

Logbook

Multi-Omics Project

Date: 30-04-2025

Written by: Jarno Jacob Duiker
Student Bio-informatics 2nd year, Hanze Honours, Faculty of Life sciences, Hanze UAS.

Contents

0.0.1	Introduction	2
0.0.2	Load libraries	2
0.1	Reading the Paper	2
0.1.1	Proposal for research question	7
	Proteomics data	8

0.0.1 Introduction

This is a logbook for the Omics liver transplant project. The goal of this project is to perform a PCA on both the proteomics data and RNA transcriptomics data, and find a link.

0.0.2 Load libraries

0.1 Reading the Paper

The paper selected for this multi-omics project is the *Bile proteome reveals biliary regeneration during normothermic preservation of human donor livers* Thorne et al. (2023)

The paper looks at the effects of NMP (Normothermic machine perfusion) on livers that are to be transplanted that may not be in the optimal shape, meaning they have high biliary injuries (Traumatic damage to the bile ducts) or lower biliary viality (a lower ammount of living cells). The research was conducted on 55 livers that were deemed suitable for transplantation analysis. The livers came from ECD (extended criteria donors) and DCD (donation after circulatory death) donors.

To analyse the livers a multitude of samples were taken. Before the NMP 46 bile duct biopsies were taken, and 50 liver tissue biopsies were taken (these biopsies were taken before the livers underwent Hypothermic oxygenated machine perfusion - DHOPE) after undergoing DHOPE and COR (controlled oxygenated re-warming) The livers were put in the NMP where metabolic processes, bloodflow and simulation of other physiological conditions were started. 30 minutes after these livers were in the NMP a Bile sample was taken, this was taken of all 55 livers. At 150 minutes another bile sampes was taken of 54 (not 55 because one sample did not contain any proteins) livers. At this point in the NMP process all livers were analysed for a multitude of things to determine organ viability. The two distinct areas that were looked at were the hepatocellular vaibility, this focused on the metabolic functionality of the liver parenchyme (functional tissue of the liver, makes up 80% of the liver volume as hepatocytes), this is assesed by criteria such as lactate clearance and bile production. The other area was biliary viability that focuses on the functional capacity of cholangiocytes (epithelial cells of the bile duct) this was assesed by specific biochemical composition of the bile Thorne et al. (2023) . After the viability assesement was concluded 35 livers were viable for further NMP treatment pre transplantation, 20 livers were deemed non-viabel and therefore had their NMP stopped then being discarded. For the 35 livers that continued another sample of bile was taken at the END of the NMP treatment. However 33 samples were taken because of complications with 2 livers at transplantation.

The Bile, a very important part of this paper. But what is this bile? According too Boyer (2013) bile consist out of a number of things ~95% of bile is water, Electrolytes, Organic Anions like Bile salts (Steroid/primary acids like cholic acid, chemodeoxcholic acid or secondary acids like deoxycholic acid and lithocholic acid) Lipids like cholesterol, Steroid hormones and estrogen. Proteins, peptides. heavy metals and others also are found in bile. For a more indepth look into what bile contains i would suggest reading table 1 in the Boyer (2013) paper this gives the known composition of bile in most species.

The BDI, BDI meaning Bile Duct Injury is a metric that shows how much damage the liver has baised on 4 criteria. These being the presence of vascular lesions (abnormal growth or malformations in the bloodvessel),

stromal necrosis (Organ tissue/cell death), Injury to the periluminal peribiliary gland and injury to the deep peribiliary glands. Using these 4 criteria a score was made called “a total histological bile duct injury score” this score ranged from 2 being the lowest to 14 being the highest with the median being 7. Using this previously mentioned median the livers were divided into 2 groups. High BDI and Low BDI using this group system a number of analysis were made like seeing if BDI had influence of biliary viability and other analyses.

PCA, Principal component analysis was used in this paper and will be used in our research as well. For this reason I made an explanation on how PCA works that is understandable for me and I will share here. PCA is an excellent tool to process datasets with a large amount of factors. This is because when comparing factors you can only plot so many otherwise the dimensions would be too large and too difficult to interpret. Using PCA you can turn for example 200 factors into 5 principal components (Principal components are ranked from most important to least important, meaning PC1 will be more important than PC3) To see how much info of the data set is explained within each principal component a SCREE plot could be used. This plot gives percentages on how much variance/info each PC contains.

Here is an example of a scree plot

```
# Drop the non-numerical column
df <- iris[, -5]

# Perform PCA on the iris dataset
pca <- prcomp(df, scale = TRUE)

# Extract the eigenvalues from the PCA object
eigenvalues <- pca$sdev^2

summary(pca)
```



```
## Importance of components:
##              PC1      PC2      PC3      PC4
## Standard deviation  1.7084 0.9560 0.38309 0.14393
## Proportion of Variance 0.7296 0.2285 0.03669 0.00518
## Cumulative Proportion 0.7296 0.9581 0.99482 1.00000
```



```
# Percentage of variance explained
plot(eigenvalues/sum(eigenvalues), type = "b",
     xlab = "Principal Component",
     ylab = "Percentage of Variance Explained")
```

In Fig 1 A scree plot was made of the PCA of the iris dataset

The scree plot shows that the first two principal components explain the most variance in the data. The third and fourth principal components explain much less variance.

Based on the scree plot, we can conclude that the first two principal components are sufficient for capturing the most important information in the data.

What is an eigen value? It is a measure of the strength or importance of its corresponding Principal Component. A larger eigenvalue means the Principal Component captures more of the total variance in the data. The sum of all eigenvalues is equal to the total variance in the original dataset. The code `eigenvalues <- pca@sdev^2` calculates the squared standard deviations, it's calculating these variances along the principal component directions, which are the eigenvalues of the data's covariance/correlation matrix. These eigenvalues are needed for understanding how much information each PC retains for this in turn helps decide how many PC will be necessary to use.

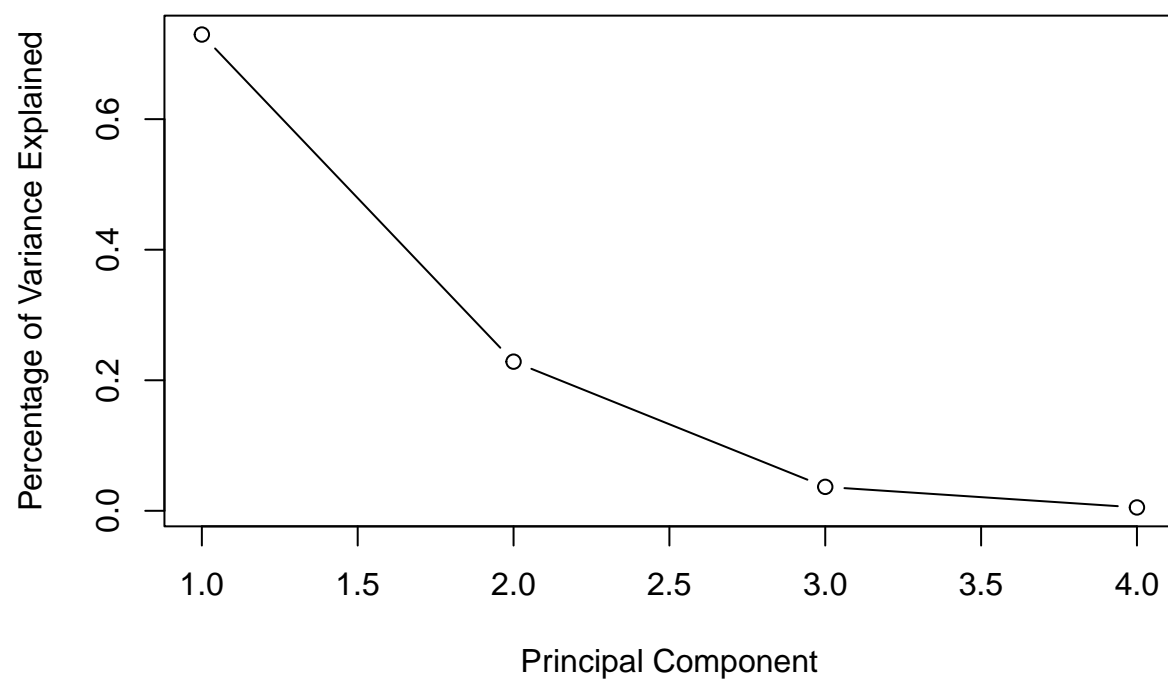
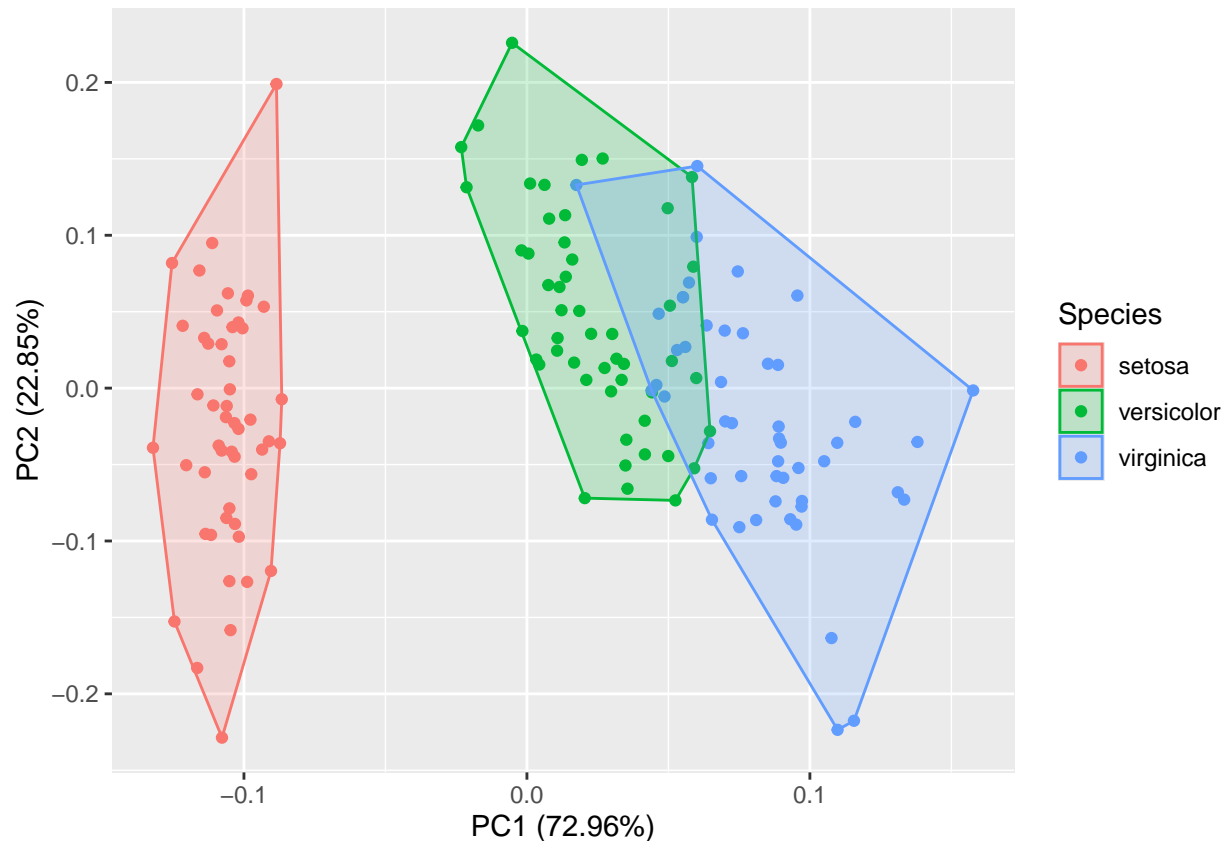


Figure 1: Fig.1 - SCREE plot

If we use PC1 and PC2 from the PCA of the iris dataset we will have about 90% of the info of the original dataset this will be more then enough. After deciding how many PC's need to be used it is important to plot them using a scatterplot for example

```
iris.pca.plot <- autoplot(pca,
  data = iris,
  frame = TRUE,
  colour = 'Species')

iris.pca.plot
```



Here a scatterplot was made of the PCA of the iris dataset. The plot shows the first two principal components (PC1 and PC2) on the x and y axes, respectively. Each point represents an observation (in this case, a flower), and the color indicates the species of the flower. The groups are shown by taking all the most outside points and drawing a line around the rest of the dots.

```
# Graph of the variables
fviz_pca_var(pca, col.var = "black")
```

This plot shows the variables and their correlation, also known as loadings. The arrows represent the original variables, and their direction and length indicate how they contribute to the principal components. Longer arrow means stronger contribution. The plot also shows the correlation between the variables. For example, Petal Length and Petal Width are highly correlated, as indicated by their close proximity and similar direction. Sepal Length is positively correlated with Petal Length and Width but not as strongly. Sepal Width is negatively correlated with both Petal Length and Width.

applying PCA to the data After reading the PCA that was done in the paper and exploring and explanation of different sides of a good PCA, I wrote down a small list of suggestions that we should keep in

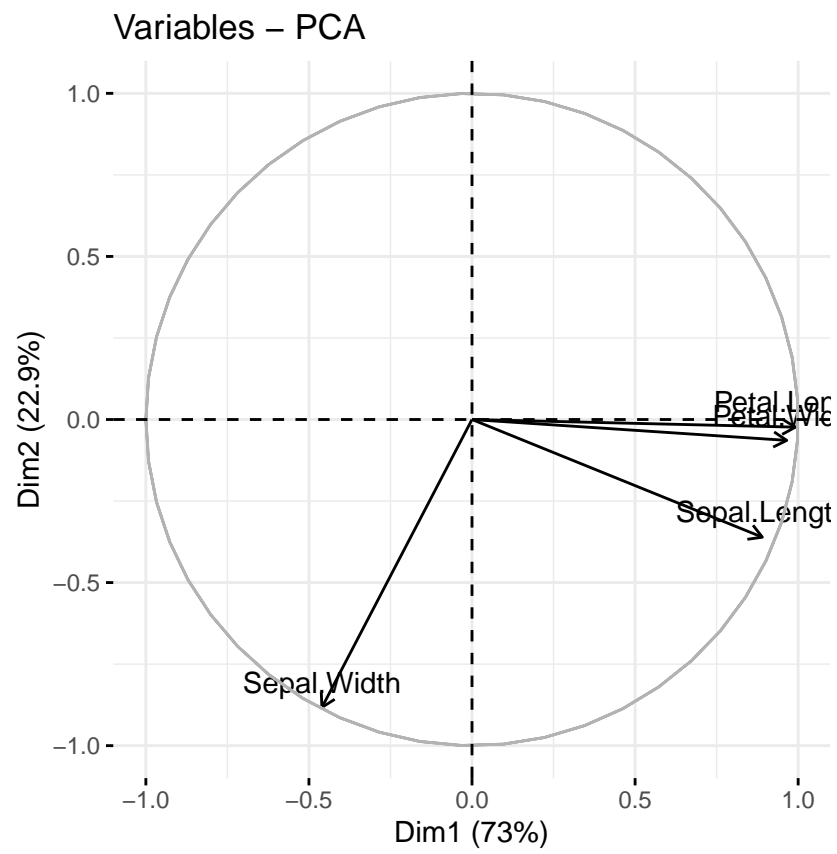


Figure 2: Fig.2 - Loading plot

mind when we are doing our PCA. This being to first do a proper analysis of the data, this means looking at the data and seeing if there are any outliers or missing values. If there are outliers or missing values, we should remove them from the dataset. After this we should scale the data, this is important because PCA is sensitive to the scale of the data. If the data is not scaled, the PCA will be biased towards the variables with larger scales. After scaling we can do a PCA on the data and plot it using a scatterplot. This will give us a good overview of how the data looks like and if there are any patterns in it. After this we can do a loading plot to see how much each variable contributes to each principal component. This will help us understand which variables are important for the PCA and which ones are not.

After this PCA, we should analyse the cluster groups by subsetting the data and looking at the different groups. This will help us understand how the different groups are related to each other and if there are any patterns in the data. After this we can do a correlation analysis to see how the different variables are related to each other. Also a heatmap can be made to see correlation. E

ANOVA?

KEGG,

Diablo

$$\max_{\alpha_h^{(1)}, \dots, \alpha_h^{(Q)}} \sum_{\substack{i,j=1 \\ i \neq j}}^Q c_{i,j} \cdot \text{cov} \left(X_h^{(i)} \alpha_h^{(i)}, X_h^{(j)} \alpha_h^{(j)} \right) \quad (1a)$$

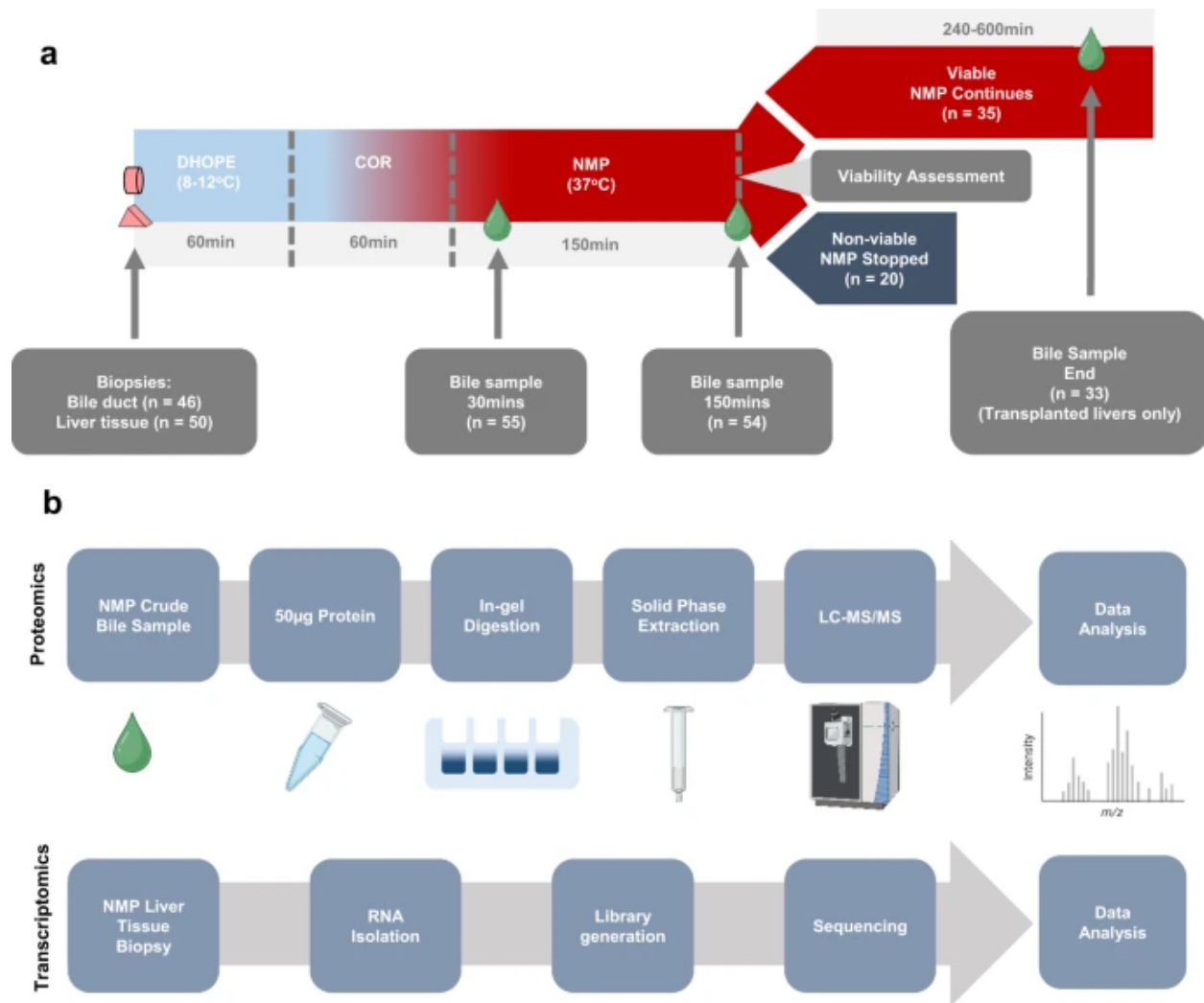
MOFA+

Transcriptomics

0.1.1 Proposal for research question

Question Can longer usage of NMP lead to benefits in preservation and functional assessment of Sub-optimal ECD livers?

Why? The paper it self calls for this as an interesting question for research, the data this paper has provided can be used with the Diablo tool to see what proteins are



Proteomics data

Boyer, James L. 2013. "Bile Formation and Secretion." *Comprehensive Physiology* 3 (3): 1035–78. <https://doi.org/10.1002/cphy.c120027>.

Thorne, Adam M., Justina C. Wolters, Bianca Lascaris, Silke B. Bodewes, Veerle A. Lantinga, Otto B. van Leeuwen, Iris E. M. de Jong, et al. 2023. "Bile Proteome Reveals Biliary Regeneration During Normothermic Preservation of Human Donor Livers." *Nature Communications* 14 (1): 7880. <https://doi.org/10.1038/s41467-023-43368-y>.