MIRCOOP-V2: DETECTING SYNERGISTIC MIRNA INTERACTIONS IN CANCER

by
ÖYKÜ ZEYNEP ASLAN

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of Master of Sciences

Sabancı University
July 2023

ABSTRACT

MIRCOOP-V2: DETECTING SYNERGISTIC MIRNA INTERACTIONS IN CANCER

ÖYKÜ ZEYNEP ASLAN

Computer Science and Engineering M.Sc. THESIS, July 2023

Thesis Supervisor: ASST. PROF. ÖZNUR TAŞTAN OKAN

Keywords: miRNA, synergy, transcription factor, web application

MiRNAs bind to their target messenger RNA (mRNA) to post-transcriptionally repress the target gene's expression through mRNA degradation and/or translational inhibition. As they are important regulators of gene expression, they take roles in different biological processes and are associated with pathologies such as cancer. While most miRNAs can repress their target only minimally, some miRNAs are known to act synergistically to strongly modulate the expression level of a common mRNA target. Harnessing this synergism in miRNA therapeutics offers the potential to enhance potency and efficacy while minimizing toxicity compared to single miRNA therapies.

Previously, a method called miRCoop was developed to identify potential synergistic miRNA pairs and their shared target mRNAs using kernel interaction tests on matched patient gene expression data of miRNAs and mRNAs. In this thesis, we present an improved version of this tool, called miRCoop-v2, which enhances several aspects of the original method. miRCoop-v2 can identify synergistic miRNA interactions involving transcription factors, explicitly incorporates miRNA arm information, utilizes a larger miRNA-mRNA target interaction dataset, and runs 30 times faster than its predecessor. This speed-up allows us to apply miRCoop-v2 to 31 different cancers to predict synergistic miRNA pairs and their shared targets. Results of the miRCoop v2 prove that it can be a valuable tool in the detection of cancer-related interactions between miRNA, TF (transcription factors), and target mRNA. Some significant examples include MYC, STAT3, MKI67, miR-19a-3p, and miR-424-5p. Even in situations where direct literature evidence for specific can-

cer may be lacking, the importance of these miRNAs and mRNAs in other cancer types suggests their potential relevance and therapeutic value across various cancer contexts.

We also have developed a web application to present the results, allowing users to access cancer-specific triplet and synergy module results, pan-cancer analyses of selected cancers, and interactive visualizations of the findings. We hope miRCoop-v2 will facilitate experimental efforts in detecting synergistic miRNA interactions. The application is accessible at http://mircoop.sabanciuniv.edu

ÖZET

ÖYKÜ ZEYNEP ASLAN

BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, TEMMUZ 2023

Tez Danışmanı: Dr. Öğr. Üyesi ÖZNUR TAŞTAN OKAN

Anahtar Kelimeler: miRNA, sinerji, transkripsiyon faktörü, web uygulaması

MiRNA'lar, hedef mesajcı RNA'ya (mRNA) bağlanarak mRNA bozunumu ve/veya translasyon inhibisyonu yoluyla hedef genin ifadesini post-transkripsiyonel olarak baskılar. Gen ifadesinin önemli düzenleyicileri oldukları için, kanser gibi farklı biyolojik süreçlerde ve patolojilerde rol oynarlar. Çoğu miRNA hedeflerini yalnızca az miktarda baskılayabilirken, bazı miRNA'ların aynı mRNA hedefin ifade düzeyini güçlü bir şekilde modüle etmek için sinerjik olarak hareket ettiği bilinmektedir. miRNA terapötiklerinde bu sinerjizmi kullanmak, tek bir miRNA tedavilerine kıyasla etkinliği artırma ve toksisiteyi en aza indirgeme potansiyeli sunar.

Daha önce, miRNA'ların ve mRNA'ların eşleştirilmiş hasta gen ifade verileri üzerinde çekirdek etkileşim testleri kullanılarak potansiyel sinerjik miRNA çiftlerini ve paylaşılan hedef mRNA'larını belirlemek için miRCoop adı verilen bir yöntem geliştirilmiştir. Bu tezde, orijinal yöntemin birçok yönünü geliştiren, miRCoop'un geliştirilmiş bir versiyonu olan miRCoop-v2'yi sunuyoruz. miRCoop-v2, transkripsiyon faktörlerini içeren sinerjistik miRNA etkileşimlerini tanımlayabilir, açıkça miRNA arm bilgisini dahil edebilir, daha geniş bir miRNA-mRNA hedef etkileşim veri kümesini kullanabilir ve önceki sürümüne göre 30 kat daha hızlı çalışabilir. Bu hızlanma sinerjik miRNA çiftlerini ve ortak hedeflerini tahmin etmek için miRCoop-v2'yi 31 farklı kansere uygulamamıza olanak sağlar. miRCoop-v2'nin sonuçları, miRNA, TF (transkripsiyon faktörleri) ve hedef mRNA arasındaki kanserle ilişkili etkileşimlerin tespitinde değerli bir araç olabileceğini kanıtlamaktadır. Bazı önemli örnekler arasında MYC, STAT3, MKI67, miR-19a-3p ve miR-424-5p bulunmaktadır. Özellikle belirli bir kanser türü için doğrudan literatür kanıtları olmayabilecek durumlarda bile, bu miRNA'ların ve mRNA'ların diğer kanser türlerindeki önemi, çeşitli kanser bağlamlarında potansiyel ilgi çekiciliğini ve terapötik değerini göster-

mektedir.

Ek olarak, sonuçları sunmak için bir web uygulaması geliştirdik, böylece kullanıcılar kanser özelinde üçlü ve sinerji modül sonuçlarına, seçilen kanserlerin pan-kanser analizlerine ve bulguların etkileşimli görselleştirmelerine erişebilirler. Umarız miRCoop-v2, sinerjik miRNA etkileşimlerini tespit etme konusundaki deneysel çabaları kolaylaştıracaktır. Uygulamaya aşağıdaki adresten erişilebilir: http://mircoop.sabanciuniv.edu

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Oznur Tastan, for her precious guidance, consistent support, and continuous encouragement throughout the entire process of this thesis. Her expertise, knowledge, and insightful feedback have been instrumental in shaping the direction and quality of this research work.

Furthermore, I would like to express my deepest gratitude to my family for their unconditional love, consistent support, and tolerance throughout this endeavor. Their encouragement and patience have been the driving force and I am forever grateful for their presence in my life.

Additionally, I would like to express my sincere appreciation to the members of the jury who dedicated their time, expertise, and valuable insights to evaluate and assess this thesis.

I would also like to extend my sincere appreciation to TUSEB for their funding support under the project numbered 4382. I also acknowledge TUBITAK BIDEB for their support.

Last but not least, I would like to express my heartfelt thanks to all the members of the Tastan Lab, and my friends Melike Tombaz and Arda Sener. I am grateful for their valuable insights and willingness to assist at every stage of the project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATONS

ACC Adrenocortical carcinoma 35, 36

BLCA Bladder Urothelial 36

BRCA Breast Invasive Carcinoma 51

CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma ... 52

CHOL Cholangiocarcinoma 33

COAD Colon adenocarcinoma 38

DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma 35, 36, 52

HNSC Head and Neck squamous cell carcinoma 44

KICH Kidney Chromophobe 34

LGG Low-grade gliomas 34, 36, 40

LIHC Liver hepatocellular carcinoma 38

LUAD Lung adenocarcinoma 45

LUSC Lung squamous cell carcinoma 51

MESO Mesothelioma 34

OV Ovarian serous cystadenocarcinoma 36

PAAD Pancreatic adenocarcinoma 39

PCPG Pheochromocytoma and Paraganglioma 52

PRAD Prostate adenocarcinoma 39

READ Rectum adenocarcinoma 33

SARC Sarcoma 33

1. INTRODUCTION

Cancer is a complex and multifaceted disease characterized by the uncontrolled proliferation and division of abnormal cells in various tissues and organs. It represents a significant global health challenge and remains one of the leading causes of death worldwide (Salehi et al., 2018; Anand et al., 2022). In the United States alone, the National Cancer Institute estimates that approximately 609,820 individuals will succumb to cancer in 2023. While significant progress has been made in cancer therapeutics, there is still much to be elucidated regarding the intricate molecular mechanisms underlying this disease.

MiRNAs, small non-coding RNAs approximately 22 nucleotides in length, play crucial regulatory roles by targeting messenger RNAs (mRNAs) (Bartel, 2004). miRNAs bind to their target messenger RNA (mRNA) sequence to post-transcriptionally repress the target gene's expression through mRNA degradation and/or translational inhibition (Bartel, 2004). These miRNAs are pivotal in various biological processes, including cell cycle regulation, cell proliferation, and their dysregulation is implicated in diseases such as cancer, where they can function as both tumor suppressors and oncogenes (Peng and Croce, 2016). Remarkably, miRNAs are known to regulate around 30% of human genes, making them an innovative therapeutic avenue for miRNA-based therapies (Garzon et al., 2009; Bueno and Malumbres, 2011; Lai et al., 2019). However, the development of miRNA-based therapeutics faces notable challenges, including limited efficacy, off-target effects, and potential toxicity. One notable limitation is the weak inhibition exhibited by certain miRNAs when regulating oncogenes, which restricts their effectiveness (Selbach et al., 2008). Consequently, high doses of miRNA mimics are often required to achieve the desired impact, resulting in unintended off-target effects and potential toxicity. The utilization of synergistic miRNAs holds promise in addressing these issues since miRNAs can exert higher repression when they work cooperatively and synergistic miRNAs enhance regulatory effects to overcome the limitations of individual miRNAs (Selbach et al., 2008; Baek et al., 2008; Lai et al., 2019).

Cooperative regulation can occur in different ways, such as multiple miRNAs tar-

geting different constituents of a functional module or directly binding to a common target mRNA. In this study, our focus is on the latter case. Despite accumulating evidence (reviewed in Chapter 2), our understanding of synergistic miRNAs that jointly regulate a single gene remains incomplete. Identifying pairs of miRNAs that exhibit synergistic effects holds the potential to advance our understanding of miRNA function and their role in regulating various gene expression programs in cancer.

This study discussed that multiple miRNAs work together by regulating the same gene (Bracken et al., 2016). To demonstrate, a pair of miRNAs can jointly target one mRNA at the same time, or one miRNA binds directly to the target gene while the other miRNA binds to the TF that regulates this gene (Bracken et al., 2016; Zhang et al., 2016). In this case, one miRNA acts directly and the other miRNA acts indirectly by regulating the transcription factor. Moreover, two miRNAs can regulate a gene through a common TF. Finally, two miRNAs can regulate the same target gene via two different transcription factors. While alternative configurations may exist, this study will concentrate on examining these four specific forms within its scope.

The study aimed to create a computational approach that integrates miRNA and RNA expression data with miRNA-mRNA interaction information to identify various types of synergistic relationships. The primary goal of these computational methods is to understand the synergistic miRNA interactions in cancer and eventually aid the design of miRNA-based therapies. The developed method was applied to several TCGA cancers, resulting in the identification of synergistic miRNA pairs and triplets with shared targets. The findings and results of this study have been made accessible through a web platform, which can be accessed at http://mircoop.sabanciuniv.edu

The thesis is organized as follows:

- Chapter 2 provides a comprehensive review of experimental evidence regarding synergistic interactions, as well as computational methods that have been developed to identify such interactions. In the final section of this chapter, we will delve into the details of miRCoop, the method that this thesis expands upon.
- We describe our approach for finding synergistic miRNA interaction in cancer in Chapter 3. In particular, the enhancements made to the miRCoop pipeline along with detailed explanations of each step of the algorithm are presented in Section 3.1, and we explain methods for detecting synergistic miRNAs tar-

geting the same gene via TFs in Section 3.2. In Section 3.4 of this chapter, the dataset used in the study and the pre-processing steps applied to the data are explained in detail.

- Chapter 4 Section 4.1 provides a general overview of the miRCoop web application with its capabilities. Analysis of the miRCoop triplets and synergy modules and the interaction networks are also presented in this chapter.
- We conclude our work and discuss future directions in Chapter 5.

2. LITERATURE REVIEW

This chapter will review experimental evidence on synergistic interactions and computational methods that are aimed at discovering them. In the last part, we will explain miRCoop, the method that this thesis extends, in more detail.

2.1 Experimental Evidence on Synergistic miRNAs

Multiple studies strongly support the presence of miRNA synergism providing compelling evidence of the collaborative interactions among miRNAs. Enright et al. (2003) presented the initial evidence of miRNA synergism, demonstrating the cooperative interaction between lin-4 and let-7 in a Drosophila experiment. Wu et al. (2010) conducted a study revealing for the first time that multiple miRNAs from different miRNA clusters can suppress the expression of p21Cip1/Waf1 by directly targeting its 3' untranslated region. In a separate study, Bandi and Vassella (2011). demonstrated the synergistic action of miR-34a and miR-15a/16, which are functionally related and control similar processes, in non-small cell lung cancer (NSCLC). Then, Lai et al. (2012) reported that miR-572 and miR-93 work together to repress p21 in melanoma cells. Furthermore, Yang et al. (2013) showed that miR-17-5p and miR-17-3p induce prostate tumor growth and invasion by targeting TIMP3. Another study by Zhao et al. (2017) highlighted the highest synergistic effect observed in glioma for miR-20a and miR-21. Additionally, Borzi et al. (2017) found that mir-486-5p is positively regulated by mir-660-5p and investigated the synergistic effect between these two miRNAs in lung cancer. Lai et al. (2018) demonstrated that miR-205-5p and miR-342-3p act cooperatively to repress E2F1 and reduce tumor chemoresistance.

2.2 Methods for Detecting miRNA Synergistic Pairs

Investigating the cooperative regulations among miRNAs and uncovering miRNA synergistic pairs have important implications for understanding gene regulatory networks and driving the development of innovative cancer therapeutic approaches. Despite their significance, we do not have the complete set of cooperative miRNAs.

Several computational methods have been developed in the literature to explore miRNA cooperativity. However, many of these tools primarily concentrate on investigating the cooperative behavior of miRNAs in regulating a specific "set of genes" that form a functional module (Xu et al., 2011; Li et al., 2014; Gennarino et al., 2012; Zhang et al., 2012; Zhao et al., 2013; Ding et al., 2015; Meng et al., 2015; Yang et al., 2015; Xu et al., 2017; Ding et al., 2018; Cantini et al., 2019). Conversely, there is limited availability of tools that specifically investigate cooperativity in regulating a single gene.

Schmitz et al. (2014) introduced a method called TriplexRNA, which aimed to identify cooperative miRNAs and their shared targets. The TriplexRNA workflow involves predicting and validating microRNA target interactions. It also incorporates secondary structure prediction, molecular dynamics simulations, and simulations of kinetic models. In another study, Meng et al. (2015) employed the functional similarity of miRNA targets and their proximity in the protein-protein interaction (PPI) network to identify synergistic miRNA pairs. Additionally, Li et al. (2014). developed a method for detecting synergistic miRNA regulatory modules by utilizing a deterministic overlapping clustering algorithm and overlapping neighborhood expansion.

One of the few studies in this field is the miRCoop algorithm developed by Olgun and Tastan (Olgun and Tastan, 2020). miRCoop was designed to find miRNA pairs that synergistically target an mRNA by simultaneously binding to the 3'UTR of an mRNA. miRCoop uses kernel-based statistical interaction tests and reveals miRNAs that strongly repress their target mRNA when acting together, but show weak or no repression when considered individually. miRCoop uses the target miRNA-mRNA information together with the matched expression profiles of miRNAs and mRNAs over a set of samples (The method is detailed in Section 2.2.1). The algorithm was previously applied to kidney and lung cancers in The Cancer Genome Atlas project (TCGA).

These studies represent significant contributions to the exploration of miRNA syner-

gistic regulations, each employing distinct methodologies to identify and characterize cooperative miRNA pairs and their shared targets.

Also, numerous studies have been conducted on the interactions between miRNAs and TFs. To demonstrate, TFmiR is a web server designed to facilitate the analysis of disease-specific transcription factor (TF) and miRNA co-regulatory networks (Hamed et al., 2015). It enables the investigation of four different types of interactions: TF-gene, miRNA-gene, TF-miRNA, and miRNA-miRNA. TFmiR offers a range of analytical features, including network analysis, detection of TF-miRNA co-regulatory motifs, network visualization, and statistical significance assessment for extracted interactions and the disease network. When a specific disease is selected, TFmiR utilizes data from the HMDD and DisGeNET to identify disease-associated miRNAs and genes. Furthermore, TFmiR identifies four distinct TF-miRNA co-regulatory motifs, where each motif consists of a TF, a miRNA, and a common target gene.

Another web server CMTCN dedicated to studying cancer-specific microRNA and transcription factor co-regulatory networks. Its pipeline involves five key steps to facilitate comprehensive analysis: gathering both predicted and experimentally validated interactions, curating genes and miRNAs that are relevant to cancer across 33 different types of cancer, identification of regulatory relationships among cancer-related transcription factors, miRNAs, and genes, detecting co-regulatory motifs involving microRNAs and TFs and construction and refinement of co-regulatory network with TCGA expression data (Li et al., 2018b).

FFLtool is a web server designed to facilitate the investigation of TF-miRNA-target Feed-Forward Loop (FFL) regulatory motifs (Xie et al., 2020). It offers two key functional modules: FFL Analysis and Browse FFLs, providing users with valuable insights into regulatory interactions within the FFL network. FFL Analysis allows users to detect candidate FFLs from a gene set defined by the user in specific tissues of interest and generates an overall graph of the FFL network. In Browse FFLs users can browse FFLs formed by differentially or specifically expressed genes, TFs, and miRNAs in various types of cancer.

Each of the developed methods exhibits its own set of advantages and disadvantages. Notably, some studies overlook the consideration of RNA expression levels, and a significant number of them lack the provision of their results through a web platform. Indeed, studies that specifically investigate miRNA-TF interactions often focus on understanding the regulatory relationships between individual miRNAs and TFs, without explicitly considering miRNA synergism.

2.2.1 miRCoop

Because the study presented in this thesis extends the previous work miRCoop. In this section, we present the miRCoop in more detail.

The miRCoop approach employs a combination of miRNA-mRNA target predictions and kernel statistical interaction tests on expression data to identify synergistic interactions between miRNA pairs and their target mRNAs. The algorithm consists of three main steps to achieve this goal. The first step involves the identification of potential miRNA-mRNA pairs by utilizing miRNA target prediction algorithms. In the second step, the candidate triplets are filtered based on the expression patterns of the miRNAs and mRNAs. The final step of the miRCoop algorithm utilizes kernel-based statistical tests (Sejdinovic et al., 2013), specifically two and three-variable tests, on the expression data.

The kernelized Lancaster three-variable interaction test in the miRCoop algorithm Step 3 is based on the concept of mapping probability distributions into a reproducing kernel Hilbert space (RKHS). This approach allows embedding probability distributions into potentially infinite-dimensional feature spaces using kernel functions (Smola et al., 2007). By employing kernel embedding, restrictive assumptions such as discrete-valued variables or linear relationships among the variables can be avoided (Song et al., 2013). The kernel approach also allows to handle small sample sizes of the available expression profiles without requiring specific distributional assumptions. The following sections will explain the steps of the algorithm in detail.

2.2.1.1 miRCoop Step 1: Identifying Candidate miRNA–Target Interactions

The initial step involves the identification of miRNAs that target the same mRNAs. This is done using target prediction algorithms such as TargetScan (Agarwal et al., 2015) and miRanda (Betel et al., 2010). By comparing the predictions from both algorithms, miRNA–mRNA pairs that are predicted by both algorithms are considered potential candidates for triplet formation. The algorithm excludes pairs where the miRNAs have overlapping binding sites on the target mRNA. After applying these filtering criteria, the algorithm generates an extensive list of candidate triplets.

2.2.1.2 miRCoop Step 2: Expression Filter

The filtering process in Step 2 of the miRCoop algorithm is based on the rationale that when both miRNAs in a candidate triplet are upregulated, the expression level of the target mRNA should be significantly lower compared to when both miRNAs are downregulated. To identify triplets that exhibit this expression pattern, the algorithm divides the tumor samples into different expression subgroups(samples where both miRNAs are upregulated and samples where both miRNAs are downregulated) based on the expression states of the potential synergistic miRNA pairs. The algorithm determines the upregulated(miRNA expression in a sample is in the fourth quartile) or downregulated(miRNA expression is in the first quartile) status of a miRNA in a sample based on its expression level relative to the distribution of expression levels across all samples. To assess the significance of the difference in mRNA expression between the two subgroups, the algorithm applies a one-sided Wilcoxon rank-sum test (Wilcoxon, 1945) and retains only those for which the null hypothesis is rejected.

2.2.1.3 miRCoop Step 3: Statistical Interaction Tests on RNA Expression Data

In Step 3, a kernel-based three-variable interaction test (Sejdinovic et al., 2013) was performed on the miRNA and mRNA expression data of the potential triplets that passed Step 2. The expression levels of the two miRNAs were represented by the random variables X and Y, while the expression level of the mRNA was represented by the random variable Z. The objective was to identify cases where the miRNAs were pairwise independent with the mRNA but formed a mutually dependent triplet. The independence of X and Z and Y and Z were tested with the Two Variable Test (Sejdinovic et al., 2013). Here, only triplets in which the null hypothesis failed to be rejected were considered. Additionally, a Three Variable Kernel Interaction Test (Sejdinovic et al., 2013) was applied to the potential triplets and only the triplets in which the null hypothesis was rejected were retained as predicted miRCoop triplets.

3. METHODS & FRAMEWORK

This chapter presents the details of the synergistic methods developed in this thesis.

3.1 Summary of Enhancements to miRCoop

Olgun and Tastan previously devised a methodology called miRCoop to identify synergistic interactions among miRNAs and their shared targets (Olgun and Tastan, 2020). The primary objective of this study is to identify miRNA pairs that exhibit weak or no individual repression but demonstrate strong repression when they act together on the target mRNA, thereby showcasing synergistic effects. The miRCoop algorithm basically consists of three steps: identification of miRNA-mRNA pairs, filtering based on miRNA and mRNA expression data, and statistical interaction tests applied to RNA expression data (The method is detailed in Section 2.2.1). Here, with this work, this method has been improved in several aspects. In order to provide the necessary background information, the following sections will explain these improvements in the miRCoop Framework.

3.1.1 Identification of miRNA Target Interactions

The initial step of miRCoop involves identifying miRNA pairs that target the same mRNAs, utilizing experimentally validated miRNA target databases. These interactions are then cross-referenced with TargetScan predictions (Agarwal et al., 2015) to obtain miRNA binding sites. From the miRNA pairs that target the same mRNA, potential triplets capable of forming synergistic interactions are constructed. In the process of forming these triplets, miRNA pairs with the same binding site are

excluded.

3.1.2 Expression Filter

Secondly, candidate triplets that passed from Step 1 were filtered based upon the following rationale: since miRNAs repress the expression level of the mRNA's, it is expected that when both miRNAs are upregulated, the expression level of the mRNA should be significantly lower compared to when both miRNA's are downregulated. In order to test the significance, the Wilcoxon rank sum test (Wilcoxon, 1945) was applied and potential triplets that were rejected at a significance level of 0.05 are fed into the algorithm Step 3.

3.1.3 Statistical Interaction Tests

In order to find synergistic triplets, a kernel-based three-variable interaction test was applied (Sejdinovic et al., 2013). This is a permutation-based non-parametric test that can be used where two independent causes individually have a weak influence on a third dependent variable, but their combined effect has a strong influence. In our algorithm we expect that each miRNA is pairwise independent with mRNA expression level, however, three of them should form a mutually dependent triplet. Here, miRNA expression levels are denoted with X and Y random variables, and mRNA expression level is denoted with the random variable Z, the following holds: $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$ whereas $\neg((X,Y) \perp\!\!\!\perp Z)$. For two variables, Lanchaster interaction is defined as:

$$\triangle_L P = P_{XY} - P_X P_Y. \tag{3.1}$$

For three variables, the Lancaster interaction is defined as:

$$\triangle_L P = P_{XYZ} - P_{XY}P_Z - P_{YZ}P_X - P_{XZ}P_Y + 2P_X P_Y P_Z. \tag{3.2}$$

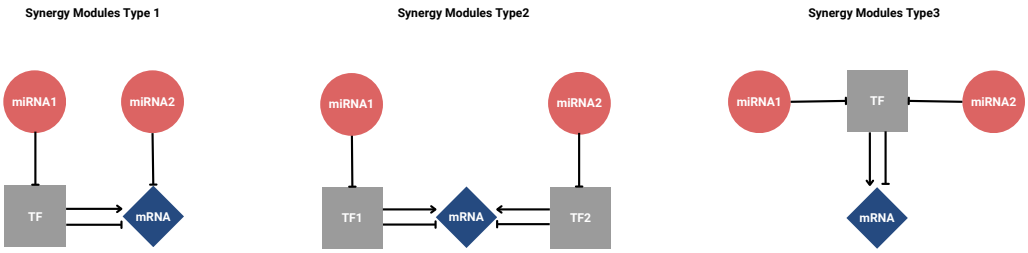
Here, P_X, P_Y and P_Z are marginal distributions, P_{XZ}, P_{XY} and P_{YZ} are pairwise marginal distributions and P_{XYZ} are the joint distributions of random variables X, Y and Z. The $\triangle_L P$ is zero whenever P can be factorized as a product of its

marginal distributions and the null hypothesis, in this case, is that the Lancaster interaction is zero. After applying two and three variable tests (Sejdinovic et al., 2013), Benjamini-Hochberg correction with a false discovery rate equal to 0.05 was applied in order to reduce false positives.

3.2 miRCoop Synergy Modules

The first objective of this study involved enhancing the miRCoop algorithm, originally designed to identify synergistic miRNAs that directly target the same gene. The improved algorithm was then employed to analyze 31 distinct cancer types. Subsequently, the study focused on developing a new algorithm capable of identifying synergistic miRNAs that target the same gene through TFs. There can be three different modes of regulation in synergistic miRNA associations with TFs as it was shown in 3.1

Figure 3.1 Detection of synergistic miRNAs targeting the same gene via TFs.



These are as follows:

- *Type 1:* One miRNA directly regulates the gene, while the second miRNA can regulate the gene indirectly by binding to the TF.
- *Type 2:* Two miRNAs can regulate the expression of a common gene by binding to two different TFs.
- *Type 3:* Two miRNAs can regulate gene expression by binding to a common single TF.

The following section provides a detailed description of the algorithm developed for

each of the three different scenarios, elucidating the specific approaches utilized in each case.

3.2.1 Step1: Identification of miRNA-mRNA Target Interactions

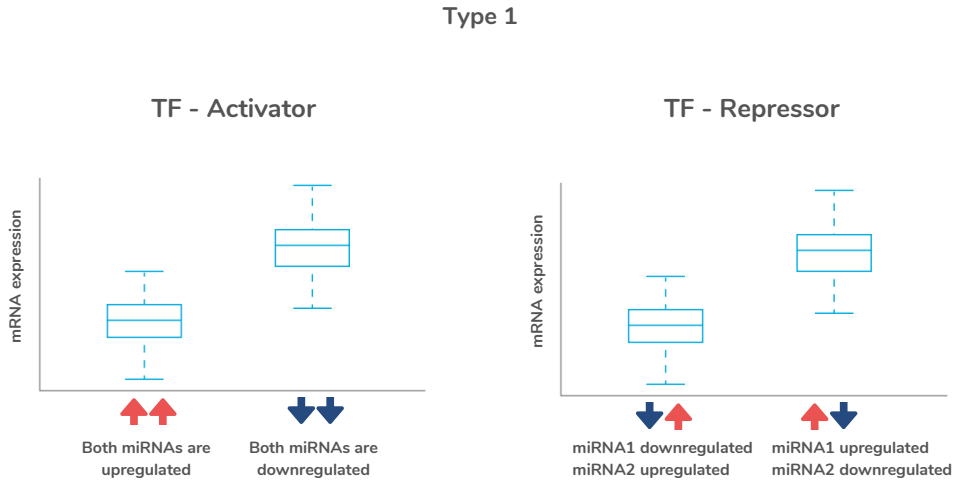
For Type 1, initially, the interaction between miRNA1, TF, and the target gene was identified. Subsequently, the algorithm detected another miRNA that also target the same gene, resulting in the creation of an interaction list. In this process, instances where direct interaction between miRNA1 and the gene occurred, were excluded from the list. For Type 2, firstly, miRNA-TF interactions targeting different TFs were established. Then, the mRNAs that these different TFs jointly target were identified. Here, miRNA interactions that have direct interactions with mRNA are not included. For Type 3, after detecting miRNAs targeting the same TF, genes targeted by TF were added to the list. For this type, miRNA interactions, which have direct interactions with mRNA, are excluded from the list of potential triplets like the other two cases.

In all steps, the intersection of experimentally detected miRNA-mRNA interactions with those from TargetScan was used. Gene-TF information was obtained from DoRothEA (Garcia-Alonso et al., 2019) in such a way that those with a confidence level of C and above were selected.

3.2.2 Step 2: Expression Filter

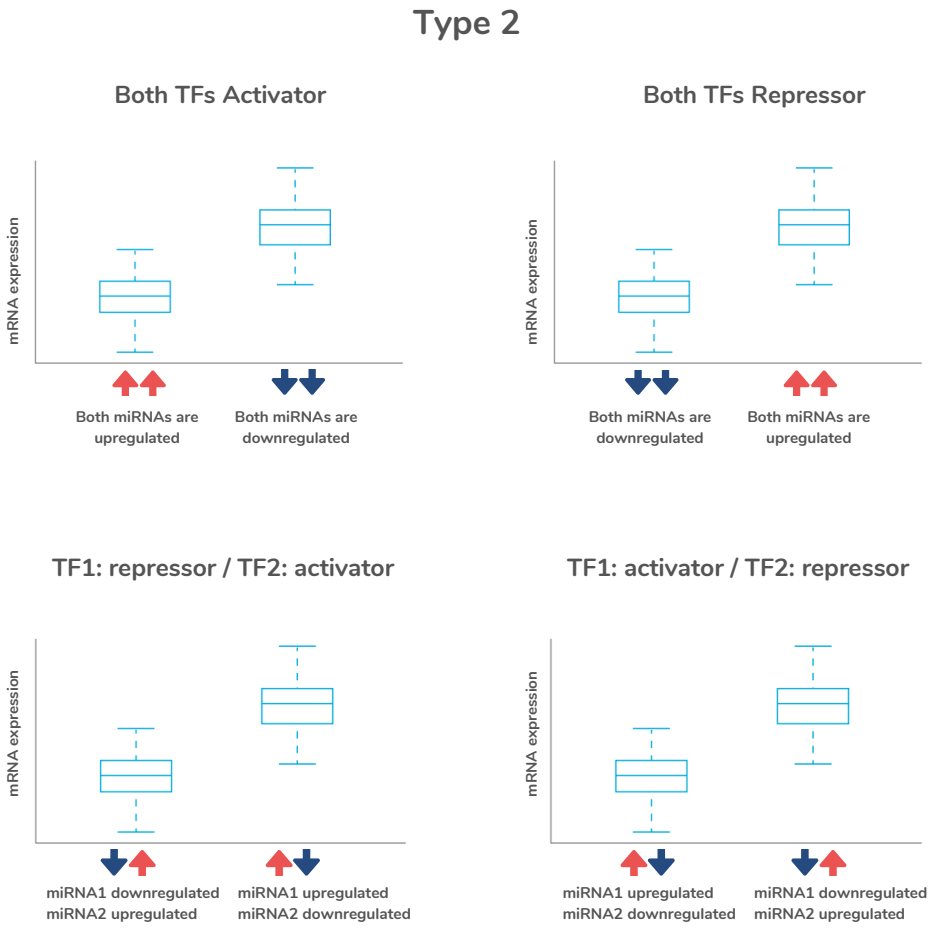
The potential triplets obtained in the first step were filtered according to the expression of miRNA and mRNAs. Patient samples were divided into different subgroups in order to identify candidate triplets. For Type 1 as it was shown in Figure 3.2 if TF is an activator, it is expected that the gene will be down-regulated in cases where the expression of both miRNAs is up-regulated and up-regulated in cases where it expression of both miRNAs is down-regulated. If TF is a repressor, the gene is expected to be up-regulated if the miRNA acting through TF is up-regulated and the other miRNA is down-regulated, and the gene is expected to be down-regulated if the miRNA acting through TF is down-regulated and the other one is up-regulated.

Figure 3.2 Overview of the second step of the algorithm for synergy modules type 1.



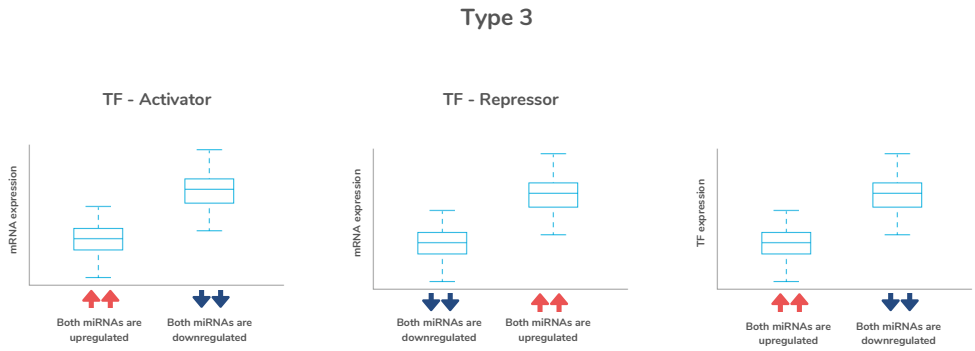
For Type 2, if both TFs are activators and miRNAs are down-regulated, the gene is expected to be up-regulated, and if miRNAs are up-regulated, the gene is expected to be down-regulated. If both TFs are repressors and miRNAs are up-regulated, the gene is expected to be up-regulated, and if the miRNAs are down-regulated, the gene is also expected to be down-regulated. If one of the TFs is a repressor, the other one is an activator, and the miRNA binding to the repressor TF is up-regulated and the other miRNA is down-regulated, the gene is expected to be up-regulated, and for the inverse, the gene is expected to be down-regulated (Figure 3.3).

Figure 3.3 Overview of the second step of the algorithm for synergy modules type 2.



Finally for Type 3, as Figure 3.4 illustrates, If TF is a repressor and miRNAs are up-regulated, the gene is expected to be up-regulated, and if miRNAs are down-regulated, the gene is also expected to be down-regulated. If TF is an activator and miRNAs are up-regulated, the gene is expected to be down-regulated, and if miRNAs are down-regulated, the gene is expected to be up-regulated. In addition, TF is expected to be down-regulated when miRNAs are up-regulated and up-regulated when miRNAs are down-regulated.

Figure 3.4 Overview of the second step of the algorithm for synergy modules type 3.



When examining all cases, a one-sided Wilcoxon rank-sum test (Wilcoxon, 1945) was applied to test the significance of the difference in mRNA level between the first and second groups. Cases, where the null hypothesis was rejected (p-value 0.05), were kept as candidates for the next step.

3.2.3 Step 3: Statistical Interaction Tests on RNA Expression Data

Each potential synergy module that passed the second step was subjected to statistical kernel interaction tests on RNA expression data. For Type 1, we are looking for situations where the gene is pairwise independent with each miRNA, situations where miRNA which acts through TF, and TF are dependent, and situations where miRNA1-miRNA2-target gene forms mutually dependent triplet. For Type 2, we are looking for situations where the miRNAs and their target TFs are pairwise dependent, miRNAs and target gene pairwise independent, and situations where miRNA1-miRNA2-gene forms mutually dependent triplet. Lastly, for Type 3, we are looking for situations where each miRNA and TF are pairwise dependent, where each miRNA and target gene are pairwise independent, and where miRNA1-miRNA2-gene forms mutually dependent triplet.

For all three different models, Two-Variable interaction tests and kernel three-variable tests were applied (Sejdinovic et al., 2013).

In the following section, we present the data sources used in the thesis in more detail.

3.3 Implementation Details and Parallelization

I acknowledge that the parallelization part is carried out by Arda Sener under the guidance of Dr. Kamer Kaya.

The statistical tests utilized in the earlier version of miRCoop were implemented using MATLAB. However, due to the relatively large dataset, certain sections of the code were observed to be slow in processing the data, which prohibited conducting the analysis on all cancer types. This limitation in scalability resulted in a bottleneck, as it required substantial computing power. In this research, we used the previous version as a reference to measure speed improvements and validate the accuracy of the outcomes.

To overcome the performance challenges, we designed and parallelized the algorithms with a high-performance computing approach. The implementation was optimized to concurrently run on multiple nodes and multiple Graphics Processing Units (GPUs) within a single node. As a result, we achieved a 1950x speedup compared to the performance of the original MATLAB code running on a CPU. This enhancement allowed for much faster and more efficient analysis of miRNA interactions and synergy modules across multiple cancer types.

After conducting a thorough profiling of the original version, we identified that the major bottlenecks were the 2-variable and 3-variable Lancaster interaction tests (Sejdinovic et al., 2013). These tests involved a large number of random permutations (in our analysis 2000 per trial) and element-wise matrix-matrix multiplications for Monte-Carlo simulations. Additionally, our previous approach involved multiple independent trials, each of which was computationally intensive.

In the current version, we have addressed these performance issues by distributing the trials to multiple nodes and GPUs. Each GPU is assigned a number of trials, and each trial is executed concurrently by assigning a permutation to a single GPU thread block. This distributed approach significantly improves the computational efficiency and allows for faster execution of the statistical tests.

During our profiling, we observed that the interaction tests were the most time-consuming part of the execution. To address this, we utilized ArrayFire (Malcolm et al., 2012) for the remaining computations, which were less computationally intensive compared to the interaction tests. ArrayFire's CUDA backend allowed us to efficiently implement the required linear algebra operations on the GPU.

By using ArrayFire, we could easily access the underlying CUDA arrays, which facilitated seamless integration with our custom CUDA kernels. As a result, we could keep the data on the GPU throughout the entire execution until we obtained the final results. This approach eliminated the need for data copy operations between the host's main memory and the GPU memory, significantly reducing the overhead and improving the overall performance.

To conduct a comprehensive experimental study and demonstrate the performance boost on various architectures and GPUs, we evaluated the code on two standalone servers and a cluster, encompassing three distinct architectures:

- The first, Arch-1, is a standalone server with an Intel Xeon Gold 6152 CPU that runs at 2.10GHz and 1TB of memory. (No GPU runs were performed on this system)
- The second, Arch-2, is a standalone server having four GPUs, two Nvidia GTX 980 (4GB memory, 2048 cores) and 2 Nvidia Titan X (12GB memory, 3584 cores) GPUs, 198GB RAM, and two Intel Xeon E5-2620 v4 processors, each with 8 CPU cores running at 2.10 GHz.
- Lastly, Arch-3 has eight Nvidia A100 GPUs with 80GB of memory, 64 cores on an AMD 7742 CPU that runs on 2.24GHz, and 512GB of memory.

The servers used in the experimental study are running different operating systems: Ubuntu 20.04.3 LTS, Ubuntu 20.04.2 LTS, and Red Hat Enterprise Linux 8. Each server's CPU code was compiled with gcc version 7.0.1, using the -O3 flag for optimization. The GPU codes were compiled with nvcc and CUDA 10.1, also using the -O3 optimization flag. The GPUs are connected to the server via PCIe 3.0 x16, and all relevant data structures are efficiently transferred and stored in the GPU memory.

3.4 Dataset and Data Processing

3.4.1 miRNA-mRNA Target Data

Table 3.1 Speed-up measurements for the miRCoop algorithm.

	Machine Specifications	Number of Nodes	Time(s)	Speed-up
Original	Single Thread	-	14024.11	1.00
Original	16 Threads	-	24140.54	0.58
Original	32 Threads	-	25379.38	0.55
Parallelized	2x Nvidia GeForce GTX 980	-	182.49	76.85
Parallelized	2x Nvidia TITAN X	-	74.98	187.04
Parallelized	Nvidia A100	-	26.11	537.12
Parallelized	2x Nvidia A100	-	14.00	984.84
Parallelized	4x Nvidia A100	-	6.96	2014.96
Parallelized	4x Nvidia A100	2 nodes each	3.51	3995.47
Parallelized	4x Nvidia A100	4 nodes each	2.35	5967.71

Interactions between miRNA and mRNA were obtained from three experimentally validated databases, namely miRTarBase v8.0 (Huang et al., 2020), Tarbase v7.0 (Vlachos et al., 2015), and miRecords v4.0(Xiao et al., 2009). The TargetScan v7.0 (Agarwal et al., 2015) algorithm, a miRNA-mRNA target prediction tool, was also utilized to acquire binding regions of the miRNAs. To ensure stringent results, predictions from TargetScan (Agarwal et al., 2015) for 6-mer sites were excluded due to their poor conservation(Olgun and Tastan, 2020). Additionally, to obtain miRNA targets from TargetScan (Agarwal et al., 2015), the Ensembl BioMart data mining tool (Kinsella et al., 2011) was employed to extract the 3'UTR, 5'UTR, and coding sequences of protein-coding genes. The miRBaseConverter, an R package (Xu et al., 2018), was used to obtain miRNA accession codes and seed sequences (version = "22"). To eliminate any potential conflicts arising from gene name aliases, gene symbols, and gene IDs, the identifiers were transformed into Entrez gene IDs (Durinck et al., 2009).

3.4.2 Expression Data

Only primary tumor samples that had both mRNA and miRNA data available were used for the analysis of mRNA and miRNA expression data. The mature miRNA expression data was obtained by retrieving reads per million mapped reads (RPM) from FireBrowse (http://firebrowse.org) and with a log2 transformation. miRNAs with NA values exceeding 20% of the samples and those that displayed no variation across samples were filtered out. To perform this filtering, miRNAs with a median absolute deviation (MAD) across samples below an absolute value of 1 were excluded.

To acquire the mRNA expression data, we retrieved mRNA expression levels using Fragments per Kilobase of transcript per Million mapped reads upper quartile normalization (FPKM-UQ) from the GDC Data Portal (https://portal.gdc.cancer.gov). In the RNA-Seq data, noncoding genes were extracted, and only protein-coding transcripts were considered for the analysis. To ensure robustness, genes that displayed low expression across a substantial number of samples were filtered out. Specifically, genes with expression values below 0.05 in more than 20% of the samples were excluded. Moreover, genes with low variability across samples, indicated by a median absolute deviation (MAD) lower than 0.5, were removed. To address the issue of zero gene expression values, a constant value of 0.25 was added to all expression values before applying a log2 transformation.

3.4.3 TF Regulation Data

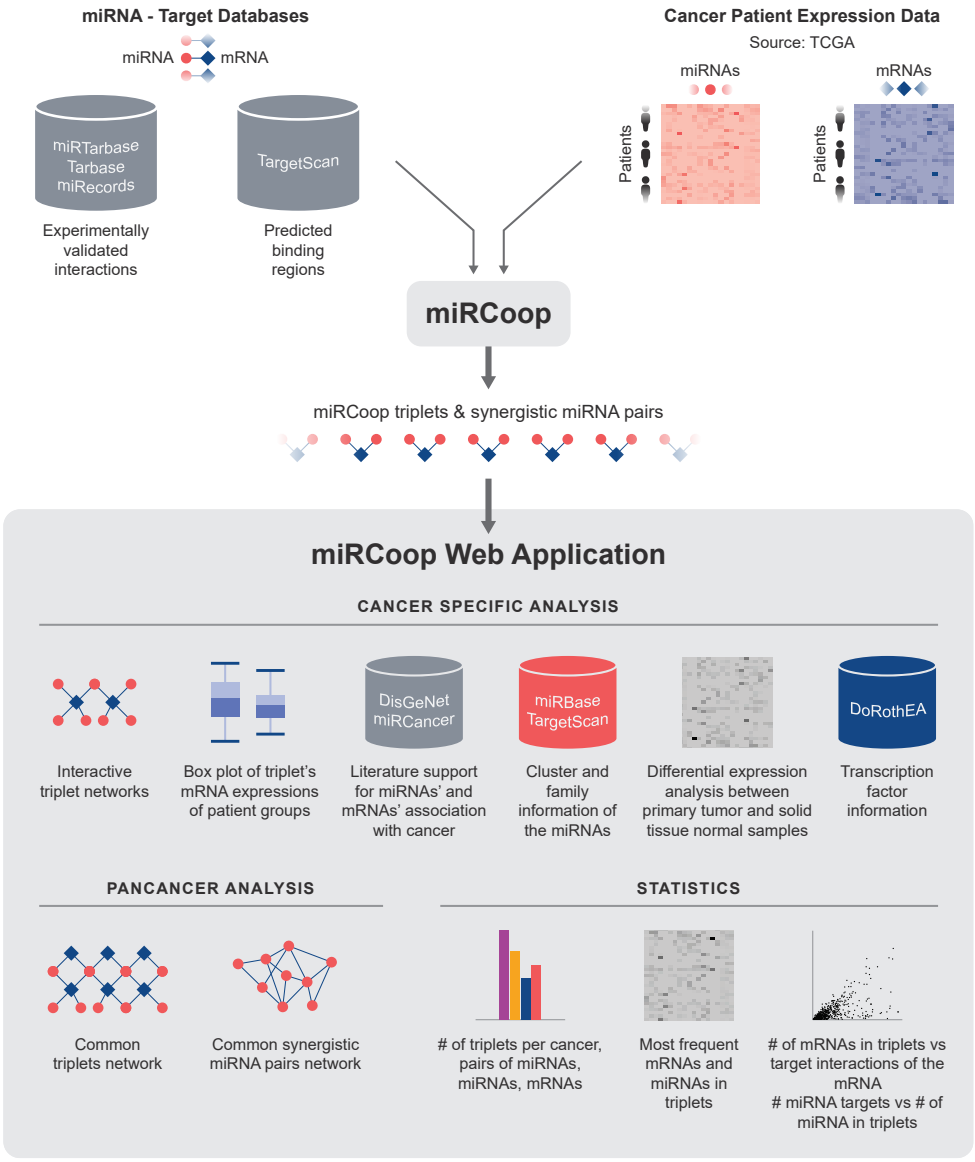
TF-Gene information was obtained from DoRothEA (Garcia-Alonso et al., 2019). The DoRothEA database encompasses TF-target interactions for 1541 human TFs. These interactions are derived from various sources, including manually curated resources, ChIP-seq binding experimental data, prediction of TF binding motifs based on gene promoter sequences, and inference from GTEx data. Each TF-target interaction in DoRothEA is accompanied by a confidence score, which is assigned based on observations from their benchmark ranging from A (highest confidence) to E (lowest confidence). In the analysis, interactions with a confidence level of C and above were used.

4. RESULTS

4.1 miRCoop Web Application

The miRCoop algorithm has been updated to calculate miRCoop triplets, and a newly developed synergy modules algorithm has been implemented to predict miRCoop synergy models. Users can access and view these miRCoop triplets and synergy modules through the web application. This platform is built using the Shiny framework (version 1.6) and incorporates various other R packages. Access to the application is available at http://mircoop.sabanciuniv.edu. The interface is designed to be easily navigated by the users without prior programming experience. An overview of the miRCoop v2 interface is depicted in Figure 4.1.

Figure 4.1 General overview of the miRCoop Web Application



The current version of the miRCoop web application comprises three main components: cancer-specific analysis, pan-cancer analysis, and statistics for both the miRCoop and miRCoop Synergy Interactions modules.

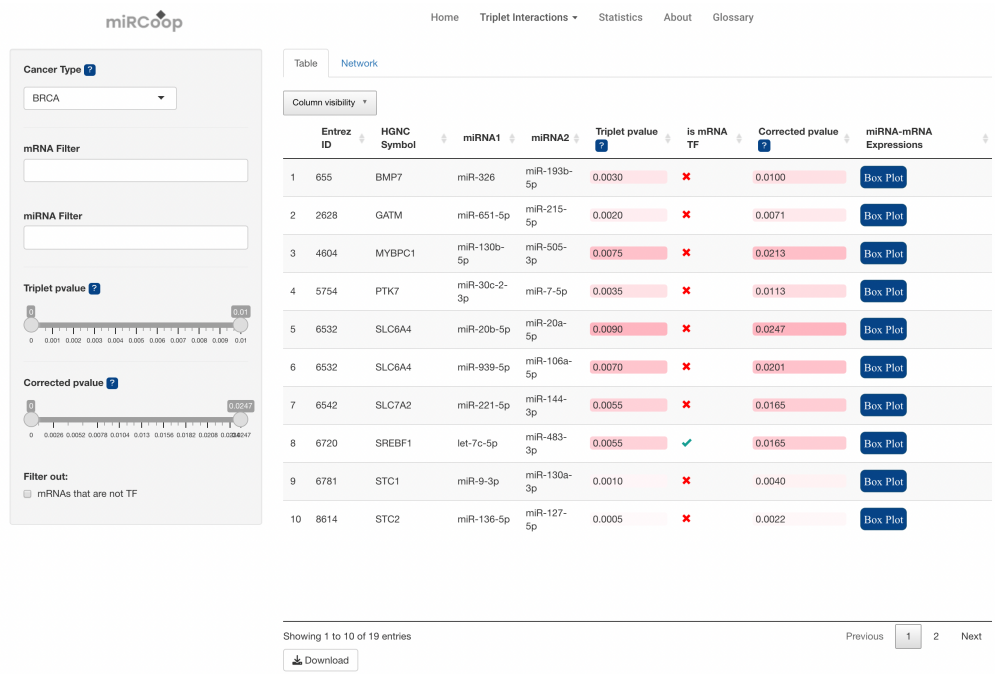
4.1.1 miRCoop Triplet Interactions

miRCoop v2, an upgraded pipeline that utilizes kernel interaction tests to identify direct synergistic pairs of miRNAs and mRNAs. miRCoop v2 was applied to analyze 31 different TCGA cancers. These results, including candidate cancer triplets, synergistic miRNA pairs, and their common targets, are now available on the miRCoop web application for public access.

4.1.1.1 Cancer Specific Analysis

Within this interface, users have the ability to explore predicted miRCoop triplets and miRNA pairs. Initially, users can choose a specific cancer type from the available 31 TCGA cancer types included in the study. This selection is made from the cancer-type section located on the left side panel. Once a cancer type is chosen, the identified triplets for that particular cancer are displayed to the user in a structured format within the main panel. This section primarily focuses on presenting the identified triplets, along with their corresponding miRNA pairs and target mRNAs (represented by both Entrez ID and HGNC Symbol). Additionally, the triplet p-values and corrected p-values, determined through the Benjamini Hochberg procedure, are provided (Figure 4.2).

Figure 4.2 Main panel for triplets. Here BRCA is selected as the cancer type.



By utilizing the column visibility button, users can enhance the displayed data with supplementary information. This additional information is as follows:

- *Experimental data source of miRNA - mRNA relationships:* From which experimentally validated databases (miRTarBase v8.0, TarBase v7.0 and miRecords v4.0) the miRNA-mRNA target interactions derived can be accessed.

- *Differential expression analysis results:* Differential expression analysis was conducted by comparing the expression levels of each miRNA and mRNA between primary tumor samples and solid tissue normal samples. The resulting data includes the log fold-change ($\log_2$FC) values and relative p-values for each miRNA-mRNA triplet. Positive $\log_2$FC values indicate upregulation in the primary tumor samples, while negative values indicate downregulation, compared to solid tissue normal samples. A p-value threshold of less than 0.05 was used to determine statistically significant differences in $\log_2$FC values.

- *Literature support of cancer-miRNA and cancer-mRNA relationship:* The literature support relevant to the relationship between specific miRNAs, mRNAs, and cancer types is also provided to the user. Information on the miRNA-cancer relationship is obtained from miRCancer (Xie et al., 2013), while data on mRNA-cancer relationships are collected from DisGeNet (Piñero et al., 2015).

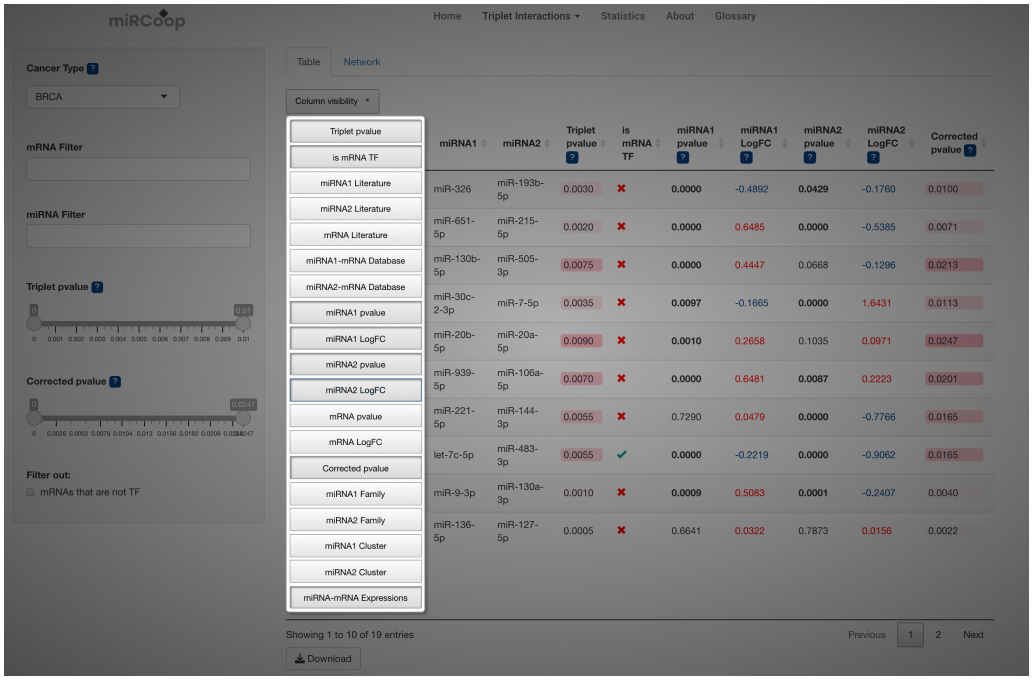
- *mRNA expressions of patients grouped by miRNA expression levels:* Patients are grouped into two according to the expression level of the miRNA pair. Two patient groups are formed for comparison: one group with downregulated expression of both miRNAs and another group with upregulated expression of both miRNAs. The difference as the log fold-change value are measured, and those triplets above a certain threshold are maintained. The mRNA levels of these two groups are expected to exhibit significant differences. We visualize these differences as a boxplot, allowing for a graphical representation of the expression levels of mRNAs in the two patient groups.

- *TF information:* Information regarding whether the mRNAs in the triplets function as a TF is also included in the data table. This additional information may help in understanding the regulatory relationships within the triplets. The TF list used in this analysis has been obtained from DoRothEA (Garcia-Alonso et al., 2019), a comprehensive resource for transcription factor-target interactions.

- *miRNA Family and Cluster:* The miRNA family and cluster information for each member of the pair can be displayed by selecting the respective columns.

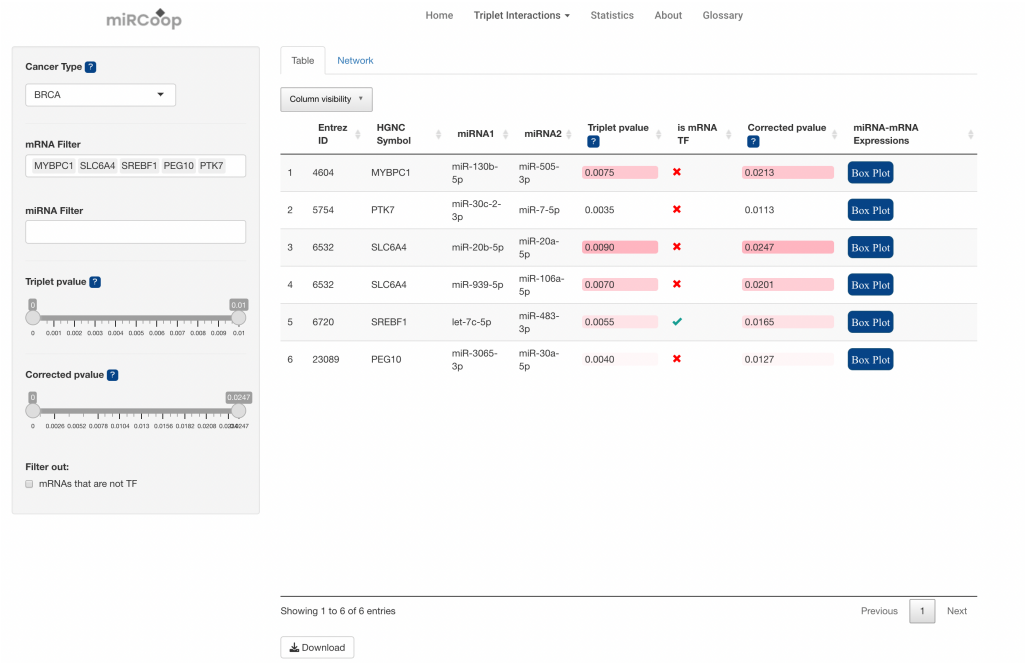
This additional information can be accessed and incorporated into the displayed data through the column visibility feature(Figure 4.3).

Figure 4.3 Additional information can be added to the main data table from the Column Visibility part.



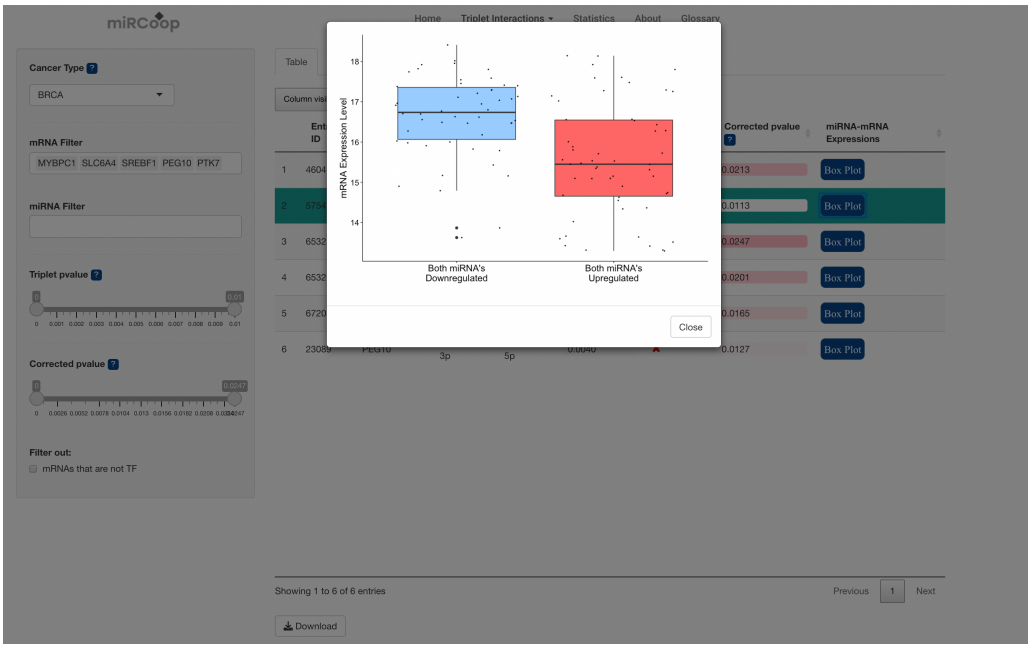
Users have various options to filter the results in this section. Users can filter the identified triplets with their respective triplet p-value, and corrected triplet p-value. Also, users can filter by whether the mRNA in triplets targeted by the miRNA pair is a transcription factor(Figure 4.4).

Figure 4.4 Example of filtering for triplets.



As part of algorithm step 2, we implemented a filtering process to group patients based on the expression level of the miRNA pair. The difference in expression is measured as a log fold change, and only patients above a specific threshold were considered for further analysis. In this context, a comparison is made between the two patient groups based on the miRNA pair's expression level. It is expected that the mRNA levels of these two groups (both miRNAs upregulated and both miRNAs downregulated) will exhibit significant differences. To visualize this difference, we provide a boxplot in the 'miRNA-mRNA Expressions' column, allowing for easy observation of the variation between the two patient groups(Figure 4.5).

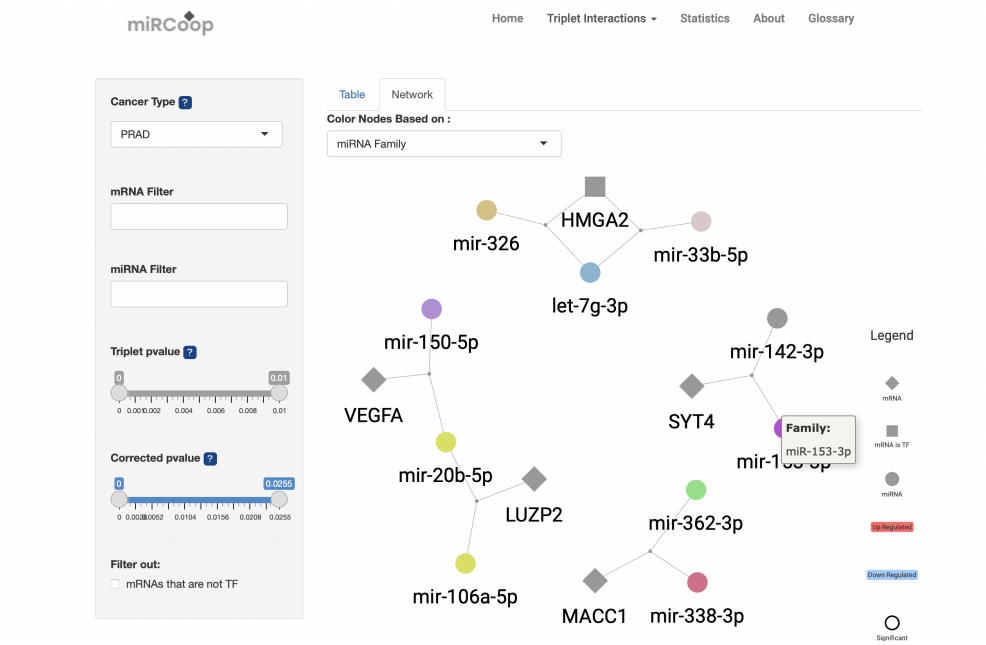
Figure 4.5 An example boxplot illustrating the grouping of mRNA expression levels based on miRNA expression levels.



For specific cancer, users have the ability to interactively visualize triplet networks. These networks consist of diamonds representing mRNAs and circles representing miRNAs from the triplets. Additionally, small grey nodes are included as dummy nodes to establish connections between the miRNAs and their target mRNA. If an mRNA within a triplet also functions as a TF, its shape changes from a diamond to a square. Node coloring within the network can be customized using three different approaches: Differential expression analysis results, which determine color based on upregulation or downregulation, miRNA family information, enabling coloring based on specific miRNA families and miRNA cluster information, allowing nodes to be colored based on miRNA clusters. When the user applies filters to the data, the network is dynamically reconstructed to reflect the updated data and filtering criteria.

When the user opts to color the networks based on family or cluster information, simply hovering the mouse over a miRNA node provides immediate insight into its corresponding family or cluster. The family information is displayed through coloring based on the family, while the cluster information is conveyed through coloring based on the cluster. This feature allows users to easily identify the family or cluster affiliation of a miRNA node by simply hovering over it with the mouse cursor(Figure 4.6).

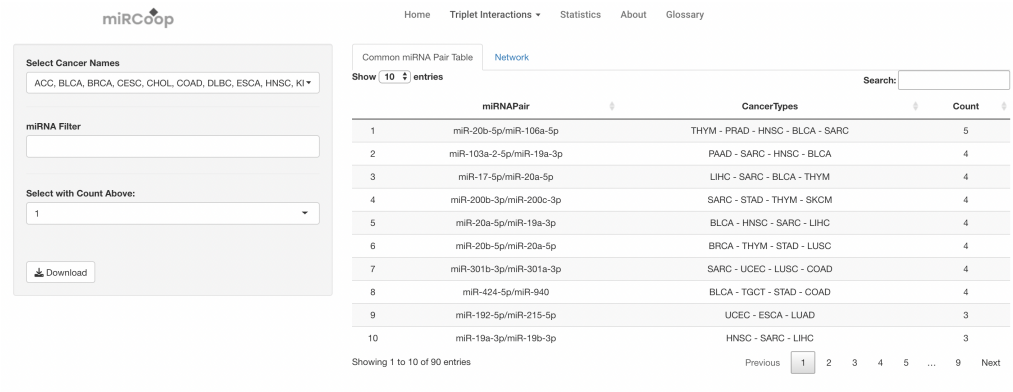
Figure 4.6 Interactive triplet network for PRAD. Network coloring option selected as miRNA family. Family information of the miRNA can be seen when hovered to node



4.1.1.2 Pan-Cancer Analysis

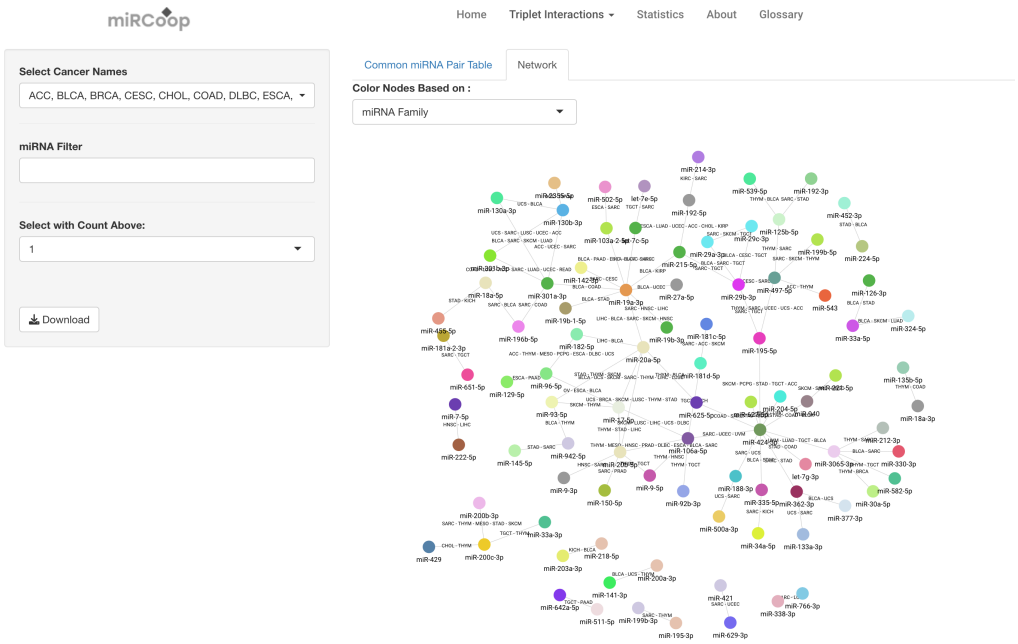
Users have the option to explore pan-cancer miRNA pairs and miRCoop triplets by selecting the corresponding sections. The interface provides several filtering options for enhanced customization. Users can filter pan-cancer triplets and synergistic pairs based on cancer type, the number of occurrences across different cancers, as well as specific miRNAs and mRNAs, all accessible from the left side panel (Figure 4.7).

Figure 4.7 Main panel for pan-cancer synergistic miRNA pairs



Additionally, users can generate interactive networks for pan-cancer miRNA pairs and triplets by navigating to the network tab, similar to the functionality available in the Cancer-Specific Triplets section. When users apply filtering options from the left panel, the network is dynamically reconstructed to reflect the updated filters and display the relevant information accordingly (Figure 4.8).

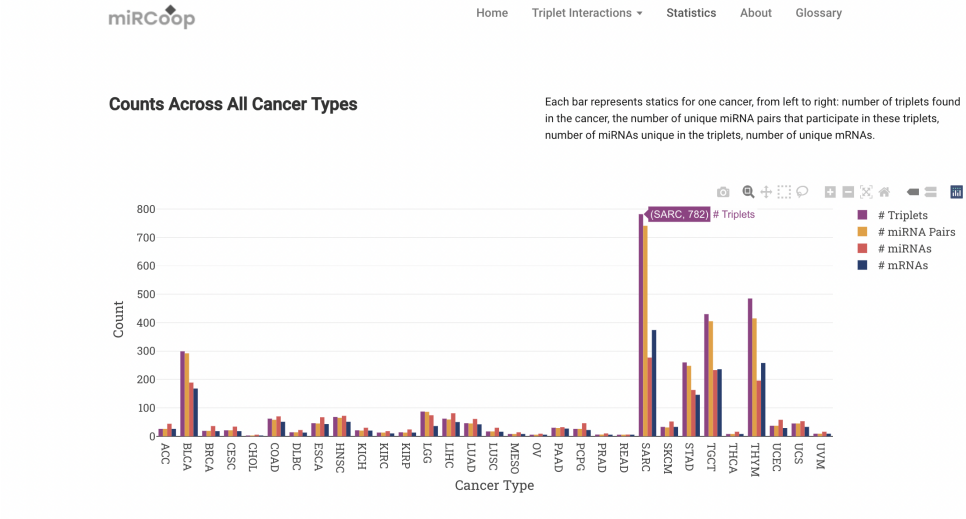
Figure 4.8 Interactive network for pan-cancer synergistic miRNA pairs



4.1.1.3 Overall Statistics on Triplets

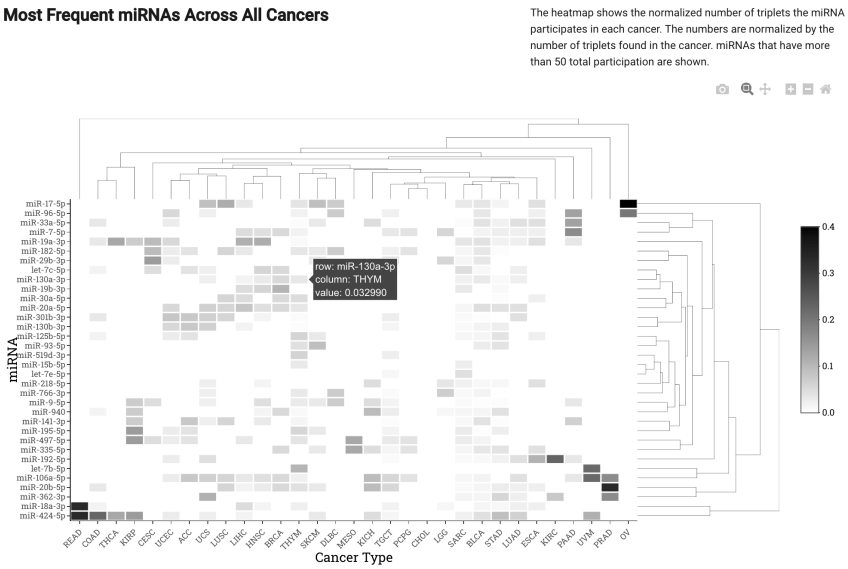
On the website, we provide various statistical insights regarding the dataset and results. This includes counts across all cancer types, such as the number of triplets in each cancer, the count of unique miRNA pairs involved in these triplets, the number of unique miRNAs present in the triplets, and the count of unique mRNAs. We provide these statistics as a bar graph (Figure 4.9). When users hover their mouse over the graph, they can view the specific numerical values associated with each category, allowing for a detailed understanding of the data.

Figure 4.9 Total counts for all cancer types are presented to users.



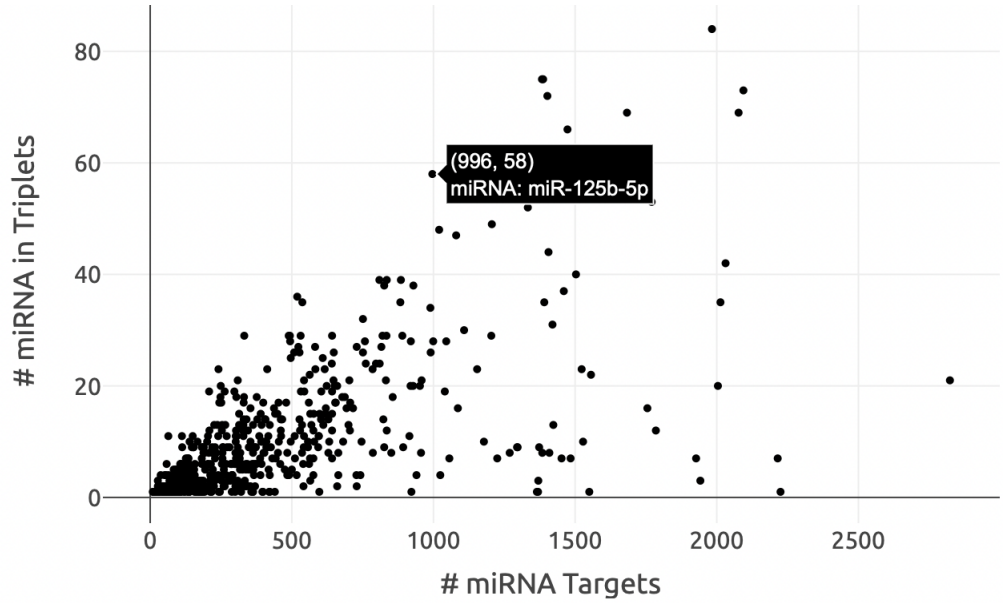
In order to analyze the distribution patterns of miRNAs and mRNAs with higher overall frequencies, we examined their presence across specific cancers or multiple cancer. When users hover their mouse over a specific cell of interest, they can view explicit information such as the miRNA/mRNA, the corresponding cancer type, and its respective value(Figure 4.10). This feature allows for a detailed exploration of the distribution and frequency of specific miRNAs and mRNAs across different cancer types.

Figure 4.10 Most frequent miRNAs across all cancer are displayed.



Users are provided with visual representations of the number of miRNA targets versus the occurrence of miRNAs within triplets. By hovering over the data points, users can easily access specific information. For instance, miRNA miR-125-5p is associated with 996 targets and is observed in 58 triplets (Figure 4.11). Additionally, the number of mRNAs present in triplets across all cancers and the corresponding count of mRNA-target interactions are also displayed.

Figure 4.11 The number of miRNAs that target the mRNA plotted against the number of triplets the miRNA participated in.



These features enable users to conveniently explore the relationships between miRNAs, mRNAs, their targets, and their presence within triplets.

4.1.2 miRCoop Synergy Modules

The miRCoop synergy modules pipeline utilizes kernel interaction tests to identify synergistic miRNAs and their targets through TFs. This pipeline has been applied to analyze various TCGA cancers. The results, including candidate synergy modules, synergistic miRNA pairs, and their common targets, are now accessible to the public through the miRCoop web application.

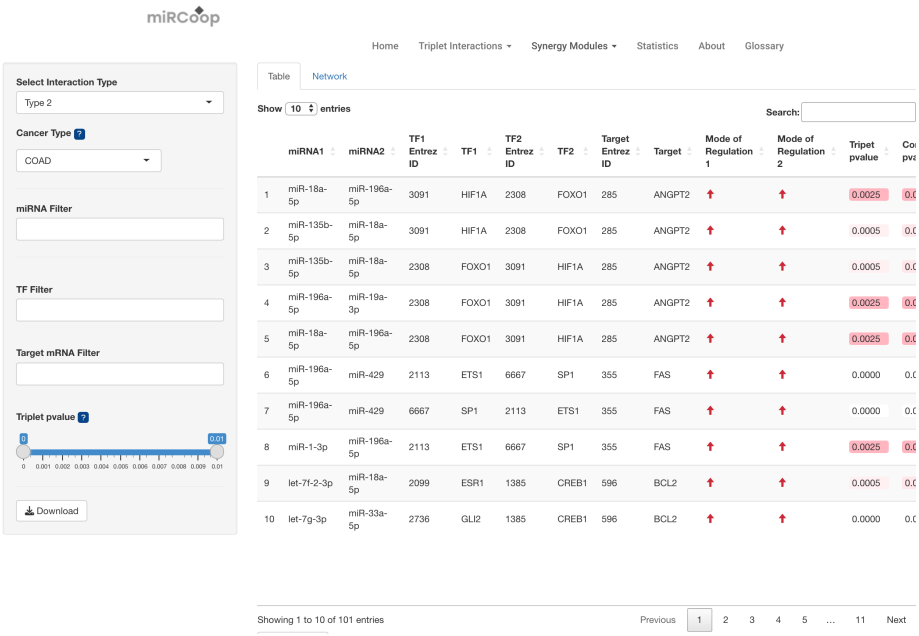
4.1.2.1 Cancer Specific Analysis

The user interface of the miRCoop web application provides users with the ability to explore predicted miRCoop synergy modules and miRNA pairs. The interface offers options to select a specific synergy module type (Type 1, Type 2, or Type

3) and a cancer type from the available TCGA cancer types. Once a cancer type is chosen, the identified synergy modules for that specific cancer are displayed in a structured format within the main panel. The displayed information includes the miRNA pairs, TFs, and target mRNAs(represented by both Entrez ID and HGNC Symbol) associated with each synergy module. Additionally, the p-values and corrected p-values of the synergy modules, determined through the Benjamini-Hochberg procedure, are provided.

Users can utilize various filtering options to narrow down the results, such as filtering by specific miRNAs, TFs, target mRNAs, and synergy module p-values. For a specific cancer, a web application allows users to interactively visualize synergy module networks. The network is dynamically reconstructed based on the applied filters and the updated data that meet the filtering criteria. This interactive visualization feature enables users to explore the connections between miRNAs, TFs, and target mRNAs within the selected cancer, providing a more comprehensive understanding of the regulatory interactions in the context of specific cancer types.

Figure 4.12 Main panel for synergy modules. Here Type 2 is selected as the interaction type and COAD is selected as the cancer type.



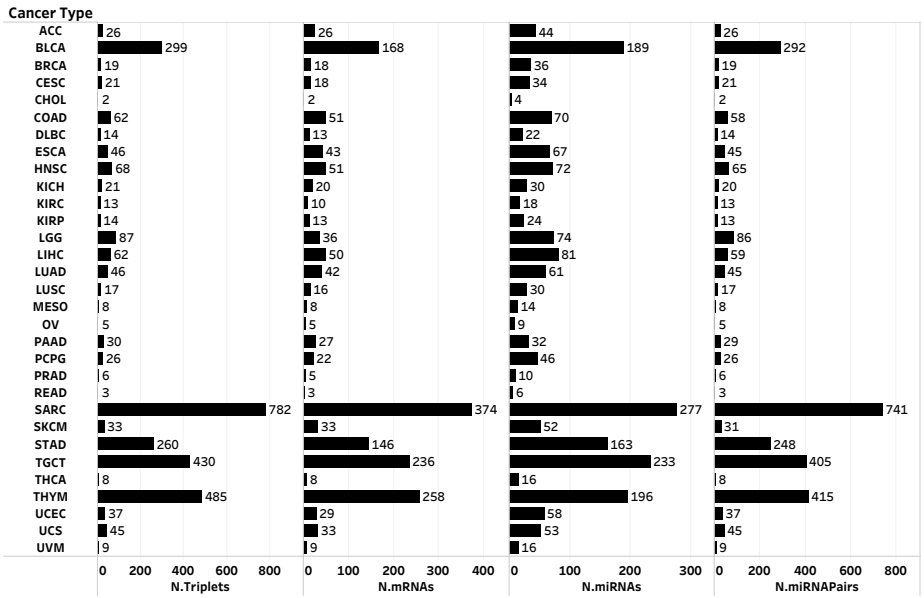
4.1.2.2 Pan-Cancer Analysis

The miRCoop web application offers users the flexibility to explore pan-cancer miRNA pairs and miRCoop synergy modules through the dedicated sections for Type 1 and Type 3 interactions. By selecting these sections, users can access the corresponding results specific to their interests. The left side panel provides various filtering options. Users can filter the pan-cancer synergy modules and synergistic pairs based on cancer type, the frequency of occurrence across different cancers, as well as specific miRNAs and mRNAs. These filtering options enable users to customize their exploration and focus on the specific aspects of interest within the pan-cancer miRNA regulatory networks. Moreover, within the miRCoop web application, users can generate interactive networks for pan-cancer miRNA pairs and synergy modules by accessing the network tab. By utilizing the filtering options available in the left panel, users can customize the displayed network to show the desired information. When users apply specific filters, such as cancer type or miRNA/mRNA selection, the network is dynamically reconstructed to reflect the updated filters and display the relevant miRNA pairs and synergy modules.

4.2 Overview of the miRCoop Triplets and Synergy Modules and the Collaborative miRNA pairs

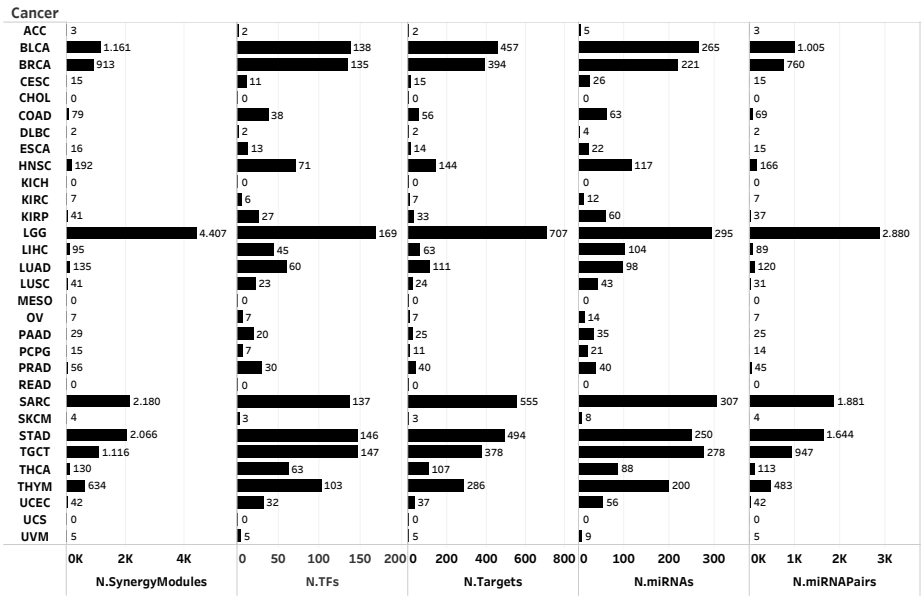
For the direct triplet interactions, the previously described miRCoop pipeline has been updated on several aspects. It has now been applied to identify cooperating pairs and triplets of miRNAs across 31 different types of cancers. In Figure 4.13, you can find a summary of the statistical information, including the number of triplets in each cancer, the count of unique synergistic miRNA pairs involved in these triplets, the number of unique miRNAs found within the triplets, and the number of unique mRNAs. The cancers that exhibited the highest numbers of triplets and miRNA pairs were SARC and THYM. Additionally, these cancers displayed a greater number of mRNAs within these triplets compared to other types of cancer. On the other hand, CHOL and READ had the lowest counts of miRNA pairs and triplets.

Figure 4.13 Each bar represents statistics for one cancer in miRCoop v2 triplets. From left to right: number of triplets, unique mRNAs, unique miRNAs, and unique miRNA pairs



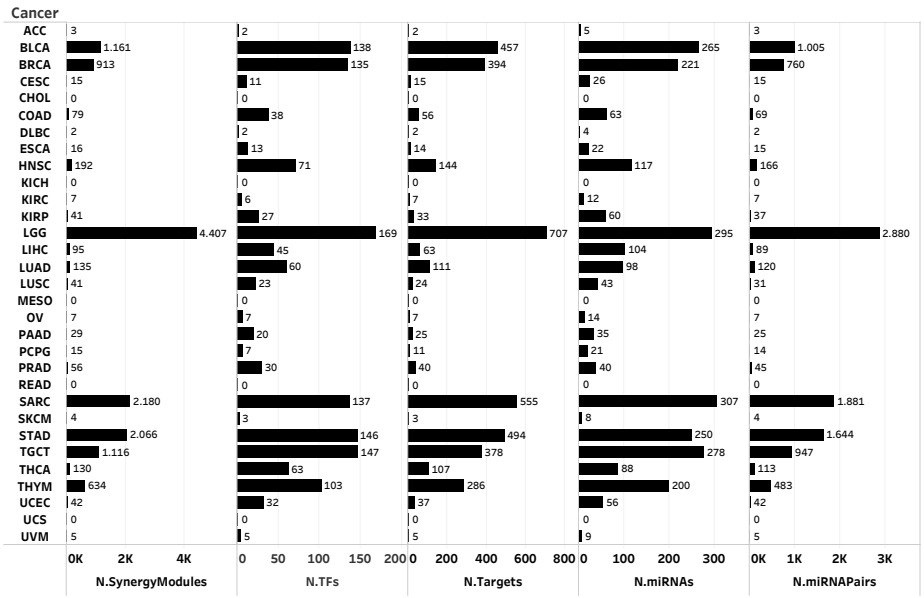
Furthermore, a novel algorithm has been developed that enables the identification of synergistic miRNAs targeting the same gene through TFs, which are called synergy modules. Summary statistical information was provided for these interactions as well. Among all cancer types, LGG displayed the highest number of synergy modules and synergistic miRNA pairs in synergy modules Type 1(Figure 4.15). SARC and STAD followed, exhibiting considerable numbers of synergy modules and synergistic miRNA pairs. However, no synergy module was observed in the cancer types CHOL, KICH, MESO, READ, and UCS.

Figure 4.14 Each bar represents statistics for one cancer in Synergy Modules Type 1. From left to right: number of synergy modules, unique TFs, unique target mRNAs, unique miRNAs, and unique miRNA pairs



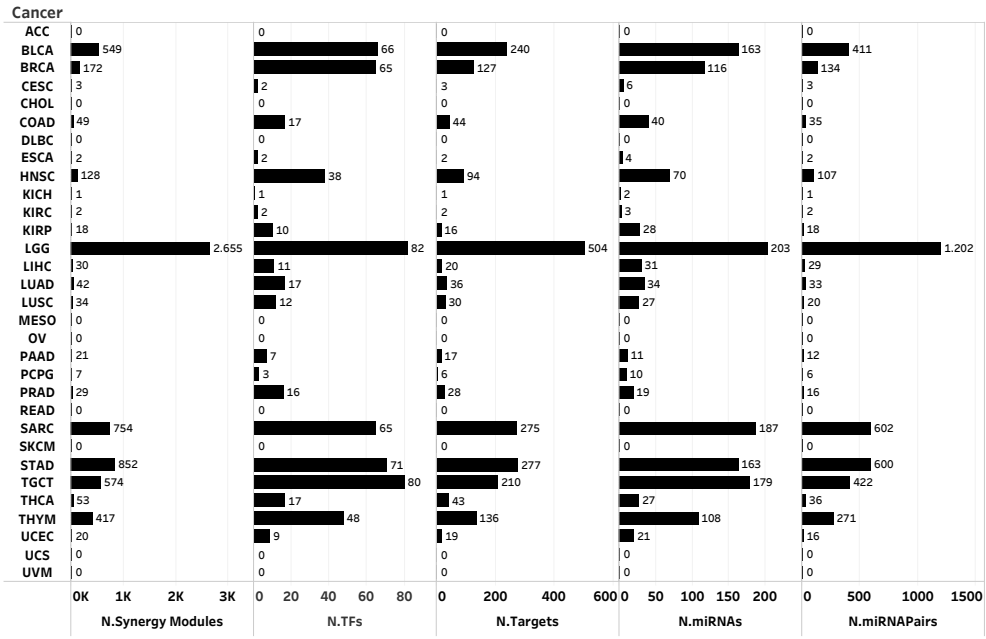
Among the cancers analyzed using the synergy module Type 2, it was observed that COAD had the highest number of identified synergy modules. On the other hand, there were several cancers, including ACC, CHOL, DLBC, KICH, MESO, READ, and UCS, where no synergy modules were detected using this module type.

Figure 4.15 Each bar represents statistics for one cancer in Synergy Modules Type 2. From left to right: number of synergy modules, unique TFs, unique target mRNAs, unique miRNAs, and unique miRNA pairs



Among the synergy module interactions in Type3, LGG exhibits the greatest number of synergy modules and unique miRNA pairs (Figure 4.16). Furthermore, STAD, SARC, TGCT, BLCA, and THYM also display significant synergy modules in Type 3 in TF-synergy module interactions. No synergy modules were observed for TF Type 3 in ACC, CHOL, DLBC, MESO, OV, READ, SKCM, UCS, and UVM.

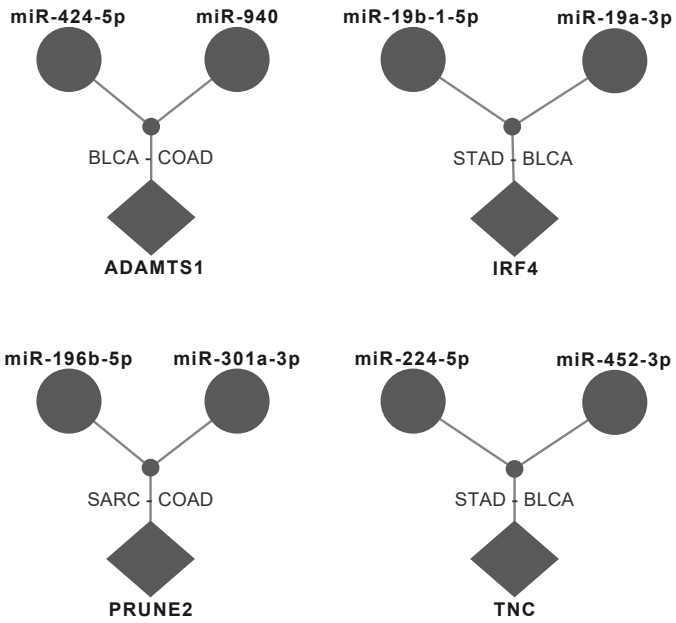
Figure 4.16 Each bar represents statistics for one cancer in Synergy Modules Type 3. From left to right: number of synergy modules, unique TFs, unique target mRNAs, unique miRNAs and unique miRNA pairs



4.3 Analysis of the Triplets and miRNA Pairs across All Cancer Types in miRCoop v2

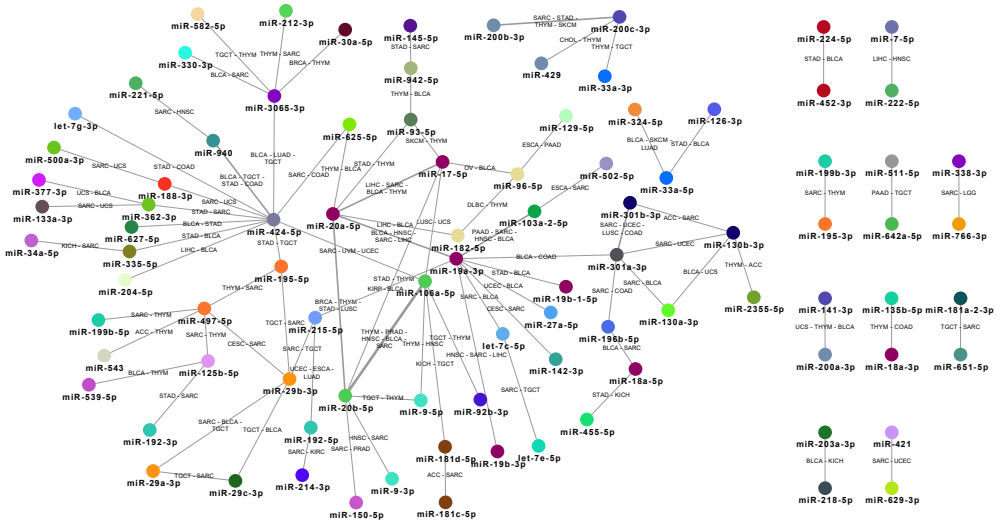
To uncover the highly interacting miRNAs and their corresponding targets, we analyzed common miRCoop triplets and synergistic miRNA pairs. The network depicted in Figure 4.17 demonstrates that four distinct triplets were consistently observed across two different cancer types. Among these triplets, two were found to be shared between STAD and BLCA.

Figure 4.17 Pan-cancer triplets. Circle nodes represent miRNAs and diamond nodes illustrate the mRNAs. Captions on the edges indicate the cancer types that synergistic triplet appears.



Furthermore, Figure 4.18 showcases synergistic pan-cancer miRNA pairs that were found to be common in two or more cancer types. Notably, the miRNA pair miR-20b-5p/miR-106a-5p emerged as the most prevalent across different cancer types. These two miRNAs are in the same family(miR-17) and cluster(miR-106a-363 cluster). This cluster has been associated with different malignancies including breast cancer, lung cancer, diffuse B-cell lymphoma, and Burkitt lymphoma (Li et al., 2018a; Olive et al., 2010; Battistella et al., 2015; Ventura et al., 2008). Also, miR-424-5p has been observed in a wide variety of cancer types (LIHC, COAD, BLCA, STAD, etc..) and has a high degree. According to a recent study, miR-424-5p plays a role in different kinds of cancers. To demonstrate miR-424-5p acts as a tumor suppressor in hepatocellular carcinoma and it was found to have a significant impact on the progression of colon cancer (Xuan et al., 2022).

Figure 4.18 The miRNA pairs that are shared across two or more cancers in triplet interactions are depicted in the network. In the visualization, the nodes represent the miRNAs, while the captions on the edges indicate the cancer types in which the synergistic pairs appear. The thickness of the edges expands according to the number of cancers in which the pair is present. Additionally, the coloring of the nodes encodes the miRNA cluster information.

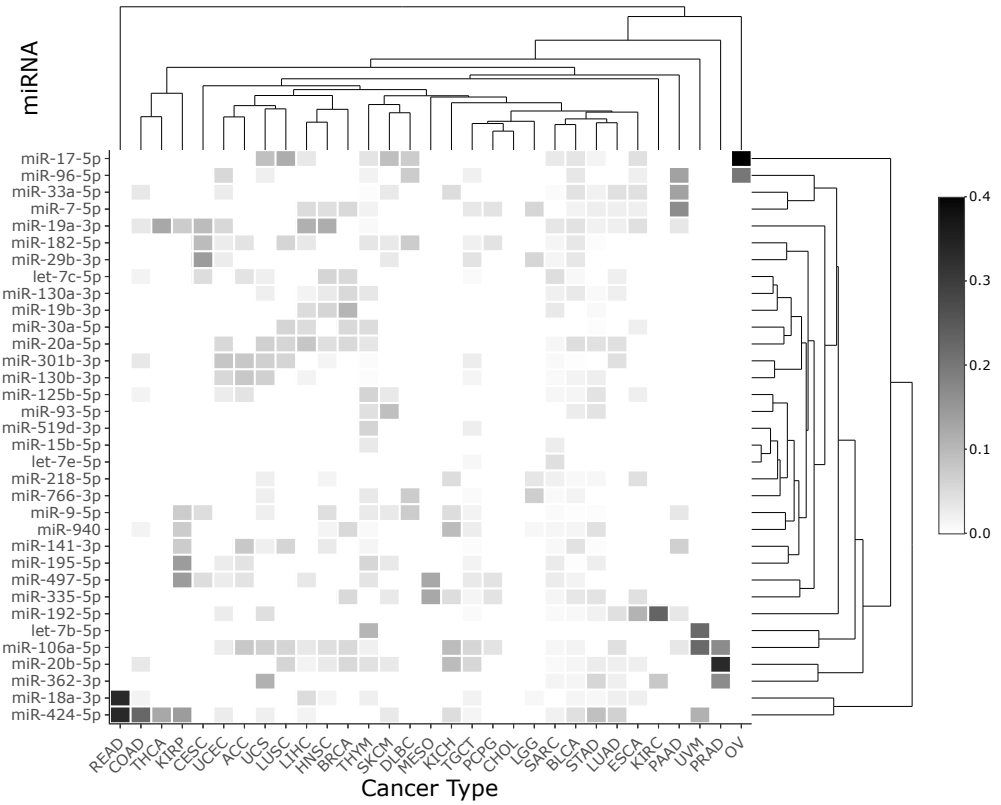


To gain insights into the distribution patterns of miRNAs and mRNAs that showed a higher overall frequency within miRCoop triplets, we conducted an analysis to determine if these molecules were predominantly associated with specific cancer types or exhibited a more widespread presence across multiple cancers. In Figures 4.19 and 4.20, we present the miRNAs and mRNAs, respectively, that participated in more than 30 and 15 miRCoop triplets. The values presented in the figures have been normalized based on the total number of triplets observed in each cancer, ensuring a fair comparison between each miRNA and mRNA.

Previous studies have provided evidence for the involvement of a regulatory loop consisting of Transforming growth factor B1 (TGFB1), miR-20b-5p, and E2F1 in the epithelial to mesenchymal transition (EMT) process in prostate cancer (Qi et al., 2020). Remarkably, our analysis identified miR-20b-5p as one of the most commonly observed miRNAs PRAD (Figure 4.19). Moreover, the downregulation of miR-7-5p in PAAD and its role as a tumor suppressor have been documented in previous studies (Zhu et al., 2018). The reduced expression of miR-7-5p in pancreatic cancer tissues is associated with poor prognosis for patients. Analysis from miRCoop v2 also showed the significant presence of miR-7-5p in the synergy modules of PAAD. Similarly, miR-424-5p, which has been implicated in the cell proliferation, migration,

and metastasis of COAD, was identified as one of the most common miRNAs in COAD by miRCoop v2 (Dai et al., 2020; Cheng et al., 2020).

Figure 4.19 The normalized count of each miRNA that participates in each specific cancer in miRCoop triplets is provided. These numbers are normalized based on the total number of triplets within each cancer. miRNAs that have participated in more than 30 triplets in total are shown. For clustering purposes, the Euclidean distance and complete linkage methods were applied.



Moreover, MKI67 was identified as the most common mRNA in LGG through miRCoop analysis as it was shown in Figure 4.20. MKI67 encodes for a nuclear protein known as Ki-67, which is widely used as a marker of cellular proliferation. In gliomas, Ki-67 labeling is utilized to measure the proliferative activity of tumor cells. The high expression of Ki-67 is associated with poor prognosis in patients with LGG (Dahlrot et al., 2021). Additionally, in THCA, one of the genes that were seen most frequently was FGF2. Downregulation of FGFR2 can serve as an early indicator for the diagnosis of thyroid cancer (Redler et al., 2013).

Figure 4.20 The normalized number of each mRNA that participates in various cancer types in miRCoop triplets is provided. These numbers are normalized based on the total number of triplets within each specific cancer. mRNAs that have more than 15 total participation are shown. To identify relationships and patterns, clustering was performed using the Euclidean distance and complete linkage methods.

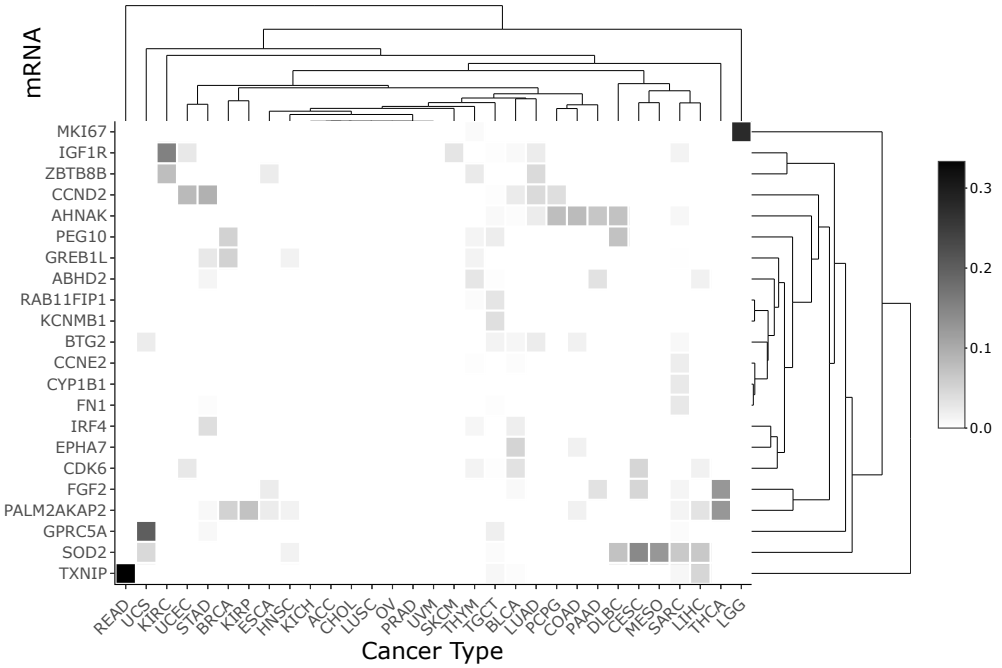
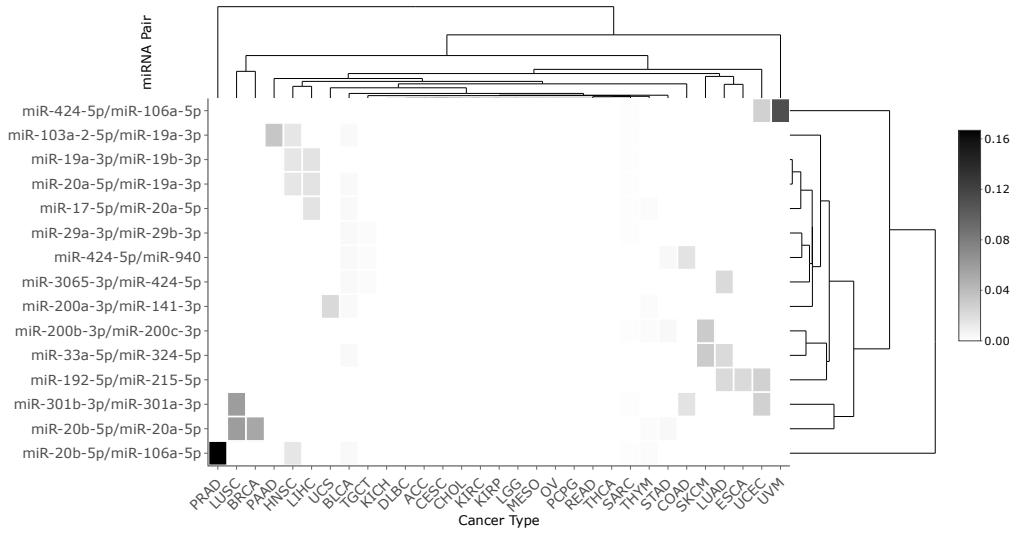


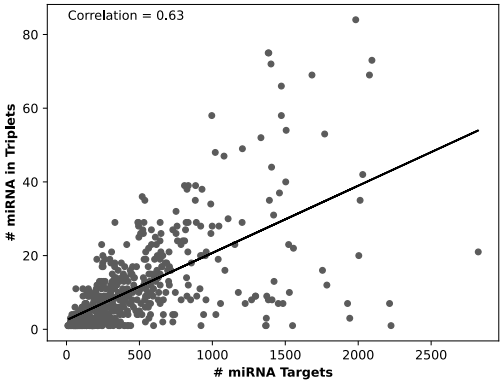
Figure 4.21 depicts the normalized frequencies of common synergistic miRNAs in each specific cancer type. The values shown in the figure have been normalized according to the total number of triplets observed in each cancer. In PRAD, the synergistic pair of miR-20b-5p and miR-106a-5p was commonly observed in the miRCoop triplets. The significance of both miRNAs in prostate cancer has been established in previous studies. For example, aberrant expression of miR-20b-5p has been closely associated with prostate cancer, where it is found to be elevated in prostate cancer cells and contributes to their proliferation (İlhan et al., 2023). Similarly, overexpression of miR-106a-5p has been shown to significantly promote the proliferation, migration, and invasion of prostate cancer cells (Luo et al., 2017). These findings highlight the potential importance of miR-20b-5p and miR-106a-5p as key regulators in prostate cancer development and progression.

Figure 4.21 The normalized number of miRNA pairs in various cancer types in mirCoop triplets is provided. These numbers are normalized based on the total number of triplets within each specific cancer. Only miRNA pairs that have common in more than 2 cancer types were included in the analysis. To identify relationships and patterns, clustering was performed using the Euclidean distance and complete linkage methods.



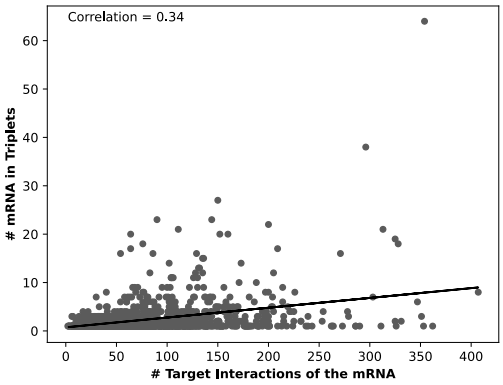
We analyzed the relationship between the number of miRNA targets and the frequency of miRNA existence in triplets across multiple cancer types. Figure 4.22 shows the number of miRNAs in triplets against the number of miRNA targets. Pearson correlation coefficient of 0.63 among the two. Interestingly, it was observed that there were miRNAs that targeted a large number of mRNAs but were not frequently found in triplets. As an example, miR-27a-3p was found to target 2225 mRNAs, indicating its potential regulatory role in a wide range of genes. However, it was only observed in a single triplet across all cancer types included in our analysis. The number of occurrences of miR-27a-3p in the triplets could have been influenced by several factors. Firstly, matched expression data for miR-27a-3p was only available in two out of the thirty-one different cancer types (LGG and SKCM). Consequently, the miRNA was restricted to these specific cancer types for analysis. Furthermore, for miR-27a-3p to be included in the triplets, it had to meet certain criteria set by our filters such as expression filtering in algorithm step 2. Additionally, it was required to exhibit weak or no interaction with its target mRNA.

Figure 4.22 A scatter plot depicting the relationship between the number of miRNAs targeting a specific mRNA and the number of triplets in which the miRNA is present reveals a Pearson correlation coefficient of 0.63 in miRCoop triplets.



The relationship between the number of target interactions of the mRNA and the frequency of mRNAs in triplets was examined across multiple cancer types. The number of mRNAs in triplets against the number of mRNA-target interactions was plotted in Figure 4.23 and observed a relatively low correlation, with a Pearson correlation coefficient of 0.343. Interestingly, we found specific examples that demonstrated variations in the occurrence of mRNAs in triplets despite having different numbers of mRNA-target interactions. For instance, SOD2 was found in a total of 64 triplets and exhibited 354 target interactions with miRNAs. In contrast, TAOK1 had 407 target interactions but was only observed in 8 triplets.

Figure 4.23 A scatter plot was generated to display the correlation between the number of mRNAs targeted by miRNA and the number of triplets in which the mRNA is found in miRCoop triplets. The analysis resulted in a Pearson correlation coefficient of 0.3429.



4.4 Analysis of the Synergy Modules and miRNA Pairs across All Cancer Types in miRCoop TF

An analysis was performed to identify common miRCoop synergy modules and synergistic miRNA pairs in order to uncover highly interacting miRNAs, TFs, and target mRNAs. Figure 4.24 illustrates the common synergy modules observed in Type 3 cancers. Among these modules, miR-20a-5b was found to form a synergistic pair with two other miRNAs (miR-17-5p and miR-18a-5p), resulting in the formation of three distinct synergy modules observed in various cancers (STAD, LIHC, BLCA, HNSC). Notably, these synergy modules involve the participation of two different transcription factors, namely STAT3 and MEF2A. Extensive research has demonstrated the importance of these miRNAs, TFs, and target genes in various cancers.

In previous studies, it was discovered that miR-20a-5p was overexpressed in hepatocellular carcinoma (HCC) and may serve as a potential biomarker for the preclinical diagnosis of HCC. Furthermore, it was proposed that the overexpression of miR-20a-5p promotes HCC cell proliferation and migration by reducing the translation of the RUNX3 (Huang et al., 2022). Additionally, aberrant upregulation of miR-20a-5p has been linked to the inhibition of WTX expression and the activation of the PI3K/AKT/mTOR signaling pathway, elucidating the underlying mechanism of progression and metastasis in gastric cancer (GC) (Huang et al., 2022).

Furthermore, the persistent activation of STAT3 has been associated with increased tumor cell proliferation, survival, and invasion, as well as the suppression of antitumor immune responses. Additionally, the persistently activated STAT3 pathway is known to mediate tumor-promoting inflammation. Given its crucial role as a signaling node in various human cancers, targeting STAT3 is expected to provide multiple benefits (Yu et al., 2009; Zou et al., 2020).

Figure 4.24 Pan-cancer synergy modules in Type 3. Circle nodes represent miRNAs, square nodes show TF, and diamond nodes illustrate the target mRNAs. The gray nodes are dummy nodes to connect the miRNA pairs to TFs. Captions on the edges indicate the cancer types in the synergy module appears. Furthermore, a dashed line is depicted between the dummy node and the target mRNA, highlighting their connection in the network.

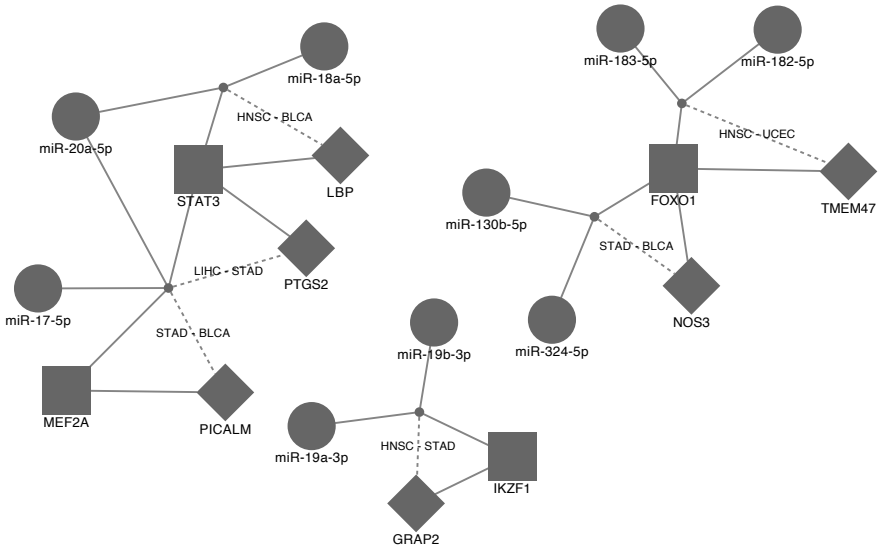
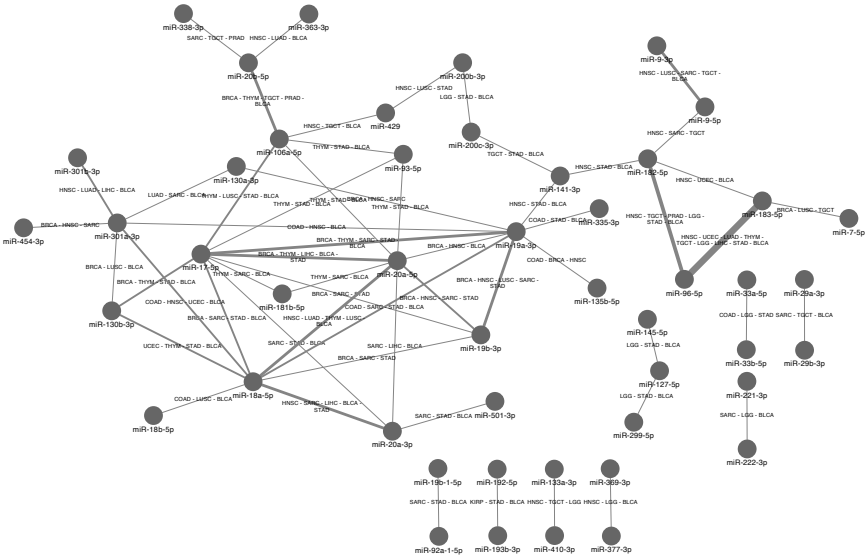


Figure 4.25 illustrates the presence of synergistic miRNA pairs that are shared across multiple cancer types in synergy modules type 3. The synergistic pair miR-96-5p and miR-183-5p was identified as the most common in multiple cancer types, occurring in 9 units. Individually, these miRNAs have been extensively studied in various types of cancer. For instance, miR-96-5p has been identified as an oncogenic miRNA with prognostic significance in head and neck squamous cell carcinoma (HNSCC) and in hepatocellular carcinoma (HCC), miR-96-5p has been found to be frequently upregulated (Vahabi et al., 2019; Iwai et al., 2018). This upregulation of miR-96-5p in HCC can increase invasiveness during tumor progression by inhibiting apoptosis. The other member of the synergistic miRNA pair, miR-183-5p, which is mostly seen in different cancer types, has been studied in previous studies. To illustrate, miR-183-5p has been found to be upregulated in non-small cell lung cancer (NSCLC). Its upregulation in NSCLC is associated with increased cell proliferation, migration, and altered cell cycle regulation. In vivo, studies have provided evidence that miR-183-5p contributes to tumor growth and metastasis (Wang et al., 2019). Also, miR-183-5p has been identified as an oncogene in LUAD, contributing to the pathogenesis of the disease through its involvement in the interaction networks of its target genes (He et al., 2018).

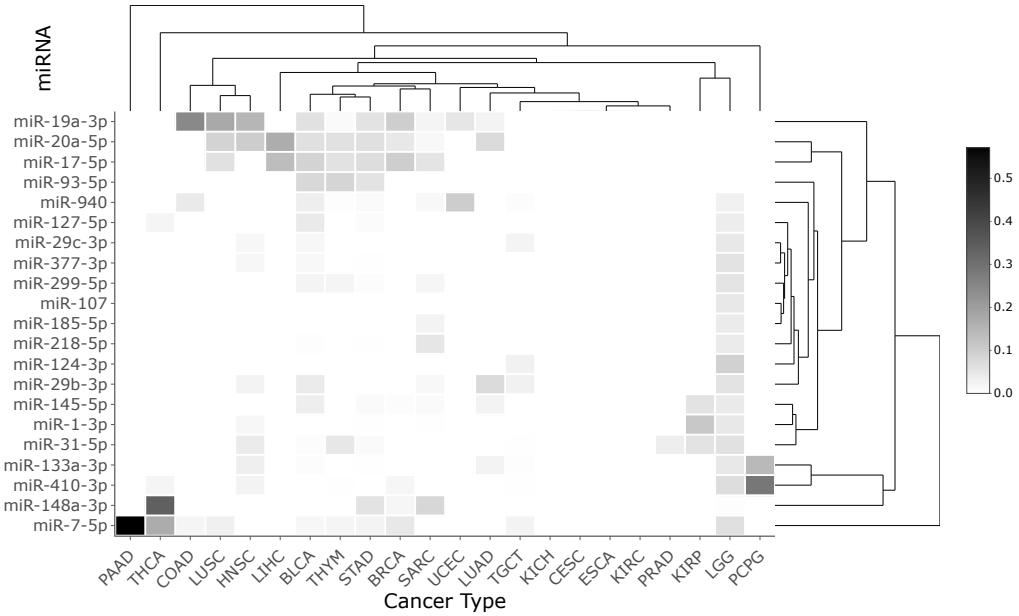
Figure 4.25 The miRNA pairs that are shared across three or more cancers are depicted in the network Synergy Modules Type 3. In the visualization, the nodes represent the miRNAs, while the captions on the edges indicate the cancer types in which the synergistic pairs appear. The thickness of the edges expands according to the number of cancers in which the pair is present.



To gain insights into the distribution patterns of miRNAs, miRNA pairs, TFs, and target mRNAs that exhibited a higher overall frequency within miRCoop triplets, we performed an analysis to assess their association with specific cancer types or their presence across multiple cancers. This analysis aimed to determine whether these molecules displayed a predominant association with particular cancer types or demonstrated a broader distribution across multiple cancer types.

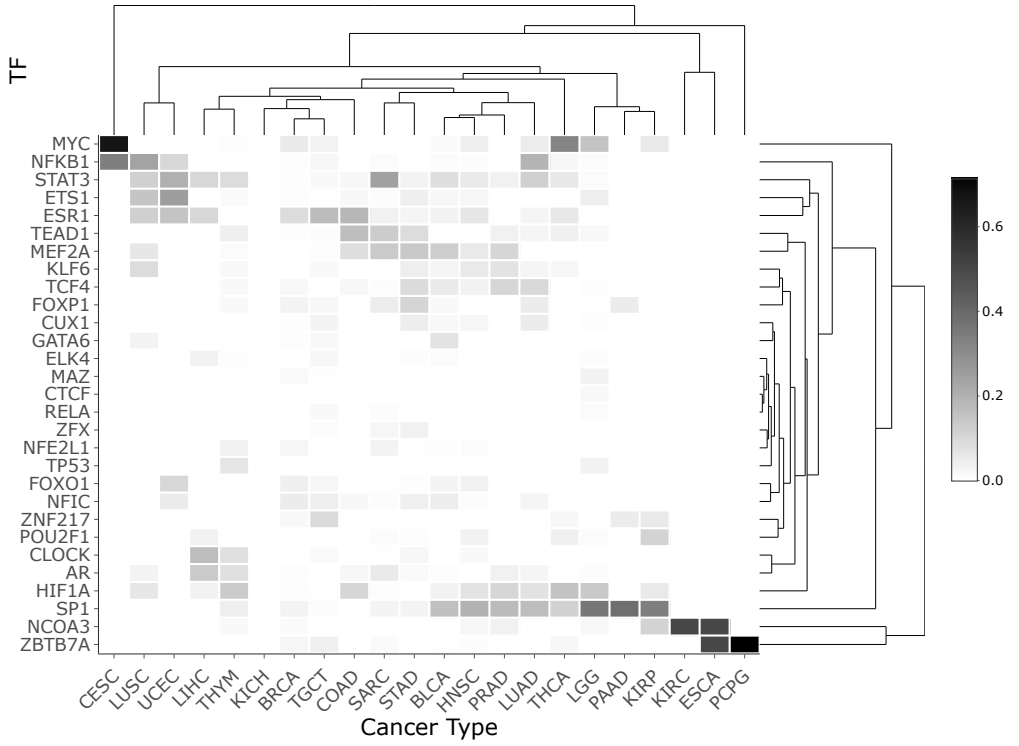
Previous studies have shown that miR-7-5p functions as a tumor suppressor in pancreatic ductal adenocarcinoma and its low expression level is associated with a poor survival rate(Zhu et al., 2018). Interestingly, in miRCoop synergy modules type 3, we found that miR-7-5p was one of the most common miRNAs identified in the PAAD (Figure 4.26). Another miRNA common in COAD in our analysis was miR-19a-3p. It is regarded as a potential therapeutic target due to its regulation of the Forkhead box F2-mediated Wnt/ß-catenin signaling pathway and its impact on the biological functions of colorectal cancer cells (Yu et al., 2020). Moreover, miR-7-5p known for its tumor-suppressive role in thyroid cancer, was also identified as one of the most common miRNAs in THCA(Liu et al., 2019).

Figure 4.26 The normalized number of miRNA in miRCoop synergy modules type 3 for various cancer types is provided. These numbers are normalized based on the total number of synergy modules within each specific cancer. miRNAs which have more than 120 participation in synergy modules in total are shown. To identify relationships and patterns, clustering was performed using the Euclidean distance and complete linkage methods.



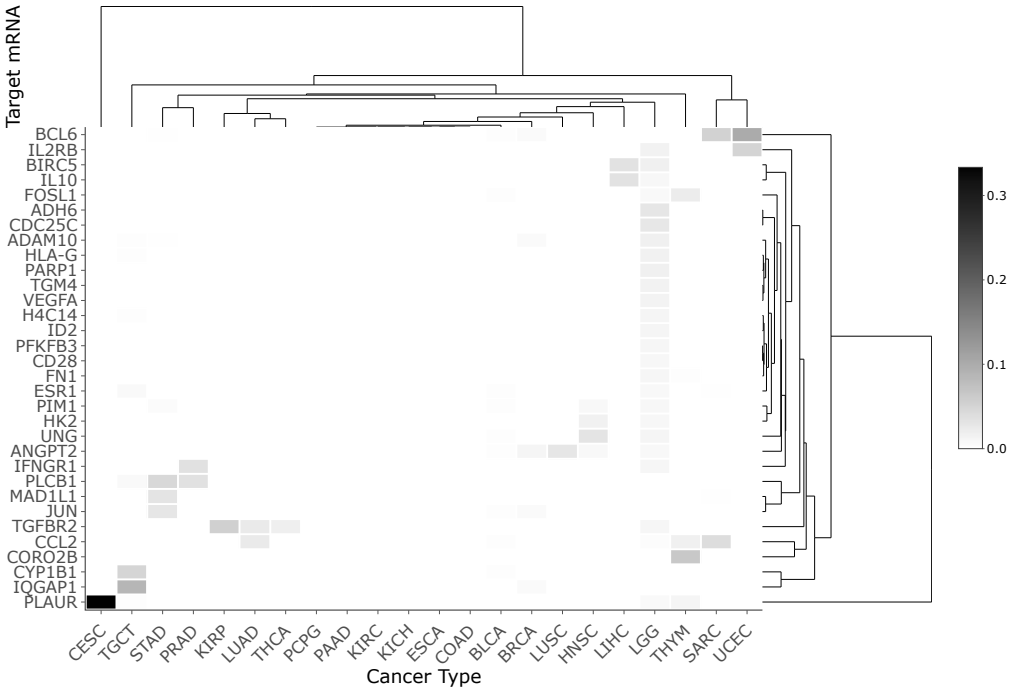
One of the most common TFs identified in COAD within miRCoop synergy modules type 3 was MYC (Figure 4.27). Previous studies have shown a strong correlation between the expression of c-Myc, a member of the MYC family, and cervical cancer tissues (Liao et al., 2014). Additionally, MYC was also commonly found in THCA within miRCoop synergy modules type 3. Upregulation of MYC contributes to the initiation of undifferentiated thyroid cancer by enhancing TRßPV-mediated repression of Pax8 expression, which serves as a potential therapeutic target (Zhu et al., 2014).

Figure 4.27 The dataset includes the normalized number of TFs in different cancer types for miRCoop synergy modules type 3, where the normalization is based on the total number of synergy modules within each specific cancer. Only TFs with over 40 participations in synergy modules across all cancers are displayed. To uncover relationships and patterns within the dataset, clustering analysis was conducted using the Euclidean distance and complete linkage methods.



CYP1B1, identified as one of the common target mRNAs in TGCT within miRCoop synergy modules type 3, exhibits high expression across a wide range of human cancers of different histogenetic types. These include breast, colon, lung, esophagus, skin, lymph node, brain, and testicular cancers (Figure 4.28) (Murray et al., 1997). Additionally, previous studies have indicated that TGFBR2 functions as a tumor suppressor, making it of prognostic value for non-small cell lung cancer (NSCLC). Interestingly, TGFBR2 is one of the common target mRNAs identified in LUAD within miRCoop synergy modules (Sardo et al., 2021).

Figure 4.28 The dataset comprises the normalized number of target mRNAs in miRCoop synergy modules type 3 for various cancer types. The normalization process is based on the total number of synergy modules within each specific cancer. Only mRNAs with more than 25 participations in synergy modules across all cancers are included in the dataset. To identify relationships and patterns within the dataset, clustering analysis was performed using the Euclidean distance and complete linkage methods.



Additionally, an analysis was conducted in Type 1 to identify common miRCoop synergy modules and synergistic miRNA pairs. The goal of this analysis was to uncover highly interacting miRNAs, TFs, and target mRNAs. In the TF synergy modules Type 1, a total of 21 pan-cancer synergy modules were identified, and their network is illustrated in Figure 4.29. Notably, miR-18a-5p exhibited interactions with multiple transcription factors (TFs) and formed synergistic pairs with different miRNAs, contributing to the formation of diverse synergy modules. It is worth mentioning that miR-18a-5p is a member of the miR-17-92a cluster, which plays a significant role in essential cellular processes such as cell cycle regulation, proliferation, and apoptosis. The dysregulation of the miR-17/92 cluster is frequently observed in both hematopoietic and solid cancers (Mogilyansky and Rigoutsos, 2013). Moreover, STAT3, along with two other different miRNAs, interacts with miR-18a-5p to form various synergy modules in the TF synergy modules. Extensive research has indicated that signal transducers and activators of transcription (STAT) proteins,

particularly STAT3, hold immense potential as targets for cancer therapy. Among the seven members of the STAT protein family, STAT3 has been identified as a crucial player in cancer progression. The oncogenic functions of STAT3, including its involvement in mitochondrial processes, epigenetic regulation, cancer stem cells (CSCs), obesity-related mechanisms, and the formation of pre-metastatic niches, underscore the significance of targeting STAT3 in cancer treatment (Yu et al., 2014).

Figure 4.29 Pan-cancer synergy modules in Type 1. The network is represented with circle nodes for miRNAs, square nodes for TFs, and diamond nodes for target mRNAs. Gray nodes are dummy nodes that connect the miRNA pairs to the target mRNAs. The captions on the edges indicate the cancer types in which the synergy module appears. Additionally, a dashed line is depicted between the dummy node and the target mRNA, emphasizing their connection within the network.

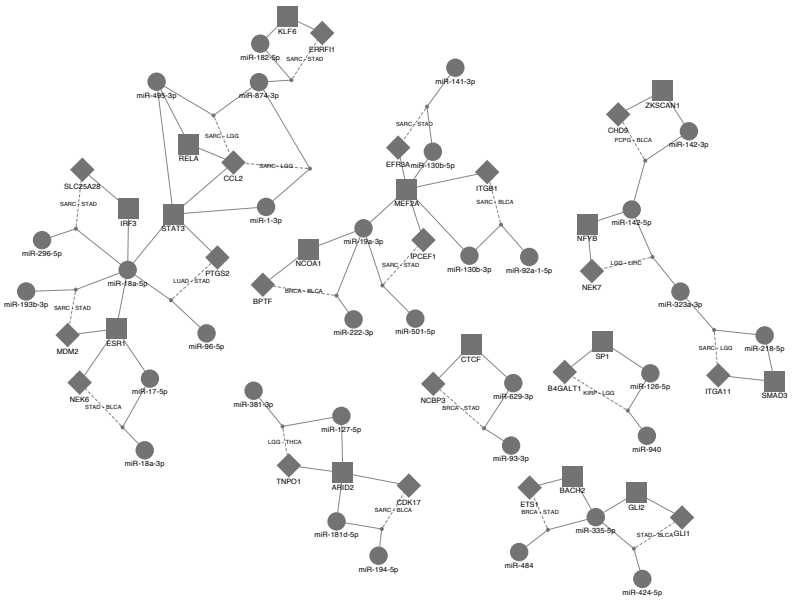
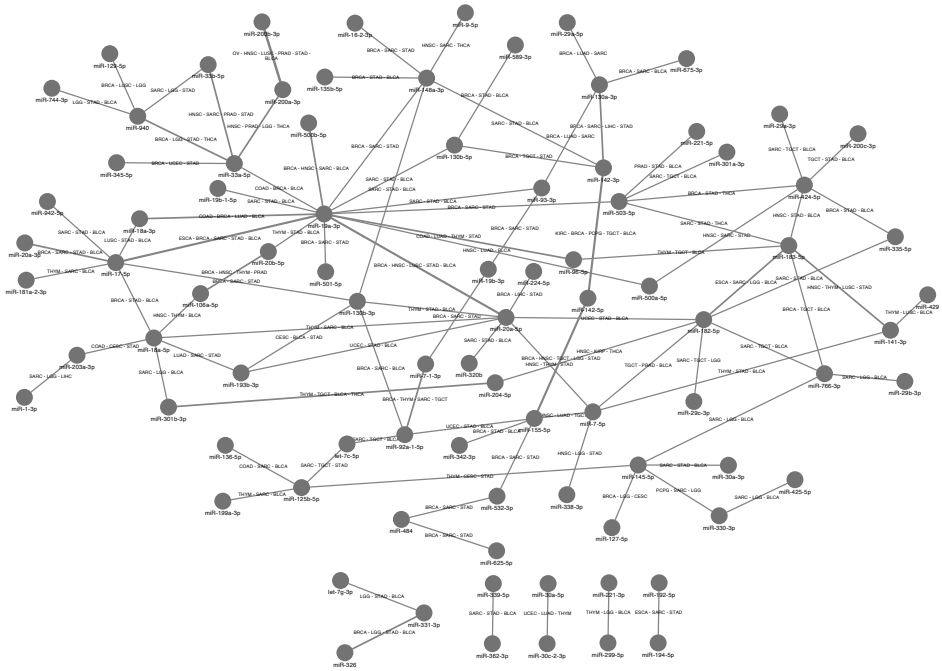


Figure 4.30 presents synergistic pan-cancer miRNA pairs that are commonly found in three or more cancer types. Among these miRNA pairs, miR-19a-3p stands out as it forms synergistic pairs with different miRNAs more frequently compared to other miRNAs. Notably, the synergistic pair of miR-19a-3p and miR-20a-5p has been observed in five different cancer types. Both miR-19a-3p and miR-20a-5p are members of the miR-17/92 cluster, which has been extensively studied for its significant impact on cancer (Mogilyansky and Rigoutsos, 2013).

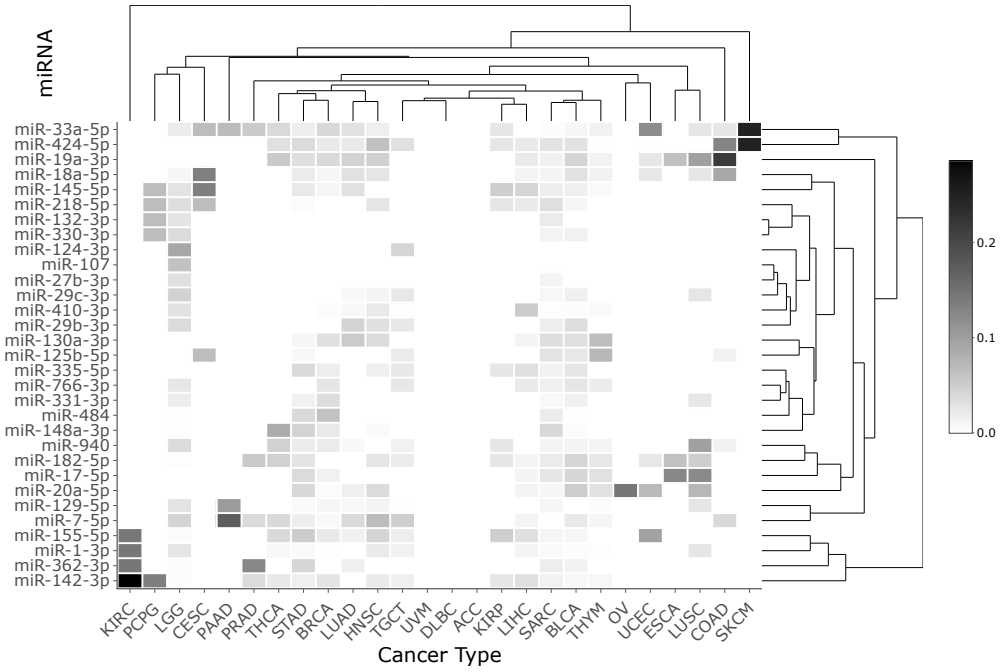
Figure 4.30 miRNA pairs that are commonly found in three or more cancers in miRCoop synergy modules type 1. Each node represents a specific miRNA, while the captions on the edges indicate the cancer types in which the synergistic pairs occur. The thickness of the edges expands based on the number of cancers in which the miRNA pair is observed.



As illustrated in Figure 4.31 miR-424-5p was identified as one of the frequently observed miRNAs in synergy modules type 1 in COAD. Consistently, this miRNA was also found to be commonly present in COAD triplets in miRCoop v2 (Figure 4.19). Previous studies have implicated miR-424-5p in various aspects of COAD, including cell proliferation, migration, and metastasis (Dai et al., 2020; Cheng et al., 2020). Another frequent miRNA in COAD within Type 1 synergy modules was miR-19a-3p. Similarly, it is commonly observed in COAD within Type 3 synergy modules. Interestingly, miR-19a-3p has been found to regulate the Wnt/ß-catenin signaling pathway mediated by Forkhead box F2. This finding suggests that miR-19a-3p may hold potential as a therapeutic target in colorectal cancer (CRC) (Yu et al., 2020). Additionally, miR-20a-5p was found to have a higher frequency in multiple cancer types such as LIHC, LUSC, and HNSC in miRCoop v2 (Figure 4.19). Similarly, in synergy modules type 3 4.26, it was commonly observed in various cancer types including LIHC, BRCA, and BLCA. However, in synergy modules type 1, miR-20a-5p exhibited a higher frequency specifically in OV. Previous studies have highlighted the effects of miR-20a-5p in different cancer types, including breast cancer, liver

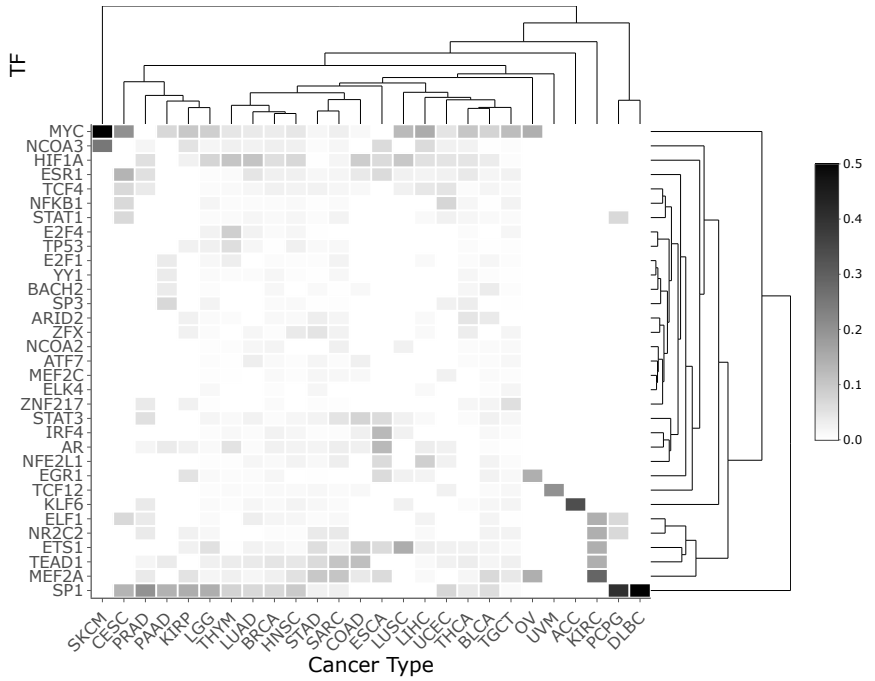
cancer, and leukemia. In the context of ovarian carcinoma, miR-20a-5p has been reported to be overexpressed and can serve as a diagnostic hallmark (Huang et al., 2022).

Figure 4.31 The normalized number of miRNAs in miRCoop synergy modules type 1 for various cancer types is presented. These numbers are normalized based on the total number of synergy modules within each specific cancer. Only miRNAs with more than 150 participations in synergy modules in total are shown. To identify relationships and patterns, clustering was performed using the Euclidean distance and complete linkage methods.



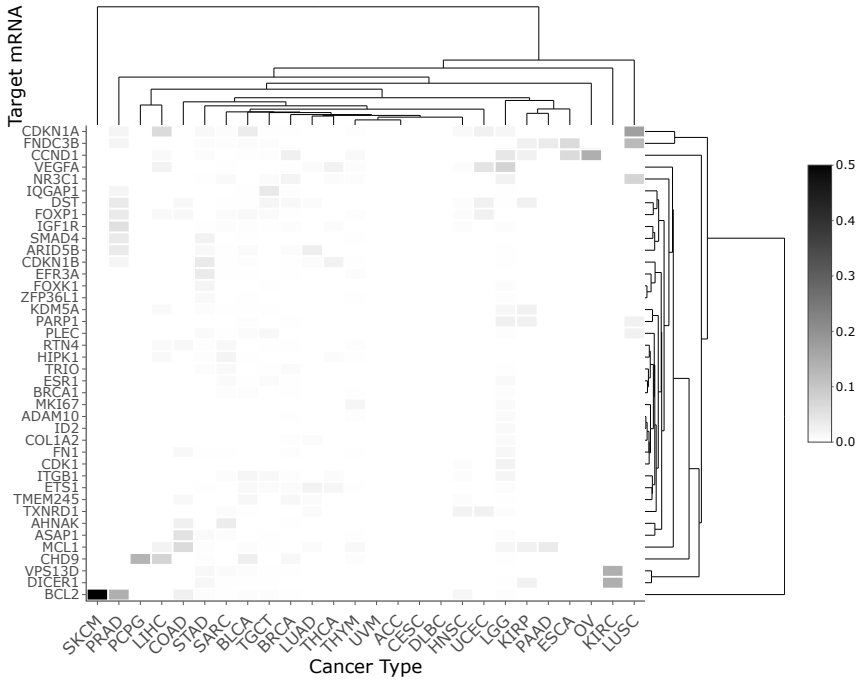
As depicted in Figure 4.32, NCOA3 stands out as one of the frequently observed TFs in various cancers, including SKCM. Previous studies have identified NCOA3 as a differentially expressed gene in metastatic melanomas and have validated its significance as a prognostic marker in primary cutaneous melanoma (Bezrookove and Kashani-Sabet, 2022). Moreover, in synergy modules type 3, MYC was frequently observed in CESC and THYM, while in synergy modules type 1, it was frequently observed in SKCM. On the other hand, SP1 was frequently observed in various cancers in synergy modules type 3. In addition to these, it was also most frequently observed in PCPG and DLBC in synergy modules type 1.

Figure 4.32 The heatmap contains the normalized counts of TFs in various cancer types for miRCoop synergy modules type 1. The normalization is based on the total number of synergy modules within each specific cancer. Only TFs with more than 100 participation in synergy modules across all cancers are included. Clustering analysis was performed using the Euclidean distance and complete linkage methods.



As depicted in Figure 4.33, CDKN1A is observed as one of the target mRNAs with a high frequency in LUSC. Remarkably, previous studies have indicated that CDKN1A functions as an oncogene, playing a role in promoting cancer cell proliferation by inhibiting apoptosis specifically in NSCLC (Zamagni et al., 2020).

Figure 4.33 The heatmap presents the normalized counts of target mRNAs in miRCoop synergy modules type 1 across different cancer types. The normalization is performed based on the total number of synergy modules within each specific cancer. Only mRNAs with more than 40 participation in synergy modules across all cancers are displayed. Clustering analysis was conducted using the Euclidean distance and complete linkage methods.



miRCoop-V2 can be a valuable tool for detecting cancer-related miRNA/TF/target mRNA interactions. Even in cases where there may not be direct literature evidence for specific cancer, the significance of these miRNAs and mRNAs in other cancer types suggests their potential importance and therapeutic potential across multiple cancer contexts. The ability to identify novel therapeutic potentials opens up new avenues for further research and development of targeted therapies for cancer treatment.

5. CONCLUSION & FUTURE WORK

MicroRNAs (miRNAs) play a vital role in cancer due to their role in the regulation of mRNAs and can act as both tumor suppressors and oncogenes. Thus, understanding their regulation is critical to dissect complex cancer biology and developing therapeutics. The development of miRNA-based therapeutics encounters challenges such as off-target effects, limited effectiveness of single miRNAs, and potential toxicity. To overcome these obstacles, the use of synergistic miRNAs has emerged as a promising solution since it was known that RNAs can act cooperatively. Cooperative miRNAs have been demonstrated to exert more potent repression of target mRNA, highlighting their collaborative and enhanced regulatory effect. While existing computational methods for identifying synergistic miRNAs often focus on specific cancer types or a limited number of them, this study aimed to refine the miRCoop algorithm. These refinements include identifying synergistic miRNA interactions involving transcription factors, explicitly incorporating miRNA arm information, utilizing a larger miRNA-mRNA target interaction dataset, and 30 times faster runtime.

The algorithm was applied to TCGA cancer patient data within the study's scope. The resulting findings were presented to users through a dedicated web application, providing a convenient platform for comprehensive data analysis. However, the developed algorithm holds the potential for broader application, extending beyond patient data to include normal tissue samples. This broader application enables a comparative analysis of RNA mechanisms between normal tissue samples and tumor samples, thereby facilitating a deeper understanding of the distinctions between these two groups.

Currently, miRCoop v2 focuses on synergistic miRNA interactions specifically for humans, as it relies on the availability of abundant matched RNA-seq and miRNA-seq data. However, it is possible to expand the database to include other species as matched RNA-seq and miRNA-seq data for those species become accessible. This could enhance the applicability and utility of miRCoop v2 for other research endeavors. The miRCoop web application currently provides users with an excellent

platform to review pre-computed results. The application can be further improved by incorporating a feature that enables users to run the algorithm using their own data. This addition could further empower researchers and users by allowing them to apply the algorithm to their specific datasets, fostering a more personalized and interactive research experience.

BIBLIOGRAPHY

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting effective microrna target sites in mammalian mrnas. *elife*, 4:e05005.

Anand, U., Dey, A., Chandel, A. K. S., Sanyal, R., Mishra, A., Pandey, D. K., De Falco, V., Upadhyay, A., Kandimalla, R., Chaudhary, A., et al. (2022). Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases*.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microrna on protein output. *Nature*, 455(7209):64–71.

Bandi, N. and Vassella, E. (2011). mir-34a and mir-15a/16 are co-regulated in non-small cell lung cancer and control cell cycle progression in a synergistic and rb-dependent manner. *Molecular cancer*, 10:1–11.

Bartel, D. P. (2004). Microrna: genomics, biogenesis, mechanism, and function. *cell*, 116(2):281–297.

Battistella, M., Romero, M., Castro-Vega, L.-J., Gapihan, G., Bouhidel, F., Bagot, M., Feugeas, J.-P., and Janin, A. (2015). The high expression of the microrna 17–92 cluster and its paralogs, and the downregulation of the target gene pten, is associated with primary cutaneous b-cell lymphoma progression. *Journal of Investigative Dermatology*, 135(6):1659–1667.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microrna targets predicts functional non-conserved and non-canonical sites. *Genome biology*, 11:1–14.

Bezrookove, V. and Kashani-Sabet, M. (2022). Ncoa3, a new player in melanoma susceptibility and a therapeutic target. *Cancer gene therapy*, 29(5):399–401.

Borzi, C., Calzolari, L., Centonze, G., Milione, M., Sozzi, G., and Fortunato, O. (2017). mir-660-p53-mir-486 network: a new key regulatory pathway in lung tumorigenesis. *International journal of molecular sciences*, 18(1):222.

Bracken, C. P., Scott, H. S., and Goodall, G. J. (2016). A network-biology perspective of microrna function and dysfunction in cancer. *Nature Reviews Genetics*, 17(12):719–732.

Bueno, M. J. and Malumbres, M. (2011). Microrna and the cell cycle. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1812(5):592–601.

Cantini, L., Bertoli, G., Cava, C., Dubois, T., Zinovyev, A., Caselle, M., Castiglioni, I., Barillot, E., and Martignetti, L. (2019). Identification of microrna clusters cooperatively acting on epithelial to mesenchymal transition in triple negative breast cancer. *Nucleic acids research*, 47(5):2205–2215.

Cheng, C., Li, H., Zheng, J., Xu, J., Gao, P., and Wang, J. (2020). Fendrr sponges mir-424-5p to inhibit cell proliferation, migration and invasion in colorectal cancer. *Technology in Cancer Research & Treatment*, 19:1533033820980102.

Dahlrot, R. H., Bangsø, J. A., Petersen, J. K., Rosager, A. M., Sørensen, M. D., Reifenberger, G., Hansen, S., and Kristensen, B. W. (2021). Prognostic role of ki-67 in glioblastomas excluding contribution from non-neoplastic cells. *Scientific reports*, 11(1):1–9.

Dai, W., Zhou, J., Wang, H., Zhang, M., Yang, X., and Song, W. (2020). mir-424-5p promotes the proliferation and metastasis of colorectal cancer by directly targeting scn4b. *Pathology-Research and Practice*, 216(1):152731.

Ding, J., Li, X., and Hu, H. (2015). Microrna modules prefer to bind weak and unconventional target sites. *Bioinformatics*, 31(9):1366–1374.

Ding, J., Li, X., and Hu, H. (2018). Ccmir: a computational approach for competitive and cooperative microrna binding prediction. *Bioinformatics*, 34(2):198–206.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184–1191.

Enright, A., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. (2003). Microrna targets in drosophila. *Genome biology*, 4:1–27.

Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome research*, 29(8):1363–1375.

Garzon, R., Calin, G. A., and Croce, C. M. (2009). Micrornas in cancer. *Annual review of medicine*, 60:167–179.

Gennarino, V. A., d'Angelo, G., Dharmalingam, G., Fernandez, S., Russolillo, G., Sanges, R., Mutarelli, M., Belcastro, V., Ballabio, A., Verde, P., et al. (2012). Identification of microrna-regulated gene networks by expression analysis of target genes. *Genome research*, 22(6):1163–1172.

Hamed, M., Spaniol, C., Nazarieh, M., and Helms, V. (2015). Tfmir: a web server for constructing and analyzing disease-specific transcription factor and mirna co-regulatory networks. *Nucleic acids research*, 43(W1):W283–W288.

He, R.-Q., Gao, L., Ma, J., Li, Z.-Y., Hu, X.-H., and Chen, G. (2018). Oncogenic role of mir-183-5p in lung adenocarcinoma: A comprehensive study of qpcr, in vitro experiments and bioinformatic analysis. *Oncology Reports*, 40(1):83–100.

Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., Tang, Y., Chen, Y.-G., Jin, C.-N., Yu, Y., et al. (2020). mirtarbase 2020: updates to the experimentally validated microrna–target interaction database. *Nucleic acids research*, 48(D1):D148–D154.

Huang, W., Wu, X., Xiang, S., Qiao, M., Cen, X., Pan, X., Huang, X., and Zhao, Z. (2022). Regulatory mechanism of mir-20a-5p expression in cancer. *Cell Death Discovery*, 8(1):262.

İlhan, A., Golestani, S., Shafagh, S. G., Asadi, F., Daneshdoust, D., Al-Naqeeb, B. Z. T., Nemati, M. M., Khalatbari, F., and Yaseri, A. F. (2023). The dual role of microrna (mir)-20b in cancers: Friend or foe? *Cell Communication and Signaling*, 21(1):1–13.

Iwai, N., Yasui, K., Tomie, A., Gen, Y., Terasaki, K., Kitaichi, T., Soda, T., Yamada, N., Dohi, O., Seko, Y., et al. (2018). Oncogenic mir-96-5p inhibits apoptosis by targeting the caspase-9 gene in hepatocellular carcinoma. *International journal of oncology*, 53(1):237–245.

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl biomarts: a hub for data retrieval across taxonomic space. *Database*, 2011.

Lai, X., Eberhardt, M., Schmitz, U., and Vera, J. (2019). Systems biology-based investigation of cooperating micrornas as monotherapy or adjuvant therapy in cancer. *Nucleic acids research*, 47(15):7753–7766.

Lai, X., Gupta, S. K., Schmitz, U., Marquardt, S., Knoll, S., Spitschak, A., Wolkenhauer, O., Pützer, B. M., and Vera, J. (2018). Mir-205-5p and mir-342-3p cooperate in the repression of the e2f1 transcription factor in the context of anticancer chemotherapy resistance. *Theranostics*, 8(4):1106.

Lai, X., Schmitz, U., Gupta, S. K., Bhattacharya, A., Kunz, M., Wolkenhauer, O., and Vera, J. (2012). Computational analysis of target hub gene repression regulated by multiple and cooperative mirnas. *Nucleic acids research*, 40(18):8818–8834.

Li, M., Zhou, Y., Xia, T., Zhou, X., Huang, Z., Zhang, H., Zhu, W., Ding, Q., and Wang, S. (2018a). Circulating micrornas from the mir-106a–363 cluster on chromosome x as novel diagnostic biomarkers for breast cancer. *Breast cancer research and treatment*, 170:257–270.

Li, R., Chen, H., Jiang, S., Li, W., Li, H., Zhang, Z., Hong, H., Huang, X., Zhao, C., Lu, Y., et al. (2018b). Cmtcn: a web tool for investigating cancer-specific microrna and transcription factor co-regulatory networks. *PeerJ*, 6:e5951.

Li, Y., Liang, C., Wong, K.-C., Luo, J., and Zhang, Z. (2014). Mirsynergy: detecting synergistic mirna regulatory modules by overlapping neighbourhood expansion. *Bioinformatics*, 30(18):2627–2635.

Liao, L.-M., Sun, X.-Y., Liu, A.-W., Wu, J.-B., Cheng, X.-L., Lin, J.-X., Zheng, M., and Huang, L. (2014). Low expression of long noncoding xloc_010588 indicates a poor prognosis and promotes proliferation through upregulation of c-myc in cervical cancer. *Gynecologic oncology*, 133(3):616–623.

Liu, X., Fu, Q., Li, S., Liang, N., Li, F., Li, C., Sui, C., Dionigi, G., and Sun, H. (2019). Lncrna foxd2-as1 functions as a competing endogenous rna to regulate tert expression by sponging mir-7-5p in thyroid cancer. *Frontiers in endocrinology*, 10:207.

Luo, B., Kang, N., Chen, Y., Liu, L., and Zhang, Y. (2017). Oncogene mir-106a promotes proliferation and metastasis of prostate cancer cells by directly targeting pten in vivo and in vitro. *Minerva medica*, 109(1):24–30.

Malcolm, J., Yalamanchili, P., McClanahan, C., Venugopalakrishnan, V., Patel, K., and Melonakos, J. (2012). Arrayfire: a gpu acceleration platform. In *Modeling and simulation for defense systems and applications VII*, volume 8403, pages 49–56. SPIE.

Meng, X., Wang, J., Yuan, C., Li, X., Zhou, Y., Hofestädt, R., and Chen, M. (2015). Cancernet: a database for decoding multilevel molecular interactions across diverse cancer types. *Oncogenesis*, 4(12):e177–e177.

Mogilyansky, E. and Rigoutsos, I. (2013). The mir-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death & Differentiation*, 20(12):1603–1614.

Murray, G. I., Taylor, M. C., McFadyen, M. C., McKay, J. A., Greenlee, W. F., Burke, M. D., and Melvin, W. T. (1997). Tumor-specific expression of cytochrome p450 cyp1b1. *Cancer research*, 57(14):3026–3031.

Olgun, G. and Tastan, O. (2020). mircoop: Identifying cooperating mirnas via kernel based interaction tests. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1760–1771.

Olive, V., Jiang, I., and He, L. (2010). mir-17-92, a cluster of mirnas in the midst of the cancer network. *The international journal of biochemistry & cell biology*, 42(8):1348–1354.

Peng, Y. and Croce, C. M. (2016). The role of micrornas in human cancer. *Signal transduction and targeted therapy*, 1(1):1–9.

Piñero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., and Furlong, L. I. (2015). Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, 2015.

Qi, J.-C., Yang, Z., Zhang, Y.-P., Lu, B.-S., Yin, Y.-W., Liu, K.-L., Xue, W.-Y., Qu, C.-B., and Li, W. (2020). mir-20b-5p, tgfbr2, and e2f1 form a regulatory loop to participate in epithelial to mesenchymal transition in prostate cancer. *Frontiers in oncology*, 9:1535.

Redler, A., Di Rocco, G., Giannotti, D., Frezzotti, F., Bernieri, M. G., Ceccarelli, S., D'Amici, S., Vescarelli, E., Mitterhofer, A. P., Angeloni, A., et al. (2013). Fibroblast growth factor receptor-2 expression in thyroid tumor progression: potential diagnostic application. *PLoS One*, 8(8):e72224.

Salehi, B., Zucca, P., Sharifi-Rad, M., Pezzani, R., Rajabi, S., Setzer, W. N., Varoni, E. M., Iriti, M., Kobarfard, F., and Sharifi-Rad, J. (2018). Phytotherapeutics in cancer invasion and metastasis. *Phytotherapy Research*, 32(8):1425–1449.

Sardo, F. L., Pulito, C., Sacconi, A., Korita, E., Sudol, M., Strano, S., and Blandino, G. (2021). Yap/taz and ezh2 synergize to impair tumor suppressor activity of tgfbr2 in non-small cell lung cancer. *Cancer Letters*, 500:51–63.

Schmitz, U., Lai, X., Winter, F., Wolkenhauer, O., Vera, J., and Gupta, S. K. (2014). Cooperative gene regulation by microrna pairs and their identification using a computational workflow. *Nucleic acids research*, 42(12):7539–7552.

Sejdinovic, D., Gretton, A., and Bergsma, W. (2013). A kernel test for three-variable interactions. *Advances in Neural Information Processing Systems*, 26.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microrna. *nature*, 455(7209):58–63.

Smola, A., Gretton, A., Song, L., and Schölkopf, B. (2007). A hilbert space embedding for distributions. In *International conference on algorithmic learning theory*, pages 13–31. Springer.

Song, L., Fukumizu, K., and Gretton, A. (2013). Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Processing Magazine*, 30(4):98–111.

Vahabi, M., Pulito, C., Sacconi, A., Donzelli, S., D'Andrea, M., Manciocco, V., Pellini, R., Paci, P., Sanguineti, G., Strigari, L., et al. (2019). mir-96-5p targets pten expression affecting radio-chemosensitivity of hnscc cells. *Journal of Experimental & Clinical Cancer Research*, 38(1):1–16.

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., Newman, J., Bronson, R. T., Crowley, D., Stone, J. R., et al. (2008). Targeted deletion reveals essential and overlapping functions of the mir-17 92 family of mirna clusters. *Cell*, 132(5):875–886.

Vlachos, I. S., Paraskevopoulou, M. D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.-L., Maniou, S., Karathanou, K., Kalfakakou, D., et al. (2015). Diana-tarbase v7. 0: indexing more than half a million experimentally supported mirna: mrna interactions. *Nucleic acids research*, 43(D1):D153–D159.

Wang, H., Ma, Z., Liu, X., Zhang, C., Hu, Y., Ding, L., Qi, P., Wang, J., Lu, S., and Li, Y. (2019). Mir-183-5p is required for non-small cell lung cancer progression by repressing pten. *Biomedicine & Pharmacotherapy*, 111:1103–1111.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. biom. bull., 1, 80–83.

Wu, S., Huang, S., Ding, J., Zhao, Y., Liang, L., Liu, T., Zhan, R., and He, X. (2010). Multiple micrornas modulate p21cip1/waf1 expression by directly targeting its 3 untranslated region. *Oncogene*, 29(15):2302–2308.

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). mirecords: an integrated resource for microrna–target interactions. *Nucleic acids research*, 37(suppl_1):D105–D110.

Xie, B., Ding, Q., Han, H., and Wu, D. (2013). mircancer: a microrna–cancer association database constructed by text mining on literature. *Bioinformatics*, 29(5):638–644.

Xie, G.-Y., Xia, M., Miao, Y.-R., Luo, M., Zhang, Q., and Guo, A.-Y. (2020). Ffltool: a web server for transcription factor and mirna feed forward loop analysis in human. *Bioinformatics*, 36(8):2605–2607.

Xu, J., Li, C.-X., Li, Y.-S., Lv, J.-Y., Ma, Y., Shao, T.-T., Xu, L.-D., Wang, Y.-Y., Du, L., Zhang, Y.-P., et al. (2011). Mirna–mirna synergistic network: construction via co-regulating functional modules and disease mirna topological features. *Nucleic acids research*, 39(3):825–836.

Xu, J., Shao, T., Ding, N., Li, Y., and Li, X. (2017). mirna–mirna crosstalk: from genomics to phenomics. *Briefings in bioinformatics*, 18(6):1002–1011.

Xu, T., Su, N., Liu, L., Zhang, J., Wang, H., Zhang, W., Gui, J., Yu, K., Li, J., and Le, T. D. (2018). mirbaseconverter: an r/bioconductor package for converting and retrieving mirna name, accession, sequence and family information in different versions of mirbase. *BMC bioinformatics*, 19(19):179–188.

Xuan, J., Liu, Y., Zeng, X., and Wang, H. (2022). Sequence requirements for mir-424-5p regulating and function in cancers. *International Journal of Molecular Sciences*, 23(7):4037.

Yang, X., Du, W. W., Li, H., Liu, F., Khorshidi, A., Rutnam, Z. J., and Yang, B. B. (2013). Both mature mir-17-5p and passenger strand mir-17-3p target timp3 and induce prostate tumor growth and invasion. *Nucleic acids research*, 41(21):9688–9704.

Yang, Y., Xing, Y., Liang, C., Hu, L., Xu, F., and Chen, Y. (2015). Crucial micrornas and genes of human primary breast cancer explored by microrna-mrna integrated analysis. *Tumor Biology*, 36:5571–5579.

Yu, F.-B., Sheng, J., Yu, J.-M., Liu, J.-H., Qin, X.-X., and Mou, B. (2020). Mir-19a-3p regulates the forkhead box f2-mediated wnt/β-catenin signaling pathway and affects the biological functions of colorectal cancer cells. *World Journal of Gastroenterology*, 26(6):627.

Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting stat3 signalling in cancer: new and unexpected biological functions. *Nature Reviews Cancer*, 14(11):736–746.

Yu, H., Pardoll, D., and Jove, R. (2009). Stats in cancer inflammation and immunity: a leading role for stat3. *Nature reviews cancer*, 9(11):798–809.

Zamagni, A., Pasini, A., Pirini, F., Ravaioli, S., Giordano, E., Tesei, A., Calistri, D., Ulivi, P., Fabbri, F., Foca, F., et al. (2020). Cdkn1a upregulation and cisplatin-pemetrexed resistance in non-small cell lung cancer cells. *International Journal of Oncology*, 56(6):1574–1584.

Zhang, J., Le, T. D., Liu, L., He, J., and Li, J. (2016). Identifying mirna synergistic regulatory networks in heterogeneous human data via network motifs. *Molecular bioSystems*, 12(2):454–463.

Zhang, T.-F., Cheng, K.-W., Shi, W.-Y., Zhang, J.-T., Liu, K.-D., Xu, S.-G., and Chen, J.-Q. (2012). Mirna synergistic network construction and enrichment analysis for common target genes in small-cell lung cancer. *Asian Pacific Journal of Cancer Prevention*, 13(12):6375–6378.

Zhao, X., Song, H., Zuo, Z., Zhu, Y., Dong, X., and Lu, X. (2013). Identification of mirna–mirna synergistic relationships in colorectal cancer. *International journal of biological macromolecules*, 55:98–103.

Zhao, Y., Cui, X., Zhu, W., Chen, X., Shen, C., Liu, Z., Yang, G., Liu, Y., and Zhao, S. (2017). Synergistic regulatory effects of microrna s on brain glioma cells. *Molecular Medicine Reports*, 16(2):1409–1416.

Zhu, W., Wang, Y., Zhang, D., Yu, X., and Leng, X. (2018). Mir-7-5p functions as a tumor suppressor by targeting sox18 in pancreatic ductal adenocarcinoma. *Biochemical and biophysical research communications*, 497(4):963–970.

Zhu, X., Zhao, L., Park, J. W., Willingham, M. C., and Cheng, S.-y. (2014). Synergistic signaling of kras and thyroid hormone receptor β mutants promotes undifferentiated thyroid cancer through myc up-regulation. *Neoplasia*, 16(9):757–769.

Zou, S., Tong, Q., Liu, B., Huang, W., Tian, Y., and Fu, X. (2020). Targeting stat3 in cancer immunotherapy. *Molecular cancer*, 19(1):1–19.

APPENDIX A

Supplementary Tables

Table A.1 The table shows the number o miRNA, the number of mRNA, and the number of tumor samples in the dataset.

	miRNA	mRNA	Tumor Samples
ACC	281	15,441	79
BLCA	349	15,887	402
BRCA	280	16,271	736
CESC	263	16,033	296
CHOL	214	15,819	36
COAD	210	15,763	218
DLBC	235	15,539	46
ESCA	250	16,841	152
HNSC	254	16,115	454
KICH	161	15,930	65
KIRC	161	16,362	250
KIRP	227	16,000	287
LGG	338	16,514	496
LIHC	294	15,364	365
LUAD	277	16,324	442
LUSC	239	16,507	337
MESO	237	16,124	81
OV	262	16,525	326
PAAD	122	16,623	177
PCPG	222	15,955	175
PRAD	85	16,418	477
READ	182	15,826	72
SARC	352	15,687	256
SKCM	351	15,726	97
STAD	287	16,716	333
TGCT	402	16,853	149
THCA	180	16,147	492
THYM	318	16,158	119
UCEC	341	15,912	403
UCS	346	16,418	55
UVM	245	15,066	77

Table A.2 Links for data sources(expression files, miRNA-target data and TF regulation data), code and documentation.

	Link
Dataset	https://zenodo.org/record/8127841
Code	https://github.com/oykuaslan/miRCoopv2
Documentation	https://mircoopwebapp.readthedocs.io/en/latest/ heightheight