Informe de análisis bioinformático y estadístico sobre los datos crudos de un estudio metabolómico

Ana Martín Ramírez- 2025

TABLA DE CONTENIDOS

1. Abstract

This study examines the variations in glucose levels in 39 patients with morbid obesity who underwent bariatric surgery. Using a dataset from a metabolomic study of surgically treated patients, a statistical analysis was conducted to evaluate the evolution of glucose, one of the most important metabolites, by comparing the levels before and after surgery. The results showed a progressive decrease in glucose levels, with stabilization and improvement following the intervention. However, no significant differences in glucose levels were observed based on gender or the type of surgery performed, suggesting that these factors do not significantly affect glucose levels. Despite the study's limitations, such as the small sample size and the lack of control over certain external factors, the findings suggest that bariatric surgery has a positive effect on glucose control. These results open new perspectives for investigating the metabolic mechanisms of bariatric surgery in patients with morbid obesity.

2. Objetivos:

Los objetivos principales de este trabajo son:

- Analizar los efectos de la cirugía bariátrica realizada a pacientes con obesidad mórbida sobre los niveles de ciertos metabolitos seleccionados a partir de un estudio metabolómico previamente realizado cuyo propósito era evaluar el papel de esta cirugía como modificador metabólico.

Para ello, se proponen los siguientes objetivos específicos:

- Seleccionar un estudio metabolómico de interés de la plataforma GitHub, descargar los datos crudos y los metadatos asociados a dicho estudio, y proceder con su análisis detallado utilizando herramientas estadísticas y bioinformáticas.
- Realizar un procesamiento preliminar de los datos crudos del estudio para asegurar su calidad antes de llevar a cabo análisis estadísticos y visualizaciones.
- Realizar un análisis exploratorio de los datos del estudio con el objetivo de obtener una visión general a través de estadísticas descriptivas y visualizaciones gráficas.
- Analizar los resultados obtenidos tras realizar los análisis estadísticos pertinentes y extraer conclusiones biológicas sobre el impacto de la cirugía bariátrica en los niveles de ciertos metabolitos de interés (glucosa) en pacientes obesos.
- Identificar las limitaciones del estudio realizado y reflexionar sobre los posibles factores que puedan haber influido en los resultados del estudio.

4. Métodos:

Los datos crudos utilizados en este estudio provienen del artículo titulado "Metabotipos de respuesta a la cirugía bariátrica independientes de la magnitud de la pérdida de peso". Este conjunto de datos se encuentra disponible en la página web del NCBI y también en un repositorio independiente en GitHub. El conjunto de datos que forman el repositorio de Github está compuesto por los siguientes archivos:

- DataInfo_S013.csv: contiene los metadatos del estudio, proporcionando información relevante sobre cada columna del archivo DataValues_S013.csv.

- DataValues_S013.csv: incluye los valores clínicos y metabolómicos de 39 pacientes en 5 momentos temporales, con mediciones detalladas de varios metabolitos y parámetros clínicos.

- AAInformation_S006.csv: proporciona información adicional sobre los metabolitos presentes en DataValues_S013.csv, detallando las características y propiedades de cada metabolito medido.

NA	SUBJECTS	SURGERY	AGE	GENDER	Group	MEDDM_T0	MEDCOL_T0	MEDINF_T0	MEDHTA_T0	GLU_T0	INS_T0	HOMA_T0	HBA1C_T0	HBA1C.mmol.mol_T0	PESO_T0	bmi_T0
1	1	by pass	27	F	1	0	0	0	1	85	11.4	2.4	NA	NA	151	62.9
2	2	by pass	19	F	2	0	0	0	0	78	12.1	2.32	NA	NA	139	47
3	3	by pass	42	F	1	0	0	0	0	75	8.41	1.56	5.4	35.51	84	29.8
4	4	by pass	37	F	2	0	0	0	0	71	12.8	2.25	5.1	32.23	136	53.1
5	5	tubular	42	F	1	0	0	0	0	82	6.01	1.22	5.6	37.69	121	46.6
6	6	by pass	24	F	2	0	0	0	0	71	9.88	1.73	5.1	32.23	148	48.8
7	7	tubular	33	F	1	0	0	0	0	80	9.2	1.82	5.6	37.69	109	43.7
8	8	tubular	55	F	1	0	0	1	0	90	3.4	0.76	5.5	36.6	109	41.8
9	9	tubular	40	F	1	0	0	0	0	92	5.43	1.23	5.7	38.78	114	44
10	10	tubular	47	M	1	0	0	0	0	84	6.98	1.45	5.5	36.6	120	40.6
11	11	tubular	33	M	1	0	0	1	0	75	13.3	2.47	5.7	38.78	171	54.4
12	12	by pass	57	F	2	0	0	0	0	108	16.8	4.47	NA	NA	135	55.5
13	13	by pass	56	F	1	0	0	0	1	101	17.1	4.26	NA	NA	124	52.3
14	14	by pass	45	F	1	0	0	0	1	105	21.3	5.53	NA	NA	119	45.9
15	15	by pass	55	F	1	0	0	0	1	139	36.6	12.6	NA	NA	154	62.5
16	16	by pass	39	F	1	0	0	0	1	106	20	5.24	5.8	39.88	162	61.7
17	17	by pass	29	F	1	0	0	0	0	159	17.6	6.91	NA	NA	146	49.9
18	18	by pass	27	F	1	0	0	0	0	103	29.5	7.49	5.8	39.88	114	47.5
19	19	by pass	41	F	2	0	0	0	0	106	13.3	3.47	NA	NA	128	53.3

Figura 1. Imagen representativa de los datos crudos contenidos en "DataValues_S013.csv"

Metodología empleada

Para el análisis de los datos, se utilizó un enfoque de procesamiento y análisis en R, una herramienta estadística y bioinformática ampliamente utilizada en la investigación ómica. El proceso siguió varios pasos secuenciales, que incluyen la carga, limpieza y alineación de datos, la creación del objeto SummarizedExperiment, y el análisis exploratorio univariante y multivariante.

- **Carga de datos**: los archivos de datos y metadatos fueron cargados en R utilizando la función read.csv (), lo que permitió importar tanto las mediciones clínicas y metabolómicas (DataValues_S013.csv) como los metadatos asociados (DataInfo_S013.csv). Posteriormente, se realizó una primera visualización de los ficheros para comprobar la información contenida en ellos y verificar sus dimensiones.

- **Procesamiento y alineación de datos y metadatos**: dado que los datos de las mediciones y los metadatos presentaban un desajuste en el número de filas, se realizaron ajustes en ambos conjuntos. Este proceso incluyó la eliminación de filas innecesarias en los datasets y la eliminación de columnas redundantes en ambos conjuntos de datos. Además, se verificó la correcta correspondencia entre las muestras y las variables, se eliminaron valores faltantes (NA) y se ajustaron las dimensiones de los dos datasets seleccionados para que coincidiesen.

- **Creación del objeto SummarizedExperiment**: una vez ajustadas las dimensiones de los datasets, se creó un dataframe para rowData, con la información sobre las muestras (filas), y otro para colData, con la información sobre las variables (columnas). A continuación, se generó el objeto SummarizedExperiment utilizando los paquetes S4Vectors y SummarizedExperiment e incorporando sus tres componentes: assays, rowData y colData. Tras la creación del objeto, se verificó y analizó la información contenida en él. Finalmente, se guardó el objeto en un archivo en formato binario (.Rda), el cual se subió posteriormente al repositorio de Github previamente creado.

- **Análisis exploratorio de los datos**: el análisis comenzó identificando las variables categóricas y numéricas en el conjunto de datos. Para las variables categóricas, se realizó un análisis de frecuencias y se visualizaron mediante gráficos de barras. Para las variables numéricas, se calcularon parámetros estadísticos como el mínimo, máximo, media, mediana y cuartiles. Se analizaron las distribuciones mediante histogramas y boxplots. Los histogramas permitieron observar la distribución de las mediciones y detectar distribuciones sesgadas, mientras que los boxplots ayudaron a identificar valores atípicos (outliers). Este análisis proporcionó una visión más detallada de las distribuciones y permitió seleccionar los metabolitos más relevantes para análisis posteriores.

- **Selección del metabolito de interés y análisis univariante**: se seleccionó la glucosa como metabolito de especial interés debido a su comportamiento atípico en los datos. Una vez hecho esto, se realizó un análisis univariante para evaluar la evolución de los niveles de glucosa a lo largo del tiempo (T0, T2, T4, T5). Para ello, se utilizaron resúmenes estadísticos y gráficos (histogramas y boxplots) que permitieron observar la tendencia y dispersión de los niveles de glucosa en los diferentes tiempos.

- **Análisis bivariado del metabolito de interés**: se compararon los niveles de glucosa entre los géneros y los tipos de cirugía bariátrica (tras la cirugía, T5) para determinar si estos factores podían influir en los niveles de glucosa. Para ello, se realizó una prueba de normalidad de Shapiro-Wilk sobre los niveles de glucosa en los diferentes grupos. Según los resultados, se seleccionaron las pruebas estadísticas apropiadas. Para los grupos que mostraron una distribución normal de los datos, se utilizó la prueba t de Student, mientras que para los grupos que no se ajustaron a la distribución normal, se aplicó el test de Wilcoxon.

Este enfoque metodológico permitió explorar y analizar cómo los diferentes factores asociados con la cirugía bariátrica afectan a los niveles de glucosa en los pacientes a lo largo del tiempo, proporcionando una comprensión más profunda de la respuesta metabólica a la intervención quirúrgica.

Herramientas estadísticas y bioinformáticas utilizadas

- **R**: programa de software para el procesamiento, visualización y análisis estadístico de los datos crudos procedentes del estudio metabolómico.
- **Paquete SummarizedExperiment**: utilizado para almacenar y gestionar los datos experimentales en un formato adecuado para datos ómicos. Este paquete proporciona una estructura organizada que facilita la integración de múltiples tipos de datos relacionados con experimentos biológicos (por ejemplo, metabolómica, transcriptómica). Los datos se almacenan de manera que se pueda acceder fácilmente a los valores de medición, las etiquetas de las muestras y la información adicional, como los metadatos asociados.

- **ggplot2**: paquete para la visualización gráfica de los datos. Se utilizó en la creación de gráficos de barras, boxplots y histogramas, lo que permitió una exploración visual de las distribuciones de los datos y la identificación de patrones importantes, como la presencia de outliers o distribuciones sesgadas.
- **Pruebas estadísticas en R (t.test, wilcox.test, shapiro.test):** estas pruebas fueron fundamentales para realizar análisis de normalidad, homogeneidad de varianzas y comparaciones de medias entre grupos. La prueba de Shapiro-Wilk (shapiro.test()) se utilizó para evaluar la normalidad de los datos y la prueba de Levene (leveneTest()) se utilizó para evaluar la homogeneidad de las varianzas entre los grupos, lo cual es crucial para determinar si se podían aplicar pruebas paramétricas, como el test t.El test t (t.test()) se aplicó para comparar medias entre grupos con distribuciones normales, mientras que el test de Wilcoxon (wilcox.test()) se empleó para realizar comparaciones entre grupos con distribuciones no normales.
- **Bioconductor**: conjunto de paquetes en R diseñados específicamente para el análisis de datos ómicos. Paquetes como S4Vectors y SummarizedExperiment fueron fundamentales en la creación y manipulación de objetos de datos estructurados. Bioconductor proporciona herramientas especializadas para trabajar con datos de alta dimensión, como los obtenidos mediante tecnologías de secuenciación, metabolómica y otras disciplinas ómicas, facilitando el análisis de datos biológicos complejos.

5. Resultados:

Análisis de los datos crudos: el análisis preliminar de los datos crudos reveló que las variables clave del estudio incluían el género, la edad, el tipo de cirugía a la que los pacientes fueron sometidos, y los niveles de varios metabolitos, incluyendo la glucosa.

Con estos datos, se creó el objeto SummarizedExperiment, el cual permitió organizar y estructurar la información de forma que fuera adecuada para el análisis estadístico posterior. Este objeto contenía los datos experimentales en un conjunto denominado counts, y metadatos en rowData y colData. En particular, los datos sobre la glucosa se seleccionaron para un análisis más profundo debido a su relevancia biológica en el contexto de la cirugía bariátrica.

Análisis exploratorio de los datos: el análisis exploratorio reveló que las variables del conjunto de datos eran predominantemente numéricas, con excepción de dos variables categóricas: "Surgery" (tipo de cirugía: "by pass" o "tubular") y "Gender" (género de los pacientes). La distribución de pacientes entre los dos tipos de cirugía mostró una mayor cantidad de pacientes en el grupo "by pass" (26 pacientes) en comparación con el grupo "tubular" (13 pacientes). Además, se observó una prevalencia notable de mujeres (27) en comparación con hombres (12) en el estudio.

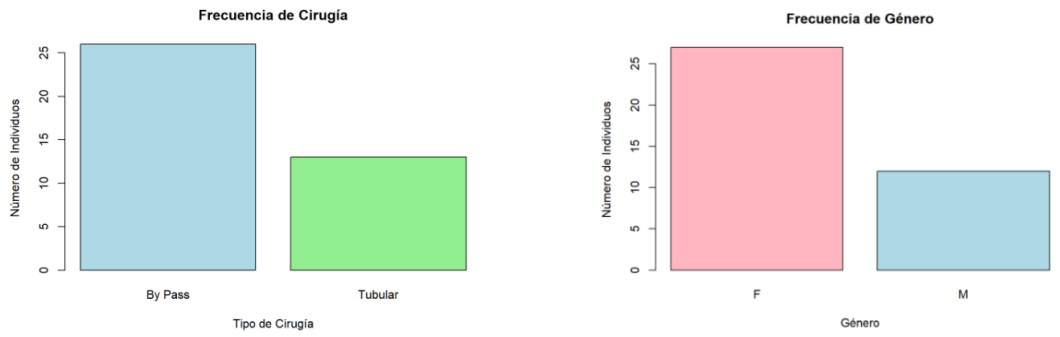


Figura 2. Gráfico de barras de las variables categóricas: tipo de cirugía y género

En cuanto al análisis de las variables numéricas, se observó que el metabolito glucosa (GLU) presentó una distribución sesgada en sus valores, especialmente antes de la cirugía (T0), lo que indicaba una mayor variabilidad y la presencia de valores atípicos que requerían un análisis más detallado. Los histogramas y boxplots de los niveles de glucosa a diferentes tiempos (T0, T2, T4 y T5) revelaron que los niveles eran elevados y variables antes de la cirugía, pero tendían a estabilizarse y mejorar con el tiempo postoperatorio.

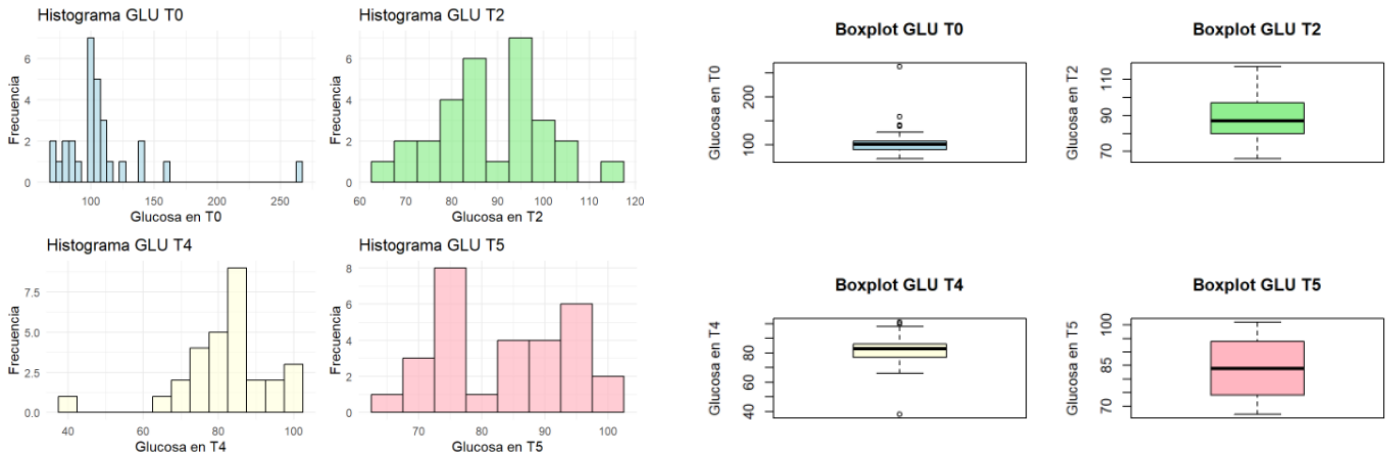


Figura 3. Histogramas y boxplots de los niveles de glucosa medidos a diferentes tiempos

El análisis univariante de los niveles de glucosa a los diferentes tiempos mostró los siguientes resultados:

- GLU_T0 (Antes de la cirugía): los niveles de glucosa en T0 presentaron una media de 108 y una mediana de 101, con un amplio rango entre 71 y 263, lo que indicaba la presencia de valores atípicos elevados. La distribución de los datos fue sesgada debido a estos valores extremos, sugiriendo que algunos pacientes podrían estar experimentando niveles elevados de glucosa, lo que podría estar relacionado con diabetes no controlada o descompensación metabólica asociada con la obesidad mórbida.

- GLU_T2 (2 semanas después de la cirugía): los niveles de glucosa en T2 presentaron una media de 88.62 y una mediana de 87, con un rango entre 66 y 117. Los datos estaban más concentrados en el rango intercuartílico (80-97) y no presentaban valores atípicos significativos, lo que indicaba una distribución más equilibrada y estable en comparación con los niveles en T0.

- GLU_T4 (4 meses después de la cirugía): los niveles de glucosa en T4 tuvieron una media de 81.9 y una mediana de 83, con un rango entre 38 y 101. Aunque la distribución fue ligeramente sesgada hacia valores más bajos, los datos mostraron una mayor concentración en el rango intercuartílico (77-86), lo que sugiere que la mayoría de los pacientes experimentaron una mejora en sus niveles de glucosa, con valores más cercanos al rango de normalidad.

- GLU_T5 (6 meses después de la cirugía): los niveles de glucosa en T5 tuvieron una media de 84.14 y una mediana de 84, con un rango de 67 a 101. Los datos mostraron una distribución equilibrada alrededor de la media, con un rango intercuartílico de 74 a 94, sin valores atípicos importantes. Esto indicó una estabilización de los niveles de glucosa en los pacientes, lo que sugiere una mejora sostenida en el control de la glucosa tras la cirugía bariátrica.

En resumen, los niveles de glucosa mostraron una disminución progresiva de T0 a T5, con una mayor variabilidad antes de la cirugía (T0), mientras que se estabilizaron y mejoraron después de la intervención. Este patrón sugiere que la cirugía bariátrica podría tener un efecto positivo en el control de la glucosa, lo que se alinea con la hipótesis de que la cirugía bariátrica mejora los parámetros metabólicos de los pacientes.

Al hacer un análisis bivariante para explorar posibles diferencias en los niveles de glucosa entre los géneros y entre los tipos de cirugía (by pass vs. tubular) tras la intervención, los resultados obtenidos fueron los siguientes:

- Para el grupo femenino, el test de Shapiro-Wilk mostró que los niveles de glucosa seguían una distribución normal ($p = 0.2716$), por lo que se utilizó la prueba t de Student para comparar las medias de glucosa entre hombres y mujeres. El valor p obtenido fue 0.8149, indicando que no existía una diferencia significativa entre las medias de glucosa entre mujeres y hombres. Para el grupo masculino, el test de Shapiro-Wilk indicó que los niveles de glucosa no seguían una distribución normal ($p = 0.02148$), por lo que se utilizó la prueba de Mann-Whitney. El valor p obtenido fue 0.7273, lo que también indicó que no existían diferencias significativas entre los géneros. En conjunto, los resultados de las pruebas estadísticas mostraron que no hubo diferencias significativas en los niveles de glucosa entre hombres y mujeres, aunque las medias de los dos grupos eran ligeramente diferentes.

- En cuanto a los tipos de cirugía, tanto el grupo de by pass como el de tubular presentaron distribuciones normales en los niveles de glucosa según el test de Shapiro-Wilk (p valor de 0,0435 para el grupo de pacientes sometido a cirugía by pass y p valor de 0,2871 para el grupo de pacientes sometidos a cirugía tubular).

La prueba de Levene indicó que las varianzas eran homogéneas entre los dos grupos (p = 0.582), por lo que se utilizó la prueba t de Student para comparar las medias de glucosa entre los dos tipos de cirugía. El valor p obtenido fue 0.6008, lo que sugirió que no existían diferencias significativas en los niveles de glucosa entre los pacientes sometidos a cirugía "by pass" y aquellos sometidos a cirugía "tubular".

En resumen, los resultados sugieren que el tipo de cirugía bariátrica no es un factor que afecte de manera significativa en los niveles de glucosa postquirúrgicos de los pacientes analizados.

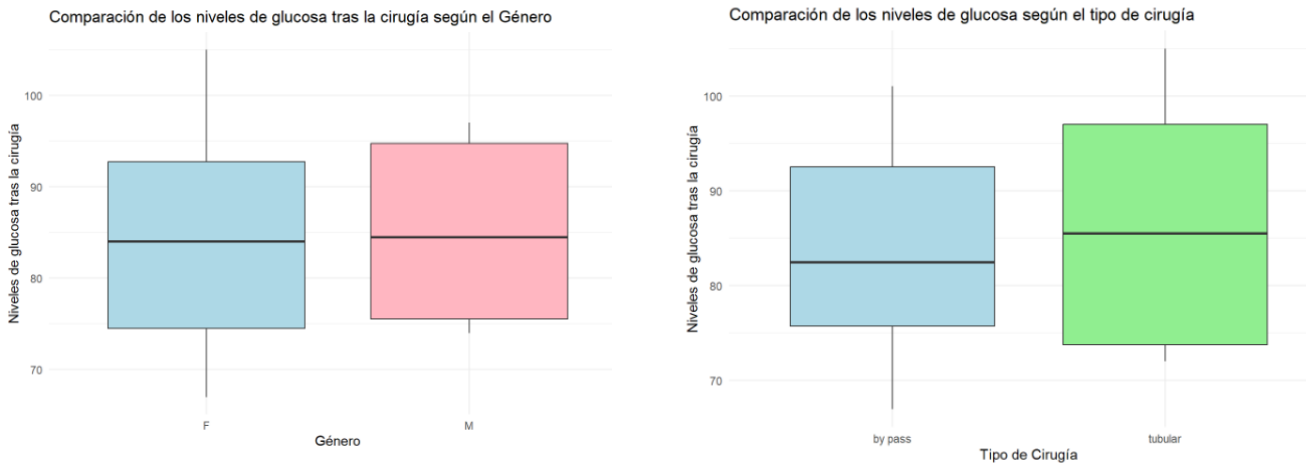


Figura 4. Boxplots para comparar los niveles de glucosa según el sexo (hombres y mujeres) y según el tipo de cirugía (by pass y tubular)

6. Discusión

Una de las principales limitaciones del presente estudio es el tamaño reducido de la muestra, con solo 39 pacientes. Este tamaño podría haber afectado la potencia estadística, dificultando la detección de diferencias significativas entre subgrupos, como los géneros o los tipos de cirugía (by pass vs. tubular). En estudios de este tipo, un tamaño de muestra pequeño aumenta el riesgo de cometer errores tipo II, es decir, no detectar diferencias reales cuando estas existen. Para obtener resultados más robustos y generalizables, sería fundamental contar con un número mayor de participantes, lo que proporcionaría una mayor potencia estadística y permitiría realizar comparaciones más precisas entre los diferentes grupos de estudio.

Otro factor importante a considerar son los posibles sesgos derivados de variables no controladas que podrían haber influido en los resultados. Por ejemplo, la adherencia al régimen postquirúrgico, que incluye la dieta y la actividad física, puede variar considerablemente entre los pacientes y tener un impacto significativo en los niveles de glucosa. Dado que la cirugía bariátrica es solo un componente de un tratamiento integral para la obesidad mórbida, aquellos pacientes que no sigan adecuadamente las indicaciones postoperatorias podrían experimentar una respuesta metabólica diferente, lo que podría haber alterado los resultados del estudio.

Además, factores como el tipo de dieta seguida, el nivel de actividad física o la presencia de comorbilidades adicionales (como diabetes tipo 2 e hipertensión) podrían haber influido en los resultados, y no se controlaron adecuadamente en este estudio. La falta de control sobre estas variables podría haber introducido sesgos, dificultando la interpretación de los efectos directos de la cirugía bariátrica sobre los niveles de glucosa.

Para mejorar la precisión de los resultados, sería necesario un diseño experimental más robusto que controle o registre adecuadamente estos factores. Además, dado que los efectos de la cirugía bariátrica pueden variar considerablemente entre individuos, un enfoque más personalizado que considere características metabólicas preexistentes podría proporcionar una comprensión más profunda de cómo esta intervención afecta a los pacientes de manera diferente.

En resumen, aunque los resultados obtenidos sugieren que la cirugía bariátrica puede tener un efecto positivo en la regulación de la glucosa, las limitaciones inherentes al diseño del estudio deben ser tomadas en cuenta. Para confirmar estos hallazgos y obtener una comprensión más clara de los mecanismos subyacentes de la mejora metabólica observada, es necesario realizar estudios adicionales con una muestra más grande, un control más exhaustivo de las variables externas y un seguimiento a largo plazo. De esta manera, se podrá validar con mayor precisión los efectos de la cirugía bariátrica en el metabolismo de los pacientes con obesidad mórbida.

7. Conclusiones

Los resultados obtenidos de este análisis proporcionan evidencia de que la cirugía bariátrica tiene un efecto positivo y estabilizador sobre los niveles de glucosa en sangre, ya que se observó una disminución progresiva en los niveles de glucosa de los pacientes tras la cirugía bariátrica. Sin embargo, no se encontraron diferencias significativas en los niveles de glucosa entre los géneros ni entre los tipos de cirugía (by pass vs. tubular). Estos resultados sugieren que, independientemente del tipo de cirugía realizada y del sexo, los pacientes mostraron una mejora en el control de la glucosa postoperatoria, lo que resalta el impacto favorable de la cirugía bariátrica sobre los parámetros metabólicos de los pacientes con obesidad mórbida.

A pesar de que se necesitan más estudios, nuestros resultados abren una nueva hipótesis en el estudio de la obesidad y proporcionan una visión integral de los cambios metabólicos después de la cirugía.

8. Referencias.

URL del repositorio de GitHub: https://github.com/anitamr/Martin-Ramirez-Ana-PEC1.git

Artículo de estudio: Palau-Rodriguez M, Tulipani S, Marco-Ramell A, Miñarro A, Jáuregui O, Sanchez-Pla A, Ramos-Molina B, Tinahones FJ, Andres-Lacueva C. Metabotypes of response to bariatric surgery independent of the magnitude of weight loss. PLoS One. 2018 Jun 1;13(6):e0198214. doi: 10.1371/journal.pone.0198214. PMID: 29856816; PMCID: PMC5983508).

Bioconductor: (Bioconductor: Open Source Software for Bioinformatics." Bioconductor. Accessed. March 25, 2025. https://www.bioconductor.org/.

ANEXO I: código de R usado para el procesamiento y análisis de los datos crudos

1. Carga de datos crudos: una vez seleccionado el conjunto de datos del estudio metabolómico desde GitHub, se cargaron los archivos de datos y metadatos en R, y se verificaron las dimensiones y el contenido de cada archivo.

```
# Cargamos los datos de las mediciones (metabolitos)
DataValues_S013 <- read.csv("DataValues_S013.csv", header = TRUE)
# Visualizamos las dimensiones y columnas
head(DataValues_S013)

dim(DataValues_S013)

## [1]  39 696

head(colnames(DataValues_S013))

## [1] "X.1"      "SUBJECTS" "SURGERY"  "AGE"      "GENDER"   "Group"

# Cargamos los metadatos
DataInfo_S013 <- read.csv("DataInfo_S013.csv", header = TRUE)
# Visualizamos las primeras filas, dimensiones y columnas
head(DataInfo_S013)

##         X  VarName    varTpe Description
## 1 SUBJECTS SUBJECTS   integer    dataDesc
## 2  SURGERY  SURGERY character    dataDesc
## 3      AGE      AGE   integer    dataDesc
## 4   GENDER   GENDER character    dataDesc
## 5    Group    Group   integer    dataDesc
## 6 MEDDM_T0 MEDDM_T0   integer    dataDesc

dim(DataInfo_S013)

## [1] 695   4

head(colnames(DataInfo_S013))

## [1] "X"           "VarName"     "varTpe"      "Description"
```

Tras cargar el conjunto de datos “DataValues_S013.csv” en R, se observó que contenía los datos crudos correspondientes al estudio metabolómico que se está analizando. Este archivo tiene un total de 39 filas y 696 columnas. Cada fila representa un paciente del estudio metabolómico, mientras que las columnas incluyen diversas características o mediciones de los sujetos, tales como edad, género, tipo de cirugía a la que fueron sometidos, y niveles de varios metabolitos analizados, como hierro, transferrina, colesterol HDL, colesterol LDL, glucosa, entre otros.

Adicionalmente, al cargar el archivo “DataInfo_S013.csv”, se verificó que contenía los metadatos del estudio. Este archivo tiene un total de 695 filas y 4 columnas. Cada fila representa una variable relacionada con las características o los niveles de los metabolitos de los sujetos en un momento determinado (por ejemplo, edad, género o niveles de algún metabolito). Las columnas de este archivo son las siguientes:

- X: número de identificación o índice.
- VarName: nombre de la variable o columna que describe algún atributo o medida relacionada con los sujetos del estudio.
- varTpe: tipo de datos de la variable (puede ser integer, character o numeric).
- Description: explica la naturaleza de los datos correspondientes a esa variable

2. Procesamiento y alineación de datos y metadatos: dado que el dataset "DataValues_S013" tiene 39 filas (muestras) y el dataset "DataInfo_S013" tiene 695 filas (metadatos de las variables), fue necesario ajustar ambos archivos para garantizar que estuvieran correctamente alineados. Las filas de los datos de "DataValues_S013" debían coincidir con las de "DataInfo_S013", por lo que se recortó "DataInfo_S013" para que tuviese 39 filas. Además, se eliminó la primera columna de ambos dataset, puesto que en ambos casos la información era redundante y los valores faltantes Na.

```
# Eliminamos la primera columna redundante de ambos datasets
DataValues_S013 <- DataValues_S013[, -1]
DataInfo_S013 <- DataInfo_S013[, -1]

# Recortamos DataInfo_S013 a 39 filas
DataInfo_S013 <- DataInfo_S013[1:39, ]

# Verificamos las dimensiones
dim(DataInfo_S013)  # Debe ser (39, 3)

## [1] 39  3

dim(DataValues_S013)  # Debe ser (39, 695)

## [1]  39 695

# Contamos los valores perdidos
sum(is.na(DataValues_S013))

## [1] 3390

# Limpiamos los datos eliminando las filas con NA
DataValues_S013_clean <- na.omit(DataValues_S013)
```

3. Creación del objeto SummarizedExperiment

   A continuación, creamos un objeto SummarizedExperiment a partir de los datos y metadatos. Un objeto SummarizedExperiment es una estructura de datos utilizada principalmente en bioinformática y análisis de datos ómicos, como los estudios de transcriptómica, metabolómica, genómica, entre otros. Este tipo de objeto es utilizado para almacenar y manejar datos de múltiples muestras de manera eficiente y flexible, y es parte del paquete SummarizedExperiment en Bioconductor.

Este objeto consta de 3 componentes principales:

- Row Data (datos de filas): son los metadatos asociados con las filas de los datos en el objeto summarizedexperiment. Los datos de las filas se almancenan generalmente en un Dataframe.

- Assays: es un contendedor de matrices de datos numéricos, donde se almacenan las mediciones cuantitativas de los experimentos.

- Col Data (datos de columnas): son los metadatos asociados con las columnas del objeto summarizedexperiment, es decir, las muestras.Estos metadatos podrían incluir información sobre las muestras, como el tratamiento, la edad del paciente, o cualquier otra característica relevante de las muestras. Al igual que los datos de las filas, estos metadatos se almacenan en un DataFrame.

El objeto summarizedExperiment presenta varias diferencias con respecto al objeto expressionset:

- ExpresssionSet: se utiliza principalmente para almacenar datos de expresión genética (como microarrays o RNA-Seq). Su estructura incluye una matriz de expresión (o "matriz de datos") junto con un conjunto de metadatos relacionados con las muestras (filas) y las características (columnas).Los metadatos de las filas (muestras) y columnas (genes o características) se almacenan en objetos separados, como pData (para muestras) y fData (para características). Originalmente fue diseñado para el análisis de expresión génica, lo que lo hace menos adecuado para datos que no sean de expresión génica.

- SummarizedExperiment: es una extensión más generalizada que puede usarse para una variedad de tipos de datos, no solo expresión genética. Además de almacenar la matriz de datos, también almacena información asociada con las filas y las columnas en formato de DataFrame, proporcionando una forma más flexible de manejar datos de múltiples tipos (por ejemplo, metabolómica, genómica, proteómica, etc.). Almacena los metadatos de las filas y columnas dentro de un solo objeto de tipo DataFrame (para las filas y las columnas), lo que proporciona una mayor flexibilidad en el manejo de los datos. a sido diseñado para manejar cualquier tipo de datos experimentales en forma de tablas, incluidos datos de metabolómica, proteómica, y genómica, y ofrece una mayor flexibilidad.

Una vez ajustadas las dimensiones de los ficheros, creamos un dataframe para rowData con la información sobre las muestras(filas) y otro para colData con la información sobre las variables (columnas). Una vez hecho esto, instalamos el paquete SummarizedExperiment y creamos el objeto SummarizedExperiment con sus 3 elementos: assays, rowData y colData.

```
# Instalamos y cargamos paquetes necesarios
BiocManager::install("S4Vectors")

BiocManager::install("SummarizedExperiment")

library(SummarizedExperiment)

library(S4Vectors)

# Creamos un DataFrame para rowData (información sobre las muestras)
rowData <- DataFrame(DataInfo_S013)

# Creamos un DataFrame para colData (información sobre las variables/metabolitos)
colData <- DataFrame(VarName = colnames(DataValues_S013))

# Creamos el objeto SummarizedExperiment
se <- SummarizedExperiment(
  assays = list(counts = as.matrix(DataValues_S013)),  # Datos de las mediciones
  rowData = rowData,  # Información sobre las muestras
  colData = colData   # Información sobre las variables
)
# Visualizamos el objeto creado
se

## class: SummarizedExperiment
## dim: 39 695
## metadata(0):
## assays(1): counts
## rownames: NULL
## rowData names(3): VarName varTpe Description
```

```
## colnames(695): SUBJECTS SURGERY ... SM.C24.0_T5 SM.C24.1_T5
## colData names(1): VarName
```

Una vez creado el objeto SummarizedExperiment, comprobamos que tenía la información correcta revisando sus componentes principales y verificando las dimensiones del objeto. Por último, guardamos el objeto en un archivo.Rda para poder subirlo posteriormente al repositorio de Github creado anteriormente.

```
# Visualizamos los primeros metadatos de las filas (rowData)
head(rowData(se))
```

```
## DataFrame with 6 rows and 3 columns
##       VarName      varTpe Description
##   <character> <character> <character>
## 1    SUBJECTS     integer    dataDesc
## 2     SURGERY   character    dataDesc
## 3         AGE     integer    dataDesc
## 4      GENDER   character    dataDesc
## 5       Group     integer    dataDesc
## 6    MEDDM_T0     integer    dataDesc
```

```
# Visualizamos los primeros metadatos de las columnas (colData)
head(colData(se))
```

```
## DataFrame with 6 rows and 1 column
##              VarName
##          <character>
## SUBJECTS    SUBJECTS
## SURGERY      SURGERY
## AGE              AGE
## GENDER        GENDER
## Group          Group
## MEDDM_T0    MEDDM_T0
```

```
# Visualizamos las dimensiones del objeto SummarizedExperiment
dim(se)
```

```
## [1]  39 695
```

```
# Verificamos que el objeto que hemos creado se encuentra el directorio de trabajo actual
getwd()
```

```
## [1] "C:/Users/bienv/Desktop/análisis de datos ómicos/PEC1"
```

```
# Guardamos el objeto SummarizedExperiment en formato binario .Rda para luego subirlo a git hub
save(se, file = "summarized_experiment.Rda")
cat("El objeto SummarizedExperiment se ha guardado en 'summarized_experiment.Rda'.")
```

```
## El objeto SummarizedExperiment se ha guardado en 'summarized_experiment.Rda'.
```

```
# Cargamos el archivo que hemos generado para asegurarnos de que todo está bien guardado
load("summarized_experiment.Rda")
```

Una vez creado el objeto SummarizedExperiment lo analizamos detenidamente y vimos que:

- Tiene un total de 39 filas y 695 columnas, lo que indica que contiene 39 muestras(pacientes) y 695 variables (metabolitos u otros atributos medidos), lo que coincide con la descripción del conjunto de datos

- Contiene 1 conjunto de datos en el componente assays, llamado counts. Esto sugiere que los datos experimentales, como los niveles de los metabolitos o las mediciones, están almacenados en este conjunto de datos bajo el nombre counts.

- Contiene el dataframe Rowdata y el dataframe colData. El dataframe Rowdata tiene 3 columnas que proporcionan información sobre las muestras como el nombre de la variable o atributo que se está estudiando (VarName), el tipo de variable (varTpe) y la descripción de la variable (dataDesc). El dataframe colData tiene una sola columna llamada VarName que almacena los nombres de las columnas de DataValues_S013, es decir, las variables o metabolitos evaluados.

- No contiene metadatos adicionales almacenados en el objeto creado, lo que podría significar que no se han añadido metadatos adicionales fuera de los rowData y colData

4. Análisis exploratorio del conjunto de datos

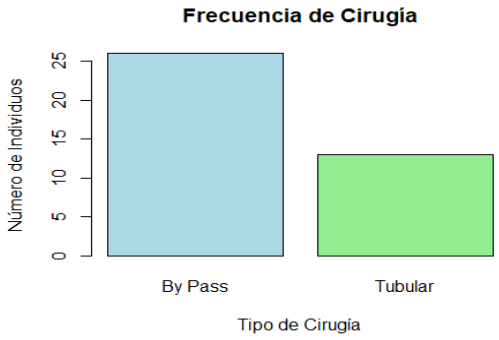
En primer lugar, se identificaron el tipo de variables (numéricas y categóricas) que componían el dataset "DataValues_S013"

```
# Obtener el tipo de cada columna en el dataset
column_types <- sapply(DataValues_S013_clean, class)

# Identificar las variables numéricas
numeric_vars <- names(column_types[column_types == "numeric" | column_types == "integer
"])

# Identificar las variables categóricas (factores o caracteres)
categorical_vars <- names(column_types[column_types == "factor" | column_types == "charact
er"])

cat("Variables categóricas:\n")

## Variables categóricas:

cat(paste(categorical_vars, collapse = ", "), "\n")

## SURGERY, GENDER
```

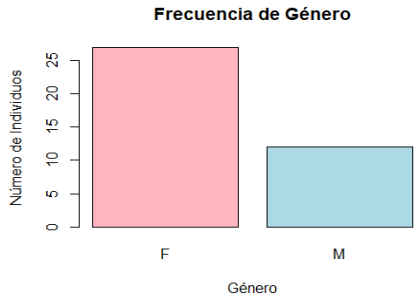
Se vio que todas las variables del dataset eran numéricas a excepción de dos variables categóricas: "SURGERY" y "GENDER". Para las variables categóricas, se realizó un análisis de frecuencias y se visualizaron mediante gráficos de barras.

```
# Análisis de frecuencia para las variables categóricas
table(DataValues_S013$SURGERY)

##
## by pass tubular
##      26      13

table(DataValues_S013$GENDER)
##  F  M
## 27 12
```

```r
# Representación gráficas de las variables categóricas
# Frecuencia de la variable 'SURGERY'
surgey_counts <- table(DataValues_S013$SURGERY)

# Crear gráfico de barras para la variable 'SURGERY'
barplot(surgey_counts,
        main = "Frecuencia de Cirugía",
        xlab = "Tipo de Cirugía",
        ylab = "Número de Individuos",
        col = c("lightblue", "lightgreen"),
        names.arg = c("By Pass", "Tubular"))
```




```r
# Frecuencia de la variable 'GENDER'
gender_counts <- table(DataValues_S013$GENDER)

# Crear gráfico de barras para la variable 'GENDER'
barplot(gender_counts,
        main = "Frecuencia de Género",
        xlab = "Género",
        ylab = "Número de Individuos",
        col = c("lightpink", "lightblue"),
        names.arg = c("F", "M"))
```



En cuanto a las variables categóricas “Surgery” y “gender”, se observó lo siguiente: la frecuencia de pacientes sometidos a cirugía de tipo “by pass” fue considerablemente mayor que la de aquellos que se sometieron a cirugía “tubular”, con 26 pacientes en el primer grupo y solo 13 en el segundo. Además, en el grupo de pacientes analizado, se registró una mayor cantidad de mujeres en comparación con los hombres, con 27 mujeres frente a 12 hombres.

Para las variables numéricas, se calcularon parámetros estadísticos como el mínimo, máximo, media, mediana y cuartiles. Además, se analizaron las distribuciones mediante histogramas y boxplots.

```r
# Cargamos la librería ggplot
library(ggplot2)

# Filtramos las variables numéricas
numeric_data <- DataValues_S013[sapply(DataValues_S013, is.numeric)]

# Resumen estadístico (mínimo, máximo, media, mediana, cuartiles)
summary_stats <- summary(numeric_data)
head(summary_stats)
```

Los histogramas permitieron observar la distribución de las mediciones y detectar distribuciones sesgadas, mientras que los boxplots ayudaron a identificar valores atípicos (outliers).

5. Selección del metabolito de interés y análisis univariado

Tras visualizar los histogramas y boxplots de todas las variables numéricas del dataset, se eligió como metabolito la glucosa para continuar con el análisis exploratorio (análisis univariado y bivariado), puesto que la distribución de dicho metabolito era sesgada y aparecían outliers en su boxplot, lo que indicaba que podría tener un comportamiento particular o estar influenciada por factores externos que merecían un análisis más detallado. Esta elección se basó en la relevancia biológica de la glucosa en el contexto de la cirugía bariátrica, ya que se esperaba que la intervención quirúrgica tuviera un impacto significativo sobre sus niveles en sangre. A partir de esta selección, se procedió a realizar un análisis univariado de la glucosa, para evaluar su comportamiento a lo largo del tiempo y determinar si existían cambios significativos en sus niveles en los distintos momentos temporales del estudio.

```r
# Cargamos las librerías necesarias
library(ggplot2)
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.4.2
```

```
##
## Adjuntando el paquete: 'gridExtra'
```

```
## The following object is masked from 'package:Biobase':
##
##     combine
```

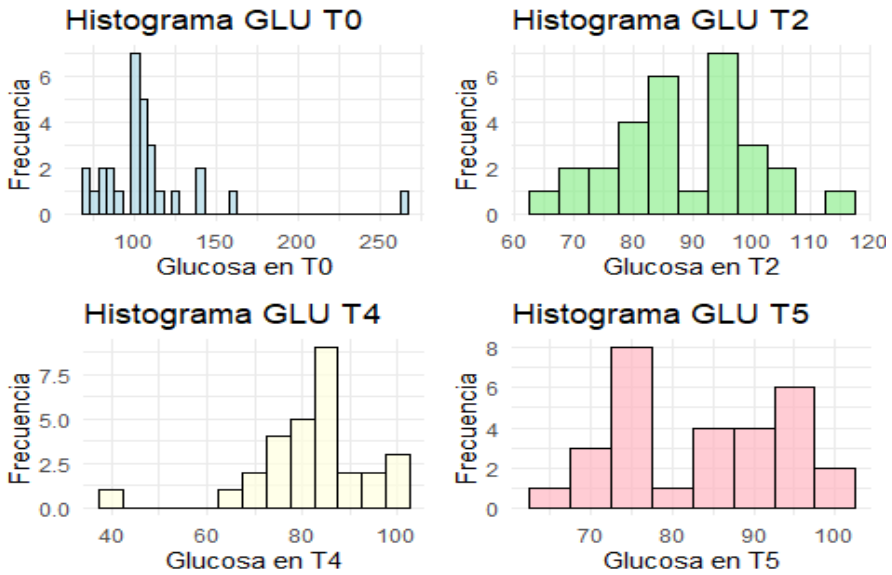
```
## The following object is masked from 'package:BiocGenerics':
##
##     combine
```

```r
# Cargamos los datos de las mediciones (metabolitos)
DataValues_S013 <- read.csv("DataValues_S013.csv", header = TRUE)

# Eliminamos las filas con valores NA en las columnas de glucosa (GLU_T0, GLU_T2, GLU_T4, GLU_T5)
DataValues_S013_clean <- na.omit(DataValues_S013[, c("GLU_T0", "GLU_T2", "GLU_T4", "GLU_T5")])

# Resumen estadístico de la variable GLU en T0, T2, T4, T5 (sin NA)
summary(DataValues_S013_clean$GLU_T0)
```

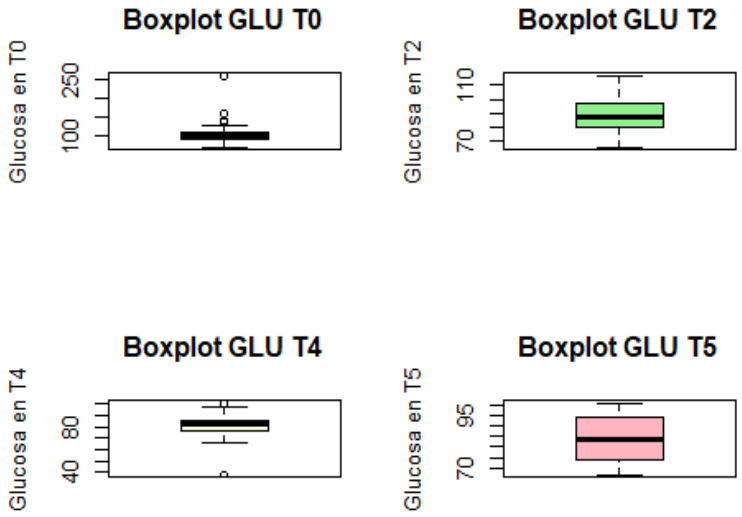
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      71      90     101     108     108     263
```

```r
summary(DataValues_S013_clean$GLU_T2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   66.00   80.00   87.00   88.62   97.00  117.00

summary(DataValues_S013_clean$GLU_T4)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    38.0    77.0    83.0    81.9    86.0   101.0

summary(DataValues_S013_clean$GLU_T5)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   67.00   74.00   84.00   84.14   94.00  101.00

# Crear los histogramas para GLU a diferentes tiempos con ggplot2
hist_T0 <- ggplot(DataValues_S013_clean, aes(x = GLU_T0)) +
  geom_histogram(binwidth = 5, fill = "lightblue", color = "black", alpha = 0.7) +
  labs(title = "Histograma GLU T0", x = "Glucosa en T0", y = "Frecuencia") +
  theme_minimal()

hist_T2 <- ggplot(DataValues_S013_clean, aes(x = GLU_T2)) +
  geom_histogram(binwidth = 5, fill = "lightgreen", color = "black", alpha = 0.7) +
  labs(title = "Histograma GLU T2", x = "Glucosa en T2", y = "Frecuencia") +
  theme_minimal()

hist_T4 <- ggplot(DataValues_S013_clean, aes(x = GLU_T4)) +
  geom_histogram(binwidth = 5, fill = "lightyellow", color = "black", alpha = 0.7) +
  labs(title = "Histograma GLU T4", x = "Glucosa en T4", y = "Frecuencia") +
  theme_minimal()

hist_T5 <- ggplot(DataValues_S013_clean, aes(x = GLU_T5)) +
  geom_histogram(binwidth = 5, fill = "lightpink", color = "black", alpha = 0.7) +
  labs(title = "Histograma GLU T5", x = "Glucosa en T5", y = "Frecuencia") +
  theme_minimal()

# Mostrar los 4 histogramas en una cuadrícula 2x2
grid.arrange(hist_T0, hist_T2, hist_T4, hist_T5, ncol = 2)
```



```r
# Boxplot para GLU en diferentes tiempos
par(mfrow = c(2, 2))  # Crear una cuadrícula de 2x2 para mostrar los gráficos

# Boxplot para GLU_T0
boxplot(DataValues_S013_clean$GLU_T0, main = "Boxplot GLU T0", ylab = "Glucosa en T0"
, col = "lightblue")

# Boxplot para GLU_T2
boxplot(DataValues_S013_clean$GLU_T2, main = "Boxplot GLU T2", ylab = "Glucosa en T2"
, col = "lightgreen")

# Boxplot para GLU_T4
boxplot(DataValues_S013_clean$GLU_T4, main = "Boxplot GLU T4", ylab = "Glucosa en T4"
, col = "lightyellow")

# Boxplot para GLU_T5
boxplot(DataValues_S013_clean$GLU_T5, main = "Boxplot GLU T5", ylab = "Glucosa en T5"
, col = "lightpink")
```



Tras el análisis estadístico de los niveles de glucosa en diferentes momentos (T0, T2, T4 y T5), se obtuvieron las siguientes conclusiones:

GLU_T0 (nivel de glucosa antes de la cirugía) presentó una media de 108 y una mediana de 101. El rango fue amplio, con un valor mínimo de 71 y un valor máximo de 263, lo que sugirió la presencia de valores atípicos. La distribución mostró una ligera asimetría hacia la derecha debido a estos valores extremos.

GLU_T2 (nivel de glucosa en T2) tuvo una media de 88.62 y una mediana de 87, con un rango de 66 a 117. Los valores estuvieron más concentrados en el rango intercuartílico (80-97), sin valores atípicos significativos, lo que indicó una distribución más equilibrada y estable.

GLU_T4 (nivel de glucosa en T4) mostró una media de 81.9 y una mediana de 83, con un rango de 38 a 101. La distribución pareció estar sesgada hacia valores más bajos, especialmente con el valor mínimo de 38. Sin embargo, el rango intercuartílico fue estrecho (77-86), lo que sugirió que la mayoría de los datos estaban agrupados en torno a un rango medio.

GLU_T5 (nivel de glucosa tras la cirugía) tuvo una media de 84.14 y una mediana de 84, con un rango de 67 a 101. Los datos estuvieron bien distribuidos alrededor de la media, con un rango intercuartílico de 74 a 94, lo que indicó una distribución equilibrada y sin valores atípicos importantes.

En resumen, los niveles de glucosa mostraron una ligera disminución de T0 a T5, con una mayor variabilidad en T0 y valores extremos, mientras que los niveles de glucosa en T2, T4 y T5 estuvieron más concentrados y equilibrados. Este patrón sugirió que los niveles de glucosa se estabilizaron o mejoraron después de la cirugía bariátrica.

Tras la observación de los histogramas y boxplot correspondientes a los niveles de glucosa en diferentes momentos (T0, T2, T4 y T5), se obtuvieron las siguientes conclusiones: - Antes de la cirugía (T0), los niveles de glucosa son más variables, con algunos pacientes mostrando niveles elevados que podrían ser indicativos de diabetes no controlada o descompensación metabólica, probablemente derivada de la condición patológica de obesidad mórbida de dichos pacientes.

- Tras la cirugía (T2 y T5), los niveles de glucosa tienden a estabilizarse, mostrando una distribución más equilibrada sin valores extremos. Esto sugiere que la cirugía bariátrica podría estar teniendo un efecto positivo en el control de la glucosa, favoreciendo una mejora metabólica en los pacientes.

6. Análisis bivariado del metabolito de interés Se realizó un análisis bivariado para comparar los niveles de glucosa entre los géneros y los tipos de cirugía. En este caso, comparamos los niveles de glucosa a tiempo 5 (tras la cirugía) según el sexo de los pacientes y según el tipo de cirugía a la que fueron sometidos.

En primer lugar, se realizó la prueba de normalidad de Shapiro-Wilk para verificar si los niveles de glucosa en los diferentes grupos de sexo seguían una distribución normal.

```
# Cargamos el paquete ggplot
library(ggplot2)
# Filtrar los datos de glucosa y género
glucose_data <- DataValues_S013$GLU_T5
gender_data <- DataValues_S013$GENDER

# Comprobar si los datos de glucosa siguen una distribución normal (opcional, pero recomenda
do)
shapiro.test(glucose_data[gender_data == "F"])  # Normalidad para el grupo femenino

##
##  Shapiro-Wilk normality test
##
## data:  glucose_data[gender_data == "F"]
## W = 0.95005, p-value = 0.2716

shapiro.test(glucose_data[gender_data == "M"])  # Normalidad para el grupo masculino

##
##  Shapiro-Wilk normality test
##
## data:  glucose_data[gender_data == "M"]
## W = 0.7884, p-value = 0.02148
```

Para el grupo femenino, el valor de W obtenido en el test de Shapiro-Wilk fue 0.95005, con un p-valor de 0.2716. Dado que el p-valor es mayor que 0.05, no se rechaza la hipótesis nula de normalidad. Esto sugiere que los niveles de glucosa en el grupo femenino siguen una distribución normal, indicando que los datos para este grupo son simétricos y no presentan una desviación significativa de la normalidad.

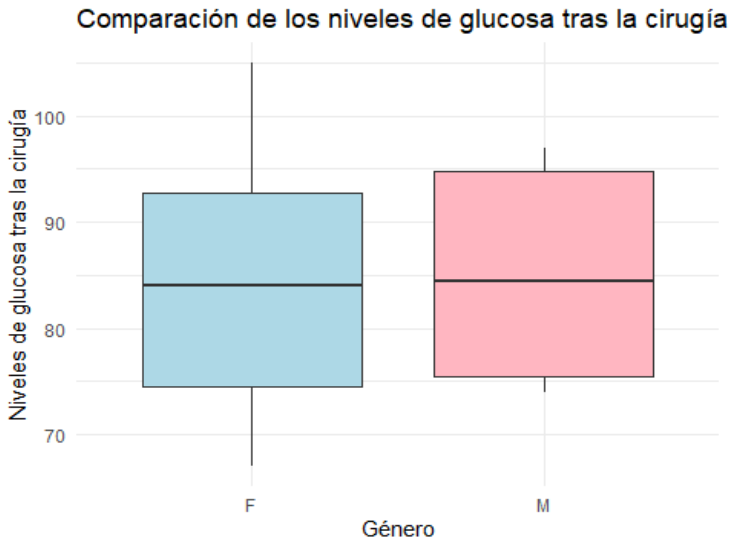
Para el grupo masculino, el valor de W obtenido fue 0.7884, y el p-valor fue 0.02148. En este caso, el p-valor es menor que 0.05, lo que implica que se rechaza la hipótesis nula de normalidad. Esto indica que los niveles de glucosa en el grupo masculino no siguen una distribución normal, lo que sugiere que los datos pueden ser asimétricos, sesgados o presentar otras características que impiden que sigan una distribución normal.

Se utilizaron diferentes pruebas según la normalidad de los datos. Para el grupo femenino, que muestra una distribución normal, se utilizó la prueba t de Student. Para el grupo masculino, no normal, se aplicó la prueba de Mann-Whitney.

```
# Comparar los niveles de glucosa entre géneros usando un test t para los datos normales del g
rupo femenino
resultado_test_t <- t.test(glucose_data[gender_data == "F"], glucose_data[gender_data == "M"]
,
                           alternative = "two.sided", var.equal = TRUE)
# Mostramos el resultado del test
resultado_test_t

##
##  Two Sample t-test
##
## data:  glucose_data[gender_data == "F"] and glucose_data[gender_data == "M"]
## t = -0.23622, df = 30, p-value = 0.8149
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.449420   8.282753
## sample estimates:
## mean of x mean of y
##  84.04167  85.12500

# Comparar los niveles de glucosa entre géneros usando una prueba no paramétrica para el gr
upo masculino
resultado_test_wilcoxon <- wilcox.test(glucose_data[gender_data == "F"], glucose_data[gender
_data == "M"],
                                       alternative = "two.sided")  # Prueba de Mann-Whitney

## Warning in wilcox.test.default(glucose_data[gender_data == "F"],
## glucose_data[gender_data == : cannot compute exact p-value with ties

# Mostramos el resultado
resultado_test_wilcoxon

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  glucose_data[gender_data == "F"] and glucose_data[gender_data == "M"]
## W = 87.5, p-value = 0.7273
## alternative hypothesis: true location shift is not equal to 0

# Gráfico de boxplot para visualizar la distribución de GLU_T5 por género
ggplot(DataValues_S013, aes(x = GENDER, y = GLU_T5, fill = GENDER)) +
  geom_boxplot() +
  labs(title = "Comparación de los niveles de glucosa tras la cirugía según el Género",
       x = "Género",
       y = "Niveles de glucosa tras la cirugía") +
  scale_fill_manual(values = c("lightblue", "lightpink")) +
```

```
  theme_minimal() +
  theme(legend.position = "none")
```

```
## Warning: Removed 7 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



- El valor p (0.8149) obtenido para las mujeres era considerablemente mayor que el umbral de significancia de 0.05, lo que indicaría que no se puede rechazar la hipótesis nula de que las medias de los niveles de glucosa entre hombres y mujeres son iguales. Además, el intervalo de confianza al 95% incluyó el valor 0, lo que refuerza la conclusión de que no existe una diferencia significativa en las medias de glucosa entre los géneros en la muestra analizada.

- El valor p (0.7273) obtenido para los hombres era considerablemente mayor que 0.05, lo que sugiere que no hay una diferencia significativa entre los niveles de glucosa en hombres y mujeres.

Por tanto, aunque las medias de los dos grupos sean ligeramente diferentes (84.04 frente a 85.13), los resultados no proporcionan evidencia suficiente para concluir que esta diferencia sea estadísticamente significativa, ya que los valores p obtenidos (0.8149 y 0.7273, respectivamente) son mucho mayores que el umbral de significancia común de 0.05.

Después, se realizó la prueba de normalidad de Shapiro-Wilk para verificar si los niveles de glucosa en los diferentes grupos de cirugía seguían una distribución normal. También se realizó la prueba de levene para evaluar si las varianzas entre grupos eran homogéneas

```
# Cargar librerías necesarias
library(ggplot2)


# Filtrar los datos según la columna 'SURGERY'
data_bypass <- DataValues_S013[DataValues_S013$SURGERY == "by pass", ]
data_tubular <- DataValues_S013[DataValues_S013$SURGERY == "tubular", ]
# Verificamos normalidad de los datos y homogeneidad de varianzas

shapiro.test(data_bypass$GLU_T5)  # Para el grupo by pass
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_bypass$GLU_T5
## W = 0.92472, p-value = 0.07435
```

```r
shapiro.test(data_tubular$GLU_T5)  # Para el grupo tubular
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_tubular$GLU_T5
## W = 0.89968, p-value = 0.2871
```

```r
# Instalamos la librería necesaria
library(car)
```

```
## Warning: package 'car' was built under R version 4.4.2
```

```
## Cargando paquete requerido: carData
```

```
## Warning: package 'carData' was built under R version 4.4.2
```

```r
# Realizamos la prueba de Levene
leveneTest(GLU_T5 ~ SURGERY, data = DataValues_S013)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.0965 0.7582
##       30
```

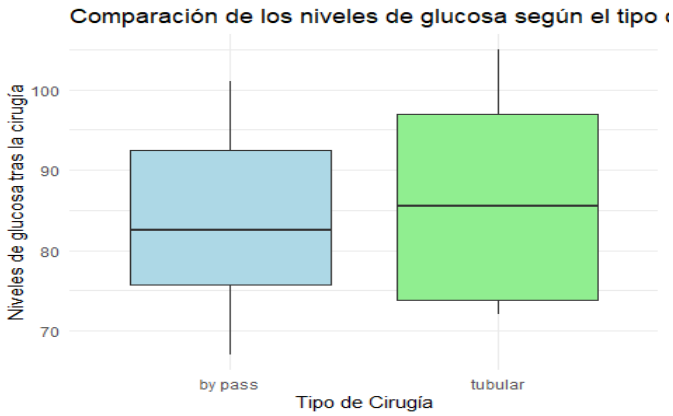
El valor p (0.0435) obtenido para el grupo de pacientes sometidos a cirugía “by pass” era mayor que el umbral común de significancia de 0.05. Esto significa que no se puede rechazar la hipótesis nula de que los datos siguen una distribución normal. El valor p (0.2871) obtenido para el grupo de pacientes sometidos a cirugía “tubular” también es mayor que 0.05, lo que indica que no se puede rechazar la hipótesis nula de normalidad. En consecuencia, se puede concluir que los niveles de glucosa tanto en el grupo de pacientes sometidos a cirugía “by pass” como en aquellos sometidos a cirugía “tubular” siguen una distribución normal, lo que permite usar métodos estadísticos paramétricos en análisis posteriores. Además, el valor p obtenido tras realizar la prueba de Levene fue de 0.582 (superior a 0.05), lo que indicaría que las varianzas de los niveles de glucosa tras la cirugía son homogéneas entre ambos grupos, por lo que se podría usar la prueba t de Student para comparar las medias de glucosa entre los dos grupos.

```r
# Cargar librerías necesarias
library(ggplot2)

# Filtrar los datos según la columna 'SURGERY'
data_bypass <- DataValues_S013[DataValues_S013$SURGERY == "by pass", ]
data_tubular <- DataValues_S013[DataValues_S013$SURGERY == "tubular", ]

# Extraemos los valores de glucosa (GLU_T5) para cada tipo de cirugía
glu_bypass <- data_bypass$GLU_T5
glu_tubular <- data_tubular$GLU_T5

# Realizar prueba t para comparar los niveles de glucosa en T5 entre los dos tipos de cirugía
t_test_result <- t.test(glu_bypass, glu_tubular, alternative = "two.sided", var.equal = TRUE)

# Mostrar el resultado de la prueba t
t_test_result
```

```
##
##  Two Sample t-test
##
## data:  glu_bypass and glu_tubular
## t = -0.52891, df = 30, p-value = 0.6008
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.74805   6.91472
## sample estimates:
## mean of x mean of y
##  83.70833  86.12500
```

```r
# Gráfico de boxplot para visualizar la distribución de GLU_T5 por tipo de cirugía
ggplot(DataValues_S013, aes(x = SURGERY, y = GLU_T5, fill = SURGERY)) +
  geom_boxplot() +
  labs(title = "Comparación de los niveles de glucosa según el tipo de cirugía",
       x = "Tipo de Cirugía",
       y = "Niveles de glucosa tras la cirugía") +
  scale_fill_manual(values = c("lightblue", "lightgreen")) +
  theme_minimal() +
  theme(legend.position = "none")
```

```
## Warning: Removed 7 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



El valor p obtenido en la prueba t de dos muestras fue de 0.6008, lo que es considerablemente mayor que el umbral común de significancia de 0.05. Esto significa que no se rechaza la hipótesis nula, lo que sugiere que no existe una diferencia significativa en las medias de los niveles de glucosa entre los dos grupos (By pass y Tubular). Por tanto, en base a los resultados obtenidos, podemos decir que no hay evidencia suficiente para afirmar que los niveles de glucosa tras la cirugía difieren significativamente entre los dos tipos de cirugía (“By pass” y “Tubular”). A pesar de que las medias de los dos grupos son ligeramente diferentes (83.71 vs. 86.13), el valor p obtenido (0.6008) es mucho mayor que el umbral de significancia (0.05), lo que indica que esta diferencia no es estadísticamente significativa.