T.C.

DOKUZ EYLÜL ÜNIVERSITESI

IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

IDENTIFICATION OF FUNCTIONALLY RELATED AND MUTUALLY EXCLUSIVE microRNAS ASSOCIATED WITH DISEASES USING DEEP NEURAL NETWORKS APPROACH

REZA ARSHINCHI BONAB

MOLECULAR BIOLOGY AND GENETICS

MASTER OF SCIENCE THESIS

IZMIR-2023

THESIS CODE: DEU.IBG.MSc-2020850028

T.C.

DOKUZ EYLÜL ÜNIVERSITESI

IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

IDENTIFICATION OF FUNCTIONALLY RELATED AND MUTUALLY EXCLUSIVE microRNAS ASSOCIATED WITH DISEASES USING DEEP NEURAL NETWORKS APPROACH

REZA ARSHINCHI BONAB

MOLECULAR BIOLOGY AND GENETICS

MASTER OF SCIENCE THESIS

IZMIR-2023

THESIS CODE: DEU.IBG.MSc-2020850028

T.C.

DOKUZ EYLÜL ÜNIVERSITESI

IZMIR INTERNATIONAL BIOMEDICINE AND GENOME INSTITUTE

IDENTIFICATION OF FUNCTIONALLY RELATED AND MUTUALLY EXCLUSIVE microRNAS ASSOCIATED WITH DISEASES USING DEEP NEURAL NETWORKS APPROACH

MOLECULAR BIOLOGY AND GENETICS

MASTER OF SCIENCE THESIS

REZA ARSHINCHI BONAB

Supervisor: Asst. Prof. ATHANASIA PAVLOPOULOU

THESIS CODE: DEU.IBG.MSc-2020850028

Dokuz Eylül University Izmir International Biomedicine and Genome Institute Department of Genomics and Molecular Biotechnology, Molecular Biology and Genetics graduate program Master of Science student Reza Arshinchi Bonab has successfully completed his Master of Science thesis titled **"Identification of functionally related and mutually exclusive microRNAs associated with diseases using Deep Neural Networks approach"** on the date of 12/06/2023.

HEAD OF THE COMMITTEE
Asst. Prof. Dr. Athanasia PAVLOPOULOU

MEMBER
Asst. Prof. Dr. Hani ALOTAİBİ

MEMBER
Prof. Dr. Ralph MEUWİSSEN

YÜKSEK LİSANS TEZ SAVUNMA SINAVI TUTANAĞI

Dokuz Eylül Üniversitesi İzmir Uluslararası Biyotıp ve Genom Enstitüsü Genom Bilimleri ve Moleküler Biyoteknoloji Anabilim Dalı, Moleküler Biyoloji ve Genetik Yüksek Lisans programı öğrencisi Reza Arshinchi Bonab **"Derin Sinir Ağları yaklaşımı kullanılarak hastalıklarla ilişkili İşlevsel olarak ilişkili ve Karşılıklı olarak özel mikroRNA'ların tanımlanması"** konulu Yüksek Lisans tezini 12/06/2023 tarihinde başarılı olarak tamamlamıştır.

BAŞKAN
Dr. Öğr. Üyesi Athanasia PAVLOPOULOU

JÜRİ ÜYESİ
Dr. Öğr. Gör. Hani ALOTAİBİ

JÜRİ ÜYESİ
Prof. Dr. Ralph MEUWİSSEN

TABLE OF CONTENTS

TABLE OF CONTENTS i
INDEX OF TABLES ii
INDEX OF FIGURES iii
LIST OF ABBREVIATIONS iv
ACKNOWLEDGEMENTS v
ÖZET 1
ABSTRACT 2
1. INTRODUCTION AND AIM 2
1.1. Statement and Importance of the Problem 2
1.2. Aim of the the Study 3
1.3. Hypothesis of the Study 4
2. GENERAL INFORMATION 5
2.1. History of microRNAs 5
2.2. miRNAs and Cancers 5
2.3. Bioinformatics approaches 6
2.4. Network-based approaches to cancer diagnosis 7
2.5. Artificial intelligence approaches to cancer diagnosis 8
2.5.1 Automated screening and detection 8
2.5.2. Genomic analysis and precision medicine 9
2.5.3. Clinical decision support systems 9
2.5.4. Identification and characterization of miRNAs 9
3. MATERIALS AND METHODOLOGY 9
3.1. Type of the Study 10
3.2. Time and Place of the Study 10
3.3. Materials and Methods of the Study 10
3.4. Variables of the Study 11
3.5. Tools for Data Collection 11
3.6. Study Plan and Calendar 12
3.7. Acquisition of miRNA-neoplasm associations 12
3.8. Neoplasm nomenclature 18
3.9. Prediction of miRNA target genes 25
3.10. Integration of miRNA-miRNA interactions 25
3.11. Construction of miRNA-neoplasm networks 28

3.12. Identification of mutually exclusive cancer-relevant miRNA modules ...32
3.13. Predictive performance of machine learning algorithms for miRNA data classification ...35
4. RESULTS ...44
4.1. Neoplasm classification ...45
4.2. Construction of miRNA-miRNA graphs per neoplasm ...45
4.3. Detection of neoplasm-specific miRNA modules ...52
5. DISCUSSION ...56
6. CONCLUSION AND FUTURE ASPECTS ...59
7. REFERENCES ...60
8. APPENDIX ...67
8.1. Ethics Committee Approval ...67
8.2. Curriculum Vitae ...67

INDEX OF TABLES

Table 1. Major classes of neoplasms, neoplasm-associated miRNAs and miRNA-miRNA interactions. ...47

Table 2. Mutually exclusive miRNA modules per neoplasm ...53

INDEX OF FIGURES

Figure 1. The role of microRNAs in human cancer; adapted from (Peng and Croce, 2016). 6

Figure 2. A graphical presentation of the overall workflow of this study. 14

Figure 3. Bar graph depicting the accuracy scores of various machine learning models.
..... 37

Figure 4. Endocrine gland neoplasm miRNA-miRNA graph.....51

Figure 5. Neoplasm-specific modules 56

LIST OF ABBREVIATIONS

HGNC	HUGO Gene Nomenclature Committee
GO	Gene Ontology
ICD-11	International Classification of Diseases and Related Health Problems, revision 11
MeSH	Medical Subject Headings
miRNA	microRNA
MLP	Multi-Layer Perceptron
SVM	Support Vector Machine
WHO	World Health Organization

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my research supervisor, Asst. Prof. Dr Athanasia Pavlopoulou for giving me the opportunity to perform my research and providing me with invaluable guidance at every stage of my research study.

I would also like to convey heartfelt appreciation to my partner Sadaf Asfa, who supported me endlessly, not only throughout our studies but also in our life.

I am very appreciative to my committee members Asst. Prof. Dr. Hani Alotaibi and Prof. Dr. Ralph Meuwissen for their valuable time and contributions.

Finally I would like to extend my profound thanks to my family, who has been supporting me emotionally and financially.

DERİN SİNİR AĞLARI YAKLAŞIMI KULLANILARAK HASTALIKLARLA İLİŞKİLİ İŞLEVSEL OLARAK İLİŞKİLİ VE KARŞILIKLI OLARAK ÖZEL mikroRNA'LARIN TANIMLANMASI

Reza ARSHINCHI BONAB, İzmir Uluslararası Biyotıp ve Genom Merkezi, Dokuz Eylül Üniversitesi Sağlık Kampüsü, Balçova 35340 – İzmir / TÜRKİYE

ÖZET

MikroRNA'lar, hastalık epigenomunun ana düzenleyici bileşenlerini temsil eder ve kanserler de dahil olmak üzere çeşitli hastalıkların doğru teşhisi ve prognozu için güçlü biyobelirteçler oluştururlar. Yüksek verimli teknolojilerin ortaya çıkışı, çok miktarda miRNA-kanser ilişkilendirme verilerinin üretilmesini kolaylaştırdı. Hesaplamalı yaklaşımlar, çeşitli kanser türleri için miRNA imzalarının tanımlanmasına yönelik bu verileri etkin bir şekilde analiz etmek ve yorumlamak için yaygın olarak kullanılmıştır. Burada, ağ tabanlı yöntemler kullanarak ana neoplastik hastalık grupları için çekirdek miRNA etkileşim kümelerini keşfetmek için yeni bir hesaplama iş akışı uyguladık. Bu amaçla, her bir ana neoplazm grubu için miRNA merkezli ağlar oluşturmak için dört kapsamlı kamuya açık kaynaktan miRNA-kanser ilişki verileri kullanıldı. Karşılık gelen miRNA-miRNA etkileşimleri, paylaşılan fonksiyonel olarak ilişkili hedef genlere dayanarak çıkarılmıştır. Oluşturulan ağların topolojik özellikleri, her ağda çekirdek modüller oluşturan yüksek düzeyde birbirine bağlı miRNA kümelerini tespit etmek için araştırıldı. Her grafikten en yüksek derecede karşılıklı ayrıcalık sergileyen modüller seçildi. Bu şekilde, ilgili hastalıklar için potansiyel imzaları temsil edebilecek neoplazmaya özgü miRNA modülleri tanımlandı.

Anahtar kelimeler: mikroRNA, etkileşimler, neoplazmalar, biyoinformatik, ağ analizi, karşılıklı münhasırlık, modüller, teşhis

IDENTIFICATION OF FUNCTIONALLY RELATED AND MUTUALLY EXCLUSIVE microRNAS ASSOCIATED WITH DISEASES USING DEEP NEURAL NETWORKS APPROACH

Reza ARSHINCHI BONAB, Izmir International Biomedicine and Genome Center, Dokuz Eylül University Health Campus, Balçova 35340 – İzmir / TURKEY

ABSTRACT

MicroRNAs represent major regulatory components of the disease epigenome and they constitute powerful biomarkers for the accurate diagnosis and prognosis of various diseases, including cancers. The advent of high-throughput technologies facilitated the generation of a vast amount of miRNA-cancer association data. Computational approaches have been utilized widely to effectively analyze and interpret these data towards the identification of miRNA signatures for diverse types of cancers. Herein, we applied a novel computational workflow to discover core sets of miRNA interactions for the major groups of neoplastic diseases by employing network- based methods. To this end, miRNA-cancer association data from four comprehensive publicly available resources were utilized for constructing miRNA-centered networks for each major group of neoplasms. The corresponding miRNA-miRNA interactions were inferred based on shared functionally related target genes. The topological attributes of the generated networks were investigated in order to detect clusters of highly interconnected miRNAs that form core modules in each network. Those modules that exhibited the highest degree of mutual exclusivity were selected from each graph. In this way, neoplasm-specific miRNA modules were identified that could represent potential signatures for the corresponding diseases.

Keywords: microRNA, interactions, neoplasms, bioinformatics, network analysis, mutual exclusivity, modules, diagnosis

1. INTRODUCTION AND AIM

1.1. Statement and Importance of the Problem

MicroRNAs (miRNAs) are endogenous, small (~22 nucleotides in length), non-protein-coding RNAs that play a critical role in regulating the expression of target mRNAs at the post-transcriptional level via complementary base-pairing of the so-called miRNA "seed" region with the 3' UTR region of the target sequence (Bartel,

2004; Kehl et al., 2017). In this way, miRNAs reduce the stability of the target mRNA and/or inhibit its translation, thereby downregulating the expression of the corresponding gene (Ambros, 2004).

The advent of high-throughput technologies has generated a vast amount of miRNA-cancer association data, which requires efficient analysis and interpretation for identifying miRNA signatures which are mutually exclusive and also specific to different types of cancers. However, the identification of core sets of miRNA interactions for different neoplastic diseases using network-based methods remains a challenging task.

The discovery of miRNA signatures associated with various cancer types holds immense potential for enhancing cancer diagnosis, prognosis, and treatment effectiveness (Adam-Artigues et al., 2021; Zhang et al., 2019). Presently, the task of identifying reliable biomarkers for cancer diagnosis and prognosis poses a significant challenge. Additionally, gaining a deeper comprehension of the molecular mechanisms underlying cancer development and progression is crucial for the advancement of novel therapeutic strategies. By utilizing computational approaches and network- based methods to scrutinize miRNA-cancer association data, it becomes possible to uncover key sets of miRNA interactions (Jin et al., 2019) that are unique to specific cancer types. Such interactions have the potential to serve as valuable biomarkers for their respective diseases.This study aims to address this important problem by applying a novel computational workflow to discover neoplasm-specific miRNA modules that could serve as potential biomarkers for the diagnosis and prognosis of different types of cancers.

The findings of this study could be possibly utilized in the clinical setting for the timely and accurate diagnosis of diverse cancer types.

1.2. Aim of the the Study

The aim of my study was to identify tumor-specific miRNAs through an integrative *in silico* approach. To this end, relevant data from publicly available highthroughput resources were exploited using a multidisciplinary bioinformatics approach. The novelty of my research proposal lies in the fact that, to our knowledge, a novel computational workflow was applied to discover core sets of miRNA

interactions for the major groups of neoplastic diseases by employing network-based methods

1.3. Hypothesis of the Study

It is hypothesized that identification of miRNA signatures specific for diverse types of cancers is critical for the accurate diagnosis and prognosis of cancer.

2. GENERAL INFORMATION

2.1. History of microRNAs

MiRNAs were initially detected in 1993 by Lee and colleagues in the nematode Caenorhabditis elegans, unveiling a new class of RNA molecules. These miRNAs, which are small in size (~19-24 nucleotides), are naturally occurring and have been conserved throughout evolution. Their primary role is to regulate gene expression at the post-transcriptional level. This is achieved through their ability to bind complementarily to target messenger RNA (mRNA), subsequently influencing the translation process and potentially impeding or modifying protein production (Lee et al., 1993)

A single miRNA can potentially regulate the expression of multiple genes, and conversely, a gene can be targeted by numerous miRNAs. Because of this property, miRNAs have been reported to be dysregulated in diverse human diseases (Chung et al., 2021; Climent et al., 2020; Kweon et al., 2020; Lam et al., 2018)

2.2. miRNAs and Cancers

Human cancers exhibit several distinctive features, commonly referred to as hallmarks. These hallmarks encompass sustained proliferative signaling, evasion of growth suppressors, activated invasion and metastasis, replicative immortality, angiogenesis, resistance to cell death, and evasion of immune destruction(Peng & Croce, 2016). Remarkably, microRNAs (miRNAs), which experience dysregulation in virtually all types of human cancers, are implicated in contributing to each of these cancer hallmarks.

Several miRNAs may function either as oncogenes (“oncomiRs”) or tumor suppressors (“suppressomiRs”), over- and under-expressed, respectively, in tumors (Zhang et al., 2019). For example, the known oncomiR miR-21 is one of the most frequently upregulated miRNAs in solid tumours (Jenike & Halushka, 2021) and also members of the mir-34 family are considered as potent tumor suppressive miRNAs (Zhang et al., 2019); MRX34, a liposomal miR-34a mimic has been evaluated in patients with advanced solid tumors (Hong et al., 2020). In addition, miR-143, miR-221/222, and miR-452 are considered potent biomarkers for tumor stratification and

diagnosis (Puerta-Gil et al., 2012). Moreover, miRNAs also have prognostic potential as they can predict cancer patient survival (Zhao et al., 2015).

The initial indications of miRNAs' role in human cancer emerged from the work conducted by Croce's research team (Peng & Croce, 2016) (Figure 1). Their investigations were centered around identifying tumor suppressors located within the chromosome 13q14 region in B-cell chronic lymphocytic leukemia cells. This pioneering research shed light on the involvement of miRNAs in cancer pathogenesis.

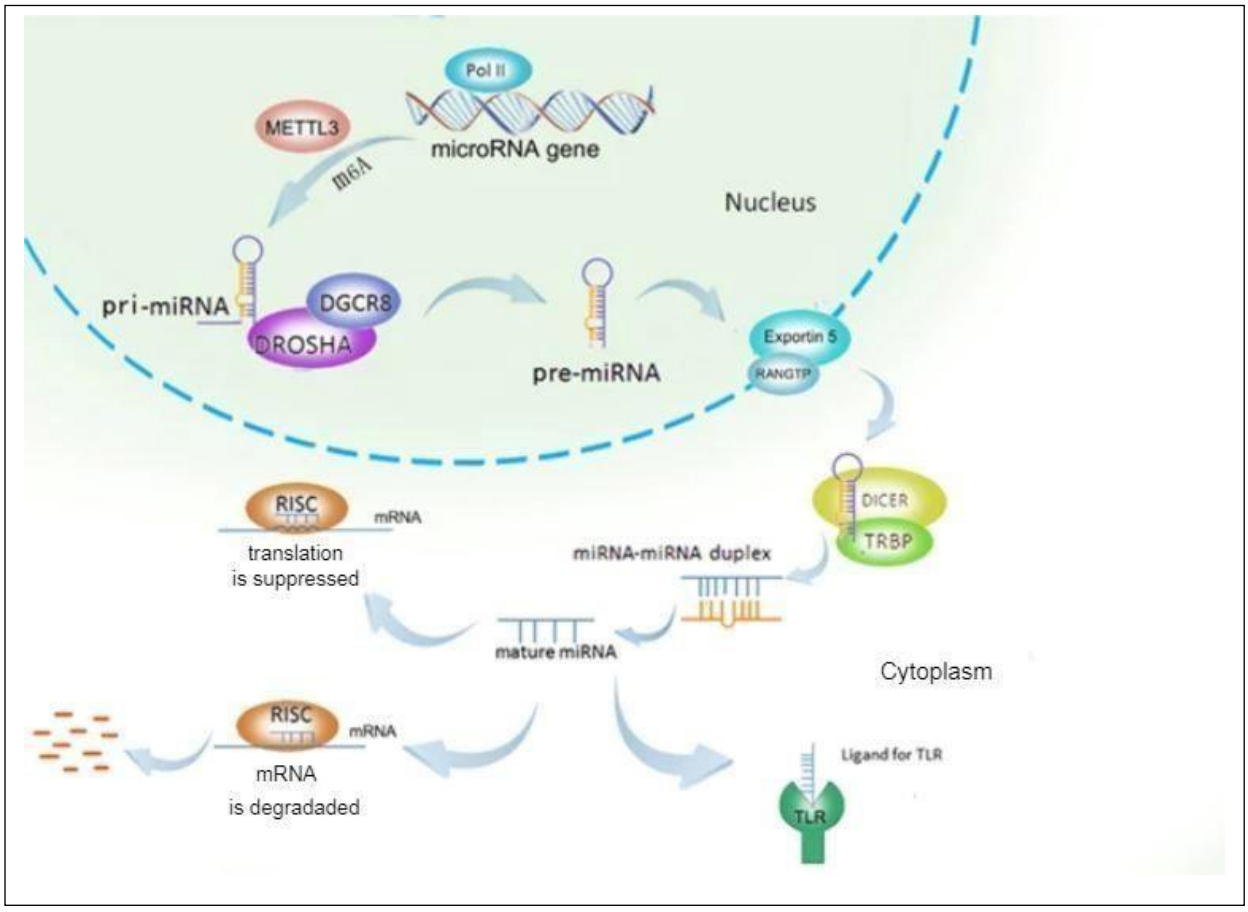


Figure 1. The role of microRNAs in human cancer; adapted from (Peng and Croce, 2016).

2.3. Bioinformatics approaches

Bioinformatics methodologies have gained significant traction in analyzing and comprehending pertinent data for the identification of miRNA signatures associated with various types and subtypes of cancers. These approaches have proven instrumental in not only recognizing existing miRNA diagnostic signatures but also in predicting them for diverse forms of cancer (Yerukala Sathipati & Ho, 2020) .

However, despite the wealth of studies on cancer-associated miRNAs and the identification of miRNAs that could serve as potential diagnostic biomarkers, these studies have largely focused on certain types/subtypes of cancers.

For example in a study by Chen and colleagues (Chen et al., 2020), it was invesigated the regulation of miRNAs in colon cancer by applying bioinformatics methods to identify miRNAs and their potential critical target genes associated with colon cancer and prognosis.

In another study, Zhao and colleagues (Zhao et al., 2015) identified microRNAs for only four cancer types, namely, lung cancer, colon cancer, gastric cancer and breast cancer.

Of note, these miRNA signatures ignore mutual exclusivity, i.e., miRNAs used to detect a particular type/subtype of cancer and not another. To the best of the authors'our knowledge, identification of neoplasm-specific miRNAs that are detected exclusively in a particular type/subtype of cancer is lacking. Therefore, there is an emerging need for identifying core miRNA signatures for characterizing diverse cancers. The findings of this study could be considered towards reducing the cases of spurious cancer diagnosis and improving neoplasm-specific detection by the application of panels of miRNAs exclusive and unique to the major types of cancers (Paranjape et al., 2009).

In the present study, an integrated bioinformatics methodology has been employed for the identification of core sets of miRNAs specific to a particular neoplastic disease, by exploiting relevant data from publicly available resources. Only highly connected miRNA modules displaying significant mutually exclusivity were selected for each group of neoplasms. The component miRNAs of those modules could potentially constitute biomarkers for distinguishing different types of cancers, which could complement and update the existing ones.

2.4. Network-based approaches to cancer diagnosis

Network-based strategies employ computational methodologies to analyze and process biological networks, including networks centered around miRNAs or protein-

protein interactions. These approaches aim to identify specific molecular markers associated with particular diseases. Particularly useful for handling vast amounts of data, such as miRNA or gene expression profiles, these methods unveil intricate interactions between various molecules that contribute to disease development and progression (Thomas et al., 2022). In the context of miRNA-based networks, networkbased techniques enable the detection of essential sets of miRNA interactions specific to different cancer types. These networks are constructed by connecting miRNAs based on shared target genes with functional relationships. Subsequently, the networks are examined to identify clusters of interconnected miRNAs, which may represent core modules in the disease process. These modules can be further characterized based on their topological properties, such as the degree of mutual exclusivity, to determine their relevance in disease diagnosis and prognosis. Network-based approaches offer numerous advantages compared to traditional single-gene analyses. They have the capacity to detect complex molecular interactions and provide a systemic understanding of the disease process (Yan et al., 2018). This, in turn, can lead to the identification of novel therapeutic targets and biomarkers. Additionally, network-based approaches can uncover crucial regulatory miRNAs that may not exhibit differential expression in disease states but play pivotal roles in disease development and progression.

2.5. Artificial intelligence approaches to cancer diagnosis

2.5.1 Automated screening and detection

Artificial intelligence (AI) algorithms have the capability to analyze medical images like mammograms, computerized tomography (CT) scans, or histopathology slides. This analysis assists radiologists and pathologists in detecting initial indications of cancer. By learning patterns and features from annotated images, AI models become equipped to conduct automated screening and identify areas of potential concern. This technology aids in early detection and intervention, empowering healthcare professionals to identify suspicious regions with improved accuracy. Consequently, AI algorithms contribute significantly to facilitating timely and effective intervention (Bi et al., 2019).

2.5.2 Genomic analysis and precision medicine

AI techniques can facilitate a comprehensive examination of genomic data to identify distinct gene mutations, biomarkers, and genetic signatures that are linked to various types of cancer. This valuable information serves as a guide for personalized treatment strategies, such as targeted therapies and precision medicine approaches. By employing AI algorithms, large-scale genomic datasets can be effectively analyzed to uncover molecular classifications or subtypes of cancer. Furthermore, these algorithms enable the prediction of treatment responses and contribute to the development of customized treatment plans. The integration of AI and genomic analysis holds significant potential in advancing the field of personalized medicine (Dlamini et al., 2020).

2.5.3 Clinical decision support systems

AI-powered clinical decision support systems offer real-time assistance to healthcare professionals by amalgamating patient-specific data, medical guidelines, and research evidence. These systems aid oncologists in deciphering intricate data, providing them with treatment suggestions, predictions of treatment outcomes, and support for personalized therapeutic decision-making. By enhancing the accuracy and efficiency of clinical decision-making in cancer care, these systems play a vital role in optimizing patient outcomes (Wang et al., 2023).

2.5.4 Identification and characterization of miRNAs

AI has become a potent instrument in the realm of miRNA exploration, bringing about a transformative impact on the identification and understanding of these vital regulatory molecules. Through the utilization of sophisticated algorithms and machine learning techniques, AI exhibits remarkable efficiency in analyzing extensive genomic and transcriptomic data. This analytical prowess enables the discovery of novel miRNAs and facilitates a deeper comprehension of their functional significance. By employing AI algorithms, characteristic structural features, sequence motifs, and thermodynamic properties that serve as indicators of potential miRNA candidates can be detected (Parveen et al., 2019).

3. MATERIALS AND METHODOLOGY

3.1. Type of the Study

The entire study was performed by employing computational methods and resources.

3.2. Time and Place of the Study

This study was conducted in Izmir Biomedicine and Genome Institute from March 2021 to July 2022.

3.3. Materials and Methods of the Study

In this study, the miRNA and neoplasm associations were retrieved from four comprehensive publicly available databases:

- miR2Disease (http://www.mir2disease.org/).
- miRCancer (http://mircancer.ecu.edu/) □ dbDEMC (https://www.biosino.org/dbDEMC.) □ HMDD v3.2 (https://www.cuilab.cn/hmdd).

In-house R scripts were used to extract the miRNA-cancer pairwise associations from those databases; duplicate entries were removed.

Both the experimentally supported and predicted mRNA targets of the neoplasm-associated miRNAs were retrieved with the usage of the following state-of-the-art software tools:

- microT_CD (https://dianalab.e-ce.uth.gr/html/dianauniverse/index.php?r=microT_CDS)
- TargetScan (https://www.targetscan.org/vert_80/)
- miRDB (https://mirdb.org/)
- PicTar (https://pictar.mdc-berlin.de/)

The miRNA-miRNA pairwise functional relationships were assembled and inferred from two diverse sources, namely (i) GOSemSim (https://www.bioconductor.org/packages/GOSemSim), and (ii) MIRGOFs (http://www.csbio.sjtu.edu.cn/bioinf/MiRGOFS/)

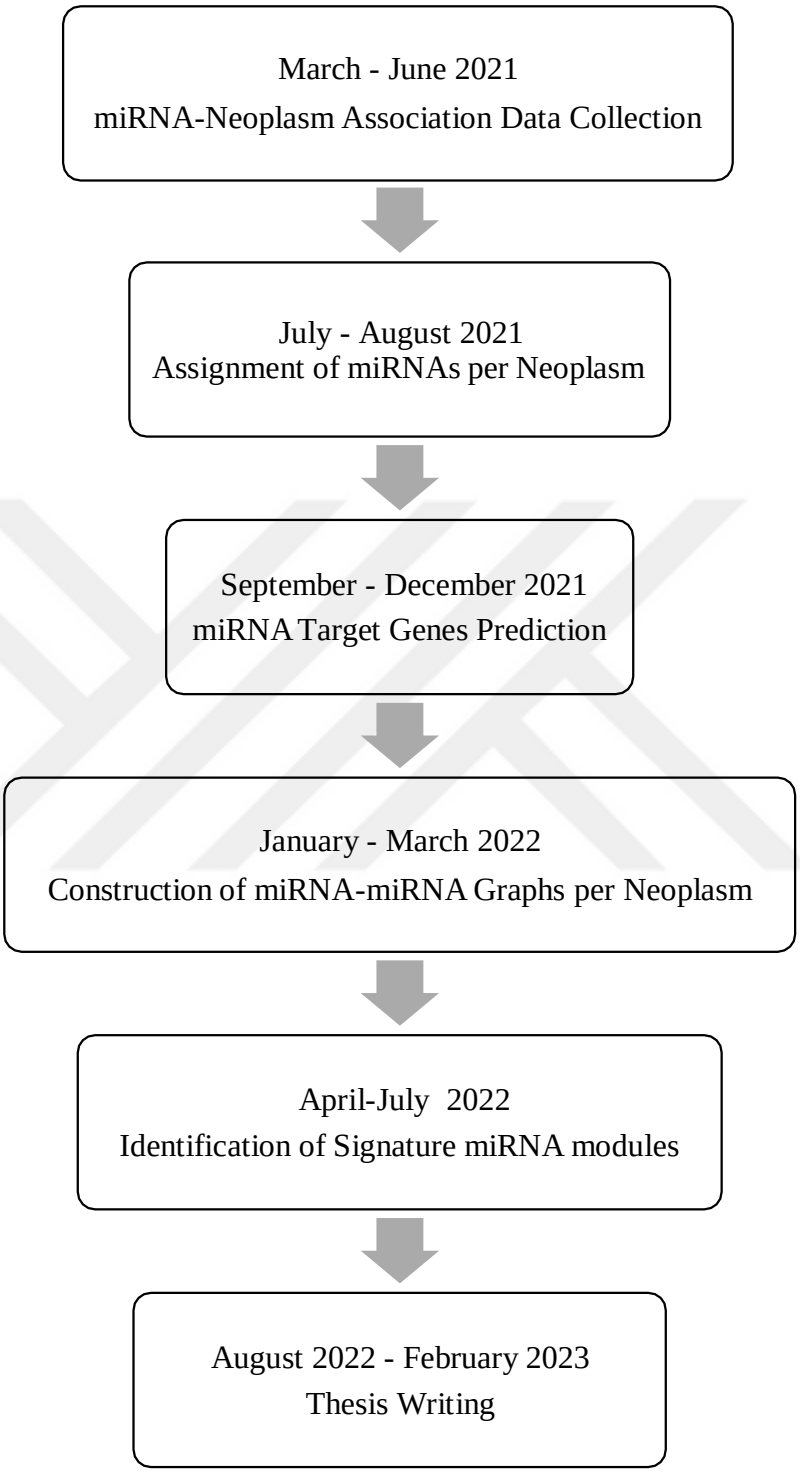
3.4. Variables of the Study

The R programming software platform (http://www.r-project.org/) (version 4.1.1) was used to extract the miRNA-cancer pairwise associations from the databases and handle the complexity of the datasets. For the network analysis process igraph package in R was used.

3.5. Tools for Data Collection

The R programming software platform (http://www.r-project.org/) (version 4.1.1) and in-house R scripts were used to extract the miRNA-cancer pairwise associations from the relevant databases.

3.6. Study Plan and Calendar



3.7. Acquisition of miRNA-neoplasm associations

I aimed to analyze the associations between miRNAs and neoplasms (tumors) and retrieve relevant information from publicly available databases (Figure 2). Four comprehensive databases were utilized for this purpose: miR2Disease, miRCancer, dbDEMC, and HMDD v3.2.

- miR2Disease, which was established by Jiang and colleagues (2009) (Jiang et al., 2009), is a manually curated repository of miRNAs that have been found to be deregulated in various human diseases. This database is a valuable resource for researchers as it provides information on the role of miRNAs in disease process.
- miRCancer, established by Xie and coworkers (2013) (Xie et al., 2013) , is another database that was used in this research. It contains information on the association between miRNAs and cancer extracted through literature mining. This database provides a comprehensive overview of the existing knowledge on the relationship between miRNAs and cancer.
- dbDEMC, established by Yang and colleagues (2010) (Yang et al., 2010), is a database that focuses specifically on the differential regulation of miRNAs in cancers. This database is an important resource for researchers as it provides information on the changes in miRNA expression levels in different types of cancer, which can provide insights into the molecular mechanisms underlying cancer development.
- HMDD v3.2, established by Huang and coworkers (2019) (Huang et al., 2019),is a repository of experimentally verified human miRNA-disease relationships. This database is considered a valuable resource for researchers as it provides information on the associations between miRNAs and diseases that have been experimentally validated.

The miRNA-cancer pairwise associations from these databases, were manually retrieved. The extracted information was then processed to remove any duplicate entries. This ensured that the resulting data were clean and accurate, allowing for a more reliable analysis of the associations between miRNAs and neoplasms.

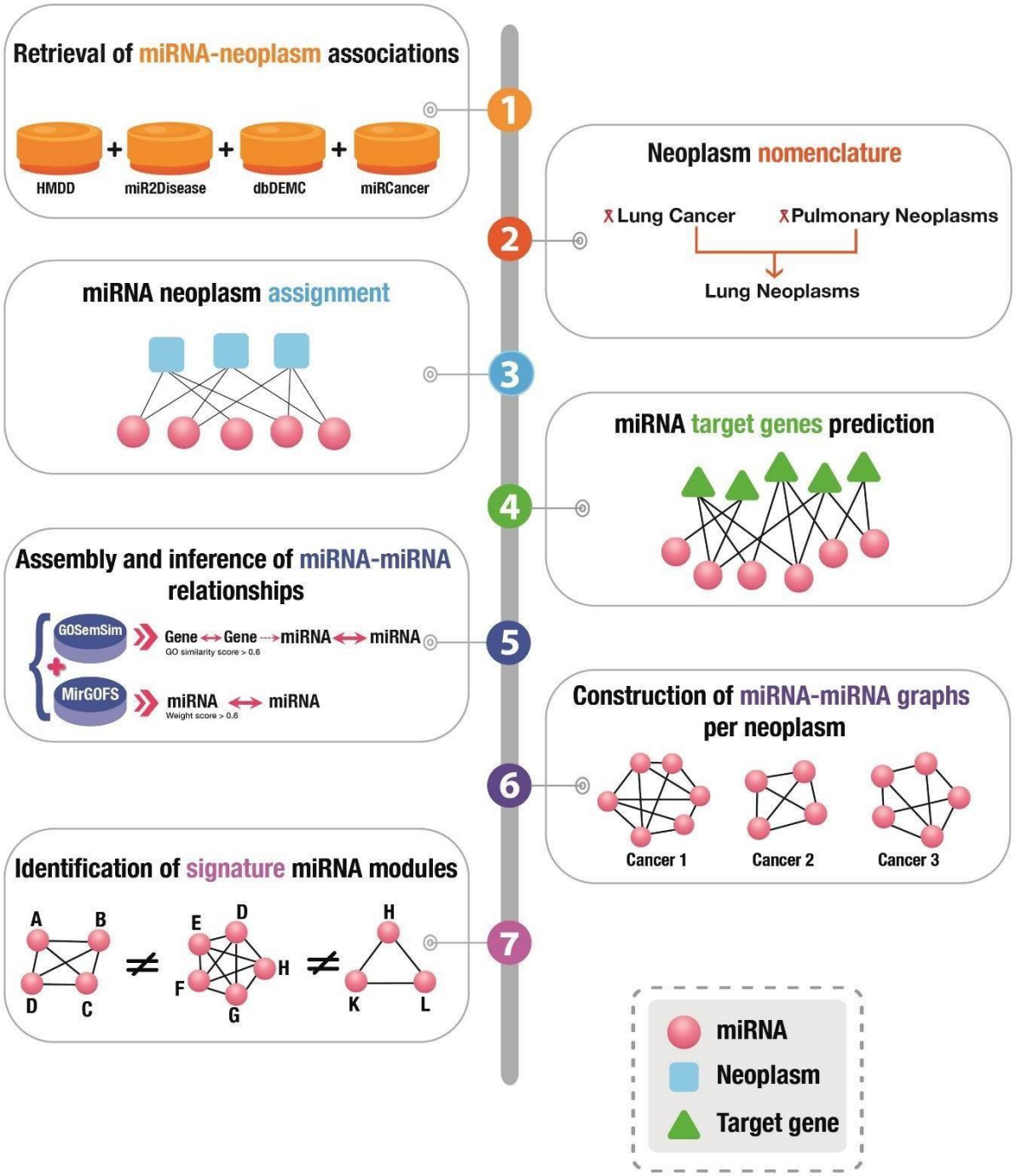


Figure 2. A graphical presentation of the overall workflow of this study. Script 1 was used to extract the required columns from different sources.

Script 1

```
#extrat target columns

library(dplyr)

install.packages("tidyverse")
library(tidyverse)

##miR2disease miR2DiseaseS = miR2Disease[c("miRNA", "CancerType",
"Expression")]       write.csv(miR2DiseaseS,       "G:/IBG/Cancer-
miRNA/miR2DiseaseS.csv")

##miRCancer miRCancerFinal = miRCancer[c("mirId",
"Cancer", "Profile")]
write.csv(miRCancerFinal, "G:/IBG/Cancer-miRNA/miRCancer-Final.csv")

#miRdbDEMC    filter    target    columns    dbdemcFinal    =
dbdemc[c("miRNA_ID", "miRBaseID", "CancerType",
"CancerSubtype", "Status")]
write.csv(dbdemcFinal, "G:/IBG/Cancer-miRNA/dbdemcFinal.csv")
##HMDD
HMDDFinal     =     HMDD[c("mir",     "disease",     "description")]
write.csv(HMDDFinal, "G:/IBG/Cancer-miRNA/hmddFinal.csv")
```

Script 2 was used to extract cancers from the HMDD dataset, remove duplicates from other sources as well, and finally combine the clean data.

Script 2

```
#Filter cancers from HMDD file

Hmddcancers_filtered = hmddFinal %>%

  filter(grepl(" Leukemia| Adenoma| Carcinoma| Myeloma| carcinoma|
Rhabdomyosarcoma|Lymphoma|Cholangiocarcinoma|
  Glioblastoma | Cancer| Glioma| Glioblastoma | Neoplasms | Hamartoma |
Insulinoma| Leiomyoma| Melanoma| Macroglobulinemia| Tumor| Teratocarcinoma
| Sarcoma|Retinoblastoma|Osteosarcoma| Myoma|Neuroblastoma|Meningioma|
  Mesothelioma|Lipoma|Cowden  Syndrome|Medulloblastoma|  Adenocarcinoma"  ,
disease))

write.csv(Hmddcancers_filtered, "G:/IBG/Cancer-
miRNA/hmddCancersFiltered.csv")

#HMDD miRNAs COLUMNS hmdd_mir = Hmddcancers_filtered ["mir"]
hmddmirunique = hmdd_mir[!duplicated(hmdd_mir), ]
write.csv(hmddmirunique,"G:/IBG/Cancer-
miRNA/hmddmirunique.csv")

#miR2Disease miRNAs Unique miR2disease_mir = mir2dis_cancers ["miRNA"]
miR2disease_mir_unique = miR2disease_mir[!duplicated(miR2disease_mir),
] write.csv(miR2disease_mir_unique,"G:/IBG/Cancer-
miRNA/miR2disease_mir_unique.csv")

#dbDEMC  miRNAs  Unique  dbdemce_mir  =
dbdemcFinal ["miRBaseID"]
dbdemce_mir_unique = dbdemce_mir[!duplicated(dbdemce_mir), ]
write.csv(dbdemce_mir_unique,"G:/IBG/Cancer-miRNA/dbdemce_mir_unique.csv")

#filtering mir2dieases: mir2diseases_cancers_filtered
= miR2DiseaseS %>%

  filter(grepl(" leukemia| adenoma| Carcinoma|carcinoma| myeloma|
rhabdomyosarcoma|lymphoma|cholangiocarcinoma|glioblastoma | Cancer| cancer
| glioma| glioblastoma | neoplasms | hamartoma | insulinoma| leiomyoma|
melanoma| macroglobulinemia| tumor| teratocarcinoma |
sarcoma|retinoblastoma|osteosarcoma| myoma|neuroblastoma|meningioma|
mesothelioma|lipoma|Cowden Syndrome|cowden syndrome |medulloblastoma|
adenocarcinoma", CancerType))
```

```
write.csv(mir2diseases_cancers_filtered, "G:/IBG/Cancer-
miRNA/mir2dis_cancers.csv")


#miRCancer miRNAs Unique miRcancer_mir = miRCancer_Final ["mirId"]
mircancer_mir_unique = miRcancer_mir[!duplicated(miRcancer_mir), ]
write.csv(mircancer_mir_unique,"G:/IBG/Cancer-
miRNA/mircancer_mir_unique.csv")

#rbine 4 sources install.packages('qpcR') library(qpcR) combined_mir <-
qpcR:::rbind.na(hmddmirunique, dbdemce_mir_unique_filtred,
mircancer_mir_unique,miR2disease_mir_unique) data_rbind
#rbind 3 sources Combined_three <-
qpcR:::rbind.na(hmddmirunique,mircancer_mir_unique,miR2disease_mir_unique)
final_three = Combined_three[!duplicated(Combined_three), ]
write.csv(final_three,"G:/IBG/Cancer-miRNA/final_three.csv")
```

3.8. Neoplasm nomenclature

The study aimed to ensure that the naming of neoplasms and genes used in the analysis was consistent and accurate. This was achieved by annotating the disease terms and gene symbols according to various standard classification systems and databases.

The International Classification of Diseases and Related Health Problems (ICD), revision 11 (ICD-11) (https://icd.who.int/en) was used for annotating the disease terms (Harrison et al., 2021). ICD-11 is maintained by the World Health Organization (WHO) (https://www.who.int/) and serves as a comprehensive worldwide resource for gathering, analyzing, and organizing diseases in a hierarchical manner. It represents a valuable asset for exploring connections among various diseases and was employed in this research to ensure standardized nomenclature for neoplasms.

In addition to ICD-11, this study also utilized NCBI's Medical Subject Headings (MeSH) (https://www.ncbi.nlm.nih.gov/mesh/) to categorize neoplastic diseases into broader classifications , the study made use of MeSH, which is an extensive hierarchical vocabulary employed for subject indexing, cataloging, and searching biomedical and life sciences information. By utilizing MeSH, the study was able to systematically organize and group neoplastic diseases, contributing to a structured approach in their analysis.

To maintain consistency in gene nomenclature, the study assigned official gene symbols to gene names according to the HUGO Gene Nomenclature Committee (HGNC) (https://www.genenames.org/). The study benefited from the reputable organization HGNC, which offers authoritative gene symbols and names, thereby serving as a valuable resource. Likewise, for the analysis of miRNAs, the study relied on the official miRNA names provided by miRBase. miRBase, an essential database, encompasses a comprehensive collection of miRNA sequences and their corresponding annotations, making it an indispensable resource for the study (Bruford et al., 2021).

In conclusion, the study aimed to ensure accuracy and consistency in the data analysis and interpretation by annotating the neoplasms and genes according to the

ICD-11, MeSH, HGNC, and miRBase, respectively. This helped to ensure that the naming of the neoplasms and genes was consistent across the four databases used in the analysis.

The entries where the miRBaseID contains the strings "mmu" or "rno" were removed by using Script 3. The miRBaseConverter package was used to convert miRNA names between different miRBase versions.

Script 3

```
#miRBAse conversion

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("miRBaseConverter")
library(miRBaseConverter)
data(miRNATest)
miRNANames = miRNATest$miRNA_Name

#Delete    MMu    and    rno    microRNAs    from    unique    dbdemc:
dbdemce_mir_unique_filtred = dbdemce_mir_unique %>%
filter(!grepl("mmu|rno", miRBaseID))

write.csv(dbdemce_mir_unique_filtred, "G:/IBG/Cancer-
miRNA/dbdemce_mir_unique_filtred.csv")


#rbind 1270 miRNAs + 1571 miRNAs, then remove duplicate
Combined_1270_1571 <- qpcR:::rbind.na(final_three,three_sources_1290)
final_three_NA = Combined_1270_1571[!duplicated(Combined_1270_1571), ]
write.csv(final_three_NA,"G:/IBG/Cancer-miRNA/final_three_NA.csv")
```

Script 4 was used to retain only the rows where 'Mesh' contains the cancer names.

Script 4

```
#grep cancers from general dataset
Adenocarcinoma = miRMeshNewfinalRepDel %>% filter(grepl('Adenocarcinoma',
  Mesh))

write.xlsx(Adenocarcinoma,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Adenocarcinoma/Ade
nocarcinomaMiR.xlsx")


#Adrenal
AdrenalGland = miRMeshNewfinalRepDel %>% filter(grepl('^Adrenal',
  Mesh))

write.xlsx(AdrenalGland,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/AdrenalGland/Adren
alGlandMiR.xlsx")

#Biliary Tract Neoplasms
BiliaryTract = miRMeshNewfinalRepDel %>% filter(grepl('^Biliary',
  Mesh))

write.xlsx(BiliaryTract,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/BiliaryTract/Bilia
ryTractdMiR.xlsx")

#Brain Neoplasms
BrainNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Brain
  Neoplasms', Mesh))
write.xlsx(BrainNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Brain/BrainMiR.xls
x")

#Breast Neoplasms
BreastNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Breast
  Neoplasms', Mesh))
write.xlsx(BreastNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Breast/BreastMiR.x
lsx")

#Colorectal Neoplasms
ColorectalNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Colorectal
  Neoplasms', Mesh))
write.xlsx(ColorectalNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Colorectal/Colorec
talMiR.xlsx")
```

```
#Endocrine Neoplasms
EndocrineNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Endocrine
  Gland Neoplasm', Mesh))
write.xlsx(EndocrineNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Endocrine/Endocri
neMiR.xlsx")

#Esophageal Neoplasms
EsophagealNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Esophageal
  Neoplasms', Mesh))
write.xlsx(EsophagealNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Esophageal/Esopha
gealMiR.xlsx")

#Eye Neoplasms
EyeNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Eye
  Neoplasms', Mesh))
write.xlsx(EyeNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Eye/EyeMiR.xlsx")

#Head and Neck Neoplasms
HeadNeckNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Head
  and Neck Neoplasms', Mesh))
write.xlsx(HeadNeckNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/HeadNeck/HeadNeck
MiR.xlsx")

#Hematologic Neoplasms
HematologicNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Hematologic
  Neoplasms', Mesh))
write.xlsx(HematologicNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Hematologic/Hemat
ologicMiR.xlsx")

#Intestinal Neoplasms
IntestinalNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Intestinal
  Neoplasms', Mesh))
write.xlsx(IntestinalNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Intestinal/Intest
inalMiR.xlsx")

#Kidney Neoplasms
KidneyNeoplasms = miRMeshNewfinalRepDel %>%
  filter(grepl('^Kidney', Mesh))
write.xlsx(KidneyNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Kidney/KidneyMiR.
xlsx")
```

```
#Leukemia
Leukemia = miRMeshNewfinalRepDel %>% filter(grepl('Leukemia',
  Mesh))
write.xlsx(Leukemia,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Leukemia/LeukemiaM
iR.xlsx")

#Liver Neoplasms
LiverNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Liver
  Neoplasms', Mesh))
write.xlsx(LiverNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Liver/LiverMiR.xls
x")

#Lung Neoplasms
LungNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Lung',
  Mesh))
write.xlsx(LungNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Lung/LungMiR.xlsx"
)

#Lymphoma
Lymphoma = miRMeshNewfinalRepDel %>% filter(grepl('Lymphoma',
  Mesh))
write.xlsx(Lymphoma,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Lymphoma/LymphomaM
iR.xlsx")

#Nerve Tissue Neoplasms
NerveTissueNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Nerve
  Tissue Neoplasms', Mesh))
write.xlsx(NerveTissueNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/NerveTissue/NerveT
issueMiR.xlsx")

#Ovarian Neoplasms
OvarianNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Ovarian
  Neoplasms', Mesh))
write.xlsx(OvarianNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Ovarian/OvarianMiR
.xlsx")

#Pancreatic Neoplasms
PancreaticNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Pancreatic
  Neoplasms', Mesh))
write.xlsx(PancreaticNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Pancreatic/Pancrea
ticMiR.xlsx")
```

```
#Pleural Neoplasms
PleuralNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Pleural
  Neoplasms', Mesh))
write.xlsx(PleuralNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Pleural/PleuralMiR.x
lsx")

#Prostatic Neoplasms
ProstaticNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Prostatic
  Neoplasms', Mesh))
write.xlsx(ProstaticNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Prostatic/ProstaticM
iR.xlsx")

#Sarcoma
Sarcoma = miRMeshNewfinalRepDel %>% filter(grepl('Sarcoma',
  Mesh))
write.xlsx(Sarcoma,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Sarcoma/SarcomaMiR.x
lsx")

#Skin Neoplasms
SkinNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Skin
  Neoplasms', Mesh))
write.xlsx(SkinNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Skin/SkinMiR.xlsx")

#Squamous Cell Carcinomas
SquamousCarcinomas = miRMeshNewfinalRepDel %>% filter(grepl('Squamous
  Cell Carcinomas', Mesh))
write.xlsx(SquamousCarcinomas,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Squamous/SquamousMiR
.xlsx")

#Stomach Neoplasms
StomachNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Stomach
  Neoplasms', Mesh))
write.xlsx(StomachNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Stomach/StomachMiR.x
lsx")

#Testicular Neoplasms
TesticularNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Testicular
  Neoplasms', Mesh))
write.xlsx(TesticularNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Testicular/Testicula
rMiR.xlsx")
```

```
#Thyroid Neoplasms
ThyroidNeoplasms = miRMeshNewfinalRepDel %>%
  filter(grepl('Thyroid Neoplasms', Mesh))
write.xlsx(ThyroidNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Thyroid/ThyroidMiR.
xlsx")


#Urinary Bladder Neoplasms
UrinaryBladderNeoplasms = miRMeshNewfinalRepDel %>% filter(grepl('Urinary',
  Mesh))
write.xlsx(UrinaryBladderNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/urinary/UrinaryMiR.
xlsx")

#Uterine Neoplasms
UterineNeoplasms = miRMeshNewfinalRepDel %>%
  filter(grepl('Uterine Neoplasms', Mesh))
write.xlsx(UterineNeoplasms,
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Uterine/UterineMiR.
xlsx")
```

3.9. Prediction of miRNA target genes

In this step, the goal was to predict the target genes of the miRNAs under investigation. The identification of these target genes is critical for understanding the role of miRNAs in neoplasm development and progression. To accomplish this, we used several state-of-the-art software tools to retrieve both experimentally supported and predicted mRNA targets of the neoplasm-associated miRNAs.

1) microT_CDS, which has implemented the target prediction algorithm DIANA-microT-CDS (Tastsoglou et al., 2023). This algorithm uses gene expression data and gene ontology information to predict potential target genes for a given miRNA.
2) TargetScan, which searches for conserved sites that match the seed region of each miRNA (Agarwal et al., 2015). Seed regions are critical for determining the specificity of miRNA-target interactions, and TargetScan takes this into account when making predictions.
3) miRDB, which predicts functional miRNA target genes using machine learning models of target gene expression data (Chen & Wang, 2020) By analyzing gene expression data, miRDB is able to determine the likelihood of a given miRNA-target interaction.
4) PicTar, which identifies common microRNA targets using a specific algorithm (Krek et al., 2005). This algorithm is designed to identify targets that are conserved across multiple species, enhancing the robustness of the predictions.

To enhance the accuracy of the predictions, only those target genes that were predicted by more than three of the five tools were included. This approach ensured that the predictions were robust and that the results were less likely to be affected by false positive predictions from a single tool.

3.10. Integration of miRNA-miRNA interactions

The next step in this study involved the identification of functional relationships between miRNAs. To this end, data were gathered from two diverse sources, GOSemSim (Yu, 2020) and MIRGOFs (Yang et al., 2018), which measure functional similarities between miRNAs based on the Gene Ontology (GO) annotations of their corresponding target genes.

GOSemSim employs an information-theoretic approach to determine the similarities between genes, utilizing Gene Ontology (GO) terms. GO terms are utilized to assign functional annotations to genes based on shared characteristics. By identifying the functional similarities between target genes, GOSemSim generates a miRNA-miRNA matrix that reflects the degree of functional similarity among their respective target genes. Higher degrees of gene relatedness correspond to higher association scores for the corresponding miRNAs.

Another resource for assessing functional similarities between miRNAs is MIRGOFs. This tool incorporates two novel features: Firstly, it employs a novel GO semantic similarity metric that considers both common ancestors and descendants of GO terms. Secondly, it calculates the similarity between sets of GO terms asymmetrically and assigns weights to each GO term based on its statistical significance.

After obtaining the miRNA pairwise interactions from both sources, we merged the data and removed duplicate miRNA-miRNA interactions. To ensure robustness in the results, only those miRNA pairwise interactions with a weight score of ≥ 0.6 (where “1” is the highest) are selected for further analysis.

The analysis of the functional relationships between miRNAs provides valuable information about the interactions between miRNAs and how they may be involved in various biological processes. By understanding these relationships, researchers can gain insight into the molecular mechanisms of miRNAs in regulating gene expression and disease development.

The *getPredictedTargets* function from the *GOSemSim* package in R was utilized with Script 5. This package is applied for semantic similarity computation among GO (Gene Ontology) terms, sets of GO terms, gene products, and gene clusters.

Script 5

```
colnames(allmiRs2657) [1]= "ID"
allmiRs <- pull(allmiRs2657,ID)

install.packages("magicfor")
library(magicfor)

options(max.print=999999)


magic_for(print, silent = FALSE)

for (i in allmiRs) { xgenes = getPredictedTargets(i , sources =
c("pictar", "diana", "targetscan","mirdb"), species = "hsa", min_src
= 2, method = "geom", both_strands = FALSE) print(xgenes)

}

result=    magic_result()    capture.output(result1,    file    =
"G:/IBG/Cancer-miRNA/result.csv")
```

The *unionG* function in R (Script 6) was designed to create a network graph from gene ontology and functional similarity data, and then identify a subgraph related to a specific cancer type. It takes two arguments: a data frame and a cancer type. This function creates three graphs from provided and constructed adjacency matrices, merges them, and extracts a subgraph consisting of nodes that are neighbors of the specified cancer type. This subgraph is converted back to a data frame, a new 'average' column is computed, and a subgraph is created that includes edges with an average weight greater than the 90th percentile of the 'average' column. The function ultimately returns this final subgraph and associated statistical measures.

Script 6

```
# In-house R function (unionG) which constructs networks and produces top
10% sub graph

unionG <- function(N,cancerName) { gN =
  graph_from_data_frame(d= N,directed = FALSE)
  networkGo <- graph_from_adjacency_matrix(matrixGo,
weighted=T,mode="undirected", diag=FALSE) networkFs <-
  graph_from_adjacency_matrix(matrixFs, weighted=T,
mode="undirected",diag=FALSE)
  UnionGraph = union(gN,networkGo,networkFs, byname = T)
  PureG <- induced_subgraph(UnionGraph, neighbors(UnionGraph,cancerName))
  PureGDf = get.data.frame(PureG)
  PureGDf$average = rowMeans(PureGDf[,c('weight_2', 'weight_3')],
na.rm=TRUE)
  Percentile <-quantile(PureGDf$average,probs = 0.90 ,na.rm = TRUE)
  PureAv = PureGDf [,-3:-4]
  PureAvG = graph_from_data_frame(d= PureAv, directed = FALSE)
  subgraphPure<- subgraph.edges(PureAvG, which(E(PureAvG)$average >
Percentile), delete.vertices = TRUE) subgraphPureDf
  = get.data.frame(subgraphPure)
  return(list(subgraphPureDf, Percentile,PureGDf,PureAv)) }
```

3.11. Construction of miRNA-neoplasm networks

In order to construct miRNA-neoplasm networks for each group of neoplasms, two weight scores were assigned to the links in the network. The initial weight score utilized in this study was derived from the MIRGOFs algorithm, developed by (Yang et al., 2018). The MIRGOFs algorithm assesses the functional similarities between miRNAs by analyzing the Gene Ontology (GO) annotations of their target genes. Notably, this algorithm incorporates a novel GO semantic similarity metric that considers both common ancestors and descendants of GO terms. Additionally, it computes the similarity between sets of GO terms in an asymmetric manner and assigns weights to each GO term based on its statistical significance.

The secondary weight score was computed based on a miRNA matrix generated using the GOSemSim algorithm, which was developed by (Yu, 2020). This algorithm employs an information-theoretic measurement to identify similarities between genes by considering their shared Gene Ontology (GO) terms. By assigning functions to genes based on their shared GO annotations, the GOSemSim algorithm constructs a

miRNA-miRNA matrix that reflects the degree of functional similarity among their respective target genes. Consequently, higher levels of gene relatedness correspond to increased association scores for the corresponding miRNAs.

The topological features of the generated networks were then investigated using the Cytoscape software platform, version 3.9.1, developed by (Shannon et al., 2003). The software platform utilized in this study serves as a robust resource for network manipulation and visualization, making it highly suitable for the analysis conducted. Furthermore, Arena3Dweb, a web-based interactive tool developed by (Karatzas et al., 2021), was employed for the visualization of multilayer networks in three-dimensional space. This tool facilitates the depiction of intricate networks in a user-friendly and interactive manner, contributing valuable insights to the analysis.

Script 7 was used to employ the unionG function to create subgraphs for each cancer

Script 7

```
Using the function unionG and igraph R package install.packages("igraph")

#Adenocarcinoma
AdenocarcinomaPair = unionG(Adenocarcinoma, 'Adenocarcinoma')
write.xlsx(AdenocarcinomaPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Adenocarcinoma/Aden
ocarcinomaPair.xlsx")

#AdrenalGland
AdrenalGlandPair = unionG(AdrenalGlandMiR,'Adrenal')
write.xlsx(AdrenalGlandPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/AdrenalGland/Adrena
lGlandPair.xlsx")

#Biliary Tract Neoplasms
BiliaryTractPairTest = unionG(BiliaryTractdMiR,'Biliary')
write.xlsx(BiliaryTractPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/BiliaryTract/Biliar
yTractPair.xlsx")

#Brain Neoplasms
BrainNeoplasmsPair = unionG(BrainNeoplasms,'Brain Neoplasms')
write.xlsx(BrainNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Brain/BrainNeoplasm
sPair.xlsx")
```

```
#Breast Neoplasms
BreastNeoplasmsPair = unionG(BreastNeoplasms,'Breast Neoplasms')
write.xlsx(BreastNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Breast/BreastNeop
lasmsPair.xlsx")

##Colorectal Neoplasms
ColorectalNeoplasmsPair =
unionG(ColorectalNeoplasms,'Colorectal Neoplasms')
write.xlsx(ColorectalNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Colorectal/Colore
ctalNeoplasmsPair.xlsx")

#Endocrine Gland Neoplasms
EndocrineNeoplasmsPair = unionG(EndocrineNeoplasms,'Endocrine
Gland Neoplasm') write.xlsx(EndocrineNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Endocrine/Endocri
neNeoplasmsPair.xlsx")

#Esophageal Neoplasms
EsophagealNeoplasmsPair =
unionG(EsophagealNeoplasms,'Esophageal Neoplasms')
write.xlsx(EsophagealNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Esophageal/Esopha
gealNeoplasmsPair.xlsx")

#Eye Neoplasms
EyeNeoplasmsPair = unionG(EyeNeoplasms,'Eye Neoplasms')
write.xlsx(EyeNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Eye/EyeNeoplasmsP
air.xlsx")

#Head and Neck
HeadNeckNeoplasmsPair = unionG(HeadNeckNeoplasms,'Head and
Neck Neoplasms') write.xlsx(HeadNeckNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/HeadNeck/HeadNeck
NeoplasmsPair.xlsx")

#Hematologic Neoplasms
HematologicNeoplasmspair = unionG(HematologicNeoplasms,'Hematologic
Neoplasms')
write.xlsx(HematologicNeoplasmspair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Hematologic/Hemat
ologicNeoplasmspair.xlsx")
```

```
#Intestinal Neoplasms
IntestinalNeoplasmsPair =
unionG(IntestinalNeoplasms,'Intestinal Neoplasms')
write.xlsx(IntestinalNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Intestinal/Intestin
alNeoplasmsPair.xlsx")

#Kidney Neoplasms
KidneyNeoplasmspair = unionG(KidneyMiR,'Kidney')
write.xlsx(KidneyNeoplasmspair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Kidney/KidneyNeopla
smspair.xlsx")

#Leukemia
LeukemiaPair = unionG(Leukemia,'Leukemia')
write.xlsx(LeukemiaPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Leukemia/LeukemiaPa
ir.xlsx")

#LiverNeoplasms
LiverNeoplasmsPair = unionG(LiverNeoplasms,'Liver Neoplasms')
write.xlsx(LiverNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Liver/LiverNeoplasm
sPair.xlsx")

#Lung Neoplasms
LungNeoplasmsPair = unionG(LungMiR,'Lung')
write.xlsx(LungNeoplasmsPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Lung/LungNeoplasmsP
air.xlsx")

#Lymphoma
LymphomaPair = unionG(Lymphoma,'Lymphoma')
write.xlsx(LymphomaPair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/Lymphoma/LymphomaPa
ir.xlsx")


#Nerve Tissue Neoplasms
NerveTissuePair = unionG(NerveTissueNeoplasms,'Nerve Tissue Neoplasms')
write.xlsx(NerveTissuePair[[1]],
"Desktop/IBG/cancer/NewCancer_Categorized/EditedVersion/NerveTissue/NerveTi
ssuePair.xlsx")
```

3.12. Identification of mutually exclusive cancer-relevant miRNA modules

The goal of this study was to identify the mutual exclusivity of miRNAs in different groups of neoplasms. To achieve this, a systematic and comprehensive approach was taken.

Firstly, unique miRNAs were detected in each group of neoplasms. This involved identifying the specific miRNAs that were present in each group and ensuring that they were not present in any other group. This step was crucial because it helped to establish a baseline for the mutual exclusivity analysis.

The structure of each miRNA-neoplasm network was then analyzed using the igraph package in R. The node degree distribution and clustering coefficient were calculated for each network (a total of thirty graphs). The node degree distribution provides information about the number of links (or edges) that are connected to each node (miRNA) in the network. This information helps to understand the relative importance of each miRNA in the network and its relationship with other miRNAs and neoplasms. The clustering coefficient, on the other hand, estimates the tendency of nodes in a graph to cluster together. This information helps to understand the overall organization of the network and the relationships between the miRNAs and neoplasms.

A simultaneous comparison was then performed between the thirty miRNA-neoplasm networks to find modules of miRNAs that are included in one neoplasm but not in the others. This was achieved using an in-house R script that was specifically designed to handle the complexity of the datasets. The script organized each dataset into a hash and then detected the mutually exclusive modules of miRNAs. This step was important because it helped to identify the miRNAs that were unique to each group of neoplasms and not present in other groups.

The results of this analysis provided valuable insights into the functional relationships between miRNAs and neoplasms, as well as the mutual exclusivity of miRNAs in different groups of neoplasms. This information can be useful for developing new treatments and therapies for neoplasms by targeting specific miRNAs. By understanding the mutual exclusivity of miRNAs in different neoplasms, researchers can identify which miRNAs are the most promising targets for therapeutic interventions. This information is critical for developing more effective and targeted treatments for

neoplasms and improving patient outcomes. Scrip 8 was used to detect mutually exclusive miRNAs.

Script 8.1

```
#first we read the mirnas list from each txt file with the name of the
neoplasm and convert to python list
```

```
Nerve=open("Nerve.txt").read().splitlines()
adenocarcinoma=open("adenocarcinoma.txt").read().splitlines()
adrenalgland=open(adrenalgland").read().splitlines()
biliary=open("biliary.txt").read().splitlines()
brain=open("brain.txt").read().splitlines()
breast=open("breast.txt").read().splitlines()
colorectal=open("colorectal").read().splitlines()
endocrine=open("endocrine.txt").read().splitlines()
esophageal=open("esophageal.txt").read().splitlines()
eye=open("eye.txt").read().splitlines()
headandneck=open("headandneck.txt").read().splitlines()
hematologic=open("hematologic.txt").read().splitlines()
intestinal=open("intestinal.txt").read().splitlines()
kidney=open("kidneyt.txt").read().splitlines()
leukemia=open("leukemia.txt").read().splitlines()
liver=open("liver.txt").read().splitlines()
lung=open("lung.txt").read().splitlines()
lymphoma=open("lymphoma.txt").read().splitlines()
ovarian=open("ovarian.txt").read().splitlines()
pancreatic=open("pancreatic.txt").read().splitlines()
pleural=open("pleural.txt").read().splitlines()
prostatic=open("prostatic.txt").read().splitlines()
sarcoma=open("sarcoma.txt").read().splitlines()
skin=open("skin.txt").read().splitlines()
squamous=open("squamous.txt").read().splitlines()
stomach=open("stomach.txt").read().splitlines()
testicular=open("testicular,txt").read().splitlines()
thyroid=open("thyroid.txt").read().splitlines()
urinary=open("urinary.txt").read().splitlines()
uterine=open("uterine.txt").read().splitlines()
```

Script 8.2

```
 #here in the mutually_exclusive fuction we place the lists we want to
compare and find the mutually_exclusive mirnas.
 Def mutually_exclusive
 (Nerve,adenocarcinoma,adrenalgland,biliary,brain,breast,
 colorectal,endocrine,esophageal,eye,headandneck,hematologic,intestinal,kid
 ney,
 leukemia,liver,lung,lymphoma,ovarian,pancreatic,pleural,prostatic,sarcoma,
 skin, squamous,stomach,testicular,thyroid,urinary,uterine):
     return
 list(set(Nerve)^set(adenocarcinoma)^set(adrenalgland)^set(biliary)^
 set(brain)^set(breast)^set(colorectal)^set(endocrine)^set(esophageal)^set(
 eye)^
 set(headandneck)^set(hematologic)^set(intestinal)^set(kidney)^set(leukemia
 )^
 set(liver)^set(lung)^set(lymphoma)^set(ovarian)^set(pancreatic)^set(pleura
 l)^
 set(prostatic)^set(sarcoma)^set(skin)^set(squamous)^set(stomach)^set(testi
 cular)^ set(thyroid)^set(urinary)^set(uterine))
```

3.13. Predictive performance of machine learning algorithms for miRNA data classification

In this study, various machine learning algorithms were implemented using the scikit-learn library to predict and classify miRNA data. The data was split into training and testing sets using the *train_test_split* function from the *sklearn.model_selection* module. The following algorithms were utilized: Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Logistic Regression.

Random Forest, Gradient Boosting, SVM, and Logistic Regression are popular machine learning algorithms that have been widely used in various domains, including cancer research and diagnostics. Each algorithm has its own characteristics and strengths when it comes to classification and prediction tasks.

Random Forest (Rigatti, 2017) is an ensemble learning technique that combines multiple decision trees to make predictions. By training each tree on a different subset of the data and using a random selection of features, Random Forest can handle high-dimensional datasets and mitigate overfitting. It excels at capturing complex relationships and interactions between variables, making it applicable in various domains.

Gradient Boosting (Griesbach et al., 2021), on the other hand, is another ensemble learning method that sequentially builds an ensemble of weak prediction models, such as decision trees. Each subsequent model is trained to correct the errors of the previous models. Gradient Boosting is adept at capturing intricate patterns and achieving high predictive accuracy. It is particularly beneficial when dealing with imbalanced datasets and supports multiple data types.

SVM, or Support Vector Machines (Huang et al., 2018), is a powerful algorithm used for both classification and regression tasks. Its objective is to find the optimal hyperplane that separates data points belonging to different classes with the widest margin. SVMs can effectively handle high-dimensional data and can handle cases where the data is not linearly separable by transforming it into a higher-dimensional space. The use of kernel functions further enhances SVMs' ability to handle non-linear relationships.

Logistic Regression (Domínguez-Almendros et al., 2011), a statistical algorithm, is commonly employed for binary classification problems. It models the relationship between input features and the probability of belonging to a specific class using a logistic function. Logistic Regression is widely adopted due to its simplicity, interpretability, and ability to handle both categorical and continuous input features. It provides insights into the significance of different features and their impact on the outcome.

XGBoost, which stands for Extreme Gradient Boosting (Wang et al., 2022), is an ensemble learning algorithm that sequentially combines multiple weak prediction models, typically decision trees. It follows a gradient boosting framework and incorporates optimization techniques to enhance model performance. XGBoost is highly proficient at capturing complex relationships and interactions within the data, enabling accurate predictions. It is recognized for its scalability, resilience against overfitting, and effectiveness in handling both numerical and categorical features.

In contrast, LightGBM (Liao et al., 2022) is a gradient boosting framework developed by Microsoft. Like XGBoost, LightGBM constructs an ensemble of decision trees in a sequential manner. However, it introduces a novel technique called Gradient-based One-Side Sampling (GOSS). GOSS selects a subset of data instances based on their gradients, resulting in faster training without compromising predictive accuracy. LightGBM is particularly efficient for large-scale datasets as it utilizes a histogram-based algorithm for quicker computation and reduced memory usage.

MLP, which stands for Multi-Layer Perceptron (Bhattacharya et al., 2022) , is a type of artificial neural network consisting of interconnected nodes (neurons) arranged in multiple layers. It is a versatile algorithm used for various machine learning tasks, including classification. MLP excels in capturing intricate non-linear relationships in the data through forward propagation and backpropagation techniques. It can learn and adapt to complex patterns, making it suitable for analyzing miRNA data. MLP models can have multiple hidden layers, enabling them to learn hierarchical representations of the data.

Although these algorithms have been successful in various domains, including cancer research, their performance can vary depending on the specific dataset and problem at hand. Factors such as data quality, feature engineering, hyperparameter

tuning, and dataset size can all influence the performance of these algorithms. It is essential to experiment with different algorithms and compare their results to determine the most suitable approach for a given problem.

In the context of the research conducted, it was observed that the performance of Random Forest, Gradient Boosting, SVM, and Logistic Regression did not yield satisfactory results. The accuracy scores obtained indicated suboptimal performance for these algorithms. Therefore, alternative approaches, such as network-based methods, were employed to capture the complex relationships and patterns present in the miRNA data, which ultimately led to more meaningful and valuable results.

The accuracy scores achieved were as follows: Random Forest Accuracy: 0.2597, Gradient Boosting accuracy: 0.22, SVM accuracy: 0.15, and Logistic Regression accuracy: 0.23, XGBoost accuracy: 0.24, LightGBM accuracy: 0.25 and MLP accuracy: 0.25 (Figure 3).

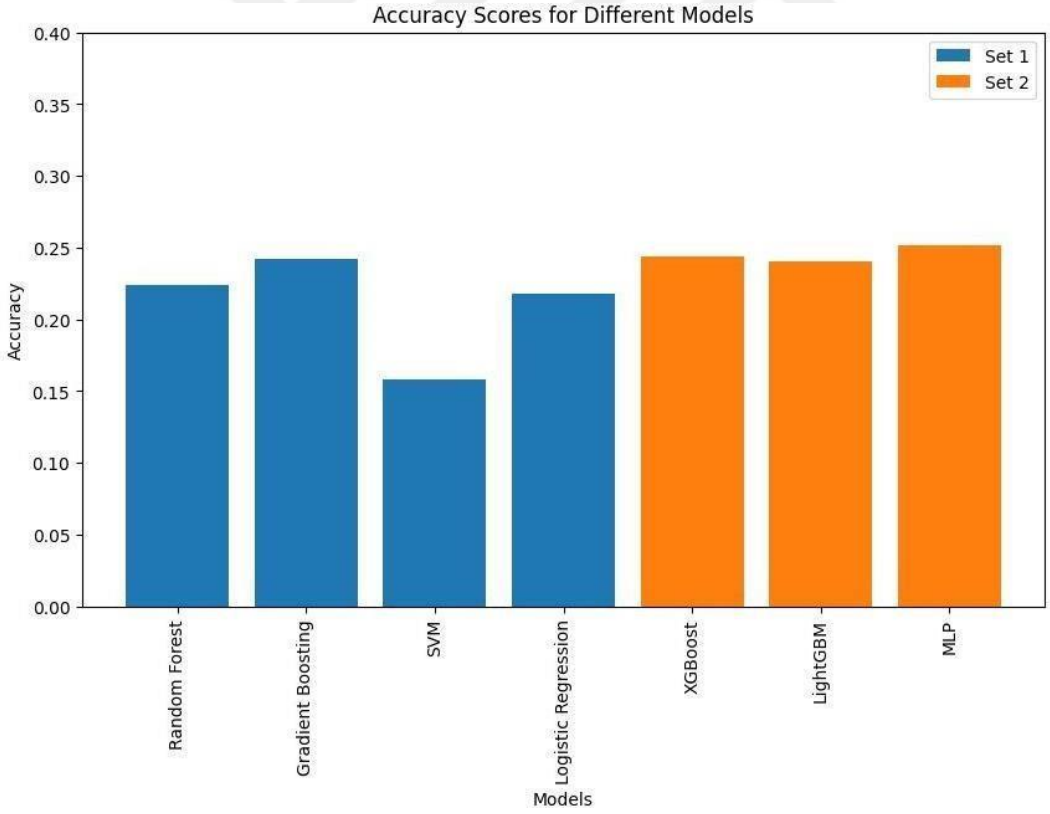


Figure 3. Bar graph depicting the accuracy scores of various machine learning models.

Python Script 9.1 was used to import the required libraries for machine learning training.

Script 9.1

```
import pandas as pd import warnings from
sklearn.model_selection import train_test_split from
sklearn.preprocessing import OneHotEncoder from
sklearn.ensemble import RandomForestClassifier,
GradientBoostingClassifier from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression from
sklearn.metrics import accuracy_score import random
warnings.filterwarnings('ignore', category=FutureWarning)

#importing the data df =
pd.read_csv('G:\\IBG\\RezaThesis\\NeoplasmMirSequence.csv')

# Preprocessing
X = df['Sequence'] # Input features (miRNA sequences) y
= df['Neoplasm'] # Target variable (neoplasm types)
# Data augmentation -
augmented_X = []
augmented_y = []

for sequence, label in zip(X, y):
    augmented_X.append(sequence)
    augmented_y.append(label) for
    _ in range(10):
        add_sequence = list(sequence) for
        i in range(len(add_sequence)):
            if random.random() < 0.1: ate with a probability of 0.1
                add_sequence[i] = random.choice(['A', 'C', 'G', 'U'])
        augmented_X.append(''.join(add_sequence))
        augmented_y.append(label)
# Perform one-hot encoding on the augmented miRNA sequences
encoder = OneHotEncoder(sparse=False, dtype=int) X_encoded
=
encoder.fit_transform(pd.Series(augmented_X).str.split('').apply(pd.Series)
.drop(columns=0))

# Split the augmented data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded, augmented_y,
test_size=0.2, random_state=42)
```

Python Script 9.2 was used for creating the Random Forest, Gradient Boosting, SVM and Logistic Regression models

Script 9.2

```
 #Random Forest rf_model = RandomForestClassifier()
rf_model.fit(X_train, y_train) rf_predictions =
rf_model.predict(X_test) rf_accuracy =
accuracy_score(y_test, rf_predictions) print("Random
Forest Accuracy:", rf_accuracy)
 #Gradient Boosting gb_model =
GradientBoostingClassifier() gb_model.fit(X_train,
y_train) gb_predictions = gb_model.predict(X_test)
gb_accuracy = accuracy_score(y_test, gb_predictions)
print("Gradient Boosting Accuracy:", gb_accuracy)
 #Support Vector Machine (SVM) svm_model = SVC()
svm_model.fit(X_train, y_train) svm_predictions =
svm_model.predict(X_test) svm_accuracy =
accuracy_score(y_test, svm_predictions) print("SVM
Accuracy:", svm_accuracy)
 #Logistic Regression lr_model =
 LogisticRegression() lr_model.fit(X_train, y_train)
 lr_predictions = lr_model.predict(X_test)
 lr_accuracy = accuracy_score(y_test,
 lr_predictions) print("Logistic Regression
 Accuracy:", lr_accuracy)


 Random Forest Accuracy: 0.2298507462686567
 Gradient Boosting Accuracy: 0.2
 SVM Accuracy: 0.15522388059701492
 Logistic Regression Accuracy: 0.2537313432835821
```

Python Script 9.3 was used to import required libraries for the machine learning models.

Script 9.3

```
import xgboost as xgb import
lightgbm as lgb
from sklearn.neural_network import MLPClassifier

# Load the dataset
df = pd.read_csv('G:\\IBG\\RezaThesis\\NeoplasmMirSequence.csv')
# Preprocessing
X = df['Sequence'] # Input features (miRNA sequences)
y = df['Neoplasm'] # Target variable (neoplasm types)
# Data augmentation - random mutations
augmented_X = [] augmented_y = []

for sequence, label in zip(X,
    y):
    augmented_X.append(sequence)
    augmented_y.append(label) for
    _ in range(100):
        add_sequence = list(sequence) for
        i in range(len(add_sequence)):
            if random.random() < 0.1:
                add_sequence[i] = random.choice(['A', 'C', 'G', 'U'])
        augmented_X.append(''.join(add_sequence))
        augmented_y.append(label)

# Perform one-hot encoding on the augmented miRNA sequences
encoder = OneHotEncoder(sparse=False, dtype=int) X_encoded
=
encoder.fit_transform(pd.Series(augmented_X).str.split('').apply(pd.Series
).drop(columns=0))

# Convert target variable to categorical
y_categorical = pd.Categorical(augmented_y)
# Split the augmented data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded,
y_categorical.codes, test_size=0.2, random_state=42)
```

Python Script 9.4 was used for the creation of XGBoost, LightGBM and MLP models .

Script 9.4

```
 #XGBoost      xgb_model      =      xgb.XGBClassifier()
xgb_model.fit(X_train,   y_train)   xgb_predictions   =
xgb_model.predict(X_test)         xgb_accuracy         =
accuracy_score(y_test,  xgb_predictions)  print("XGBoost
Accuracy:", xgb_accuracy)
 #LightGBM lgb_model = lgb.LGBMClassifier()
 lgb_model.fit(X_train, y_train) lgb_predictions =
 lgb_model.predict(X_test) lgb_accuracy =
 accuracy_score(y_test, lgb_predictions)
 print("LightGBM Accuracy:", lgb_accuracy)

 #Multi-layer    Perceptron    (Neural    Network)    mlp_model    =
 MLPClassifier(hidden_layer_sizes=(100,),             max_iter=1000)
 mlp_model.fit(X_train,       y_train)       mlp_predictions       =
 mlp_model.predict(X_test)  mlp_accuracy  =  accuracy_score(y_test,
 mlp_predictions) print("MLP Accuracy:", mlp_accuracy)


 XGBoost Accuracy: 0.25236079452946925
 LightGBM Accuracy: 0.24910452621295995
 MLP Accuracy: 0.2448713774014979
```

Script 10 Python code was used to generate a bar graph using the matplotlib library in Python to visually represent the accuracy scores of various machine learning models.

Script 10

```
import matplotlib.pyplot as plt
#Accuracy values for the first set of models
models1 = ['Random Forest', 'Gradient Boosting', 'SVM', 'Logistic
Regression'] accuracies1 = [0.22388059701492538, 0.2417910447761194,
0.1582089552238806, 0.21791044776119403]
#Accuracy values for the second set of models models2 =
['XGBoost',    'LightGBM',    'MLP']    accuracies2    =
[0.24356887007489417,              0.2406382285900358,
0.25170954086616737]

#Plotting accuracy scores plt.figure(figsize=(10,
6))

#First set of models
plt.bar(models1, accuracies1, label='Set 1')
#Second set of models
plt.bar(models2, accuracies2, label='Set 2')

plt.ylim(0, .4) # Set y-axis limits
plt.xlabel("Models") plt.ylabel("Accuracy")
plt.title("Accuracy Scores for Different Models")
plt.xticks(rotation='vertical') plt.legend()
#Show the plot
plt.show()
```

Several factors may have contributed to the suboptimal performance of these machine learning methods in this specific context.

First and foremost, it is crucial to consider the unique characteristics of miRNA data. MiRNA sequences are relatively short and highly intricate, hence posing challenges for traditional machine learning algorithms in identifying meaningful patterns and relationships. These algorithms may struggle to extract relevant features or adequately model the complex interplay among miRNAs.

Furthermore, the limited size of the dataset could have impacted the performance of the algorithms. Machine learning models typically require substantial

amounts of data to generalize well and avoid overfitting. Insufficient data may result in decreased model robustness and generalizability.

In contrast, the network-based approaches employed in this study demonstrated greater effectiveness in yielding meaningful outcomes. Several factors contribute to this advantage. Firstly, network-based approaches excel in capturing the intricate relationships and interactions among miRNAs. The biology of miRNAs involves intricate regulatory mechanisms, where miRNAs collaborate within interconnected modules. Network analysis enables a comprehensive exploration of these interactions, offering a holistic perspective on miRNA behavior and revealing crucial regulatory patterns.

Additionally, network-based approaches incorporate various factors, such as node degree and centrality, which contribute to the detection of significant miRNA modules. By considering the connectivity and importance of miRNAs within the network, these approaches are better able to identify modules that exhibit mutual exclusivity and possess a certain degree of autonomy. Such modules hold significant biological relevance and offer valuable insights into neoplastic diseases.

Through the utilization of domain-specific knowledge and the unique characteristics of miRNA networks, the network-based approaches employed in this study facilitated a more specialized and targeted analysis. These approaches surpass the limitations of traditional machine learning algorithms by capturing the intricate relationships and dependencies present in miRNA data.

Therefore, due to the inherent complexity and regulatory nature of miRNAs, the utilization of network-based approaches offered a more effective means of identifying miRNA modules and gaining valuable insights from the respective data. These approaches leverage the interconnectedness of miRNAs and consider factors such as node degree and centrality, enabling a more comprehensive understanding of neoplasm-specific miRNA behavior and facilitating the detection of important regulatory modules.

4. RESULTS

The goal of this study was to investigate the functional connections between miRNAs and various groups of neoplastic diseases. It is well established that functionally linked diseases often share common epigenetic mechanisms to regulate the genes that are responsible for causing the disease. MiRNAs play a crucial role in gene regulation and they often work together to regulate multiple target genes in neoplastic diseases. This leads to the formation of densely connected communities of miRNAs in the disease regulome.

This study aimed to uncover the interactions between miRNAs in different groups of neoplastic diseases, with a focus on the functional relationships of their target genes. By exploring the miRNA-miRNA interaction networks for each group of neoplasms, the researchers aimed to identify core sets of miRNAs that could potentially be used as signatures for each type of cancer.

To achieve this goal, the study aimed to find tightly interconnected modules of miRNAs in each neoplasm that have the highest possible degree of mutual exclusivity, based on network topological properties. This involved analyzing the structure of each miRNA-neoplasm network, calculating the node degree distribution and clustering coefficient, and performing a simultaneous comparison between the networks to identify mutually exclusive modules of miRNAs. The results of this analysis would provide valuable insights into the functional relationships between miRNAs and neoplasms, as well as the mutual exclusivity of miRNAs in different groups of neoplasms.

This information can be useful for developing new treatments and therapies for neoplasms by targeting specific miRNAs. By identifying the core sets of miRNAs that are unique to each type of cancer, researchers can develop targeted therapies that are more effective and less harmful than current treatments. Overall, the study aimed to make significant contributions to the field of cancer research by uncovering the interactions between miRNAs in different groups of neoplastic diseases and identifying potential signatures for each type of cancer.

4.1. Neoplasm classification

In this study, all neoplasms were categorized into thirty major groups in order to allow for a more organized and focused analysis. This process involved two steps. Firstly, the ICD-11 standard hierarchy for medical terms was utilized to classify all neoplasms. This provided a standardized framework for organizing the diseases into distinct groups.

Next, the MeSH vocabulary was used to assign the diseases to broader classes. The MeSH vocabulary is a comprehensive, controlled vocabulary that is used in the medical and scientific fields to categorize and index articles, books, and other resources. The ICD-11 terms for each neoplasm were assigned to a corresponding MeSH term in order to further classify the diseases into broader categories.

For example, if the ICD-11 term for a particular neoplasm was ‘adenocarcinoma of lung’, this would be assigned to the broader MeSH term ‘adenocarcinoma’. This process was repeated for all neoplasms, resulting in a clear and organized classification system.

It is important to note that ambiguous terms were removed from the analysis during this process. This ensured that the data used in the subsequent stages of the analysis was accurate and reliable, and would not affect the results of the study. The use of both the ICD-11 standard hierarchy and the MeSH vocabulary allowed for a comprehensive and organized categorization of all neoplasms, which was crucial for the success of this study.

4.2. Construction of miRNA-miRNA graphs per neoplasm

The construction of the miRNA-neoplasm bipartite graphs in this study was based on a comprehensive integration of experimentally known miRNA-cancer associations that were obtained from four diverse resources. These resources were selected carefully to provide the most accurate and comprehensive data on the interactions between miRNAs and different types of neoplasms.

The miRNA-miRNA interactions for each of the thirty neoplasms were inferred from two sources, the direct and indirect method. The direct method relied on the

miRGOFS database which contains pairwise miRNA associations based on the functional relationships between the miRNA target genes. These relationships were determined using shared GO terms. On the other hand, the indirect method involved detecting miRNA targets through the consensus of multiple miRNA-gene interaction databases, including both experimentally supported and predicted interactions.

To determine the degree of association between two miRNAs, the study utilized the GOSemSim tool. This tool calculated the similarities between each pair of miRNAs based on the functional annotation of their target genes in the Gene Ontology (GO) database. The more semantically related two GO terms were, the higher the degree of association between the corresponding genes.

Finally, the miRNA-miRNA pairs from both sources were combined and any duplicates were removed. This resulted in a total of 2604 unique miRNAs detected across the thirty neoplasms (Table 1), which formed the foundation for constructing the miRNA-neoplasm bipartite graphs and analyzing the miRNA-miRNA interaction networks in each group of neoplasms.

Table 1. Major classes of neoplasms, neoplasm-associated miRNAs and miRNA-miRNA interactions.

Neoplasm	Unique miRNAs	miRNA pairs
Adenocarcinoma	551	51455
Adrenal Gland Neoplasms	431	22150
Biliary Tract Neoplasms	1050	147665
Brain Neoplasms	1679	267717
Breast Neoplasms	1762	288200
Colorectal Neoplasms	1821	304757
Endocrine Gland Neoplasm	39	241
Esophageal Neoplasms	1806	301059
Eye Neoplasms	125	2885
Head and Neck Neoplasms	1249	155197
Hematologic Neoplasms	54	492
Intestinal Neoplasms	403	28209
Kidney Neoplasms	1252	152880
Leukemia	286	15275
Liver Neoplasms	1463	266584

Lung Neoplasms	1736	282743
Lymphoma	1330	177172
Nerve Tissue Neoplasms	211	7005
Ovarian Neoplasms	1404	245655
Pancreatic Neoplasms	1701	272487
Pleural Neoplasms	112	2928
Prostatic Neoplasms	1110	164209
Sarcoma	1378	235334
Skin Neoplasms	1197	143565
Squamous Cell Carcinomas	884	118173
Stomach Neoplasms	1660	260636
Testicular Neoplasms	178	4690
Thyroid Neoplasms	1108	126140
Urinary Bladder Neoplasms	1528	224985
Uterine Neoplasms	329	18475

The study identified a group of miRNAs that were found in all thirty types of cancers analyzed. These miRNAs are hsa-miR-101-3p, hsa-miR-145-5p, and hsa-miR-223-3p. In addition, the researchers also found two well-studied miRNAs, miR-21

and miR-34a, in 29 out of the 30 types of cancer, with the exception of testicular cancers.

The study further identified fifteen unique miRNAs that were found in ten specific types of cancers. These miRNAs are: hsa-miR-291-5p in brain cancers, hsa-miR-103a-1-5p in colorectal cancers, hsa-miR-6836-3p in head and neck cancers, hsa-miR-199a-1-5p and hsa-miR-199a-2-5p in liver cancers, hsa-miR-886-5p, hsa-miR-886-3p, and hsa-mir-320a-3p in lung cancers, hsa-miR-526a-5p in ovarian cancers, hsa-miR-3689c in skin cancers, hsa-miR-7156-3p in stomach cancers, hsa-miR-5686, hsa-miR-199a-3p, and hsa-miR-199b-3p in thyroid cancers, and hsa-miR-498-3p in urinary bladder cancers.

The study analyzed the networks of thirty different types of cancers and found that the networks are highly dense, with an average estimated clustering coefficient of approximately 0.89, where 1 represents the highest value. This result suggests that the nodes in the networks tend to cluster together and form close connections with each other. Despite the stringent criteria used to determine the connections, the number of significant links remains large, which makes it difficult to visually identify any distinct modules within the networks.

Based on topological analyses, the node degrees, or the number of connections of a given node to its neighboring nodes, in all thirty networks follow a random Poisson distribution. This is in contrast to biological networks, which are typically scale-free and dominated by a few highly connected hubs. The "promiscuous" nature of miRNAs, which can potentially target numerous genes, likely results in a large number of corresponding miRNA connections.

These findings indicate that the dense and complex nature of the miRNA networks in different cancers can make it difficult to identify distinct modules and patterns, but also highlight the potential role that miRNAs may play in the development and progression of cancer.

In order to create more robust and less dense miRNA-miRNA interaction networks, we focused on the strongest interactions, which were identified by selecting only the links with weights above the 90th percentile for each network. This

resulted in the use of the highest scoring miRNA-miRNA interactions for network construction and subsequent module detection.

After this process, the number of unique miRNAs across all neoplasms was reduced to 1,893. The number of miRNAs per neoplasm varied greatly, ranging from 39 in endocrine gland neoplasms to 1,821 in colorectal neoplasms. The pruning process also significantly reduced the number of pairwise miRNA-miRNA associations.

It is important to note that this reduction was more pronounced in the more populated networks, as the number of potential miRNA connections increases rapidly with an increase in the number of miRNAs. This can be seen in the endocrine gland neoplasm miRNA network, where the use of only the upper 10th percentile of miRNA-miRNA links resulted in a less dense network with a reduced number of nodes (miRNAs) and connections (miRNA relations) (Figure 4). This highlights the effectiveness of the researchers' approach in creating more robust and less dense networks.

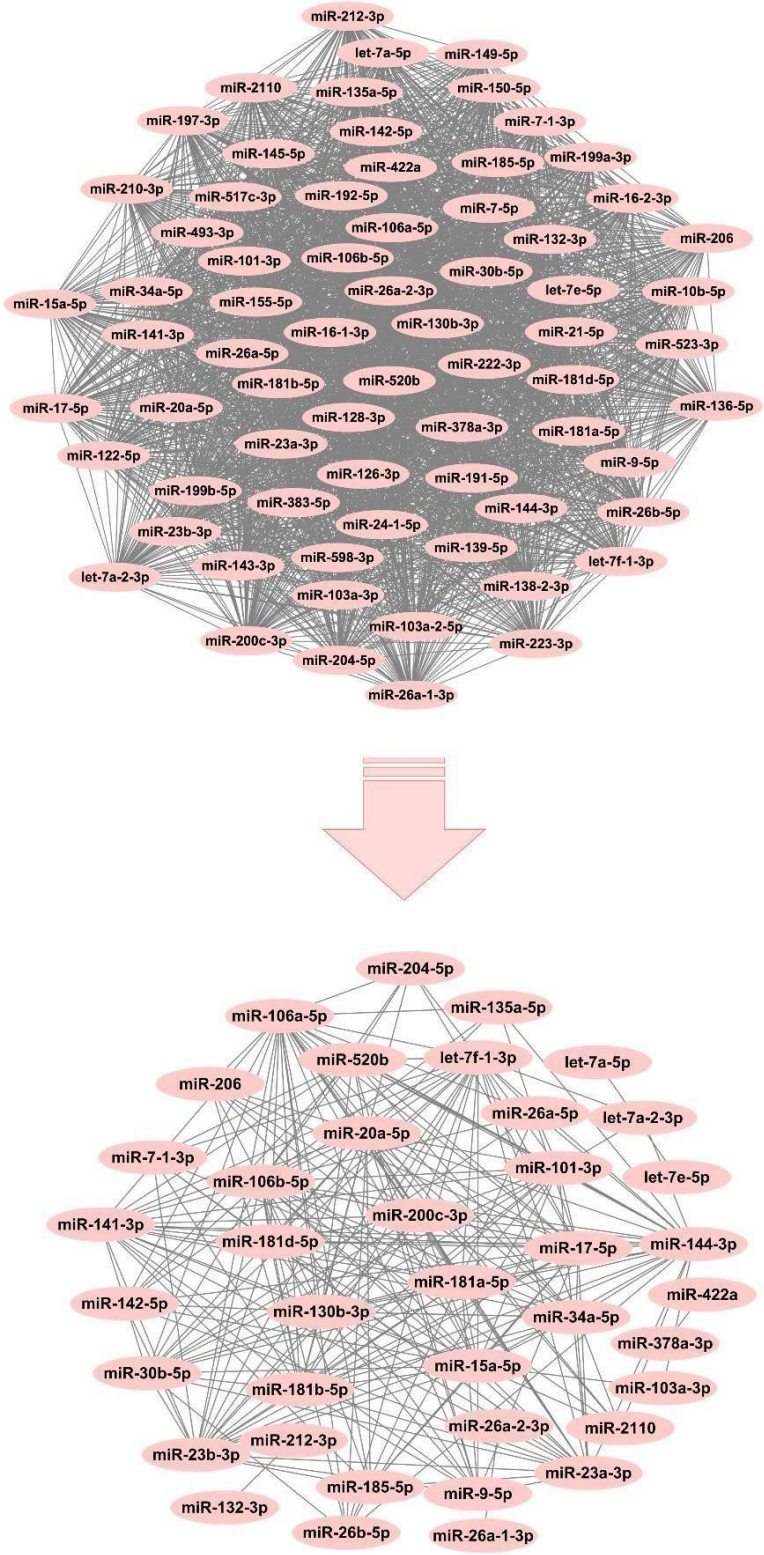


Figure 4. Endocrine gland neoplasm miRNA-miRNA graph. The nodes represent miRNAs and the connecting lines indicate the functional relations between miRNAs. In the original network (up) there are 70 nodes and 2,415 pairs, whereas in the pruned/filtered network (down) there are 39 nodes and 241 links.

4.3. Detection of neoplasm-specific miRNA modules

The aim of this study was to uncover the structure of miRNA interactions within neoplasms by identifying subgraphs, referred to as modules, that were not contained by any other subgraphs. These modules were considered of significant biological importance as they were thought to represent semi-autonomous systems (Peláez & Carthew, 2012)

To determine the most relevant modules, a two-fold approach was used. Firstly, we considered the level of mutual exclusivity, or how distinct the modules were from one another. Secondly, we evaluated the level of connectivity within each module, ensuring that the miRNAs within each module were more highly connected to each other than to other miRNAs in the overall network.

We could detect modules that met these criteria in 17 neoplasms (Table 2), but were unable to find any in 12 other neoplasms under study. The detected modules ranged in size from three to 23 miRNAs and varied greatly in terms of their composition and mutual exclusivity. For example, there were modules consisting of a large number of miRNAs in esophageal (23), colorectal (20), and breast neoplasms (17), as well as modules with a moderate number of miRNAs in liver (8) and lung neoplasms (7). There were also small cliques of three miRNAs found in sarcoma and squamous cell carcinomas (Figure 5).

Finally, we conducted an extensive search of the scientific literature by searching PubMed for any reported combinations of the miRNAs within the neoplasm-specific modules; no reported combinations of any two miRNAs that composed the neoplasm-specific modules were found.

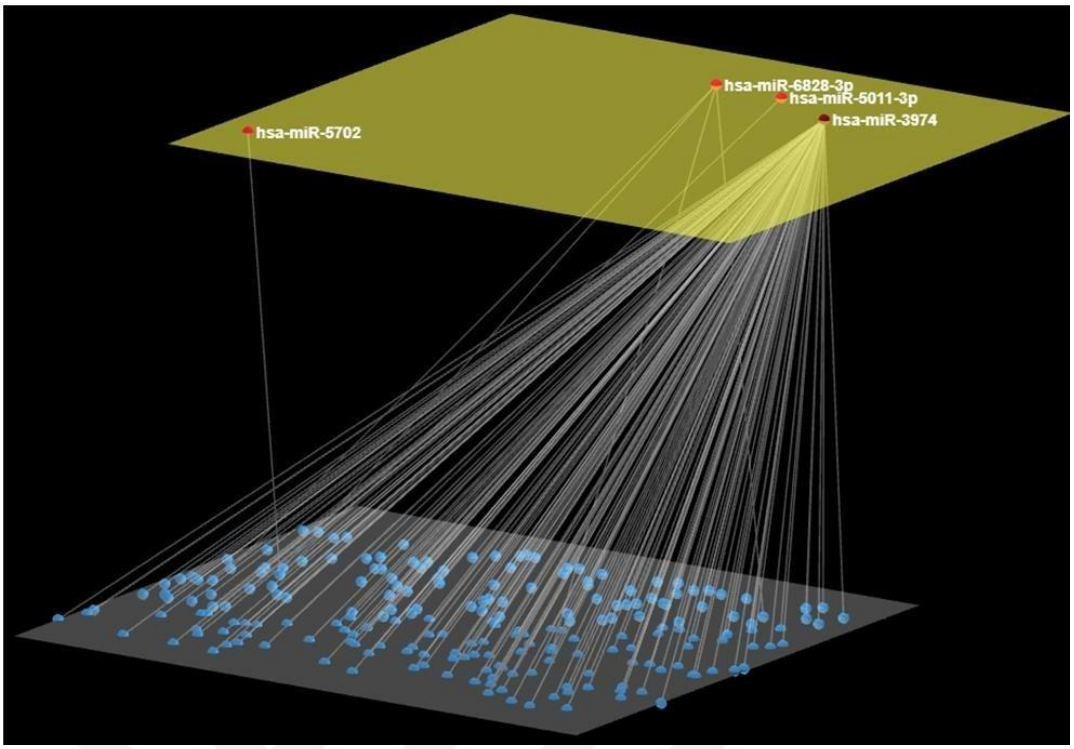
Table 2. Mutually exclusive miRNA modules per neoplasm.

Neoplasm	miRNA modules
Brain neoplasms	hsa-miR-3689e, hsa-miR-4422, hsa-miR-4432, hsa-miR-4720-3p, hsa-miR-5002-3p, hsa-miR- 548ao-5p, hsa-miR-5688, hsa-miR-6511b-5p, hsa- miR-6719-3p, hsa-miR-6753-3p, hsa-miR-6811-5p, hsa-miR-7107-3p, hsa-miR-8062, hsa-miR-8067, hsa-miR-8068

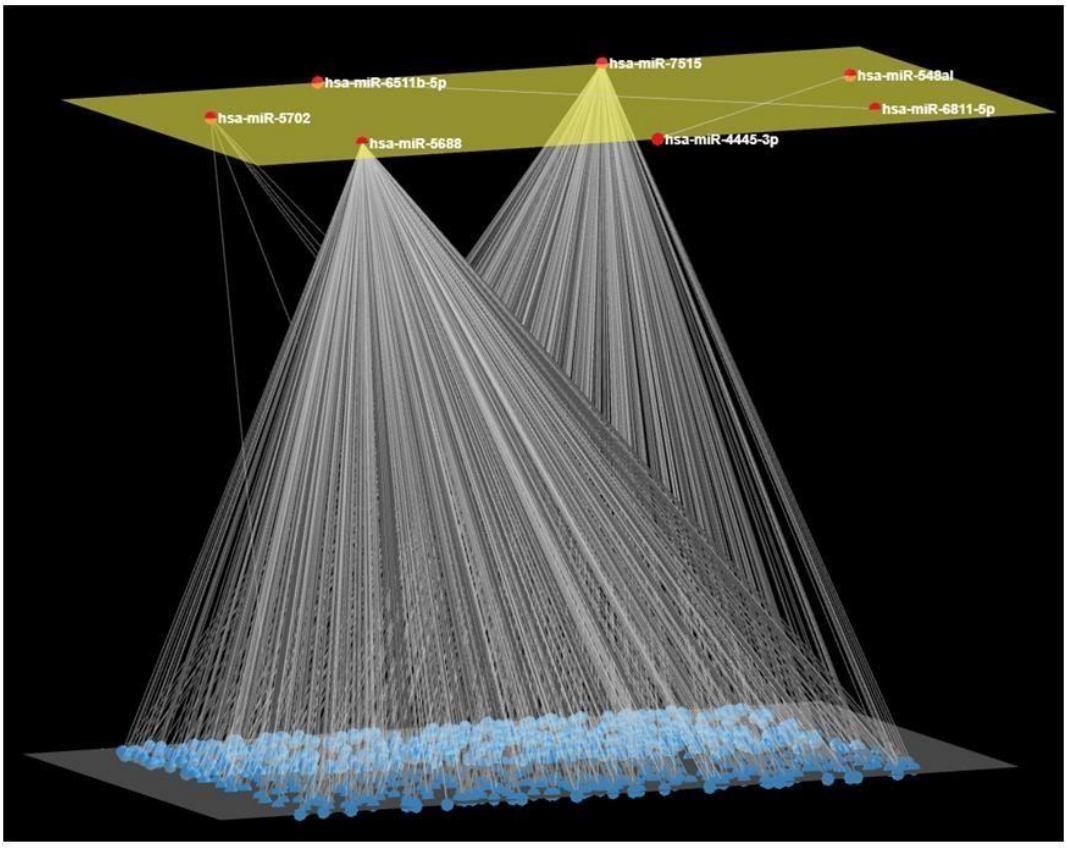
Breast neoplasms	hsa-miR-1296-3p, hsa-miR-4273, hsa-miR-4422, hsa-miR-4445-3p, hsa-miR-4694-3p, hsa-miR-4720-3p, hsa-miR-5002-3p, hsa-miR-5011-3p, hsa-miR-548ae-5p, hsa-miR-548al, hsa-miR-548ao-5p, hsa-miR-5699-5p, hsa-miR-6724-5p, hsa-miR-6748-5p, hsa-miR-6753-3p, hsa-miR-6773-5p, hsa-miR-7107-3p
Colorectal neoplasms	hsa-miR-4273, hsa-miR-4293, hsa-miR-4422, hsa-miR-4432, hsa-miR-4445-3p, hsa-miR-4694-3p, hsa-miR-4720-3p, hsa-miR-5002-3p, hsa-miR-5011-3p, hsa-miR-548aj-5p, hsa-miR-548al, hsa-miR-6511b-5p, hsa-miR-6724-5p, hsa-miR-6753-3p, hsa-miR-6770-5p, hsa-miR-6773-5p, hsa-miR-6811-5p, hsa-miR-7107-3p, hsa-miR-7154-5p, hsa-miR-8068
Esophageal neoplasms	hsa-miR-1296-3p, hsa-miR-3974, hsa-miR-4273, hsa-miR-4445-3p, hsa-miR-4720-3p, hsa-miR-5002-3p, hsa-miR-548aj-5p, hsa-miR-548al, hsa-miR-548ao-5p, hsa-miR-5688, hsa-miR-5699-5p, hsa-miR-6511b-5p, hsa-miR-6719-3p, hsa-miR-6724-5p, hsa-miR-6748-5p, hsa-miR-6770-5p, hsa-miR-6773-5p, hsa-miR-6811-5p, hsa-miR-6828-3p, hsa-miR-7154-5p, hsa-miR-7515, hsa-miR-8062, hsa-miR-8067
Head and neck neoplasms	hsa-miR-3974, hsa-miR-5011-3p, hsa-miR-5702, hsa-miR-6828-3p
Kidney neoplasms	hsa-miR-1296-3p, hsa-miR-3689e, hsa-miR-4422, hsa-miR-5699-5p
Liver neoplasms	hsa-miR-3689e, hsa-miR-6724-5p, hsa-miR-6773-5p, hsa-miR-7154-5p, hsa-miR-7159-5p, hsa-miR-8062, hsa-miR-8067, hsa-miR-8068
Lung neoplasms	hsa-miR-4445-3p, hsa-miR-548al, hsa-miR-5688, hsa-miR-5702, hsa-miR-6511b-5p, hsa-miR-6811-5p, hsa-miR-7515

Ovarian neoplasms	hsa-miR-4273, hsa-miR-526a-5p, hsa-miR-6770-5p, hsa-miR-7156-5p, hsa-miR-7515
Pancreatic neoplasms	hsa-miR-3974, hsa-miR-4273, hsa-miR-4293, hsa-miR-4694-3p, hsa-miR-522-5p, hsa-miR-523-5p, hsa-miR-5702, hsa-miR-6770-5p, hsa-miR-6828-3p, hsa-miR-7156-5p, hsa-miR-7159-5p, hsa-miR-7515
Prostatic neoplasms	hsa-miR-1296-3p, hsa-miR-522-5p, hsa-miR-5699-5p, hsa-miR-6748-5p
Sarcoma	hsa-miR-5688, hsa-miR-7515, hsa-miR-8067
Skin neoplasms	hsa-miR-3689c, hsa-miR-3689e, hsa-miR-522-5p, hsa-miR-548ae-5p, hsa-miR-548ao-5p, hsa-miR-5688, hsa-miR-6753-3p, hsa-miR-6828-3p, hsa-miR-7107-3p, hsa-miR-7154-5p, hsa-miR-8062, hsa-miR-8068
Squamous cell carcinomas	hsa-miR-4293, hsa-miR-4694-3p, hsa-miR-548aj-5p
Stomach neoplasms	hsa-miR-4720-3p, hsa-miR-5002-3p, hsa-miR-5699-5p, hsa-miR-5702, hsa-miR-6719-3p, hsa-miR-6828-3p
Thyroid neoplasms	hsa-miR-4293, hsa-miR-4422, hsa-miR-5011-3p, hsa-miR-6753-3p, hsa-miR-7107-3p
Urinary bladder neoplasms	hsa-miR-6719-3p, hsa-miR-6724-5p, hsa-miR-6773-5p, hsa-miR-8067

(A)


hsa-miR-6828-3p
hsa-miR-5011-3p
hsa-miR-3974
hsa-miR-5702

(B)


hsa-miR-7515
hsa-miR-6511b-5p
hsa-miR-548al
hsa-miR-6811-5p
hsa-miR-5702
hsa-miR-5688
hsa-miR-4445-3p

(C)

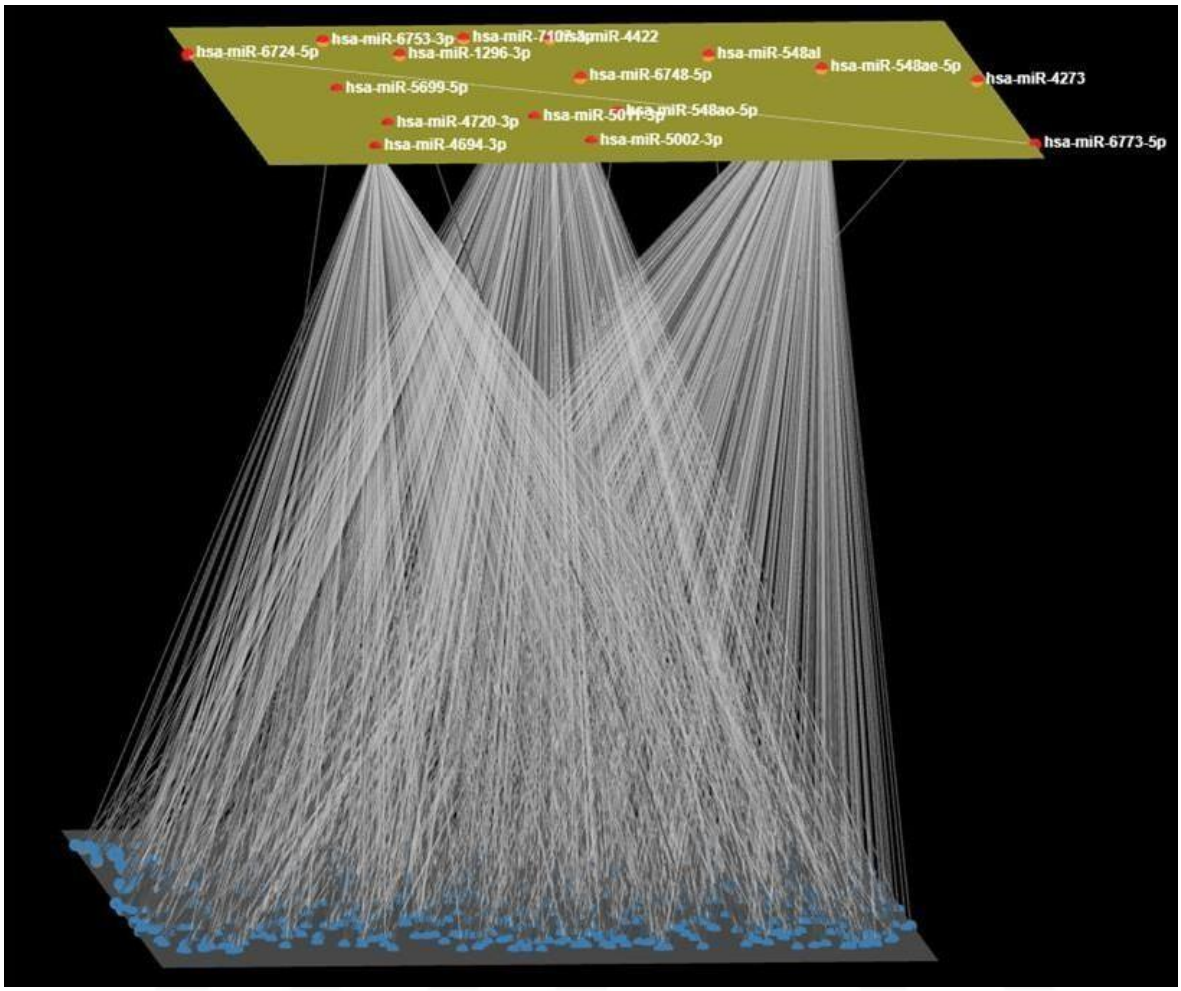


Figure 5. Neoplasm-specific modules. Depiction of modules consisting of (A) small (head and neck neoplasms), (B) moderate (lung neoplasms), and (C) large number (breast neoplasms) of miRNAs. The module miRNAs are shown in the upper layer, and their associations, both between them and with other miRNAs in the neoplasm network, are displayed in the lower layer. Arena3D was used for visualization.

5. DISCUSSION

The study by (Nalluri et al., 2017) aimed to discover preserved miRNA-miRNA interactions across different types of cancers by using a network inference method called miRsig. This approach leveraged miRNA expression profiles to identify these interactions. Another study by (Xu et al., 2011) utilized differential expression profiles of miRNAs and mRNAs to detect cancer-specific miRNAs in target-deregulated networks. (R. Song et al., 2015) also focused on identifying co-regulating miRNAs in lung cancer based on shared target genes that were functionally linked.

Despite these studies, a key aspect that has not been considered in previous research is the mutual exclusivity between the neoplastic diseases being investigated.

The current study aims to fill this gap by focusing on the mutual exclusivity relations between neoplastic diseases. The goal is to detect distinct subgraphs or modules within the neoplasm graphs that cannot be contained by another subgraph. These modules, which represent local clusters, are considered semi autonomous and therefore of high biological importance (Pelaez and Carthew, 2012). The top scoring mutually exclusive modules per neoplastic disease were selected based on the criteria that they exhibited significant mutual exclusivity and their constituent miRNAs were more highly connected to each other than to other miRNAs in the entire graph.

The results showed that distinguishable modules were detected in 17 out of the 17 neoplasms studied. The modules consisted of 3 to 23 miRNAs and only a few miRNAs were found in up to five cancer modules. A search on PubMed for combinations of any two miRNAs that compose the neoplasm-specific modules revealed no reported findings. This highlights the unique nature of the mutual exclusivity relations discovered in this study and the potential biological significance of these modules in the context of neoplastic diseases.

The present study aimed to uncover the distinct groups of miRNAs that play a role in neoplastic diseases, such as cancer. To achieve this goal, we gathered data from four different databases that contained various types of information about the relationships between miRNAs and cancers, including differential expression profiles, circulating biomarkers, genome-wide association studies (GWAS), and knockdown studies. The collected data was processed to eliminate duplicate entries and homogenize the neoplasm terms.

Next, we inferred miRNA connections by analyzing the shared functionality of their target genes. This was based on the assumption that miRNAs contribute to post-transcriptional gene regulation in a synergistic manner by co-ordinately regulating target genes. To uncover the miRNA-oriented networks and find the modules of tightly connected miRNAs per neoplasm, the researchers prioritized the top-ranking miRNA pairs based on the amount of available cancer-associated miRNA information.

Herein, mutual exclusivity was included as a prerequisite in order to detect sets of densely connected miRNAs that could represent the signature component for a certain type of cancer and not for another. This approach would be particularly useful

for the accurate detection of specific types of cancers. Likewise, in a prominent study by Ciriello et al. (2012), mutual exclusivity analysis of diverse genomic alterations was performed in order to identify oncogenic pathway modules.

The environmental context, that is, the connection patterns of the individual miRNAs, is very important, since in different niches (e.g., normal versus diseased state) the same miRNA may behave differently, including its differential expression status and connections. Therefore, both the miRNAs *per se* and the connection patterns among the component miRNAs of a module should be taken into consideration when extracting miRNA signatures. This is suggestive of the cooperativity between the respective miRNAs, so as to exert their regulatory effect on their targets.

In this study, miRNA modules were identified in seventeen neoplastic disease groups, composed of three to twenty-three members. The unique miRNA hsa-miR-3689c was found in the skin neoplasms module, and was also shown to be overexpressed in vulvar squamous cell carcinoma (X. Yang & Wu, 2016). Moreover, another unique miRNA, miR-526a-5p, was included in the module of ovarian neoplasms, and was found to be significantly differentially expressed in ovarian cancer (Penyige et al., 2019) (Table 2). Of the signature miRNAs (Table 2), hsa-miR-6770-5p, is suggested to have prognostic value in triple-negative breast cancer (TNBC) (Wu et al., 2020), and hsa-miR-1296-3p was found to be downregulated in prostate cancer (Majid et al., 2010). In addition, genetic variation in the target site of miR-4273-5p (Table 2) was associated with colorectal cancer susceptibility (Lee et al., 2016).

Of particular note, members of the let-7 (let-7a, 7b, 7c, 7d, 7e, 7f, 7g, 7i, miR-98, and miR-202) (Ma et al., 2021; Mizuno et al., 2018), miR-10 (miR-10a/b, miR-99a/b, miR-100 and miR-125) (Lund, 2010; Tehler et al., 2011), miR-15 (miR-15a/b, miR-16-1, miR-16-2, miR-195, and miR-497) (Aqeilan et al., 2010; Liu et al., 2019), miR-183 (miR-96, miR-182, and miR-183) (Ma et al., 2016; Pavlakis et al., 2017), miR-200 (miR-200a, 200b, 200c, miR-429, and miR-141)(Cavallari et al., 2021; Y. Chen & Zhang, 2017; G.-L. Huang et al., 2019), miR-221/222(Garofalo et al., 2012), miR-23 (miR-23a, miR-27a, and miR-24-2) (Chhabra et al., 2010), miR-29 (miR-29a, 29b, 29c) (H. Jiang et al., 2014; B. Yan et al., 2015), miR-30 (miR-30a, 30b, 30c, 30d, 30e)(K.

Song et al., 2020), miR-34 (miR-34a, 34b, 34c) (Agostini & Knight, 2014; Hermeking, 2010; Kalfert et al., 2020)miR-99 (miR-99a, 99b, and miR-100) (Eniafe & Jiang, 2021) families, miR-17/92 (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92a-1) cluster or 'oncomiR-1'(Concepcion et al., 2012; Mendell, 2008), and miR-21(Bautista-Sánchez et al., 2020; Jenike & Halushka, 2021), which are well-documented to be associated with multiple types/subtypes of cancers, occur with a high frequency across all neoplasms. Of particular note, none of those miRNAs were detected in any of the neoplasm-specific signature modules in Table 2. Due to their omnipresence in the major neoplastic diseases, the above miRNAs cannot represent signature components for specific types of cancers. Moreover, similar miRNA signatures were not detected despite thorough literature searches, further highlighting the importance of the findings of this study, as novel miRNA-miRNA interactions related to groups of neoplastic diseases were discovered

6. CONCLUSION AND FUTURE ASPECTS

In this study, bioinformatics and network based approaches were applied to miRNA-cancer associations, and finally the core sets of mutually exclusive and functionally related miRNAs for a broad range of Neoplasms were identified. In this way, neoplasm-specific miRNA modules were identified that could represent potential signatures for the corresponding diseases. Findings of this study could potentially be used in the clinical settings for the cancer diagnostic, prognostic and treatment purposes.

7. REFERENCES

Adam-Artigues, A., Garrido-Cano, I., Carbonell-Asins, J. A., Lameirinhas, A., Simón, S., Ortega-Morillo, B., Martínez, M. T., Hernando, C., Constância, V., Burgues, O., Bermejo, B., Henrique, R., Lluch, A., Jerónimo, C., Eroles, P., & Cejalvo, J. M. (2021). Identification of a two-MicroRNA signature in plasma as a novel biomarker for very early diagnosis of breast cancer. *Cancers (Basel)*, *13*(11), 2848.

Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, *4*.

Agostini, M., & Knight, R. A. (2014). miR-34: From bench to bedside. *Oncotarget*, *5*(4), 872–881.

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*(7006), 350–355.

Aqeilan, R. I., Calin, G. A., & Croce, C. M. (2010). miR-15a and miR-16-1 in cancer: Discovery, function and future perspectives. *Cell Death Differ.*, *17*(2), 215–220.

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, *116*(2), 281–297.

Bautista-Sánchez, D., Arriaga-Canon, C., Pedroza-Torres, A., De La Rosa-Velázquez, I. A., González-Barrios, R., Contreras-Espinosa, L., Montiel-Manríquez, R., Castro-Hernández, C., Fragoso-Ontiveros, V., Álvarez-Gómez, R. M., & Herrera, L. A. (2020). The promising role of miR-21 as a cancer biomarker and its importance in RNA-based therapeutics. *Mol. Ther. Nucleic Acids*, *20*, 409–420.

Bhattacharya, S., Bennet, L., Davidson, J. O., & Unsworth, C. P. (2022). Multi-layer perceptron classification & quantification of neuronal survival in hypoxic-ischemic brain image slices using a novel gradient direction, grey level co-occurrence matrix image training. *PloS One*, *17*(12), e0278874. https://doi.org/10.1371/journal.pone.0278874

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., & Aerts, H. J. W. L. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J. Clin.*, *69*(2), 127–157.

Bruford, E. A., Antonescu, C. R., Carroll, A. J., Chinnaiyan, A., Cree, I. A., Cross, N. C. P., Dalgleish, R., Gale, R. P., Harrison, C. J., Hastings, R. J., Huret, J.-L., Johansson, B., Le Beau, M., Mecucci, C., Mertens, F., Verhaak, R., & Mitelman, F. (2021). HUGO Gene Nomenclature Committee (HGNC) recommendations for the designation of gene fusions. *Leukemia*, *35*(11), 3040–3043.

Cavallari, I., Ciccarese, F., Sharova, E., Urso, L., Raimondi, V., Silic-Benussi, M., D'Agostino, D. M., & Ciminale, V. (2021). The miR-200 family of microRNAs: Fine tuners of epithelial-mesenchymal transition and circulating cancer biomarkers. *Cancers (Basel)*, *13*(23), 5874.

Chen, W., Gao, C., Liu, Y., Wen, Y., Hong, X., & Huang, Z. (2020). Bioinformatics analysis of prognostic miRNA signature and potential critical genes in colon cancer. *Front. Genet.*, *11*, 478.

Chen, Y., & Wang, X. (2020). MiRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.*, *48*(D1), D127–D131.

Chen, Y., & Zhang, L. (2017). Members of the microRNA-200 family are promising therapeutic targets in cancer. *Exp. Ther. Med.*, *14*(1), 10–17.

Chhabra, R., Dubey, R., & Saini, N. (2010). Cooperative and individualistic functions of the microRNAs in the miR-23a 27a 24-2 cluster and its implication in human diseases. *Mol. Cancer*, *9*(1), 232.

Chung, M.-J., Son, J.-Y., Park, S., Park, S.-S., Hur, K., Lee, S.-H., Lee, E.-J., Park, J.-K., Hong, I.-H., Kim, T.-H., & Jeong, K.-S. (2021). Mesenchymal stem cell and MicroRNA therapy of musculoskeletal diseases. *Int. J. Stem Cells*, *14*(2), 150–167.

Climent, M., Viggiani, G., Chen, Y.-W., Coulis, G., & Castaldi, A. (2020). MicroRNA and ROS crosstalk in cardiac and pulmonary diseases. *Int. J. Mol. Sci.*, *21*(12), 4370.

Concepcion, C. P., Bonetti, C., & Ventura, A. (2012). The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer J.*, *18*(3), 262–267.

Dlamini, Z., Francies, F. Z., Hull, R., & Marima, R. (2020). Artificial intelligence (AI) and big data in cancer and precision oncology. *Comput. Struct. Biotechnol. J.*, *18*, 2300–2311.

Domínguez-Almendros, S., Benítez-Parejo, N., & Gonzalez-Ramirez, A. R. (2011). Logistic regression models. *Allergologia Et Immunopathologia*, *39*(5), 295–305. https://doi.org/10.1016/j.aller.2011.05.002

Eniafe, J., & Jiang, S. (2021). MicroRNA-99 family in cancer and immunity. *Wiley Interdiscip. Rev. RNA*, *12*(3), e1635.

Garofalo, M., Quintavalle, C., Romano, G., Croce, C. M., & Condorelli, G. (2012). miR221/222 in cancer: Their role in tumor progression and response to therapy. *Curr. Mol. Med.*, *12*(1), 27–33.

Griesbach, C., Säfken, B., & Waldmann, E. (2021). Gradient boosting for linear mixed models. *The International Journal of Biostatistics*, *17*(2), 317–329. https://doi.org/10.1515/ijb-2020-0136

Harrison, J. E., Weber, S., Jakob, R., & Chute, C. G. (2021). ICD-11: An international classification of diseases for the twenty-first century. *BMC Med. Inform. Decis. Mak.*, *21*(Suppl 6), 206.

Hermeking, H. (2010). The miR-34 family in cancer and apoptosis. *Cell Death Differ.*, *17*(2), 193–199.

Hong, D. S., Kang, Y.-K., Borad, M., Sachdev, J., Ejadi, S., Lim, H. Y., Brenner, A. J., Park, K., Lee, J.-L., Kim, T.-Y., Shin, S., Becerra, C. R., Falchook, G., Stoudemire, J.,

Martin, D., Kelnar, K., Peltier, H., Bonato, V., Bader, A. G., … Beg, M. S. (2020). Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *Br. J. Cancer*, *122*(11), 1630–1637.

Huang, G.-L., Sun, J., Lu, Y., Liu, Y., Cao, H., Zhang, H., & Calin, G. A. (2019). MiR-200 family and cancer: From a meta-analysis view. *Mol. Aspects Med.*, *70*, 57–71.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics*, *15*(1), 41–51. https://doi.org/10.21873/cgp.20063

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., Zhou, Y., & Cui, Q. (2019). HMDD v3.0: A database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.*, *47*(D1), D1013–D1017.

Jenike, A. E., & Halushka, M. K. (2021). miR-21: A non-specific biomarker of all maladies. *Biomark. Res.*, *9*(1), 18.

Jiang, H., Zhang, G., Wu, J.-H., & Jiang, C.-P. (2014). Diverse roles of miR-29 in cancer (review). *Oncol. Rep.*, *31*(4), 1509–1516.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., & Liu, Y. (2009). miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, *37*(Database issue), D98-104.

Jin, S., Zeng, X., Fang, J., Lin, J., Chan, S. Y., Erzurum, S. C., & Cheng, F. (2019). A network-based approach to uncover microRNA-mediated disease comorbidities and potential pathobiological implications. *NPJ Syst. Biol. Appl.*, *5*(1), 41.

Kalfert, D., Ludvikova, M., Pesta, M., Ludvik, J., Dostalova, L., & Kholová, I. (2020). Multifunctional roles of miR-34a in cancer: A review with the emphasis on head and neck squamous cell carcinoma and thyroid cancer with clinical implications. *Diagnostics (Basel)*, *10*(8), 563.

Karatzas, E., Baltoumas, F. A., Panayiotou, N. A., Schneider, R., & Pavlopoulos, G. A. (2021). Arena3Dweb: Interactive 3D visualization of multilayered networks. *Nucleic Acids Res.*, *49*(W1), W36–W45.

Kehl, T., Backes, C., Kern, F., Fehlmann, T., Ludwig, N., Meese, E., Lenhof, H.-P., & Keller, A. (2017). About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget*, *8*(63), 107167–107175.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M., & Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nat. Genet.*, *37*(5), 495–500.

Kweon, M., Kim, J. Y., Jun, J. H., & Kim, G. J. (2020). Research trends in the efficacy of stem cell therapy for hepatic diseases based on MicroRNA profiling. *Int. J. Mol. Sci.*, *22*(1), 239.

Lam, I. K. Y., Chow, J. X., Lau, C. S., & Chan, V. S. F. (2018). MicroRNA-mediated immune regulation in rheumatic diseases. *Cancer Lett.*, *431*, 201–212.

Lee, A.-R., Park, J., Jung, K. J., Jee, S. H., & Kim-Yoon, S. (2016). Genetic variation rs7930 in the miR-4273-5p target site is associated with a risk of colorectal cancer. *Onco. Targets. Ther.*, *9*, 6885–6895.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854.

Liao, H., Zhang, X., Zhao, C., Chen, Y., Zeng, X., & Li, H. (2022). LightGBM: An efficient and accurate method for predicting pregnancy diseases. *Journal of Obstetrics and Gynaecology: The Journal of the Institute of Obstetrics and Gynaecology*, *42*(4), 620–629. https://doi.org/10.1080/01443615.2021.1945006

Liu, T., Xu, Z., Ou, D., Liu, J., & Zhang, J. (2019). The miR-15a/16 gene cluster in human cancer: A systematic review. *Journal of Cellular Physiology*, *234*(5), 5496–5506. https://doi.org/10.1002/jcp.27342

Lund, A. H. (2010). MiR-10 in development and cancer. *Cell Death & Differentiation*, *17*(2), Article 2. https://doi.org/10.1038/cdd.2009.58

Ma, Y., Liang, A.-J., Fan, Y.-P., Huang, Y.-R., Zhao, X.-M., Sun, Y., & Chen, X.-F. (2016). Dysregulation and functional roles of miR-183-96-182 cluster in cancer cell proliferation, invasion and metastasis. *Oncotarget*, *7*(27), 42805–42825. https://doi.org/10.18632/oncotarget.8715

Ma, Y., Shen, N., Wicha, M. S., & Luo, M. (2021). The Roles of the Let-7 Family of MicroRNAs in the Regulation of Cancer Stemness. *Cells*, *10*(9), Article 9. https://doi.org/10.3390/cells10092415

Majid, S., Dar, A. A., Saini, S., Chen, Y., Shahryari, V., Liu, J., Zaman, M. S., Hirata, H., Yamamura, S., Ueno, K., Tanaka, Y., & Dahiya, R. (2010). Regulation of Minichromosome Maintenance Gene Family by MicroRNA-1296 and Genistein in Prostate Cancer. *Cancer Research*, *70*(7), 2809–2818. https://doi.org/10.1158/0008-5472.CAN-09-4176

Mendell, J. T. (2008). MiRiad Roles for the miR-17-92 Cluster in Development and Disease. *Cell*, *133*(2), 217–222. https://doi.org/10.1016/j.cell.2008.04.001

Mizuno, R., Kawada, K., & Sakai, Y. (2018). The Molecular Basis and Therapeutic Potential of *Let-7* MicroRNAs against Colorectal Cancer. *Canadian Journal of Gastroenterology and Hepatology*, *2018*, e5769591. https://doi.org/10.1155/2018/5769591

Nalluri, J. J., Barh, D., Azevedo, V., & Ghosh, P. (2017). miRsig: A consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep39684

Paranjape, T., Slack, F. J., & Weidhaas, J. B. (2009). MicroRNAs: Tools for cancer diagnostics. *Gut*, *58*(11), 1546–1554. https://doi.org/10.1136/gut.2009.179531

Parveen, A., Mustafa, S. H., Yadav, P., & Kumar, A. (2019). Applications of machine learning in miRNA discovery and target prediction. *Curr. Genomics*, *20*(8), 537–544.

Pavlakis, E., Tonchev, A. B., Kaprelyan, A., Enchev, Y., & Stoykova, A. (2017). Interaction between transcription factors PAX6/PAX6-5a and specific members of miR183-96-182 cluster, may contribute to glioma progression in glioblastoma cell lines. *Oncology Reports*, *37*(3), 1579–1592. https://doi.org/10.3892/or.2017.5411

Peláez, N., & Carthew, R. W. (2012). Chapter nine - Biological Robustness and the Role of MicroRNAs: A Network Perspective. In E. Hornstein (Ed.), *Current Topics in Developmental Biology* (Vol. 99, pp. 237–255). Academic Press. https://doi.org/10.1016/B978-0-12-387038-4.00009-4

Peng, Y., & Croce, C. M. (2016). The role of MicroRNAs in human cancer. *Signal Transduct. Target. Ther.*, *1*(1).

Penyige, A., Márton, É., Soltész, B., Szilágyi-Bónizs, M., Póka, R., Lukács, J., Széles, L., & Nagy, B. (2019). Circulating miRNA Profiling in Plasma Samples of Ovarian Cancer Patients. *International Journal of Molecular Sciences*, *20*(18), Article 18. https://doi.org/10.3390/ijms20184533

Puerta-Gil, P., García-Baquero, R., Jia, A. Y., Ocaña, S., Alvarez-Múgica, M., Alvarez-Ossorio, J. L., Cordon-Cardo, C., Cava, F., & Sánchez-Carbayo, M. (2012). MiR-143, miR-222, and miR-452 are useful as tumor stratification and noninvasive diagnostic biomarkers for bladder cancer. *The American Journal of Pathology*, *180*(5), 1808–1815. https://doi.org/10.1016/j.ajpath.2012.01.034

Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine (New York, N.Y.)*, *47*(1), 31–39. https://doi.org/10.17849/insm-47-01-31-39.1

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Song, K., Jiang, Y., Zhao, Y., Xie, Y., Zhou, J., Yu, W., & Wang, Q. (2020). Members of the miR-30 family inhibit the epithelial-to-mesenchymal transition of non-small-cell lung cancer cells by suppressing XB130 expression levels. *Oncology Letters*, *20*(4), 1–1. https://doi.org/10.3892/ol.2020.11929

Song, R., Catchpoole, D. R., Kennedy, P. J., & Li, J. (2015). Identification of lung cancer miRNA–miRNA co-regulation networks through a progressive data refining approach. *Journal of Theoretical Biology*, *380*, 271–279. https://doi.org/10.1016/j.jtbi.2015.05.025

Tastsoglou, S., Alexiou, A., Karagkouni, D., Skoufos, G., Zacharopoulou, E., & Hatzigeorgiou, A. G. (2023). DIANA-microT 2023: Including predicted targets of virally encoded miRNAs. *Nucleic Acids Research*, gkad283. https://doi.org/10.1093/nar/gkad283

Tehler, D., Høyland-Kroghsbo, N. M., & Lund, A. H. (2011). The miR-10 microRNA precursor family. *RNA Biology*, *8*(5), 728–734. https://doi.org/10.4161/rna.8.5.16324
Thomas, J. P., Ölbei, M., Brooks-Warburton, J., Korcsmaros, T., & Modos, D. (2022). Analysing miRNA-target gene networks in inflammatory bowel disease and other complex diseases using transcriptomic data. *Genes (Basel)*, *13*(2), 370.

Wang, L., Chen, X., Zhang, L., Li, L., Huang, Y., Sun, Y., & Yuan, X. (2023). Artificial intelligence in clinical decision support systems for oncology. *Int. J. Med. Sci.*, *20*(1), 79–86.

Wang, R., Zhang, J., Shan, B., He, M., & Xu, J. (2022). XGBoost Machine Learning Algorithm for Prediction of Outcome in Aneurysmal Subarachnoid Hemorrhage. *Neuropsychiatric Disease and Treatment*, *18*, 659–667. https://doi.org/10.2147/NDT.S349956

Wu, X., Ding, M., & Lin, J. (2020). Three_microRNA expression signature predicts survival in triple_negative breast cancer. *Oncology Letters*, *19*(1), 301–308. https://doi.org/10.3892/ol.2019.11118

Xie, B., Ding, Q., Han, H., & Wu, D. (2013). miRCancer: A microRNA-cancer association database constructed by text mining on literature. *Bioinformatics (Oxford, England)*, *29*(5), 638–644. https://doi.org/10.1093/bioinformatics/btt014

Xu, J., Li, C.-X., Lv, J.-Y., Li, Y.-S., Xiao, Y., Shao, T.-T., Huo, X., Li, X., Zou, Y., Han, Q.-L., Li, X., Wang, L.-H., & Ren, H. (2011). Prioritizing Candidate Disease miRNAs by Topological Features in the miRNA Target–Dysregulated Network: Case Study of Prostate Cancer. *Molecular Cancer Therapeutics*, *10*(10), 1857–1866. https://doi.org/10.1158/1535-7163.MCT-11-0055

Yan, B., Guo, Q., Fu, F., Wang, Z., Yin, Z., Wei, Y., & Yang, J. (2015). The role of miR-29b in cancer: Regulation, function, and signaling. *OncoTargets and Therapy*, *8*, 539–548. https://doi.org/10.2147/OTT.S75899

Yan, J., Risacher, S. L., Shen, L., & Saykin, A. J. (2018). Network approaches to systems biology analysis of complex disease: Integrative methods for multi-omics data. *Brief. Bioinform.*, *19*(6), 1370–1381.

Yang, X., & Wu, X. (2016). MiRNA expression profile of vulvar squamous cell carcinoma and identification of the oncogenic role of miR-590-5p. *Oncology Reports*, *35*(1), 398–408. https://doi.org/10.3892/or.2015.4344

Yang, Y., Fu, X., Qu, W., Xiao, Y., & Shen, H.-B. (2018). MiRGOFS: A GO-based functional similarity measurement for miRNAs, with applications to the prediction of miRNA subcellular localization and miRNA-disease association. *Bioinformatics (Oxford, England)*, *34*(20), 3547–3556. https://doi.org/10.1093/bioinformatics/bty343

Yang, Z., Ren, F., Liu, C., He, S., Sun, G., Gao, Q., Yao, L., Zhang, Y., Miao, R., Cao, Y., Zhao, Y., Zhong, Y., & Zhao, H. (2010). dbDEMC: A database of differentially expressed miRNAs in human cancers. *BMC Genomics*, *11 Suppl 4*(Suppl 4), S5. https://doi.org/10.1186/1471-2164-11-S4-S5

Yerukala Sathipati, S., & Ho, S.-Y. (2020). Novel miRNA signature for predicting the stage of hepatocellular carcinoma. *Sci. Rep.*, *10*(1), 14452.

Yu, G. (2020). Gene Ontology Semantic Similarity Analysis Using GOSemSim. *Methods in Molecular Biology (Clifton, N.J.)*, *2117*, 207–215. https://doi.org/10.1007/978-1-0716-0301-7_11

Zhang, J., Bing, Z., Yan, P., Tian, J., Shi, X., Wang, Y., & Yang, K. (2019). Identification of 17 mRNAs and a miRNA as an integrated prognostic signature for lung squamous cell carcinoma. *J. Gene Med.*, *21*(8), e3105.

Zhao, X.-M., Liu, K.-Q., Zhu, G., He, F., Duval, B., Richer, J.-M., Huang, D.-S., Jiang, C.-J., Hao, J.-K., & Chen, L. (2015). Identifying cancer-related microRNAs based on gene expression data. *Bioinformatics*, *31*(8), 1226–1234.

8. APPENDIX

8.1. Ethics Committee Approval



İZMİR BİYOTIP VE GENOM MERKEZİ
GİRİŞİMSEL OLMAYAN ARAŞTIRMALAR ETİK KURULU (İBG-GOEK)
KARARI

Toplantı Tarihi	: 24/09/2021	**Toplantı Günü**	: Cuma
Toplantı Sayısı	: 10	**Toplantı Saati**	: 10:30

Sayın None Athanasia PAVLOPOULOU,

2021-031 Protokol No'lu; sorumlusu olduğunuz "Derin Sinir Ağları yaklaşımı kullanılarak hastalıklarla ilişkili İşlevsel olarak ilişkili ve Karşılıklı olarak özel mikroRNA'ların tanımlanması." başlıklı araştırmanın uygulanmasında etik açıdan sakınca olmadığına oy birliği ile karar verilmiştir.

Bilgilerinizi ve gereğini rica ederiz.

Prof. Dr. H. Alper BAĞRIYANIK **Başkan** (Katılmadı)	**Prof. Dr. Hilmi ORHAN** **Başkan Yardımcısı** (Katılmadı)
Prof. Dr. Sedef AKGÜNGÖR **Üye**	**Prof. Dr. İnci ALACACIOĞLU** **Üye**
Prof. Dr. Gülgün OKTAY **Üye** (Katılmadı)	**Dr. Serap ERKEK ÖZHAN** **Üye**
Prof. Dr. Şermin GENÇ **Üye** (Katılmadı)	**Doç. Dr. Canan Aslı YILDIRIM** **Üye**
Doç. Dr. Can KÜÇÜK **Üye**	**Dr. Öğr. Üyesi Yavuz OKTAY** **Üye**

8.2. Curriculum Vitae

REZA ARSHİNCHİ BONAB

Year of Birth 1994

Address: İzmir Uluslararası Biyotıp ve Genom Enstitüsü (İBG-İzmir)

Balçova 35340 İzmir

Telephone: +

e-mail:

University	Faculty/Institute	Field	Degree	Year of Graduation
Dokuz Eylul University, Turkey	İzmir International Biomedicine and Genome Institute	Molecular Biology and Genetics	M.Sc.	20223
Seraj Higher Education Institute, Iran	Faculty of Engineering	Information Technology Engineering	B.Sc.	2016

Skills

Adobe Illustrator
Bash scripting
Cytoscape
Convolutional Neural Networks
Machine Learning
Python
R programming language
Recurrent Neural Networks

Languages

Persian (native)
English (Fluent)

Research Interests

Bioinformatics
Systems Biology
Deep Learning
Cancer Computational Biology

Publications in ISI Journals

1) Asfa S, Toy HI, Arshinchi Bonab R, Chrousos GP, Pavlopoulou A, Geronikolou SA (2023) “Soft Tissue Ewing Sarcoma Cell Drug Resistance Revisited: A Systems Biology approach.” *International Journal of Environmental Research and Public Health* 20:6288.
2) Özbek M, Toy HI, Takan I, Asfa S, Arshinchi Bonab R, Karakülah G, Kontou PI, Geronikolou SA, Pavlopoulou A (2022) “A Counterintuitive Neutrophil-Mediated Pattern in COVID-19 Patients Revealed through Transcriptomics Analysis” *Viruses* 15:104
3) Arshinchi Bonab R, Asfa S, Kontou P, Karakülah,G, Pavlopoulou A (2022) “Identification of neoplasm-specific signatures of miRNA interactions by employing a systems biology approach” *PeerJ* 10:e14149