



UNIVERSITÀ DEL PIEMONTE ORIENTALE

Piemonte Orientale University

Department of Translational Medicine

PhD Program in Medical Sciences and Biotechnologies

XXXVIII Cycle

Thesis

**A COMPUTATIONAL PIPELINE FOR MULTI-OMICS REGULATORY
NETWORK INFERENCE IN VASCULAR AND GLIOBLASTOMA MODELS**

Supervisor: Professor Davide Corà
Co-tutor: Professor Giuliana Pelicci

PhD Candidate: Teresa Gravina

Coordinator: Professor Marisa Gariglio

Academic Years: 2022/2026

INDEX

1. LIST OF ABBREVIATIONS	4
2. SUMMARY	7
3. INTRODUCTION	10
3.1 Regulatory network to study disease	11
3.2 Multilevel regulatory network	13
3.2.1 Transcription factors and ChIP-Seq analysis.....	14
3.2.2 Non-coding RNA: miRNAs and circRNAs.....	16
3.2.3 Protein coding genes and transcriptomic analysis by mRNA-Seq.....	20
3.3 Regulatory network in endothelium	21
3.3.1 TFEB	21
3.3.2 TFEB in endothelium.....	23
3.4 Regulatory network in Glioblastoma	24
3.4.1 Glioblastoma tumor initiating cells	25
3.4.2 LSD1	26
4. AIM OF THE STUDY	29
5. MATERIALS AND METHODS	32
5.1 Computational environment: Linux and programming languages	33
5.2 Promoter definition of genes and miRNAs	33
5.3 Databases and biological resources	34
5.3.1 JASPAR	34
5.3.2 miRBase	34
5.3.3 TargetScan and miRTarBase	35
5.3.4 CSCD2.....	35
5.3.5 UCSC Genome Browser	36
5.3.6 ENCODE.....	36
5.4 Bioinformatics tools	37
5.4.1 BEDTools	37
5.4.2 IPA	37
5.4.3 ToppGene Suite	38
5.4.4 Cytoscape and NetworkX.....	38
5.5 Overview of the methods used for the construction of the regulatory network in endothelium – first project	39
5.5.1 Sample preparation	40

5.5.2 mRNA-Seq analysis	40
5.5.3 miRNA-Seq analysis	41
5.5.4 TFEB ChIP-Seq analysis	41
5.5.5 Bioinformatics data integration.....	42
5.5.6 Functional analysis of DEGs	42
5.5.7 Experimental validation	43
5.6 Overview of the methods used for the construction of the regulatory network in GBM – second project.....	44
5.6.1 Faletti et al., 2021 dataset	45
5.6.2 mRNA-Seq analysis	45
5.6.3 LSD1 ChIP-Seq analysis	46
5.6.4 ATAC-Seq analysis.....	46
5.6.5 Bioinformatics data integration.....	46
5.6.6 Wang et al., 2021 dataset.....	48
6. RESULTS.....	50
6.1 Results from the regulatory network in endothelium – first project	51
6.1.1 The TFEB-regulated transcriptional program in HUVEC.....	53
6.1.2 The TFEB-regulated miRNA program in HUVEC	55
6.1.3 Regulatory network of TFEB-associated miRNAs and their target genes	58
6.1.4 Experimental validation of miR-222	60
6.2 Results from the regulatory network in GBM – second project	62
6.2.1 Combined Analysis of LSD1 Binding Sites and ATF4 Motifs in Gene and miRNA Promoter Regions.....	63
6.2.2 Transcriptomic analysis of differentially expressed genes following LSD1 silencing in BT165	65
6.2.3 Identification of miR-124-3p as an upstream regulator of DEGs	69
6.2.4 LSD1-Dependent regulation of miR-124-3p and its target genes	70
6.2.5 Evidence from Wang et al. 2021 shows miR-124-3p downregulation in GBM.....	72
6.2.6 Multi-omics analysis of the Wang et al., 2021 dataset	74
7. DISCUSSION.....	80
7.1 Discussion of the endothelial regulatory network project	81
7.2 Discussion of the glioblastoma regulatory network project.....	86
8. CONCLUSION	91
9. ADDITIONAL WORKS	93
9.1 Bioinformatic investigation of diacylglycerol kinases in acute myeloid leukemia.....	94

9.2 Metagenomic analysis of gut microbiota alterations across disease progression in an Alzheimer's disease mouse model	96
10. BIBLIOGRAPHY	98
11. LIST OF PUBLICATIONS	116

1. LIST OF ABBREVIATIONS

Abbreviation	Full Form
3' UTR	3'-UnTranslated Region
AD	Activating domain / Alzheimer's disease
AML	Acute myeloid leukemia
bHLH	Basic helix-loop-helix domain
cCRE	Candidate regulatory element
ChIP-Seq	Chromatin Immunoprecipitation followed by Sequencing
CLEAR elements	Coordinated Lysosomal Expression and Regulation elements
CiRNA	Circular intronic RNA
CSCs	Cancer Stem Cells
cRE	Cis-regulatory element
DEM	Differentially expressed miRNA
DEG	Differentially expressed gene
DGK	Diacylglycerol kinase
EcircRNA	Exonic circular RNA
ElciRNA	Exon-intron circular RNA
EC	Endothelial cell
FBL	Feedback loop
FFL	Feed-forward loop
GBM	Glioblastoma
GBM TICs	Glioblastoma Tumor Initiating Cells
HUVEC	Human Umbilical Vein Endothelial Cells
IPA	Ingenuity Pathway Analysis
KDM	Histone lysine demethylase
LSD	Lysosomal storage disease
miRNA	MicroRNA
miRNA-Seq	MicroRNA Sequencing
mRNA	Messenger RNA
mRNA-Seq	mRNA Sequencing
MIT/TFE	Microphthalmia/Transcription Factor E

MITF	Microphthalmia-associated Transcription Factor
ncRNA	Non-coding RNA
NGS	Next Generation Sequencing
NSCLC	Non-small cell lung cancer
Pri-miRNA	Primary microRNA
Pre-miRNA	Precursor microRNA
T-ALL	T-cell acute lymphoblastic leukemia
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TFEB	Transcription factor EB
TFEC	Transcription Factor EC
TFE3	Transcription Factor E3
TICs	Tumor Initiating Cells
TSS	Transcription Start Site
WHO	World Health Organization
Zip	Leucine zipper domain

2. SUMMARY

Regulatory networks are fundamental tools for studying complex diseases, as they allow for the integration and interpretation of interactions between genes, proteins, and non-coding RNAs, providing a systemic view of the molecular mechanisms underlying physiological and pathological processes in a living cell. In this thesis, we describe the development and application of a bioinformatics pipeline for the study of gene regulatory networks using multi-omics data, in humans. The pipeline is applied to two specific cases: the transcription factor TFEB in endothelial cell and the epigenetic regulator LSD1 in Glioblastoma.

In the first project, dedicated to the study of the transcription factor TFEB in HUVECs, a regulatory network integrating protein-coding target genes and miRNAs was first reconstructed based on ChIP-Seq data. Various TFEB binding sites were identified in the promoter regions of both genes and pri-miRNAs, suggesting a direct role of the factor in their transcriptional regulation. The analysis of TFEB-bound differentially expressed genes and miRNAs highlighted a final set of 7 key miRNA candidates, including miR-222, which is involved in controlling the transition from the G1 to the S phase of the cell cycle through the negative regulation of the E2F1 and CDK1 genes. These results reveal a novel TFEB-mediated control of endothelial activity via miRNA-dependent post-transcriptional regulatory circuits, highlighting the importance of non-coding RNAs in modulating endothelial function and angiogenesis.

In the second project, we focused on the role of the epigenetic factor LSD1 and its ATF4 cofactor in Glioblastoma Tumor Initiating Cells (GBM TICs), using the BT165 cell line, representative of the proneural subtype. The integration of LSD1 ChIP-Seq and mRNA-Seq data allowed the identification of thousands of genomic regions bound by LSD1, many of which are co-occupied by ATF4. Ingenuity Pathway Analysis (IPA) identified miR-124-3p as an upstream regulator, suggesting a model in which LSD1 suppresses its expression. Consistently, LSD1 inhibition resulted in increased miR-124-3p levels and reduced its target genes, including DNAJC1, CD164, TRIB3, and MAGT1, known to promote proliferation, invasiveness, and therapeutic resistance in GBM. Integration of public data (Wang et al., 2021) confirmed the downregulation of miR-124-3p in GBM patients, supporting the hypothesis of an LSD1–miR-124-3p axis involved in tumorigenesis and the maintenance of stem-like properties. A further extension of the analysis allowed the definition of an epigenetic-post-transcriptional regulatory network that suggests a cooperative role for non-coding RNAs (miRNAs and circRNAs) in controlling LSD1-driven transcriptional programs.

A final part of this thesis is dedicated to the presentation of two additional research projects carried out during the doctoral period, namely a bioinformatic investigation of the role of diacylglycerol kinases in acute myeloid leukemia and a metagenomic analysis of gut microbiota alterations across disease progression in an Alzheimer's disease mouse model. Both the works were based on the computational analysis of big omic data.

Overall, the results highlight the value of multi-omics integration strategies in studying regulatory networks and infer biological properties of living systems. We confirm the crucial role of non-coding RNAs in modulating transcription factor function and the progression of complex diseases such as cancer. The developed pipeline therefore represents a useful tool for future investigations aimed at identifying novel therapeutic targets and molecular biomarkers, with potential translational applications in oncology and vascular disease.

3. INTRODUCTION

3.1 Regulatory network to study disease

Living beings are characterized by a wide variety of molecules, such as proteins, metabolites and non-coding RNA, which perform different biological functions. These molecules interact with each other through various modalities, including physical interactions, transcriptional induction and enzymatic activation.¹ The overall set of all these molecular interactions is called interactome, which in human is potentially characterized by more than 100,000 molecules, including approximately 25,000 protein-coding genes, 1000 metabolites and numerous other types of RNA.^{1,2,3}

Given the complexity and the large number of molecular actors involved in biological processes, an effective way to analyze their interactions is represented by networks. This type of approach, known as “network medicine”, is widely used in the study of the molecular bases of diseases.^{1,4} Furthermore, the development of modern omics technologies, such as genomics, transcriptomics, epigenomics, proteomics, metabolomics, and lipidomics, has significantly expanded the amount and diversity of available data. The integration of these heterogeneous data through multi-omics approaches allows for the construction of complex biological networks, which more accurately reflect the reality of biological systems.^{5,6}

From a mathematical point of view, a network consists of two fundamental components: nodes and edges. In the context of biology, nodes represent biological factors, such as proteins, RNAs, metabolites, genes, while edges represent interactions between two or more nodes (Figure 3.1A).^{1,4} Networks can be direct or indirect, depending on whether the interactions have an explicit directionality (for example, in the case of a gene regulating the transcription of another) or not (as in the physical interactions between two proteins).⁶ A key concept in biological networks is the one of hubs, nodes with higher number of connections than average, which often play a crucial role in the network.^{2,4}

The human interactome is an example of a scale-free network, characterized by a distribution of the number of connections per node that follows a power law distribution, so that most nodes have few connections, while a minority, represented by hubs, are highly connected. This type of organization is not random and gives robustness to the network, but at the same time makes it vulnerable to the loss of hubs.^{1,4} Figure 3.1B and Figure 3.1C illustrate respectively a scale-free network and a random network.

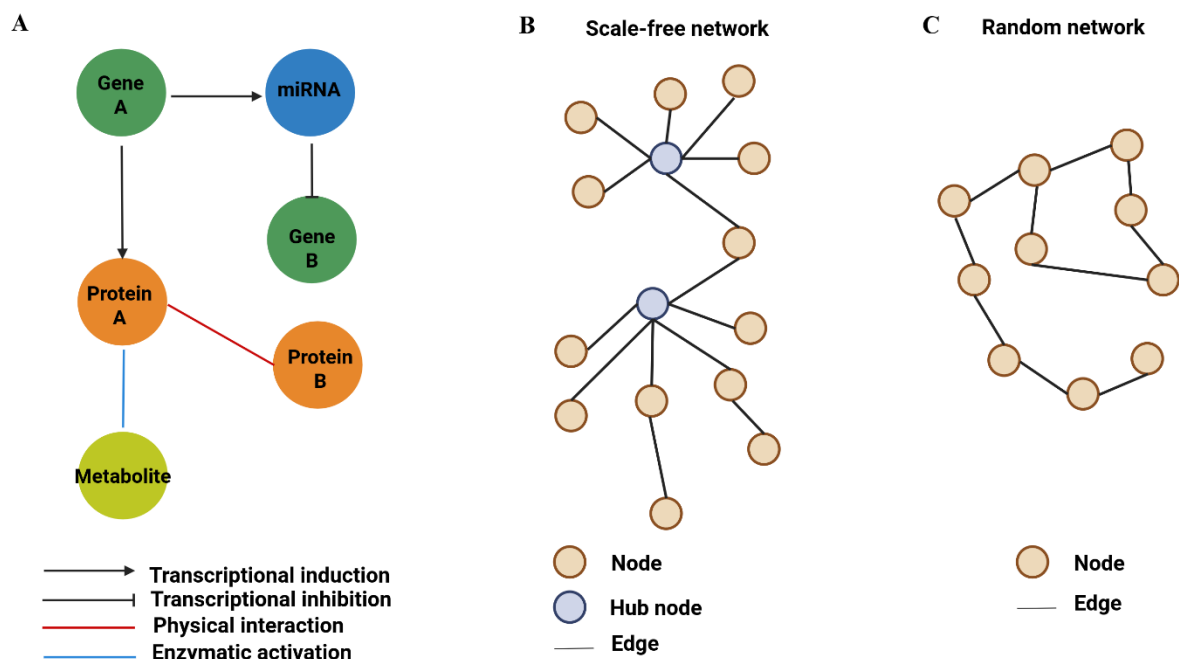


Figure 3.1. Representation of biological networks and their topological properties. (A) Example of a possible biological network representing different type interactions of biological molecules. Nodes represent various biological factors, such as genes (green), miRNA (blue), proteins (orange), and metabolites (yellow). Edges represent different types of interactions: transcriptional induction (black arrow), transcriptional inhibition (black blunt arrow), physical interaction (red line), and enzymatic activation (light blue line). (B) Example of a scale-free network, typical of biological networks. It is characterized by a power-law degree distribution in which most nodes have few connections while a few nodes, called hubs (highlighted in blue), are highly connected. (C) Example of a random network in which edges are distributed randomly among nodes, leading to a more uniform degree distribution and the absence of highly connected hubs. (Image created with BioRender).

Network medicine can be applied to represent and analyze different types of biological networks, thus defining specific aspects of the functional organization of cells. The main categories of molecular networks include protein–protein interaction (PPI) networks, in which nodes represent proteins linked by direct physical interactions; metabolic networks, in which nodes correspond to metabolites that are connected if involved in the same biochemical reactions; and regulatory networks, which establish directional relationships between transcription factors (TFs) and target genes, or between modifying enzymes, such as kinases, and their substrates, through post-translational mechanisms. Another class of networks is constituted by non-coding RNA networks, in particular those involving microRNAs (miRNAs) and circular RNAs (circRNAs), known for their role in the post-transcriptional regulation of gene expression. These networks allow to integrate information on RNA–mRNA, RNA–DNA and RNA–protein interactions, broadening the understanding of complex molecular mechanisms.^{1,2}

The use of disease networks has found numerous applications in the study of complex diseases, including for example neurodegenerative disorders and cancer, thus providing a characterization of the altered molecular modules and regulatory mechanisms underlying the disease phenotypes.^{7,8,9,10}

3.2 Multilevel regulatory network

In this work, we developed a pipeline based on the construction and analysis of complex regulatory networks in cancer, derived from the integration of epigenomic and transcriptomic data, with a specific focus on gene regulation mechanisms mediated by epigenetic factors and non-coding RNAs, including miRNAs and circRNAs. This multilevel approach is particularly suitable for the study of complex mechanisms implicated in tumorigenesis, thus allowing the integration of different sources of regulation acting on interconnected levels of gene expression.

Gene regulation is a complex process mediated by numerous elements including TFs and non-coding RNAs, such as miRNAs. TFs exert their function by binding to specific DNA sequences in gene promoters, thereby modulating the expression of target genes through transcription activation or repression. Instead, microRNAs regulate gene expression at the post-transcriptional level, acting mainly as repressors of the expression of their targets.^{11,12}

Given the multilevel and interconnected nature of these interactions, a widely used approach to describe them is represented by the construction of regulatory networks, in particular feed-forward loops (FFLs). In these models, a TF can directly regulate a miRNA and, together with it, a common target gene.^{12,13,14} Depending on the main regulatory node, different types of FFL can be distinguished: in the TF-FFL, the TF represents the main regulator and controls both the partner miRNA and the common target gene; in miRNA-FFL, instead, the miRNA plays the central role, regulating the TF and their shared target; finally, in the composite FFL, TF and miRNA regulate each other, establishing a more complex cross-regulatory network (Figure 3.2).

In addition to FFL, feedback loops (FBL) can also be observed, in which, for example, a TF activates a miRNA that in turn inhibits it, generating a self-regulatory circuit.

These regulatory network models, based on the interaction between TFs and miRNA, have been widely used in the study of several tumors, including acute myeloid leukemia (AML), T-cell acute lymphoblastic leukemia (T-ALL), breast cancer and glioblastoma (GBM).¹⁴

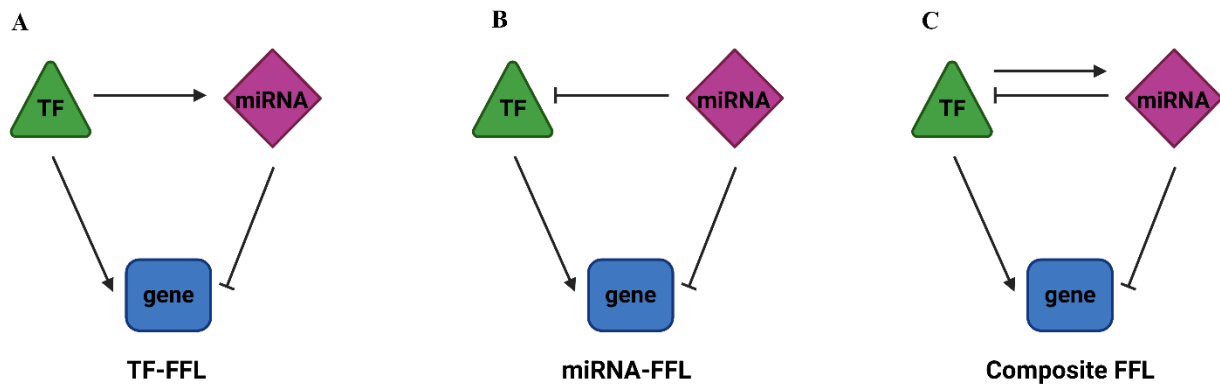


Figure 3.2. Feed forward loop examples. TF is depicted with a green triangle, miRNA with a purple diamond, target gene with a blue rectangle. (A) Feed forward loop centered on transcription factor. (B) Feed forward loop centered on miRNA. (C) Composite feed forward loop. (Image adapted from Zhang et al., 2015; Image created with BioRender).

3.2.1 Transcription factors and ChIP-Seq analysis

TFs are key elements in the regulation of gene expression, which bind to short DNA sequences called cis-regulatory elements (cREs), usually located in the promoter regions of genes or in distal regions such as enhancers. In particular, TFs can act as activators binding to specific DNA region present, for example, in the promoter of a gene, thus allowing the recruitment of RNA polymerase II, which is necessary for the initiation of transcription (Figure 3.3A). Some TFs can also interact with corepressors and thus inhibit transcription.^{15,16,17}

Each TF is characterized by its own binding motif, which is a specific sequence recognized on the DNA. These motifs are collected in databases such as JASPAR and ENCODE, where they are represented in the form of a logo sequence, characterized by a score indicating the probability of binding for each base.^{18,19,20}

TF dysregulation could be associated with tumorigenesis. In particular, they can contribute to pathological processes such as metastasis, uncontrolled cell proliferation, angiogenesis, genomic instability and metabolic alterations. For this reason, it is essential to study the mechanisms of gene regulation mediated by TFs, since their alteration may represent a primary cause in the onset and progression of different types of cancer. Understanding these mechanisms and the binding sites to which TFs bind may pave the way for the development of targeted drugs, with potential therapeutic applications.^{17,21,22}

One of the most used techniques to identify the binding sites of transcription factors in the genome is the Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq), which allows to map with high resolution the interactions between proteins and DNA within chromatin.^{14,16}

Furthermore, ChIP-Seq can also be used to identify modified histone regions, such as acetylated or methylated ones, which are indicators of actively transcribed or silenced chromatin.

ChIP-Seq is based on the combination of chromatin immunoprecipitation followed by high-throughput sequencing. The experimental procedure involves the crosslinking of the TF of interest to DNA, fragmentation of chromatin, and immunoprecipitation using a TF-specific antibody. The recovered DNA fragments are then purified and subjected to high-throughput sequencing. Bioinformatic analysis of a ChIP-Seq is composed of some key steps: the sequenced reads are aligned to the reference genome using software such as Bowtie, Bowtie2, or BWA, and the output is a SAM or BAM (compressed binary SAM) file, containing alignment information for each read. The next step is peak calling, which allows the identification of enriched peaks to the reference genome (Figure 3.3B).^{23,24}

The resulting peaks encompass different regions of the genome, which may include promoters of protein-coding genes and non-coding RNAs, such as miRNAs, enhancer regions, etc. To understand whether a miRNA or a protein-coding gene is positively or negatively regulated by a TF, it is interesting to see from a transcriptional point of view how their expression changes, through a miRNA-Seq and RNA-Seq analysis, respectively.

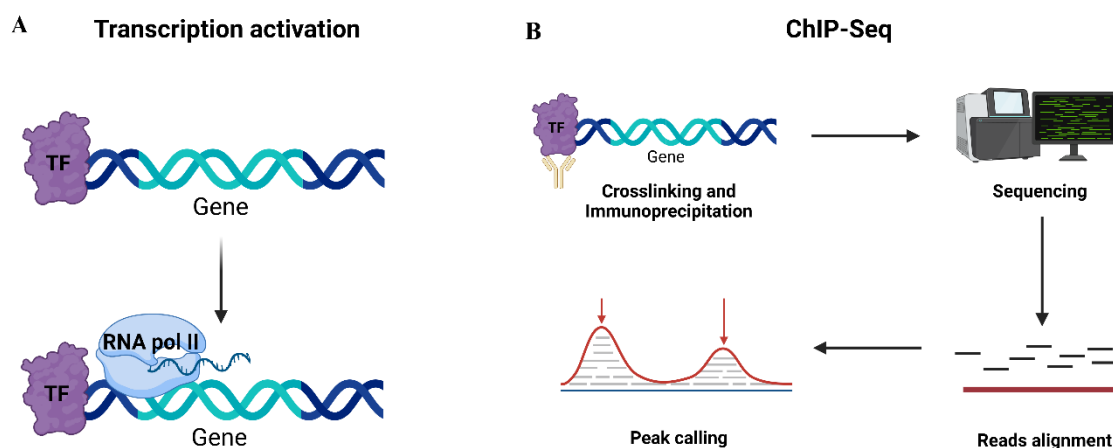


Figure 3.2. Transcriptional activation mechanism and ChIP-seq workflow. (A) Example of transcriptional activation mediated by transcription factor binding to the promoter of a gene, resulting in recruitment of RNA polymerase II and initiation of transcription, which leads to the synthesis of nascent mRNA. (B) Experimental phase of ChIP-seq, in which the transcription factor is stabilized on DNA by crosslinking with formaldehyde, followed by immunoprecipitation using an antibody specific for the factor itself. The bound DNA is then purified and sequenced. In the subsequent bioinformatics analysis phase, the reads are aligned to the reference genome and peak calling is performed to identify the binding sites of the transcription factor on a genome-wide scale. (Image created with BioRender).

3.2.2 Non-coding RNA: miRNAs and circRNAs

Protein-coding genes represent approximately 1.5% of the human genome and for a long time, research attention was focused almost exclusively on proteins, while large portions of non-coding DNA were considered "junk DNA". In recent years, however, the crucial role played by these non-coding regions, in particular non-coding RNAs (ncRNAs), in the regulation of gene expression and epigenetic mechanisms has become increasingly relevant.

NcRNAs are divided into two main categories: short non-coding RNA (such as miRNAs and piwi-interacting RNAs) and long non-coding RNAs (such as lncRNAs and circRNAs). These regulatory molecules are involved in various cellular processes, acting through different mechanisms, such as translation inhibition, modulation of transcription or interaction with other biomolecules.^{25,26,27}

Numerous studies have shown that ncRNAs also play a significant role in tumorigenesis: alterations in their expression or function have been associated with both the activation of oncogenes and the inhibition of tumor suppressor genes.²⁷

Short non-coding RNA, such as miRNAs, could be profiled with microRNA Sequencing (miRNA-Seq), in which, since these molecules are typically 18–24 nucleotides long, small-RNA library preparation protocols include size selection and specialized adapter ligation steps to ensure selective enrichment of short RNA species and accurate quantification of miRNA expression.

In this thesis work we focused on the analysis of miRNAs and circRNAs to investigate respectively their target genes and target miRNAs, using a miRNA-Seq and high sequencing depth RNA-Seq analysis. In the first project described in this thesis work we focused only on miRNAs, while in the second project we focused on both miRNAs and circRNAs.

miRNAs are short non-coding RNA of 19-25 nucleotides length, that play a pivotal role in the epigenetic regulation of gene expression. They bind to the 3'-UnTranslated Region (3'-UTR) of target mRNAs, thus inhibiting gene expression through mRNA destabilization or translational inhibition.¹² The 3' UTR of 60% of human protein-coding genes is characterized by miRNAs-binding sites, meaning that miRNAs are involved in several cellular pathways like differentiation, metabolism, proliferation, development, apoptotic cell death, viral infection, and oncogenesis.²⁸

The miRNAs sequences are scattered throughout the human genome and approximately 1,900 human miRNAs loci are annotated on miRBase: they can be in non-coding regions, in exonic or intronic regions of protein-coding genes, and quite often different miRNAs loci are close each other forming polycistronic transcription unit.^{29,30}

miRNA biogenesis (Figure 3.4) occurs mainly through a canonical pathway, which involves a complex multi-step process: the miRNA is first transcribed in the nucleus by the RNA polymerase II, giving rise to the primary miRNA (pri-miRNA), a non-coding RNA usually more than 1 kb long. Pri-miRNA is subjected to a first maturation process into the nucleus, operated by the Microprocessor complex (Drosha and DGCR8), forming the precursor miRNA (pre-miRNA) with a hairpin like structure. The pre-miRNA is exported by Exportin-5 (XPO5) inside the cytoplasm, where it is processed by the Dicer enzyme, an RNase III endonuclease. Dicer enzyme cleaves the pre-miRNA hairpin structure near its terminal loop, generating the double-stranded miRNA duplex composed of the mature miRNA strand and its complementary passenger strand. One strand of the duplex, known as the guide strand, is selectively incorporated into Argonaute (AGO) proteins within the RISC complex, while the passenger strand is typically degraded. The mature miRNA in RISC mediates gene silencing through mRNA cleavage or translational repression.^{12,28,29,30,31}

In addition to the canonical miRNA biogenesis pathway, non-canonical pathways have also been described^{30,31}, which increase the complexity of their processing and allow the generation of miRNAs independently of the Drosha enzyme, thus bypassing the action of the Microprocessor complex (Drosha-DGCR8). An example is represented by mirtrons, derived from debranched introns that, once removed by splicing, assume a secondary conformation similar to that of canonical pre-

miRNAs. These are recognized as pre-miRNAs and are exported via XPO5 into the cytoplasm, where they undergo processing by the Dicer enzyme, similarly to what occurs in the canonical pathway. There are also pre-miRNAs that are generated directly as primary transcripts by RNA polymerase II and that have a 5' cap. These precursors are directly exported into the cytoplasm via the Exportin-1 (XPO1) and processed by Dicer to generate mature miRNAs. Finally, other ncRNAs, such as some snoRNAs (small nucleolar RNAs) and tRNAs, can act as alternative miRNA precursors even in the absence of canonical Drosha-mediated processing.

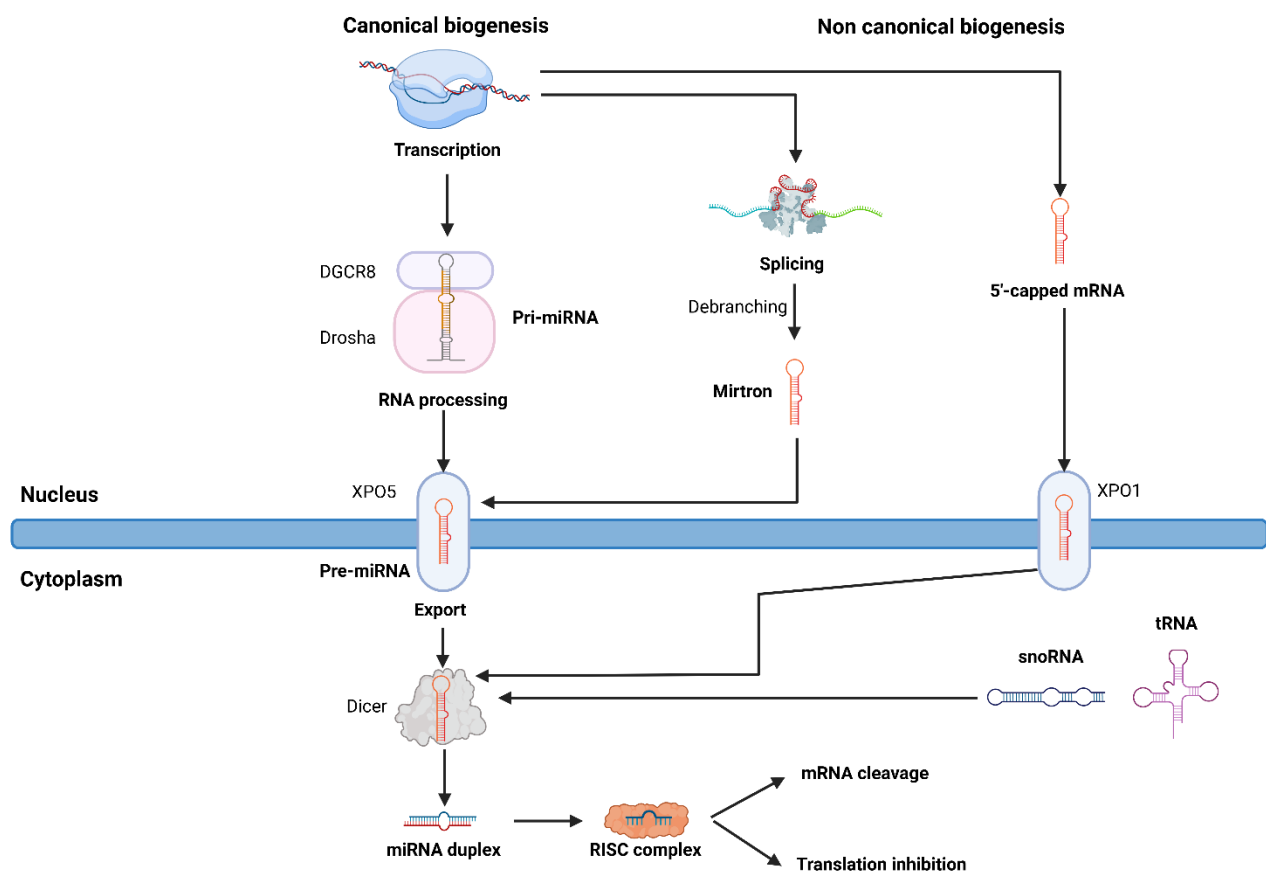


Figure 3.3. Canonical and non canonical miRNA biogenesis. In the canonical biogenesis pathway, the enzymes Drosha and Dicer act sequentially: Drosha processes the nuclear pri-miRNA into pre-miRNA, while Dicer, in the cytoplasm, converts the pre-miRNA into a miRNA duplex. This is then incorporated into the RISC complex, which guides the mature miRNA toward its targets, inducing mRNA cleavage or translation inhibition.

In the non-canonical pathways Drosha enzyme is bypassed: the non-canonical precursor is directly recognized by Dicer, which still generates a functional miRNA duplex. Also in this case, the output converges in the canonical pathway, with loading into the RISC and post-transcriptional regulation of target genes. (Image adapted from Kim et al., 2025; Image created with BioRender).

circRNAs are a particular class of long non-coding RNAs, characterized by a ring structure resulting from a process known as back-splicing. In this mechanism, the 5' end of an exon is joined to the 3'

end of the same or another upstream exon, resulting in the formation of a circular molecule. Unlike linear RNAs, circRNAs do not have a 5' capped end or a 3' polyadenylated tail.^{32,33,34,35,36} This circular conformation gives them high stability, making them less susceptible to degradation by cellular exosomes. For this reason, a progressive accumulation of circRNAs is observed with age, even if their synthesis occurs at a relatively slow rate.^{33,36}

Based on their biogenesis and composition, circRNAs can be divided into three main categories^{33,35,36,37} (Figure 3.5):

1. Exonic circRNAs (EcircRNAs): consisting exclusively of one or more exons (usually one to five). They represent over 80% of identifying circRNAs and are predominant in most human cells.
2. Circular intronic RNAs (ciRNAs): containing exclusively intronic sequences and deriving from introns that are not properly eliminated during the canonical splicing process.
3. Exon-intron circRNAs (ElciRNAs): comprising both exonic and intronic sequences.

circRNAs perform a wide range of cellular functions, mainly acting as “sponge” for miRNAs. By binding complementarily to miRNAs, circRNAs inhibit their activity, thus preventing the downregulation of target genes mediated by miRNAs themselves. In addition to this function, recent studies have shown that circRNAs can also bind to proteins, acting as a “sponge protein”, and alter gene expression by regulating the activity of protein-coding genes. These interactions place circRNA at the center of the epigenetic regulation process.^{33,34,35,36,37}

Given their ability to interact with miRNAs and regulatory proteins, circRNAs are involved in numerous pathological processes, particularly in tumors.^{36,37} For example, circRNAs that bind tumor suppressor miRNAs can activate genes that promote tumor cell growth and survival, while those that sequester oncogenic miRNAs can have protective effects, restoring genes that control apoptosis and cell division. Furthermore, circRNAs can interact with regulatory proteins, modifying gene activity and signaling pathways often altered in cancer. Numerous scientific evidence documents a key role in several forms of tumors, including breast cancer, cervical cancer, colorectal cancer, gastric cancer, hepatocellular carcinoma, lung cancer, osteosarcoma, ovarian cancer, renal cell carcinoma and skin cancer.³⁵

In this thesis, we focused on glioblastoma (second project), analyzing the role of circRNAs as miRNA sponges, in order to identify which miRNAs are sequestered and inhibited by specific circRNAs potentially involved in the progression of this neoplasm.

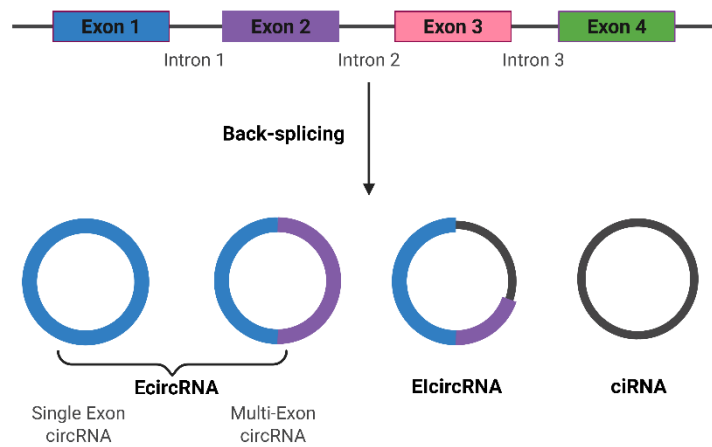


Figure 3.4. circRNA biogenesis and classification. circRNAs originate from a back-splicing process and are divided into three main categories, illustrated in the figure. EcircRNAs are made up exclusively of exons: in the first example on the left, a single exon circRNA is shown, consisting of only exon 1 (colored in blue), while immediately next to it is shown a multi-exon circRNA, which includes exons 1 and 2. Proceeding from left to right, there is an example of ElcircRNA, containing both exons (exon 1 and exon 2) and an intron, indicating a mixed structure. Finally, the ciRNA, consisting exclusively of introns, is represented in the image as a circular structure with a thinner gray line. (Image adapted from Pisignano et al., 2023. Image created with BioRender).

3.2.3 Protein coding genes and transcriptomic analysis by mRNA-Seq

After analyzing the main regulators of gene expression, such as TFs through ChIP-Seq and miRNAs through miRNA-Seq, it is essential to also consider protein-coding genes, which represent the final targets of these regulatory pathways. The analysis of the expression of these genes through mRNA Sequencing (mRNA-Seq) allows the evaluation of the overall functional effect of transcriptional and post-transcriptional regulatory mechanisms, providing an integrated view of gene control in different biological conditions.

mRNA-Seq is a technique based on Next Generation Sequencing (NGS) that allows both qualitative and quantitative measurement of the transcriptome, thus defining the set of messenger RNAs (mRNA) present in a cell at a given time. In this way, it is possible to determine whether a gene is actually expressed and at what level, in a specific biological context.

This method has revolutionized the study of the transcriptome, proving particularly useful in the oncology field. In fact, it allows the identification of genes differentially expressed between cells in normal conditions (wild type) and tumor cells. This has opened new perspectives in the study of cancer, contributing to the development of precision medicine. Thanks to the analysis of the transcriptome, it has been possible to distinguish different tumor subtypes and identify specific therapeutic targets, favoring the development of targeted drugs.³⁸

3.3 Regulatory network in endothelium

Endothelial cells (ECs) cover the inner surface of blood vessels and play a crucial role in numerous biological processes, including vascular development, inflammation, coagulation, and angiogenesis. A widely used model for studying endothelial function is represented by HUVEC (Human Umbilical Vein Endothelial Cells), primary cells isolated from the human umbilical vein. These cells retain the functional characteristics of endothelial cells and are therefore used in research for studying various pathologies, including cardiovascular disease, inflammatory processes, cancer, metabolic and neurological disorders.^{39,40}

In the first project of this thesis, we analyzed the regulatory networks in a HUVEC model with the aim of understanding the role of the transcription factor TFEB in the regulation of protein-coding genes and miRNAs.⁴¹

3.3.1 TFEB

Transcription factor EB (TFEB) is a transcription factor belonging to the Microphthalmia/Transcription Factor E (MIT/TFE) family, which also includes three other factors: Microphthalmia-associated Transcription Factor (MITF), Transcription Factor EC (TFEC), and Transcription Factor E3 (TFE3). All members of this family are characterized by the presence of two conserved domains: the basic helix–loop–helix (bHLH) domain and the leucine zipper (Zip) domain, which derive from a common ancestor that emerged through gene duplication events.^{42,43,44}

Specifically, TFEB consists of 476 amino acids and shares with other family members the bHLH and Zip domains, which are essential for the formation of homodimers and heterodimers with MITF, TFE3, or TFEC. It also shares with MITF and TFEC the activating domain (AD), which is essential for the regulation of gene expression. TFEB also exhibits a glutamine-rich domain, which contributes to the recruitment of transcription coactivators, and a proline-rich domain, which confers structural rigidity to the protein and may facilitate specific interactions with other regulatory factors.^{43,44} In Figure 3.6A is depicted TFEB structure.

TFEB is normally found in the cytoplasm of cells in an inactive state and is rapidly degraded. However, following post-translational modifications, such as dephosphorylation and acetylation, it translocates to the nucleus where it binds to the promoter regions of target genes containing Coordinated Lysosomal Expression and Regulation (CLEAR) elements, thereby activating the

transcription of these genes (Figure 3.6B).^{43,44,45,46} The genes targeted by TFEB are associated with various biological processes such as lysosomal catabolic activity, autophagy, exo- and endocytosis, and phagocytosis, thereby influencing cellular degradative processes and intracellular clearance.^{43,47} Beyond its role in degradative pathways, TFEB also influences cell division and motility, metabolism, aging, immune responses, and vascular development, and it is involved in various sets of diseases, such as cancer, lysosomal storage diseases (LSDs), neurodegenerative diseases, metabolic disorders, and cardiovascular diseases.^{44,48,49,50,51,52,53}

Mutations and alterations in TFEB expression were initially identified in renal cell carcinoma, pancreatic cancer, and melanoma, where they are associated with abnormal regulation of the cell cycle, metabolism, and tumor microenvironment. Subsequent evidence has also highlighted TFEB involvement in other malignancies, such as colorectal cancer, gastric cancer, non-small cell lung cancer (NSCLC), and breast cancer. Furthermore, TFEB may contribute to tumor angiogenesis, promoting the formation of new blood vessels necessary for the supply of oxygen and nutrients to the tumor, thus promoting tumor growth and metastatic spread.^{54,55}

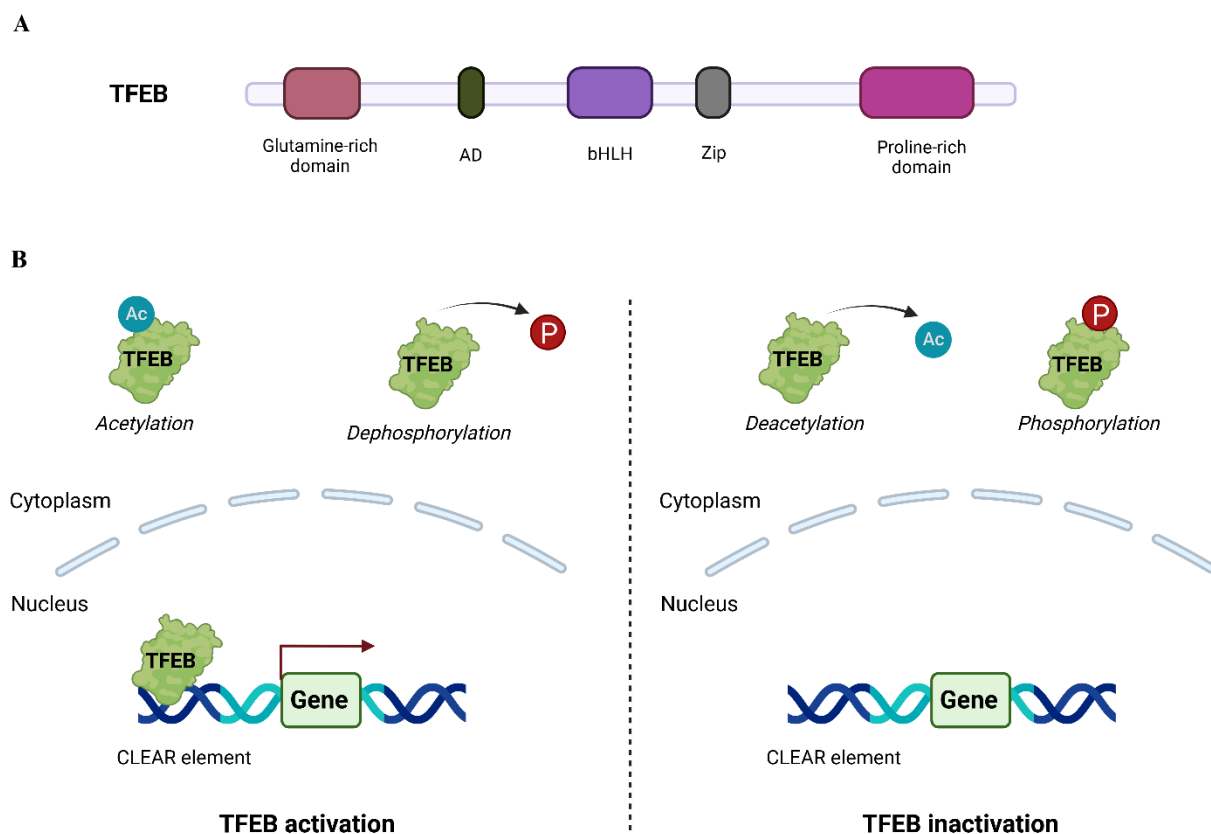


Figure 3.6. TFEB structure and mechanism of regulation. (A) Schematic representation of TFEB structure characterized by glutamine-rich domain, AD, bHLH, Zip and proline-rich domain. (B) TFEB regulatory mechanism. When acetylated or dephosphorylated, TFEB translocates from the cytoplasm to the nucleus, where it binds to CLEAR elements and activates the transcription of target genes. Conversely, when deacetylated or phosphorylated, nuclear translocation is inhibited, and consequently, TFEB is unable to activate gene expression. (Image created with BioRender).

3.3.2 TFEB in endothelium

TFEB plays a key role in regulating endothelial cells, influencing crucial processes such as angiogenesis, inflammatory response, and the progression of atherosclerosis. These functions make it a promising target for the development of new drugs in future vascular therapies.⁵⁶

Studies conducted on embryos and neonatal mice suggest that TFEB plays a key role in vascular endothelial development. Specifically, TFEB has been shown to directly modulate the promoter region of the cyclin-dependent kinase 4 (CDK4) gene, which is involved in cell cycle progression, and regulate the expression of vascular endothelial growth factor receptor 2 (VEGFR2) through the control of the miR-15a/16-1 cluster, highlighting a possible mechanism of miRNA-mediated gene regulation.⁵⁰

To further explore TFEB's involvement in miRNA regulation in endothelial cells, in this thesis we used the HUVEC cell model, a widely recognized model for studying endothelial function. The results obtained show that TFEB is indeed capable of regulating the expression of specific mRNAs and miRNAs in HUVEC.⁴¹

3.4 Regulatory network in Glioblastoma

Glioblastoma multiforme (GBM) is a primary aggressive malignant tumor of the central nervous system, originating from neuroglial cells and with the ability of invading the brain parenchyma.^{57,58,59,60} The most commonly affected sites are the temporal lobe (33% of cases) and the frontal lobe (32% of cases) of the brain.⁶⁰

According to the latest World Health Organization (WHO) classification, GBM is considered a grade IV glioma and includes only tumors with wild-type IDH status, meaning that the IDH1 and IDH2 genes are not mutated. In contrast, gliomas with IDH mutations, typically harboring a single (heterozygous) point mutation, are now classified as astrocytomas, regardless of histological characteristics.⁶¹

Although a rare tumor, GBM is the most common and most aggressive form of malignant tumor of the central nervous system, with a higher incidence in men than women (4.07 cases per 100,000 population),⁶² and diagnosis occur predominantly in individuals over the age of 65.^{60,63}

Despite aggressive treatment, which includes surgical resection followed by radiotherapy and adjuvant chemotherapy with temozolomide, median survival remains extremely low, ranging between 12 and 15 months. For this reason, understanding the molecular mechanisms and pathways involved in the pathogenesis of GBM is crucial for the development of more targeted and effective therapies, with the aim of improving prognosis and patient survival.^{58,59,60,63}

Large-scale multi-omic studies, such as those performed by The Cancer Genome Atlas (TCGA) project and the analysis conducted by Wang et al., 2021, have highlighted profound heterogeneity in GBM at the genomic, epigenetic, transcriptomic, and proteomic levels.⁶⁴ Specifically, genomic analyses have identified three main molecular pathways frequently altered in glioblastoma: RTK, p53, and RB. Alterations in the RTK (Receptor Tyrosine Kinase) pathway promote cell survival and proliferation; mutations in the TP53 gene impair apoptosis mechanisms, promoting cell invasion and migration; dysfunction of the RB pathway leads to deregulation of the cell cycle, contributing to the uncontrolled proliferation of tumor cells (Figure 3.7). In summary, the coordinated disruption of these three fundamental pathways underlies the tumorigenic process of GBM.⁵⁹

Another level of complexity is related to transcriptomics, which has allowed GBM to be classified into three distinct molecular subtypes: proneural, mesenchymal, and classical.

The proneural subtype is characterized by a molecular profile like that of oligodendrocytes, with the expression of specific genes such as DLL3, PDGFRA, OLIG2, SOX, DCX, ASCL1, and TCF. This subtype is more common in young patients and shows a poor response to conventional therapies. After

relapse, it may undergo a transformation to the mesenchymal subtype, following downregulation of PTEN and knockdown of NF1.

The mesenchymal subtype presents markers typical of astrocytes and angiogenesis, with mutations in the NF1 gene and alterations in the AKT and NF- κ B pathways. It is characterized by a greater infiltration of immune cells, such as M1 and M2 macrophages and glial cells, which is the reason why patients belonging to this subtype respond better to immunotherapy.

Both the proneural and mesenchymal subtypes possess stem cell characteristics, due to the presence of Tumor Initiating Cells (TICs), which are responsible for treatment resistance and relapse in patients.

The classical subtype, unlike the proneural and mesenchymal subtypes, does not present stem cell characteristics and is instead defined by amplification of the EGFR gene and deletion of CDKN2A.⁶⁵

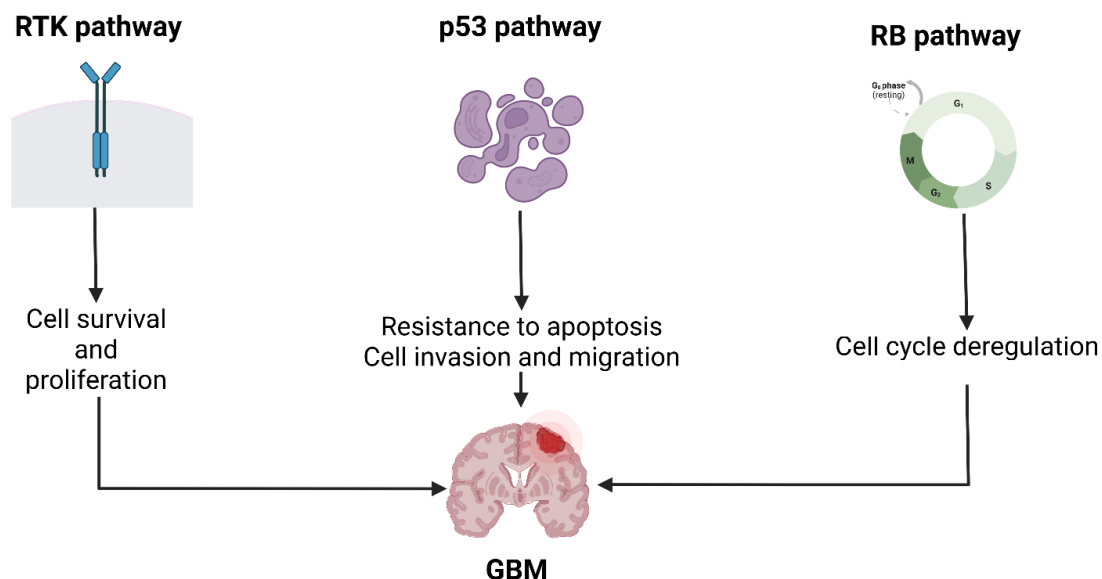


Figure 3.7. Deregulated pathways in GBM. Schematic representation of the three main pathways altered in GBM. (Image created with BioRender).

3.4.1 Glioblastoma tumor initiating cells

Cancer Stem Cells (CSCs) or Tumor Initiating cells (TICs) have been identified in numerous tumor types and exhibit stem-like and self-regenerative properties, allowing the regeneration of tumor mass even after aggressive chemotherapy. Two main theoretical models explain their origin: the stochastic model and the hierarchical model. According to the stochastic model, differentiated tumor cells, following the accumulation of genetic mutations, can undergo a process of

dedifferentiation, acquiring stem-like characteristics; whereas the hierarchical model hypothesizes that accumulation of genetic mutations in normal stem cells cause the transformation into cancer stem cells, giving rise to a subpopulation resistant to therapies. These cells are therefore responsible for the recurrences observed in cancer patients, making their study crucial for the development of drugs aimed at their eradication.^{66,67}

Cell subpopulations with these characteristics, named Glioblastoma Tumor Initiating Cells (GBM TICs), have also been identified in GBM. These subpopulations are thought to be responsible for frequent relapses despite the use of aggressive therapies and are also responsible for reduced patient survival. Indeed, in murine xenografts, a high number of TICs correlates with reduced survival.⁶⁸

The main signaling pathways involved in the survival and proliferation of GBM TICs, and underlying the high heterogeneity of this tumor, are Wnt/ β -catenin, PI3K/AKT, and Notch.^{69,70}

Surface biomarkers commonly associated with these cells include CD133, CD44, CD15, and CD36, while transcription factors include SOX2, OCT-4, NANOG, and PAX6. These elements underscore the crucial role of epigenetic mechanisms in the maintenance of TICs.⁷⁰ Recently, LSD1 (also known as KDM1A) emerged as an important epigenetic regulator responsible for the maintenance and survival of GBM TICs.⁷¹

3.4.2 LSD1

Histone lysine demethylases (KDMs) were discovered about twenty years ago and comprise eight families, divided into two main classes: FAD-dependent amine oxidases and JmjC domain-containing enzymes.

The first class includes only the KDM1 family, consisting of the LSD1 (KDM1A) and LSD2 (KDM1B) epigenetic factors, which require FAD as a cofactor to carry out the demethylation reaction. All other families belong to the Jumonji C domain (JmjC) enzymes class, which depend on Iron (II) and α -ketoglutarate for their catalytic activity.^{72,73}

In the second project of this thesis work, we focused specifically on the epigenetic factor LSD1, which plays a key role in the maintenance and survival of GBM TICs, that are responsible for the frequent relapses observed in GBM patients.

LSD1 is a protein composed of 852 amino acids and structurally organized into three functional domains:

- the SWIRM domain, which is located at the end of the N-terminal and mediates LSD1's interaction with other proteins;
- the AOL domain, responsible for the catalytic activity;
- the Tower domain, which interrupts the AOL domain and allows interaction with LSD1 cofactors (Figure 3.8A).^{74,75}

LSD1 can form complexes with various mediators including protein complexes, receptors, long noncoding RNA, microRNAs, non-histone substrates, and transcription factors. Depending on the context and the molecular partners involved, when LSD1 demethylates histone proteins can act as either an activator or a repressor of target gene transcription (Figure 3.8B).

In addition to histones, LSD1 can also demethylate several non-histone proteins, such as p53, MYPT1, DNMT1, STAT3, E2F1, MEFD2, MTA1, ER α , HSP90, HIF-1 α , and AGO2.^{75,76}

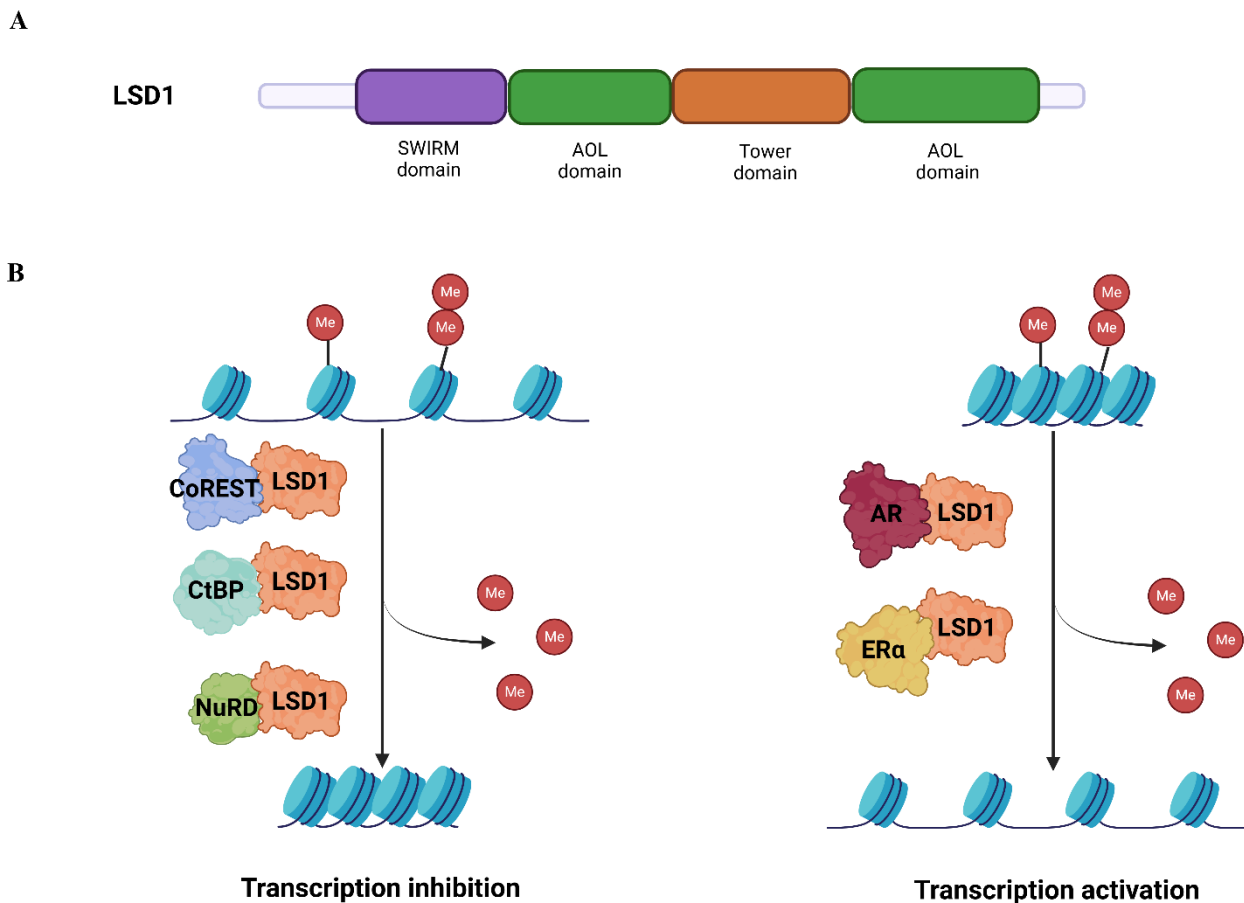


Figure 3.8. LSD1 structure and mechanism of regulation. (A) Schematic representation of LSD1 structure with the SWIRM domain, AOL domain separated by Tower domain. (B) Based on the partner LSD1 can activate or repress gene transcription. When LSD1 interacts with CoREST, CtBP and NuRD proteins demethylates H3K4me1/2 repressing gene transcription, whereas if it collaborates with the androgen or estrogen nuclear receptors demethylates H3K9me1/2 activating gene transcription. (Image adapted from Gu et al., 2020; Image created with BioRender).

This wide variety of substrates translates into remarkable functional heterogeneity, linking LSD1 to processes such as stemness maintenance, differentiation, cell motility, epithelial-mesenchymal transition, autophagy, senescence, neurodegenerative diseases, and metabolism.⁷⁶ Numerous studies have also demonstrated its involvement in various tumor types. In particular, in GBM it is essential for GBM TIC survival, making it a highly attractive therapeutic target: its eradication could reduce relapses and increase patient survival.

Indeed, both genetic and pharmacological inhibition of LSD1 with the oral LSD1 inhibitor DDP_38003, are associated with a reduction in the stemness and self-renewal capacity of GBM TICs. In PDX models of GBM, LSD1 inhibition reduces tumor growth and blocks stemness, exerting significant antitumor effects.⁷¹ LSD1 inhibitors therefore represent a promising strategy for the treatment of GBM, as demonstrated by several ongoing studies.^{77,78}

A global transcriptomic analysis of LSD1-silenced GBM TICs conducted by Faletti et al., 2021, which are an example of proneural subtype, highlighted the deregulation of numerous genes involved in the integrated stress response (ISR), crucial for cell survival under stressful conditions. These include ASNS, CHAC1, DDIT3/CHOP, GFG15, NARS, PSAT1, SLC7A11, TRIB3, and XPOT, also known as targets of the transcription factor ATF4. Inhibition of LSD1 also reduces ATF4 activation, confirming ATF4 role as an LSD1 cofactor.⁷¹

The work conducted by Faletti et al., 2021 revealed that LSD1 plays a key role in the maintenance of GBM TICs, and the protein-coding genes regulated by LSD1 have also been identified.⁷¹ Based on these findings, in the second project of this thesis, we focused on analyzing LSD1 ChIP-Seq data to identify not only the protein coding genes regulated by the enzyme, but also potential miRNAs involved in regulatory networks.

This observation led to the hypothesis that LSD1 may act not only on coding genes but also on non-coding RNAs, particularly miRNAs and circRNAs. In the absence of miRNA-Seq and circRNA-Seq data for the GBM TICs model, we therefore extended the analysis to the proneural subtype of GBM patients, with the aim of identifying specific ncRNAs potentially regulated by LSD1. We finally built a network of interactions to better understand their functional correlations, integrating the data with public resources obtained from Wang et al., 2021⁶⁴ to add a broader and clinically relevant perspective.

4. AIM OF THE STUDY

Understanding complex biological diseases, such as cancer, is particularly challenging due to the multitude of molecular factors involved and the need to integrate information across different levels of omics investigation (genomics, transcriptomics, proteomics, metabolomics, lipidomic, and epigenomics). The analysis of each dataset provides only a partial view, whereas their integration represents a crucial step to more precisely identify the regulatory mechanisms and molecules implicated in the disease. In this context, biological networks emerge as an effective tool both for integrating heterogeneous data and for intuitively visualizing the underlying molecular interactions.

The primary objective of this thesis is to develop a reproducible and flexible bioinformatics pipeline capable of integrating data from different omics levels to construct regulatory networks mediated by transcription factors and ncRNAs, applying it in two distinct research projects.

The first project involved the study of the epigenetic factor TFEB in HUVEC cells, a widely used endothelial model for investigating key biological processes such as angiogenesis, lysosomal biogenesis, and cancer. The aim of this project was to investigate the role of TFEB in HUVEC, by integrating ChIP-Seq, miRNA-Seq, and mRNA-Seq data to construct a regulatory network linking TFEB, the miRNAs it regulates, and their target genes expressed. This approach provided a representation of the potential TFEB-mediated regulation of miRNAs, a still underexplored aspect, particularly in endothelial systems.

The second project, developed in collaboration with Professor Giuliana Pelicci's group at the IEO, focused on glioblastoma (GBM). Based on a previous work by Faletti et al., 2021, it emerged that LSD1 plays a key role in the maintenance of GBM TICs, tumor cells with stem-like characteristics and involved in the continuous relapses of patients, which derive from a patient with the proneural subtype. Furthermore, it was possible to identify the protein-coding genes regulated by LSD1. Based on these results, the primary objective of this second project was to identify not only the protein coding genes regulated by LSD1 in cooperation with ATF4, but also a candidate miRNA potentially involved in the regulatory network and its target genes, through mRNA-Seq and LSD1 ChIP-Seq analysis performed in GBM TICs.

Starting from this observation, our hypothesis is that LSD1 may regulate not only protein-coding genes, but also ncRNAs, particularly miRNAs and circRNAs. However, in the absence of miRNA-Seq and circRNA-Seq data in the same cell model used by Faletti et al., 2021, we extended the analysis to public data from GBM patients obtained from a study conducted by Wang et al., 2021, focusing

only on the proneural subtype. The goal was to identify specific ncRNAs (miRNAs and circRNAs) potentially regulated by LSD1. To this end, we constructed a network of interactions to better understand their functional correlations and include a broader and more clinically relevant perspective. We also assessed the presence of LSD1 around the promoter region of these ncRNAs.

The common thread of both projects is therefore the construction and application of a bioinformatics pipeline aimed at multi-omics integration for the identification of regulatory networks mediated by transcription factors. This approach not only provides versatile and reproducible tools for different biological contexts but also contributes to elucidating the molecular mechanisms underlying complex diseases such as cancer, highlighting the role of transcription factors and ncRNAs as central regulatory nodes.

5. MATERIALS AND METHODS

5.1 Computational environment: Linux and programming languages

This thesis project was conducted using a bioinformatics approach. Specifically, it was developed in a computing environment based on the Ubuntu Linux operating system, widely used in computational biology for the development and execution of bioinformatics software. Linux, an open-source distribution derived from Unix, provides a flexible and tool-rich platform for processing complex data and allows the integration of various programming languages through the shell.

This project primarily employed Perl, Python, and R, each chosen for its unique characteristics in data manipulation and analysis.

Perl, particularly effective at handling strings, was used for the manipulation and parsing of ASCII files.

Python, using widely used scientific libraries such as NumPy and pandas, was used for processing text files and managing data in matrix format.

Finally, R was used as a statistical and graphical platform for data analysis, particularly for performing statistical tests and creating graphical representations such as heatmaps and volcano plots, which are essential for the biological interpretation of the results.

5.2 Promoter definition of genes and miRNAs

One of the goals of both the projects of this thesis was to identify the genes and miRNAs characterized by the presence of transcription factors in proximity to their promoter region. Indeed, the presence of binding sites or peaks in these regions increases the probability that a specific transcription factor can effectively regulate the expression of that gene or miRNA.

For this reason, it was first essential to define the promoter regions of genes and miRNAs. In this thesis, we used the human genome assembly hg38. Specifically, promoter regions were defined as the interval from -2,500 to +2,500 nucleotides relative to the Transcription Start Site (TSS) (Figure 5.1).

The human gene annotation for the hg38 assembly was obtained from Ensembl BioMart;⁷⁹ subsequently, using a Perl script, we defined the region between -2,500 and +2,500 nucleotides around the TSS of each gene.

For miRNAs, the annotation was obtained from the work of De Rie et al., 2017,⁸⁰ which provides a list of pri-miRNA promoters aligned to the hg19 reference genome. Since our work is based on hg38

assembly, these coordinates were converted using the LiftOver tool available on the UCSC Genome Browser.⁸¹ Then, a Perl script was used to define the promoter regions between -2,500 and +2,500 nucleotides around the TSS of each miRNA.

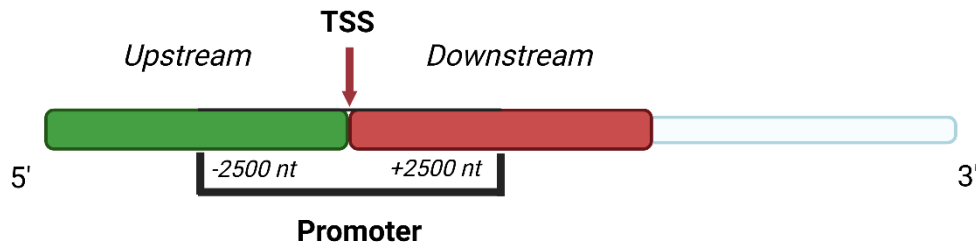


Figure 5.1. Schematic representation of the definition of the promoter regions for genes and miRNAs. For both genes and miRNAs, promoter regions were defined as the genomic interval spanning from -2,500 to +2,500 nucleotides relative to their TSS. (Image created with BioRender).

5.3 Databases and biological resources

5.3.1 JASPAR

JASPAR is an online open-access database of transcription factor motifs, represented as Position Frequency Matrices (PFMs), which describe the probability of occurrence of each nucleotide at a given binding site. For each transcription factor, it is also possible to obtain the genomic coordinates of the motif's occurrence sites across the entire reference genome.¹⁸

In the first project related to the regulation of endothelial networks JASPAR was used to identify TFEB motifs at the level of the promoter regions of miRNAs, using the MA0692.1 matrix profile.

In the second project, focused on glioblastoma, motifs from ATF4, a known cofactor of LSD1, were retrieved from JASPAR. Since ChIP-seq data for ATF4 were not available for our cell model, the MA0833.3 matrix from the JASPAR database was used.

5.3.2 miRBase

MiRBase is an open-access database that collects information on miRNAs, including sequences, genomic coordinates, bibliographic references, precursor hairpin structures, and the nomenclature of mature miRNAs. In the naming system, precursors are referred to as "mir," while mature forms are referred to as "miR".⁸²

The latest available release (2018, version 22) includes data from 271 species and, for the human genome, contains annotations for 1,917 precursors and 2,654 mature sequences.

In both projects analyzed, the MiRBase v.22 database was used to link human pre-miRNAs to their respective mature miRNAs.

5.3.3 TargetScan and miRTarBase

TargetScan and miRTarBase are two widely used open-access databases for studying interactions between miRNAs and their target genes.

TargetScan predicts potential miRNA–mRNA binding sites based on the complementarity of the seed region.⁸³ In this thesis TargetScan version 8.0 was used to predict target genes associated with the identified miRNAs in both the first and second projects.

miRTarBase is a database that collects information on interactions between miRNAs and target genes, experimentally validated through various techniques, including reporter assays, Western blots, microarray analyses, and next-generation sequencing experiments.⁸⁴ Some interactions in the database are classified as "weak"; in our study, only those with robust validity were considered, thus excluding interactions marked as "weak." In this work, miRTarBase version 9.0 was used only in the first phase of the second project, with the aim of identifying target genes for the candidate miR-124-3p. Only target genes shared between the TargetScan predictions, and the experimental validations reported in miRTarBase were selected.

5.3.4 CSCD2

CSCD2 is a free database that collects interactions identified in both tumor samples and normal tissues between circRNAs and miRNAs, or between circRNAs and RNA-binding proteins (RBPs). This tool thus allows exploration of circRNAs involved in tumor processes. Specifically, it integrates existing data sources for its predictions, but the reported interactions are computationally generated based on known binding motifs and binding sites.⁸⁵

In the second project of this thesis, CSCD2 was used to predict interactions between the miRNAs and the circRNAs identified in the study by Wang et al., 2017, considering only the interactions predicted by both TargetScan and miRanda.

5.3.5 UCSC Genome Browser

UCSC Genome Browser is one of the most widely used bioinformatics tools, developed at the University of California Santa Cruz. It is a freely accessible platform that enables graphical and interactive genome visualization, along with a wide range of complementary biological information. UCSC represents a complex set of tools and databases accessible through a dedicated web portal. These resources provide a vast amount of sequence data, with a particular focus on vertebrate genomes, especially the human genome.

The Genome Browser is divided into two fundamental components: raw genomic sequences and functional and molecular annotations, imported from external sources and represented as tracks. Each track is a graphical representation of specific biological information aligned to the reference genome. Example of track types include gene annotations, indicating the structure of exons and introns; alignments of transcripts (mRNAs) and DNA repetitive elements; gene expression data, such as from RNA-seq or microarray experiments; epigenomic information, such as DNA methylation profiles and histone modifications (acetylation, methylation, etc.).⁸¹

In the context of this thesis, it was used to visualize and analyze the promoter regions of selected miRNAs.

5.3.6 ENCODE

The ENCODE (Encyclopedia of DNA Elements) project was launched in 2007, with the aim of systematically identifying and cataloging all functional elements present in the human genome, using various sequencing techniques.

The analyses conducted within this project have provided insights into the functional aspects of transcription, particularly in relation to chromatin organization, histone modifications, and promoter regulation mechanisms.

The ENCODE database collects and makes available numerous experimental data, including those derived from transcription factor ChIP-Seq, DNase-Seq, ATAC-Seq, mRNA-Seq, and many other types of genomic experiments.^{19,20}

In the first project, H3K27ac and H3K4me3 peaks relative to the human reference genome hg38 in HUVEC cells were retrieved from the ENCODE portal, using the identifiers ENCFF426QFH for H3K27ac and ENCFF992YLK for H3K4me3, respectively.

In the second project, information on candidate regulatory elements (cCREs) was obtained from the ENCODE database and explored using the UCSC Genome Browser, with the aim of identifying specific activated elements.

5.4 Bioinformatics tools

5.4.1 BEDTools

BEDtools is an open-source software developed for Linux systems that brings together several useful tools for manipulating and analyzing genomic data. Among these, the *bedtools intersect* command allows to compare two *.bed* files to identify shared genomic intervals (Figure 5.2).⁸⁶

In this thesis, *bedtools intersect* command was used in both projects to intersect LSD1 and TFEB peaks, obtained from ChIP-Seq experiments, with the genomic coordinates of genes and miRNAs referred to the human hg38 genome. The analysis was conducted using the *-wao* option, and only peaks with an overlap of more than 100 nucleotides were considered for further analysis.

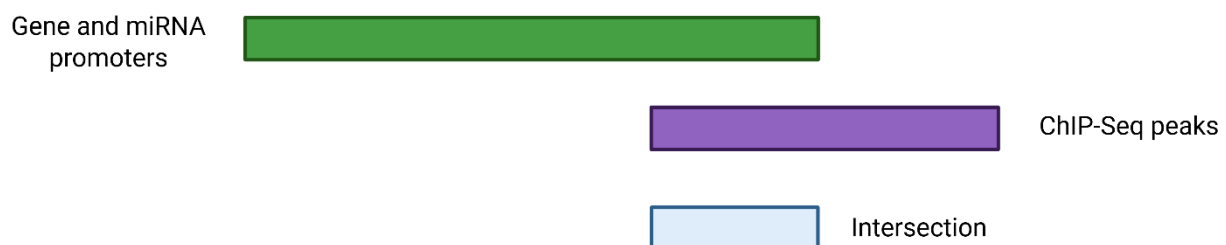


Figure 5.2. Example of overlapping genomic intervals identified by *bedtools intersect*. (Image created with BioRender)

5.4.2 IPA

Ingenuity Pathway Analysis (IPA) is a software developed to interpret high-throughput gene-expression data, such as RNA-Seq data, by integrating them with existing biological knowledge. Its primary goal is to identify the possible upstream causes and potential downstream effects that changes in expression may have on cell biology and the organism.

Among the various available tools, IPA also allows for the identification of upstream regulatory molecules of the genes included in the dataset, which could explain the observed expression

variations. Furthermore, the software identifies upstream regulators by relating them to the genes in the dataset through direct or indirect connections.⁸⁷

In this thesis, IPA was used in the second project. Specifically, the differentially expressed genes between silenced LSD1 and control samples resulting from an RNA-Seq experiment were used as input with the aim of identifying a candidate miRNA as a potential upstream regulator.

5.4.3 ToppGene Suite

ToppGene is a web-based software used for functional analysis giving as input a set of genes.⁸⁸ It was used in the first project to perform a functional analysis giving as input the list of genes differentially expressed among silenced TFEB and control samples. We focused on selected gene ontology (GO) biological process terms, which were ranked based on their statistical significance, determined through p-values corrected using the Benjamini–Hochberg method.

5.4.4 Cytoscape and NetworkX

Cytoscape, a software, and NetworkX, a python package, are commonly used for graphical reconstruction of molecular interaction networks, such as protein-protein, protein-DNA, miRNAs-target genes interactions. The molecules are indicated as nodes, while the interactions between them as edges.^{89,90}

In this thesis, they were specifically used to graphically reconstruct the regulatory networks mediated by the transcription factors and miRNAs.

5.5 Overview of the methods used for the construction of the regulatory network in endothelium – first project

In Figure 5.3 is presented a schematic representation of an overview of the methods used in the first project related to the construction of the regulatory network mediated by TFEB in HUVEC cell model.

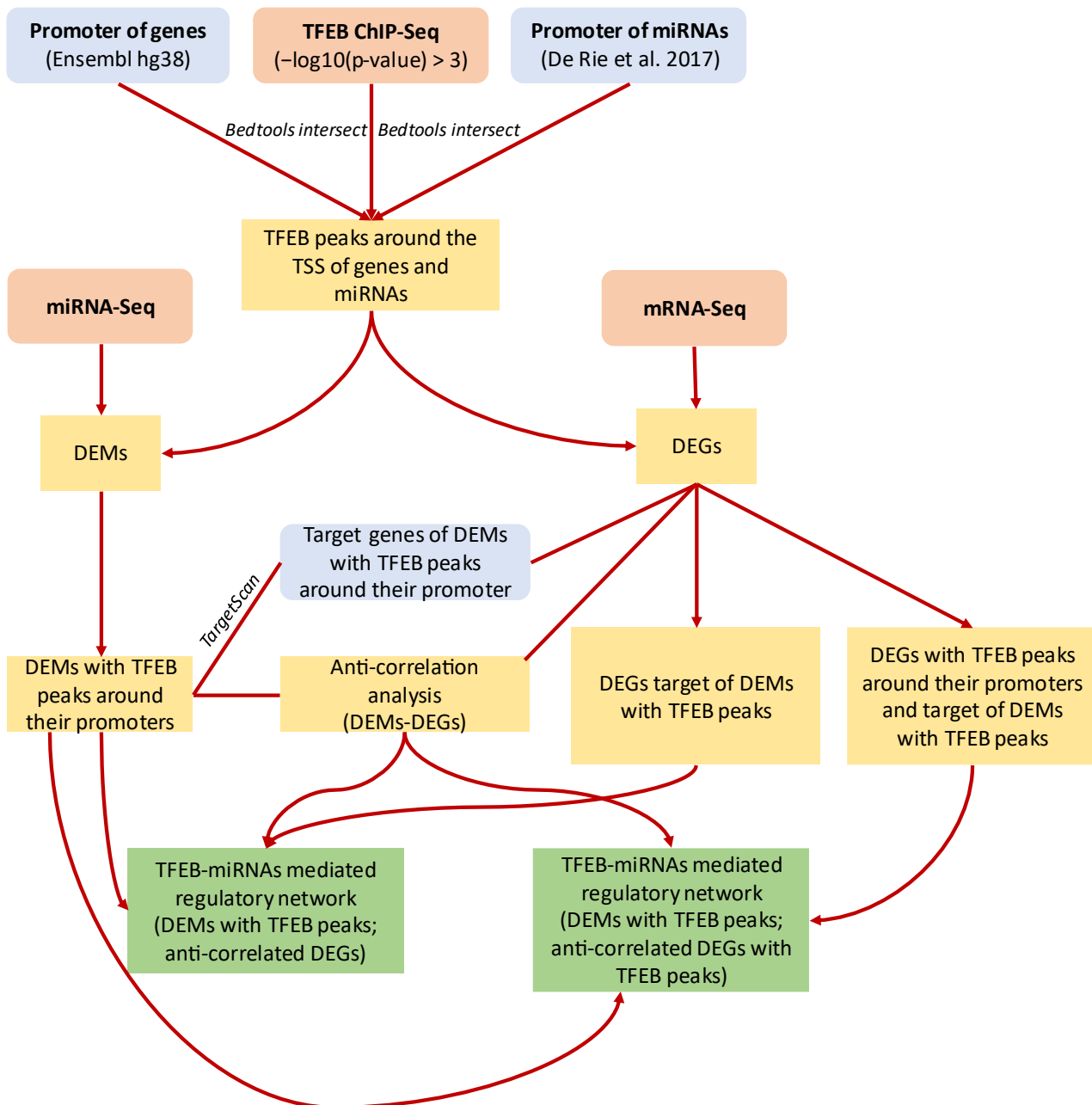


Figure 5.3. Flowchart of the methods used in the first project. The blue rectangles represent the gene and miRNA promoters, along with the list of DEM target genes. The orange rectangles refer to the omics data used, the yellow ones to the intermediate results, while the green rectangles correspond to the final output. Specifically, the gene and miRNA promoters were defined considering the regions between -2,500 and +2,500 bp with respect to the respective TSS and subsequently intersected with TFEB binding peaks, obtained by ChIP-Seq in HUVEC cells, using bedtools intersect. A differential expression analysis was conducted from the miRNA-Seq and mRNA-Seq data, allowing the identification of DEMs and DEGs. Only DEMs associated with the presence of TFEB in their promoter were selected, and their target genes were identified using the TargetScan database. Subsequently, an anti-correlation analysis was performed between the TFEB-associated DEMs and the DEGs. This procedure resulted in the construction of two regulatory networks: one containing the DEM–TFEB pairs (with TFEB in their promoter regions) and the DEGs; and another consisting of the DEMs with TFEB in their promoter regions and the DEGs also associated with TFEB in their promoter regions.

5.5.1 Sample preparation

All laboratory experiments were performed at the Candiolo Cancer Institute under the supervision of Dr. Gabriella Doronzo.

In vitro studies were conducted on human umbilical vein endothelial cells (HUVECs). The TFEB gene was silenced by lentiviral transduction with a specific shRNA. In each experiment, TFEB-silenced HUVEC cells (shRNA-TFEB) were compared to control cells transduced with a nonspecific shRNA (WT).

5.5.2 mRNA-Seq analysis

RNA-libraries were prepared using TruSeq Stranded Total RNA kit. Libraries were sequenced with an Illumina NovaSeq 6000, using a paired-end protocol.

The quality of the Fastq files generated was assessed using FastQC software.⁹¹ Subsequently, the sequences were subjected to a trimming process to remove TruSeq adapters using Cutadapt v3.5⁹² in paired-end mode, thus obtaining high-quality Fastq files.

To quantify the genes expressed in all samples, the RSEM v1.3.1 pipeline⁹³ was used, employing the STAR aligner⁹⁴ in paired-end mode, with the human reference genome hg38. The gene annotations provided by the Ensembl database⁷⁹ were used as a reference for the analysis of gene expression levels.

Finally, the identification of differentially expressed genes (DEGs) between TFEB-silenced HUVEC cells (shRNA-TFEB) and control cells (WT) was performed using DESeq2,⁹⁵ applying a $|\log_2FC|$

expression change as statistical significance criteria. > 1 and an FDR value < 0.05 . For each experimental condition, two biological replicates were included.

5.5.3 miRNA-Seq analysis

Small-RNA-sequencing was performed according to the Illumina TruSeq small RNA protocol and libraries were sequenced using an Illumina NovaSeq 6000.

The quality of the generated Fastq files was assessed using FastQC software.⁹¹ Fastq files were processed using a pipeline based on miRDeep2 v.2.0.1.2,⁹⁶ applying default parameters to both the mapper.pl and miRDeep2.pl scripts, using the hg38 human genome as a reference. The miRBase database was used to annotate miRNAs and associate mature miRNA identifiers with those of their precursors miRNAs.⁸² The output generated by the pipeline was subsequently filtered considering as expressed only the miRNAs with a normalized count > 1 in at least one sample.

Identification of differentially expressed miRNAs (DEMs) between TFEB-silenced (shRNA-TFEB) and control (WT) HUVEC cells was performed using DESeq2 on raw counts.⁹⁵ DEMs with expression change $|\log_2FC| > 0.5$ and FDR value < 0.1 were considered significant.

For each experimental condition, three biological replicates were included.

5.5.4 TFEB ChIP-Seq analysis

Fastq files from the TFEB ChIP-Seq experiment in HUVEC cells were downloaded from the NCBI GEO database (accession number GSE88896) using the SRA links SRR4429371 (TFEB ChIP-seq) and SRR4429370 (Input_IgG) using the fastq-dump command. The data used are from the work of Doronzo et al. 2019.⁵⁰

The quality of the Fastq files obtained was assessed with FastQC.⁹¹ Subsequently, the sequences were aligned to the hg38 reference genome using the Bowtie2 aligner.⁹⁷ TFEB enrichment peaks were identified with MACS2,⁹⁸ comparing the HUVEC TFEB ChIP sample with the corresponding input file, applying the following parameters: *-f BAM -g hs -n output_name -B -p 0.01 --call-summits --fe-cutoff 1.0*.

TFEB peaks with a $-\log_{10}(\text{p-value}) > 3$ were considered statistically significant.

5.5.5 Bioinformatics data integration

TFEB binding peaks obtained from ChIP-Seq data were intersected with gene and pre-miRNA promoter regions using *bedtools intersect* to identify genes and miRNAs characterized by the presence of TFEB at their promoter regions.⁸⁶ For subsequent analyses, only peaks with an overlap of more than 100 nucleotides were considered.

DEMs associated with TFEB-occupied promoter sites were further annotated for the presence of specific binding motifs using the JASPAR database, specifically the MA0692.1 motif profile.¹⁸

DEM target genes were identified using TargetScan Human version 8.0.⁸³ Only DEM target genes that were also DEGs between the shRNA-TFEB and WT samples were considered. An anti-correlation analysis was then conducted, selecting only the DEM–DEG pairs that showed opposing expression trends, namely: $[\log_2FC_miRNA > 0 \text{ and } \log_2FC_gene < 0]$, or $[\log_2FC_miRNA < 0 \text{ and } \log_2FC_gene > 0]$.

From this analysis, two distinct regulatory networks were constructed:

1. A network comprising the interactions between TFEB, DEMs, and DEGs regardless of the presence of TFEB in the DEG promoters.
2. A network including only the connections where TFEB binding was verified in the promoter regions of DEGs.

These regulatory networks between TFEB, miRNAs, and genes were generated with Cytoscape.⁸⁹

Volcano plots were generated using the *scatterplot* function in R, while heatmaps were obtained using the *pheatmap()* package in R. Finally, the graphical representation of selected miRNA promoter region was created using the UCSC Genome Browser.⁸¹ H3K27ac and H3K4me3 peaks in HUVEC were downloaded from the ENCODE portal.^{19,20}

5.5.6 Functional analysis of DEGs

The upregulated and downregulated genes were subjected to a functional analysis using ToppGene, considering only selected gene ontology (GO) biological terms.⁸⁸

5.5.7 Experimental validation

MicroRNA-222 and its target genes, E2F2 and CDK1, were experimentally validated by real-time PCR by Dr. Gabriella Doronzo at the Candiolo Cancer Institute.

For mRNA analysis, RNA was reverse-transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). For miRNA expression analysis, reverse transcription was performed using the same kit in the presence of miRNA-specific stem-loop primers provided in the TaqMan MicroRNA Assays (Thermo Fisher Scientific), which allow selective conversion of mature miRNAs into cDNA. Real-time PCR was performed on a CFX96 Real-Time PCR Detection System (Bio-Rad) using TaqMan Universal PCR Master Mix or miRNA-specific TaqMan MicroRNA Assays. This technique is based on the use of a fluorescent probe that specifically binds to the target sequence. During PCR amplification, the probe is cleaved by the DNA polymerase, resulting in the emission of a fluorescent signal that increases proportionally with the amount of amplified DNA. This allows a sensitive and specific quantification of both mRNA and miRNA expression levels. TBP and RNU44 were used as endogenous reference genes for normalization of mRNA and miRNA data, respectively. The following Taqman assays (ThermoFisher Scientific, Waltham, MA, USA) were used: TFEB (human, Hs00292981_m1), E2F1 (human Hs00153451_m1), and CDK1 (human Hs00938777_m1); pri-miR-222 (Hs03303011_pri); miR-222-3p (002276); miR-222-5p (002097); TBP (human Hs00427620) and RNU44 (001094).

5.6 Overview of the methods used for the construction of the regulatory network in GBM – second project

The second project of this thesis, focused on GBM, was divided into two phases: a first phase based on the analysis of the data published by Faletti et al., 2021⁷¹ and a second phase, described in paragraph 4.6.6, in which the data from Wang et al., 2021⁶⁴ were used. Figure 4.4 graphically summarizes the analytical workflow relating to the multi-omics integration of Faletti et al., 2021 dataset.

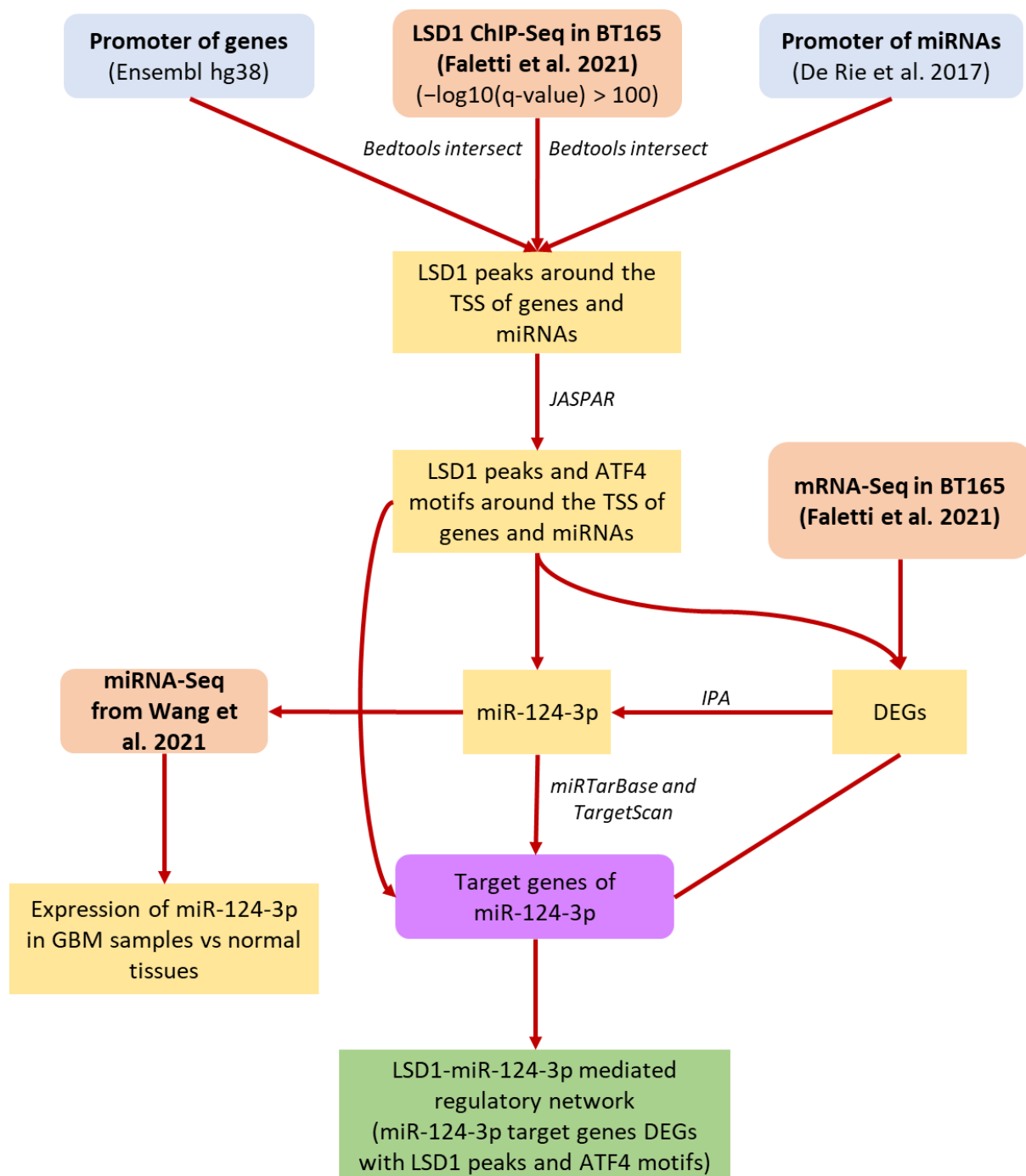


Figure 5.4. Flowchart of the methods used for the integration of Faletti et al., 2021 data. The blue rectangles represent the gene and miRNA promoters, the orange rectangles represent the omics data used, the yellow rectangles represent the intermediate results, the purple rectangle represents the predicted miRNA, and the green rectangle represents the final result. In particular, the gene and miRNA promoters, defined as the region between -2,500 and +2,500 bp around the TSS, were intersected with the LSD1 peaks obtained from ChIP-Seq in BT165 using bedtools intersect. Only genes and miRNAs characterized by the simultaneous presence of an LSD1 peak and an ATF4 motif, identified using JASPAR, were then selected. From the mRNA-Seq analysis conducted in BT165 (LSD1-silenced vs. control), DEGs were calculated and subsequently used as input for IPA, which allowed us to predict miR-124-3p as a potential upstream regulator. This miRNA was associated with LSD1 peaks and the presence of an ATF4 motif in the promoter region. For miR-124-3p, target genes were then defined by integrating information from miRTarBase and TargetScan. To build the final network, only miR-124-3p target genes that were also DEGs between LSD1-silenced and control conditions and that presented both LSD1 peaks and an ATF4 motif at the promoter level were considered. Finally, miR-124-3p expression was assessed in the data published by Wang et al., 2021, comparing GBM samples with normal tissues. This dataset served as the starting point for the second part of the project (section 4.6.6).

5.6.1 Faletti et al., 2021 dataset

The starting point for this second project was the analysis of omics data generated and published by Professor Pelicci's group at the European Institute of Oncology (IEO).⁷¹ These data derive from the BT165 (GBM#22) cell line, a model of GBM TICs with stemness properties, obtained from a patient with proneural GBM. The mRNA-Seq, LSD1 ChIP-Seq, and ATAC-Seq raw data were downloaded from NCBI GEO under accession GSE181379 and used as the basis for the multi-omics integration analyses performed in the first part of this work.

5.6.2 mRNA-Seq analysis

The mRNA-Seq experiment generated by Faletti et al. 2021 on the BT165 cell line was performed under two experimental conditions: LSD1-silenced (sh71) and control (shNT), each with two biological replicates.⁷¹

Fastq files quality was assessed with FastQC.⁹¹ Gene quantification was performed using the RSEM v1.3.1 pipeline,⁹³ using STAR⁹⁴ as the reference aligner in single-end mode, with the human hg38 genome as a reference. Gene annotation was obtained from the Ensembl database.⁷⁹ Genes with a TPM > 1 in at least one sample were considered expressed.

The analysis of differentially expressed genes between the LSD1-silenced and control conditions was conducted with EdgeR,¹⁰⁰ selecting genes with $|\log_2FC| > 1.2$ and $FDR \leq 0.05$ as DEGs. The identified

genes were used as input for a Core Analysis using QIAGEN IPA (QIAGEN Inc., <https://digitalinsights.qiagen.com/IPA>) to identify potential upstream regulators.⁸⁷ In particular, our attention was paid to miR-124-3p (and other miRNAs w/seed AAGGCAC) found as upstream regulator.

5.6.3 LSD1 ChIP-Seq analysis

The quality of Fastq files for the input and vehicle of the LSD1 ChIP-Seq experiment was assessed with FastQC.⁹¹ Subsequently, the files were aligned to the hg38 reference genome using Bowtie2.⁹⁷ Duplicate reads were removed using *samtools fixmate* and *samtools markdup* commands.¹⁰¹ Peak identification was performed with MACS2, using the following command: *macs2 callpeak -t Vehicle.alllineato.sorted.rmdup.bam -c Input.alllineato.sorted.rmdup.bam -f AUTO -g hs -n Vehicle.out.macs2 -B -q 0.05*.⁹⁸

Only LSD1 peaks with a $-\log_{10}(\text{q-value}) > 100$ were considered for further analysis.

5.6.4 ATAC-Seq analysis

Fastq files from the ATAC-Seq were assessed with FastQC.⁹¹ Reads were trimmed using Cutadapt⁹² and then aligned to the hg38 human reference with Bowtie2,⁹⁷ without mitochondrial DNA. Duplicated reads were removed using *samtools fixmate* and *samtools markdup* command.¹⁰¹ Peaks were called with HMMRATAC with standard parameters and were filtered using score ≥ 10 .¹⁰²

5.6.5 Bioinformatics data integration

LSD1 peaks obtained by ChIP-Seq in the BT165 cell line were intersected with the promoter regions of pre-miRNAs and genes (defined as the range around the TSS between -2500 and +2500 bp) using *bedtools intersect*.⁸⁶ Only peaks with a nucleotide overlap >100 were considered for further analysis. From the work of Faletti et al. 2021,⁷¹ we know that ATF4 acts as a cofactor for LSD1 in the regulation of protein-coding genes; however, in the absence of ChIP-Seq data for ATF4 in the BT165 cell line, we used the ATF4 motifs present in JASPAR (matrix profile MA0833.3).¹⁸ This allowed us to

determine the number of genes and pre-miRNAs characterized by the presence of ATF4 motifs in their promoter regions.

For differentially expressed genes (DEGs) between LSD1-silenced and control conditions, we associated the ATF4 motif score values from JASPAR. The 25th and 75th percentiles were calculated from the score distribution using Pandas' *pd.cut()* function, and each gene was then classified into three groups:

- Low: score < 25th percentile.
- Medium: score between 25th and 75th percentile.
- High: score > 75th percentile.

A heatmap was generated with *pheatmap()* package in R, including a double annotation: the ATF4 score classification (Low/Medium/High, from JASPAR) and the presence of LSD1 peaks (ChIP-Seq in BT165) in the promoter regions of the DEGs.

The DEGs obtained between LSD1-silenced and controls were analyzed with IPA, which identified miR-124-3p (along with other miRNAs seeded with AAGGCAC) as an upstream regulator.⁸⁷ The immature forms of the miRNAs were analyzed for the presence of LSD1 peaks and ATF4 motifs, and the graphical representation of the promoters of these pre-miRNAs was obtained using UCSC Genome Browser.⁸¹ Furthermore, to assess chromatin accessibility, ATAC-Seq and LSD1 ChIP-Seq data in BT165 were uploaded as custom tracks to UCSC Genome Browser.

Attention was then focused specifically on miR-124-3p. Its target genes were identified by integrating information from miRTarBase⁸⁴ and TargetScan;⁸³ subsequently, only targets effectively expressed in mRNA-Seq (TPM > 1 in at least one sample) were retained for further analysis. To construct the final network, we considered only miR-124-3p targets that were also DEGs between LSD1-silenced and control and characterized by LSD1 peaks and ATF4 motifs in their promoter regions. The network was created using Cytoscape.⁸⁹

Finally, since miRNA-Seq data were not available in the BT165 cell line, we used those published by Wang et al., 2021,⁶⁴ comparing miR-124-3p expression between GBM samples and normal tissues. Expression levels in log₂(TPM) were represented with a boxplot, generated in R with the *boxplot()* function. Wilcoxon test was used to compare the expression value of miR-124-3p from Wang et al., 2021 in GBM samples and normal tissue samples.

5.6.6 Wang et al., 2021 dataset

In the second phase of the second project, a multi-omics analysis was conducted based on data published by Wang et al., 2021, including mRNA-Seq, miRNA-Seq, and circRNA-Seq data. All datasets were downloaded from the supplementary files of the original publication.⁶⁴

Only GBM samples belonging to the proneural subtype, for which all three types of omics data were available, were selected. In total, 29 tumor samples were considered, common to mRNA-Seq, miRNA-Seq, and circRNA-Seq.

For control samples, 8 normal tissue samples available for the mRNA-Seq and miRNA-Seq datasets were used; for the circRNA-Seq data, however, no control samples were available.

For each dataset, only transcripts expressed in at least one sample were retained, applying the following filtering thresholds: FPKM > 1 for mRNA-Seq and circRNA-Seq data, and TPM > 1 for miRNA-Seq data.

The identification of differentially expressed genes (DEGs) and miRNAs (DEMs) was performed using the Wilcoxon test, using significance criteria of $|\log_2FC| > 2$ and FDR < 0.01 for DEGs, and $|\log_2FC| > 2$ and FDR < 0.05 for DEMs.

For circRNAs, due to the absence of control samples, differential expression analysis was not performed. Therefore, circRNAs expressed in at least one sample with FPKM > 1 were considered, limiting the analysis to those annotated in the CSCD2 database and filtered according to the Miranda and TargetScan criteria.⁸⁵

Subsequently, a Spearman anticorrelation analysis was performed between the DEMs and DEGs using the *spearmanr* Python package and considering only miRNA–gene interactions annotated in the TargetScan Human database.⁸³ Pairs with a Spearman correlation coefficient $\rho < -0.6$ and a p-value < 0.05 were selected as significant. CircRNAs predicted to target the DEMs identified in this analysis were subsequently retrieved from the CSCD2 database.⁸⁵

To assess the possible epigenetic involvement of LSD1, we analyzed whether the identified DEGs and DEMs exhibited LSD1 binding peaks near their respective promoter regions, defined as the range between –2,500 and +2,500 bp from the Transcription Start Site (TSS). To this end, we used LSD1 ChIP-Seq data obtained from BT165 cells and published by Faletti et al., 2021, and the intersection between the genomic coordinates was calculated using the *bedtools intersect* software.⁸⁶

Heatmaps representing DEGs and DEMs annotated with the presence or absence of LSD1 peaks were generated using the *pheatmap()* package in R.

Finally, a regulatory network centered on LSD1 was constructed by integrating DEM–DEG, circRNA–DEM, and LSD1 interactions using the *NetworkX* package in Python.⁹⁰

In Figure 4.5 is represented the schematic workflow used in this second phase of the project.

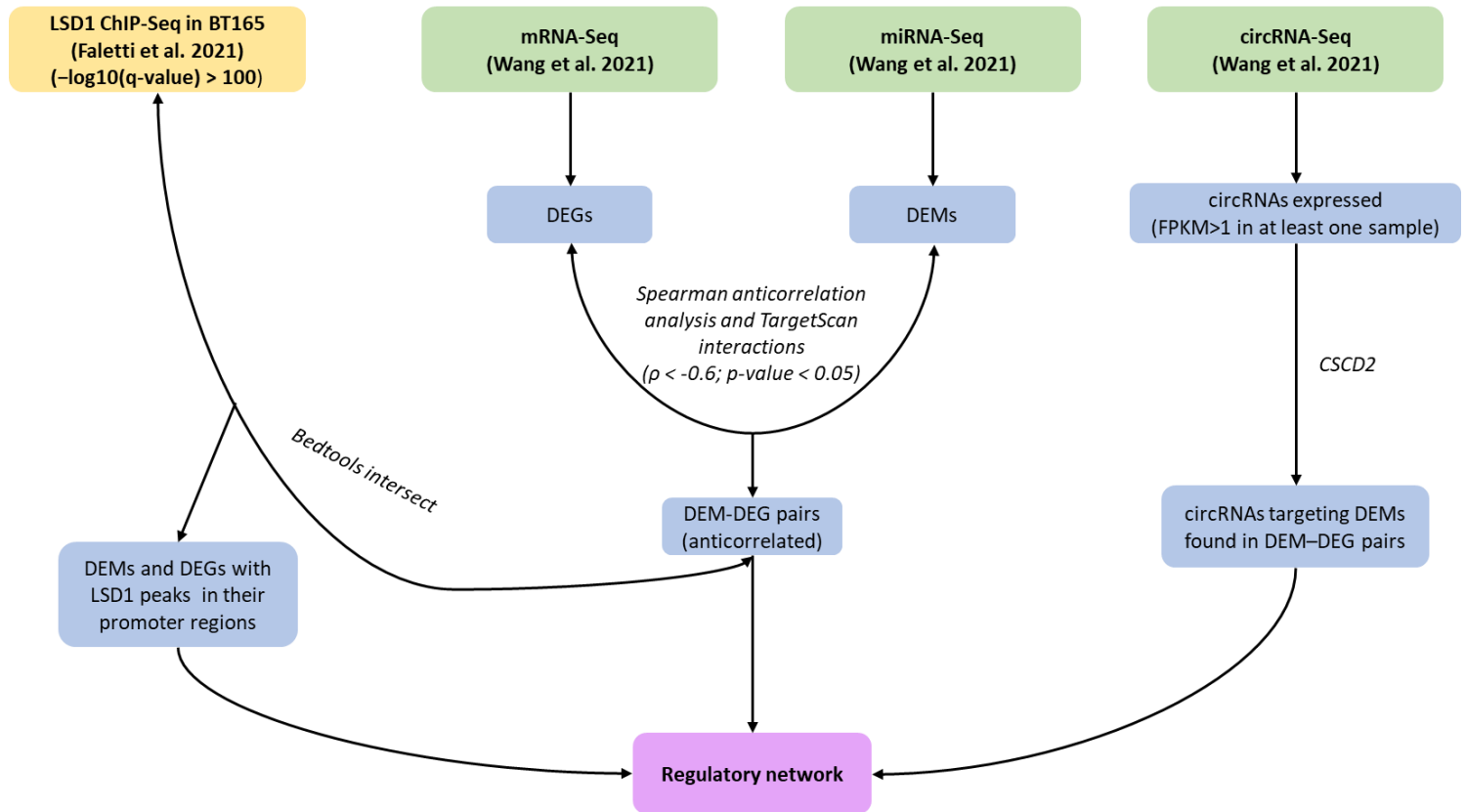


Figure 5.5. Flowchart of the methods used for the integration of Wang et al., 2021 data. The yellow rectangle represents the LSD1 ChIP-Seq dataset from Faletti et al., 2021; green rectangles represent the omics data from Wang et al., 2021; blue rectangles indicate intermediate results, and purple rectangle represents the final output of the analysis.

6. RESULTS

6.1 Results from the regulatory network in endothelium – first project

In the first project of this thesis, a bioinformatics pipeline was developed based on the integration of various omics data (ChIP-Seq, mRNA-Seq, and miRNA-Seq) with the aim of understanding the role of TFEB in HUVEC, and its function in the regulation of miRNAs (Figure 6.1).

The main goal of the project was to investigate the role of TFEB in the regulation of genes and miRNAs. To this end, the promoter regions of genes and miRNAs were defined, considering a range of -2500 and +2500 bps with respect to their TSS. TSS gene coordinates were obtained from Ensembl Biomart,⁷⁹ while miRNA coordinates were obtained from De Rie et al., 2019.⁸⁰ Overall, 111,050 TSSs were obtained from Ensembl Biomart for protein-coding genes, 141,155 for non-coding RNAs, and 1,130 TSSs for miRNAs from De Rie et al., 2019.

ChIP-Seq analysis of TFEB in HUVEC, described in section 5.5.4 of the Methods, identified 26,615 binding peaks on the human hg38 genome, considering those with a $-\log_{10}(\text{p-value}) > 3$. The intersection of the TFEB peak coordinates and those of the TSSs highlighted 5,588 genes and 298 mature miRNAs characterized by the presence of TFEB in the promoter region, and therefore potentially regulated by the transcription factor.

Subsequently, using mRNA-Seq and miRNA-Seq data, we identified genes and miRNAs differentially expressed (DEGs and DEMs) following TFEB silencing. The integration of these results allowed the construction of a TFEB-driven regulatory network, which is essential for understanding the role of ncRNAs in the endothelial model.

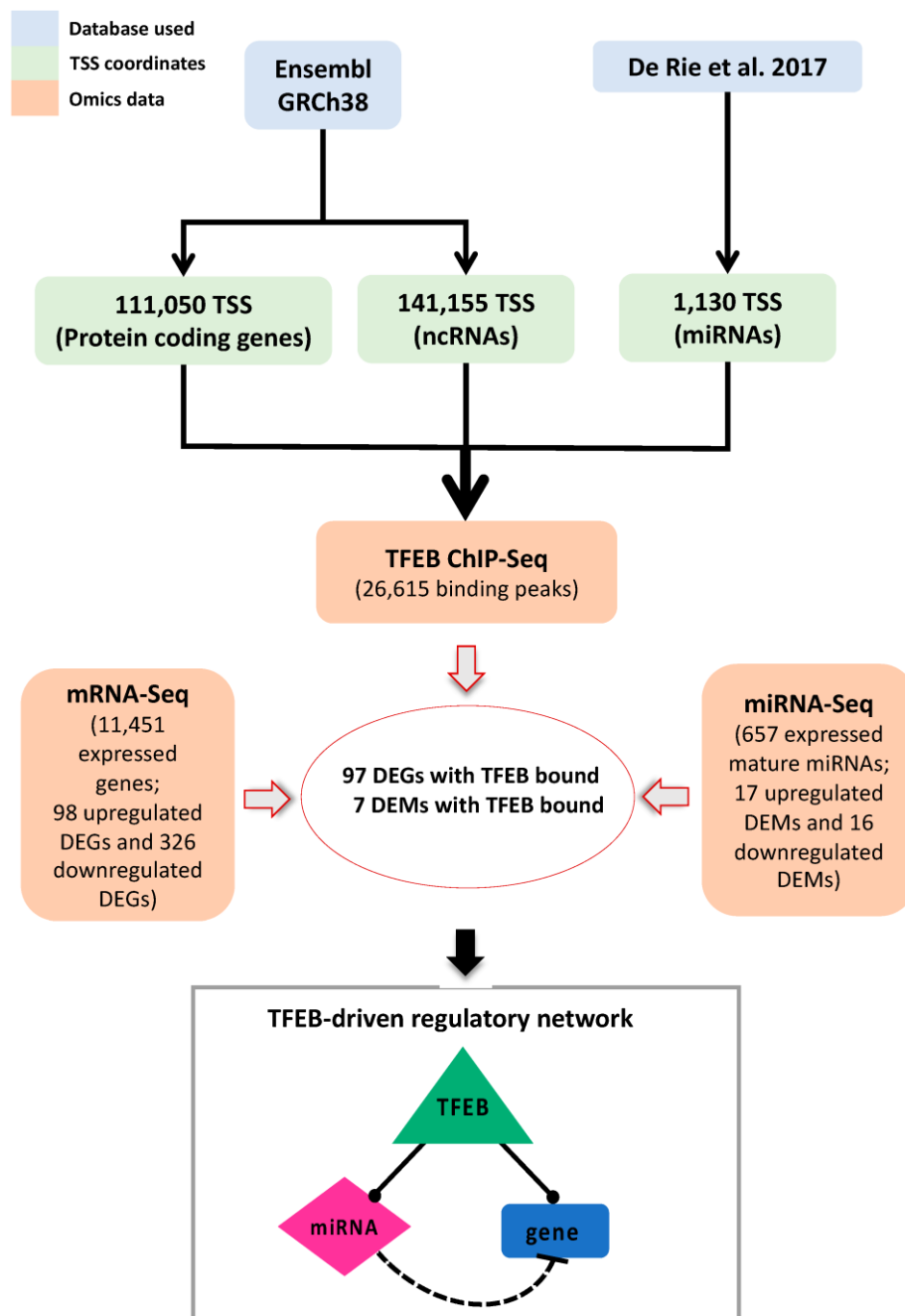


Figure 6.1. Schematic representation of the bioinformatics workflow used to define the TFEB-driven regulatory network in HUVECs. The green rectangles represent the TSS of non-coding RNA and miRNAs obtained from the Ensembl Biomart and De Rie et al. 2017 database (blue rectangles). The orange rectangles represent the omics data used in the work (TFEB ChIP-Seq, miRNA-Seq and mRNA-Seq). The circle in the middle of the figure represents the integration of these data, thus indicating the DEGs and DEMs characterized by TFEB at the level of their promoter regions. The results obtained were graphically represented with a comprehensive regulatory network, with TFEB as the master regulator.

6.1.1 The TFEB-regulated transcriptional program in HUVEC

To investigate the TFEB-regulated transcriptional program in HUVEC cells, we analyzed transcriptome changes resulting from TFEB silencing. To this end, an mRNA-seq experiment was performed comparing two experimental conditions: HUVEC treated with a TFEB-specific shRNA and control cells infected with a non-targeting shRNA. After transcript quantification, only genes with TPM values > 1 in at least one sample were considered expressed, thus defining 11,451 expressed genes. Among these, 424 were found to be differentially expressed (DEGs) between TFEB-silenced cells and controls, of which 98 were upregulated and 326 were downregulated, as shown in the volcano plot in Figure 6.2A.

Integrating TFEB ChIP-seq data revealed that 97 of these 424 DEGs exhibit transcription factor binding peaks in their promoter regions. The unsupervised hierarchical clustering complete set of DEGs is represented in the heatmap in Figure 6.2B, which also reports the annotation for the presence of TFEB peaks in the corresponding promoters.

Figure 6.2C shows a Venn diagram representing the relationship between differentially expressed genes, those associated with TFEB peaks, and their intersection, which represents DEGs directly related to TFEB binding at the level of their promoter regions.

To understand the functional role of the 97 DEGs characterized by TFEB in their promoter regions, we distinguished between up- and down-regulated DEGs and conducted a functional enrichment analysis using ToppGene, focusing exclusively on Gene Ontology (GO) terms.⁸⁸ The results showed that down-regulated genes are mainly associated with processes related to the cell cycle and mitosis, while up-regulated genes are related to angiogenesis and vascular development, as shown in Figure 6.2D.

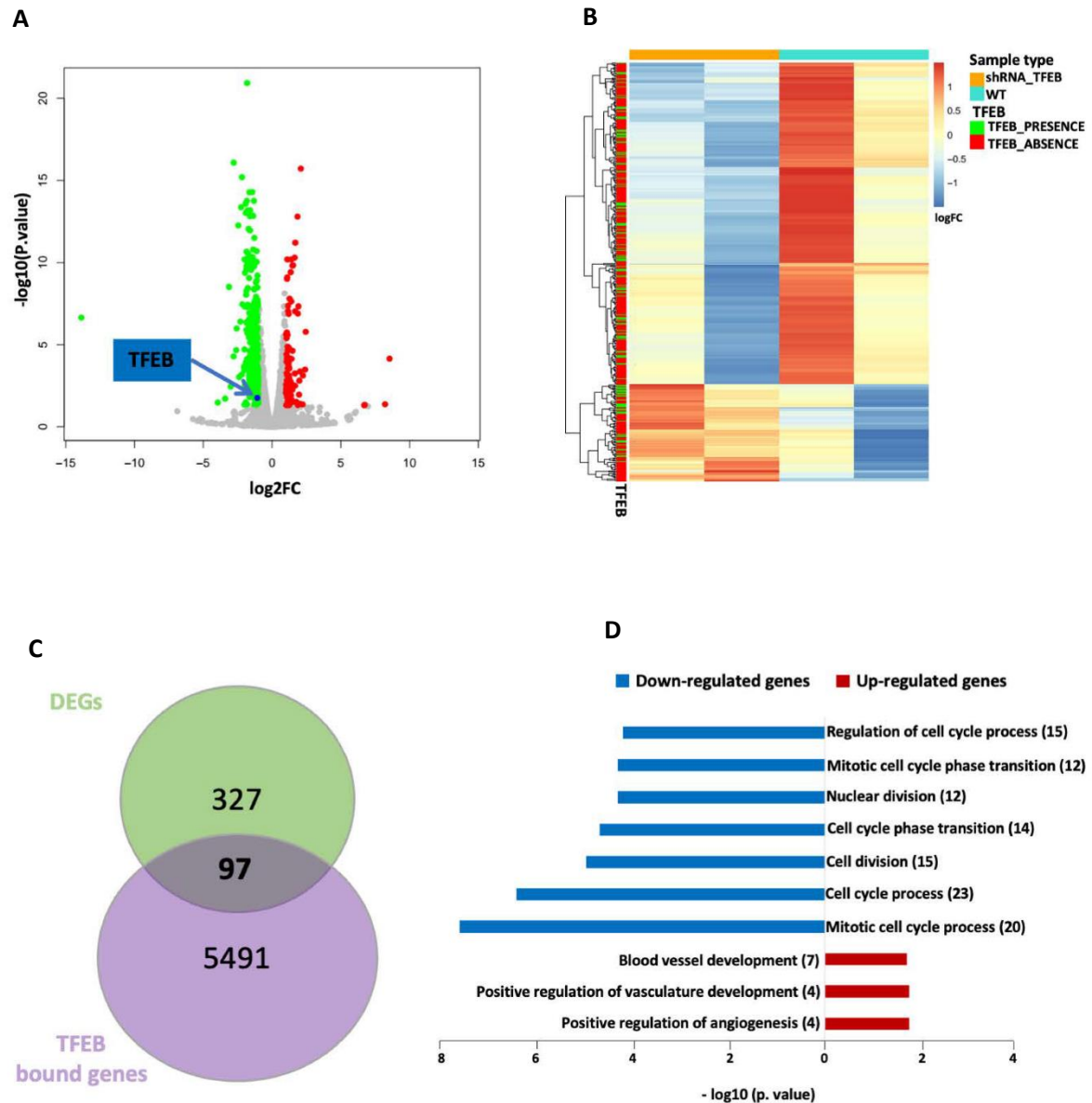


Figure 6.2. Transcriptomic analysis by mRNA-seq highlights the role of TFEB in modulating genes in HUVECs. (A) Volcano plot illustrating the differentially expressed genes (DEGs) between TFEB-silenced HUVECs (shRNA-TFEB) and control cells (WT). Downregulated genes are shown in green, upregulated genes in red, and the TFEB gene itself is indicated in blue. (B) Heatmap displaying the unsupervised hierarchical clustering of DEGs in TFEB-silenced (orange) versus control (cyan) HUVECs. Each gene is annotated to indicate the presence (green) or absence (red) of at least one TFEB binding peak in its promoter region. (C) Venn diagram showing the intersection between DEGs in TFEB-silenced and control cells with the full set of genes carrying TFEB binding peaks at their promoters. (D) Representative GO terms from the functional enrichment analysis performed with ToppGene on DEGs. Terms enriched among downregulated genes are shown in blue, while those associated with upregulated genes are in red. The number of genes per term is reported in parentheses. GO terms are ranked according to $-\log_{10}(p\text{-value})$, adjusted with the Benjamini–Hochberg correction.

6.1.2 The TFEB-regulated miRNA program in HUVEC

To further investigate the role of miRNAs under the same experimental conditions used for mRNA-seq analysis, a miRNA-Seq experiment was conducted comparing TFEB-silenced HUVEC cells with control cells. Differential analysis identified 33 differentially expressed miRNAs (DEMs) between the two groups. Subsequently, integration with TFEB ChIP-Seq data allowed us to assess the presence of transcription factor binding sites in the promoter regions of the DEMs, revealing that 7 of the 33 miRNAs exhibit TFEB peaks. The complete list of DEMs and the annotation for the presence of TFEB in the promoter regions are shown in Table 6.1.

Differentially expressed miRNAs (DEMs)	TFEB
hsa-miR-27a-5p	PRESENCE
hsa-miR-222-5p	PRESENCE
hsa-miR-665	ABSENCE
hsa-miR-181a-3p	PRESENCE
hsa-miR-4741	ABSENCE
hsa-miR-181b-3p	PRESENCE
hsa-miR-92a-1-5p	ABSENCE
hsa-miR-487a-3p	ABSENCE
hsa-miR-143-3p	ABSENCE
hsa-miR-431-5p	ABSENCE
hsa-miR-1307-5p	ABSENCE
hsa-miR-370-3p	ABSENCE
hsa-miR-29b-1-5p	PRESENCE
hsa-miR-433-3p	ABSENCE
hsa-miR-494-3p	ABSENCE
hsa-miR-339-3p	PRESENCE
hsa-miR-432-5p	ABSENCE
hsa-miR-493-5p	ABSENCE
hsa-miR-1185-1-3p	ABSENCE
hsa-miR-25-5p	ABSENCE
hsa-miR-409-3p	ABSENCE
hsa-miR-193a-3p	PRESENCE

hsa-miR-323b-3p	ABSENCE
hsa-miR-410-3p	ABSENCE
hsa-miR-503-5p	ABSENCE
hsa-miR-100-3p	ABSENCE
hsa-miR-887-5p	ABSENCE
hsa-miR-3187-3p	ABSENCE
hsa-miR-26a-2-3p	ABSENCE
hsa-miR-3085-3p	ABSENCE
hsa-miR-496	ABSENCE
hsa-miR-671-5p	ABSENCE
hsa-miR-193b-5p	ABSENCE

Table 6.1. List of differentially expressed miRNAs identified in TFEB-silenced HUVECs with promoter binding information from TFEB ChIP-seq. The first column reports the name of each differentially expressed miRNAs (DEMs), while the second column indicates the presence or absence of TFEB binding peaks in the corresponding promoter regions, as identified by TFEB ChIP-Seq analysis. The 7 DEMs characterized by TFEB peaks in their promoter regions are highlighted in yellow.

To determine whether the promoters of the 33 miRNAs were associated with regions of open chromatin, the H3K27ac and H3K4me3 signals available in the ENCODE database were analyzed.^{19,20} The results showed that 29 of the 33 miRNAs exhibited H3K27ac and H3Kme3 signals at their promoter level. These data are illustrated in Figures 6.3A and 6.3B: the first represents a heatmap annotated with the presence of TFEB and epigenetic signals from ENCODE, while the second provides a Venn diagram showing the intersection between the DEMs and the miRNAs characterized by TFEB peaks in their respective promoter regions.

Finally, Figure 6.3C shows a heatmap for a subset of the 33 DEMs, consisting of only the 7 miRNAs associated with TFEB peaks in the promoter. Interestingly, all these miRNAs (hsa-miR-181a-3p, hsa-miR-181b-3p, hsa-miR-29b-1-5p, hsa-miR-222-5p, hsa-miR-27a-5p, hsa-miR-193a-3p, and hsa-miR-339-3p) were significantly up-regulated after TFEB silencing.

Given the biological significance of these findings, we chose to focus our attention on miR-222, which is known to negatively regulate the angiogenic activity of the stem cell factor (SCF) by targeting its receptor, c-Kit.⁹⁹ Figure 6.3D illustrates the genomic region encompassing the miR-222 locus on the reference human hg38 genome, as depicted by the UCSC Genome Browser.⁸¹ This region not only shows a TFEB binding peak in the promoter region of the pri-miRNAs but also displays additional signals from the H3K27ac and H3K4me3 HUVEC tracks obtained from

ENCODE,^{19,20} supporting the notion of an actively transcribed region and highlighting the complex regulation by TFEB in the endothelial environment.

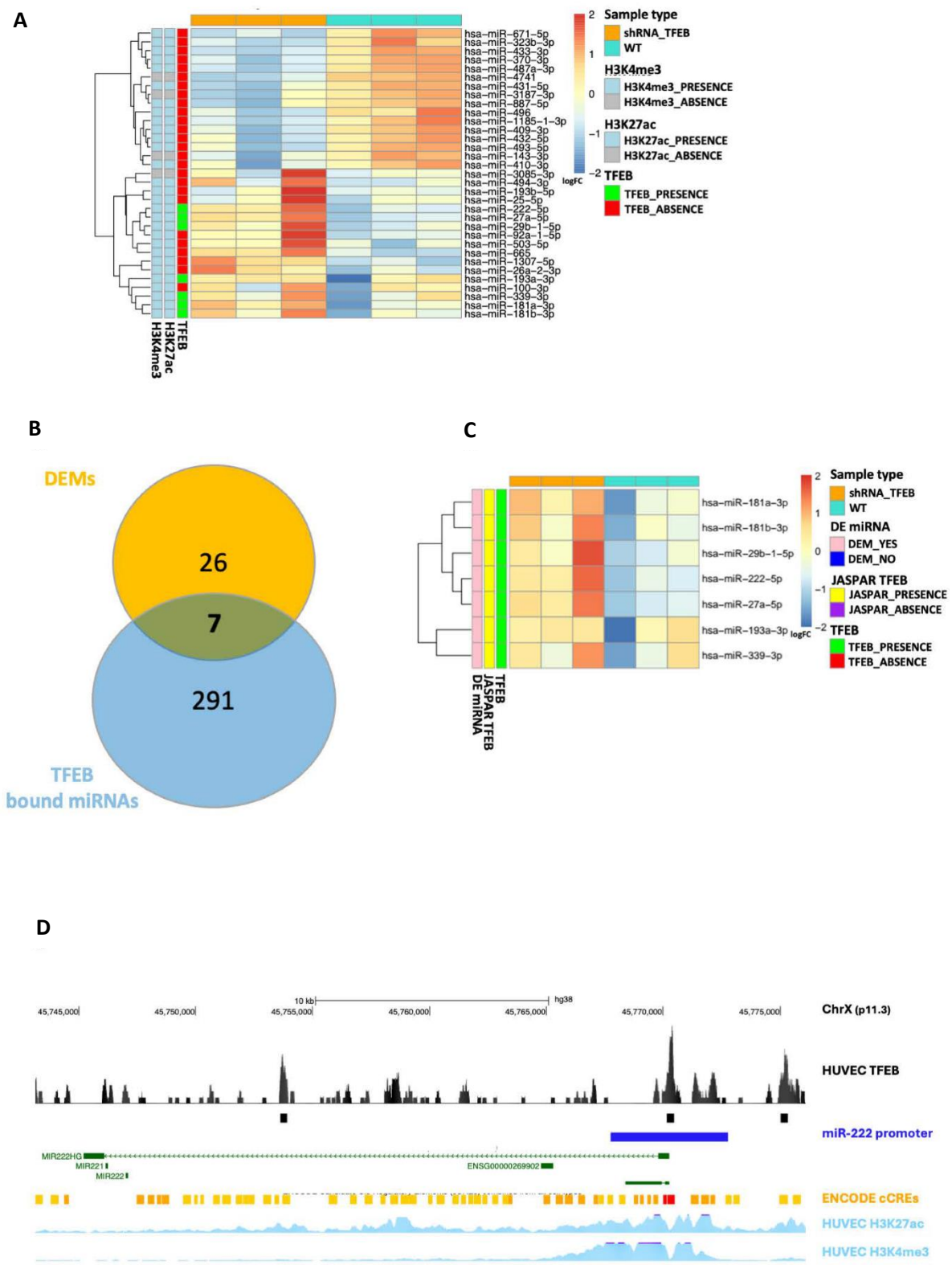


Figure 6.3. TFEB-dependent regulation of miRNAs in HUVECs. (A) Heatmap representing the unsupervised hierarchical clustering of differentially expressed miRNAs (DEMs) between TFEB-silenced HUVECs (shRNA-TFEB, orange) and control cells (WT, cyan). For each miRNA, the annotation indicates the presence (green) or absence (red) of TFEB binding in the promoter region, as well as the presence (light blue) or absence (light grey) of the histone marks H3K27ac and H3K4me3 according to ENCODE HUVEC data. (B) Venn diagram illustrating the overlap between DEMs identified in TFEB-silenced versus control cells and the set of miRNAs with TFEB peaks in their promoters. (C) Heatmap showing the clustering of the subset of 7 DEMs that display TFEB binding at their promoter regions. Each of these miRNAs is annotated with the presence (green) or absence (red) of the TFEB from ChIP-Seq, but also for the presence (yellow) or absence (purple) of the TFEB binding motif, as predicted from the JASPAR database. (D) Genomic view of the miR-222 locus in the human hg38 reference genome (UCSC Genome Browser). The panel highlights the host gene (green) and the putative promoter region (blue, defined as -2500 to +2500 bp around the TSS). The TFEB binding signal in HUVECs is shown at the top (black), together with H3K27ac and H3K4me3 tracks from ENCODE HUVECs (bottom).

6.1.3 Regulatory network of TFEB-associated miRNAs and their target genes

In subsequent analyses, we focused exclusively on the 7 DEMs characterized by the presence of at least one TFEB peak at their promoter regions, with the aim of identifying their target genes. To this end, we used the TargetScan Human database,⁸³ considering only the target genes present in the database and expressed by RNA-seq analysis following TFEB silencing. Subsequently, only the DEM–target gene pairs with anti-correlated expression were selected, obtaining a total of 3,073 genes anticorrelated to the seven DEMs. Among these, 152 genes were also found to be differentially expressed (DEGs) when comparing TFEB-silenced HUVECs with controls. Figure 6.4A shows a Venn diagram representing the intersection between the DEGs identified in the RNA-seq and the target genes of the 7 miRNAs identified by TargetScan and expressed in the transcriptomic analysis. Figure 6.4B illustrates the regulatory network generated by the 7 DEMs, all associated with a TFEB binding peak in their promoter regions, together with their 152 differentially expressed target genes identified after TFEB silencing.

Among these 152 genes, 44 also displayed TFEB binding at their promoters, whereas the remaining 108 did not. To provide a more detailed view, Figure 6.4C presents a TFEB-driven miRNA–target gene interaction network, highlighting specifically those target genes directly bound by TFEB.

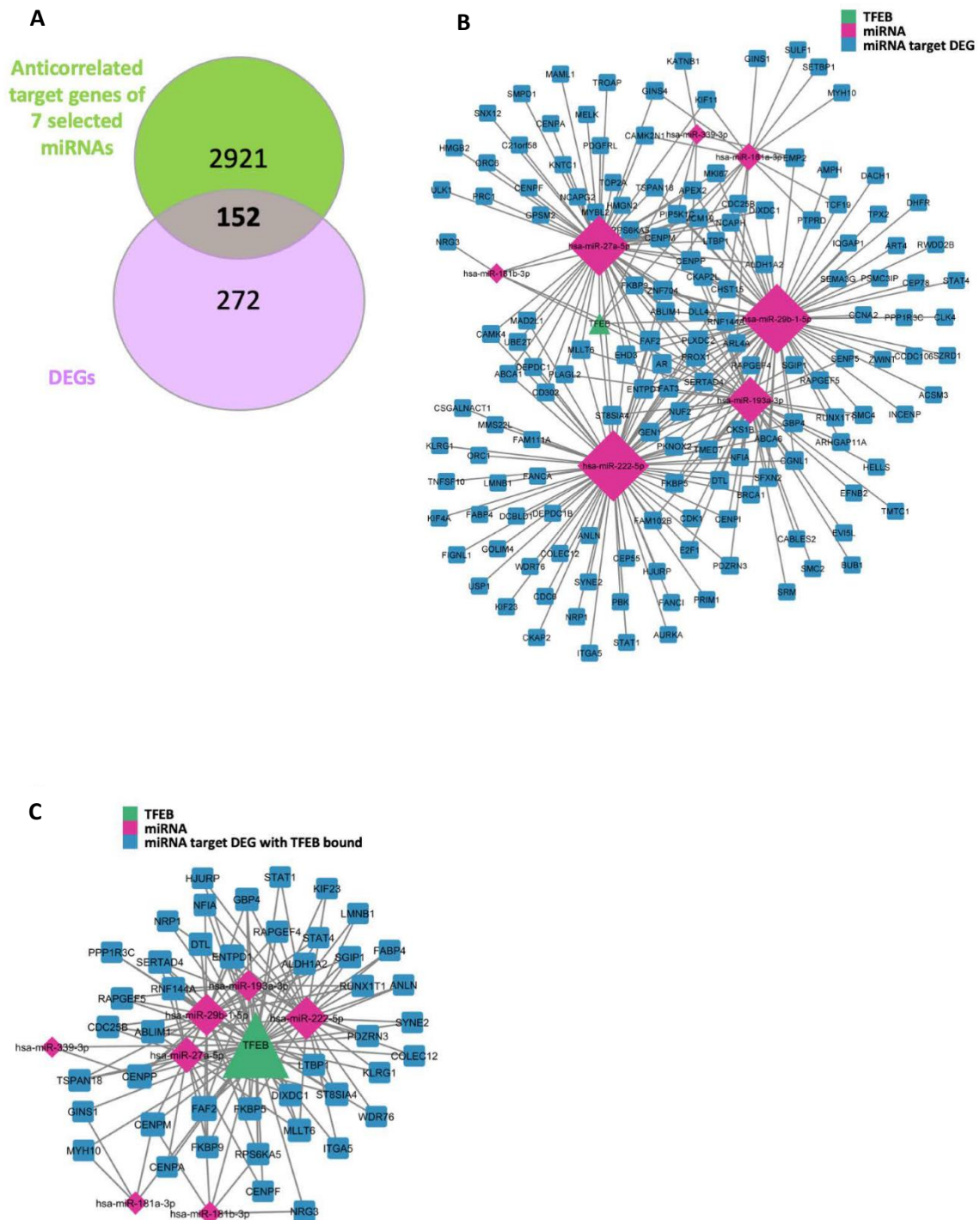


Figure 6.4. TFEB-miRNA-mediated regulatory networks. (A) Venn diagram summarizing the intersection between the anti-correlated target genes of the seven DEMs in TFEB-silenced versus control HUVECs and the set of DEGs identified by RNA-seq. (B) Schematic network representation including TFEB, the seven DEMs characterized by TFEB peaks in their promoters, and their 152 differentially expressed target genes. In the scheme, TFEB is shown as green triangles, miRNAs as magenta diamonds, and target genes as blue squares. A fraction of these genes also shows TFEB binding in their promoters. (C) Subnetwork focusing specifically on TFEB, the seven DEMs, and the 44 differentially expressed target genes with TFEB binding at their promoter regions.

6.1.4 Experimental validation of miR-222

After defining the regulatory networks, we selected miR-222 for experimental validation, given its well-established role in negatively modulating angiogenic activity reported in the literature.⁹⁹ The experimental validation of this project was conducted by Dr. Gabriella Doronzo.

The miRNA-seq analysis revealed that TFEB silencing resulted in increased expression of miR-222 (Figure 6.3A–D). In parallel, bioinformatics integration with the TFEB ChIP-seq dataset highlighted a TFEB binding peak in the promoter region of miR-222, suggesting that TFEB may directly regulate its transcription (Figure 6.3D). To test this hypothesis, we measured the expression of both pri-miR-222 and the mature isoforms miR-222-3p/5p in control and TFEB-silenced HUVECs. As expected, TFEB knockdown caused a marked upregulation of both the primary transcript and the mature forms (Figure 6.5A).

Anti-correlation analysis further identified several differentially expressed genes as putative miR-222 targets, including E2F1 and CDK1, two well-known regulators of cell cycle progression. In line with the increase of miR-222 upon TFEB silencing, both E2F1 and CDK1 were downregulated in HUVECs (Figure 6.5B), supporting the predicted interactions.

Altogether, these results demonstrate that TFEB controls critical cell cycle regulators in endothelial cells indirectly, through the modulation of miR-222 and its downstream targets.

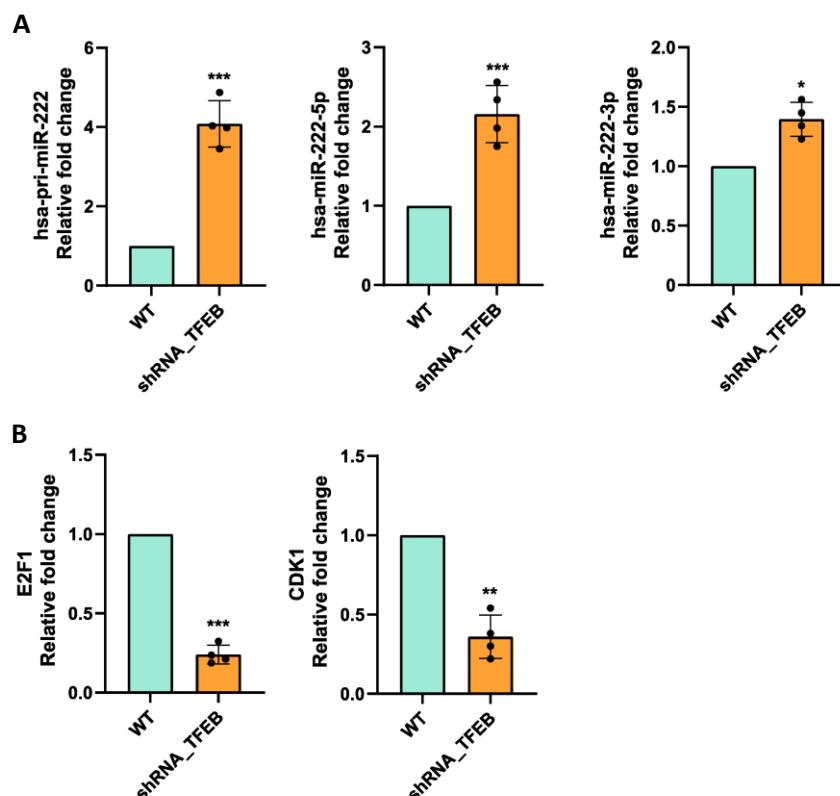


Figure 6.5. TFEB–miR-222–E2F1/CDK1 regulatory axis in HUVECs. (A, B) Quantitative PCR (qPCR) assessment of hsa-pri-miR-222 and the mature isoforms hsa-miR-222-5p and -3p (a), as well as E2F1 and CDK1 (b), in TFEB-depleted (shRNA-TFEB) and control (WT) HUVECs. Expression levels are shown as fold changes relative to WT cells and normalized to the housekeeping gene TBP (n = 4, mean ± SEM; * p-value < 0.01, ** p-value < 0.001, *** p-value < 0.0001, Student's t-test).

Figure 6.6 shows a summary scheme of the TFEB/miR-222/E2F1-CDK1 regulatory network. According to our hypothesis, TFEB under physiological conditions directly represses miR-222, allowing the expression of the E2F1 and CDK1 genes and promoting endothelial cell proliferation. Following TFEB silencing, miR-222 is upregulated, leading to the inhibition of E2F1 and CDK1 and a consequent slowing of cell cycle progression.

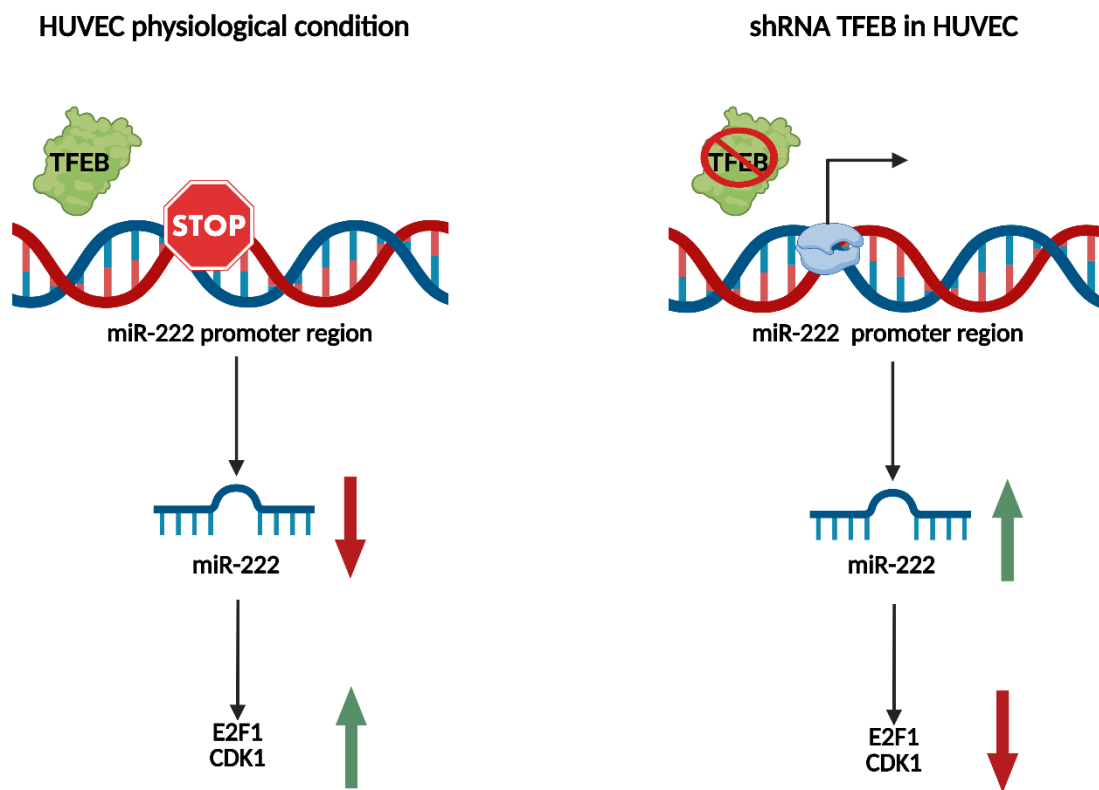


Figure 6.6. Regulatory scheme of the TFEB/miR-222/E2F1-CDK1 pathway in HUVECs. Under physiological conditions, TFEB (green) represses miR-222, allowing the expression of its targets E2F1 and CDK1 and normal endothelial proliferation. After TFEB silencing, miR-222 is upregulated and inhibits E2F1 and CDK1, slowing cell cycle progression. (Image created with BioRender).

6.2 Results from the regulatory network in GBM – second project

The second project of this thesis focuses on the construction of a regulatory network in glioblastoma (GBM), with particular attention to LSD1, an epigenetic regulator that plays a key role in maintaining stemness in GBM TICs. GBM TICs are responsible for the frequent relapses observed in patients, and understanding the regulatory mechanisms that sustain them is a crucial aspect of GBM research.

This project is divided into two main phases:

- **Phase 1 – Analysis of data from Faletti et al., 2021**

The starting point was the analysis of multi-omics data (LSD1 ChIP-Seq, mRNA-Seq, and ATAC-Seq) generated in the BT165 cell model, derived from a patient with the proneural subtype, published by Faletti et al., 2021. This analysis allowed us to identify genes and miRNAs characterized by the presence of LSD1 in their promoter regions, along with its known cofactor ATF4. Furthermore, from the mRNA-Seq data, by comparing the LSD1-silenced condition with the control condition and through the analysis of differentially expressed genes, it was possible to predict miR-124-3p as a potential upstream regulator using IPA.

These results confirmed the complex and already known role of LSD1 on protein-coding genes, but also highlighted the need to further explore the non-coding dimension.

- **Phase 2 – Analysis of data from Wang et al., 2021**

Since miRNA-Seq data are not available for BT165 cells and starting from the hypothesis that non-coding RNAs (particularly miRNAs and circRNAs) may also play a role in GBM tumorigenesis, we extended the analysis using open-access multi-omics data made available by Wang et al., 2021.

Specifically, we considered RNA-Seq, miRNA-Seq, and circRNA-Seq from 29 samples of proneural subtype GBM, consistent with our BT165 model.

This second phase allowed us to include circRNAs in the analysis and hypothesize the construction of an LSD1-mediated miRNA–circRNA regulatory network in proneural GBM.

6.2.1 Combined Analysis of LSD1 Binding Sites and ATF4 Motifs in Gene and miRNA Promoter Regions

Analysis of LSD1 ChIP-Seq data in the BT165 cell line, aligned to the human reference genome hg38, revealed a total of 56,959 LSD1 binding peaks with statistical significance ($-\log_{10}(\text{q-value}) > 100$), distributed across diverse genomic regions, including exonic, intronic, intergenic, and promoter regions.

To assess genes and pre-miRNAs potentially regulated by LSD1, these peaks were intersected with promoter regions, defined as the range between -2,500 and +2,500 bp relative to the TSS, considering only nucleotide overlaps greater than 100 bp. This analysis revealed 29,795 LSD1 peaks in the promoters of 24,604 genes, while 972 peaks were detected near the promoters of 765 pre-miRNAs, corresponding to 1,164 mature miRNAs.

Since the study by Faletti et al., 2021, demonstrated that the transcription factor ATF4 acts as a cofactor for LSD1, the same promoter regions (-2,500/+2,500 bp) of the genes and pre-miRNAs associated with LSD1 were also analyzed for the presence of ATF4 binding motifs, using the JASPAR database, which defines a canonical 10-nucleotide motif. Predictive analysis revealed that 24,519 genes and 759 pre-miRNAs (equivalent to 1,154 mature miRNAs) simultaneously exhibit an LSD1 binding peak and an ATF4 binding motif in their promoter regions (Figure 6.7).

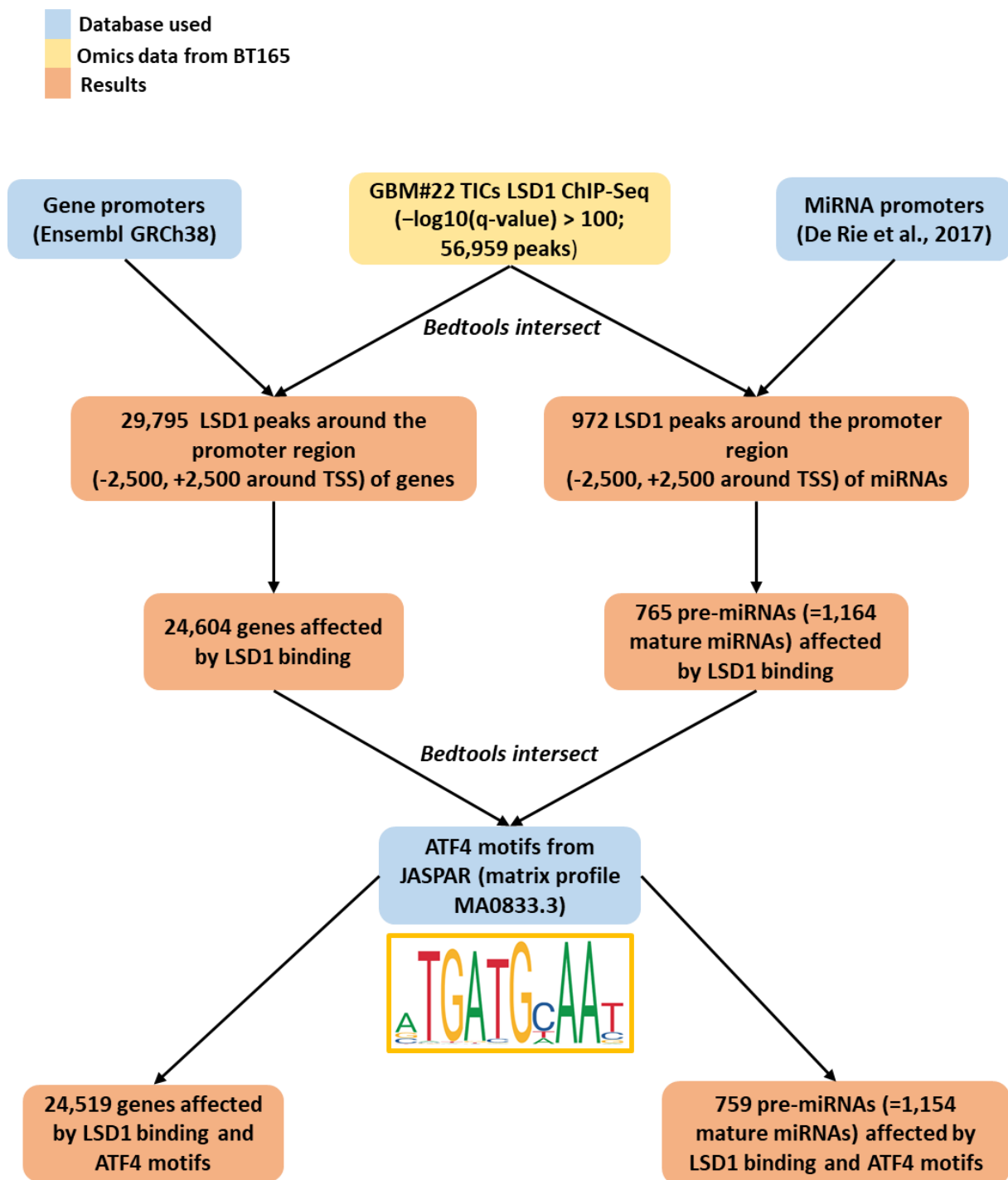


Figure 6.7. Schematic representation of the integration of LSD1 ChIP-Seq and ATF4 motif analysis at gene and miRNA promoters. The blue rectangles represent the promoter regions of genes and pre-miRNAs, including the ATF4-binding motifs predicted by JASPAR. The yellow rectangles indicate LSD1-binding peaks detected by ChIP-Seq in the BT165 cell line. The orange rectangles indicate the genes and miRNAs characterized by LSD1 peaks and ATF4 motifs in their promoter regions.

6.2.2 Transcriptomic analysis of differentially expressed genes following LSD1 silencing in BT165

To investigate transcriptomic changes following LSD1 silencing in the BT165 cell line, we analyzed the mRNA-Seq dataset published by Faletti et al., 2021, using hg38 as the reference genome. After quantification, only genes with TPM expression > 1 in at least one sample were considered, resulting in a total of 16,941 expressed genes. Differential analysis identified 95 DEGs between LSD1-silenced and controls. Of these, 74 genes exhibit LSD1 binding peaks in their promoter regions, while 93 genes exhibit the presence of ATF4 motifs. In particular, the 74 genes with LSD1 peaks also exhibit ATF4 motifs with distinct scores: 10 genes with a low score, 37 with a medium score, and 27 with a high score. Table 6.2 reports the list of the 95 DEGs with the annotation for the presence or absence of LSD1 peaks and ATF4 motifs (low, medium or high) in their promoter regions, while Figure 6.8 shows the unsupervised hierarchical clustering of the 95 DEGs, characterized by the annotation of the presence of LSD1 and ATF4 motifs.

DEGs	LSD1	ATF4
ENSG00000004487_KDM1A	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000006283_CACNA1G	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000008282_SYPL1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000008283_CYB561	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000018625_ATP1A2	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH
ENSG00000041982_TNC	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000063854_HAGH	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000064300_NGFR	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000064666_CNN2	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000064886_CHI3L2	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000064989_CALCRL	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000069482_GAL	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000070669_ASNS	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000072682_P4HA2	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000089169_RPH3A	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000091972_CD200	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000100095_SEZ6L	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000100889_PCK2	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000101255_TRIB3	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000102158_MAGT1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000111845_PAK1IP1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000111981_ULBP1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000113356_POLR3G	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000113732_ATP6V0E1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000114209_PDCD10	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM

ENSG00000116990_MYCL	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000117054_ACADM	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000122203_KIAA1191	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000125355_TMEM255A	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000128923_MINDY2	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000128965_CHAC1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000130513_GDF15	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000130818_ZNF426	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000135046_ANXA1	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH
ENSG00000135069_PSAT1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000135535_CD164	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000136770_DNAJC1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000136874_STX17	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000138944_SHISAL1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000139269_INHBE	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000139324_TMTCC3	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000139433_GLTP	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000140105_WARS1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000140287_HDC	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000150893_FREM2	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000151012_SLC7A11	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000152377_SPOCK1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000153214_TMCM87B	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000159658_EFCAB14	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000162692_VCAM1	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000163577_EIF5A2	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000164106_SCRG1	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH
ENSG00000165169_DYNLT3	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000165323_FAT3	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000167315_ACAA2	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000169241_SLC50A1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000172340_SUCLG2	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000175197_DDIT3	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000175582_RAB6A	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000176018_LYSMD3	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000178741_COX5A	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000180287_PLD5	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000180573_H2AC6	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000181019_NQO1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000182118_FAM89A	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000185947_ZNF267	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000186063_AIDA	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000186951_PPARA	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000187735_TCEA1	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000197008_ZNF138	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000198521_ZNF43	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000198692_EIF1AY	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH

ENSG00000199673_SNORD16	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000201654_RNU6-7	LSD1_ABSENCE	JASPAR_PRESENCE_LOW
ENSG00000202337_RNU6-8	LSD1_ABSENCE	JASPAR_PRESENCE_LOW
ENSG00000205133_TRIQK	LSD1_PRESENCE	JASPAR_PRESENCE_HIGH
ENSG00000229509	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000235513_L3MBTL2-AS1	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH
ENSG00000250091_DNAH10OS	LSD1_ABSENCE	JASPAR_PRESENCE_LOW
ENSG00000256879_PDE3A-AS1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000257365_FNTB	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000257379	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000258800	LSD1_ABSENCE	JASPAR_PRESENCE_HIGH
ENSG00000259976	LSD1_ABSENCE	JASPAR_PRESENCE_LOW
ENSG00000260335	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000267127	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000270249	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000271153_RPL23AP88	LSD1_ABSENCE	JASPAR_PRESENCE_LOW
ENSG00000275140_SEC22B3P	LSD1_ABSENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000275342_PRAG1	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000288905_LSP1P5	LSD1_PRESENCE	JASPAR_PRESENCE_LOW
ENSG00000290226	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000291080_ZNF252P	LSD1_PRESENCE	JASPAR_PRESENCE_MEDIUM
ENSG00000293516_ZNF271P	LSD1_ABSENCE	JASPAR_ABSENCE
ENSG00000305485	LSD1_ABSENCE	JASPAR_ABSENCE

Table 6.2. List of differentially expressed genes identified in LSD1-silenced vs control in BT165 with LSD1 and ATF4 annotation. The first column reports the name of each DEG, the second column indicates the presence or absence of LSD1 binding peaks in the corresponding promoter regions obtained from LSD1 ChIP-Seq, and the third column indicates the presence or absence of ATF4 motifs obtained from JASPAR. The 74 DEGs characterized by LSD1 peaks and ATF4 motifs in their promoter regions are written in red.



Figure 6.8. Heatmap showing the unsupervised hierarchical clustering of 95 DEGs with LSD1 and ATF4 annotations. The two experimental groups, LSD1 silenced (LSD1 shRNA) and control (shNT), are indicated in orange and cyan, respectively. For each gene, the annotation related to the presence (green) or absence (red) of LSD1 binding peaks in the promoter region is reported. The annotation related to ATF4 is also shown, indicating the score of the binding motif: low, medium, and high, represented by different shades of yellow, while absence is indicated in purple.

6.2.3 Identification of miR-124-3p as an upstream regulator of DEGs

The 95 DEGs between LSD1-silenced and controls were used as input for a Core Analysis of IPA, thus obtaining miR-124-3p as an upstream regulator (and other miRNAs with AAGGCAC seed). The forms of this mature miRNA were explored and only two precursors (hsa-mir-124-1 and hsa-mir-124-3) are characterized by the presence of LSD1 peaks and ATF4 motifs in their promoter region. In particular, in Figure 6.9 the promoter regions of these pre-miRNAs are shown on the UCSC Genome Browser.

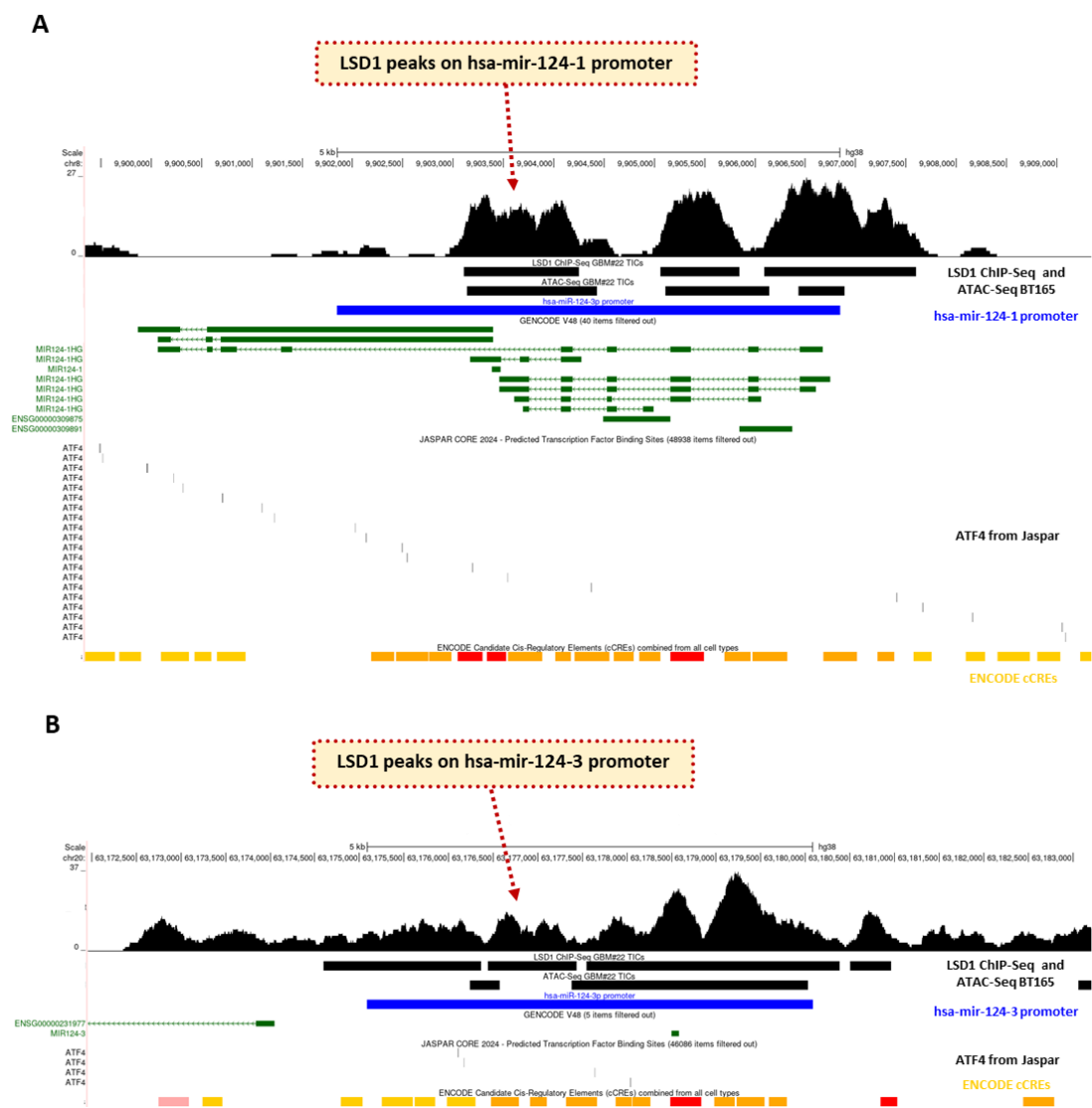


Figure 6.9. Schematic representation of the promoter regions of hsa-mir-124-1 and hsa-mir-124-3 in UCSC Genome Browser. Visualization of the promoter regions of the precursors hsa-mir-124-1 (A) and hsa-mir-124-3 (B) using UCSC Genome Browser. In both panels are shown LSD1 binding peaks derived from ChIP-seq in BT165 (black track), ATAC-Seq (black track), the promoter region of pre-miRNAs (blue track), the ATF4 motifs identified by JASPAR, and the cCREs annotated by the ENCODE consortium.

6.2.4 LSD1-Dependent regulation of miR-124-3p and its target genes

The analysis performed with IPA has revealed that miR-124-3p acts as an upstream regulator of genes differentially expressed between LSD1-silenced cells and controls. We therefore focused on identifying the target genes of this miRNA to define an LSD1-mediated regulatory network.

Specifically, the target genes of miR-124-3p were obtained by intersecting miRTarBase and TargetScan databases. For miRTarBase, we selected only the miR-124-3p target genes with interactions classified as non “weak”, obtaining a total of 1,474 target genes. For TargetScan, instead, the entire database was used without filters, including 1,820 target genes for the same miRNA.

The integration of the two databases allowed the identification of 518 common target genes of miR-124-3p. Among these, only the genes expressed in the mRNA-seq of Faletti et al. 2021 were selected with TPM >1 in at least one sample, resulting in a final list of 468 target genes (Figure 6.10A).

Among these 468 genes, 12 were differentially expressed between LSD1-silenced and controls and also exhibited peaks of LSD1 and the ATF4 motif in their promoter region. Figure 6.10B shows a heatmap of these 12 DEGs, annotated with LSD1 and ATF4, which were all downregulated following LSD1 silencing.

Finally, this result was represented through a regulatory network (Figure 6.10C), constructed with Cytoscape. The network shows that LSD1 regulates miR-124-3p, which in turn downregulates the 12 DEGs, which exhibit both LSD1 and the ATF4 motif in their promoter region.

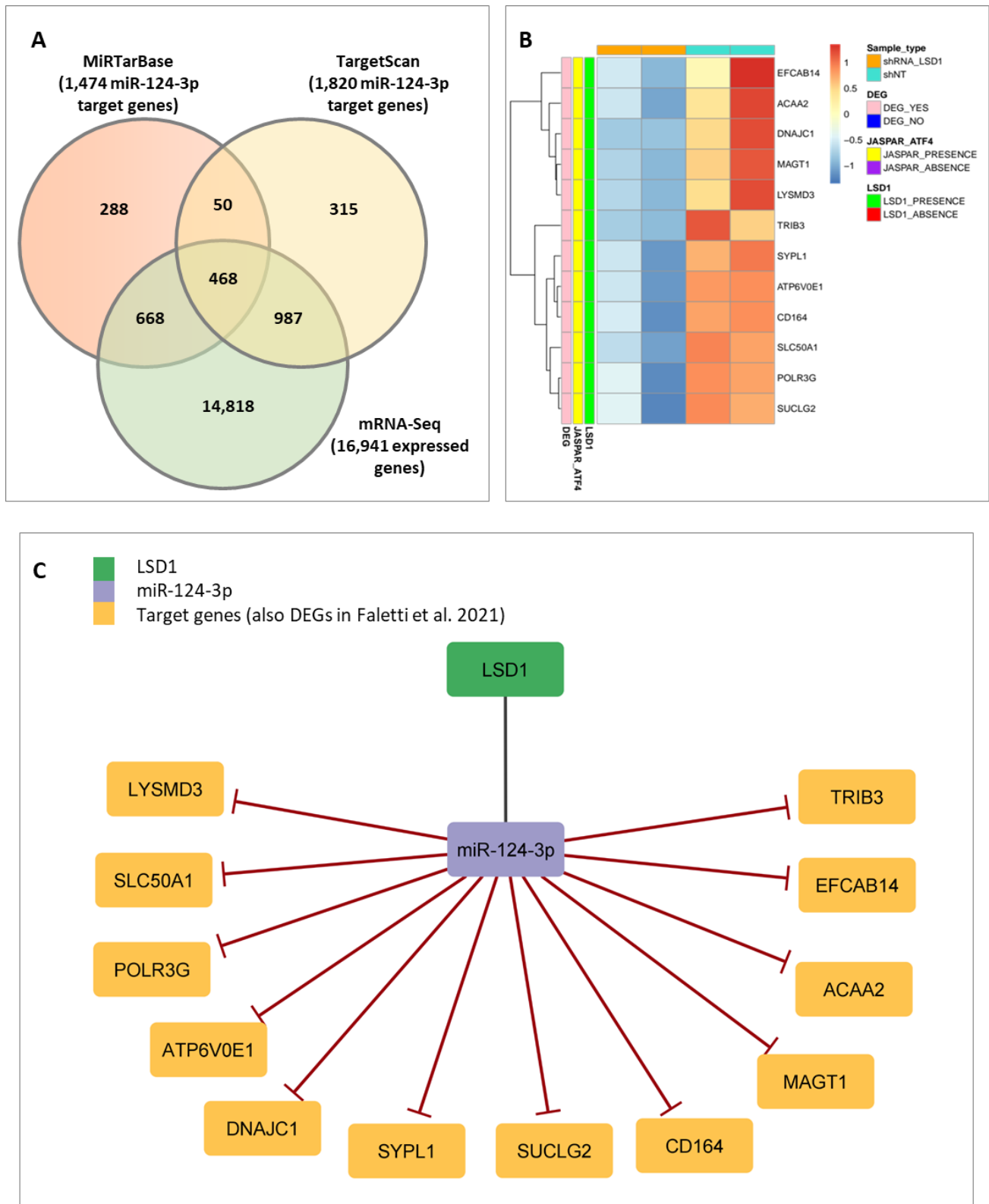


Figure 6.10. LSD1-dependent regulatory network involving miR-124-3p and its target genes. (A) Venn diagram showing target genes shared between miRTarBase and TargetScan and genes expressed in mRNA-seq from Faletti et al. 2021. (B) Heatmap showing the hierarchical clustering of the 12 DEGs between LSD1-silenced (LSD1 shRNA, orange) and controls (shNT, cyan). Genes are annotated as DEGs (pink) or non-DEGs (blue), based on the presence of LSD1 (green) or absence (red), and based on the presence of the ATF4 motif (yellow) or absence (purple). (C) LSD1-ATF4-miRNA mediated regulatory network mediated of miR-124-3p. LSD1 is highlighted with a green rectangle, the miRNA in purple, and the target genes in orange.

6.2.5 Evidence from Wang et al. 2021 shows miR-124-3p downregulation in GBM

Since miRNA-Seq data are not available for the BT165 cell model, to assess miR-124-3p expression in GBM, we downloaded a miRNA-Seq dataset published by Wang et al., 2021. Specifically, in this study, the authors identified five clusters of 50 miRNAs each, differentially expressed compared to normal tissues and miR-124-3p was also included among these.

Figure 6.11 shows the expression level of miR-124-3p in GBM samples compared to normal tissues analyzed in the study by Wang et al., 2021. The comparison between the two groups, performed using the Wilcoxon test, shows a significant downregulation of miR-124-3p in tumor samples (p-value < 0.0001).

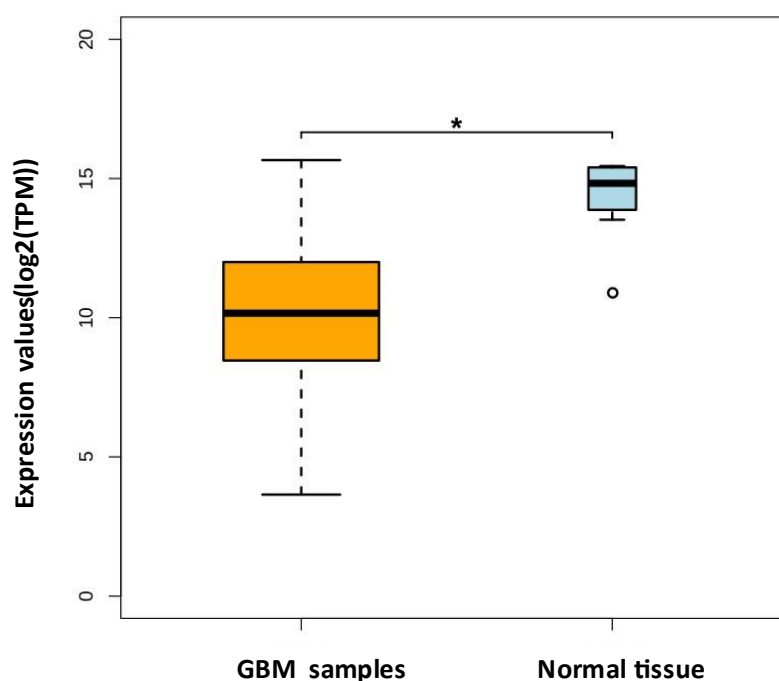


Figure 6.11. MiR-124-3p expression in glioblastoma samples and normal tissues. MiR-124-3p expression levels, measured in $\log_2(\text{TPM})$, in GBM samples compared to normal control tissues. The orange boxplot represents GBM samples, while the light blue boxplot represents normal tissues. The data show a significant difference in expression between the two groups (p-value < 0.0001).

This finding is consistent with what was observed in our cell model. In particular, Figure 6.9B shows a heatmap in which, following LSD1 silencing via shRNA, target genes are downregulated compared to controls. This strengthens the hypothesis that miR-124-3p is normally downregulated in GBM,

thus allowing the overexpression of its target genes. Conversely, LSD1 inhibition leads to increased miR-124-3p expression, resulting in the repression of its targets.

In summary, we hypothesize that LSD1 inhibition promotes the transcriptional activation of miR-124-3p, while under physiological conditions, LSD1 contributes to its repression (Figure 6.12).

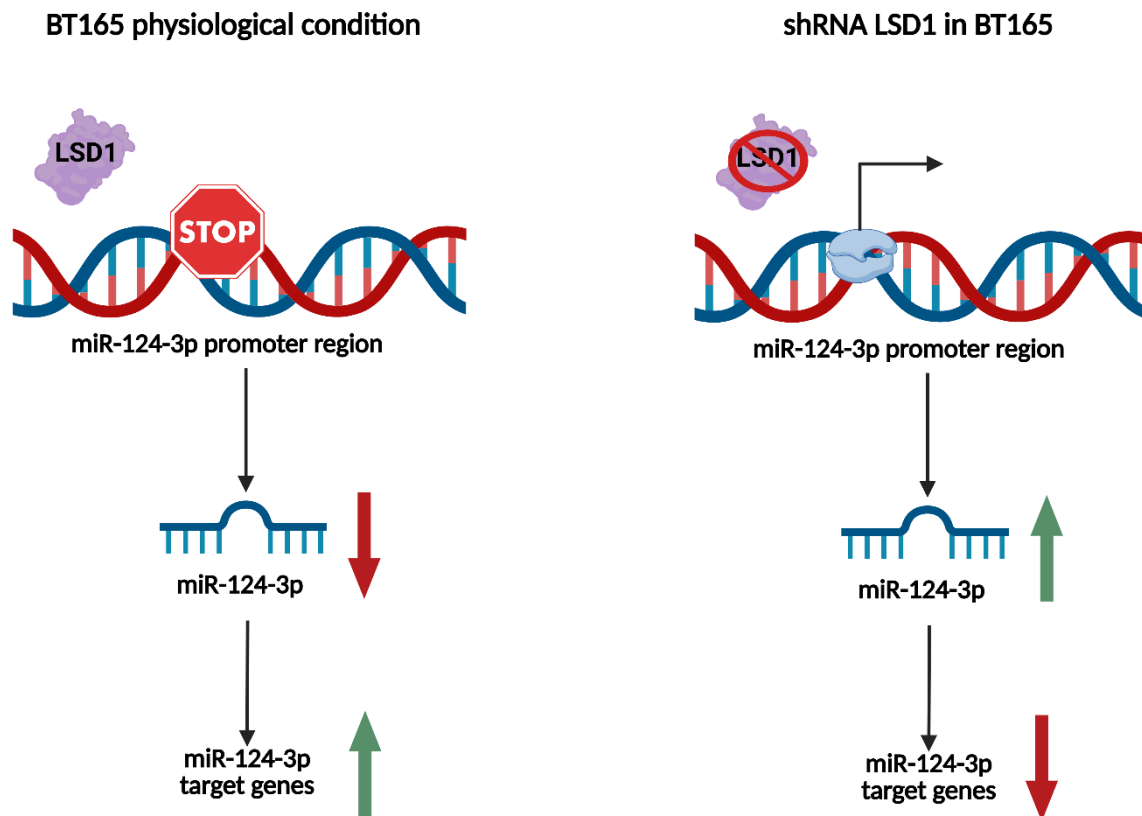


Figure 6.12. Hypothetical model of the role of LSD1 in the transcriptional regulation of miR-124-3p in GBM. According to our hypothesis, under physiological conditions in BT165 cells, LSD1 (purple) represses the transcription of miR-124-3p. Consequently, miR-124-3p is downregulated and unable to inhibit the expression of its target genes, which are therefore upregulated. Conversely, when LSD1 is silenced via shRNA, miR-124-3p is transcribed and acts to repress the expression of its target genes. Image created with BioRender.

6.2.6 Multi-omics analysis of the Wang et al., 2021 dataset

In the second phase of the second project, due to the lack of miRNA-Seq and circRNA-Seq data for the BT165 cell line used in the first phase, mRNA-Seq, miRNA-Seq, and circRNA-Seq data published by Wang et al. (2021) and obtained from GBM patient samples were analyzed.

To explore the molecular mechanisms specific to the proneural subtype, corresponding to the model represented by the BT165 cell line, only GBM samples belonging to this subtype were selected, resulting in a total of 29 tumor samples.

For the mRNA-Seq and miRNA-Seq datasets, the availability of eight normal tissue samples as controls made it possible to perform a differential expression analysis, which identified 6069 DEGs and 124 DEMs.

Regarding the circRNA-Seq data, due to the absence of control samples, a differential expression analysis could not be performed. Therefore, circRNAs expressed with FPKM > 1 in at least one sample and annotated in the CSCD2 database were considered, yielding a total of 1888 circRNAs.

Subsequently, a Spearman anticorrelation analysis was conducted between DEMs and DEGs, considering as significant those pairs with a correlation coefficient < -0.6. This analysis identified 39 DEM–DEG interactions, corresponding to 18 DEMs and 37 DEGs (Table 6.3).

Based on the 18 DEMs identified, circRNAs potentially capable of targeting (sponging) these miRNAs were subsequently searched for, resulting in 213 circRNA–DEM interactions, corresponding to 204 circRNAs interacting with 15 of the 18 DEMs.

Finally, to assess the potential epigenetic involvement of LSD1, LSD1 ChIP-Seq data obtained from the BT165 cell line (Faletti et al., 2021) were analyzed to verify the presence of binding peaks in the promoter regions of the identified DEGs and DEMs ($\pm 2,500$ bp from the TSS).

The analysis revealed that 13 of the 18 DEMs and 20 of the 37 DEGs showed LSD1 binding peaks in their respective promoter regions (Table 6.4 and Table 6.5).

The overall analysis workflow is illustrated in Figure 6.13.

DEMs	DEGs
hsa-miR-139-5p	MAML2
hsa-miR-128-3p	MAML2
hsa-miR-218-5p	SOX11
hsa-miR-106b-5p	NFAM1
hsa-miR-10a-5p	CD180
hsa-miR-491-5p	WNK4
hsa-miR-10a-5p	TREM2

hsa-miR-16-2-3p	NAT2
hsa-miR-16-2-3p	STEAP3
hsa-miR-16-2-3p	ANO6
hsa-miR-10a-5p	ITGB2
hsa-miR-128-3p	ZNF850
hsa-miR-106b-5p	PTGER4
hsa-miR-138-1-3p	MPZ
hsa-miR-224-5p	TP53INP1
hsa-miR-128-3p	YBX1
hsa-miR-16-2-3p	CCL20
hsa-miR-130b-5p	STEAP2
hsa-miR-210-3p	IGSF6
hsa-miR-10a-5p	IGSF6
hsa-miR-183-3p	CALB1
hsa-miR-15b-3p	ITGBL1
hsa-miR-106b-5p	SLC2A5
hsa-miR-10a-5p	KCNE3
hsa-miR-106b-5p	CYBB
hsa-miR-106b-5p	GPR183
hsa-miR-485-5p	MYCN
hsa-miR-10a-5p	LILRA2
hsa-miR-4325	RGS4
hsa-miR-106b-5p	IER3
hsa-miR-106b-5p	CSF2RA
hsa-miR-106b-5p	NCF2
hsa-miR-7977	DLGAP3
hsa-miR-106b-5p	GBP2
hsa-miR-128-3p	SIX1
hsa-miR-6842-3p	GYLTL1B
hsa-miR-218-5p	XRCC2
hsa-miR-196a-5p	TMEM154
hsa-miR-218-5p	MEX3A

Table 6.3. List of 39 DEM–DEG interactions. The first column represents the DEMs targeting the corresponding DEGs shown in the second column.

DEMs	LSD1
hsa-miR-106b-5p	LSD1_PRESENCE
hsa-miR-10a-5p	LSD1_PRESENCE
hsa-miR-128-3p	LSD1_PRESENCE
hsa-miR-130b-5p	LSD1_PRESENCE
hsa-miR-138-1-3p	LSD1_PRESENCE
hsa-miR-139-5p	LSD1_PRESENCE
hsa-miR-15b-3p	LSD1_PRESENCE
hsa-miR-16-2-3p	LSD1_PRESENCE
hsa-miR-183-3p	LSD1_PRESENCE
hsa-miR-196a-5p	LSD1_PRESENCE
hsa-miR-210-3p	LSD1_PRESENCE

hsa-miR-218-5p	LSD1_PRESENCE
hsa-miR-224-5p	LSD1_ABSENCE
hsa-miR-4325	LSD1_ABSENCE
hsa-miR-485-5p	LSD1_ABSENCE
hsa-miR-491-5p	LSD1_PRESENCE
hsa-miR-6842-3p	LSD1_ABSENCE
hsa-miR-7977	LSD1_ABSENCE

Table 6.4. List of 18 DEMs found in the DEM-DEG interactions with LSD1 annotation.

DEGs	LSD1
ANO6	LSD1_PRESENCE
CALB1	LSD1_ABSENCE
CCL20	LSD1_ABSENCE
CD180	LSD1_ABSENCE
CSF2RA	LSD1_ABSENCE
CYBB	LSD1_ABSENCE
DLGAP3	LSD1_PRESENCE
GBP2	LSD1_PRESENCE
GPR183	LSD1_ABSENCE
GYLTL1B	LSD1_ABSENCE
IER3	LSD1_PRESENCE
IGSF6	LSD1_ABSENCE
ITGB2	LSD1_PRESENCE
ITGBL1	LSD1_ABSENCE
KCNE3	LSD1_ABSENCE
LILRA2	LSD1_ABSENCE
MAML2	LSD1_PRESENCE
MEX3A	LSD1_PRESENCE
MPZ	LSD1_ABSENCE
MYCN	LSD1_PRESENCE
NAT2	LSD1_ABSENCE
NCF2	LSD1_ABSENCE
NFAM1	LSD1_PRESENCE
PTGER4	LSD1_PRESENCE
RGS4	LSD1_ABSENCE
SIX1	LSD1_PRESENCE
SLC2A5	LSD1_PRESENCE
SOX11	LSD1_PRESENCE
STEAP2	LSD1_PRESENCE
STEAP3	LSD1_PRESENCE
TMEM154	LSD1_PRESENCE
TP53INP1	LSD1_PRESENCE
TREM2	LSD1_ABSENCE
WNK4	LSD1_ABSENCE
XRCC2	LSD1_PRESENCE
YBX1	LSD1_PRESENCE

Table 6.5. List of 37 DEGs found in the DEM-DEG interactions with LSD1 annotation

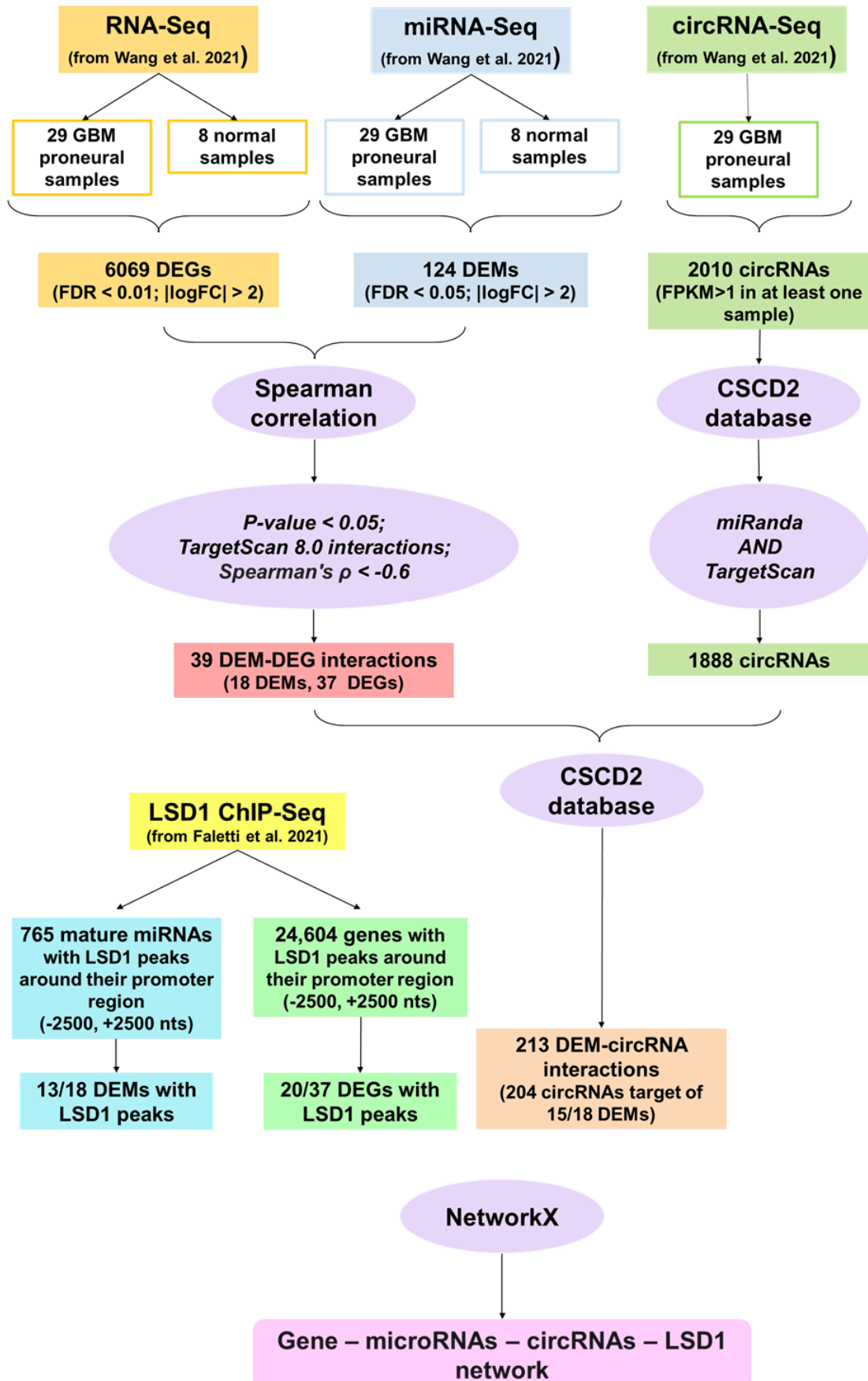
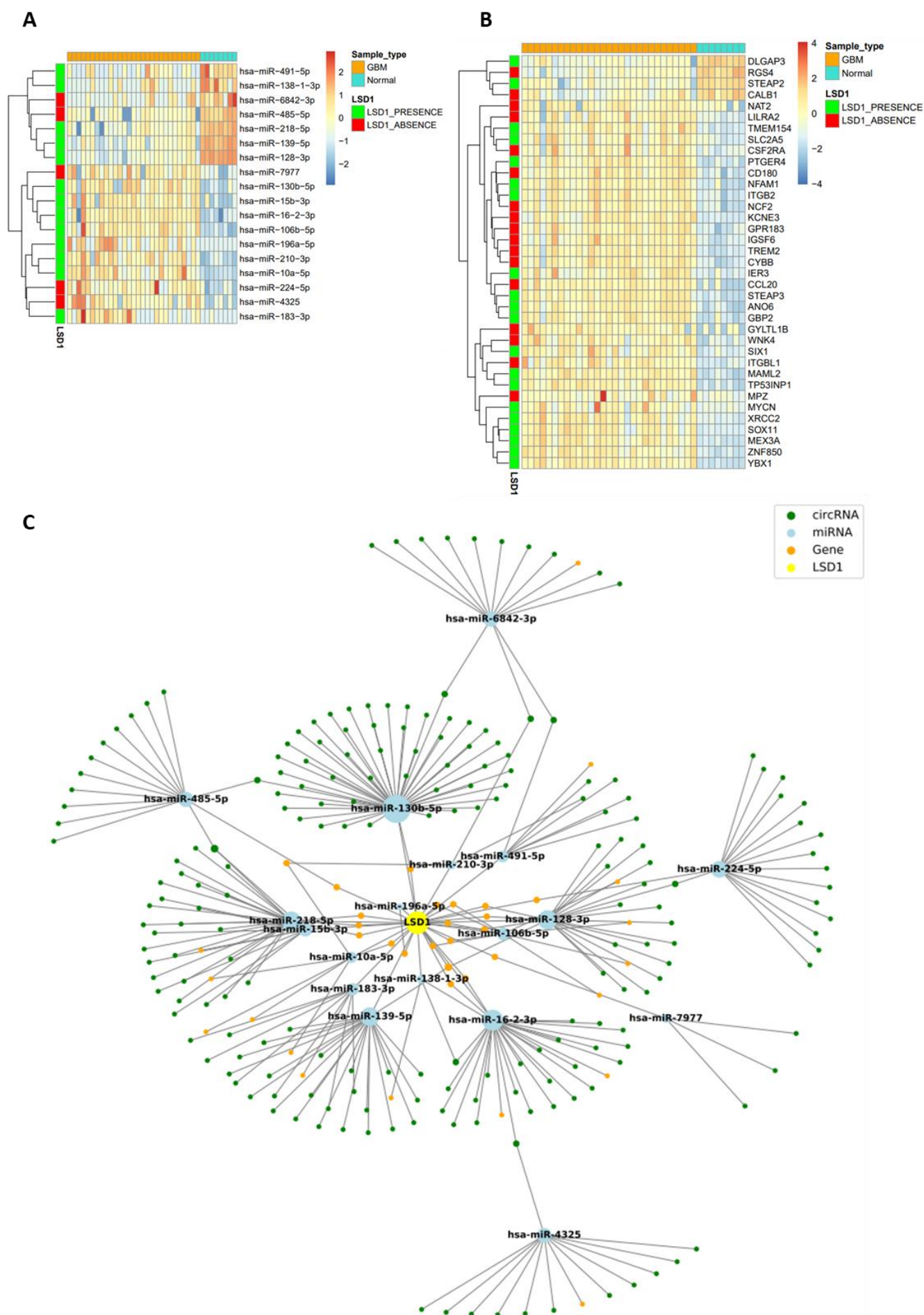


Figure 6.13. Schematic representation of the workflow used for the data analysis of Wang et al., 2021. The yellow rectangles represent the mRNA-Seq data from which 6069 DEGs were identified, the light blue rectangles indicate the miRNA-Seq data and the 124 DEMs, the green rectangles show the circRNA-Seq data and the circRNAs obtained by integrating the dataset with the CSCD database represented in purple. The Spearman anticorrelation analysis in purple allowed to identify 39 DEM–DEG interactions (red rectangle). The LSD1 ChIP-Seq data from Faletti et al., 2021 is represented in yellow, the light blue indicates the DEMs with LSD1 peaks and the green the DEGs with LSD1 peaks. The graphical integration of all this information allowed to obtain the gene–miRNA–circRNA–LSD1 regulatory network (pink).

Heatmaps depicting the 18 DEMs and 37 DEGs differentially expressed between GBM and control tissues, annotated with the presence of LSD1 peaks, are shown in Figures 6.14A and 6.14B, respectively.

The final regulatory network, integrating the 39 DEM–DEG interactions, the 204 circRNA–DEM interactions, and the LSD1–DEM and LSD1–DEG connections, was constructed using the NetworkX package in Python and is presented in Figure 6.14C.

Figure 6.14. Analysis of LSD1-centered interactions among DEMs and DEGs in GBM and normal tissues. (A) Heatmap of the 18 DEMs and (B) of the 37 DEGs identified between glioblastoma (GBM, in orange) and normal tissue samples (in cyan), involved in the 39 DEM–DEG interactions. For each DEM and DEG, the presence of LSD1 in their promoter regions is indicated in green, while its absence is shown in red. (C) LSD1-centered regulatory network illustrating the DEM–DEG, circRNA–DEM, LSD1–DEM, and LSD1–DEG interactions. LSD1 is shown in yellow, circRNAs in green, miRNAs in blue, and genes in orange.



7. DISCUSSION

7.1 Discussion of the endothelial regulatory network project

TFEB is a transcriptional regulator that plays a key role in autophagy, lysosomal processes, and the maintenance of endothelial homeostasis. Specifically, its activity contributes to the stability and function of the endothelium, which are essential for maintaining vascular integrity. In this first project of the thesis, we explored the molecular mechanisms through which TFEB exerts its transcriptional function in endothelial cells, using HUVEC cells as an experimental model.

To achieve this goal, we developed an integrative approach based on bioinformatics analyses of high-throughput sequencing data, including ChIP-Seq, mRNA-Seq, and miRNA-Seq. Specifically, the study of gene expression and miRNA profiles in TFEB-silenced HUVECs allowed us to systematically delineate the regulatory network associated with this transcription factor, highlighting the specific contribution of miRNAs in modulating its activities. Our result not only provides a useful resource for future functional analyses, but also highlights how interactions between transcription factors and miRNA-mediated regulatory circuits constitute a crucial level of molecular control in endothelial cells. The data obtained demonstrates the broad involvement of miRNAs in modulating TFEB function, confirming observations already reported in different endothelial models.^{103,104,105} In this context, our study contributes to expanding our knowledge of the impact of noncoding RNAs in regulating endothelial function, with particular reference to TFEB activity.

Our results revealed overlap between TFEB ChIP-Seq data from HUVECs and a set of annotated promoter regions in the genome, including both mRNA-coding genes and pri-miRNAs. Significant TFEB binding peaks were found in the promoters of genes and miRNAs differentially expressed following TFEB silencing, indicating the factor's direct involvement in their transcriptional regulation. However, the TSS catalog used does not represent a complete and up-to-date repertoire of promoter regions; therefore, it is plausible that more recent versions may reveal additional regulatory interactions mediated by TFEB.¹⁰⁶ This is particularly relevant for miRNAs, which are known to use alternative promoters to those close to the TSS of the pri-miRNAs considered in this study.¹⁰⁷

Another aspect concerns the distribution of binding peaks with respect to the regulatory elements identified by the ENCODE consortium. As shown in Figure 7.1, only a limited fraction of TFEB peaks coincide with promoter elements, while the majority are associated with distal enhancers. This finding suggests that the regulatory network modulated by TFEB in endothelial cells is much more extensive than that evidenced by promoters alone, and that distal or indirect interactions may play a crucial role, especially under conditions of factor silencing.

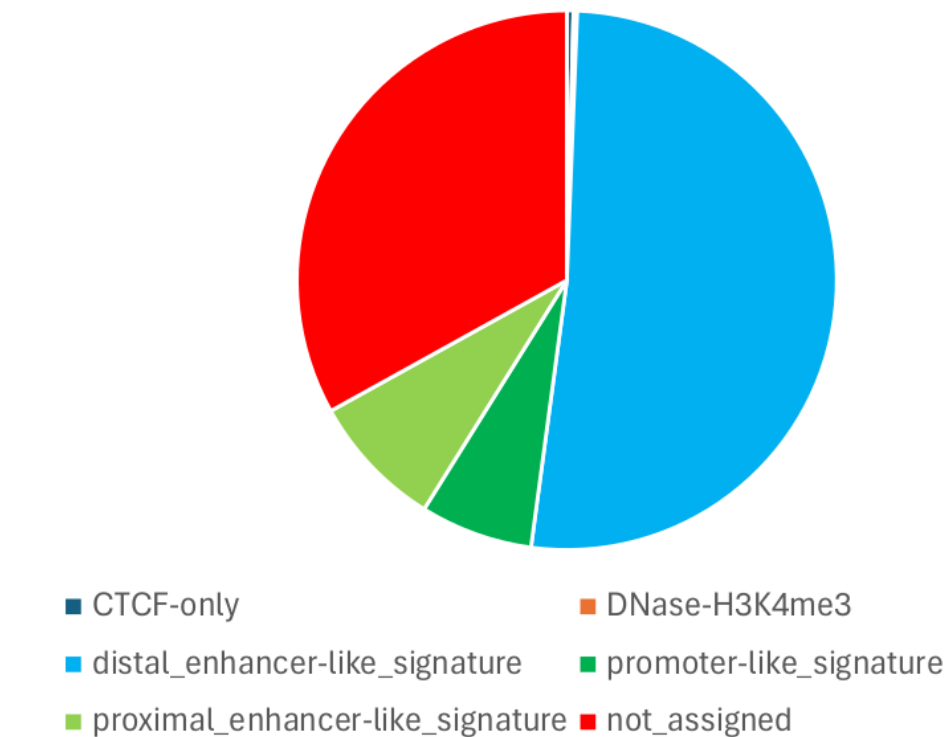


Figure 7.1. Overlap between TFEB peaks in HUVECs and ENCODE cCREs. The pie chart shows the overlap between the catalog of TFEB peaks identified in human endothelial cells (HUVECs) and candidate cis-regulatory elements (cCREs) in the human genome, as annotated by the ENCODE project. The percentages of overlaps between TFEB peaks and each class of cCREs are also reported, considering only cases where the overlap exceeds 100 nucleotides.

Figure 7.2 clarifies the possible regulatory mechanisms suggested by the distribution of TFEB binding peaks. Panel A illustrates promoter-based regulation, in which TFEB binds to promoter regions of protein-coding genes and miRNAs, consistent with direct transcriptional control. Panel B depicts a putative enhancer-mediated regulatory mechanism, in which TFEB binding to distal enhancer elements may influence target gene expression through chromatin looping.

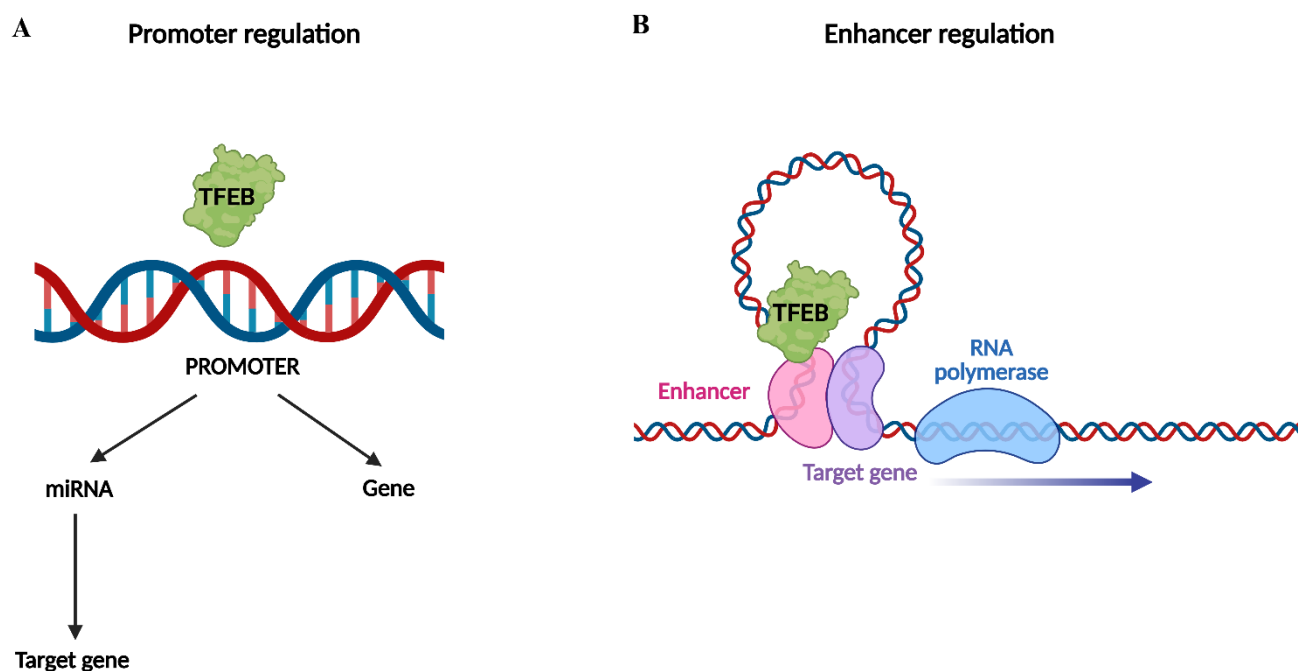


Figure 7.2. Promoter-based and putative enhancer-mediated regulation by TFEB in endothelial cells. (A) TFEB (green) binding to promoter regions of protein-coding genes and miRNAs, consistent with direct transcriptional regulation. (B) TFEB binding to distal enhancer elements (pink) that may regulate target gene expression (violet) through chromatin looping. (Image created with BioRender).

A final observation concerns the fact that a portion of the identified genes and miRNAs appear to be negatively regulated by TFEB. This possibility is supported by previous evidence, including a direct example of negative regulation of the MYO1C gene in HUVECs.⁵⁰ However, the presence of repressed targets could also result from indirect mechanisms or the action of specific transcriptional cofactors capable of modulating TFEB activity.

In the final phase of this first project, we focused on the seven differentially expressed miRNAs (DEMs) that exhibit TFEB binding peaks in their promoter regions following TFEB silencing: miR-181a-3p, miR-181b-3p, miR-29b-1-5p, miR-222-5p, miR-27a-5p, miR-193a-3p, and miR-339-3p. These miRNAs play critical roles in various endothelial cell (EC) functions and vascular biology, including vascular inflammation, angiogenesis, endothelial integrity, and vascular remodeling. Below, we summarize the specific contribution of each miRNA to vascular health based on current literature:

- miR-181a-3p decreases the progression of atherosclerosis by modulating vascular inflammation. In ECs, it exerts an anti-inflammatory effect by downregulating adhesion molecules such as ICAM-1 and VCAM-1, as well as the NF-κB essential modulator (NEMO);¹⁰⁸
- miR-339-3p is upregulated in vascular smooth muscle cells (VSMCs) in response to angiotensin II type 1 receptor autoantibodies (AT1-AA), contributing to vascular inflammation;¹⁰⁹
- miR-181b-3p is overexpressed in chronic obstructive pulmonary disease (COPD) and is associated with pulmonary endothelial dysfunction. In vitro, its overexpression in HUVECs reduces endothelial sprouting, impairing angiogenesis;¹¹⁰
- miR-29b-1-5p is downregulated in damaged endometrial stromal cells (ESCs) co-cultured with Wharton's jelly mesenchymal stem cells (WJ-MSCs), promoting angiogenesis through upregulation of RAPB1;¹¹¹
- miR-222-5p is known to inhibit the angiogenic activity of stem cell factor (SCF) by targeting c-Kit.⁹⁹ It is also upregulated in VSMCs treated with oxidized LDL, contributing to atherosclerosis development;¹¹²
- miR-27a-5p is reduced in exosomes derived from Tsp-1-expressing microglia, leading to increased Smad3 expression in ECs and reduced retinal neovascularization, suggesting a protective role in maintaining vascular homeostasis and preventing pathological angiogenesis;¹¹³
- miR-193a-3p is implicated in the pathogenic mechanisms of vascular conditions such as obstructive sleep apnea (OSA), aortic dissection (AD), and angiogenesis, regulating endothelial functions in HUVECs, VSMCs, and circulating endothelial colony-forming cells (ECFCs).^{114,115,116}

Table 7.1 recapitulates the functions and mechanisms/targets of the seven miRNAs discussed in this study.

miRNA	Function	Mechanism / Target
miR-181a-3p	Anti-inflammatory; slows atherosclerosis progression	Reduces ICAM-1, VCAM-1, and NEMO expression in ECs
miR-339-3p	Promotes vascular inflammation	Upregulated in VSMCs in response to AT1-AA

miR-181b-3p	Impairs angiogenesis in pulmonary endothelium	Reduces endothelial sprouting in HUVECs; associated with COPD
miR-29b-1-5p	Promotes angiogenesis	Upregulation of RAPB1 in ESCs co-cultured with WJ-MSCs
miR-222-5p	Anti-angiogenic; contributes to atherosclerosis	Targets c-Kit; upregulated in VSMCs treated with ox-LDL
miR-27a-5p	Maintains vascular homeostasis; suppresses pathological angiogenesis	Downregulated in Tsp-1 microglia exosomes; increases Smad3 in ECs
miR-193a-3p	Regulates endothelial and VSMC functions; involved in vascular pathology	Influences HUVECs, VSMCs, and circulating ECFCs; implicated in OSA and aortic dissection

Table 7.1. Summary of the seven differentially expressed miRNAs (DEMs) with TFEB promoter binding peaks identified in this study, highlighting their functions in vascular biology and known molecular mechanisms or targets.

Out of the 7 DEMs identified, we directed our attention to miR-222 and its downstream targets, E2F1 and CDK1, because they are implicated in a particularly specific function of TFEB in endothelial cells: the control of the transition from the G1 phase to the S phase of the cell cycle.⁵⁰ Following TFEB silencing, we observed an increase in both the primary transcript and mature miR-222, while E2F1 and CDK1 levels decreased. This pattern strongly suggests a direct regulatory link between TFEB and miR-222 in HUVECs, highlighting a previously unrecognized mechanism by which TFEB influences endothelial cell behavior (Figure 6.5).

The subset of miRNAs and genes depicted in Figure 6.4a represent only a fraction of those altered after TFEB silencing. To explore the potential functional implications, we applied ToppGene to identify enriched biological processes, thus revealing an enrichment for pathways related to cell cycle progression and DNA replication, reinforcing the role of these genes in proliferative control. Although TFEB ChIP-Seq experiments have been conducted in other cell types, including HeLa, human EndoC- β H1, and HEK293 lines,^{45,46,117,118,119} prior studies have not thoroughly examined TFEB's capacity to regulate miRNAs. Our study, therefore, constitutes the first comprehensive mapping of a TFEB-driven miRNA regulatory network in HUVECs, unveiling a novel perspective on how TFEB orchestrates endothelial function through non-coding RNA-mediated mechanisms.

7.2 Discussion of the glioblastoma regulatory network project

In the second project of this thesis, we investigated the role of the epigenetic regulator LSD1 and its ATF4 cofactor in glioblastoma tumor-initiating cells (GBM TICs), using BT165 as cellular model, representing an example of proneural subtype. LSD1 is a well-established epigenetic regulator in GBM and has been shown to contribute to the maintenance of stem-like properties in GBM TICs, supporting self-renewal and tumorigenic potential. Faletti et al., 2021 demonstrated that ATF4 acts as a cofactor for LSD1, suggesting that the LSD1–ATF4 complex may orchestrate transcriptional networks critical for GBM TIC biology.⁷¹

The integration of LSD1 ChIP-Seq and mRNA-Seq data from GBM#22 TICs allowed the identification of thousands of LSD1-bound genomic regions, including promoters of genes and pri-miRNAs. Motif analysis revealed that the majority of the LSD1-bound promoters also contained ATF4 motifs, supporting a model in which LSD1 and ATF4 cooperate to regulate gene transcription. Differential expression analysis in LSD1-silenced versus control, followed by Ingenuity Pathway Analysis (IPA), highlighted miR-124-3p as a putative upstream regulator. Only hsa-mir-124-1 and hsa-mir-124-3 displayed both LSD1 and ATF4 binding at their promoter regions, indicating a direct regulatory relationship.

MiR-124-3p is widely recognized in literature as a tumor suppressor in GBM, and it is generally downregulated in tumor tissues compared to normal brain, as reported in several recent works. Experimental studies in vitro and in vivo have shown that its overexpression in GBM cell lines (including U87, U251, T98G, and C6) reduces proliferation, migration, invasiveness, angiogenesis, and stemness traits. Notably, regulation of targets such as CDK6, PPPR13L, SNAI2, RhoG, NRP-1, SDCBP, and EPHA2 underscores its impact on key pathways controlling cell proliferation, cell cycle, motility, and stem-like features.^{120,121,122,123,124,125,126,127} Additional epigenetic mechanisms, including ALKBH5-mediated m6A modification, further modulate stemness via inhibition of EPHA2, suggesting novel regulatory elements that could be therapeutically exploited.¹²⁶

Beyond molecular effects, clinical evidence indicates that miR-124-3p is also reduced systemically, as observed in circulating exosomes from patients with high-grade gliomas, supporting its relevance as a non-invasive biomarker for disease diagnosis and monitoring. Moreover, its overexpression increases the sensitivity of GBM cells to temozolomide, by downregulating DNA repair genes and potentially enhancing standard treatment efficacy.¹²⁷ In general, miR-124-3p has a central role in controlling proliferation, stemness, and therapeutic response in GBM, and suggest that modulation

of this microRNA, through epigenetic or pharmacological strategies, may represent a promising approach to limit tumor progression and improve treatment outcomes.

Our results highlighted the role of miR-124-3p as a tumor suppressor downstream of LSD1. In GBM#22 TICs under basal conditions (without LSD1 silencing), the predicted target genes of miR-124-3p (LYSMD3, SLC50A1, POLR3G, ATP6V0E1, DNAJC1, SYPL1, SUCLG2, CD164, MAGT1, ACAA2, EFCAB14, and TRIB3) are upregulated, supporting the hypothesis that miR-124-3p is downregulated in these cells. Upon LSD1 silencing, miR-124-3p is upregulated, leading to downregulation of these target genes, suggesting that LSD1 may suppress miR-124-3p expression in GBM TICs.

The 12 target genes of miR-124-3p participate in cell cycle regulation, transcriptional control, and maintenance of stem cell properties, indicating that the LSD1–miR-124-3p axis plays a key role in the characteristics of tumor-initiating cells in glioblastoma. Among these, DNAJC1, SYPL1, CD164, MAGT1, and TRIB3 are overexpressed in GBM and promote cell proliferation, invasiveness, survival, and resistance to therapies, constituting potential biomarkers and specific therapeutic targets.^{128,129,130,131,132} The other targets have been implicated in several tumors: LYSMD3 in papillary thyroid cancer,¹³³ SLC50A1 in breast cancer,^{134,135} POLR3G in lung, liver, and bladder cancers,^{136,137,138} ATP6V0E1 in esophageal cancer,^{139,140} SUCLG2 in thyroid and prostate cancers,^{141,142} and ACAA2 in ovarian cancer.¹⁴³ EFCAB14, however, has no relevant evidence in oncology. These data suggest that, despite originating from different tumor contexts, upregulation of these genes supports general pro-tumor functions, such as proliferation, metabolism, and progression. A summary of the oncogenic landscape of the 12 predicted targets of miR-124-3p is presented in Table 7.2.

Target gene	Associated cancer(s)	Role/Function
DNAJC1	Glioblastoma	Upregulated; promotes proliferation, migration, epithelial-to-mesenchymal transition, extracellular matrix remodeling, and macrophage infiltration; poor survival ¹²⁸
SYPL1	Glioblastoma	Upregulated; regulates extracellular vesicle biogenesis and trafficking, microenvironment plasticity ¹²⁹

CD164	Glioblastoma	Upregulated; enhances proliferation, migration, invasiveness via Akt/Beclin-1 and autophagy regulation ¹³⁰
MAGT1	Glioma	Upregulated; promotes growth and radioresistance, activates ERK/MAPK and PD-L1 → immune evasion ¹³¹
TRIB3	Glioblastoma	Upregulated; inhibits autophagy, favoring tumor survival and therapy resistance ¹³²
LYSMD3	Papillary thyroid carcinoma	Overexpressed; poor prognosis; LysM-domain protein involved in recognition and intracellular trafficking ¹³³
SLC50A1	Breast cancer	Overexpressed; promotes metabolism, growth; poor outcome ^{134,135}
POLR3G	NSCLC, hepatocellular carcinoma, bladder cancer	Overexpressed; promotes epithelial-to-mesenchymal transition, proliferation, invasiveness, metastasis, regulates Wnt/MAPK pathways ^{136,137,138}
ATP6V0E1	Esophageal squamous cell carcinoma	Upregulated via NAT10-mediated ac4C; enhances lysosomal acidification, E-cadherin degradation, epithelial-to-mesenchymal transition, metastasis ^{139,140}
SUCLG2	Thyroid carcinoma, prostate carcinoma	Overexpressed; modulates TCA cycle and lipid metabolism; poor prognosis ^{141,142}
ACAA2	Ovarian carcinoma	Overexpressed; regulates fatty acid metabolism, promotes proliferation, migration, invasion ¹⁴³
EFCAB14	Not reported	No relevant oncological evidence available

Table 7.2. Summary of miR-124-3p predicted target genes and their cancer-related functions.

To further validate the relevance of miR-124-3p in GBM, we analyzed patient-derived miRNA-Seq data from Wang et al., 2021.⁶⁴ Despite being derived from bulk GBM tissue rather than TICs, this analysis revealed significant downregulation of miR-124-3p in GBM samples compared to normal tissue, particularly within the proneural subtype, which corresponds to the molecular identity BT165. While these data provide additional support for our model, differences in cellular composition and context between patient samples and TICs represent a limitation of this comparison.

Importantly, our study is based on integrative bioinformatics analysis combining ChIP-Seq, mRNA-Seq, and miRNA target prediction, and no experimental validation of miR-124-3p targets was performed. Nevertheless, the identification of consistent regulatory patterns and differentially expressed genes provides a strong rationale for future functional validation, which will be crucial to confirm the predicted LSD1–ATF4–miR-124-3p network in GBM TICs. Future perspectives include generating a miRNA-Seq dataset specifically on BT165, allowing a direct assessment of miR-124-3p expression and activity of its predicted targets in the same cellular context, thereby strengthening the mechanistic link between LSD1, ATF4, and miR-124-3p.

Overall, our study strengthens the hypothesis that LSD1 inhibition can restore miR-124-3p levels, reducing the expression of pro-tumor genes and modulating the properties of tumor-initiating cells in GBM and that non-coding RNA are involved in GBM tumorigenesis. This finding led us to hypothesize that also other miRNAs, as well as circRNAs acting as sponges, might also play important roles in GBM, and that exploring their expression and interactions could uncover novel regulatory mechanisms. Since miRNA-Seq and circRNA-Seq data were not available for the BT165 cell line, we downloaded Wang et al., 2021 omics data to understand the contribution of non-coding RNAs to GBM pathogenesis. Given that BT165 represents an example of proneural subtype of GBM, we specifically selected patient-derived samples belonging to the same molecular subtype to maintain biological consistency with our experimental model.

The analysis of Wang et al., 2021 data allowed the identification of a substantial number of DEMs and a diverse set of expressed circRNAs. Thanks to the integration of these omics data with the LSD1 ChIP-Seq in BT165 a preliminary but informative regulatory network encompassing miRNA-mRNA anticorrelations, circRNA-miRNAs interactions and epigenetic regulation by LSD1 was created.

This integrative approach revealed that many miRNAs and circRNAs may be involved in GBM, and several of them show potential regulation by LSD1, either directly or through intermediary RNA layers. These results support the hypothesis that non-coding RNAs, including miRNAs and circRNAs,

could contribute to tumorigenic pathways in GBM and may be integrated into epigenetic regulatory programs orchestrated by LSD1.

However, a key limitation of this second phase is the lack of data homogeneity. The different sources of data coming from GBM patient samples rather than the BT165 cell line may introduce biological variability, and the absence of circRNA controls precluded proper differential expression analysis. Therefore, the findings from this phase should be interpreted as preliminary and hypothetical rather than conclusive.

Future studies should aim to validate these findings directly in the BT165 model through the generation of matched miRNA-Seq and circRNA-Seq datasets. This would allow a more coherent and solid investigation of the epigenetic post-transcriptional regulatory axis involving LSD1, miRNAs, and circRNAs in GBM.

Overall, this analysis provides a valuable framework for future research and supports the idea that miRNA and circRNA deregulation may cooperate with epigenetic dysregulation to sustain GBM pathogenesis, particularly within the proneural subtype.

8. CONCLUSION

In conclusion, the results obtained in the two main projects of this thesis underscore the importance of a systematic and uniform approach based on the integration of multi-omics data for the study of regulatory networks. In the case of TFEB in HUVEC endothelial cells, the reconstruction of the regulatory network highlighted how this transcription factor controls not only protein-coding genes but also specific miRNAs, such as miR-222, involved in the regulation of cell cycle progression and endothelial activity. Similarly, in the study of LSD1 in GBM TICs, the integration of ChIP-Seq and RNA-Seq data identified miR-124-3p as a key regulator of the LSD1–ATF4 axis, suggesting a functional role for miRNAs in modulating transcriptional programs and promoting stem-like and tumorigenic properties.

Although available information on circular RNAs is still limited, their integration into the network analyzed in the second project suggests a potential regulatory role, opening new avenues for future studies. Overall, the work highlights how short non-coding RNAs, and miRNAs, represent key elements in modulating the functions of transcription factors and complex pathological processes. The developed bioinformatics pipeline is a powerful and versatile tool for analyzing regulatory networks, useful for identifying new therapeutic targets and biomarkers, with potential translational applications in oncology and vascular diseases.

9. ADDITIONAL WORKS

During these 3 years of PhD, in addition to the two main projects described in the previous chapters, additional research studies were conducted.

One study analyzed the role of diacylglycerol kinases (DGKs) in acute myeloid leukemia (AML), with the aim of understanding the molecular mechanisms regulating disease progression and evaluating their potential as therapeutic targets.¹⁴⁴

Another study explored the dynamics of the gut microbiota in a mouse model of Alzheimer's disease, using a metagenomic approach to characterize the functional changes associated with the pathology.¹⁴⁵

9.1 Bioinformatic investigation of diacylglycerol kinases in acute myeloid leukemia

This study aimed to understand the role of diacylglycerol kinases (DGKs), a family of ten isoenzymes involved in regulating lipid signaling, in acute myeloid leukemia (AML). The primary objective was to identify isoforms with potential biological and clinical relevance, using public transcriptomic data available on IPA.

The bioinformatic approach included the analysis of three datasets available via IPA: TCGA-LAML, BeatAML, and TARGET. These data allowed to compare the expression levels of the different DGK isoforms within the AML patient population and assess their potential association with prognosis. To this end, Kaplan–Meier survival curves were generated, comparing patients with high, low, or normal expression for each isoform.

Analysis of TCGA-LAML data showed that high DGKA expression correlates with a reduction in overall survival, suggesting an unfavorable role in disease progression. Conversely, high expression of the DGKG isoform is associated with a better prognosis. Other isoforms, such as DGKE, DGKD, and DGKZ, showed no significant associations with survival. Isoforms such as DGKH and DGKQ, which are not overexpressed in TCGA-LAML, showed similar, non-significant effects (Figure 9.1).

Analysis of independent cohorts confirmed some trends but also highlighted differences: in the BeatAML dataset, high DGKH expression was associated with reduced survival, while in the TARGET dataset, which primarily includes pediatric patients with higher survival rates, a negative correlation was observed between survival and high DGKH levels.

To further investigate the molecular mechanisms, tools such as cBioPortal and Metascape were used to analyze gene co-expression networks and identify pathways related to each isoform. The analysis highlighted isoform-specific functions: DGKA is associated with immune signaling genes,

DGKD with vesicle trafficking and autophagy pathways, while DGKG and DGKZ are related to carbohydrate metabolism and neutrophil activity. Cell enrichment analysis suggested opposing roles of the isoforms in myeloid differentiation, with DGKA bound to activated T lymphocytes and DGKG to CD34+ hematopoietic stem cells.

These results support the idea that DGKs exert isoform-specific functions in leukemia, influencing both cellular transformation and differentiation fate. The identification of DGKA and DGKG as opposing prognostic markers highlights the clinical relevance of these isoforms and the need for targeted therapeutic approaches.

Laboratory experiments on AML cell lines (HL-60 and HEL) confirmed the functional importance of DGKs, suggesting that simultaneous inhibition of multiple isoforms may be more effective than selective targeting of DGKA.

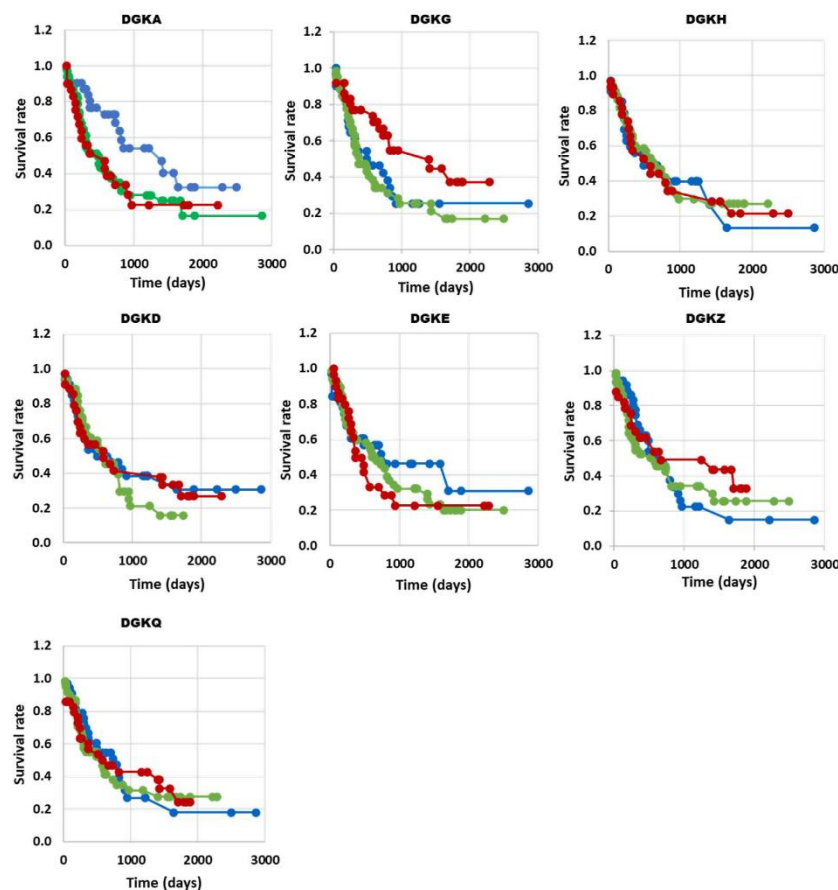


Figure 9.1. Correlation between DGK isoform expression and overall survival of AML patients in the TCGA dataset. Patients were divided into three groups based on the expression levels of each isoform: high (red), normal (green), and low (blue). For the DGKA isoform, 26 patients showed high expression, 50 medium expression, and 30 low expression; the association with overall survival was significant ($p = 0.04$), indicating a negative impact of high expression. Similarly, for DGKG, 30 patients had high expression, 48 medium, and 27 low, with a significant correlation favoring survival ($p = 0.03$). The other isoforms did not show statistically significant associations with survival: DGKH (25 high, 50 medium, 29 low; $p = 0.99$), DGKD (28 high, 47 medium, 31 low; $p = 0.76$), DGKE (28 high, 53 medium, 26 low; $p = 0.51$), DGKZ (23 high, 45 medium, 31 low; $p = 0.55$), and DGKQ (28 high, 47 medium, 29 low; $p = 0.92$). Image adapted from Gravina et al., 2023.

9.2 Metagenomic analysis of gut microbiota alterations across disease progression in an Alzheimer's disease mouse model

This study explored changes in the gut microbiota in a mouse model of Alzheimer's disease (AD) using a next-generation metagenomic approach based on clade-specific taxonomic markers. The hypothesis is that the gut-brain axis may contribute to disease progression in neurodegenerative processes or represent a biological marker.

The 3xTgAD transgenic mouse model, which reproduces several neuropathological and behavioral features typical of human Alzheimer's disease, was used, comparing it to healthy wild-type mice at three different developmental time points (2, 6, and 12 months), corresponding to distinct phases of disease progression. The mice's feces were subjected to shotgun metagenomic sequencing and analyzed with MetaPhlAn 4 software, a tool that accurately quantifies the composition of microbial communities through clade-specific genetic markers, distinguishing bacterial genomes at the species (SGBs), family (FGBs), order (OGBs), and class (CGBs) levels. Specifically, the study separated known SGBs (kSGBs) from those still unclassified (uSGBs), with the aim of assessing how much of the gut microbiota in mice with Alzheimer's disease is still unknown and potentially relevant to understanding the disease.

The results reveal that the composition of the gut microbiota in 3xTgAD mice undergoes significant changes with age and in relation to the presence of the disease. The main variations are concentrated in the *Bacteroidetes* and *Firmicutes* phyla, which constitute the dominant groups in the gut microbial community. These changes involve both the relative abundance of specific known taxa and the emergence or disappearance of numerous unclassified taxa, the so-called uSGBs. One of the most notable observations is the large number of unknown microbial species that show different patterns in Alzheimer's-affected mice compared to controls, suggesting that a significant portion of the microbial variability occurring in the disease belongs to clades not yet described in the taxonomic literature. From a temporal perspective, the study highlights that changes in the microbial community are not static but evolve over the course of the disease, with a progressive remodeling of the microbiota that accompanies aging. In 3xTgAD mice, the changes become more marked in the more advanced stages (12 months), when neurodegeneration and cognitive deficits are most evident.

Figure 9.2 illustrates these dynamics, showing the changes in the relative abundance of the main microbial taxa in 3xTgAD mice compared to wild-type controls at the three observation time points. In particular, the *Bacteroidetes* and *Firmicutes* phyla exhibit opposing trends between groups, and

age amplifies the differences in microbial composition. This visual representation reinforces the idea of a progressive microbiota imbalance associated with both aging and the disease, suggesting a functional link between gut dysbiosis and the progression of Alzheimer's symptoms.

These findings confirm the involvement of the gut microbiota in the pathophysiology of Alzheimer's, highlighting that gut bacteria may help modulate inflammatory and immune processes known to play a role in disease progression. The observation of specific microbial patterns related to age and genotype opens the possibility of using certain taxa as early biomarkers or indicators of Alzheimer's disease progression, as well as developing strategies to modulate the microbiota, for example through diet, prebiotics, or probiotics, for therapeutic or preventive purposes.

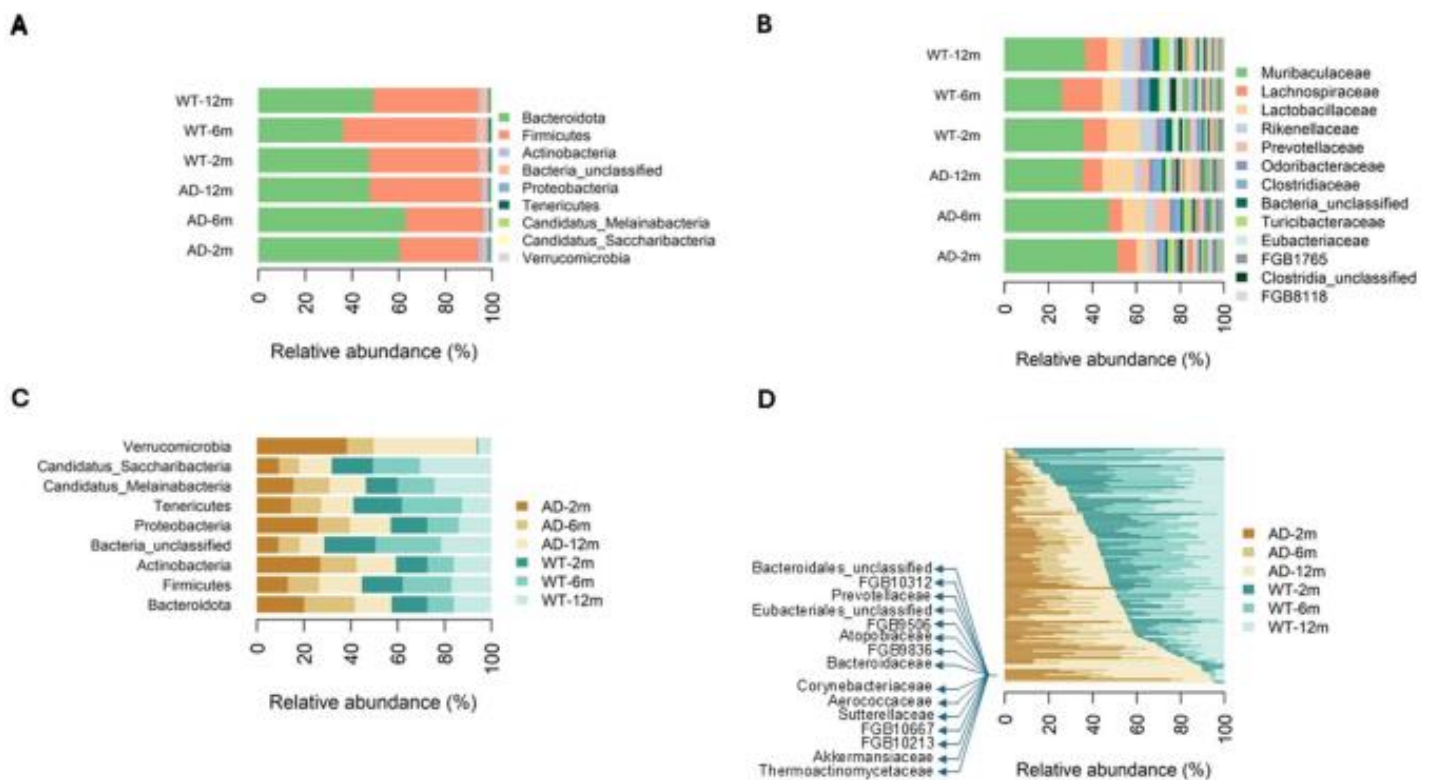


Figure 9.2. Top expressed taxa in each experimental group. A) Barplot representing the average relative abundance for phylum taxonomic level in AD microbiomes vs. their time-matched control counterparts. B) Barplot showing the average relative abundance for family taxonomic level in AD microbiomes vs. their time-matched control counterparts. C) Barplot representing each detected phylum in AD microbiomes and time-matched control. D) Barplot representing each detected family in AD microbiomes and time-matched control. Image adapted from Favero et al., 2024.

10. BIBLIOGRAPHY

1. Chan, S.Y., and Loscalzo, J. (2012). The Emerging Paradigm of Network Medicine in the Study of Human Disease. *Circulation Research* 111, 359–374. <https://doi.org/10.1161/CIRCRESAHA.111.258541>.
2. Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12, 56–68. <https://doi.org/10.1038/nrg2918>.
3. Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., et al. (2009). An empirical framework for binary interactome mapping. *Nat Methods* 6, 83–90. <https://doi.org/10.1038/nmeth.1280>.
4. Loscalzo, J., and Barabasi, A. (2011). Systems biology and the future of medicine. *WIREs Mechanisms of Disease* 3, 619–627. <https://doi.org/10.1002/wsbm.144>.
5. Cheng, C., Yan, K.-K., Hwang, W., Qian, J., Bhardwaj, N., Rozowsky, J., Lu, Z.J., Niu, W., Alves, P., Kato, M., et al. (2011). Construction and Analysis of an Integrated Regulatory Network Derived from High-Throughput Sequencing Data. *PLoS Comput Biol* 7, e1002190. <https://doi.org/10.1371/journal.pcbi.1002190>.
6. Somers, J., Fenner, M., Kong, G., Thirumalaisamy, D., Yashar, W.M., Thapa, K., Kinali, M., Nikolova, O., Babur, Ö., and Demir, E. (2023). A framework for considering prior information in network-based approaches to omics data analysis. *Proteomics* 23, 2200402. <https://doi.org/10.1002/pmic.202200402>.
7. Recabarren, D., and Alarcón, M. (2017). Gene networks in neurodegenerative disorders. *Life Sciences* 183, 83–97. <https://doi.org/10.1016/j.lfs.2017.06.009>.
8. Ruffini, N., Klingenberg, S., Heese, R., Schweiger, S., and Gerber, S. (2022). The Big Picture of Neurodegeneration: A Meta Study to Extract the Essential Evidence on Neurodegenerative Diseases in a Network-Based Approach. *Front. Aging Neurosci.* 14, 866886. <https://doi.org/10.3389/fnagi.2022.866886>.
9. Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat Methods* 10, 1108–1115. <https://doi.org/10.1038/nmeth.2651>.

10. Li, R., Jiang, S., Li, W., Hong, H., Zhao, C., Huang, X., Zhang, Z., Li, H., Chen, H., and Bo, X. (2019). Exploration of prognosis-related microRNA and transcription factor co-regulatory networks across cancer types. *RNA Biology* 16, 1010–1021. <https://doi.org/10.1080/15476286.2019.1607714>.
11. Le, T.D., Liu, L., Zhang, J., Liu, B., and Li, J. (2015). From miRNA regulation to miRNA-TF co-regulation: computational approaches and challenges. *Briefings in Bioinformatics* 16, 475–496. <https://doi.org/10.1093/bib/bbu023>.
12. Cora', D., Re, A., Caselle, M., and Bussolino, F. (2017). MicroRNA-mediated regulatory circuits: outlook and perspectives. *Phys. Biol.* 14, 045001. <https://doi.org/10.1088/1478-3975/aa6f21>.
13. Guzzi, P.H., Di Martino, M.T., Tagliaferri, P., Tassone, P., and Cannataro, M. (2015). Analysis of miRNA, mRNA, and TF interactions through network-based methods. *J Bioinform Sys Biology* 2015, 4. <https://doi.org/10.1186/s13637-015-0023-8>.
14. Zhang, H.-M., Kuang, S., Xiong, X., Gao, T., Liu, C., and Guo, A.-Y. (2015). Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics* 16, 45–58. <https://doi.org/10.1093/bib/bbt085>.
15. Veitia, R.A. (2025). Rethinking transcription factor dynamics and transcription regulation in eukaryotes. *Trends in Biochemical Sciences* 50, 376–384. <https://doi.org/10.1016/j.tibs.2025.01.012>.
16. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* 172, 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>.
17. Islam, Z., Ali, A.M., Naik, A., Eldaw, M., Decock, J., and Kolatkar, P.R. (2021). Transcription Factors: The Fulcrum Between Cell Development and Carcinogenesis. *Front. Oncol.* 11, 681377. <https://doi.org/10.3389/fonc.2021.681377>.
18. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* 52, D174–D182. <https://doi.org/10.1093/nar/gkad1059>.

19. The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>.
20. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research* 48, D882–D889. <https://doi.org/10.1093/nar/gkz1062>.
21. Hasan, A., Khan, N.A., Uddin, S., Khan, A.Q., and Steinhoff, M. (2024). Deregulated transcription factors in the emerging cancer hallmarks. *Seminars in Cancer Biology* 98, 31–50. <https://doi.org/10.1016/j.semcancer.2023.12.001>.
22. Bushweller, J.H. (2019). Targeting transcription factors in cancer — from undruggable to reality. *Nat Rev Cancer* 19, 611–624. <https://doi.org/10.1038/s41568-019-0196-7>.
23. Mundade, R., Ozer, H.G., Wei, H., Prabhu, L., and Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* 13, 2847–2852. <https://doi.org/10.4161/15384101.2014.949201>.
24. Nakato, R., and Sakata, T. (2021). Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* 187, 44–53. <https://doi.org/10.1016/j.ymeth.2020.03.005>.
25. Panni, S., Lovering, R.C., Porras, P., and Orchard, S. (2020). Non-coding RNA regulatory networks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1863, 194417. <https://doi.org/10.1016/j.bbagrm.2019.194417>.
26. Loganathan, T., and Doss C, G.P. (2023). Non-coding RNAs in human health and disease: potential function as biomarkers and therapeutic targets. *Funct Integr Genomics* 23, 33. <https://doi.org/10.1007/s10142-022-00947-4>.
27. Yan, H., and Bu, P. (2021). Non-coding RNA in cancer. *Essays in Biochemistry* 65, 625–639. <https://doi.org/10.1042/EBC20200032>.
28. Ahir, B.K., Ozer, H., Engelhard, H.H., and Lakka, S.S. (2017). MicroRNAs in glioblastoma pathogenesis and therapy: A comprehensive review. *Critical Reviews in Oncology/Hematology* 120, 22–33. <https://doi.org/10.1016/j.critrevonc.2017.10.003>.

29. Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 15, 509–524. <https://doi.org/10.1038/nrm3838>.
30. Kim, H., Lee, Y.-Y., and Kim, V.N. (2025). The biogenesis and regulation of animal microRNAs. *Nat Rev Mol Cell Biol* 26, 276–296. <https://doi.org/10.1038/s41580-024-00805-0>.
31. Shang, R., Lee, S., Senavirathne, G., and Lai, E.C. (2023). microRNAs in action: biogenesis, function and regulation. *Nat Rev Genet* 24, 816–833. <https://doi.org/10.1038/s41576-023-00611-y>.
32. Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.-L., Yang, L., and Chen, L.-L. (2016). The Biogenesis of Nascent Circular RNAs. *Cell Reports* 15, 611–624. <https://doi.org/10.1016/j.celrep.2016.03.058>.
33. Mumtaz, P.T., Taban, Q., Dar, M.A., Mir, S., Haq, Z.U., Zargar, S.M., Shah, R.A., and Ahmad, S.M. (2020). Deep Insights in Circular RNAs: from biogenesis to therapeutics. *Biol Proced Online* 22, 10. <https://doi.org/10.1186/s12575-020-00122-8>.
34. Yang, L., Wilusz, J.E., and Chen, L.-L. (2022). Biogenesis and Regulatory Roles of Circular RNAs. *Annu. Rev. Cell Dev. Biol.* 38, 263–289. <https://doi.org/10.1146/annurev-cellbio-120420-125117>.
35. Malviya, A., and Bhuyan, R. (2023). The recent advancements in circRNA research: From biogenesis to therapeutic interventions. *Pathology - Research and Practice* 248, 154697. <https://doi.org/10.1016/j.prp.2023.154697>.
36. Pisignano, G., Michael, D.C., Visal, T.H., Pirlog, R., Lodomery, M., and Calin, G.A. (2023). Going circular: history, present, and future of circRNAs in cancer. *Oncogene* 42, 2783–2800. <https://doi.org/10.1038/s41388-023-02780-w>.
37. Conn, V.M., Chinnaiyan, A.M., and Conn, S.J. (2024). Circular RNA in cancer. *Nat Rev Cancer* 24, 597–613. <https://doi.org/10.1038/s41568-024-00721-7>.
38. Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D., and Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. *J Hematol Oncol* 13, 166. <https://doi.org/10.1186/s13045-020-01005-x>.

39. Medina-Leyte, D.J., Domínguez-Pérez, M., Mercado, I., Villarreal-Molina, M.T., and Jacobo-Albavera, L. (2020). Use of Human Umbilical Vein Endothelial Cells (HUVEC) as a Model to Study Cardiovascular Disease: A Review. *Applied Sciences* 10, 938. <https://doi.org/10.3390/app10030938>.
40. Duranova, H., Kuzelova, L., Borotova, P., Simora, V., and Fialkova, V. (2024). Human Umbilical Vein Endothelial Cells as a Versatile Cellular Model System in Diverse Experimental Paradigms: An Ultrastructural Perspective. *Microscopy and Microanalysis* 30, 419–439. <https://doi.org/10.1093/mam/ozae048>.
41. Gravina, T., Favero, F., Rosano, S., Parab, S., Diaz Alcalde, A., Bussolino, F., Doronzo, G., and Corà, D. (2024). Integrative Bioinformatics Analysis Reveals a Transcription Factor EB-Driven MicroRNA Regulatory Network in Endothelial Cells. *IJMS* 25, 7123. <https://doi.org/10.3390/ijms25137123>.
42. Napolitano, G., and Ballabio, A. (2016). TFEB at a glance. *Journal of Cell Science* 129, 2475–2481. <https://doi.org/10.1242/jcs.146365>.
43. Takla, M., Keshri, S., and Rubinsztein, D.C. (2023). The post-translational regulation of transcription factor EB (TFEB) in health and disease. *EMBO Reports* 24. <https://doi.org/10.15252/embr.202357574>.
44. Zhang, L., Li, Z., Zhang, L., Qin, Y., and Yu, D. (2023). Dissecting the multifaced function of transcription factor EB (TFEB) in human diseases: From molecular mechanism to pharmacological modulation. *Biochemical Pharmacology* 215, 115698. <https://doi.org/10.1016/j.bcp.2023.115698>.
45. Sardiello, M., Palmieri, M., Di Ronza, A., Medina, D.L., Valenza, M., Gennarino, V.A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R.S., et al. (2009). A Gene Network Regulating Lysosomal Biogenesis and Function. *Science* 325, 473–477. <https://doi.org/10.1126/science.1174447>.
46. Palmieri, M., Impey, S., Kang, H., Di Ronza, A., Pelz, C., Sardiello, M., and Ballabio, A. (2011). Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Human Molecular Genetics* 20, 3852–3866. <https://doi.org/10.1093/hmg/ddr306>.
47. Settembre, C., Di Malta, C., Polito, V.A., Arencibia, M.G., Vetrini, F., Erdin, S., Erdin, S.U., Huynh, T., Medina, D., Colella, P., et al. (2011). TFEB Links Autophagy to Lysosomal Biogenesis. *Science* 332, 1429–1433. <https://doi.org/10.1126/science.1204592>.

48. Ariano, C., Riganti, C., Corà, D., Valdembri, D., Mana, G., Astanina, E., Serini, G., Bussolino, F., and Doronzo, G. (2022). TFEB controls integrin-mediated endothelial cell adhesion by the regulation of cholesterol metabolism. *Angiogenesis* 25, 471–492. <https://doi.org/10.1007/s10456-022-09840-x>.
49. Astanina, E., Bussolino, F., and Doronzo, G. (2022). Transcription factor EB controls both motogenic and mitogenic cell activities. *FEBS Letters* 596, 1973–1980. <https://doi.org/10.1002/1873-3468.14442>.
50. Doronzo, G., Astanina, E., Corà, D., Chiabotto, G., Comunanza, V., Noghero, A., Neri, F., Puliafito, A., Primo, L., Spanpanato, C., et al. (2019). TFEB controls vascular development by regulating the proliferation of endothelial cells. *The EMBO Journal* 38. <https://doi.org/10.15252/emboj.201798250>.
51. Agostini, F., Agostinis, R., Medina, D.L., Bisaglia, M., Greggio, E., and Plotegher, N. (2022). The Regulation of MiTF/TFE Transcription Factors Across Model Organisms: from Brain Physiology to Implication for Neurodegeneration. *Mol Neurobiol* 59, 5000–5023. <https://doi.org/10.1007/s12035-022-02895-3>.
52. Jiao, F., Zhou, B., and Meng, L. (2023). The regulatory mechanism and therapeutic potential of transcription factor EB in neurodegenerative diseases. *CNS Neurosci Ther* 29, 37–59. <https://doi.org/10.1111/cns.13985>.
53. Tan, A., Prasad, R., Lee, C., and Jho, E. (2022). Past, present, and future perspectives of transcription factor EB (TFEB): mechanisms of regulation and association with disease. *Cell Death Differ* 29, 1433–1449. <https://doi.org/10.1038/s41418-022-01028-6>.
54. Ariano, C., Costanza, F., Akman, M., Riganti, C., Corà, D., Casanova, E., Astanina, E., Comunanza, V., Bussolino, F., and Doronzo, G. (2023). TFEB inhibition induces melanoma shut-down by blocking the cell cycle and rewiring metabolism. *Cell Death Dis* 14. <https://doi.org/10.1038/s41419-023-05828-7>.
55. Astanina, E., Bussolino, F., and Doronzo, G. (2021). Multifaceted activities of transcription factor EB in cancer onset and progression. *Molecular Oncology* 15, 327–346. <https://doi.org/10.1002/1878-0261.12867>.
56. Doronzo, G., Astanina, E., and Bussolino, F. (2021). The Oncogene Transcription Factor EB Regulates Vascular Functions. *Front. Physiol.* 12. <https://doi.org/10.3389/fphys.2021.640061>.

57. De Luca, C., Virtuoso, A., Papa, M., Cirillo, G., La Rocca, G., Corvino, S., Barbarisi, M., and Altieri, R. (2024). The Three Pillars of Glioblastoma: A Systematic Review and Novel Analysis of Multi-Omics and Clinical Data. *Cells* 13, 1754. <https://doi.org/10.3390/cells13211754>.
58. Pouyan, A., Ghorbanlo, M., Eslami, M., Jahanshahi, M., Ziaei, E., Salami, A., Mokhtari, K., Shahpasand, K., Farahani, N., Meybodi, T.E., et al. (2025). Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. *Mol Cancer* 24. <https://doi.org/10.1186/s12943-025-02267-0>.
59. Lan, Z., Li, X., and Zhang, X. (2024). Glioblastoma: An Update in Pathology, Molecular Mechanisms and Biomarkers. *IJMS* 25, 3040. <https://doi.org/10.3390/ijms25053040>.
60. Khan, M., Nasim, M., Feizy, M., Parveen, R., Gull, A., Khan, S., and Ali, J. (2024). Contemporary strategies in glioblastoma therapy: Recent developments and innovations. *Neuroscience* 560, 211–237. <https://doi.org/10.1016/j.neuroscience.2024.09.022>.
61. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* 23, 1231–1251. <https://doi.org/10.1093/neuonc/noab106>.
62. Price, M., Ballard, C., Benedetti, J., Neff, C., Cioffi, G., Waite, K.A., Kruchko, C., Barnholtz-Sloan, J.S., and Ostrom, Q.T. (2024). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2017–2021. *Neuro-Oncology* 26, vi1–vi85. <https://doi.org/10.1093/neuonc/noae145>.
63. Tan, A.C., Ashley, D.M., López, G.Y., Malinzak, M., Friedman, H.S., and Khasraw, M. (2020). Management of glioblastoma: State of the art and future directions. *CA A Cancer J Clinicians* 70, 299–312. <https://doi.org/10.3322/caac.21613>.
64. Wang, L.-B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* 39, 509–528.e20. <https://doi.org/10.1016/j.ccell.2021.01.006>.

65. Xu, C., Hou, P., Li, X., Xiao, M., Zhang, Z., Li, Z., Xu, J., Liu, G., Tan, Y., and Fang, C. (2024). Comprehensive understanding of glioblastoma molecular phenotypes: classification, characteristics, and transition. *Cancer Biol Med*, 1–19. <https://doi.org/10.20892/j.issn.2095-3941.2023.0510>.
66. Barbato, L., Bocchetti, M., Di Biase, A., and Regad, T. (2019). Cancer Stem Cells and Targeting Strategies. *Cells* 8, 926. <https://doi.org/10.3390/cells8080926>.
67. Walcher, L., Kistenmacher, A.-K., Suo, H., Kitte, R., Dluczek, S., Strauß, A., Blaudszun, A.-R., Yevsa, T., Fricke, S., and Kossatz-Boehlert, U. (2020). Cancer Stem Cells—Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. *Front. Immunol.* 11, 1280. <https://doi.org/10.3389/fimmu.2020.01280>.
68. Richichi, C., Osti, D., Del Bene, M., Fornasari, L., Patanè, M., Pollo, B., DiMeco, F., and Pelicci, G. (2016). Tumor-initiating cell frequency is relevant for glioblastoma aggressiveness. *Oncotarget* 7, 71491–71503. <https://doi.org/10.18632/oncotarget.11600>.
69. Agosti, E., Antonietti, S., Ius, T., Fontanella, M.M., Zeppieri, M., and Panciani, P.P. (2024). Glioma Stem Cells as Promoter of Glioma Progression: A Systematic Review of Molecular Pathways and Targeted Therapies. *IJMS* 25, 7979. <https://doi.org/10.3390/ijms25147979>.
70. Ghosh, N., Chatterjee, D., and Datta, A. (2025). Tumor heterogeneity and resistance in glioblastoma: the role of stem cells. *Apoptosis* 30, 1695–1729. <https://doi.org/10.1007/s10495-025-02123-y>.
71. Faletti, S., Osti, D., Ceccacci, E., Richichi, C., Costanza, B., Nicosia, L., Noberini, R., Marotta, G., Furia, L., Faretta, M.R., et al. (2021). LSD1-directed therapy affects glioblastoma tumorigenicity by deregulating the protective ATF4-dependent integrated stress response. *Sci. Transl. Med.* 13, eabf7036. <https://doi.org/10.1126/scitranslmed.abf7036>.
72. Diao, W., Zheng, J., Li, Y., Wang, J., and Xu, S. (2022). Targeting histone demethylases as a potential cancer therapy (Review). *Int J Oncol* 61, 103. <https://doi.org/10.3892/ijo.2022.5393>.
73. Gray, Z.H., Honer, M.A., Ghatalia, P., Shi, Y., and Whetstine, J.R. (2025). 20 years of histone lysine demethylases: From discovery to the clinic and beyond. *Cell* 188, 1747–1783. <https://doi.org/10.1016/j.cell.2025.02.023>.

74. Karakaidos, P., Verigos, J., and Magklara, A. (2019). LSD1/KDM1A, a Gate-Keeper of Cancer Stemness and a Promising Therapeutic Target. *Cancers* 11, 1821. <https://doi.org/10.3390/cancers11121821>.
75. Gu, F., Lin, Y., Wang, Z., Wu, X., Ye, Z., Wang, Y., and Lan, H. (2020). Biological roles of LSD1 beyond its demethylase activity. *Cell. Mol. Life Sci.* 77, 3341–3350. <https://doi.org/10.1007/s00018-020-03489-9>.
76. Majello, B., Gorini, F., Saccà, C.D., and Amente, S. (2019). Expanding the Role of the Histone Lysine-Specific Demethylase LSD1 in Cancer. *Cancers* 11, 324. <https://doi.org/10.3390/cancers11030324>.
77. Stitzlein, L.M., Gangadharan, A., Walsh, L.M., Nam, D., Espejo, A.B., Singh, M.M., Patel, K.H., Lu, Y., Su, X., Ezhilarasan, R., et al. (2023). Comparison of pharmacological inhibitors of lysine-specific demethylase 1 in glioblastoma stem cells reveals inhibitor-specific efficacy profiles. *Front. Neurol.* 14, 1112207. <https://doi.org/10.3389/fneur.2023.1112207>.
78. Marotta, G., Osti, D., Zaccheroni, E., Costanza, B., Faletti, S., Marinaro, A., Richichi, C., Mesa, D., Rodighiero, S., Soriani, C., et al. (2025). Metabolic traits shape responses to LSD1-directed therapy in glioblastoma tumor-initiating cells. *Sci. Adv.* 11, eadt2724. <https://doi.org/10.1126/sciadv.adt2724>.
79. Dyer, S.C., Austine-Orimoloye, O., Azov, A.G., Barba, M., Barnes, I., Barrera-Enriquez, V.P., Becker, A., Bennett, R., Beracochea, M., Berry, A., et al. (2025). Ensembl 2025. *Nucleic Acids Research* 53, D948–D957. <https://doi.org/10.1093/nar/gkae1071>.
80. The FANTOM Consortium, De Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., et al. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol* 35, 872–878. <https://doi.org/10.1038/nbt.3947>.
81. Perez, G., Barber, G.P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, C.M., et al. (2025). The UCSC Genome Browser database: 2025 update. *Nucleic Acids Research* 53, D1243–D1249. <https://doi.org/10.1093/nar/gkae974>.
82. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research* 47, D155–D162. <https://doi.org/10.1093/nar/gky1141>.

83. McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. *Science* 366, eaav1741. <https://doi.org/10.1126/science.aav1741>.
84. Huang, H.-Y., Lin, Y.-C.-D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., et al. (2022). miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Research* 50, D222–D230. <https://doi.org/10.1093/nar/gkab1079>.
85. Feng, J., Chen, W., Dong, X., Wang, J., Mei, X., Deng, J., Yang, S., Zhuo, C., Huang, X., Shao, L., et al. (2022). CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Research* 50, D1179–D1183. <https://doi.org/10.1093/nar/gkab830>.
86. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
87. Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30, 523–530. <https://doi.org/10.1093/bioinformatics/btt703>.
88. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* 37, W305–W311. <https://doi.org/10.1093/nar/gkp427>.
89. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
90. Hagberg, A.A., Schult, D.A. and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pp.11-15, Aug 2008.
91. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
92. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.

93. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>.
94. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
95. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
96. Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research* 40, 37–52. <https://doi.org/10.1093/nar/gkr688>.
97. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
98. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
99. Poliseno, L., Tuccoli, A., Mariani, L., Evangelista, M., Citti, L., Woods, K., Mercatanti, A., Hammond, S., Rainaldi, G. (2006). MicroRNAs modulate the angiogenic properties of HUVECs. *Blood*, 108, 3068–3071. <https://doi.org/10.1182/blood-2006-01-012369>.
100. Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L., and Smyth, G.K. (2025). edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research* 53, gkaf018. <https://doi.org/10.1093/nar/gkaf018>.
101. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.

102. Tarbell, E.D., and Liu, T. (2019). HMMRATAC: a Hidden Markov Modeler for ATAC-seq. *Nucleic Acids Research* 47, e91–e91. <https://doi.org/10.1093/nar/gkz533>.
103. Rosano, S., Corà, D., Parab, S., Zaffuto, S., Isella, C., Porporato, R., Hoza, R.M., Calogero, R.A., Riganti, C., Bussolino, F., et al. (2020). A regulatory microRNA network controls endothelial cell phenotypic switch during sprouting angiogenesis. *eLife* 9, e48095. <https://doi.org/10.7554/eLife.48095>.
104. Fernández-Hernando, C., and Suárez, Y. (2018). MicroRNAs in endothelial cell homeostasis and vascular disease. *Current Opinion in Hematology* 25, 227–236. <https://doi.org/10.1097/MOH.0000000000000424>.
105. Annese, T., Tamma, R., De Giorgis, M., and Ribatti, D. (2020). microRNAs Biogenesis, Functions and Role in Tumor Angiogenesis. *Front. Oncol.* 10, 581007. <https://doi.org/10.3389/fonc.2020.581007>.
106. Moreau, P.R., Tomas Bosch, V., Bouvy-Liivrand, M., Öunap, K., Örd, T., Pulkkinen, H.H., Pölönen, P., Heinäniemi, M., Ylä-Herttuala, S., Laakkonen, J.P., et al. (2021). Profiling of Primary and Mature miRNA Expression in Atherosclerosis-Associated Cell Types. *ATVB* 41, 2149–2167. <https://doi.org/10.1161/ATVBAHA.121.315579>.
107. Tong, Z., Cui, Q., Wang, J., and Zhou, Y. (2019). TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Research* 47, D253–D258. <https://doi.org/10.1093/nar/gky1023>.
108. Su, Y., Yuan, J., Zhang, F., Lei, Q., Zhang, T., Li, K., Guo, J., Hong, Y., Bu, G., Lv, X., et al. (2019). MicroRNA-181a-5p and microRNA-181a-3p cooperatively restrict vascular inflammation and atherosclerosis. *Cell Death Dis* 10, 365. <https://doi.org/10.1038/s41419-019-1599-9>.
109. Li, Y., Long, Y., Zhi, X., Hao, H., Wang, X., Liu, H., and Wang, L. (2023). miR-339-3p promotes AT1-AA-induced vascular inflammation by upregulating NFATc3 protein expression in vascular smooth muscle cells. *ABBS* 55, 295–303. <https://doi.org/10.3724/abbs.2023009>.
110. Green, C.E., Clarke, J., Bicknell, R., and Turner, A.M. (2021). Pulmonary MicroRNA Changes Alter Angiogenesis in Chronic Obstructive Pulmonary Disease and Lung Cancer. *Biomedicines* 9, 830. <https://doi.org/10.3390/biomedicines9070830>.

111. Shi, Q., Sun, B., Wang, D., Zhu, Y., Zhao, X., Yang, X., and Zhang, Y. (2020). Circ6401, a novel circular RNA, is implicated in repair of the damaged endometrium by Wharton's jelly-derived mesenchymal stem cells through regulation of the miR-29b-1-5p/RAP1B axis. *Stem Cell Res Ther* 11, 520. <https://doi.org/10.1186/s13287-020-02027-5>.
112. Liu, Y., Jiang, G., Lv, C., and Yang, C. (2022). miR -222-5p promotes dysfunction of human vascular smooth muscle cells by targeting RB1. *Environmental Toxicology* 37, 683–694. <https://doi.org/10.1002/tox.23434>.
113. Luo, Q., Jiang, Z., Jiang, J., Wan, L., Li, Y., Huang, Y., Qiu, J., Yu, K., and Zhuang, J. (2023). Tsp-1+ microglia attenuate retinal neovascularization by maintaining the expression of Smad3 in endothelial cells through exosomes with decreased miR-27a-5p. *Theranostics* 13, 3689–3706. <https://doi.org/10.7150/thno.84236>.
114. Chen, Q., Lin, G., Huang, J., Chen, L., Liu, Y., Huang, J., Zhang, S., and Lin, Q. (2020). Inhibition of miR-193a-3p protects human umbilical vein endothelial cells against intermittent hypoxia-induced endothelial injury by targeting FAIM2. *Aging* 12, 1899–1909. <https://doi.org/10.18632/aging.102729>.
115. Wang, W., Wang, Y., Piao, H., Li, B., Zhu, Z., Li, D., Wang, T., and Liu, K. (2021). Bioinformatics Analysis Reveals MicroRNA-193a-3p Regulates ACTG2 to Control Phenotype Switch in Human Vascular Smooth Muscle Cells. *Front. Genet.* 11, 572707. <https://doi.org/10.3389/fgene.2020.572707>.
116. Khoo, C.P., Roubelakis, M.G., Schrader, J.B., Tsaknakis, G., Konietzny, R., Kessler, B., Harris, A.L., and Watt, S.M. (2017). miR-193a-3p interaction with HMGB1 downregulates human endothelial cell proliferation and migration. *Sci Rep* 7, 44137. <https://doi.org/10.1038/srep44137>.
117. Pasquier, A., Pastore, N., D'Orsi, L., Colonna, R., Esposito, A., Maffia, V., De Cegli, R., Mutarelli, M., Ambrosio, S., Tufano, G., et al. (2023). TFEB and TFE3 control glucose homeostasis by regulating insulin gene expression. *The EMBO Journal* 42, e113928. <https://doi.org/10.15252/embj.2023113928>.
118. Gambardella, G., Staiano, L., Moretti, M.N., De Cegli, R., Fagnocchi, L., Di Tullio, G., Polletti, S., Braccia, C., Armirotti, A., Zippo, A., et al. (2020). GADD34 is a modulator of autophagy during starvation. *Sci. Adv.* 6, eabb0205. <https://doi.org/10.1126/sciadv.abb0205>.

119. Luo, J.-F., Wang, S., Fu, J., Xu, P., Shao, N., Lu, J.-H., and Ming, C. (2024). Integration of transcriptional and epigenetic regulation of TFEB reveals its dual functional roles in Pan-cancer. *NAR Cancer* 6, zcae043. <https://doi.org/10.1093/narcan/zcae043>.
120. Suzhi, Z., Liang, T., Yuexia, P., Lucy, L., Xiaoting, H., Yuan, Z., and Qin, W. (2015). Gap Junctions Enhance the Antiproliferative Effect of MicroRNA-124-3p in Glioblastoma Cells. *Journal Cellular Physiology* 230, 2476–2488. <https://doi.org/10.1002/jcp.24982>.
121. Lin, J., Wen, X., Zhang, X., Sun, X., Yunzhi, L., Peng, R., Zhu, M., Wang, M., Zhang, Y., Luo, W., et al. (2018). miR-135a-5p and miR-124-3p Inhibit Malignancy of Glioblastoma by Downregulation of Syndecan Binding Protein. *j biomed nanotechnol* 14, 1317–1329. <https://doi.org/10.1166/jbn.2018.2579>.
122. Zhang, G., Chen, L., Khan, A.A., Li, B., Gu, B., Lin, F., Su, X., and Yan, J. (2018). miRNA-124-3p/neuropilin-1(NRP-1) axis plays an important role in mediating glioblastoma growth and angiogenesis. *Intl Journal of Cancer* 143, 635–644. <https://doi.org/10.1002/ijc.31329>.
123. Zottel, A., Šamec, N., Kump, A., Raspor Dall'Olio, L.R., Pužar Dominkuš, P., Romih, R., Hudoklin, S., Mlakar, J., Nikitin, D., Sorokin, M., et al. (2020). Analysis of miR-9-5p, miR-124-3p, miR-21-5p, miR-138-5p, and miR-1-3p in Glioblastoma Cell Lines and Extracellular Vesicles. *IJMS* 21, 8491. <https://doi.org/10.3390/ijms21228491>.
124. Cai, S., Shi, C.-J., Lu, J.-X., Wang, Y.-P., Yuan, T., and Wang, X.-P. (2021). miR-124-3p inhibits the viability and motility of glioblastoma multiforme by targeting RhoG. *Int J Mol Med* 47, 69. <https://doi.org/10.3892/ijmm.2021.4902>.
125. Oliosio, D., Caccese, M., Santangelo, A., Lippi, G., Zagonel, V., Cabrini, G., Lombardi, G., and Dehecchi, M.C. (2021). Serum Exosomal microRNA-21, 222 and 124-3p as Noninvasive Predictive Biomarkers in Newly Diagnosed High-Grade Gliomas: A Prospective Study. *Cancers* 13, 3006. <https://doi.org/10.3390/cancers13123006>.
126. Tuoheti, M., Li, J., Zhang, C., Gao, F., Wang, J., and Wu, Y. (2024). MiR-124-3p inhibits cell stemness in glioblastoma via targeting EPHA2 through ALKBH5-mediated m6A modification. *Human Cell* 38, 10. <https://doi.org/10.1007/s13577-024-01129-z>.

127. Golchin, M.M., Arefian, E., Fekrirad, Z., and Tabar, G.H. (2025). miR-124-mediated temozolomide sensitivity and DNA repair modulation in Glioblastoma Multiforme. *Neuroscience* 573, 52–63. <https://doi.org/10.1016/j.neuroscience.2025.03.010>.
128. Zhang, H., Zheng, W., Chen, X., Sa, L., Huo, Y., Zhang, L., Shan, L., and Wang, T. (2024). DNAJC1 facilitates glioblastoma progression by promoting extracellular matrix reorganization and macrophage infiltration. *J Cancer Res Clin Oncol* 150, 315. <https://doi.org/10.1007/s00432-024-05823-1>.
129. Boccacino, J.M., Dos Santos Peixoto, R., Fernandes, C.F.D.L., Cangiano, G., Sola, P.R., Coelho, B.P., Prado, M.B., Melo-Escobar, M.I., De Sousa, B.P., Ayyadury, S., et al. (2024). Integrated transcriptomics uncovers an enhanced association between the prion protein gene expression and vesicle dynamics signatures in glioblastomas. *BMC Cancer* 24, 199. <https://doi.org/10.1186/s12885-024-11914-6>.
130. Wang, C.-C., Hueng, D.-Y., Huang, A.-F., Chen, W.-L., Huang, S.-M., and Yi-Hsin Chan, J. (2019). CD164 regulates proliferation, progression, and invasion of human glioblastoma cells. *Oncotarget* 10, 2041–2054. <https://doi.org/10.18632/oncotarget.26724>.
131. Wu, Y., Wang, H., and Wei, D. (2022). Oncogenic magnesium transporter 1 upregulates programmed death-1-ligand 1 expression and contributes to growth and radioresistance of glioma cells through the ERK/MAPK signaling pathway. *Bioengineered* 13, 9575–9587. <https://doi.org/10.1080/21655979.2022.2037214>.
132. Tang, Z., Chen, H., Zhong, D., Wei, W., Liu, L., Duan, Q., Han, B., and Li, G. (2020). TRIB3 facilitates glioblastoma progression via restraining autophagy. *Aging* 12, 25020–25034. <https://doi.org/10.18632/aging.103969>.
133. Meng, K., Hu, X., Zheng, G., Qian, C., Xin, Y., Guo, H., He, R., Ge, M., and Xu, J. (2022). Identification of prognostic biomarkers for papillary thyroid carcinoma by a weighted gene co-expression network analysis. *Cancer Medicine* 11, 2006–2019. <https://doi.org/10.1002/cam4.4602>.
134. Wang, Y., Shu, Y., Gu, C., and Fan, Y. (2019). The novel sugar transporter SLC50A1 as a potential serum-based diagnostic and prognostic biomarker for breast cancer. *CMAR* Volume 11, 865–876. <https://doi.org/10.2147/CMAR.S190591>.

135. Zhang, Q., Fang, Y., She, C., Zheng, R., Hong, C., Chen, C., and Wu, J. (2022). Diagnostic and prognostic significance of SLC50A1 expression in patients with primary early breast cancer. *Exp Ther Med* 24, 616. <https://doi.org/10.3892/etm.2022.11553>.
136. Li, P., Yang, X., Tang, H., Zhou, Z., and Liu, B. (2025). Cancer-associated fibroblasts-secreted lactate promotes RNA polymerase III subunit G-mediated epithelial–mesenchymal transition in non-small cell lung cancer by increasing m6A modification of zinc finger protein 384. *Journal of Cell Communication and Signaling* 19, e70037. <https://doi.org/10.1002/ccs3.70037>.
137. Yang, M., Chen, X., Liu, Y., Wang, M., Chen, L., Qi, H., Sun, M., Chen, J., Dong, H., and Xie, T. (2025). The anti-metastatic effect of elemene in human hepatocellular carcinoma is potentially mediated by inhibiting a novel driver-circBPTF. *Phytomedicine* 143, 156808. <https://doi.org/10.1016/j.phymed.2025.156808>.
138. Liu, X., Zhu, L., Li, D., and Chen, X. (2024). The upregulation of POLR3G correlates with increased malignancy of bladder urothelium. *Eur J Med Res* 29, 381. <https://doi.org/10.1186/s40001-024-01980-8>.
139. Zhan, Y., Deng, C., Tang, L., Li, S., Xu, T., Wei, X., Zhang, X., Zheng, C., Deng, L., Shao, C., et al. (2025). NAT10 Increases Lysosomal Acidification to Promote Esophageal Cancer Metastasis via ac4C Acetylation of ATP6V0E1 mRNA. *Advanced Science* 12, e02931. <https://doi.org/10.1002/adv.202502931>.
140. Wang, H., Ma, S., Yang, Z., Niu, R., Zhu, H., Li, S., Gao, S., Li, Z., and Tian, Y. (2025). Revolutionizing ESCC prognosis: the efficiency of tumor-infiltrating immune cells (TIIC) signature score. *Discov Onc* 16, 65. <https://doi.org/10.1007/s12672-024-01709-3>.
141. Ling, B., Tian, M., Wang, J., Luo, W., Yu, J., Wan, X., Dong, Q., Gao, M., Zhao, M., Zheng, X., et al. (2025). Integrative Multi-Omics Analysis Reveals Key Metabolic Regulators and Prognostic Biomarkers in Pediatric and Adult Thyroid Cancer. *J. Cancer* 16, 3551–3566. <https://doi.org/10.7150/jca.117034>.
142. Hua, B., Yang, Q., Song, S., Li, W., Xu, B., and Gu, Y. (2025). Exploring the prognostic significance and therapeutic potential of SUCLG2 in prostate cancer. *Front. Genet.* 16, 1592779. <https://doi.org/10.3389/fgene.2025.1592779>.

143. Leng, Y., Tian, T., Tang, B., Ma, Y., Li, Z., Shi, Q., Liu, J., Zhou, Y., Wang, W., Huang, C., et al. (2024). The oncogenic role and regulatory mechanism of ACAA2 in human ovarian cancer. *Molecular Carcinogenesis* 63, 1362–1377. <https://doi.org/10.1002/mc.23729>.
144. Gravina, T., Boggio, C.M.T., Gorla, E., Racca, L., Polidoro, S., Centonze, S., Ferrante, D., Lunghi, M., Graziani, A., Corà, D., et al. (2023). The role of diacylglycerol kinases in acute myeloid leukemia. *Biomedicines* 11, 1877. <https://doi.org/10.3390/biomedicines11071877>.
145. Favero, F., Re, A., Dason, M.S., Gravina, T., Gagliardi, M., Mellai, M., Corazzari, M., and Corà, D. (2024). Characterization of gut microbiota dynamics in an Alzheimer’s disease mouse model through clade-specific marker-based analysis of shotgun metagenomic data. *Biology Direct* 19, 100. <https://doi.org/10.1186/s13062-024-00541-7>.

11. LIST OF PUBLICATIONS

1. **Gravina, T.**, Boggio, C.M.T., Gorla, E., Racca, L., Polidoro, S., Centonze, S., Ferrante, D., Lunghi, M., Graziani, A., Corà, D., and Baldanzi, G. (2023). Role of diacylglycerol kinases in acute myeloid leukemia. *Biomedicines* 11, 1877. <https://doi.org/10.3390/biomedicines11071877>.
2. **Gravina, T.**, Favero, F., Rosano, S., Parab, S., Diaz Alcalde, A., Bussolino, F., Doronzo, G., and Corà, D. (2024). Integrative bioinformatics analysis reveals a transcription factor EB-driven microRNA regulatory network in endothelial cells. *Int. J. Mol. Sci.* 25, 7123. <https://doi.org/10.3390/ijms25137123>.
3. Favero, F., Re, A., Dason, M.S., **Gravina, T.**, Gagliardi, M., Mellai, M., Corazzari, M., and Corà, D. (2024). Characterization of gut microbiota dynamics in an Alzheimer's disease mouse model through clade-specific marker-based analysis of shotgun metagenomic data. *Biol. Direct* 19, 100. <https://doi.org/10.1186/s13062-024-00541-7>.
4. Gorla, E., Cartella, M. C., Borghetti, E., Lovati, G., Racca, L., **Gravina, T.**, Biazzo, G., Bonello, G., Malacarne, V., De Giorgis, V., Corà, D., Manfredi, M., Massarotti, A., Graziani, A., and Baldanzi, G. (2025). Acute Myeloid Leukemia: A Key Role of DGK α and DGK ζ in Cell Viability. *Cells*, 14(21), 1721. <https://doi.org/10.3390/cells14211721>.
5. **Gravina, T.**, Favero, F., Faletti, S., Osti, D., Baldanzi, G., Pelicci, G. and Corà, D. A microRNA Regulatory Network Driven by the Combination of the Epigenetic Factor LSD1 and its Cofactor ATF4 in GBM. Manuscript in preparation.