Universidad Zaragoza
1474

Trabajo Fin de Máster

Función de los microARNs del tejido adiposo subcutáneo en la regulación de la esteatosis hepática.

Role of microRNAs from subcutaneous adipose tissue in the regulation of hepatic steatosis.

Autor

José Andrés Castillo Rivas

Directores

Silvia Lorente Cebrián - UNIZAR

José Miguel Arbonés Mainar - IACS

Master in Biophysics and Quantitative Biotechnology

FACULTAD DE CIENCIAS
2024–2025

Acknowledgments

Abstract

Background Hepatic steatosis is characterized by the excessive accumulation of triglycerides within hepatocytes. This condition is associated with the progression of metabolic dysfunction-associated steatotic liver disease (MASLD), which may be influenced by soluble molecules secreted from subcutaneous white adipose tissue (scWAT). Among these molecules, microRNAs (miRNAs) are thought to play a critical role in regulating the molecular pathways underlying the pathogenesis of MASLD.

Aim To profile miRNA expression in the subcutaneous white adipose tissue of patients with hepatic steatosis.

Methods Subcutaneous white adipose tissue (scWAT) samples were collected from 78 obese patients enrolled in the FATe cohort and subsequently frozen. The patients were categorized into steatosis groups based on liver fat content: "<5%", "5–33%", "33–66%", and ">66%". Extracted total RNA was profiled using ..... Pairwise comparisons were performed using Student's t-test for variables with a Gaussian distribution, the Mann–Whitney U test for non-Gaussian variables, and the chi-square test for categorical variables.

Results A total of 374 miRNA differentially expressed between obese patiens. Of these, two remained significant after multiple test correction (hsa-miR-372-3p and hsa-miR-144-3p). Predicted target genes for these miRNAs include insulin receptor pathway components (IGF1, IGFR13), cytokines (CCL3, IL6), ghrelin/obestatin gene, and inflammation-related genes (NFKB1, RELB, FAS). e

Conclusions miRNA expression from VAT may contribute to the pathogenesis of NAFLD – a finding which may distinguish relatively simple steatosis from NASH. This could help identify potential targets for pharmacological treatment regimens and candidate biomarkers for NASH.

Table of contents

List of Tables

List of Figures

Chapter 1

Introduction and Antecedents

1.1 Overview of Hepatic Steatosis and Metabolic dysfunction- associated steatotic liver disease

Hepatic steatosis, defined by the excessive accumulation of triglycerides within hepatocytes, arises from a range of factors[1]. These include drug-induced effects, such as those caused by steroids and chemotherapy, infections like hepatitis C virus, and both nutritional and metabolic causes[2]. This condition is closely linked to steatotic liver disease (SLD), which encompasses two primary metabolic categories: alcoholic-related liver disease (ALD) and metabolic dysfunction-associated steatotic liver disease (MASLD)[3].

MASLD and its more advanced form, metabolic dysfunction-associated steatohepatitis (MASH), were previously known as nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH), respectively[4]. MASLD is diagnosed in adults who exhibit hepatic steatosis, identified through imaging techniques, blood biomarkers, or liver histology, in conjunction with being overweight or obese[5], or in the presence of type 2 diabetes mellitus (T2DM) or at least two metabolic risk abnormalities (hypertension, hyperlipidaemia or insulin resistance)[6,7]. Although in many cases it may be asymptomatic in its early stages, this condition can progress to more severe stages depending on its cause[8]. It encompasses a spectrum of liver pathologies ranging from simple hepatic steatosis to metabolic dysfunction-associated steatohepatitis (MASH), which is marked by lobular inflammation and hepatocellular ballooning, potentially progressing to fibrosis and cirrhosis, ultimately leading to liver failure[9].

The prevalence of MASLD is affecting more than a third of the adult population worldwide, making it the most common chronic liver disease globally[10]. Among adults, the prevalence (MASLD) is approximately 30%[11]. MASLD is particularly prevalent in overweight or obese individuals, with a global prevalence of approximately 50%, rising to nearly 60% in individuals with type 2 diabetes (T2D), a condition that affects up to 10% of the world's adult population[12].

Over the last 30 years, the global prevalence of MASLD has experienced significant growth, rising from 17.6% in 1990 to 23.4% in 2019, reflecting an average annual increase of approximately 1.0%. As of 2019, it was estimated that there were around 1.66 billion prevalent cases of MASLD worldwide[13]. This condition is widespread, with the highest rates reported in South America (44.4%), Middle East East and North Africa (36.5%) and followed by South Asia (33.8%), South-East Asia (33.1%), North America and Australia (31.2%), East Asia (29.7%), Asia Pacific regions (28.0%)

and Western Europe (25.1%)[11,14]. Approximately one-quarter of the European population is affected by this liver disease[7]. In Europe, the prevalence of NAFLD variates between countries ranging from 5% to 44%[15]. Data from Spain reflect similar rates, indicating a NAFLD prevalence of 25.8% in the adult population[16].

This growing prevalence of MASLD is closely intertwined with the global rise in obesity and metabolic syndrome, conditions largely influenced by the complex interplay of adipose tissue function and dysfunction. Understanding the role of adipose tissue in metabolic regulation is therefore critical for elucidating the pathophysiology of MASLD.

1.2 Role of Adipose Tissue in Metabolic Regulation

1.2.1 Metabolic and Endocrine Functions of Adipose Tissue

Adipose tissue (AT) is a complex and dynamic organ with both metabolic and endocrine functions[17]. It plays a pivotal role in energy balance, insulin sensitivity, immune responses and overall health[18]. Its role extends beyond simple fat storage, influencing whole-body physiology and contributing to various pathologies, notably obesity and its associated complications like MASLD[19].

1.2.2 Types and Location of Adipose Tissue

Adipose tissue in mammals exists in three primary forms: white adipose tissue (WAT), brown adipose tissue (BAT), and beige or brite (brown-in-white) adipose tissue. Each type is distinguished by its unique functions and cellular composition[20]. WAT primarily stores energy in the form of lipids, while BAT specializes in heat production through the uncoupling of oxidative phosphorylation, a process critical for thermogenesis[21]. Beige AT, a metabolically flexible tissue, can transition between energy storage and thermogenesis depending on physiological needs, highlighting its role in adaptive responses[22].

The anatomical distribution of WAT further underscores its functional diversity. WAT is primarily located in subcutaneous depots, beneath the skin, and visceral depots, surrounding internal organs such as the liver and intestines[23]. Additionally, smaller WAT depots are present in areas like bone marrow and muscle tissue, contributing to localized metabolic regulation[21].

White adipose tissue (WAT) is classified into two major subtypes based on location and function: subcutaneous white adipose tissue (scWAT), which constitutes over 80% of total body fat and serves as the primary site for long-term energy storage, with its distribution and metabolic activity linked to protective effects on overall metabolic health; and visceral white adipose tissue (visWAT), which accounts for 10–20% of total body fat in men and 5–10% in women, is more metabolically active, and is strongly associated with adverse health outcomes such as insulin resistance and inflammation[24,25].

1.3 Subcutaneous White Adipose Tissue (scWAT)

Subcutaneous White Adipose Tissue (scWAT) is the most abundant type of adipose tissue, found in various locations such as under the skin and in clustered regions of the body, including the upper (deep and shallow abdomen) and lower (gluteofemoral) body areas[26].

Under physiological conditions, scWAT serves as a metabolically inert, long-term triglyceride storage site[27]. Acting as a buffer for excess energy, scWAT protects other organs from ectopic lipid deposition and contributes to specific metabolic benefits[20]. This buffering role allows scWAT to mitigate the impact of excess dietary lipid consumption while also compensating for energy deficits during fasting, starvation, or strenuous exercise[27].

Structurally, scWAT is composed predominantly of adipocytes, which constitute approximately 50% of its cellular content[28]. In addition to adipocytes, scWAT includes vascular cells, fibroblasts, adipocyte precursors, multipotent mesenchymal stem-like cells, nerve processes, and immune cells such as macrophages, lymphocytes, eosinophils, and mast cells. These components are embedded within an extracellular matrix that provides structural support[28]. Collectively, these cells secrete a wide array of signaling molecules, including adipokines such as leptin, adiponectin, and resistin, which play essential roles in maintaining metabolic homeostasis[22].

1.4 Metabolic Functions of WAT

The main role of white adipose tissue (WAT) is to manage energy balance by storing and releasing fatty acids (FAs) according to variations in energy availability[22]. Furthermore, WAT produces a range of hormones and cytokines, collectively referred to as adipokines, which are crucial for regulating numerous physiological functions[29].

1.4.1 Lipids Storage and Mobilization

WAT maintains energy balance by storing and releasing fatty acids (FAs), a process controlled by the interplay between lipogenesis and lipolysis. This equilibrium is essential for sustaining energy homeostasis during periods of fasting or exercise[22].

Lipogenesis is the process of synthesizing new lipids from excess glucose or dietary fatty acids. This process occurs in the cytoplasm of adipocytes and is tightly regulated by various hormones and enzymes[30]. Insulin, the key hormone involved in this process, facilitates the uptake of glucose and fatty acids into adipocytes and activates essential enzymes for lipid synthesis, including acetyl CoA-carboxylase (ACC) and fatty acid synthase (FAS)[31].

Lipolysis, on the other hand, is responsible for breaking down stored lipids in adipose tissue to release energy for peripheral organs. This process is particularly important during fasting or exercise when glucose levels are low, prompting the body to utilize stored fat for energy[32]. Lipolysis is regulated by lipases, which are activated by signals from the sympathetic nervous system, primarily mediated by norepinephrine, with some influence from epinephrine. The primary lipases involved include adipose triglyceride lipase (ATGL), hormone-sensitive lipase (HSL), and monoacylglycerol lipase (MGL)[33].

1.4.2 Endocrin Function

Additionally, WAT secretes various hormones and cytokines, collectively known as adipokines, which play essential roles in regulating multiple physiological processes such metabolic, inflammatory, and immune processes throughout the body[29]. The endocrine function of WAT is influenced by nutritional status, physical activity, hormonal levels, and environmental signals, and is closely related to its metabolic and storage functions[34].

Current understanding of adipose tissue (AT)-derived adipokines encompasses over 100 proteins that interact with various cells and tissues[22]. Leptin and adiponectin are the most prevalent and well-studied adipokines. Leptin, the first adipokine to be discovered, has been shown to influence appetite and energy expenditure, serving as an important feedback mechanism to the brain regarding the size and condition of adipose tissue[35]. In contrast, adiponectin promotes insulin sensitivity and fatty acid oxidation in skeletal muscle and the liver, contributing to the maintenance of glucose and lipid homeostasis[36].

1.5 Dysregulation of Adipose Tissue and Its Implications

However, the function of scWAT is closely linked to body fat distribution, and under pathological conditions, its capacity to store lipids becomes overwhelmed. This dysfunction impairs its ability to store lipids appropriately, leading to hypoxia, infiltration of pro-inflammatory macrophages, and ectopic fat accumulation in other tissues like liver (Ipsen et al., 2018). These alterations directly contribute to the dysregulation of adipokine secretion, fostering a pro-inflammatory state and systemic insulin resistance, which are hallmark features in patients with NAFLD-associated diseases (Ipsen et al., 2018).

The adipose tissue expandability hypothesis is central to understanding the pathogenesis of obesity-related comorbidities, including non-alcoholic fatty liver disease (NAFLD). This hypothesis posits that the body's ability to store excess calories in scWAT is limited and varies greatly between individuals [9,47]. When scWAT reaches its maximal storage capacity, AT can no longer store lipids effectively, leading to the redirection of lipid flux to other organs. This results in ectopic fat accumulation, primarily in visWAT and the liver, causing insulin resistance and related metabolic complications through lipotoxic and inflammation-driven mechanisms

The imbalance between energy intake and the storage capacity of AT is a crucial factor in the development of NAFLD. Excessive caloric intake, especially when coupled with the limited expandability of scWAT, leads to fat deposition in the liver. This hepatic fat accumulation, or steatosis, is the hallmark of NAFLD and sets the stage for further liver damage.

1.6 Crosstalk

Peroxisome Proliferator Activated Receptor Gamma (*PPARG*), patatin like phospholipase domain containing 2 (*PNPLA2*), Diacylglycerol O-Acyltransferase 2 (*DGAT2*), Fas Cell Surface Death Receptor (*FAS*), Acetyl-CoA Carboxylase Alpha (*ACACA*)

1.7 MicroRNAs y Regulación del Tejido Adiposo

1.7.1 microRNAS (miRNAs): biogenesis and function

MicroRNAs (miRNAs) have emerged as essential and versatile post-transcriptional regulators of gene expression, playing a significant role in various biological processes including the cell cycle, proliferation, lipid metabolism, inflammation, fibrosis among others[37]. The dysregulation of miRNA has been implicated in the pathogenesis of several many diseases and metabolic disorders, including NAFLD[38]. Specific miRNAs are known to be differentially expressed in metabolic tissues, including adipose tissue, in response to pathological states[37]. Their potential as biomarkers for disease progression and therapeutic targets has garnered significant interest in recent years[38].

Given the increasing prevalence of NAFLD and the limited availability of non-invasive diagnostic tools, identifying specific miRNAs as biomarkers in subcutaneous white adipose tissue could provide valuable insights into the underlying molecular mechanisms and aid in risk stratification. This study aims to bridge this knowledge gap by analyzing the differential expression of miRNAs in scWAT across varying degrees of steatosis, predicting their target genes, and exploring their potential functional roles.

Los miRNAs están implicados en procesos clave como la adipogénesis, la inflamación, la lipólisis y la resistencia a la insulina, todos fundamentales en la fisiopatología del tejido adiposo. ### MiRNAs en el contexto de MASLD: Evidencias crecientes sugieren que los miRNAs desempeñan un papel crucial en la comunicación entre el tejido adiposo y el hígado, influyendo en la progresión de la esteatosis hepática. scWAT communicates with the liver and other metabolic organs through a network of signaling molecules, including adipokines, free fatty acids, and cytokines, which influence metabolic homeostasis (Ghesmati et al., 2024). Dysregulation of this crosstalk in the context of obesity and insulin resistance contributes to ectopic fat deposition and inflammation, exacerbating hepatic steatosis (Koenen et al., 2021).

1.7.2 Basic nomenclature of the miRNA

1.7.3 microRNAS related to MASLD

1.7.4 miRNAs as Biomarkers

1.7.5 Functional and Cellular Complexity of White Adipose Tissue

Beyond its anatomical distribution, the composition of WAT reveals a remarkable heterogeneity that underscores its functional complexity. While adipocytes are the primary cellular component, WAT also encompasses a rich network of blood vessels, nerve terminals, and various cell types, including immune cells. This cellular diversity plays a crucial role in supporting the tissue's metabolic and endocrine functions, further highlighting its central role in whole-body energy regulation (Kwok et al., 2016).

Chapter 2

Objectives

2.0.1 5. Study Hypothesis and Objectives

- Hypothesis based on the differential regulation of miRNAs according to steatosis grade and their impact on specific metabolic pathways.
- Main objectives, such as:
  - Identifying differentially expressed miRNAs in subcutaneous adipose tissue of patients with varying degrees of steatosis.
  - Predicting target genes of these miRNAs and analyzing their role in steatosis progression.
  - Assessing the feasibility of using these miRNAs as diagnostic or prognostic biomarkers.

Chapter 3

Methodological Framework

3.1 Study Population

The analyzed data originates from the FATe cohort[39], a longitudinal study of obese patients undergoing bariatric surgery at Miguel Servet University Hospital (HUMS, Zaragoza, Spain). For this research, 78 patients with varying degrees of adiposity were selected based on subcutaneous adipose tissue (scWAT) samples, which are registered at the regional Biobank (Biobanco Aragón) and approved by the CEICA ethics committee. Patients with alcohol or drug abuse, autoimmune diseases, chronic inflammatory conditions, or infectious diseases (HIV, HBV, HCV) were excluded during the screening process. The FATe cohort is characterized by a range of clinical and demographic variables, including sex, age, body mass index (BMI), steatosis, metabolic dysfunction-associated steatohepatitis (MASH), hepatocytic ballooning, lobular inflammation, diabetes, hyperlipidemia, and metabolic dysfunction-associated steatotic liver disease (MASLD).

3.2 Collection and characterization of subcutaneous white adipose tissue

Subcutaneous white adipose tissueAdipose tissue (scWAT) biopsies (~3 cm^3) from the subcutaneous depot were obtained during laparoscopic bariatric surgery using a bipolar/ultrasonic device (Thunderbeat). The samples were extracted through a 12 mm trocar (Applied Medical) inserted into the left hypochondrium. An experienced pathologist evaluated pathological features such as steatosis, lobular inflammation, hepatocellular ballooning, and fibrosis according to the criteria established by the Nonalcoholic Steatohepatitis Clinical Research Network ([40]).

3.3 RNA isolation

Total RNA was extracted from frozen biopsies of subcutaneous white adipose tissue (scWAT) and cell cultures using TRIzol (#T9424, Sigma Aldrich) following the manufacturer's instructions. For lysing adipose tissue samples, 1 ml of TRIzol was added per sample in a homogenizer, while for cell cultures, 1 ml of TRIzol was used per 10 cm^2 of the culture plate, along with a scraper. The resulting cell lysates or tissue disaggregates were transferred to a vial and incubated for 5 minutes at room temperature to dissociate nuclear components. Subsequently, 0.2 ml of 100% chloroform

per ml of TRIzol was added. The mixture was shaken vigorously, incubated for 15 minutes at room temperature, and then centrifuged for 15 minutes at 12,000 g and 4 ºC.

Following centrifugation, three distinct phases were formed. The aqueous phase, which contained the ribonucleic acids, was collected, and 0.5 ml of 100% isopropanol was added to precipitate the RNA. This mixture was mixed, incubated on ice for 10 minutes, and then centrifuged at 12,000 g for 15 minutes at 4 ºC. Afterward, the supernatant was carefully decanted, and the RNA pellet was resuspended in 1 ml of 75% ethanol for washing. The pellet was homogenized and centrifuged at 7,500 g for 5 minutes at 4 ºC. The supernatant was discarded, and the pellet was allowed to dry for 10 minutes at room temperature. Finally, the RNA was resuspended in DEPC-treated water. To eliminate any genomic DNA contamination, all RNA samples were treated with RNase-Free DNase (Life Technologies). The concentration and purity of the RNA were assessed by measuring absorbance at 260/280 nm and 260/230 nm using a Nanodrop 2000 (Thermo Fisher). The quality of the extracted RNA was visualized using an agarose gel.

3.4 RNA Sequencing

RNA integrity was evaluated using the RNA Integrity Number (RIN) on the Agilent 2200 TapeStation with the RNA ScreenTape assay. Stranded mRNA libraries were prepared with the Novogene NGS RNA Library Prep Set (PT042), which included mRNA isolation using poly-T oligo-attached magnetic beads, cDNA synthesis, adapter ligation, and PCR amplification. Libraries that passed quality control checks were sequenced (2x150 bp) on the Illumina Novaseq X Plus platform.

3.5 Analysis of sRNA-seq Data with *nf-core/smrnaseq*

For the analysis of small RNA sequencing (sRNA-seq) data, version 2.4.0 of the *nf-core/smrnaseq* pipeline[41] was used, which is specifically designed for the automated processing of miRNAs data.

3.5.1 Execution of the *nf-core/smrnaseq* Pipeline

The installation of *nf-core/smrnaseq* was carried out following the instructions provided by the authors in *nf-core*[42], available at https://nf-co.re/smrnaseq/2.4.0 .

To ensure the proper installation and execution of the pipeline, the following key components were installed beforehand:

1. *Nextflow*: Version 24.04.4 of Nextflow was used, following the detailed instructions at https://nf-co.re/usage/installation.
2. *Java Runtime Environment (JRE)*: Version 11.0.25 of the Java Runtime Environment was installed, as it is required for compatibility with Nextflow and the *nf-core/smrnaseq* pipeline.

To ensure reproducibility and streamline the pipeline execution, one of the available Docker containers was utilized. These containers provide the necessary instructions and configurations required to run the pipeline. The configuration is specified at runtime using the `profile` argument. For this analysis, the Docker image *nf-core/smrnaseq*, available at Docker Hub, was employed.

The pipeline was executed on a server with 8 CPUs, 16 GB of RAM, and a Linux operating system. The following command was used in the terminal, which configures the main options, including the reference genome, input data, and output file location:

```
nextflow run nf-core/smrnaseq -r 2.4.0
-profile docker,ci
--genome GRCh38
--input '/home/joshoacr13/Documentos/TFM/nfcore-smrnaseq/input/samples.csv'
--fasta 'https://github.com/nf-core/test-datasets/raw/smrnaseq/reference/genome.fa'
--mirtrace_species 'hsa'
--outdir /home/joshoacr13/Documentos/TFM/nfcore-smrnaseq/workdir
-resume -c /home/joshoacr13/Documentos/TFM/nfcore-smrnaseq/nextflow_memory.config
--save_intermediates FALSE
```

The pipeline was executed three times to accommodate the large number of samples, processing 26 samples per run.

3.5.2 Description of the Parameters Used

- `-profile docker,ci`: Runs the pipeline inside a Docker container to ensure reproducibility and sets up a continuous integration (CI) profile.
- `--genome GRCh38`: Specifies the human genome (version GRCh38) as the reference for sequence mapping.
- `--input`: Provides the path to the CSV file containing metadata and the paths to the FASTQ files.
- `--fasta`: URL to the FASTA file of the reference genome.
- `--mirtrace_species hsa`: Defines the species as Homo sapiens (hsa) for miRNAs analysis with miRTrace.
- `--outdir`: Sets the working directory for the processed results.
- `-resume`: Allows continuation of a previous analysis without restarting from the beginning.
- `-c`: Specifies a custom configuration file (nextflow_memory.config) to adjust resource usage.
- `--save_intermediates FALSE`: Prevents the storage of intermediate files to save disk space.

3.5.3 Analysis Workflow and Tools Used

The *nf-core/smrnaseq* pipeline performs the following steps:

1. **Quality Control** An initial quality assessment of the raw reads was conducted using *FastQC* (version 0.12.1)[43]. Additionally, 3' adapter trimming was performed using *fastp* (version 0.23.4)[44], followed by quality and length filtering. A second quality assessment of the trimmed reads was conducted with *FastQC*.
2. **miRNA Quality Control**: A more specific quality control process for miRNA sequencing was implemented, whereby samples that failed to meet the minimum quality thresholds established by *miRTrace* (version 1.0.1)[45] were excluded from further analysis. This tool incorporates the following steps to ensure the integrity and reliability of the data:

- **Verify Read Length Distribution**: The majority of reads fell within the expected range of 18–24 nucleotides, indicative of high-quality small RNA data.

- **Identify Contaminants**: Potential contaminants such as tRNA, rRNA, and other non-target molecules were flagged.
- **Taxonomic Classification**: Reads were classified taxonomically to ensure that most sequences originated from the organism of interest (*Homo sapiens*).

3. **miRNAs Quantification**:

- **Alignment**: The filtered reads were aligned against mature miRNA sequences in the miRBase database using *Bowtie1* (version 1.3.1)[46]. Unmapped reads were aligned against "hairpin" sequences to identify miRNA precursors.
- **Post-Alignment Processing**: *SAMtools* (version 1.16.1)[47] was used to process the mapping results.
- **Quantification and Normalization**: Initial quantification was performed with *edgeR* (version 4.4)[48], generating normalized count tables (TMM) for detected miRNAs. Exploratory graphs were generated, including a multidimensional scaling (MDS) analysis to cluster samples and a heatmap to evaluate similarities among them.

4. **IsomiR Annotation**: The collapsed reads were processed with *mirtop* (version 0.4.28)[49] to identify miRNA variants (isomiRs). This analysis allows for the mapping and annotation of variants related to length and sequence modifications of mature miRNAs. The *mirtop* tool employs the *Blending Analysis* technique to process and integrate miRNA data, ultimately generating a count matrix that accurately represents the expression levels of these molecules in the analyzed samples. The application of *Blending Analysis* allows for the generation of a more robust and comprehensive count matrix, thus facilitating subsequent differential expression analysis. This method includes the following essential steps:

- **Read Grouping**: The miRNA reads are grouped from the processed data, ensuring that different variants and reference sequences are integrated coherently.
- **Adjustment for Variants**: Both miRNA variants (isomiRs) and standard reference sequences extracted from databases such as miRBase are considered. This adjustment is fundamental to obtain an accurate representation of miRNA expression in the analyzed samples.

5. **Analysis and Visualization of Results**: The overall pipeline metrics, encompassing quality assessments, mapping statistics, and expression analysis results, were consolidated and summarized using *MultiQC* (version 1.25.1)[50]. Visualization of the results was performed using the *ggplot2* package in R[51], which facilitated the creation of clear and informative graphical representations.

3.6 Differential expression analysis according to steatosis using *DEseq2*

For the differential expression analysis, the R statistical software[52], version 4.4.1 (2024-06-14) (https://cran.r-project.org/), was used. This analysis was performed using the RStudio integrated development environment (IDE)[53], version 2023.12.0+369, designed for Ubuntu Jammy (https://www.rstudio.com/).

The script used to perform the differential expression analysis is available in the file "`miRNA_steatosis.qmd`" which can be accessed at the following link: https://github.com/joshoandres13/miRNAs.

The analysis began by loading essential R packages: *tidyverse* (version 2.0.0)[54] for data manipulation, *isomiRs*[55] (version 1.32.1) for analyze isomirs and miRNAs from small RNA-seq, *DESeq2*

(version 1.44.0)[56] for differential expresion analysis, *org.Hs.eg.db* (version 3.19.1)[57] for gene annotation.

3.6.1 Data Preparation

Metadata were imported, and sample identifiers were stablished as row names. The variable *Steatosis* is categorized in four distinct groups. Isomirs count data were subsequently obtained where rows represent the identified isomiRs and columns correspond to the experimental samples.

Using isomirs count data and metadata, an object of class `IsomirDataSeq` was created. This object enables efficient management of information derived from small RNA sequencing studies, streamlining differential expression analyses and facilitating the interpretation of biological findings.

3.6.2 Filtering and Processing of isomiRs

The filtering process enables the grouping of isomiRs into distinct categories, associating them with a single variant of a miRNA. This grouping is essential for ensuring consistency and accuracy in differential expression analyses. To reduce technical noise and highlight biologically meaningful signals, a stringent filtering criterion was applied: only isomiRs with a minimum of 20 counts in at least 40 samples were retained.

3.6.3 Differential expression in scWAT

In this study, the *DESeq2* package was used to perform differential expression analysis on scWAT samples with varying degrees of liver steatosis. The *DESeq2* object was configured using the Likelihood Ratio Test (LRT), enabling the analysis to account for all four steatosis categories: <5% (no steatosis), 5–33% (mild steatosis), 33–66% (moderate steatosis), and >66% (severe steatosis). This approach allowed for a comprehensive assessment of gene expression changes across the entire spectrum of steatosis progression.

3.6.3.1 Testing for differential expression

The criteria for identifying significant differentially expressed miRNAs across the steatosis categories involved applying a false discovery rate (FDR) cutoff of less than 0.05. Following the identification of significant miRNAs, the expression patterns of the selected miRNAs were further analyzed by representing their normalized counts in boxplots across the four steatosis groups. This approach enabled the observation of expression variations of these miRNAs across different steatosis degrees, offering insights into their potential roles in steatosis progression. Differences between steatosis groups were evaluated using the Kruskal-Wallis test

3.7 Target mRNA Selection and Validation Using *multiMiR*

The selected miRNAs were used for the identification of mRNA targets through the *multiMiR* bioinformatics package[58,59], version 2.4.0 in R. *multiMiR* facilitates a systematic search and annotation of miRNA targets, providing functional analysis to elucidate biological mechanisms. For this analysis, only validated interaction data were utilized.

3.7.1 Filtering Parameters

The validated target table provided by *multiMiR* was used during the selection process. Key columns included:

1. **database**: Source database of validated interactions, such as *miRTarBase*, *TarBase*, or *miRecords*.
2. **mature_mirna_id**: Standard format identifier for the miRNA.
3. **target_symbol**: Target gene symbol.
4. **experiment**: Experimental methods used for validation, including luciferase assays, Western blot, or qRT-PCR.
5. **support_type**: Level of experimental support, such as "Functional MTI" (miRNA-mRNA functional interaction).
6. **pubmed_id**: References to PubMed articles reporting the interaction.
7. **type**: Specifies whether the interaction is "validated" or "predicted."

3.7.2 Selection Criteria

To ensure reliable results, databases were filtered according to update criteria and the following selection parameters:

- Databases up-to-date at the time of analysis were prioritized (*miRTarBase* and *TarBase*).
- Only interactions classified as "validated" were included.
- Interactions backed by robust experimental methods, such as luciferase assays or Western blot, were prioritized.
- Interactions with functional support ("Functional MTI") and verifiable references in PubMed were selected.

This approach ensured the identification of mRNA targets with high reliability and experimental backing, facilitating the analysis of potential regulatory functions of the studied miRNAs.

3.7.3 Functional Analysis

To explore the biological functions associated with the validated target genes, an enrichment analysis was performed using the *KEGGREST* package[60] in R. Gene symbols for validated target genes associated with selected miRNAs were extracted using *multiMiR* previously, with duplicates removed to ensure a unique gene list. The Entrez IDs of these genes were mapped to specific metabolic pathways in KEGG, using parameters that controlled the false discovery rate (FDR) with the Benjamini-Hochberg method and significance cutoff values for `qvalue` and `pvalue` set at 0.05. The results included a bar plot displaying the top 5 significantly enriched KEGG pathways, highlighting their statistical significance and the number of genes associated with each pathway. This analysis provided insights into key metabolic pathways and biological processes involving miRNA-regulated target genes.

3.8 Functional validation of miRNAs

3.8.1 Cell Culture and Transfections with miRNA Mimics

The human hepatoma HepG2 cell line (American Type Culture Collection, ATCC® HB-8065™; Manassas, VA, USA) was cultured in an incubator at 37 ºC and 5% CO_2 using high-glucose Dulbecco's Modified Eagle Medium L-GlutaMAX (DMEM) (Gibco, Thermo Fisher Scientific Inc., Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, Thermo Fisher Scientific Inc., Brazil). Experiments were carried out when the cells reached 70-80% confluence.

For the experiments, HepG2 cells were plated at a density of 100,000 cells per well in 12 well-plates for gene expression assays, in DMEM high-glucose (1 g/L) L-GlutaMAX supplemented with 10% FBS.

HepG2 cells were reverse-transfected with Lipofectamine RNAiMAX Reagent (Thermo Fisher Scientific Inc.) and 50 nM of the following mirVana™ miRNA mimics (Thermo Fisher Scientific Inc.): a scramble sequence as a negative control (mirVana™ miRNA Mimic, Negative Control #1), negative control (mirVana™ miRNA Inhibitor, Negative Control #1), *hsa-miR-144-3p* (5′ – *UACAGUAUAGAUGAUGUACU* – 3′; assay ID H11051), and *hsa-miR-372-3p* (5′ – *AAAGUGCUGCGACAUUUGAGCGU* – 3′; assay ID MC10165). The miRNA mimics were diluted in Opti-MEM I Reduced Serum Medium (Gibco, Thermo Fisher Scientific Inc.) and added to the wells according to the manufacturer's instructions. Following this, Lipofectamine was added to the wells containing the diluted miRNA mimics and incubated for 15 minutes at room temperature to form miRNA mimic–lipofectamine complexes. HepG2 cells diluted in DMEM with 10% FBS were then plated into the wells. The cells were incubated for 24 hours to assess transfection efficiency. To induce lipid accumulation and simulate steatosis[61], HepG2 cells were treated with oleic acid 24 hours after miRNA mimic transfections, mixed in DMEM GlutaMAX-I supplemented with 10% FBS, at a final concentration of 0.5 mM, which reflects the physiological range of fatty acids used to mimic hepatic steatosis[62].

3.8.2 RNA Isolation and Gene Expression Analyses

HepG2 cells were frozen on dry ice and stored at -80 °C until RNA extraction. Cells used for miRNA expression analyses were washed previously with phosphate-buffered saline (PBS) to completely remove potential unabsorbed miRNA mimics. Total RNA was extracted from frozen cell cultures using TRIzol (#T9424, Sigma Aldrich) following the manufacturer's instructions. For lysing cell cultures, 1 ml of TRIzol was used per 10 cm^2 of the culture plate, along with a scraper. The resulting cell lysates were transferred to a vial and incubated for 5 minutes at room temperature to dissociate nuclear components. Subsequently, 0.2 ml of 100% chloroform per ml of TRIzol was added. The mixture was shaken vigorously, incubated for 15 minutes at room temperature, and then centrifuged for 15 minutes at 12,000 g and 4 ºC.

Following centrifugation, three distinct phases were formed. The aqueous phase, which contained the ribonucleic acids, was collected, and 0.5 ml of 100% isopropanol was added to precipitate the RNA. This mixture was mixed, incubated on ice for 10 minutes, and then centrifuged at 12,000 g for 15 minutes at 4 ºC. Afterward, the supernatant was carefully decanted, and the RNA pellet was resuspended in 1 ml of 75% ethanol for washing. The pellet was homogenized and centrifuged at 7,500 g for 5 minutes at 4 ºC. The supernatant was discarded, and the pellet was allowed to dry for 10 minutes at room temperature. Finally, the RNA was resuspended in DEPC-treated water. To eliminate any genomic DNA contamination, all RNA samples were treated with RNase-Free DNase

(Life Technologies). RNA concentration were determined using Qubit 4 Fluorometer (Thermo Fisher Scientific Inc.).

mRNA expression was evaluated in HepG2 cells transfected with miRNA mimics for 48 h. RNA was reverse transcribed using PrimeScript Reverse Transcriptase (Takara Bio), with 100 ng of RNA utilized for each reaction in a total volume of 10 μl. The process was carried out using an Applied Biosystems 2720 Thermal Cycler, following this protocol: 10 minutes at 25 °C, 2 hours at 37 °C, and finally, 5 minutes at 85 °C. Ten nanograms of the cDNA product were amplified using Quantitative-real time PCR (qPCR) in a total reaction volume of 15 μl with SYBR Select Master Mix (Applied Biosystems), to which 0.5 μl of gene-specific primers at a concentration of 10 μM was added. The primers utilized are detailed in Table 3.1 cDNA amplification was performed on a StepOnePlus system (Applied Biosystems) with the following protocol: an initial step at 95 °C for 10 minutes, followed by 40 cycles of 15 seconds at 95 °C and 1 minute at 60 °C, concluding with 15 seconds at 95 °C, 1 minute at 60 °C, and a final 15 seconds at 95 °C. The gene β-actin was used as a housekeeping control to normalize gene expression levels ($\triangle Cq$). Comparisons between gene expression levels after miRNA mimic transfections vs. scramble-sequence-transfected cells (negative control) were established using the the $2 - \triangle\triangle CT$ method, determining relative gene expression[63].

Table 3.1: Primers designed for qPCR mRNA gene expression analysis. Abbreviations: *ACTB-2* (Actin) , *ACACA* (Acetyl-CoA Carboxylase Alpha), *DGAT2* (Diacylglycerol O-Acyltransferase 2), *FAS* (Fatty Acid Synthase), *PNPLA2* (Patatin-like phospholipase domain-containing protein 2), *PPARG* (Peroxisome Proliferator Activated Receptor Gamma)

Gene Name	**Forward Primer**	**Reverse Primer**
ACTB-2	ACCGAGCGCGGCTACAG	CTTAATGTCACGCACGATTTCC
PPARG	AGATGACAGCGACTTGGCAAT	ACTCAGGGTGGTTCAGCTTC
ATGL (PNPLA2)	TGGAGACTGAGGAGAACAAG	ATCCCTGCTTGCACATCTCT
FAS	AAGGACCTGTCTAGGTTTGATGC	TGGCTTCATAGGTGACTTCCA
DGAT2	AGTGGCAATGCTATCATCATCGT	TCTTCTGGACCCATCGGCCCCAGGA
ACCA	AGGTGCCTAGAGGGTTGAAGA	TCGGCCCTGCTTTACTAGGT

3.9 Statistical Analysis

The results are expressed as median [interquartile range] and number of cases (%). Pairwise group comparisons for continuous variables were calculated using Student's t-test for variables with a Gaussian distribution and the Mann–Whitney U test for data that do not follow this distribution. Categorical variables were analyzed using the chi-square test.

Chapter 4

Results

4.1 Phenotypic Characterization

Table 4.1 describes the phenotypical features of FATe patients selected for this study. This cohort encomprisses 78 subjects (24% males; 76% females), aged between 22 to 61 with obesity as regards of BMI levels of 46.11 ± 6.13 kg/m^2. These patients had an overall presence of general variables/descriptors of metabolic syndrome features of particular interest those related to liver/hepatic status (steatosis, lobular inflammation and the prevalence of associated metabolic diseases). According to the hepatic steatosis scale, 35.9% of patients had less than 5% liver fat, indicating a normal or minimal steatosis state. A total of 32.1% exhibited mild steatosis (5–33%), while 25.6% showed moderate fat accumulation (33–66%). Only 6.4% of patients presented severe steatosis (>66%), with no significant differences between men and women (p= 0.818).

Table 4.1: Clinical characteristics of the FATe cohort. Data are presented as number of cases (%) or median [interquartile range]. Differences between groups were tested with the Mann–Whitney U test and chi-square test; *BMI*: Body Mass Index (kg/m^2); *MASH*: Metabolic dysfunction-associated steatohepatitis.

Characteristic	**Overall (n=78)**	**Female (n=59)**	**Male (n=19)**	**p**
Age (years) median	47.03 [13.75]	46.53 [13]	48.58 [14]	0.408
Body Mass Index (kg/m^2) median	46.11 [8.70]	45.68 [8.76]	47.42 [8.70]	0.284
MASLD Activity Score Category (%):				0.510
- 0	22 (28.2)	16 (27.1)	6 (31.6)	
- 1	16 (20.5)	13 (22.0)	3 (15.8)	
- 2	17 (21.8)	14 (23.7)	3 (15.8)	
- 3	9 (11.5)	7 (11.9)	2 (10.5)	
- 4	9 (11.5)	7 (11.9)	2 (10.5)	
- >= 5	5 (6.4)	2 (3.4)	3 (15.8)	
Hepatic Steatosis Scale (%):				0.818
- < 5%	28 (35.9)	21 (35.6)	7 (36.8)	
- 5-33%	25 (32.1)	20 (33.9)	5 (26.3)	
- > 33-66%	20 (25.6)	15 (25.4)	5 (26.3)	
- > 66%	5 (6.4)	3 (5.1)	2 (10.5)	
Hepatocytic ballooning Category (%):				0.489

- None	58 (74.4)	45 (76.3)	13 (68.4)	
- Few cells	13 (16.7)	10 (16.9)	3 (15.8)	
- Many cells	7 (9.0)	4 (6.8)	3 (15.8)	
Lobular Inflammation Category (%):				0.596
- No foci	52 (66.7)	41 (69.5)	11 (57.9)	
- < 2 foci/200x	19 (24.4)	14 (23.7)	5 (26.3)	
- 2-4 foci/200x	4 (5.1)	2 (3.4)	2 (10.5)	
- > 4 foci/200x	3 (3.8)	2 (3.4)	1 (5.3)	
Diabetes:				1.000
- Yes (%)	21 (26.9)	16 (27.1)	5 (26.3)	
- No (%)	57 (73.1)	43 (72.9)	14 (73.7)	
Hyperlipidemia:				0.006
- Yes (%)	27 (34.6)	15 (25.4)	12 (63.2)	
- No (%)	51 (65.4)	44 (74.6)	7 (36.8)	
Non-alcoholic steatohepatitis:				0.249
- NASH (%)	12 (15.4)	7 (11.9)	5 (26.3)	
- Non-NASH (%)	66 (84.6)	52 (88.1)	14 (73.7)	

Regarding lobular inflammation, most patients (66.7%) had no inflammatory foci, while 24.4% exhibited fewer than two foci per microscopic field, and 3.8% showed severe inflammation (>4 foci). Hepatocyte ballooning analysis revealed that 74.4% of patients did not display significant damage, although 9.0% showed severe ballooning.

Among most prevalent comorbidities, 26.9% of patients were diagnosed with diabetes, and 34.6% presented hyperlipidemia, with the latter being significantly more prevalent in men (63.2%) compared to women (25.4%, p= 0.006). Additionally, 15.4% of patients were classified with metabolic dysfunction-associated steatohepatitis (MASH), although this proportion showed no statistically significant differences between genders (p= 0.249). Collectively, these data highlight the heterogeneity in the clinical characteristics of the cohort, emphasizing the complexity of the relationship between obesity and liver disease in this group of patients.

4.2 Quality Control (QC) and reads preprocessing

All QC results from the various steps in the *nf-core/smrnaseq* pipeline are summarized in the following figures and tables. The evaluation of samples, both before and after processing, was conducted using *fastp*, a rapid tool designed for preprocessing RNA sequencing data. This tool includes features for adapter trimming, quality filtering, and report generation, making it essential for assessing sample quality. Overall, Table 4.2 presents a descriptive analysis that includes key metrics related to the sequencing data.

Table 4.2: Descriptive statistics of the analyzed metrics with fastp. *% Duplication*: Duplication rate before filtering; *Reads After Filtering*: Total reads after filtering in millions; *% GC* content: GC content after filtering; *% PF*: Percent reads passing filter; *% Adapter*: Percentage adapter-trimmed reads

	Mean	**sd**	**Median**	**Minimum**	**Maximum**	**Range**
% Duplication	98.30513	0.68	98.45	94.31	99.09	4.78

Reads After Filtering (M)	25.23152	5.67	25.82	1.51	35.47	33.96
% GC content	46.71982	1.68	46.43	43.26	50.89	7.63
% PF	99.02358	1.21	99.36	90.13	99.76	9.63
% Adapter	99.38240	0.35	99.45	96.96	99.62	2.66

The metrics obtained from the fastp analysis provide a detailed overview of the preprocessing performance across the samples. The percentage of duplicated reads before filtering was notably high, with a mean of 98.31% (±0.68), a median of 98.45%, and a range between 94.31% and 99.09%. After filtering, the number of reads retained per sample averaged 25.23 million (±5.67M) suggesting a reasonable amount of reads obtained, with a median of 25.82M and a range from 1.51M to 35.47M.

The GC content showed consistency across the samples, with an average of 46.72% (±1.68), a median of 46.43%, and a range between 43.26% and 50.89%. The percentage of pass-filtered (PF%) reads was high, with an average of 99.02% (±1.21), reaching a maximum of 99.76%, as shown in Figure 4.1.

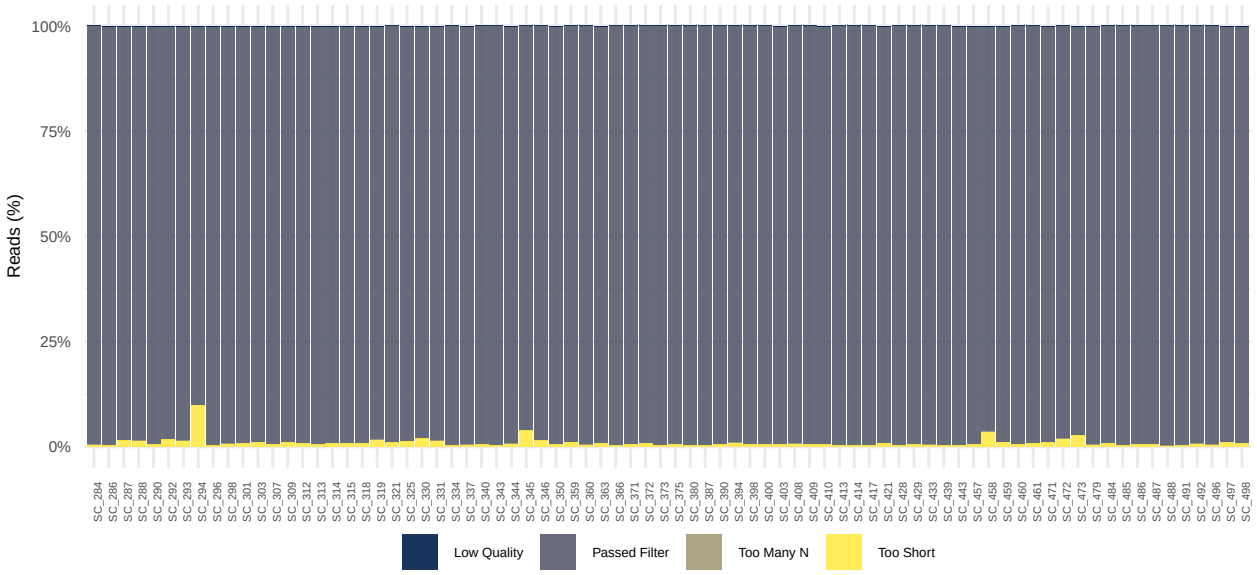

Figure 4.1: Fastp: Filtered Reads

The adapters used in library construction, specifically the *Illumina Universal Adapter* type, are optimized for the amplification and sequencing of diverse sample types. The adapter trimming was highly effective, achieving a mean success rate of 99.38% (±0.35) indicating a low level of adapter contamination in most samples, with a range from 96.96% to 99.62% (Table 4.2). In line with these findings, Figure 4.2 illustrates the adapter content before and after processing with *fastp*, where the mean trimming percentage of 0.05 ± 0.06% for scWAT samples indicates that adapters were effectively removed from all samples. These results confirm the efficiency of the preprocessing steps, ensuring that high-quality reads were retained while effectively eliminating low-quality sequences and adapters.

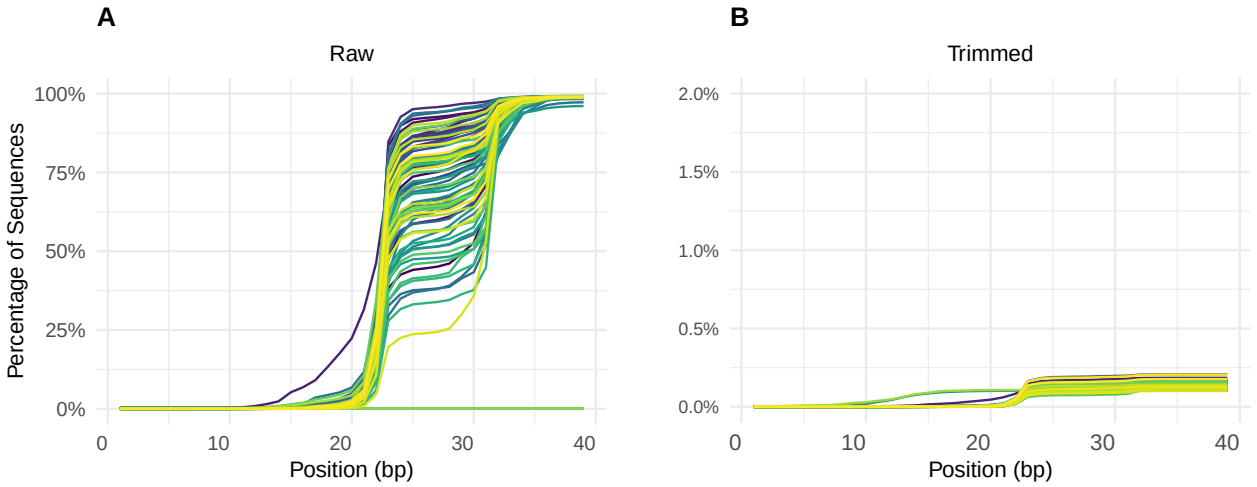

Figure 4.2: Adapters Content (%) across all bases before (A) and after (B) using *fastp* (version 0.23.4)

Figure 4.3 shows the analysis of the mean quality value of sequences across all bases, both before (*A*) and after (*B*) processing with *fastp*, it is evident from that the sequences from all samples consistently fall within an acceptable quality range, with mean Phred scores exceeding 30. This indicates a high base-calling accuracy, with an error probability of less than 0.1%, ensuring robust data integrity throughout the process. However, there is a noticeable drop in quality towards the extremes of the readings, particularly at position 31 bp. This decline can be attributed to the effects of adapter trimming in small RNA sequencing (smRNA-seq). Even though adapters have been removed, if they were present at the end of the sequence, their trimming may result in reduced quality for the subsequent bases. Bases near the adapter might have lower quality, leading to a set of bases that, while technically valid, originate from a lower-quality region. Additionally, it is common in smRNA-seq to encounter contaminants from various sources, such as rRNA or tRNA, which introduce background noise and can further diminish the quality of the readings obtained.

The Figure 4.3 illustrates the metric of the average quality value of each sequence across all samples. Before processing with *fastp* (*C*), the Phred scores consistently exceed 30, with most samples achieving scores around 35. This indicates that the average sequence quality is optimal, ensuring high reliability and low error rates. However, after trimming (*D*) a slight decrease in both Phred scores and read counts is observed. This reduction in the number of reads is expected after strict preprocessing with *fastp*, especially if contaminated or low-quality reads were present.

In Figure 4.3 (*E*), prior to trimming, the GC content distribution does not follow a normal pattern. Multiple peaks are observed, likely due to the presence of adapters, which typically have a fixed GC content that generates specific peaks. Additionally, since this is a *small RNA-seq* experiment, contamination from rRNA or tRNA is common, as these molecules often have a different GC content compared to other small RNAs. After processing with *fastp* (*F*), the GC content decreases, indicating that most of these contaminant sequences have been removed, leaving of reads with anomalies. This remaining may correspond to sequences that are difficult to classify or contaminants that are not easily eliminated by *fastp*.

Additionally, no issues with the presence of ambiguous bases (Ns) are detected in any of the samples before and after trimming. Figure 4.3 (*G* and *H*).

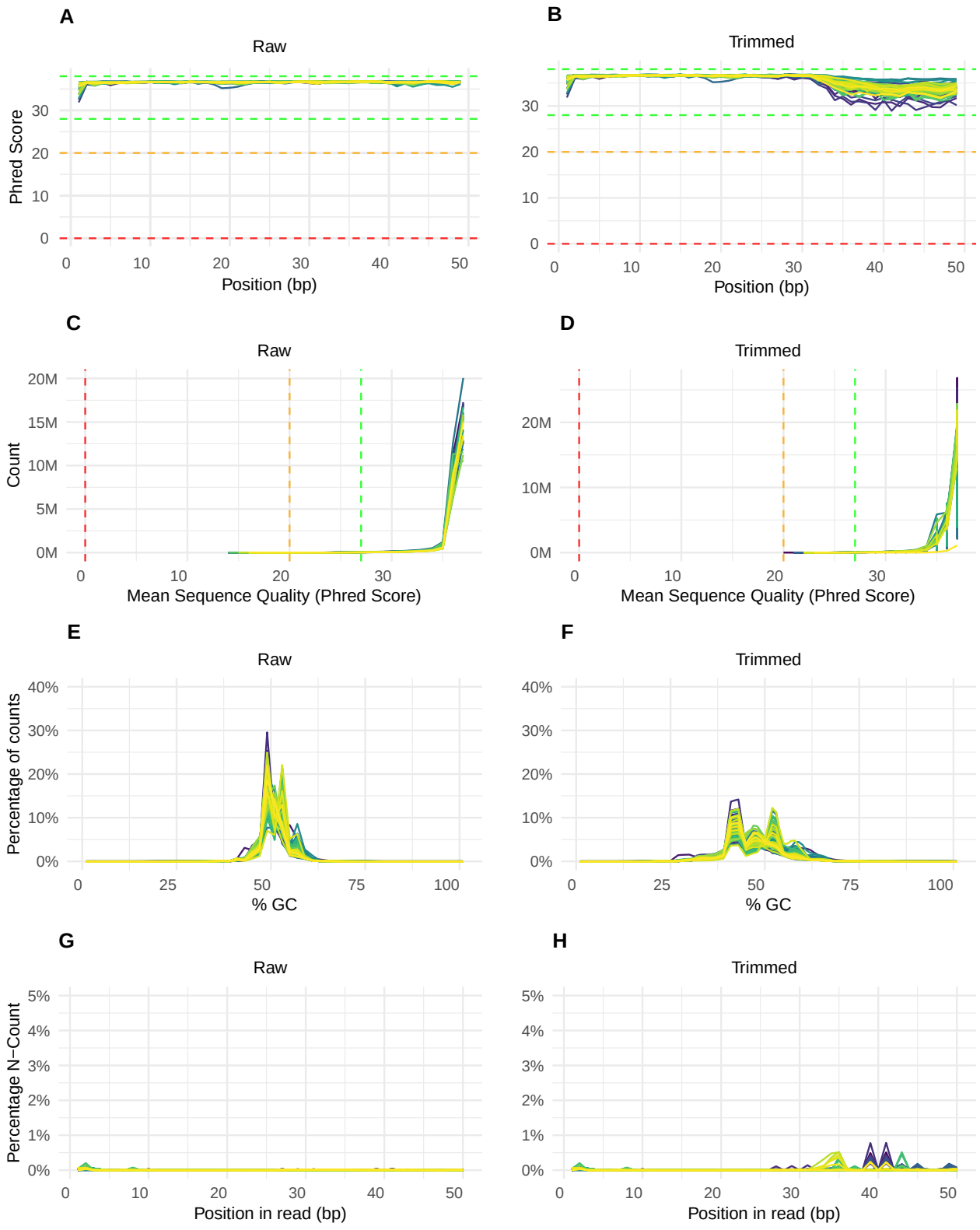

Figure 4.3: Quality Control Analysis. *A* and *B*: Mean quality values of sequences across all bases after and before using fastp (v0.23.4); *C* and *D*: Per Sequence Quality Scores across all bases after and before using fastp; *E* and *F*: Per Sequence GC Content Raw after and before using fastp; *G* and *H*: Read N content after and before using fastp.

However, the distribution of sequence lengths is irregular across all samples Figure 4.4. After trimming, the majority of sequences cluster around a length of 20-25 nucleotides, although a smaller subset of sequences with lengths between 29-32 nucleotides is also observed. This indicates the presence of different types of small RNAs. Generally, miRNAs are found in the range of 20-24 bp, with a peak at 22 bp potentially indicating an abundant population of miRNAs. In contrast, tRNAs can vary in length but are commonly found at lengths that may include higher peaks, such as at 31 bp. This may reflect the presence of tRNAs derived from the degradation of double-stranded RNA or transposons.

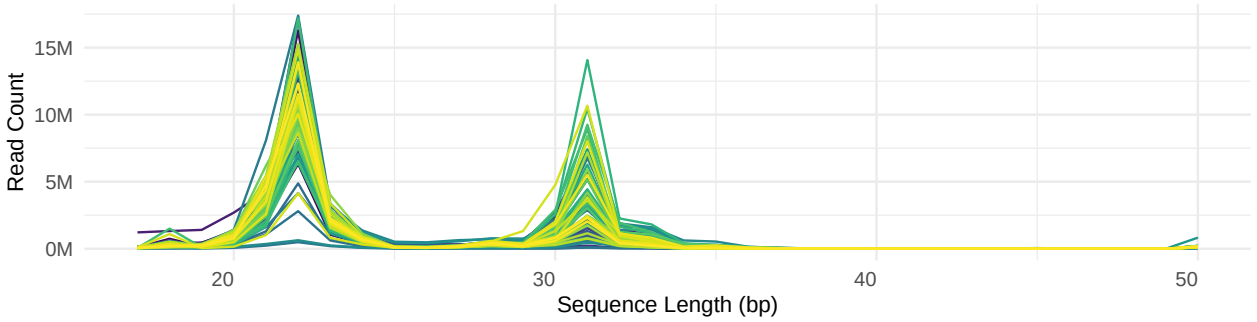

Figure 4.4: FastQC: Sequence Length Distribution

In the Figure 4.5 shows the total number of reads per sample is around 25 million, both in Raw and Trimmed, with means of 25.47 and 25.23 million, respectively. The interquartile ranges (IQR) of these measurements are similar, approximately 4.2 million, reflecting a high consistency among the samples.

Regarding duplicate reads, they dominate the data, with means of 25.10 million in Raw and 24.95 million in Trimmed, and an IQR of about 4.2 million in both conditions. This indicates that more than 90% of the sequences are duplicated, a high value but expected in smRNAseq samples, given the nature of the short reads of 20 to 24 nucleotides.

In contrast, unique reads are significantly less frequent, with means of 0.38 million in Raw and 0.28 million in Trimmed. The interquartile ranges for these measurements are also low, around 0.13 million in Raw and 0.10 million in Trim.

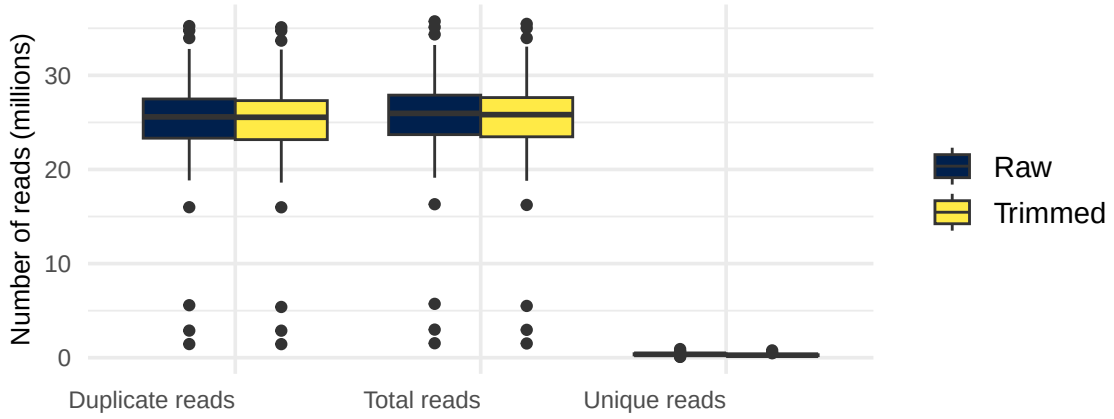

Figure 4.5: Number of reads from small RNA-seq. Total reads before and after trimming of adapters

4.3 miRNA Quality Control

The *nf-core/smrnaseq* pipeline performs a quality analysis specific to smRNAseq data using *miRTrace*. This analysis assesses sequencing quality, identifies the presence of miRNA and undesired sequences from tRNA, rRNA, or Illumina artifact sequences, and identifies clade-specific miRNA profiles based on a comprehensive catalog of previously identified miRNA families.

In the annotation step Figure 4.6, the mapped reads against reference databases revealed that a mean of 65.06% of analyzed sequences per sample corresponded to miRNA precursors, with a range of 22.51% to 91.77%. Other categories included 21.16% tRNA sequences, 9.16% unknown sequences, 4.55% rRNA sequences, and 0.05% artifacts.

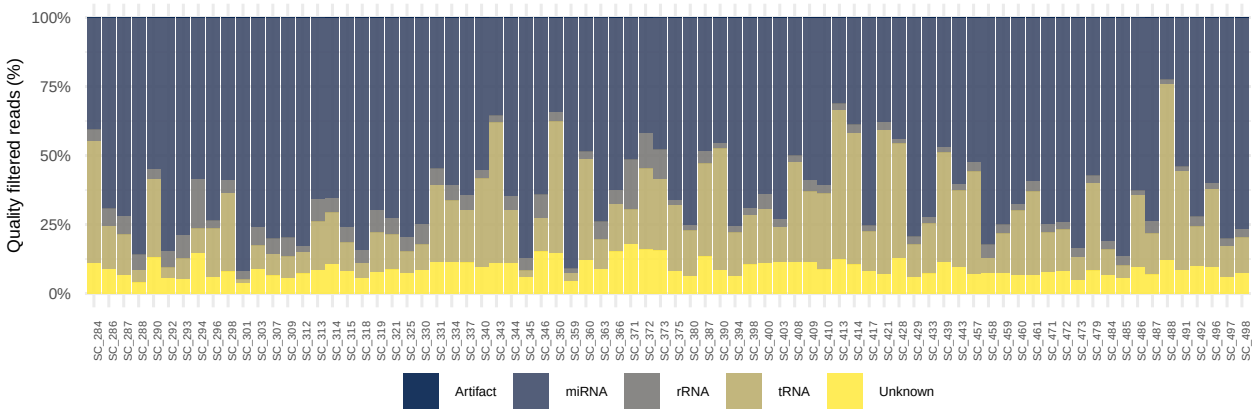

Figure 4.6: *miRTrace* (v1.0.1) Analysis: RNA Categories

During the contamination assessment step Figure 4.7, mapping miRNA precursor sequences against the clade-specific miRNA catalog showed that a mean of 94.14% of analyzed sequences belonged to the human category, with a range of 67.21% to 99.94%. Minor contributions from other clades, such as Rodentia (2.07%), Dicots (2.73%), Insects (0.58%), and Monocots (0.38%), were detected at low proportions. These identifications could result from contamination, such as incorrect index assignment during sample demultiplexing, or may have a biological origin.

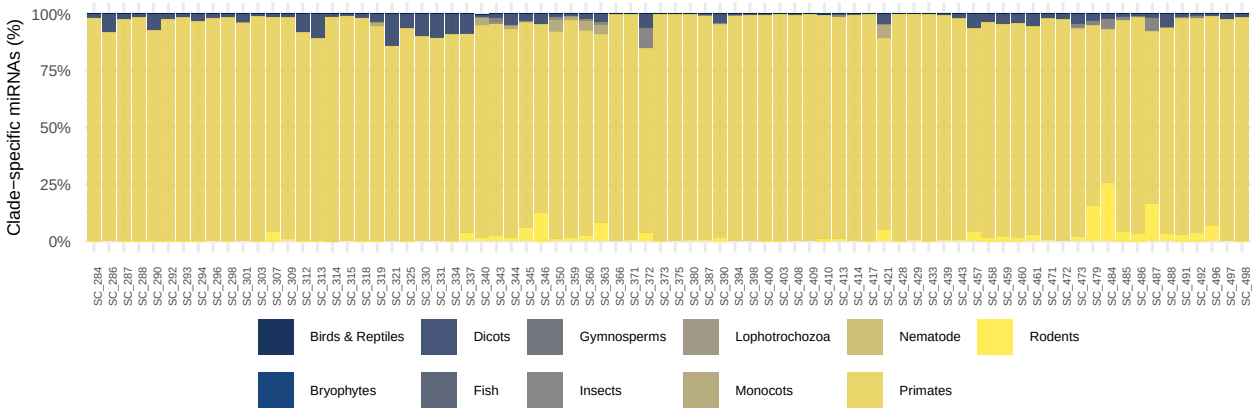

Figure 4.7: *miRTrace* (v1.0.1) Analysis: Contamination Check

4.4 miRNA Quantification

In this step of the *nf-core/smrnaseq pipeline*, the alignment of reads is conducted sequentially against databases of mature miRNAs, precursor miRNAs, and a combined database of both. These alignments enable the identification and quantification of miRNAs. The statistics obtained for mature miRNAs, precursor miRNAs, and the combination of both against the reference genome are presented in Table 4.3. On average, 52.07% of the mature miRNAs were mapped, with a range between 17.72% and 79.75%.

For precursor miRNAs, an average of 32.33% of sequences was identified, ranging from 6.31% to 68.83%. Subsequently, the reads of mature and precursor miRNAs were aligned against a reference genome, not for miRNA identification and quantification but as a quality control for the sequences. In this regard, the mean percentage of aligned reads was around 43% of the total reads. Across the sample set, minimum values of 10.28% and maximum values of 73.31% were observed. For more details, the metrics for each sample can be found in https://github.com/joshoandres13/miRNAs.

Table 4.3: Descriptive statistics of alignment with samtools of all samples. *TM*: Mean of Total Mapped (reads); *TU*: Mean of Total Unmapped (reads): *Mean M*: Mean Mapped (%); *Max M*: Max Mapped (%); *Min M*: Min Mapped (%)

Group	TM	TU	Mean M	Max M	Min M
mature	13166950	12194855	52.07	79.75	17.72
mature_hairpin	3644435	8836242	32.33	68.83	6.31
mature_hairpin_genome	4334118	4565712	43.87	73.31	10.28

4.5 IsomiR Annotation

Mirtop (v0.4.28) was used for the annotation of miRNAs and isomiRs. In figure Figure 4.8, the *mean isomiR read counts* are presented, which refer to an average calculation that helps describe how the reads of isomiRs are distributed within a dataset. Among the annotated isomiR sequences per sample, the *reference miRNA* accounts for 96.51%, with ranges varying from 88.70% to 98.13%. On the other hand, the distributions of isomiR variants are as follows: *3' Isoform* (1.36%), *3' Addition* (1.23%), *5' Isoform* (0.64%), *SNVs in the Central Offset Region* (0.05%), *SNVs in the Central Region* (0.06%), *Supported SNVs in the Central Region* (0.05%), and *SNV in Seed Region* (0.05%), all of which are below 1.40%.

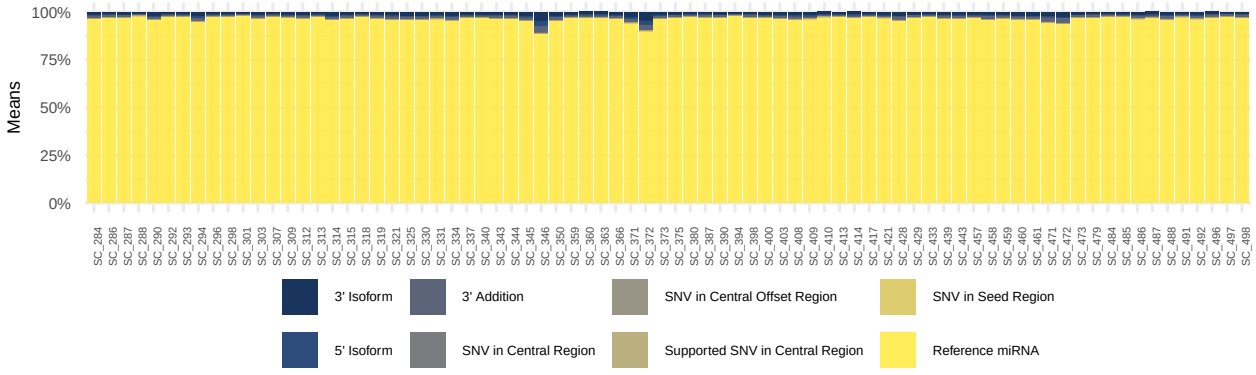

Figure 4.8: Annotation of miRNAs and isomiRs with *mirtop* (v0.4.28): Mean isomiR read counts

4.6 Differential expression analysis

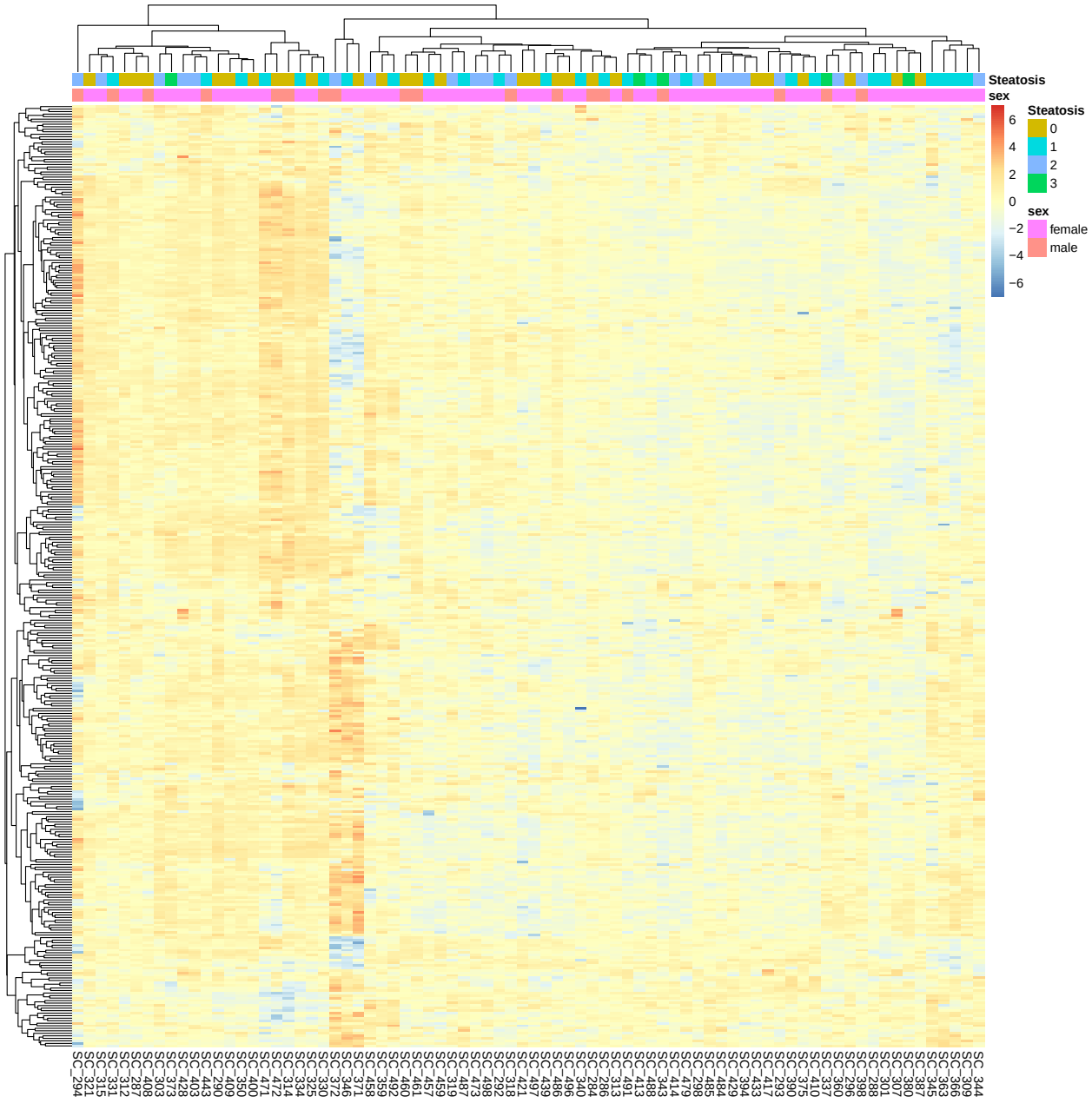

Figure 4.9: The heatmap illustrates the expression of 374 individual miRNA sequences across the analyzed samples. In this visualization, shades of red indicate increased miRNA expression, whereas shades of blue denote reduced or absent miRNA expression. Although a substantial number of miRNAs were identified, no distinct grouping patterns emerged among the analyzed samples, suggesting heterogeneity in miRNA expression profiles across the dataset.

Differential expression analysis is a bioinformatics and statistical technique used to identify genes, proteins, or other biomolecules that exhibit significant differences in expression levels between two or more biological conditions. For this analysis, the sequences of miRNAs of reference were

utilized, as annotated in the previous analysis.

Under the hood of *DESeq2*, we applied the *regularized logarithm* (*rlog*) transformation for normalization, which incorporates a prior on the sample differences[64]. This step is crucial because the standard logarithmic transformation can be sensitive to low expression values, whereas *rlog* addresses this issue by introducing regularization that helps stabilize variability between samples. This method facilitates seamless integration into subsequent analyses, as demonstrated in Figure 4.9.The figure shows a total of 374 miRNAs were found to be expressed differentially in the overall cohort when patients (samples) were categorized by sex and steatosis.

Since we were primarily interested in the transcriptional changes that may precede the initiation of the steatosis process, we focused on contrasting individuals across different levels of steatosis. In this experiment, the goal is to identify miRNAs that are expressed differentially at various levels of steatosis. In our differential expression analysis using *DESeq2* with the LRT model, we initially identified 169 miRNAs as upregulated and 205 downregulated in the subcutaneous white adipose tissue (scWAT) of individuals according their steatosis degree. Comparing these four groups and applying a false discovery rate (FDR) cutoff of less than 0.05, the number of statistically siginificant differentially expressed miRNAs was greatly reduced to 2 (Figure 4.10). Of these, we found *hsa-miR-372-3p* is upregulated and *hsa-miR-144-3p* is downregulated (Table 4.4).

Table 4.4: Differentially expressed miRNAs in subcutaneous white adipose tissue (scWAT). *FDR* : False Discovery Rate

	log2FoldChange	**pvalue**	**FDR**
hsa-miR-144-3p	-3.319	p< 0.0002	0.04
hsa-miR-372-3p	1.085	p< 0.0002	0.04

Figure 4.10: Plot showing differentially expressed miRNAs (red) in subcutaneous white adipose tissue (scWAT) according to the four groups of steatosis

Following their identification, we further analyzed the expression patterns of these two differentially expressed miRNAs by representing their normalized counts in boxplots across the four

steatosis groups Figure 4.11. This approach allowed us to observe how the expression of these significant miRNAs varies across different degrees of steatosis. The boxplots highlight the consistency of expression changes across groups, reinforcing the relevance of these miRNAs in the context of steatosis severity. The miRNA *hsa-miR-144-3p* exhibits a decrease in expression as the level of steatosis increases, whereas *hsa-miR-372-3p* shows a trend of increasing expression across the different levels of steatosis.

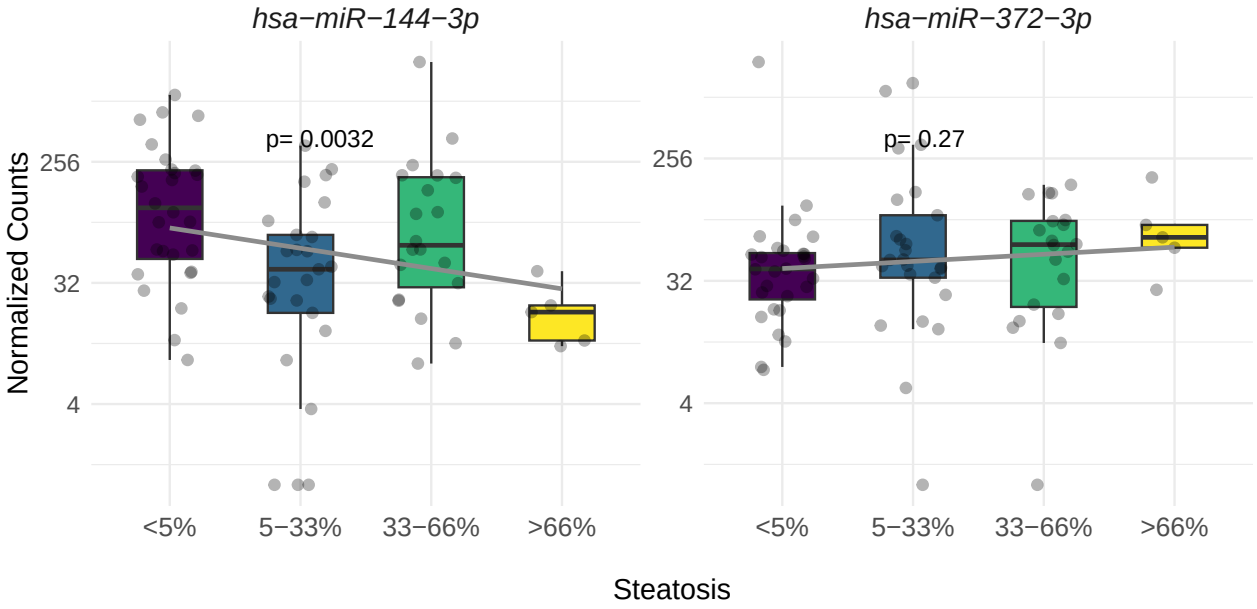

Figure 4.11: Boxplots of the differentially expressed miRNAs in subcutaneous white adipose tissue (scWAT) according to the four groups of steatosis. Each box represents the interquartile range (IQR) of the normalized counts, with the line inside the box indicating the median. The whiskers extend to show the range of the data, excluding outliers, which are displayed as individual points. p: p-value fot the Kruskal–Wallis test for the comparison between groups.

4.7 Target mRNA Selection and Validation

4.7.1 *hsa-miR-372-3p*

Potential target genes of the selected miRNAs were identified using the *multiMiR* package. This analysis included only validated interaction data, ensuring the biological relevance of the identified mRNAs. A subsequent filtering process was applied to select targets based on the type of experimental validation (e.g., luciferase assays, Western blot, or qRT-PCR) and functional support. As shown in Table 4.5, 24 target genes were identified after filtering.

Table 4.5: Selected interactions after filtering by database, experiment type (including luciferase assays, Western blot, or qRT-PCR), functional support (Functional MTI) and validated type for *hsa-miR-372-3p*.

Gene	Ensembl Identifier	Type
LATS2	ENSG00000150457	validated
TGFBR2	ENSG00000163513	validated

NFIB	ENSG00000147862	validated
CDKN1A	ENSG00000124762	validated
VEGFA	ENSG00000112715	validated
TNFAIP1	ENSG00000109079	validated
TRPS1	ENSG00000104447	validated
MBNL2	ENSG00000139793	validated
RHOC	ENSG00000155366	validated
NR4A2	ENSG00000153234	validated
ERBB4	ENSG00000178568	validated
CDK2	ENSG00000123374	validated
LEFTY1	ENSG00000243709	validated
BTG1	ENSG00000133639	validated
WEE1	ENSG00000166483	validated
CCNA1	ENSG00000133101	validated
DKK1	ENSG00000107984	validated
PHLPP2	ENSG00000040199	validated
ATAD2	ENSG00000156802	validated
ADAMTS9	ENSG00000163638	validated
CADM2	ENSG00000175161	validated
ZBTB7A	ENSG00000178951	validated
TXNIP	ENSG00000265972	validated
KLF13	ENSG00000275746	validated

The target genes identified through multiMiR were subjected to functional enrichment analysis using KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways[60]. The KEGG pathway enrichment analysis identified five significantly enriched pathways ($p < 0.05$) (Figure 4.12). The most prominent pathways include:

- *Cellular senescence* (KEGG ID: hsa04218): associated with cellular aging and the permanent arrest of the cell cycle in response to genetic damage, oxidative stress, or oncogenic signals.
- *Cell cycle* (KEGG ID: hsa04110): involved in the regulation and progression of the cell cycle, including critical checkpoints that ensure the fidelity of cell division.
- *Hepatitis B* (KEGG ID: hsa05161): related to infection by the hepatitis B virus and the inflammatory and carcinogenic processes it can trigger in the liver.
- *PI3K-Akt signaling pathway* (KEGG ID: hsa04151): a key pathway regulating cellular processes such as proliferation, survival, and metabolism, frequently deregulated in cancer.
- *Pancreatic cancer* (KEGG ID: hsa05212): associated with molecular and cellular alterations characteristic of this cancer type, including aberrant signaling and apoptosis resistance.

These pathways are primarily linked to essential processes such as cell cycle regulation, intracellular signaling in cancer, and cellular responses to damage and infection.

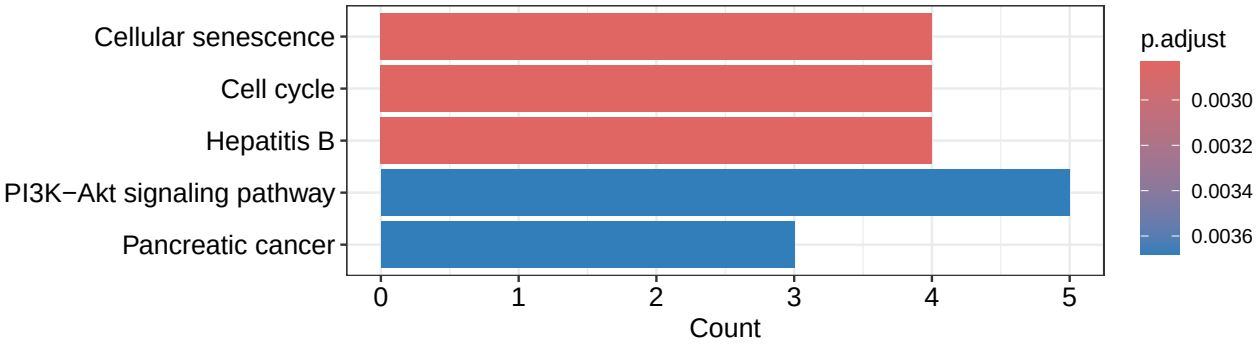

Figure 4.12: Top 5 Significant KEGG Pathways of *hsa-miR-372-3p*

4.7.2 *hsa-miR-144-3p*

Table 4.6 shows 22 target genes were identified for *hsa-miR-372-3p*, filtered by validated interactions and experimental evidence.

Table 4.6: Selected interactions after filtering by database, experiment type (including luciferase assays, Western blot, or qRT-PCR), functional support (Functional MTI) and validated type for *hsa-miR-144-3p*.

Gene	Ensembl Identifier	Type
NOTCH1	ENSG00000148400	validated
PLAG1	ENSG00000181690	validated
ZEB1	ENSG00000148516	validated
ZEB2	ENSG00000169554	validated
IRS1	ENSG00000169047	validated
MAP3K8	ENSG00000107968	validated
EZH2	ENSG00000106462	validated
APP	ENSG00000142192	validated
PTGS2	ENSG00000073756	validated
MET	ENSG00000105976	validated
ETS1	ENSG00000134954	validated
TGFB1	ENSG00000105329	validated
CFTR	ENSG00000001626	validated
FGG	ENSG00000171557	validated
MTOR	ENSG00000198793	validated
SMAD4	ENSG00000141646	validated
NFE2L2	ENSG00000116044	validated
PBX3	ENSG00000167081	validated
TTN	ENSG00000155657	validated
TUG1	ENSG00000253352	validated
PTEN	ENSG00000284792	validated
XIST	ENSG00000229807	validated

The target genes identified through multiMiR were subjected to functional enrichment analysis

using KEGG where five significantly enriched pathways ($p < 0.05$) (Figure 4.13). The most prominent pathways include:

- *MicroRNAs in cancer* (KEGG ID: hsa05206): Highlighting the role of miRNAs in the regulation of gene expression, particularly in pathways associated with tumorigenesis and cancer progression.
- *Hepatocellular carcinoma* (KEGG ID: hsa05225): Specifically linked to liver cancer and the molecular mechanisms underlying its development and progression.
- *FoxO signaling pathway* (KEGG ID: hsa04068): A critical pathway regulating oxidative stress response, apoptosis, and metabolism, which plays a significant role in cancer and aging.
- *Gastric cancer* (KEGG ID: hsa05226): Associated with molecular and cellular alterations characteristic of gastric tumorigenesis.
- *Cellular senescence* (KEGG ID: hsa04218): Related to aging and the permanent arrest of the cell cycle due to stress signals or damage.

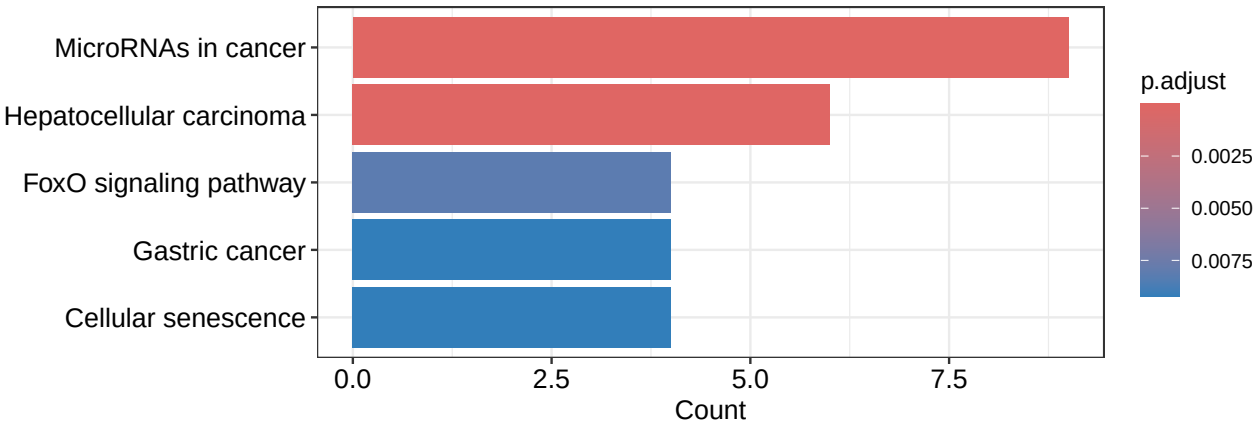

Figure 4.13: Top 5 Significant KEGG Pathways of *hsa-miR-144-3p*

These pathways reflect key biological processes, including tumorigenesis, stress response, and cell cycle regulation, underscoring the importance of miRNAs in these mechanisms. The findings provide valuable insights into potential molecular interactions and pathways relevant to the studied system.

4.8 The *hsa-miR-372-3p* and *hsa-miR-144-3p* Modulate Lipid Metabolism Genes in a HepG2 Steatosis Model.

The gene expression was evaluated in HepG2 cells transfected for 24 h and treated with Oleic Acid (OA) after 24 h to assess the impact of miRNA in a cell model mimicking hepatic steatosis. Specifically, we investigated the relative expression of *PPARG* (Peroxisome Proliferator Activated Receptor Gamma), a key regulator of adipogenesis and insulin sensitivity; *PNPLA2* (Patatin Like Phospholipase Domain Containing 2), which encodes the Adipose Triglyceride Lipase (*ATGL*) involved in triglyceride breakdown; *DGAT2* (Diacylglycerol O-Acyltransferase 2), a critical enzyme

for triglyceride synthesis; *FAS* (Fas Cell Surface Death Receptor), a mediator of apoptosis and cellular stress; and *ACACA* (Acetyl-CoA Carboxylase Alpha), the rate-limiting enzyme in de novo fatty acid synthesis. These genes play pivotal roles in lipid metabolism and, to a lesser extent, glucose metabolism, making them relevant targets for understanding the molecular mechanisms underlying hepatic steatosis.

The *hsa-miR-372-3p* mimic significantly downregulated the mRNA expression of the target gene *ACACA* (-91.40% ± 3.39; $p < 0.05$) in OA-treated HepG2 cells. A similar, although not statistically significant, decrease was observed in the mRNA levels of *FAS* (-90.50% ± 4.30; $p < 0.093$), suggesting a trend towards reduced expression. In contrast, the mRNA levels of *DGAT2* (-73.7% ± 6.70; $p = 0.462$) and *PNPLA2* (-55.10% ± 19.56; $p = 0.243$) did not show significant downregulation. Notably, the mRNA levels of *PPARG* remained unchanged compared to the negative control cells (Figure 4.14).

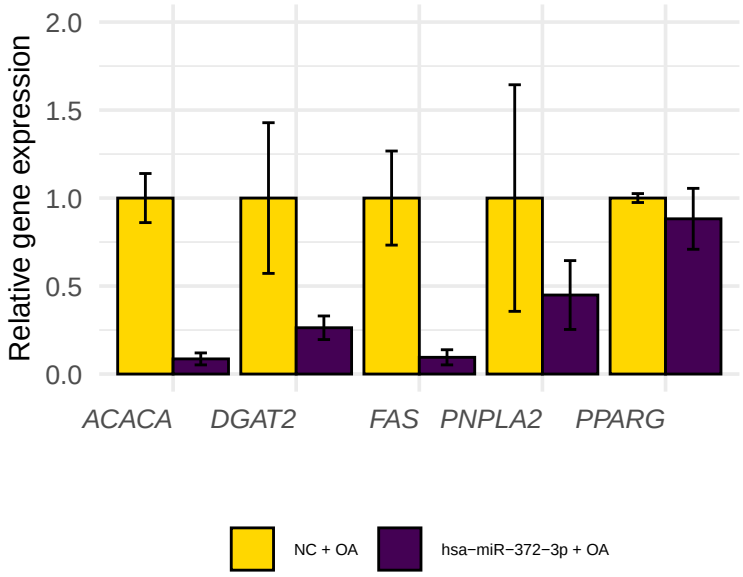

Figure 4.14: Gene expression analysis of HepG2 cells transfected with *hsa-miR-372-3p* mimic and treated with oleic acid to mimic in vitro hepatic steatosis was performed. The mRNA levels of protein-coding genes (*ACACA*, *DGAT2*, *FAS*, *PNPLA2*, *PPARG*), which are involved in glucose and lipid metabolism, were assessed. HepG2 cells were transfected with 50 nM of mirVana mimic *hsa-miR-372-3p* (5'-AAAGUGCUGCGACAUUUGAGCGU-3') along with a randomized sequence as a control (Negative Control 1). Twenty-four hours post-transfection, the cells were treated with oleic acid (0.5 mM) for 24 hours. The results are presented as mean relative gene expression ± standard error of the mean (SEM) (n = 3). Abbreviations: NC (negative control), OA (oleic acid), *ACACA* (Acetyl-CoA Carboxylase Alpha), *DGAT2* (Diacylglycerol O-Acyltransferase 2), *FAS* (Fatty Acid Synthase), *PNPLA2* (Patatin-like phospholipase domain-containing protein 2), *PPARG* (Peroxisome Proliferator Activated Receptor Gamma).

On the other hand, inhibition of *hsa-miR-144-3p* upregulated the expression of the putative target genes. The relative gene expression levels of the evaluated targets in HepG2 cells transfected and treated with Oleic Acid (OA) were as follows: *ACACA* (96.80% ± 53.79; $p = 0.186$), *FAS* (135% ± 160.34; $p = 0.429$), *DGAT2* (24.13% ± SD; $p = 0.128$), *PNPLA2* (48.83% ± 14.32; $p = 0.140$), and *PPARG* (64.09% ± 41.25; $p = 0.239$) (Figure 4.14). Although the expression levels varied across

genes, none of these changes were statistically significant ($p > 0.05$). These findings suggest that the regulatory effect of *hsa-miR-144-3p* on these genes may be limited under the experimental conditions employed.

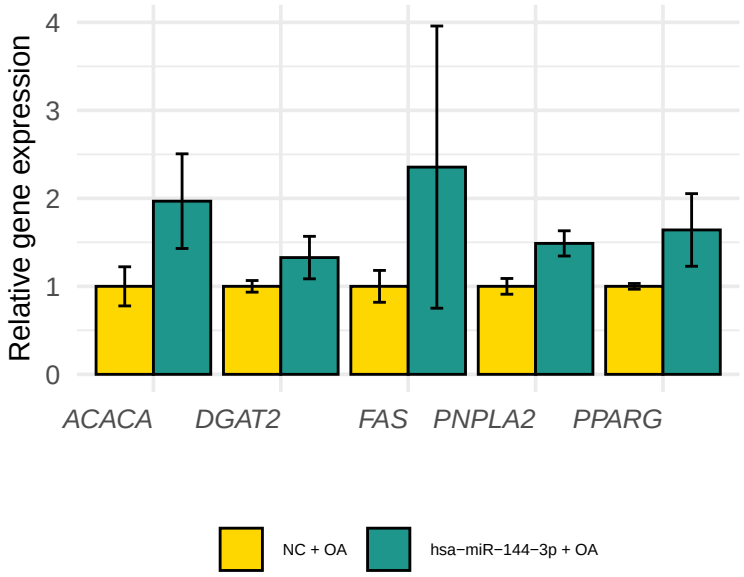

Figure 4.15: Gene expression analysis of HepG2 cells transfected with *hsa-miR-144-3p* inhibitor and treated with oleic acid to mimic in vitro hepatic steatosis was performed. The mRNA levels of protein-coding genes (*ACACA*, *DGAT2*, *FAS*, *PNPLA2*, *PPARG*), which are involved in glucose and lipid metabolism, were assessed. HepG2 cells were transfected with 50 nM of mirVana mimic *hsa-miR-144-3p* (5'-UACAGUAUAGAUGAUGUACU-3') along with a randomized sequence as a control (Negative Control 1). Twenty-four hours post-transfection, the cells were treated with oleic acid (0.5 mM) for 24 hours. The results are presented as mean relative gene expression ± standard error of the mean (SEM) (n = 3). Abbreviations: NC (negative control), OA (oleic acid), *ACACA* (Acetyl-CoA Carboxylase Alpha), *DGAT2* (Diacylglycerol O-Acyltransferase 2), *FAS* (Fatty Acid Synthase), *PNPLA2* (Patatin-like phospholipase domain-containing protein 2), *PPARG* (Peroxisome Proliferator Activated Receptor Gamma).

Chapter 5

Discussion

Chapter 6

Conclusions

References

1. Idilman, I. S., Ozdeniz, I., & Karcaaltincaba, M. (2016). Hepatic steatosis: Etiology, patterns, and quantification. *Seminars in Ultrasound, CT and MRI*, *37*, 501–510.
2. Angulo, P. (2007). Obesity and nonalcoholic fatty liver disease. *Nutrition Reviews*, *65*(suppl_1), S57–S63.
3. Wong, R. J. (2024). Epidemiology of metabolic dysfunction-associated steatotic liver disease (MASLD) and alcohol-related liver disease (ALD). *Metabolism and Target Organ Damage*, *4*(4), N–A.
4. Mary, E., Jeffrey, V., Ratziu, V., Sven, M., Arun, J., Kanwal, F., Romero, D., Manal, F., Quentin, M., Arab, J. P., et al. (2024). A multisociety delphi consensus statement on new fatty liver disease nomenclature. *Annals of Hepatology*, *29*(1), 1–15.
5. Thyfault, J. P., & Rector, R. S. (2020). Exercise combats hepatic steatosis: Potential mechanisms and clinical implications. *Diabetes*, *69*(4), 517–524.
6. Chan, W.-K., Chuah, K.-H., Rajaram, R. B., Lim, L.-L., Ratnasingam, J., & Vethakkan, S. R. (2023). Metabolic dysfunction-associated steatotic liver disease (MASLD): A state-of-the-art review. *Journal of Obesity & Metabolic Syndrome*, *32*(3), 197.
7. Younossi, Z. M. (2019). Non-alcoholic fatty liver disease–a global public health perspective. *Journal of Hepatology*, *70*(3), 531–544.
8. Ma, Y., Wang, J., Xiao, W., & Fan, X. (2024). A review of MASLD-related hepatocellular carcinoma: Progress in pathogenesis, early detection, and therapeutic interventions. *Frontiers in Medicine*, *11*, 1410668.
9. Sarwar, R., Pierce, N., & Koppe, S. (2018). Obesity and nonalcoholic fatty liver disease: Current perspectives. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 533–542.
10. Miao, L., Targher, G., Byrne, C. D., Cao, Y.-Y., & Zheng, M.-H. (2024). Current status and future trends of the global burden of MASLD. *Trends in Endocrinology & Metabolism*.
11. Younossi, Z. M., Koenig, A. B., Abdelatif, D., Fazel, Y., Henry, L., & Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*, *64*(1), 73–84.
12. Younossi, Z. M., Golabi, P., Avila, L. de, Paik, J. M., Srishord, M., Fukui, N., Qiu, Y., Burns, L., Afendy, A., & Nader, F. (2019). The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. *Journal of Hepatology*, *71*(4), 793–801.
13. Paik, J. M., Henry, L., Younossi, Y., Ong, J., Alqahtani, S., & Younossi, Z. M. (2023). The burden of nonalcoholic fatty liver disease (NAFLD) is rapidly growing in every region of the world from 1990 to 2019. *Hepatology Communications*, *7*(10), e0251.
14. Younossi, Z. M., Golabi, P., Paik, J. M., Henry, A., Van Dongen, C., & Henry, L. (2023). The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. *Hepatology*, *77*(4), 1335–1347.
15. Younossi, Z., Anstee, Q. M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., & Bugianesi,

E. (2018). Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention. *Nature Reviews Gastroenterology & Hepatology*, *15*(1), 11–20.
16. Caballería, L., Pera, G., Auladell, M. A., Torán, P., Muñoz, L., Miranda, D., Alumá, A., Casas, J. D., Sánchez, C., Gil, D., et al. (2010). Prevalence and factors associated with the presence of nonalcoholic fatty liver disease in an adult population in spain. *European Journal of Gastroenterology & Hepatology*, *22*(1), 24–32.
17. Kershaw, E. E., & Flier, J. S. (2004). Adipose tissue as an endocrine organ. *The Journal of Clinical Endocrinology & Metabolism*, *89*(6), 2548–2556.
18. Rosen, E. D., & Spiegelman, B. M. (2014). What we talk about when we talk about fat. *Cell*, *156*(1), 20–44.
19. Cinti, S. (2007). The adipose organ. *Adipose Tissue and Adipokines in Health and Disease*, 3–19.
20. Zwick, R. K., Guerrero-Juarez, C. F., Horsley, V., & Plikus, M. V. (2018). Anatomical, physiological, and functional diversity of adipose tissue. *Cell Metabolism*, *27*(1), 68–83.
21. Cinti, S. (2019). Anatomy and physiology of the nutritional system. *Molecular Aspects of Medicine*, *68*, 101–107.
22. Lopez-Yus, M., Hörndler, C., Borlan, S., Bernal-Monterde, V., & Arbones-Mainar, J. M. (2024). Unraveling adipose tissue dysfunction: Molecular mechanisms, novel biomarkers, and therapeutic targets for liver fat deposition. *Cells*, *13*(5), 380.
23. Ghesmati, Z., Rashid, M., Fayezi, S., Gieseler, F., Alizadeh, E., & Darabi, M. (2024). An update on the secretory functions of brown, white, and beige adipose tissue: Towards therapeutic applications. *Reviews in Endocrine and Metabolic Disorders*, *25*(2), 279–308.
24. Jialal, I., & Devaraj, S. (2018). Subcutaneous adipose tissue biology in metabolic syndrome. *Hormone Molecular Biology and Clinical Investigation*, *33*(1), 20170074.
25. Ibrahim, M. M. (2010). Subcutaneous and visceral adipose tissue: Structural and functional differences. *Obesity Reviews*, *11*(1), 11–18.
26. Kwok, K. H., Lam, K. S., & Xu, A. (2016). Heterogeneity of white adipose tissue: Molecular basis and clinical implications. *Experimental & Molecular Medicine*, *48*(3), e215–e215.
27. McQuaid, S. E., Humphreys, S. M., Hodson, L., Fielding, B. A., Karpe, F., & Frayn, K. N. (2010). Femoral adipose tissue may accumulate the fat that has been recycled as VLDL and nonesterified fatty acids. *Diabetes*, *59*(10), 2465–2473.
28. Koenen, M., Hill, M. A., Cohen, P., & Sowers, J. R. (2021). Obesity, adipose tissue and vascular dysfunction. *Circulation Research*, *128*(7), 951–968.
29. Scherer, P. E. (2006). Adipose tissue: From lipid storage compartment to endocrine organ. *Diabetes*, *55*(6), 1537–1545.
30. Song, Z., Xiaoli, A. M., & Yang, F. (2018). Regulation and metabolic significance of de novo lipogenesis in adipose tissues. *Nutrients*, *10*(10), 1383.
31. Carpentier, A. C. (2021). 100th anniversary of the discovery of insulin perspective: Insulin and adipose tissue fatty acid metabolism. *American Journal of Physiology-Endocrinology and Metabolism*, *320*(4), E653–E670.
32. Grabner, G. F., Xie, H., Schweiger, M., & Zechner, R. (2021). Lipolysis: Cellular mechanisms for lipid mobilization from fat stores. *Nature Metabolism*, *3*(11), 1445–1465.
33. Duncan, R. E., Ahmadian, M., Jaworski, K., Sarkadi-Nagy, E., & Sul, H. S. (2007). Regulation of lipolysis in adipocytes. *Annu. Rev. Nutr.*, *27*(1), 79–101.
34. Scheja, L., & Heeren, J. (2019). The endocrine function of adipose tissues in health and cardiometabolic disease. *Nature Reviews Endocrinology*, *15*(9), 507–524.
35. Friedman, J. (2019). *Leptin and the endocrine control of energy balance. Nat. Metab. 1, 754–764.*
36. Karbowska, J., & Kochan, Z. (2006). Role of adiponectin in the regulation of carbohydrate and lipid metabolism. *Journal of Physiology and Pharmacology*, *57*, 103.

37. Santulli, G. (2015). *MicroRNA: Basic science: From molecular biology to clinical practice* (Vol. 887). Springer.
38. Arrighetti, N., & Beretta, G. L. (2021). miRNAs as therapeutic tools and biomarkers for prostate cancer. *Pharmaceutics*, *13*(3), 380.
39. Torres-Perez, E., Valero, M., Garcia-Rodriguez, B., Gonzalez-Irazabal, Y., Calmarza, P., Calvo-Ruata, L., Ortega, C., Garcia-Sobreviela, M. P., Sanz-Paris, A., Artigas, J. M., et al. (2015). The FAT expandability (FATe) project: Biomarkers to determine the limit of expansion and the complications of obesity. *Cardiovascular Diabetology*, *14*, 1–8.
40. Kleiner, D. E., Brunt, E. M., Van Natta, M., Behling, C., Contos, M. J., Cummings, O. W., Ferrell, L. D., Liu, Y.-C., Torbenson, M. S., Unalp-Arida, A., et al. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*, *41*(6), 1313–1321.
41. Peltzer, A., Trigila, A., Pantano, L., Ewels, P., Wang, C., Espinosa-Carrasco, J., Schcolnicov, N., Mohr, C., bot, nf-core, Menden, K., Patel, H., Sturm, G., CKComputomics, Cabus, L., Keys, K. L., Guizard, S., Garcia, M. U., Syme, R., Talbot, A., … Tommaso, P. D. (2024). *Nf-core/smrnaseq: v2.4.0 - 2024-10-14 - gray zinc dalmation patch*. Zenodo. https://doi.org/10.5281/ZENODO.3456879
42. Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278.
43. Andrews, S. et al. (2010). *FastQC: A quality control tool for high throughput sequence data*. Cambridge, United Kingdom.
44. Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890.
45. Kang, W., Eldfjell, Y., Fromm, B., Estivill, X., Biryukova, I., & Friedländer, M. R. (2018). miRTrace reveals the organismal origins of microRNA sequencing data. *Genome Biology*, *19*, 1–15.
46. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, 1–10.
47. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008
48. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P., & Smyth, G. K. (2024). edgeR 4.0: Powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv*. https://doi.org/10.1101/2024.01.21.576131
49. Desvignes, T., Loher, P., Eilbeck, K., Ma, J., Urgese, G., Fromm, B., Sydes, J., Aparicio-Puerta, E., Barrera, V., Espin, R., et al. (2020). Unification of miRNA and isomiR research: The mirGFF3 format and the mirtop API. *Bioinformatics*, *36*(3), 698–703.
50. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047. https://doi.org/10.1093/bioinformatics/btw354
51. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org
52. R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
53. Posit team. (2023). *RStudio: Integrated development environment for r*. Posit Software, PBC. http://www.posit.co/
54. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K.,

Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

55. Pantano, L., & Escaramis, G. (2024). *isomiRs: Analyze isomiRs and miRNAs from small RNA-seq*. https://doi.org/10.18129/B9.bioc.isomiRs
56. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8
57. Carlson, M. (2024). *Org.hs.eg.db: Genome wide annotation for human*.
58. Ru, Y., Kechris, K. J., Tabakoff, B., Hoffman, P., Radcliffe, R. A., Bowler, R., Mahaffey, S., Rossi, S., Calin, G. A., Bemis, L., & Theodorescu, D. (2014). The multiMiR r package and database: Integration of microRNA target interactions along with their disease and drug associations. *Nucleic Acids Research*, *42*(17), e133. https://doi.org/10.1093/nar/gku631
59. Ru, Y., Mulvahill, M., Mahaffey, S., & Kechris, K. (n.d.). *multiMiR: Integration of multiple microRNA-target databases with their disease and drug associations*. https://github.com/KechrisLab/multiMiR
60. Tenenbaum, D., & Maintainer, B. P. (2024). *KEGGREST: Client-side REST access to the kyoto encyclopedia of genes and genomes (KEGG)*. https://doi.org/10.18129/B9.bioc.KEGGREST
61. Gómez-Lechón, M. J., Donato, M. T., Martínez-Romero, A., Jiménez, N., Castell, J. V., & O'Connor, J.-E. (2007). A human hepatocellular in vitro model to investigate steatosis. *Chemico-Biological Interactions*, *165*(2), 106–116.
62. Belfort, R., Harrison, S. A., Brown, K., Darland, C., Finch, J., Hardies, J., Balas, B., Gastaldelli, A., Tio, F., Pulcini, J., et al. (2006). A placebo-controlled trial of pioglitazone in subjects with nonalcoholic steatohepatitis. *New England Journal of Medicine*, *355*(22), 2297–2307.
63. Rao, X., Huang, X., Zhou, Z., & Lin, X. (2013). An improvement of the 2^ (–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, *3*(3), 71.
64. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 1–21.