.0187	0.0093	0.0238	0.0109	0.0245	0.0128	0.03	0.0083	0.0231	0.0073	0.018	0.0068
Postprandial levels at glucose oral load day: L-Threonine_1 (a.u)	-0.0031	0.0051	-0.0014	0.0064	-0.0024	0.0047	-0.0012	0.005	-0.0005	0.0052	0.0026	0.0064
Basal levels at lipids oral load day: L-Threonine_1 (a.u)	0.0216	0.0093	0.0225	0.0083	0.022	0.0116	0.0316	0.0061	0.0234	0.0076	0.0239	0.0111
Postprandial levels at proteins oral load day: L-Threonine_1 (a.u)	0.011	0.0075	0.0139	0.0068	0.0127	0.0092	0.0114	0.0113	0.0147	0.0058	0.0096	0.0051
Basal levels at glucose oral load day: Glycerol_2 (a.u)	0.0134	0.0146	0.0233	0.0055	0.0238	0.0134	0.0313	0.0105	0.016	0.0061	0.018	0.0068
Postprandial levels at lipids oral load day: Glycerol_2 (a.u)	-0.0042	0.0091	-0.0069	0.0051	-0.0074	0.0063	-0.0103	0.0072	-0.0019	0.0067	-0.0051	0.005
Basal levels at glucose proteins oral load day: Glycerol_2 (a.u)	0.0154	0.012	0.028	0.0135	0.0219	0.0073	0.0291	0.0084	0.0202	0.0099	0.0181	0.0087
Postprandial levels at glucose oral load day: Glycerol_2 (a.u)	-0.0003	0.0084	-0.0008	0.0072	-0.0013	0.0046	0.0006	0.0079	-0.0035	0.0063	-0.0011	0.0065
Basal levels at lipids oral load day: Glycerol_2 (a.u)	0.0142	0.0086	0.0303	0.0115	0.0197	0.011	0.0305	0.0155	0.02	0.0035	0.0205	0.0096
Postprandial levels at proteins oral load day: Glycerol_2 (a.u)	-0.0047	0.0071	-0.0083	0.0079	-0.0052	0.0082	-0.0043	0.0123	-0.0073	0.0035	-0.0054	0.0055
Basal levels at glucose oral load day: L-Serine_1 (a.u)	0.0281	0.0042	0.0213	0.0032	0.0326	0.0086	0.0225	0.006	0.0214	0.0073	0.0232	0.0066
Postprandial levels at lipids oral load day: L-Serine_1 (a.u)	-0.003	0.0054	-0.0023	0.0041	-0.006	0.0075	-0.0004	0.0038	-0.0015	0.0063	-0.0038	0.0051
Basal levels at glucose proteins oral load day: L-Serine_1 (a.u)	0.0269	0.0062	0.0232	0.0076	0.0256	0.0069	0.0214	0.0092	0.0221	0.0056	0.0227	0.0076
Postprandial levels at glucose oral load day: L-Serine_1 (a.u)	0.0009	0.0031	-0.0006	0.0064	-0.0008	0.0083	0.0032	0.0092	-0.0009	0.0067	-0.001	0.0061
Basal levels at lipids oral load day: L-Serine_1 (a.u)	0.0249	0.0083	0.0271	0.0187	0.0297	0.0096	0.025	0.0097	0.0238	0.0074	0.0219	0.0062
Postprandial levels at proteins oral load day: L-Serine_1 (a.u)	0.0138	0.0087	0.0084	0.0092	0.0057	0.0091	0.0061	0.0081	0.0086	0.0043	0.006	0.0036
Basal levels at glucose oral load day: L-Proline_3 (a.u)	0.0245	0.0066	0.0224	0.0134	0.0252	0.0053	0.0264	0.005	0.0246	0.0071	0.0248	0.0061
Postprandial levels at lipids oral load day: L-Proline_3 (a.u)	-0.0028	0.0038	-0.0017	0.0038	-0.0052	0.0035	-0.0041	0.0041	-0.0041	0.0047	-0.0023	0.0021
Basal levels at glucose proteins oral load day: L-Proline_3 (a.u)	0.0232	0.0071	0.0217	0.0108	0.0206	0.0069	0.0299	0.0028	0.0272	0.0085	0.0201	0.0044
Postprandial levels at glucose oral load day: L-Proline_3 (a.u)	0.0053	0.007	0.0051	0.0047	0.0063	0.0067	0.0038	0.0037	0.0016	0.0058	0.0031	0.0043
Basal levels at lipids oral load day: L-Proline_3 (a.u)	0.026	0.0053	0.0237	0.0082	0.0241	0.0086	0.0279	0.0036	0.0271	0.0035	0.0261	0.0077
Postprandial levels at proteins oral load day: L-Proline_3 (a.u)	0.0093	0.0063	0.0106	0.0035	0.0082	0.0062	0.0106	0.0062	0.0138	0.0071	0.0094	0.0078
Basal levels at glucose oral load day: beta-Glucose_1 (a.u)	0.1843	0.0431	0.237	0.0432	0.2215	0.049	0.2697	0.0393	0.2791	0.0386	0.2673	0.0273
Postprandial levels at lipids oral load day: beta-Glucose_1 (a.u)	0.0563	0.0646	0.0836	0.0555	0.0331	0.071	0.1724	0.0757	0.1	0.0777	0.1456	0.0703
Basal levels at glucose proteins oral load day: beta-Glucose_1 (a.u)	0.1638	0.0412	0.2162	0.0451	0.2143	0.0553	0.2961	0.0215	0.2676	0.017	0.2624	0.0363
Postprandial levels at glucose oral load day: beta-Glucose_1 (a.u)	-0.0072	0.014	-0.0026	0.0172	-0.0125	0.0208	-0.0199	0.0103	-0.0088	0.0117	-0.0093	0.0214
Basal levels at lipids oral load day: beta-Glucose_1 (a.u)	0.1751	0.0427	0.2334	0.0394	0.225	0.045	0.2962	0.0142	0.2541	0.0248	0.2652	0.0379

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	9		8		9		8		10		9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Postprandial levels at proteins oral load day: beta-Glucose_1 (a.u)	-0.0126	0.0169	0.0002	0.0153	-0.0234	0.0272	0.0035	0.0272	0.0133	0.0162	0.0072	0.0132
Basal levels at glucose oral load day: D-Glucose_2 (a.u)	0.2767	0.0732	0.3368	0.0713	0.3221	0.0794	0.4084	0.0404	0.4046	0.0423	0.3847	0.0544
Postprandial levels at lipids oral load day: D-Glucose_2 (a.u)	0.0785	0.0869	0.1128	0.0717	0.0439	0.0588	0.2101	0.0769	0.1283	0.0856	0.1876	0.0818
Basal levels at glucose proteins oral load day: D-Glucose_2 (a.u)	0.259	0.0797	0.3243	0.0587	0.3062	0.069	0.4311	0.0397	0.4068	0.032	0.3822	0.0682
Postprandial levels at glucose oral load day: D-Glucose_2 (a.u)	-0.0157	0.027	-0.0039	0.0394	-0.0152	0.0533	-0.0108	0.031	-0.014	0.0238	-0.0006	0.0502
Basal levels at lipids oral load day: D-Glucose_2 (a.u)	0.2523	0.0499	0.342	0.062	0.3216	0.0571	0.4141	0.0477	0.3934	0.0396	0.3918	0.0763
Postprandial levels at proteins oral load day: D-Glucose_2 (a.u)	-0.0135	0.0196	0.0023	0.0234	-0.0213	0.0384	0.0266	0.0254	-0.0035	0.0306	0.0059	0.0442
Basal levels at glucose oral load day: 1-Methylhistidine_1 (a.u)	0.005	0.0048	0.0132	0.003	0.0102	0.0056	0.016	0.0032	0.0145	0.003	0.0146	0.0038
Postprandial levels at lipids oral load day: 1-Methylhistidine_1 (a.u)	-1.00E-04	5.00E-04	-0.0014	0.0018	-0.0011	0.0013	-0.0002	0.0034	0.0001	0.0013	-0.0006	0.0023
Basal levels at glucose proteins oral load day: 1-Methylhistidine_1 (a.u)	0.0056	0.0067	0.0139	0.0036	0.009	0.0058	0.0165	0.0033	0.0155	0.003	0.0154	0.0037
Postprandial levels at glucose oral load day: 1-Methylhistidine_1 (a.u)	-0.0013	0.0021	-0.001	0.0025	0.0008	0.0027	-0.0001	0.0017	-0.0003	0.0021	-0.0008	0.0015
Basal levels at lipids oral load day: 1-Methylhistidine_1 (a.u)	0.0052	0.005	0.0132	0.0042	0.0107	0.0068	0.015	0.0037	0.0171	0.0038	0.015	0.0054
Postprandial levels at proteins oral load day: 1-Methylhistidine_1 (a.u)	0.001	0.002	0.0041	0.0011	0.0007	0.0026	0.0053	0.0023	0.0036	0.0025	0.0016	0.0022
Basal levels at glucose oral load day: Tyrosine_1 (a.u)	0.0266	0.0063	0.0339	0.0121	0.0319	0.0086	0.0388	0.0078	0.0322	0.0082	0.0344	0.0065
Postprandial levels at lipids oral load day: Tyrosine_1 (a.u)	-0.0038	0.0038	-0.0063	0.0041	-0.0052	0.0062	-0.004	0.0045	-0.0045	0.0056	-0.0036	0.0042
Basal levels at glucose proteins oral load day: Tyrosine_1 (a.u)	0.0288	0.0106	0.0299	0.0078	0.033	0.0088	0.0399	0.0068	0.0357	0.0089	0.0348	0.0071
Postprandial levels at glucose oral load day: Tyrosine_1 (a.u)	-0.0046	0.0052	-0.001	0.0076	-0.0053	0.0043	-0.0041	0.005	-0.0041	0.0065	-0.0016	0.0041
Basal levels at lipids oral load day: Tyrosine_1 (a.u)	0.0271	0.0047	0.0336	0.0105	0.0301	0.0074	0.0419	0.0054	0.0328	0.0072	0.0379	0.0067
Postprandial levels at proteins oral load day: Tyrosine_1 (a.u)	0.019	0.007	0.0289	0.0085	0.0149	0.0117	0.0207	0.0113	0.0269	0.0124	0.0162	0.0062
Basal levels at glucose oral load day: L-Phenylalanine_2 (a.u)	0.0306	0.0057	0.0291	0.0053	0.0339	0.0064	0.0347	0.0043	0.0362	0.0055	0.0336	0.0051
Postprandial levels at lipids oral load day: L-Phenylalanine_2 (a.u)	-0.0028	0.0026	-0.0029	0.0028	-0.0051	0.004	-0.0001	0.0038	-0.0045	0.0025	-0.0014	0.0029
Basal levels at glucose proteins oral load day: L-Phenylalanine_2 (a.u)	0.03	0.0045	0.0283	0.0058	0.0316	0.0066	0.0367	0.003	0.0344	0.0064	0.0379	0.0065
Postprandial levels at glucose oral load day: L-Phenylalanine_2 (a.u)	-0.0033	0.0033	-0.0016	0.0032	-0.0023	0.0049	-0.0019	0.0033	-0.0011	0.0034	-0.0055	0.0053
Basal levels at lipids oral load day: L-Phenylalanine_2 (a.u)	0.031	0.0044	0.0304	0.0071	0.0309	0.0043	0.0363	0.005	0.0347	0.0052	0.036	0.0077
Postprandial levels at proteins oral load day: L-Phenylalanine_2 (a.u)	0.0083	0.0042	0.0113	0.0027	0.009	0.0062	0.0124	0.0045	0.012	0.0052	0.0073	0.0016
Basal levels at glucose oral load day: L-Tryptophan_4 (a.u)	0.0113	0.0029	0.0136	0.0032	0.0139	0.0034	0.0156	0.0035	0.0162	0.0048	0.0156	0.0026
Postprandial levels at lipids oral load day: L-Tryptophan_4 (a.u)	-5.00E-04	2.00E-03	-0.0005	0.0018	-0.0007	0.0025	0	0.0032	-0.0018	0.0033	-0.0007	0.0027
Basal levels at glucose proteins oral load day: L-Tryptophan_4 (a.u)	0.0135	0.004	0.0164	0.0025	0.0151	0.003	0.0164	0.0023	0.0179	0.0032	0.0181	0.0034
Postprandial levels at glucose oral load day: L-Tryptophan_4 (a.u)	-0.0013	0.0025	-0.0029	0.0007	-0.0028	0.0018	-0.0015	0.0018	-0.0021	0.0027	-0.0033	0.0028
Basal levels at lipids oral load day: L-Tryptophan_4 (a.u)	0.0123	0.0029	0.016	0.0037	0.0152	0.0032	0.0155	0.0028	0.0161	0.0028	0.0166	0.0048
Postprandial levels at proteins oral load day: L-Tryptophan_4 (a.u)	0.0021	0.0026	0.002	0.0024	0.0014	0.0026	0.0036	0.0038	0.0045	0.0029	0.0014	0.0037
Basal levels at glucose oral load day: Formate_1 (a.u)	0.0195	0.0066	0.0124	0.0021	0.0108	0.006	0.0162	0.0043	0.0107	0.002	0.0129	0.002
Postprandial levels at lipids oral load day: Formate_1 (a.u)	-0.0016	0.0018	0.0006	0.0019	0	0.0017	-0.0004	0.0031	0.0002	0.0018	-0.0003	0.0014
Basal levels at glucose proteins oral load day: Formate_1 (a.u)	0.0183	0.0061	0.013	0.0024	0.0102	0.0045	0.0161	0.0021	0.0111	0.0012	0.0131	0.0013
Postprandial levels at glucose oral load day: Formate_1 (a.u)	0	0.0033	0.0001	0.0021	-0.0001	0.0015	-6.00E-04	2.00E-03	-0.001	0.0015	-0.0001	0.0014
Basal levels at lipids oral load day: Formate_1 (a.u)	0.0195	0.0071	0.0118	0.0025	0.0102	0.0051	0.0168	0.0057	0.0097	0.0018	0.0135	0.0022
Postprandial levels at proteins oral load day: Formate_1 (a.u)	0.0023	0.0019	0.0016	0.0015	0.0011	0.0015	-0.0018	0.005	0.0007	0.0022	-0.0001	0.0013

8.7 Descriptive tables of the microbiome

8.7.1 Descriptive table of the pjylum Relative abundance

Variable	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	8		8		7		8		8		7	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Firmicutes	73.4686	19.4854	76.3222	10.55	70.4499	13.6843	75.1677	12.8644	70.7162	12.793	75.6603	12.9987
Bacteroidetes	13.698	14.1474	15.3537	11.2163	21.7426	11.5428	12.8227	10.7925	21.3867	9.2557	13.6272	7.569
Proteobacteria	1.4068	1.6908	0.9125	0.6621	1.3164	1.197	3.1519	6.8972	1.3445	1.5001	4.2092	7.0686
Actinobacteria	6.9028	4.6793	5.5297	3.3605	2.2597	1.2734	6.2568	6.8764	5.7679	8.934	5.7817	3.6947
Verrucomicrobia	3.2983	8.2859	1.4552	2.2555	2.6165	4.364	1.8594	3.7896	0.552	0.8951	0.0603	0.0827
Candidatus.Saccharibacteria	0.0229	0.0167	0.0913	0.101	0.0156	0.0122	0.0759	0.1135	0.0303	0.0227	0.0424	0.0295
Chloroflexi	0	0	0	0	0	0	0	0	1.00E-04	4.00E-04	0	0
Cyanobacteria.Chloroplast	0.0057	0.0074	0.0047	0.0079	0.0313	0.0707	0.0042	0.0048	0.0257	0.0508	0.0045	0.0063
Elusimicrobia	3.00E-04	8.00E-04	0	0	0	0	0	0	0	0	0	0
Euryarchaeota	0.1892	0.3911	0.1687	0.3794	0.4592	0.6478	0.0368	0.0769	0.0028	0.0075	0.4065	0.6099
Fusobacteria	0.0058	0.0101	0.0065	0.0111	7.00E-04	8.00E-04	0.0012	0.0016	0.0039	0.0069	0.0022	0.0028
Lentisphaerae	3.00E-04	8.00E-04	0	0	3.00E-04	8.00E-04	0.0005	0.0015	0	0	0	0
Nitrospirae	0	0	0	0	0	0	0	0	0	0	1.00E-04	2.00E-04
Parcubacteria	0	0	0	0	0	0	0	0	0	0	3.00E-04	8.00E-04
FSR1	0	0	0	0	1.00E-04	3.00E-04	0	0	0	0	0	0
Synergistetes	0.0188	0.0463	0	0	0.0514	0.1045	0.0199	0.0538	0.003	0.0085	0.0004	0.0012
Tenericutes	0	0	0	0	0	0	0	0	3.00E-04	8.00E-04	3.00E-04	8.00E-04

14. Table: Numerical descriptive table of the phylum

8.7.2 Descriptive table of the genus: Relative abundance

	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	8		8		7		8		8		7	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Abiotrophia	6.00E-04	9.00E-04	0.0013	0.0022	2.00E-04	6.00E-04	0.0015	0.0027	0.0005	0.0013	0.0017	0.0026
Acetanaerobacterium	0.0117	0.0156	0.0045	0.0061	0.0012	0.002	1.00E-04	4.00E-04	0.0035	0.0045	0.0041	0.0046
Acidaminococcus	0.0084	0.0156	0.7027	1.9139	0.0673	0.1782	0.1222	0.2959	0.2183	0.4967	0.2506	0.5792
Acidovorax	0.0009	0.0026	0	0	0	0	0	0	0	0	0	0
Acinetobacter	0.0004	0.0013	0	0	0	0	0	0	0	0	3.00E-04	9.00E-04
Actinomyces	0.0297	0.0154	0.0753	0.0925	0.0201	0.0166	0.0655	0.0944	0.0318	0.0249	0.053	0.035
Adlercreutzia	2.00E-04	5.00E-04	2.00E-04	5.00E-04	1.00E-04	3.00E-04	0.001	0.0029	0.0087	0.0236	2.00E-04	3.00E-04
Aerococcus	0	0	2.00E-04	5.00E-04	0	0	0	0	0	0	0	0
Aestuariispira	2.00E-04	6.00E-04	0	0	0	0	0	0	0	0	0	0
Aggregatibacter	0	0	0	0	0	0	0	0	0.0008	0.0022	0	0
Akkermansia	2.7154	6.8303	1.1232	1.6917	2.4571	4.3041	1.5098	3.0561	0.4814	0.7826	0.0521	0.0712
Alistipes	0.7839	1.2002	0.9815	1.1379	2.7203	3.2499	1.1771	1.5888	2.5137	2.1854	2.8335	3.0913
Allisonella	0.0031	0.0088	0	0	0.0072	0.019	0.0026	0.0073	0.0136	0.0249	0.0267	0.0551
Alloprevotella	0.3127	0.881	0	0	0.5221	0.6666	4.00E-04	1.00E-03	0.0838	0.1614	0.3206	0.5727
Anaerobacter	1.00E-04	3.00E-04	0	0	0	0	0	0	6.00E-04	7.00E-04	0	0
Anaerococcus	0.0013	0.0016	0.0005	0.0014	0.0008	0.0021	0.0013	0.0018	0.0028	0.0078	2.00E-04	6.00E-04
Anaerofilum	0.0034	0.0045	0.0023	0.0024	0.0069	0.0056	0.0024	0.0028	0.0023	0.0039	0.0021	0.0029
Anaerofustis	0.0014	0.004	0.0011	0.0013	0	0	0.0024	0.0037	3.00E-04	1.00E-03	0.0036	0.0088
Anaeroglobus	0	0	0	0	0	0	0	0	2.00E-04	4.00E-04	0	0
Anaerorhabdus	0	0	0	0	0	0	1.00E-04	4.00E-04	0	0	0	0
Anaerospora	0	0	0	0	2.00E-04	5.00E-04	0	0	0	0	0	0
Anaerosporobacter	0	0	0	0	2.00E-04	6.00E-04	0	0	0	0	0	0
Anaerostipes	0.7464	0.7847	0.268	0.3101	0.66	0.6147	0.384	0.4794	0.5364	0.7455	0.1265	0.1323
Anaerotruncus	0.0227	0.0197	0.0125	0.0168	0.1542	0.3623	0.0235	0.0325	0.0166	0.0179	0.0105	0.0156
Anaerovorax	1.00E-04	4.00E-04	4.00E-04	7.00E-04	0	0	2.00E-04	6.00E-04	1.00E-04	3.00E-04	0	0
Asaccharobacter	0.0013	0.0031	2.00E-04	4.00E-04	2.00E-04	5.00E-04	0.0005	0.0015	1.00E-04	3.00E-04	0.0007	0.0017
Atopobium	0.002	0.0024	0.014	0.0124	0.0042	0.0051	0.0094	0.0163	0.003	0.003	0.004	0.0024
Bacillus	0.0007	0.0019	0	0	1.00E-04	4.00E-04	4.00E-04	9.00E-04	0	0	2.00E-04	5.00E-04
Bacteroides	9.6784	10.7641	11.5552	10.9127	8.7778	7.7975	8.978	8.4657	12.3026	8.7938	6.2632	4.7497
Barnesiella	0.3496	0.651	0.7901	0.8864	0.6999	0.6653	0.1861	0.295	0.3926	0.4526	0.3243	0.4638
Bdellovibrio	0	0	0.0006	0.0016	0	0	0	0	0	0	0	0
Bifidobacterium	3.7876	3.1955	2.5467	1.7158	0.9263	0.8022	3.7728	4.6833	2.054	2.8196	2.579	2.7656
Bilophila	0.0449	0.0615	0.027	0.0258	0.0932	0.1371	0.1069	0.1702	0.0477	0.0569	0.0291	0.0214
Blautia	2.5605	0.9455	1.6278	0.7773	1.4855	1.0734	2.5138	2.7089	1.9715	1.1286	2.6221	3.3061
Brevibacillus	3.00E-04	6.00E-04	0	0	3.00E-04	9.00E-04	0.001	0.0016	0.0007	0.0011	6.00E-04	6.00E-04
Bulleidia	0	0	0	0	1.00E-04	2.00E-04	1.00E-04	2.00E-04	1.00E-04	3.00E-04	0	0
Burkholderia	0.0011	0.0023	0	0	1.00E-04	3.00E-04	2.00E-04	5.00E-04	3.00E-04	6.00E-04	0	0
Buttiauxella	0	0	0	0	0	0	0	0	0.0028	0.0079	0.0034	0.0082
Butyricicoccus	0.1102	0.0765	0.1999	0.169	0.2684	0.2409	0.8296	2.0431	0.3944	0.4404	0.1875	0.1621
Butyricimonas	0.0723	0.1702	0.0259	0.0401	0.1728	0.189	0.0477	0.0742	0.0273	0.0306	0.0375	0.0436
Butyrivibrio	0	0	0	0	1.00E-04	4.00E-04	0	0	0	0	0	0
Campylobacter	0.002	0.0056	0.0022	0.0041	0	0	0.0006	0.0011	0.0013	0.002	4.00E-04	1.00E-03
Candidatus Carsonella	0	0	0	0	0	0	1.00E-04	4.00E-04	0	0	0	0
Capnocytophaga	2.00E-04	6.00E-04	2.00E-04	5.00E-04	0	0	3.00E-04	1.00E-03	0.001	0.0029	4.00E-04	1.00E-03
Carnobacterium	0	0	0.0008	0.0024	0	0	0	0	0	0	0	0
Catabacter	0	0	0	0	0.001	0.0021	0.0016	0.0024	0.0009	0.0021	1.00E-04	2.00E-04
Catenibacterium	3.00E-04	5.00E-04	0	0	0.5572	1.474	2.192	6.1793	0.148	0.4185	0	0
Catonella	0.0006	0.0017	0	0	1.00E-04	2.00E-04	0	0	0	0	0	0
Centipeda	0	0	0	0	0	0	0	0	2.00E-04	4.00E-04	0	0
Christensenella	0.0199	0.0315	0.0015	0.0034	0.0034	0.0039	0.0012	0.0028	0.0026	0.0068	0.0011	0.0014

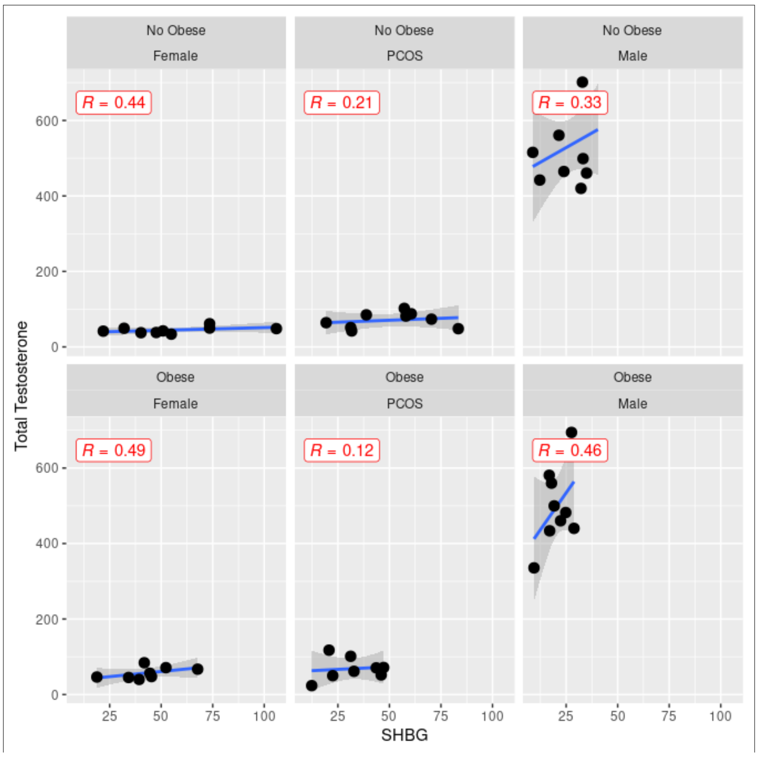
15. Table: Numerical descriptive table of the genus

	Control Females				**PCOS**				**Control Males**			
	No Obese		**Obese**		**No Obese**		**Obese**		**No Obese**		**Obese**	
	8		**8**		**7**		**8**		**8**		**7**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Citrobacter	0.0009	0.0025	2.00E-04	4.00E-04	0	0	0	0	0.1316	0.3715	0.0473	0.1155
Cloacibacillus	0.0141	0.0371	0	0	0	0	0.0154	0.0409	0	0	2.00E-04	5.00E-04
Clostridium IV	0.9084	0.7399	0.9857	0.8061	1.4764	1.8195	1.7202	3.3	0.7672	0.7911	0.7838	0.8618
Clostridium sensu stricto	0.044	0.0689	0.1392	0.1442	0.1061	0.0877	0.0962	0.2152	0.2861	0.3753	0.3281	0.4232
Clostridium XI	0.0141	0.0399	3.00E-04	9.00E-04	0.001	0.0027	0.0006	0.0017	1.00E-04	3.00E-04	0.0041	0.0108
Clostridium XlVa	0.1626	0.1645	0.3189	0.2539	0.4125	0.3785	0.9851	1.5534	0.1823	0.1381	0.2974	0.1685
Clostridium XlVb	0.0821	0.0771	0.0747	0.1228	0.1969	0.415	0.357	0.5145	0.1252	0.1635	0.3474	0.5524
Clostridium XVIII	2.3826	4.5196	1.127	1.5033	0.6856	1.1051	0.9942	1.1512	1.51	1.7649	3.3979	5.859
Collinsella	0.3469	0.435	0.2579	0.2937	0.1524	0.095	0.1482	0.28	1.1399	2.5015	0.7747	0.8087
Coprobacillus	0.0063	0.0109	0.0062	0.0146	0.003	0.0057	0.0012	0.002	0.0005	0.0013	0.0135	0.0354
Coprobacter	0.0137	0.0306	0.0187	0.0268	0.0909	0.1403	0.0024	0.0047	0.0768	0.1987	0.0088	0.0146
Coprococcus	7.5148	11.8567	5.2211	3.9874	2.4411	1.3576	4.5118	4.5907	2.5408	1.9797	9.1771	9.0369
Corynebacterium	0.0007	0.0015	5.00E-04	1.00E-03	0.0007	0.0015	3.00E-04	6.00E-04	0.0008	0.0014	0.0015	0.0029
Cronobacter	0	0	2.00E-04	5.00E-04	0	0	0	0	0.0007	0.0021	3.00E-04	6.00E-04
Cryptobacterium	0	0	2.00E-04	4.00E-04	1.00E-04	3.00E-04	1.00E-04	3.00E-04	5.00E-04	9.00E-04	2.00E-04	5.00E-04
Cupriavidus	3.00E-04	9.00E-04	0	0	0	0	0	0	0	0	0	0
Defluviitalea	0	0	0	0	2.00E-04	5.00E-04	0	0	0	0	0	0
Desulfobaculum	2.00E-04	6.00E-04	0	0	2.00E-04	5.00E-04	1.00E-04	4.00E-04	0	0	0	0
Desulfovibrio	0.0082	0.0201	0.0223	0.0631	0.2886	0.2944	0	0	0.0782	0.1775	0.0244	0.0342
Devosia	0	0	0	0	0	0	0	0	1.00E-04	2.00E-04	0	0
Dialister	0.6181	1.309	1.7572	2.1019	0.8159	1.4639	0.959	1.7538	0.8328	1.0211	0.926	1.3143
Dorea	1.4788	1.2564	2.1765	2.666	0.8192	0.6405	2.7713	2.304	1.2031	1.0766	2.0462	1.9191
Eggerthella	0.0195	0.0318	0.0181	0.0429	0.0093	0.0207	0.1695	0.4385	0.0474	0.1177	0.0045	0.0107
Eisenbergiella	0.0343	0.0413	0.0018	0.0036	0.0061	0.0071	0.0064	0.0137	0.0083	0.0146	0.002	0.0029
Elusimicrobium	0.0004	0.0011	0	0	0	0	0	0	0	0	0	0
Enterobacter	0.004	0.0112	0.0019	0.0037	0	0	0	0	0.0009	0.0024	0.0256	0.0655
Enterococcus	0.0042	0.0115	0.0248	0.0685	0.0007	0.0018	0.0215	0.0603	0.0025	0.0044	0.0175	0.0433
Enterorhabdus	0.0007	0.0019	2.00E-04	6.00E-04	0.001	0.0018	0	0	0	0	0.0011	0.002
Erysipelotrichaceae_incertae_sedis	0.0413	0.0984	0.042	0.0764	0.0111	0.022	0.152	0.3204	0.0533	0.1227	0.1163	0.2926
Escherichia/Shigella	0.7892	1.2899	0.155	0.1103	0.245	0.4368	2.2897	6.2838	0.0865	0.1047	0.1409	0.2313
Ethanoligenens	2.00E-04	5.00E-04	0	0	0	0	0	0	0	0	0	0
Eubacterium	0.0096	0.0136	0.0096	0.011	0.0017	0.0016	0.006	0.0064	0.0036	0.0045	0.0114	0.01
Ezakiella	0.0008	0.0011	0	0	0.0016	0.0025	2.00E-04	5.00E-04	0.0004	0.0012	0.0009	0.0018
Facklamia	0	0	0	0	0	0	0	0	0.0004	0.0012	0	0
Faecalibacterium	5.9046	5.4038	10.8876	7.1283	13.2867	7.279	6.2729	7.1589	17.6768	10.4369	10.3088	7.1011
Faecalicoccus	0.0028	0.0067	0.0006	0.0013	0	0	0	0	0	0	1.00E-04	3.00E-04
Faecalitalea	0	0	0	0	0.0055	0.0141	0.0006	0.0018	0	0	0	0
Finegoldia	0	0	0	0	0	0	0	0	0.0021	0.0037	2.00E-04	6.00E-04
Flavonifractor	0.2003	0.263	0.0949	0.1054	0.1326	0.1521	0.2012	0.2605	0.136	0.1474	0.1289	0.1345
Fructobacillus	0	0	3.00E-04	9.00E-04	0	0	0	0	0	0	0	0
Fusicatenibacter	6.7116	3.6573	4.3223	3.9071	1.16	1.0529	5.3603	4.5066	3.1252	2.4967	1.9693	1.7566
Fusobacterium	0.0039	0.0085	0.0061	0.0107	7.00E-04	8.00E-04	0.0011	0.0017	0.0016	0.0024	0.002	0.004
Gardnerella	0.0004	0.0011	0	0	0	0	0	0	0	0	0	0
Gemella	0.0187	0.0289	0.0323	0.0353	0.0055	0.0061	0.0125	0.0155	0.0142	0.0162	0.0163	0.0095
Gemmiger	4.1077	4.0261	2.4289	1.7945	3.2485	3.1054	2.058	1.6868	4.0007	1.9445	3.6051	2.1515
Gordonibacter	0.0008	0.0012	0.0114	0.0295	0.0026	0.004	0.0067	0.0112	0.0029	0.005	0.0046	0.0074
Granulicatella	0.0086	0.0094	0.0305	0.0378	0.0026	0.0022	0.0281	0.0452	0.0067	0.012	0.008	0.0077
Haemophilus	0.0033	0.0036	0.1837	0.5063	0.0033	0.0038	0.0011	0.0018	0.2474	0.3628	0.2897	0.7504
Hafnia	0	0	2.00E-04	4.00E-04	0	0	0	0	0	0	0.1449	0.3819
Halomonas	0	0	0	0	0	0	0	0	0	0	1.00E-04	2.00E-04

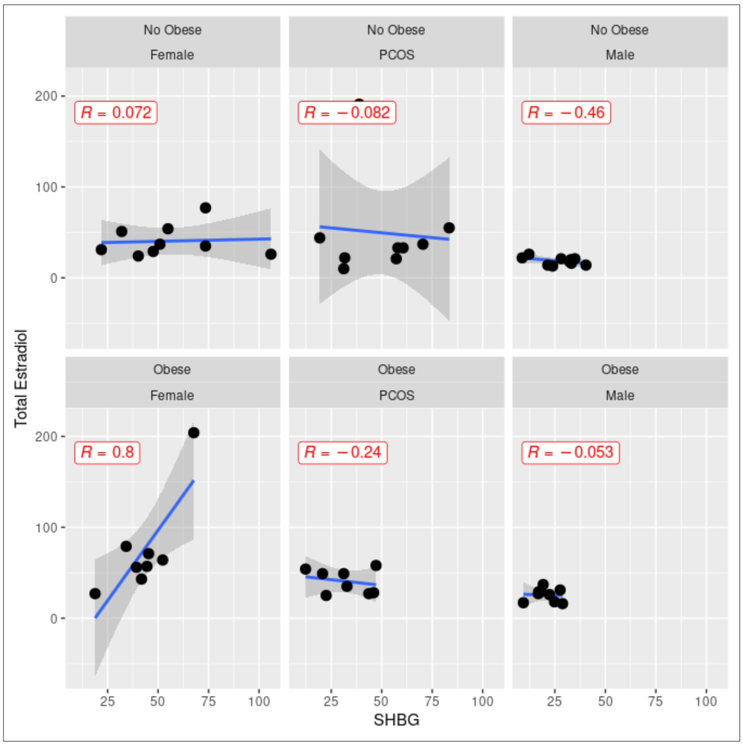
	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	8		**8**		**7**		**8**		**8**		**7**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Hespellia	0	0	0	0	1.00E-04	3.00E-04	0	0	0	0	1.00E-04	2.00E-04
Holdemanella	0.2072	0.5198	0.0607	0.171	0.1128	0.1967	1.5858	2.6552	0.3373	0.4784	1.7177	2.1652
Holdemania	0.0352	0.0414	0.0207	0.022	0.0166	0.0157	0.0078	0.0084	0.0245	0.0281	0.0445	0.0742
Howardella	0.0239	0.0677	0	0	0.011	0.0292	0.0113	0.0229	0.0104	0.0197	0.0158	0.0418
Hungatella	0.0024	0.0063	5.00E-04	1.00E-03	0.0016	0.0021	0.0014	0.0031	0.0008	0.0012	2.00E-04	3.00E-04
Hydrogenoanaerobacterium	0	0	2.00E-04	4.00E-04	0.0009	0.0018	0	0	1.00E-04	3.00E-04	3.00E-04	7.00E-04
Intestinibacter	0.0431	0.0464	0.054	0.0521	0.0903	0.0748	0.0629	0.155	0.0179	0.0261	0.126	0.264
Intestinimonas	0.0086	0.0133	0.0103	0.017	0.0233	0.0318	0.0048	0.0054	0.0023	0.0026	0.0068	0.0084
Janibacter	2.00E-04	5.00E-04	0	0	0	0	0	0	0	0	0	0
Kandleria	0	0	0	0	0.0681	0.1802	1.0349	2.9175	0.0398	0.1126	0	0
Kingella	0	0	0	0	0	0	0	0	0	0	2.00E-04	6.00E-04
Klebsiella	0.0008	0.0017	0.0051	0.0081	0	0	0.0005	0.0012	0.0363	0.1008	0.3273	0.7337
Kocuria	0	0	0	0	0	0	1.00E-04	3.00E-04	0	0	0	0
Lachnoanaerobaculum	0.001	0.0028	0.0004	0.0012	0	0	2.00E-04	5.00E-04	1.00E-04	3.00E-04	0.0012	0.0021
Lachnospira	0	0	3.00E-04	6.00E-04	1.00E-04	4.00E-04	0	0	1.00E-04	3.00E-04	0	0
Lachnospiracea_incertae_sedis	1.7306	0.7801	3.9915	6.1569	2.7637	2.1711	4.9696	5.493	2.3	1.3694	3.8491	3.8637
Lactobacillus	0.1288	0.2829	0.0681	0.1622	0.1574	0.4096	0.0227	0.0353	0.0533	0.1394	0.0555	0.0919
Lactococcus	0.0168	0.0362	0.01	0.0153	0.0915	0.156	0.0367	0.0971	0.0246	0.0542	0.093	0.1962
Lactonifactor	3.00E-04	6.00E-04	0.0012	0.0034	0	0	0.005	0.0142	0.0024	0.0065	0.004	0.0055
Lautropia	0	0	0	0	0	0	0	0	2.00E-04	6.00E-04	0	0
Leclercia	0	0	0	0	0	0	0	0	0	0	1.00E-04	3.00E-04
Leptotrichia	1.00E-04	3.00E-04	0.0004	0.0012	0	0	0	0	0.0019	0.0035	1.00E-04	2.00E-04
Leuconostoc	0.0082	0.0224	0.0017	0.0033	0.0019	0.0037	0.0038	0.0056	0.0034	0.0059	0.0042	0.0056
Limnohabitans	0	0	0	0	0	0	0	0	1.00E-04	2.00E-04	0	0
Lysinibacillus	0	0	0.0004	0.0012	0	0	0	0	0	0	0	0
Massilia	1.00E-04	2.00E-04	0	0	0	0	0	0	0	0	0	0
Megamonas	0.002	0.0023	0.003	0.0086	0.0034	0.0091	0.0011	0.0031	0.1361	0.3323	0.0367	0.0732
Megasphaera	2.00E-04	6.00E-04	0.0073	0.0201	3.00E-04	4.00E-04	0.001	0.0029	0.0987	0.2609	0.3998	0.6883
Methanobrevibacter	0.1938	0.4088	0.176	0.3897	0.4693	0.6621	0.0388	0.0803	0.0031	0.0082	0.4327	0.6403
Methanosphaera	0	0	0	0	0.0023	0.0062	0	0	0	0	0	0
Methylobacterium	0	0	2.00E-04	4.00E-04	0	0	0	0	0	0	0.0006	0.0014
Mitsuokella	3.00E-04	9.00E-04	0	0	0.0155	0.041	0	0	0.0102	0.0191	0.0579	0.1523
Mobiluncus	0	0	0	0	1.00E-04	3.00E-04	0	0	0	0	1.00E-04	2.00E-04
Mogibacterium	0	0	3.00E-04	9.00E-04	0	0	4.00E-04	1.00E-03	0	0	1.00E-04	2.00E-04
Morganella	0	0	0	0	0	0	0	0	0	0	3.00E-04	9.00E-04
Morococcus	1.00E-04	2.00E-04	0	0	0	0	0	0	0	0	0	0
Murdochiella	3.00E-04	9.00E-04	2.00E-04	4.00E-04	4.00E-04	7.00E-04	4.00E-04	1.00E-03	0	0	1.00E-04	3.00E-04
Murimonas	2.00E-04	6.00E-04	0	0	2.00E-04	4.00E-04	2.00E-04	5.00E-04	0	0	0	0
Negativicoccus	0	0	0	0	0	0	0	0	1.00E-04	4.00E-04	0	0
Neisseria	0.0107	0.0184	0.0015	0.0043	0.0018	0.002	5.00E-04	1.00E-03	0.002	0.0024	0.0022	0.0013
Odoribacter	0.0843	0.1218	0.1414	0.2237	0.1046	0.1088	0.0295	0.0413	0.0479	0.0361	0.0974	0.1161
Oligella	0	0	0	0	0.001	0.0027	0	0	0	0	0	0
Olsenella	0.0135	0.0196	0.0088	0.0135	0.0118	0.0199	0.0053	0.0119	0.0066	0.01	0.0029	0.0032
Oribacterium	0	0	0.0085	0.021	0.0011	0.0018	9.00E-04	1.00E-03	0.0017	0.0032	0.001	0.0024
Oscillibacter	0.5634	0.796	0.2933	0.226	1.6441	2.0034	0.3935	0.5168	0.4974	0.5581	0.3063	0.2293
Oxalobacter	0.0027	0.005	0.0084	0.0147	0.012	0.0177	0.001	0.0029	0.0023	0.0048	0.0014	0.0018
Papillibacter	1.00E-04	4.00E-04	0	0	0	0	0	0	0	0	0	0
Parabacteroides	0.9526	1.1267	0.6244	0.4721	1.4154	1.1428	0.2871	0.2573	0.8579	0.3918	0.8907	0.7726
Paraprevotella	0.0353	0.0852	0.0166	0.0456	0.1786	0.3025	0.1235	0.2393	0.6424	1.1487	0.077	0.1333
Parasporobacterium	0.0005	0.0013	0	0	0	0	0	0	0	0	0	0

	Control Females				PCOS				Control Males			
	No Obese		Obese		No Obese		Obese		No Obese		Obese	
	8		8		7		8		8		7	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Parasutterella	0.0857	0.1953	0.0855	0.0974	0.0155	0.0293	0.0989	0.1887	0.064	0.1473	0.04	0.0948
Parcubacteria_genera_incertae_sedis	0	0	0	0	0	0	0	0	0	0	3.00E-04	8.00E-04
Parvimonas	0.0093	0.0148	0.0103	0.0107	0.003	0.0028	0.0037	0.0036	0.0051	0.0089	0.0018	0.002
Pediococcus	0	0	0	0	0	0	0	0	0	0	1.00E-04	3.00E-04
Peptococcus	2.00E-04	5.00E-04	2.00E-04	6.00E-04	0.0037	0.0098	0.0006	0.0018	2.00E-04	5.00E-04	0.001	0.0025
Peptoniphilus	0.0012	0.0021	0	0	9.00E-04	1.00E-03	0.0011	0.0023	0.0045	0.0128	3.00E-04	9.00E-04
Peptostreptococcus	0.0011	0.0025	0.0012	0.0028	0.0007	0.0011	0	0	0.006	0.0138	0	0
Phascolarctobacterium	0.1857	0.3174	0.0624	0.1589	0.9551	1.7912	0.1704	0.419	0.7367	0.9211	0.2395	0.442
Phocaeicola	0	0	0	0	0	0	2.00E-04	6.00E-04	0	0	1.00E-04	2.00E-04
Photobacterium	3.00E-04	7.00E-04	0	0	0	0	0	0	0	0	0	0
Porphyromonas	0.0076	0.008	0.0031	0.0055	0.004	0.0045	0.0025	0.0027	0.0033	0.0042	0.0022	0.0018
Prevotella	0.1395	0.2789	0.2813	0.5021	4.696	6.5095	0.8766	1.4635	2.6111	5.6607	2.129	3.6209
Propionibacterium	0.0039	0.0064	0.0004	0.0011	0.0022	0.0029	0.0006	0.0014	1.00E-04	2.00E-04	0.0006	0.0016
Pseudobutyrivibrio	0	0	0	0	0	0	0.0073	0.0207	0	0	0	0
Pseudoflavonifractor	0.0152	0.0227	0.0107	0.015	0.0126	0.023	0.0444	0.1209	0.0019	0.002	0.0119	0.0204
Pseudomonas	0.0039	0.0054	2.00E-04	5.00E-04	2.00E-04	4.00E-04	1.00E-04	3.00E-04	0.0047	0.0134	8.00E-04	2.00E-03
Psychrobacter	0	0	0	0	2.00E-04	6.00E-04	0	0	0	0	0	0
Pyramidobacter	0.0009	0.0025	0	0	0.0441	0.0911	0	0	0.0023	0.0065	2.00E-04	5.00E-04
Raoultella	0	0	0.0053	0.0151	0	0	0	0	0.0015	0.0018	0.263	0.6956
Rheinheimera	3.00E-04	9.00E-04	0	0	0	0	0	0	0	0	0	0
Rhodococcus	0	0	0	0	0	0	0	0	1.00E-04	3.00E-04	0	0
Robinsoniella	0	0	0	0	0.0004	0.0011	0	0	0.0007	0.0019	0	0
Romboutsia	0.1344	0.16	0.2898	0.3327	0.0929	0.0932	0.2096	0.4049	0.3874	0.3721	1.1256	2.0417
Roseburia	5.1109	4.2389	7.5853	4.7923	7.8715	9.3382	7.0034	7.0095	8.1267	2.5849	5.5054	5.5144
Rothia	0.0169	0.0217	0.0592	0.1046	0.012	0.0095	0.0273	0.0466	0.026	0.0213	0.0554	0.0866
Ruminobacter	0	0	0	0	0	0	0	0	0	0	1.00E-04	3.00E-04
Ruminococcus	11.4259	16.3908	11.182	12.9546	7.7713	7.0268	10.1912	12.1082	4.9595	5.6386	4.7389	4.287
Ruminococcus2	1.901	1.6792	1.7923	1.6629	2.2631	3.4468	1.7304	1.6049	1.8301	1.9611	1.4415	1.0655
Saccharibacteria_genera_incertae_sedis	0.024	0.0187	0.0987	0.1104	0.0155	0.0126	0.0787	0.117	0.0304	0.0215	0.0457	0.0283
Saccharofermentans	2.00E-04	6.00E-04	0	0	0	0	0	0	0	0	0	0
Salmonella	0	0	0	0	0	0	0	0	0	0	0.0009	0.0023
Scardovia	0	0	3.00E-04	9.00E-04	3.00E-04	6.00E-04	0.0012	0.0024	0	0	6.00E-04	1.00E-03
Selenomonas	0	0	0	0	1.00E-04	2.00E-04	0	0	0	0	0	0
Senegalimassilia	0.0374	0.0574	0.1376	0.3072	0.2298	0.4523	0.1368	0.321	0.0409	0.0957	0.118	0.1382
Serratia	0	0	0	0	0	0	0	0	0	0	2.00E-04	6.00E-04
Shuttleworthia	0	0	4.00E-04	7.00E-04	0	0	1.00E-04	3.00E-04	0	0	0	0
Slackia	0.0446	0.0717	0.0333	0.0918	0.0086	0.0187	0.0744	0.1399	0.0243	0.0537	0.0525	0.0598
Sneathia	0.0013	0.0037	0	0	0	0	0	0	1.00E-04	3.00E-04	2.00E-04	4.00E-04
Solobacterium	0	0	0.0029	0.0042	0	0	0.0016	0.0021	4.00E-04	5.00E-04	6.00E-04	6.00E-04
Sphingomonas	0.0007	0.0019	0	0	0	0	0	0	0	0	0	0
Sporobacter	0.0047	0.0057	0.0008	0.0011	0.0037	0.0053	0.0034	0.0067	0.0031	0.0044	0.0023	0.0027
SR1_genera_incertae_sedis	0	0	0	0	1.00E-04	4.00E-04	0	0	0	0	0	0
Staphylococcus	0.0102	0.0164	0.0019	0.0022	0.0024	0.0023	5.00E-04	9.00E-04	0.0015	0.002	0.0021	0.0024
Stenotrophomonas	0.0005	0.0012	0	0	0	0	2.00E-04	5.00E-04	0	0	1.00E-04	2.00E-04
Stomatobaculum	0	0	0.0008	0.0024	2.00E-04	6.00E-04	3.00E-04	9.00E-04	4.00E-04	6.00E-04	0.0005	0.0012
Streptococcus	1.2792	0.8321	1.2715	1.1806	0.743	0.476	1.0411	0.839	1.6687	1.368	1.576	1.75
Streptophyta	0.0065	0.0091	0.0052	0.0078	0.0312	0.0696	0.005	0.006	0.0259	0.048	0.0049	0.0076
Subdivision5_genera_incertae_sedis	0	0	2.00E-04	5.00E-04	0	0	0	0	0	0	0	0
Subdoligranulum	3.00E-04	1.00E-03	3.00E-04	6.00E-04	0.0011	0.0029	0	0	0	0	3.00E-04	7.00E-04
Succinatimonas	0	0	0	0	0	0	0	0	0	0	0.0004	0.0011
Succiniclasticum	0	0	0.3768	1.0581	0	0	0	0	0.1599	0.4521	0.0772	0.1864
Succinivibrio	0	0	0	0	0.2575	0.6812	2.00E-04	6.00E-04	0	0	0.0175	0.0462
Sutterella	0.0704	0.1867	0.0178	0.0298	0.1015	0.1115	0.0259	0.0541	0.0493	0.0656	0.0582	0.0761
Terrisporobacter	0.0225	0.0385	0.0322	0.0392	0.0222	0.0344	0.0021	0.0026	0.0194	0.0321	0.0213	0.0123
Tetragenococcus	0	0	0	0	0	0	2.00E-04	6.00E-04	0	0	0	0
Turicibacter	0.0076	0.01	0.0086	0.0105	0.0209	0.022	0.0033	0.0037	0.0245	0.0599	0.0153	0.0163
Vampirovibrio	0	0	0	0	0.0005	0.0013	1.00E-04	2.00E-04	0	0	0	0
Varibaculum	0	0	0	0	2.00E-04	5.00E-04	0	0	0	0	0	0
Veillonella	0.0142	0.0183	0.3751	0.9166	0.0127	0.0098	0.0884	0.2388	0.1332	0.2031	0.0514	0.1073
Victivallis	3.00E-04	8.00E-04	0	0	3.00E-04	7.00E-04	3.00E-04	7.00E-04	0	0	0	0
Weeksella	0	0	0	0	0.0011	0.003	0	0	0	0	0	0
Weissella	0.0016	0.0032	0.0086	0.0203	0.002	0.0035	3.00E-04	5.00E-04	0.0153	0.0408	0.0132	0.0338
Yersinia	0	0	0	0	0	0	0	0	0	0	0.0083	0.022

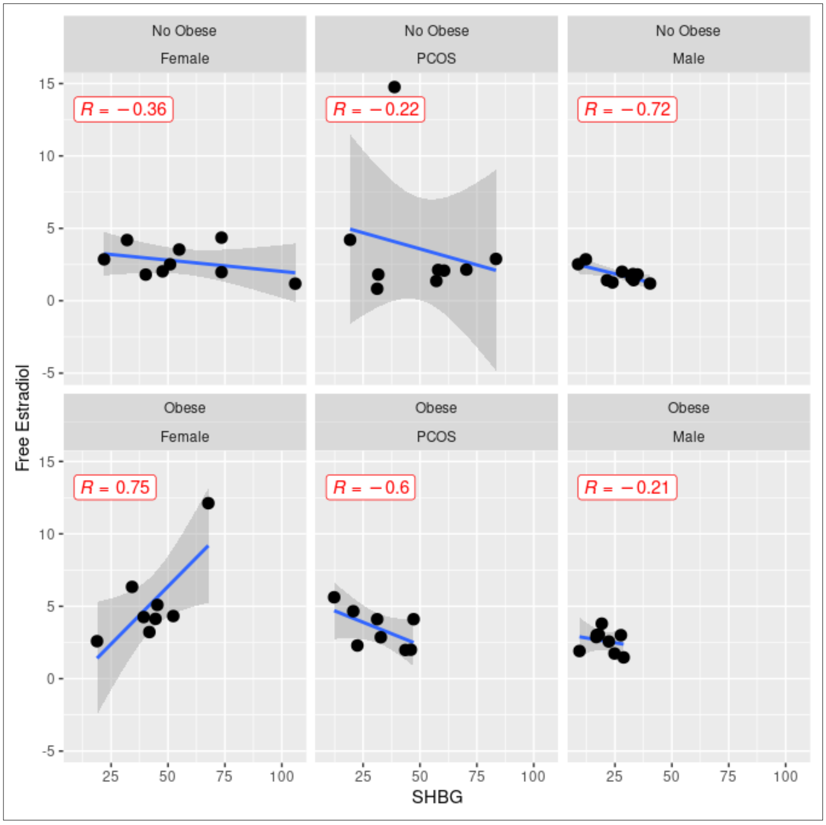
8.8 Bivariate Analysis of the Clinical Data



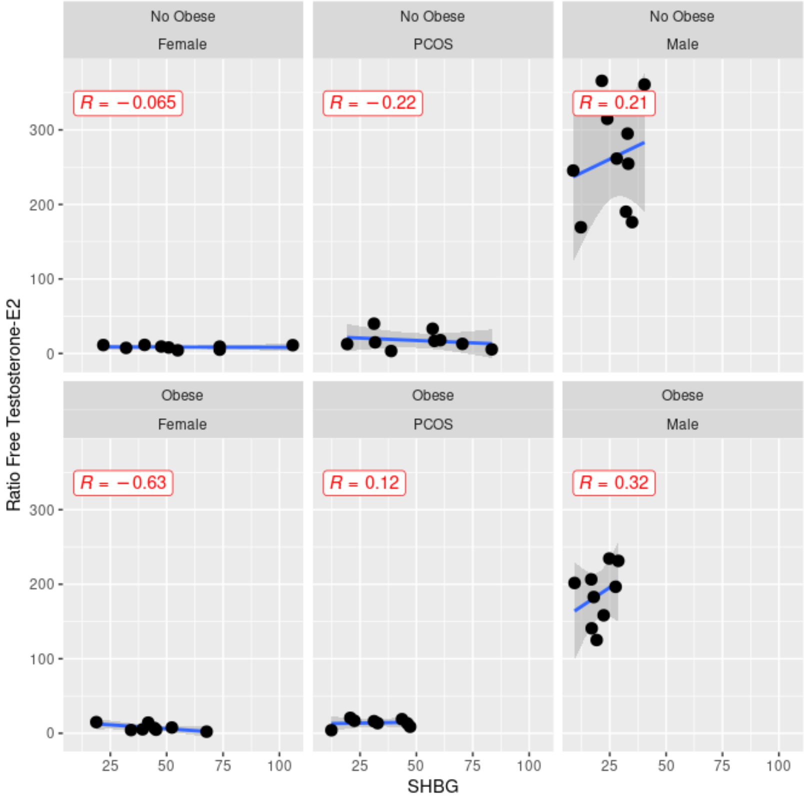
53. Figure: Ordinary Least Squares regression between the



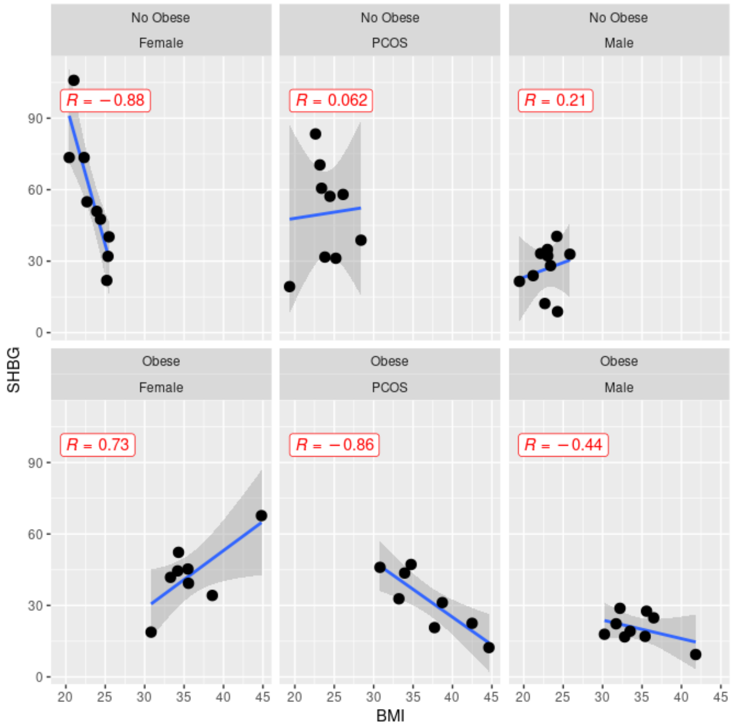
total estradiol and the SHBG



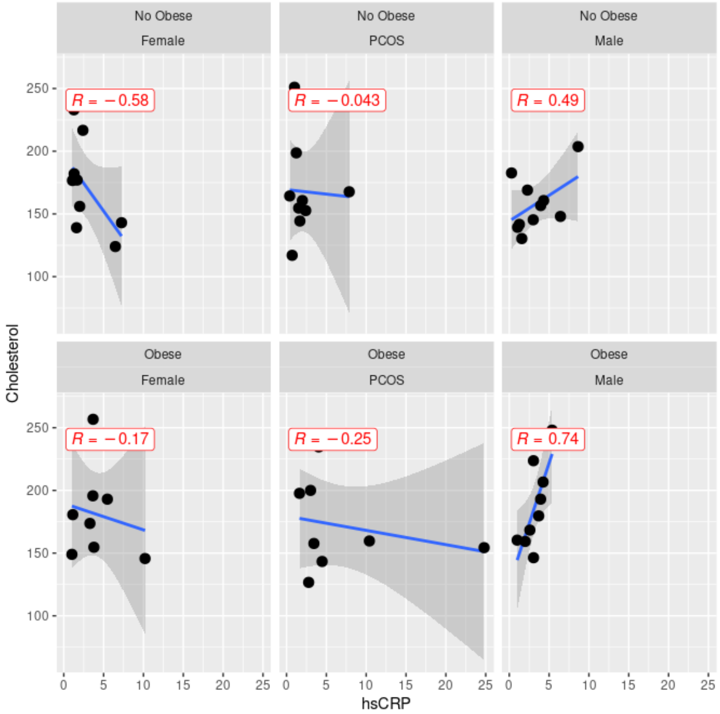
54. *Figure: Ordinary Least Squares regression between the free estradiol and the SHBG*



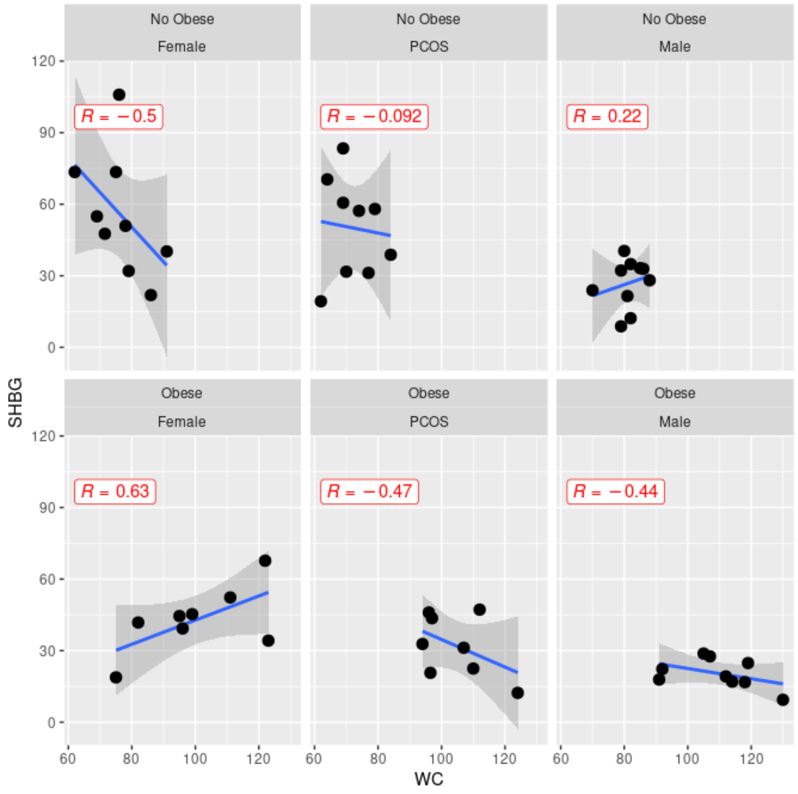
55. *Figure: Ordinary Least Squares regression between the ratio of free sex steroids.*



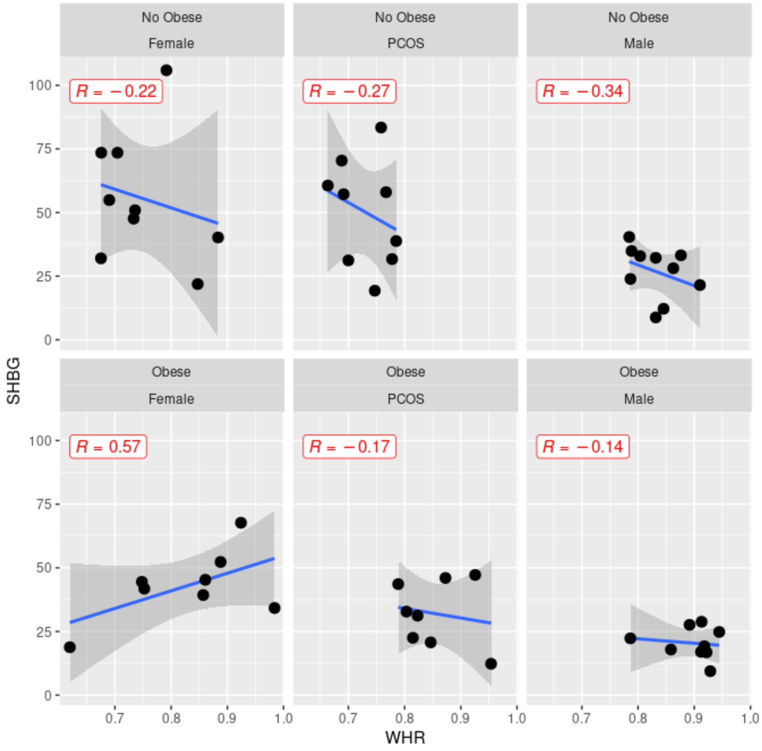
56. Figure: Ordinary Least Squares regression between SHBG and BMI



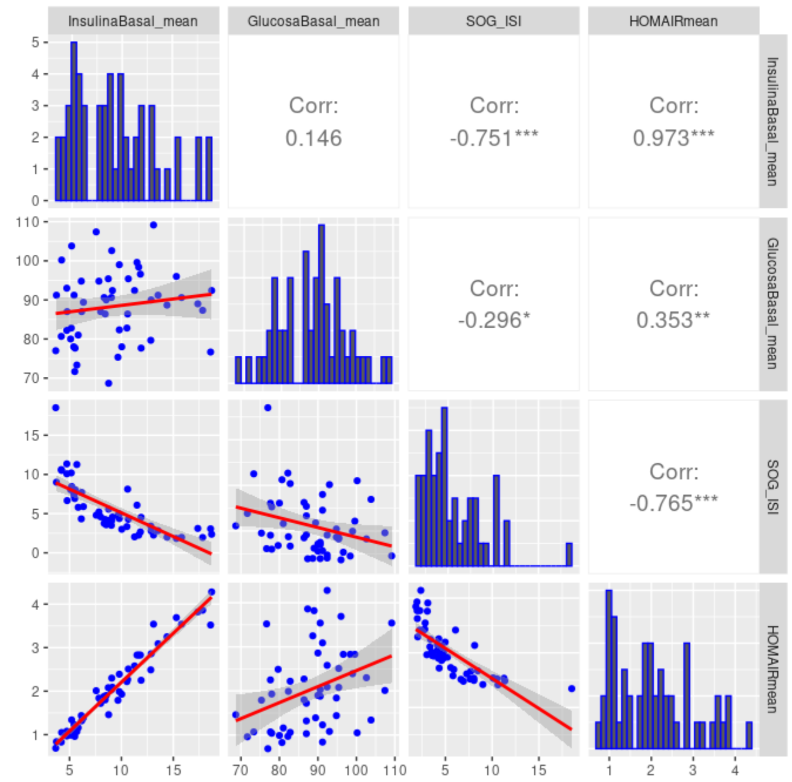
57. Figure: Ordinary Least Squares regression between cholesterol and hsCRP



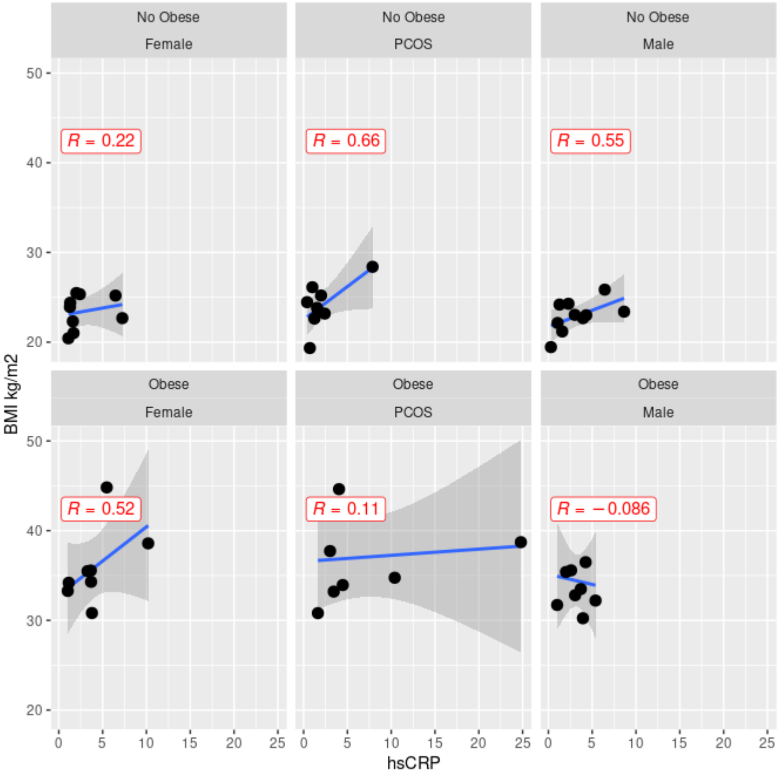
58. Figure: Ordinary Least Squares regression between SHBG and WC



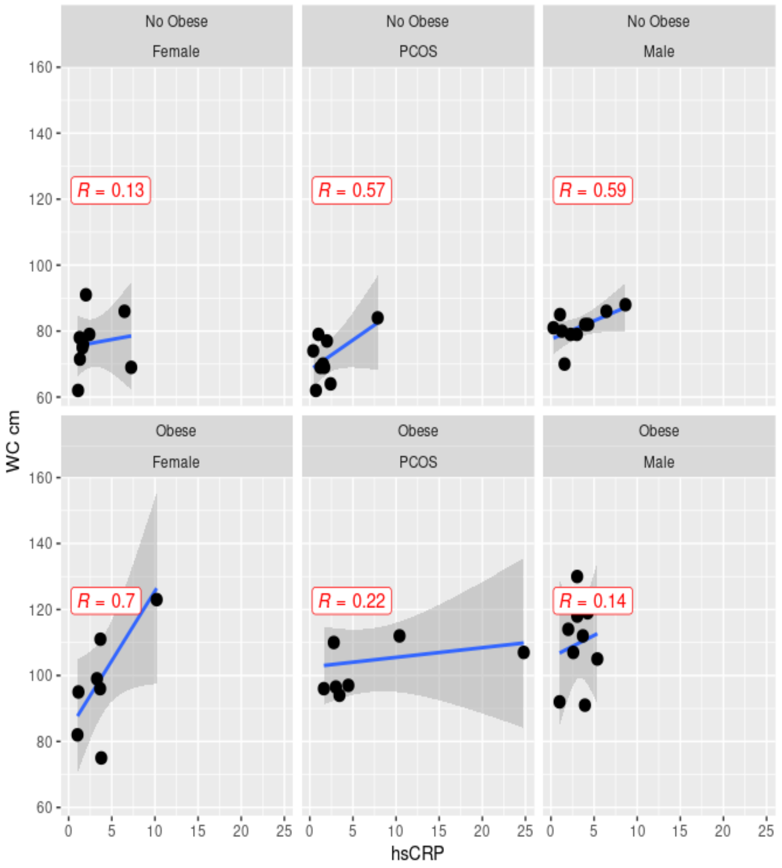
59. Figure: Ordinary Least Squares regression between the SHBG and WHR



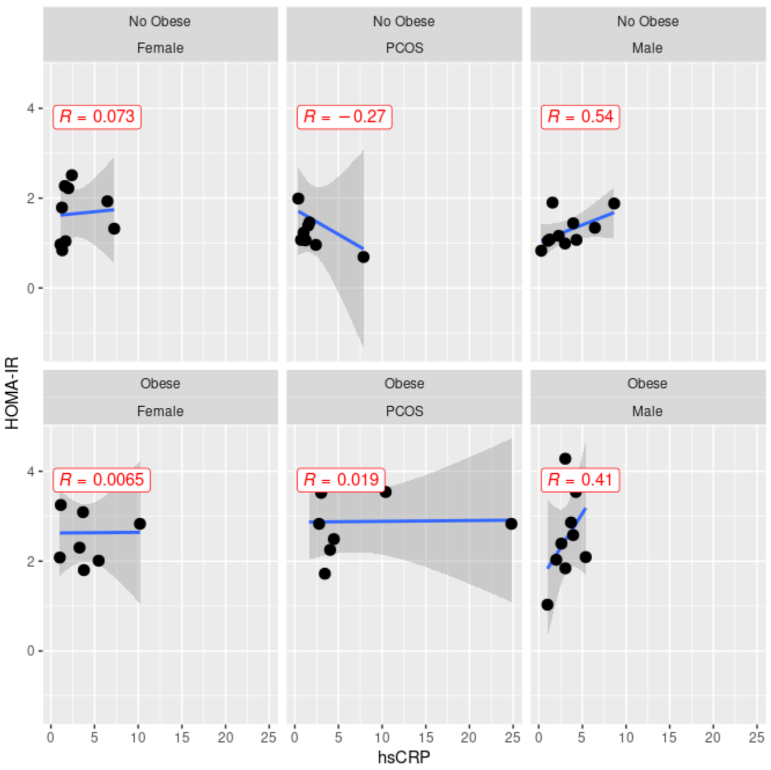
60. Figure: Scatter matrix correlation and OLS between the mean basal values across all days of Insulin, Glucose, ISI and HOMA-IR



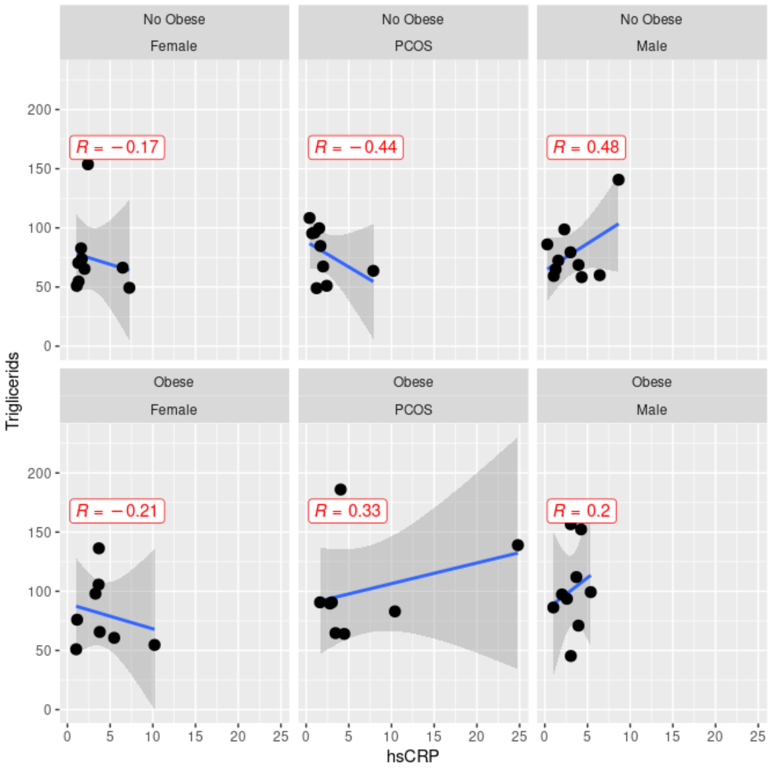
61. Figure: Ordinary Least Squares regression between BMI and hsCRP



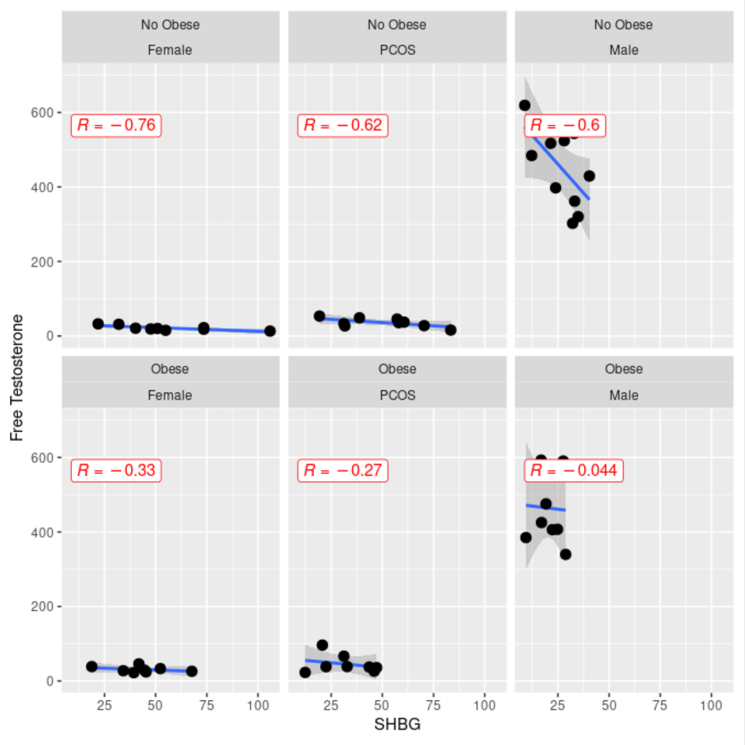
62. Figure: Ordinary Least Squares regression between WC and hsCRP



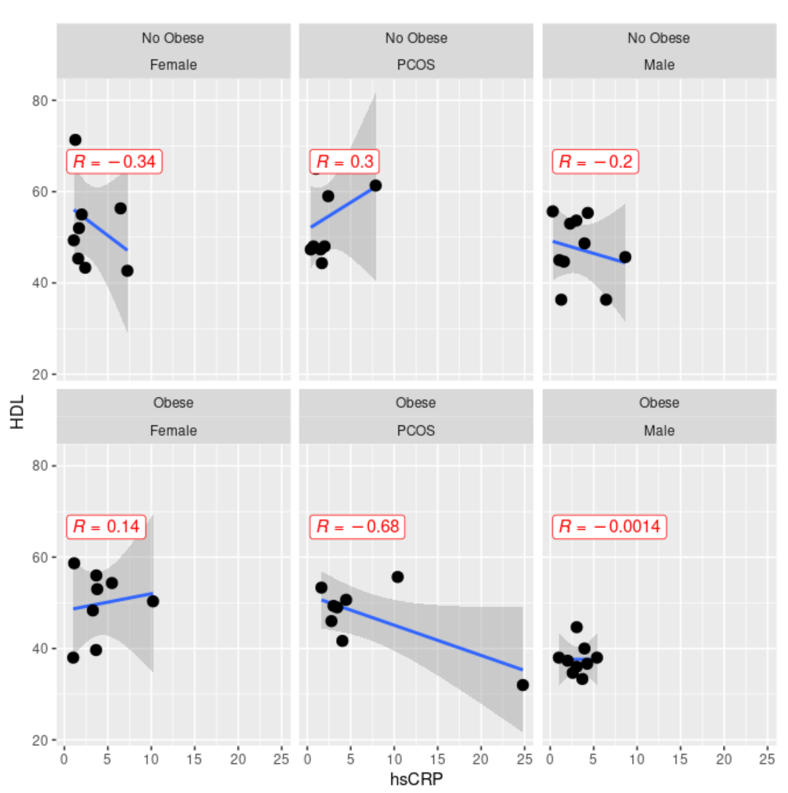
63. Figure: Ordinary Least Squares regression between HOMA-IR and hsCRP



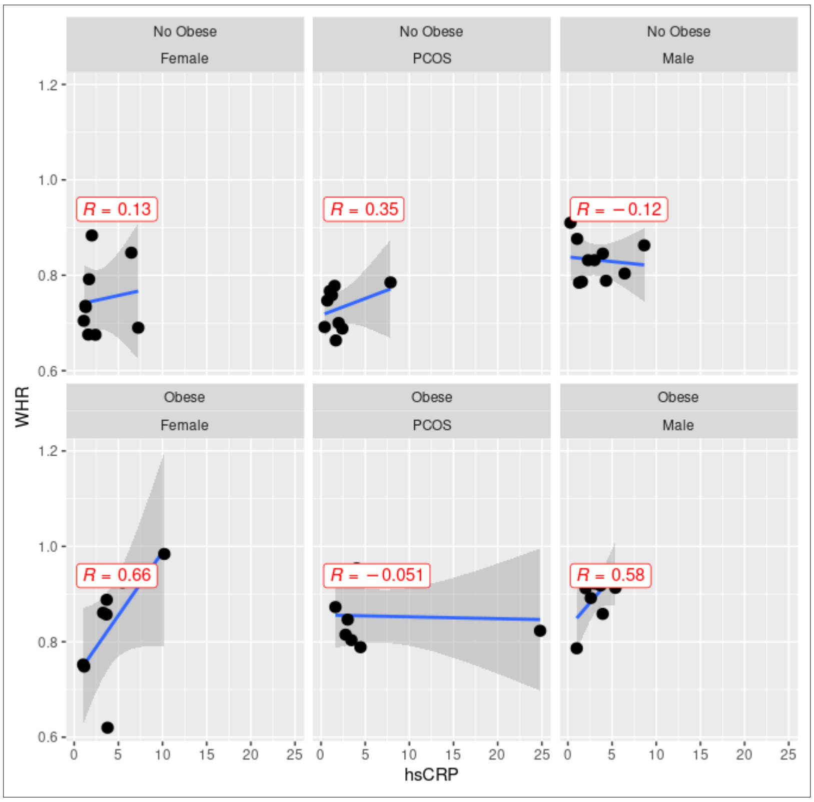
64. Figure: Ordinary Least Squares regression between triglycercids and hsCRP



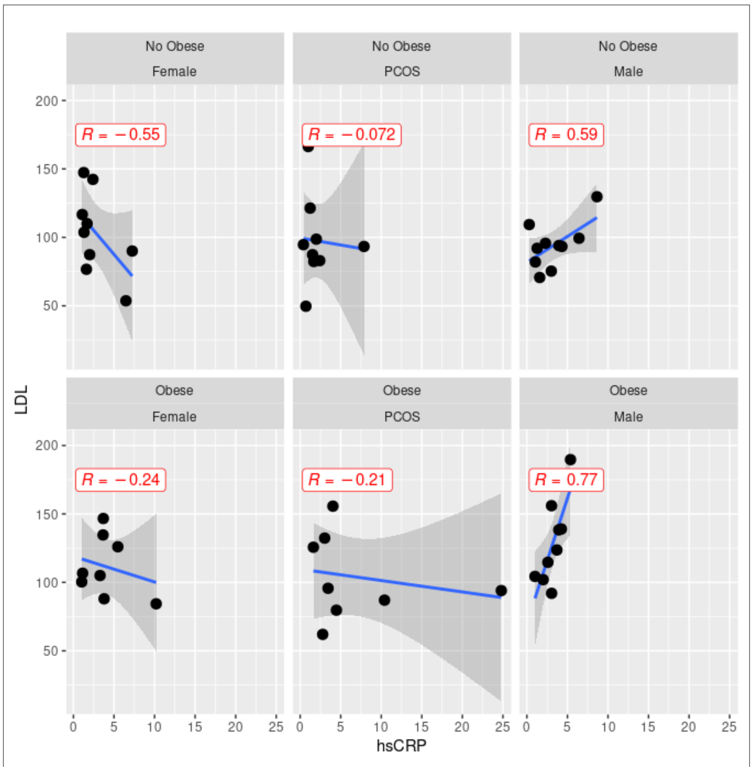
65. Figure: Ordinary Least Squares regression between th free testosterone and th SHBG



66. Figure: Ordinary Least Squares regression between HDL-cholesterol and hsCRP

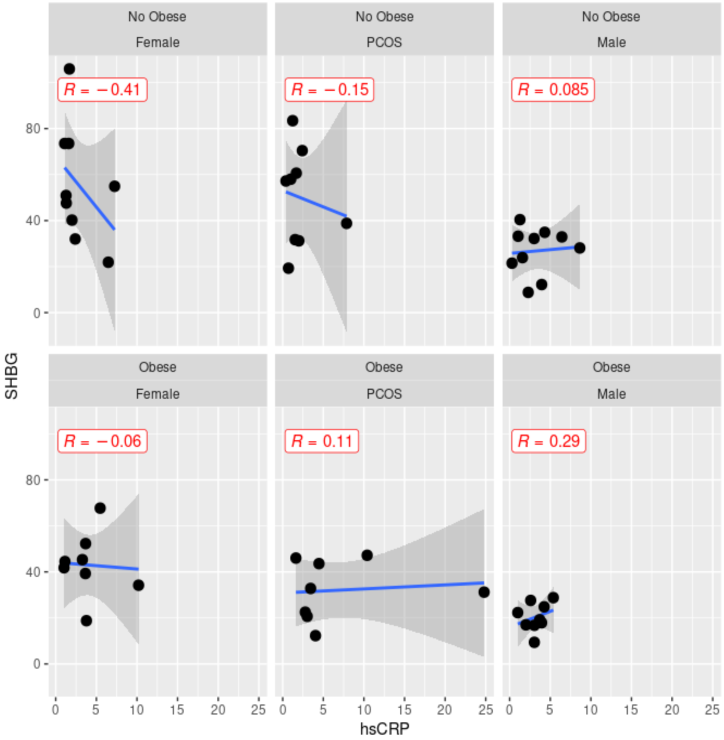


67. Figure: Ordinary Least Squares regression between WHR and hsCRP



68. Figure: Ordinary Least Squares regression between LDL-cholesterol and hsCRP

69. Figure: Ordinary Least Squares regression between



the SHBG and hsCRP