miRCoop: Identifying Cooperating miRNAs via Kernel Based Interaction Tests

Gulden Olgun and Oznur Tastan

Abstract—Although miRNAs can cause widespread changes in expression programs, single miRNAs typically induce mild repression on their targets. Cooperativity among miRNAs is reported as one strategy to overcome this constraint. Expanding the catalog of synergistic miRNAs is critical for understanding gene regulation and for developing miRNA-based therapeutics. In this study, we develop miRCoop to identify synergistic miRNA pairs that have weak or no repression on the target mRNA individually, but when act together, induce strong repression. miRCoop uses kernel-based statistical interaction tests, together with miRNA and mRNA target information. We apply our approach to patient data of two different cancer types. In kidney cancer, we identify 66 putative triplets. For 64 of these triplets, there is at least one common transcription factor that potentially regulates all participating RNAs of the triplet, supporting a functional association among them. Furthermore, we find that identified triplets are enriched for certain biological processes that are relevant to kidney cancer. Some of the synergistic miRNAs are very closely encoded in the genome, hinting a functional association among them. In applying the method on tumor data with the primary liver site, we find 3105 potential triplet interactions. We believe miRCoop can aid our understanding of the complex regulatory interactions in different health and disease states of the cell and can help in designing miRNA-based therapies. Matlab code for the methodology is provided in https://github.com/guldenolgun/miRCoop.

Index Terms—synergistic miRNAs, miRNA cooperativity, miRNA interactions, kernel interaction test

◆

1 INTRODUCTION

MICRORNAS (miRNAs) are small non-coding RNAs that play extensive regulatory roles across different biological processes and pathologies such as cancer ([23], [49]). By binding the complementary sequence of the target messenger RNAs (mRNAs), miRNAs repress their target mRNA gene expression ([10]). Despite the widespread changes they have on the protein levels, single miRNAs typically induce modest repression on their targets ([7], [54]). Cooperation among miRNAs is one strategy to boost their activity. This cooperativity can be realized in the form of multiple miRNAs collectively co-regulating a cohort of genes that are part of a functional module, or multiple miRNAs can jointly co-target an individual gene to achieve cooperative control of a target gene (reviewed in [14]). In this work, we focus on the latter case.

There is a substantial amount of experimental evidence on the cooperativity of miRNAs, where a stronger degree of repression of a single gene is achieved with multiple miRNAs than is possible by the individual miRNAs. Enright et al. [21] are the first to show the cooperativity of lin–4 and let–7 in Drosophila experimentally. Krek et al. [37] find that in a murine pancreatic cell line that miR–375, miR–124, and let–7b cooperate to repress the MTPN gene. Similarly, Lai et al. [38] report synergistic interaction of miR–93 and miR–572 in repressing the p21 gene in human SK–Mel–147 melanoma cells. Yang et al. [66] shows that miR–17–5p and miR–17–

3p can synergistically induce prostate tumor growth and invasion by repressing the same target TIMP3. Cherone et al. [4] experimentally show that functionally related miR-138/miR137 does not involve mRNA expression, but by co-targeting, they repress the gene expression five to tenfold. Recently, Briskin et al. [5] show that TNRC6 facilitates the cooperative binding of miRNA AGO complexes. Others also show similar lines of evidence on the synergistic interactions of miRNAs to regulate a mutual target ([25], [39], [59], [60], [62]).

Despite the accumulating evidence, our understanding of synergistic miRNAs that regulate a single gene is still rudimentary. Determining which pairs of miRNAs synergize and which mRNA is the target of this cooperativity is critical for understanding how miRNAs function and how they regulate various gene expression programs. Furthermore, synergism can be exploited in designing miRNA based therapeutics, as pointed in a recent survey by Lai et al. [39]. The single miRNA treatments require high amounts of miRNAs, which can provoke off-target effects while therapeutics based on synergistic miRNAs have the potential to overcome these unintended effects by enabling lower levels of each of the miRNAs to be administered ([39]).

In this study, we develop miRCoop for detecting synergistic miRNAs and their mutual targets using the expression data and miRNA target information. We focus on a specific type of synergistic relationship: the miRNA pairs which have weak or no repression on the target mRNA individually, but when together strongly repress its expression. Please note that this does not cover all the synergistic relationships, i.e., the case where miRNAs have moderate expression themselves can also be in a synergistic relationship. Although there are methods that address the

- *G. Olgun, Department of Computer Engineering at Bilkent Universtiy, Ankara, Turkey, 06800.*
  E-mail: gulden@cs.bilkent.edu.tr
- *O. Tastan, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, 34956, Turkey.*
  E-mail: otastan@sabanciuniv.edu

cooperativity of miRNAs - discussed below - there is no work that specifically addresses this specific type of synergistic relationship. We focus on this particular case because this set of miRNAs can uncover more novel RNAs that are important for a disease or condition which cannot be captured by merely checking the miRNAs individual effect, and they are important for miRNA-based therapies.

Several related computational methods exist in the literature on miRNA cooperativity. However, the majority of these tools inspect the cooperativity of miRNAs in regulating *a set of genes* which are part of a functional module ([2], [3], [6], [13], [18], [19], [26], [41], [45], [55], [64], [69]) in contrast to the cooperativity for regulating *a single gene*. Hon et al. [30] is one of the early work that analyzes the combinatorial interactions of miRNAs and their targets. Although the study includes an analysis of the miRNAs that mutually target a single gene, this work does not aim at finding miRNAs that bind their target concurrently in a synergistic manner. The target relationships are established solely based on the predicted miRNA-mRNA interactions without expression data.

TriplexRNA aims at detecting synergistic miRNAs that bind to a mutual target gene [52]. TriplexRNA assesses the synergism of the pairs of miRNAs in the human genome using miRNA target predictions, structural predictions, molecular dynamics simulations, and mathematical modeling methods. The main limitation of this method is that it does not incorporate the gene expression profiles. The gene regulation changes drastically in different conditions; thus, the identified miRNAs are putative triplets and are not context-dependent. CancerNet [45] makes use of the expression data but not for testing the statistical interactions of the triplets. The miRNA-mRNA expression profiles in each cancer are used to remove the targets not expressed from the protein-protein interaction (PPI) network and filter miRNA-mRNA interactions that are not negatively correlated. CancerNet detects synergistic miRNA pairs based on the functional similarity of the miRNAs' targets and the proximity of these targets in the protein-protein interaction network. This functional association does not necessarily point to synergism in regulating a single gene, but instead, it refers to a group of genes regulated together by the same pair of miRNAs.

Finally, Zhang et al. [70] and Zhang et al. [76] uses gene expression data and apply causal inference methods to find synergistic pairs. In Zhang et al. [70], miRNA knockout is simulated using gene expression data to discover synergistic miRNAs. In a more recent study, Zhang et al. [76] extend this work to incorporate the simulation of multiple knockouts of miRNAs. Zhang et al.'s method [70] relies on the IDA method, Interventional calculus when DAG is Absent. Similarly miRsyn, [76] makes use of the joint-IDA method to identify the synergistic miRNAs. IDA learns causal structure by the PC-algorithm [65] and measures the effects with Pearl's do-calculus method [16]. In this method, the variables are assumed to follow a multivariate Gaussian distribution, which might not hold in the case of expression data. IDA also assumes there are no hidden effects on the mRNAs; however, the main regulators of the mRNA are actually the transcription factors. The miRsyn method also includes a filtering step that removes transcripts from the

analysis that are not significantly associated with survival. miRsyn filters the miRNAs and mRNA that are significantly different in their survival using a Cox regression model, which limits the applicability as patient survival data are not available in many situations and the participants of the synergistic relations are not necessarily associated with survival. Another limitation is that the two works do not check if the binding sites of the candidate miRNAs are overlapping or not. If they are overlapping, it is unlikely that the miRNAs can bind at the same time.

Apart from these methods, other proposed algorithms have been proposed to identify cooperative miRNAs that focus on identifying cooperative modules. RFCM3 [8] identifies functional miRNA-mRNA modules by maximizing the relation of mRNAs with miRNAs and functional similarity between selected mRNAs. Similarly, ProModule [9] detects miRNA prognostic modules. However, both methods focus on identifying a set of miRNAs that functionally work as a module without necessarily co-targeting a single gene. Our approach, therefore, is motivated with a different aim.

We cast the problem of finding synergistic miRNAs and their mutual target as a statistical dependence discovery problem. Using matched miRNA and mRNA expression data and considering the expression level of each miRNA and mRNA as a random variable, miRCoop finds a specific three-way statistical interaction among these random variables. Here, the pairs of miRNAs have weak or no statistical dependence with the mRNAs individually, yet when taken jointly, they are dependent on the mRNA. This case is similar to the example wherein adding sugar to coffee or stirring the coffee alone has weak or no effect on the coffee's sweetness, but the joint presence of the two has a strong effect ([53]). This specific case has not been the focus of previous work. In testing whether the triplets of RNAs have such relationships, we resort to the kernelized Lancaster three-variable interaction developed by Sejdinovic et al. [53]. This test relies on mapping the probability distributions into a reproducing kernel Hilbert space (RKHS). The effectiveness of embedding probability distributions are demonstrated in different machine learning and inference tasks (reviewed in [58] and [46]). The key idea of embedding probability distributions is to implicitly map probability distributions into potentially infinite-dimensional feature spaces using kernel functions [57]. Kernel embedding of the probability distributions avoids the restrictive assumptions on the distributions such as variables to be discrete-valued, or assumptions on the relationships among the variables such as the linear relationship between variables ([58]). The kernel trick helps us to deal with small sample sizes for the available expression profiles, and it does not require any distributional assumptions. In addition to the expression profiles, miRCoop makes use of the putative target information among the miRNAs and mRNAs.

We apply miRCoop to kidney cancer using tumor samples from The Cancer Genome Atlas (TCGA) project. This analysis discovers 66 putative triplets. We use several validation strategies to check the existence of a regulatory relationship among the triplets' participants. Investigating the triplets' RNAs reveals that most of them are individually associated with kidney cancer, as well. Furthermore, we observe that some miRNA pairs are encoded very close on

the genome, hinting a tight regulation at the genome level and supporting a functional link among these miRNAs. To check other coordinated regulatory evidence, we investigate whether a shared transcription factor potentially coordinates the miRNAs and RNAs. Interestingly, we observe that this holds for almost all identified triplets. When we explore the relevance of the identified triplets to kidney cancer, our functional enrichment analysis on the largest connected component of miRNA–miRNA–mRNA network highlights critical processes related to kidney cancer, part of which are key driver genes. To show miRCoop consistency, we further test miRCoop on liver cancer and we observe similar results as identified in kidney triplets.

2 MATERIALS AND METHODS

miRCoop relies on miRNA–mRNA target predictions and kernel statistical tests on the expression data. The overall pipeline of miRCoop is summarized in Fig 1. The first step curates a list of possible triplets based on miRNA target predictions and their predicted mRNAs. In the second step, the candidate triplets are pruned based on expression data. In the final stage, we test the remaining triplets for the statistical dependence relationships. Below, we detail these three steps. Following the miRCoop descriptions, we provide information on the null model used for triplets and the data collection and pre-processing steps when miRCoop is applied to kidney cancer.

2.1 miRCoop Step 1: Identifying Candidate miRNA–mRNA Target Interactions

miRCoop first identifies pairs of miRNAs that can bind to a common target mRNA (Fig 1 Step 1). A pair of miRNAs and their mutual target mRNA constitute a candidate triplet. As the interactions of newly cataloged miRNAs are not included in experimentally validated miRNA target interaction databases, we resort to available target prediction algorithms. For this, we use TargetScan ([1]) and miRanda ([12]). When using TargetScan, the 6–mer sites' predictions are filtered since they are classified as poorly conserved ([1]). Similarly, when running miRanda, to reduce false-positive predictions, we apply stringent parameter settings (a pairing score of > 155 and an energy score of < -20). We consider only the miRNA–mRNA pairs that are predicted by both algorithms. The triplets where miRNA pairs have an overlapping binding site on the target mRNA are also excluded.

Although the target information prunes many combinations, there is still an extensive list of candidate triplets at the end of this step. For example, when we apply to kidney cancer (see Results), there are $31,184$ potentially interacting miRNA–mRNA pairs, and there are $163,550$ candidate miRNA–miRNA–mRNA triplets.

2.2 miRCoop Step 2: Expression Filter

We further narrow down the candidate list based on the expression profiles of the miRNAs and mRNAs measured over a set of samples, Fig 1 (Step 2). miRNAs repress the expression of their target mRNA genes. Therefore, we expect that if both miRNAs are upregulated, the mRNA expression

will be lower compared to the case where both miRNAs are down-regulated. To filter for triplets where this expression pattern holds, we divide the samples into different expression subgroups based on the expression states of the potential synergistic miRNA pairs. The first group comprises the samples where both miRNAs are upregulated, and the second group is the subset of samples where both miRNAs are downregulated. In this analysis, miRNA expression in a sample is considered upregulated if it is in the fourth quartile of all samples, and it is deemed to be downregulated if it is in the first quartile. Having grouped the samples, for every triplet in the candidate list, we check whether the mRNA expression levels are significantly lower than the mRNA expression levels compared to the second group. To test for significance, we apply the one-sided Wilcoxon rank-sum test ([61]). This test is used to determine whether two groups of expression values are selected from populations having the same mRNA expression distribution. We apply this test to all possible triplets formed at the end of Step 1 and retain only those for which the null hypothesis is rejected at 0.05 significance level.

2.3 miRCoop Step 3: Testing Statistical Interactions of RNAs with Kernel-Based Interaction Tests

For every triplet candidate that passed through Step 2, we perform statistical interaction tests on RNA expression data to find putative miRCoop triplets, Fig 1 (Step 3). We are looking for a high order interaction between the three RNA entities, that is, while each miRNA expression levels have weak or no interaction on the mRNA expression individually, the simultaneous interaction of the two miRNAs with the mRNA expression levels is strong.

We denote the two miRNAs' expression levels as two random variables: X and Y and define the expression level of the mRNA to be tested as a third random variable denoted as Z. We are interested in cases where the miRNAs are pairwise independent with mRNA: $X \perp\!\!\!\perp Y$ and $Y \perp\!\!\!\perp Z$ but form a mutually dependent triplet because of the following holds: $(\neg((X,Y) \perp\!\!\!\perp Z))$. Note that this also corresponds to checking for a V-structure, a directed graphical model with two independent parents pointing to a single child, where the miRNAs' expression levels (X and Y) are the parent variables, and the child variable is the mRNA expression level (Z) ([53]).

To ensure that the miRNAs have weak influence individually, we first test for pairwise independence among X and Z and Y and Z. In each of these tests, the null hypothesis that the two variables are independent. We continue with triplets, where the null hypothesis is accepted. On these triplets, we apply three-variable kernel interaction tests proposed by [53] and consider only triplets that reject the null hypothesis. Here, we provide a brief description of these kernel-based tests and later detail how we use this method to discover miRCoop triplets.

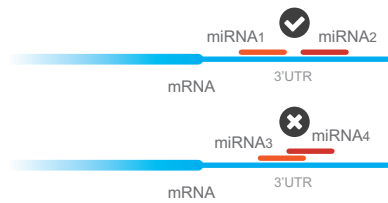#### 2.3.1 Kernel Three-Variable Lancaster Interaction Test

Interaction effects occur when the impact of one variable depends on the value of another variable. In this case, the repression effect of the miRNAs on the mRNA depends on the value of the cooperative miRNA's expression. An
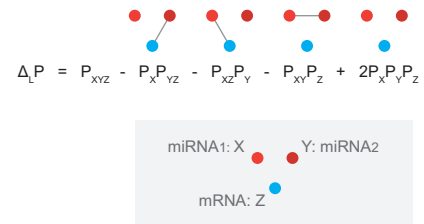
Fig. 1: Overview of the methodology. Each box represents a step in the methodology. The first step identifies miRNA pairs that can target the same RNA based on miRNA target predictions. The second step illustrates how the RNAs are filtered based on expression profiles. The third step illustrates the idea of cooperative RNAs. $\triangle_L P$ denotes the Lancaster interaction on the triple of random variables (X, Y, Z) is defined as the signed measure. P_{XYZ} denotes the joint distributions of X, Y, Z random variables, while P_X, P_Y and P_Z are marginal distributions. If the joint distribution P_{XYZ} factorizes into a product of marginals in any way, then $\triangle_L P$ vanishes. This is used as a test statistic to test for independence.

interaction measure is used to test for the existence of such a statistical dependency. An interaction measure associated with a multidimensional probability distribution P is a signed measure that vanishes whenever P can be factorized as a product of its marginal distributions ([40]). When there are two variables, X, and Y, the Lancaster interaction measure is defined as:

$$\triangle_L P = P_{XY} - P_X P_Y. \tag{1}$$

For three random variables, the Lancaster interaction is defined as:

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

In the above equations, while P_X, P_Y and P_Z are marginal distributions of the random variables X, Y and Z respectively. P_{XYZ} denotes the joint distributions of X, Y, Z random variables, P_{XZ}, P_{XY} and P_{YZ} are pairwise marginal distributions. The $\triangle_L P$ is zero whenever P can be factorized as a product of its marginal distributions. A statistical dependency test can be constructed where the Lancaster interaction is the test statistic ([40], [53]). The null hypothesis, in this case, is that the Lancaster interaction is zero. The null distribution of the test statistic can be estimated using a permutation-based approach where the indices of the variable(s) of interest are shuffled ([53]).

Sejdinovic et al. [53] formulate a kernel-based three-variable Lancaster interaction test by embedding the probability distributions into reproducing kernel Hilbert spaces (RHKS). In the general formulation of [53], the three variables to be tested form a random vector of triplets $(X, Y, Z) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$, where p is the dimension. There is only one measurement per sample in our case, $p = 1$ as for each miRNA and mRNA.

The key idea of embedding probability distributions is to implicitly map probability distributions into potentially infinite dimensional feature spaces using the positive definite *kernel functions* [46]. The positive definiteness of the kernel function guarantees the existence of a dot product space H and a feature mapping $\phi : \mathcal{X} \to H$ such that $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_H$. This dot product can be calculated implicitly with the kernel function without mapping ϕ explicitly. This is commonly referred as the *kernel trick* [50]. The mean embedding of a random variable X is defined as: $\mu_X := \mathbb{E}_X k(X, \cdot) \in \mathcal{F}_k$. In [53], given observations X_i, an estimate of the mean embedding is $\tilde{\mu}_X = \frac{1}{n} \sum_{i=1}^n k(X_i, \cdot)$. Given an *i.i.d.* sample $(X_i, Y_i, Z_i)_{i=1}^n$, the norm of the mean embedding μ_L is empirically estimated using empirically centered kernel matrices. The empirically centered kernel matrix $\tilde{K}$ for the kernel k is given by:

$$\tilde{K}_{ij} = \langle k(X_i, \cdot) - \tilde{\mu}_X, k(X_j, \cdot) - \tilde{\mu}_X \rangle, \tag{3}$$

Sejdinovic et al. [53] arrive at the following test statistics for the two variable and three-variable dependency test using the kernel trick:

$$\| \triangle_L \hat{P} \|_{k \otimes l}^2 = \| \hat{P}_{XY} - \hat{P}_X \hat{P}_Y \|_{k \otimes l}^2 = \frac{1}{n^2} (\tilde{K} \circ \tilde{L})_{++} \tag{4}$$

$$\| \triangle_L \hat{P} \|_{k \otimes l \otimes m}^2 = \frac{1}{n^2} (\tilde{K} \circ \tilde{L} \circ \tilde{M})_{++} \tag{5}$$

Here, k denotes the kernel function defined in domain $\mathcal{X}$ that maps P_X to H. Similarly, l and m are kernel functions defined in domains of $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Using the kernel trick, computations on the probability distributions can be calculated implicitly via kernel matrix operations. The kernel matrix for the kernel k is defined as the $n \times n$ matrix $K_{ij} = k(X_i, X_j)$ computed over all sample pairs, i, j. $\circ$ denotes an Hadamard (element-wise) product of matrices.

$\otimes$ indicates the Kronecker product. $\mathbf{1}$ denotes an $n \times 1$ vector and $A_{++} = \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} A_{ij}$ for a matrix A. $\tilde{K}$ denotes the corresponding centered matrix of K, where $\tilde{K} = HKH$ and $H = I - \frac{1}{n}\mathbf{11}$ and I is the identity matrix.

The same rules apply to the other gram matrices, L and M. Given an i.i.d. sample $(x_i, y_i, z_i)_{i=1}^{n}$, the centered kernel matrices can be computed over the samples. Using equations 5, the norm of the embeddings can be computed using these centered matrices. We refer readers to [53] for the derivation of these equations.

2.3.2 Applying Kernel-based Interaction Tests

We form a random vector of triplets $(X, Y, Z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. In the pairwise interaction tests, where we test for independence of miRNA expression levels with the mRNA (X-Z and Y-Z), we consider triplets where the pairwise independence test fails to be rejected ($\alpha = 0.01$). For the three-variable independence interaction test, we only consider triplets, where the null hypothesis of the three variable interaction is rejected ($\alpha = 0.01$). We use the Gaussian kernel as defined:

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|_2^2}{2\sigma^2}\right) \qquad (6)$$

, where σ is a non-negative real number that determines the width of the kernel. We set σ using the median heuristic ([24]). In both tests, the null distribution is estimated by repeatedly permuting the indices of the mRNA expressions of the samples (Z variable), where we set the number of permutations to $2,000$.

2.4 A Regression-Based Method

We use a simple alternative strategy to detect synergistic miRNAs using linear regression, which forms a baseline. Starting from the candidate triplets after Step 2 in the miRCoop pipeline, we build three regression models. In the first two models, we model the relationship of miRNAs with the mRNA separately by fitting a linear equation (Eq 7 and Eq 8), where X indicates the first miRNA's expression level and Y indicates the second one's expression. In the third model, in addition to the main effects of the individual miRNAs, we include an interaction term ($X * Y$) to measure synergy. A significant, negative, and large in magnitude regression coefficient will indicate a synergy. These three models are given below:

$$Z = \beta_0 + \beta_x X + \varepsilon \qquad (7)$$
$$Z = \beta_0 + \beta_y Y + \varepsilon \qquad (8)$$
$$Z = \beta_0 + \beta_1 X + \beta_2 Y + \beta_{x,y}(X * Y) + \varepsilon \qquad (9)$$

In the first two models, a significant, negative, and small in magnitude coefficient suggests a weak or non-existing linear relationship between the miRNA and the mRNA. To decide whether the magnitude of a coefficient is large or small, we need to set a cut-off value. The coefficients of miRNAs (see Supplementary File 1 Fig S3 and S4) are distributed within the range -2 and 2. We pick -0.25, and this choice corresponds to an approximately 25 percentile for both kidney and liver data.

We expect to see a negative correlation between the interacting miRNA-mRNA. Moreover, since we are interested in cases where the miRNA itself shows weak interaction, we keep the triplets for which both miRNA's regression coefficients are small. Thus, we only consider cases where $-0.25 < \beta_x < 0$ and $-0.25 < \beta_y < 0$ in the first two models (Equations 7 and 8). Having filtered out triplets for which the above conditions on β_x and β_y does not hold, we fit the third model (Equation 9). For this, we look for cases where $\beta_{x,y}$ in Eq. 9 is negative and large. In all cases, we only considered triplets with significant regression coefficients (p-value < 0.05). The p-value is for the t-statistic of the hypothesis test with the null hypothesis that the corresponding coefficient is equal to zero against the alternative hypothesis that it is non-zero.

2.5 The Null Model for miRCoop Triplets

Having identified the triplets, to check the validity of different comparisons with the biological data, we use a random sampling-based approach to obtain a random set of triplets that constitutes the null model. We use these randomized triplets to obtain an empirical null distribution of the test statistic of interest in each case. We construct the randomized triplets as follows: let $S = t_1, t_2, ..., t_n$ be the set of miRCoop triplets identified. Let the i-th triplet member of S denoted by $t_i = (x_i, y_i, z_i)$, which include the miRNA pairs, x_i, y_i and the mRNA z_i. We sample a collection of B random triplet sets of each with size m. Let $S^b = \{t_1^b, t_2^b, \ldots, t_n^b\}$ denote the b-th random triplet set. For the b-th sampling, t_i^b is generated as follows: the mRNA z_i is retrieved, and one miRNA pair of the triplet (x_i or y_i) is selected at random and retained, the paired miRNA triplet is though selected uniformly at random from the set of miRNAs that can potentially target z_i, R_z, as identified at the end of Step 2 of the miRCoop pipeline excluding the miRNA triplets (Fig 1). By repeating this procedure for all samplings $\forall b \in \{1, \ldots, B\}$, we obtain a collection of B many triplet sets, $S_{b b=i}^{B}$ the random counterparts of S. We must note that this procedure constructs a rather stringent null model. Note that if we were to select the mRNAs and miRNAs all at random, the miRNAs and mRNAs in these random triplets would likely be unrelated to each other. With such a null model, attaining significance would be essentially trivial.

We calculate the p-value of the observed statistic empirically based on this null distribution. To assess the statistical significance of the statistic of interest (such as the number of TFs (See sections 3.4, 3.3, 3.6)), we obtain an empirical p-value by calculating the fraction of the cases in B random samplings where the test statistic is as extreme or more extreme than the observed statistic. The null model is constructed with $1,000$ samplings.

2.6 Data Collection and Processing

We apply miRCoop on the patient–derived kidney cancer samples. We retrieve the precomputed mRNA and miRNA expression levels from GDC Data Portal on Sept 2nd 2017. mRNAs were quantified using RNA-seq technology while miRNA's are using miRNA-seq technology. The details of miRNA transcriptomic profiling are described in [17]. The

details of data processing can be accessed at https://docs.gdc.cancer.gov/ and Supplementary File 1 Table S1 presents the details of the data set that are utilized in the analysis. We utilize the available Reads per million mapped reads (RPM) values and Fragments per Kilobase of transcript per Million mapped reads upper quartile normalization (FPKM–UQ) for miRNA and mRNA expression data, respectively. We only use matched primary tumor solid samples that concurrently include mRNA and miRNA expression data, containing 995 tumor patient samples. To filter out the non-coding RNA genes from the RNASeq expression data and create an mRNA list, we used GENCODE v27 annotations ([29]).

We eliminate mRNAs with 3′UTRs shorter than 12 nucleotides since two miRNAs cannot bind on such a short sequence at the same time. 3′UTR sequences of mRNAs are obtained using ENSEMBL BioMart ([20]). Genes with expression values lower than 0.05 in more than 20% of the samples are filtered. Expression values are added with a constant 0.25 to deal with the 0 gene expression values, and data are $\log 2$ transformed. We eliminate RNAs that do not vary across samples. For this purpose, we filter RNAs with a median absolute deviation (MAD) below 0.5. In the final set, we have 12,113 mRNAs and 349 miRNAs in kidney tumor samples.

We also test the miRCoop on liver cancer. We collect mRNA and miRNA expression profiles of liver tumor samples from the GDC Data Portal on April 29th 2020. We use Reads per million mapped reads (RPM) values in miRNA-Seq, and Fragments per Kilobase of transcript per Million mapped reads upper quartile normalization (FPKM–UQ) in RNASeq for miRNA and mRNA expression data, respectively. We only consider 400 tumor samples whose both mRNA and miRNA expression data are available. Also, we remove the non-coding RNA genes from the RNASeq data using GENCODE v27 annotations. We apply the same pre-processing step explained above. After the pre-processing step, 8,588 mRNA and 205 miRNA have remained, and we consider those mRNAs and miRNAs for further analysis.

3 RESULTS

3.1 Overview of the synergistic miRNAs and their targets in Kidney Cancer

We applied the methodology summarized in Fig 1 to find synergistic miRNAs and their targeting mRNAs. The number of candidate triplets that remain after applying each of the filtering steps is provided in Supplementary File 1 Fig S1. The target prediction step produces 31,184 potentially interacting miRNA–mRNA pairs and 163,550 miRNA–miRNA–mRNA interaction combinations. Following the remaining steps, we arrive at potential 66 synergistic triplets, which also contains 66 different synergistic miRNA pairs. The triplets include 75 unique miRNAs. The summary statistics of the expression values of these miRNAs are provided in Supplementary File 4. These triplets include 41 different mRNAs, 27 of which are reported to be cancer driver genes for kidney cancer as reported in IntOGen ([27]). The complete list of the identified triplets is listed in Supplementary File 2, and the network for the potential miRNA-miRNA pairs is provided in Fig 2. In this analysis, we use 0.01 p-value cutoffs for kernel based tests. However;

miRNA	miRNA	Distance	miRNA Family
miR-132	miR-212	262	miR-132
miR-501	miR-502	4792	miR-500
miR-500a	miR-532	5192	miR-500

TABLE 1: The list of identified synergistic miRNA pairs in kidney cancer that are encoded close in the genome and their family information.

we also provide the set of miRCoop triplets for different alpha cutoffs. Those sets are listed in Supplementary File 6 but in the following sections, we will consider the results obtained by 0.01 p-value thresholds.

3.2 Analysis of the Identified Triplets

There is no available experimentally identified set of synergistic miRNAs to validate the triplets we identified. Thus, we devise several strategies to investigate the miRCoop triplets in light of external biological information. To check the statistical significance of each of these findings, we use a null model by generating random synergistic triplets, which we detail in Methods Section 2.4. Below we describe the biological evidence relevant to the identified triplet set.

3.3 Triplets Coded Close on the Genome

Recent studies reveal that miRNAs, which regulate similar processes, are located nearby on the genome ([69]). Others report that miRNAs that are proximally encoded are generally co-expressed, [11], and this indicates a functional link. Thus, we investigate whether the identified synergistic pairs are close on the genome. Specifically, 10 kb of the transcription start and end sites within the genome are considered spatially close. We retrieve known transcription start and end sites of the miRNA from the Ensembl using biomaRt package [20]. We observe that three of such pairs are very close, and these pairs form two different clusters (Table 1). We find this number is not significantly high (p-value = 0.163). The reason could be that most reported since the synergistic miRNAs are not necessarily acting cis. However, those that we observe close in the genome are interesting. These clustered miRNAs are part of the miR–500 and miR–132 miRNA family. There is evidence that miR–500 family takes part in kidney cancer ([22], [44], [56]). Genomic locations of these synergistic miR–500 family miRNAs are illustrated in Fig 3.

3.4 Common TF Regulating the RNAs of the Triplets

Coordinated regulation of genes by a shared transcription factor often indicates a functional link among the genes. To this end, we checked if the identified triplets were potentially co-regulated by a common TF. To be able to conduct this analysis, we first curated a set of regulatory interactions for TFs regulating miRNA and mRNA.

Since none of the TF–miRNA databases are comprehensive, we use two strategies to curate TFs regulating miRNAs. We first use data from available databases; this includes TransmiR ([34]), PuTmiR ([33]) and miRTrans ([32]) databases. miRTrans database provides a list of TFs–miRNA interactions for different tissues and cell types; we only consider the renal cell lines HRE, human Renal
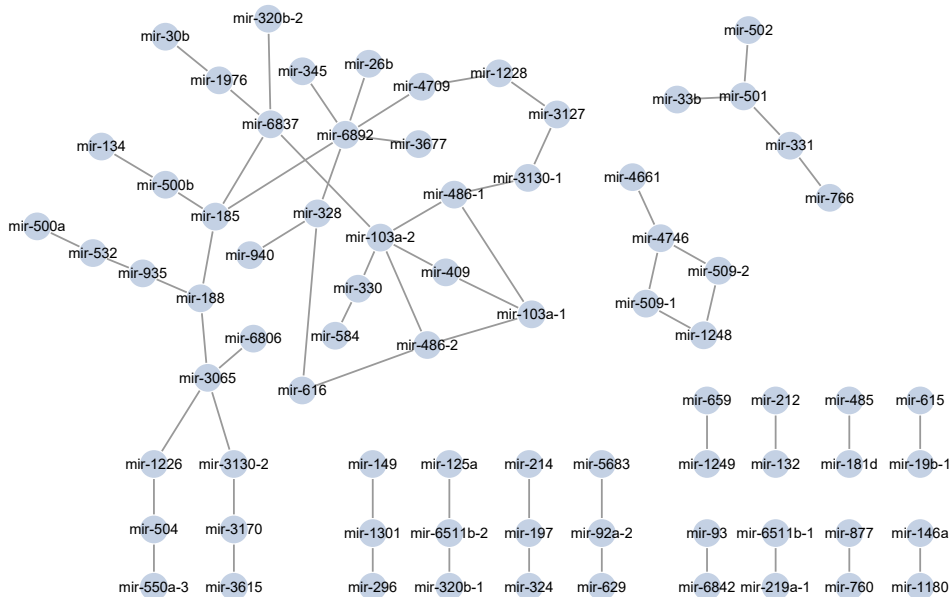
Fig. 2: The network shows the miRNA synergistic pairs discovered for kidney cancer. Nodes are miRNAs and edges exist between them if they are participate in at least one miRCoop triplet.
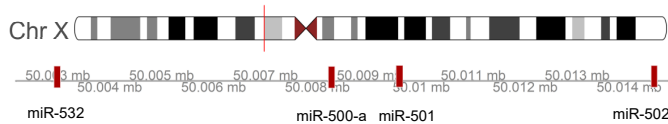

Fig. 3: The red marks on the genome axis indicate the miRNA positions of the synergistic miRNAs that belong to the miR-500 family. The figure was created with Gviz package in R ([28]).

Epithelial Cells, and RPTEC, hTERT immortalized human renal proximal tubular epithelial cell lines. Secondly, we compile the list by analyzing ChIP-Seq data made available by the ENCODE project ([31]) and TF motifs. We retrieve the transcription start sites from Ensembl employing the biomaRt package [20] and $10,000$ upstream regions of transcription start of the gene in question using the R package GenomicRanges ([71]). We check if this region overlaps with the ChIP-Seq peaks in three kidney-related cell lines (HRE, HEK293B, HEK293) using the GLANET tool [75]. We compiled the TF–miRNA interactions from the databases and discovered through analyzing ChIP-Seq data. For the list of TFs regulating mRNAs, we conducted the same analysis on the ChIP-Seq data. Furthermore, we scan the promoter regions with the TFs' binding motifs with JASPAR ([72]) and Hocomoco v.11 ([73]) using the FIMO ([74]) tool with default parameter settings. If the TF has a reported overlapping peak with the promoter and there is a matching motif inside the peak, we consider this as a TF–mRNA interaction.

Finally, we check if there is at least one common TF that regulates all RNAs (the miRNAs and the mRNA) that participate in a triplet interaction. In checking this, we disregard POL2 as it will overlap with promoter regions without any relevance to the functional cooperation. We observe that for

64 of the 66 triplets, all triplets participants are co-regulated by at least one common TF. This number was significantly high at test level (p-value = 0.04). The list of the TF–triplet relations is provided in Supplementary File 3.

We inspect a subset of the TF–triplet regulations more in detail. Recent studies report that the transcription factors, HIF1A ([35]), TFEB, TFE3 and MITF ([51]) are related to the kidney cancer. We check if there are any miRCoop triplets regulated by these transcription factors. These four TFs are linked to a total of 27 triplets (out of 66), and the sub-network includes triplets that are coordinated by multiple of these TFs, Fig 4. Additionally, we observe that the mRNA pairs that are encoded very closely on the genome (see Section 3.3) are among them and have shared TFs. For example, TFE3 potentially regulates both miR–501, miR–502, miR–500a, and miR–532. Apart from this, interestingly, TFs related to the Xp11 translocation renal cell carcinoma disease (TFEB, TFE3 and MITF) ([51]) regulate the same triplets: (miR103a.1, miR486.1, PDZD8) and (miR103a.1, miR330, AMOT).

3.5 Functional Enrichment Analysis

Synergistic miRNAs are likely to regulate genes with the same or similar functions ([64]). To investigate the potential biological process related to the synergistic miRNAs, we conduct a functional enrichment analysis on them. To this end, we first build a miRNA–miRNA–mRNA network. In this network, nodes are miRNAs and mRNAs, and an edge exists between two RNAs whenever they are connected through at least one miRCoop triplet. A connected component analysis reveals that the network includes 13 connected components, Supplementary File 1 Fig S2. Interestingly, the largest connected component covers a very large fraction of the all miRNAs: of the 75 miRNAs that is part of the triplets, 41 of them were in this largest connected component (p-value = 0.032) (Fig 5 (a)) indicating that the cooperative
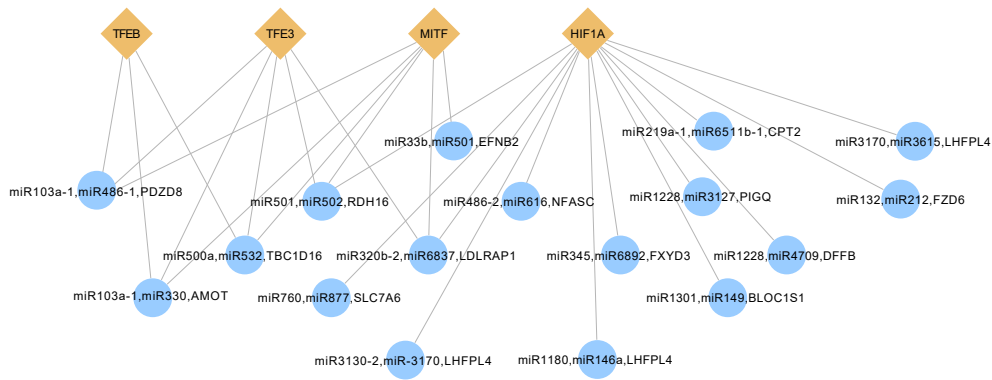
Fig. 4: The network for the synergistic triplets regulated by common TFs. Each node denotes a triplet or a TF while an edge represents an interaction between them. TFs are represented by the diamond-shaped yellow symbol, and triplets are represented by the circle-shaped blue symbol.

miRNAs are directly or indirectly linked to each other through their targets.

Gene ontology (GO) enrichment analysis is GO analysis is conducted with clusterProfiler [68] package in R. We use the default background set that is all human genes annotated by GO Biological Process in the org.Hs.eg.db database [15]. We set a p-value cutoff of 0.05, an FDR cutoff 0.05, and used Bonferroni multiple hypothesis test correction.

The GO enrichment analysis reveals that the set of mRNAs are enriched in processes that are associated with kidney cancer. For example, metabolic networks such as the receptor metabolic processes, the negative regulation of hydrolase activity, and the regulation of metal ion transport are all found to be enriched (Fig 5 (b)). Among these, sodium and guanylate related processes are particularly reported as important in kidney cancer ([36], [47], [48]). Also, the genes that are already known as the drivers of renal cancer are annotated with these processes. An example is Angiomotin (AMOT). [42], reports that AMOT is an oncogene in renal cancer. AMOT family members are also reported to promote cancer initiation in other cancers and act as a tumor suppressor in some others ([43], [67]).

Interestingly, the brain related biological processes such as layer formation in the cerebral cortex, synapse organization, embryonic brain development are found to be enriched (Fig 5(b)). It is reported that brain metastasis is common for renal cancer, and in 15 % of cases, the tumor metastasizes to the brain ([63]). Additionally, brain metastasis is the main cause of morbidity and mortality in kidney cancer ([63]). These findings suggest that mRNAs regulated by the synergistic triplets may participate in the mechanisms that promote brain metastasis.

3.6 Comparison with Other Methods

We first check the overlap of the identified triplets with that of other computational methods available in the literature: CancerNet ([45]), TriplexRNA ([52]) and miRsyn [76] and a simple regression-based model (See Section 2.4). CancerNet provides cancer-specific pre-computed synergistic pairs of

miRNAs. We obtain synergistic CancerNet miRNA pairs from Kidney Chromophobe, Kidney Renal Clear Cell Carcinoma, and Kidney Renal Papillary Cell Carcinoma results, which make a big pool of CancerNet synergistic miRNA pairs. Of the 66 synergistic miRNA pairs, miRCoop identifies, 35 of them are also suggested by CancerNet. This overlap is highly significant based on the sampling-based null model with a sampling size of $1,000$ (p-value ≤ 0.001). We also check the overlap with the synergistic pairs provided by TriplexRNA. We collect the TEC predictions, which is one the largest pool of synergistic data for comparison among the reported four sets. However, we do not find a significant overlap. One reason for this minimal overlap could be that TriplexRNA triplets are not cancer-specific while the triplets we identified are. miRCoop does not have any overlapping triplets with the miRsyn and the regression-based model. The sizes of the overlaps are summarized in the Venn diagram in Fig 6 (A).

We also compare miRCoop with miRsyn [76]. We run miRSyn on the same data using the same data described in Section 2.6 using suggested parameters. When considered only the pairwise synergistic pairs, miRsyn return $5,284$ miRNA-miRNA-mRNA triplets that involves $1,079$ synergistic miRNAs. There was no overlap of the triplets with miRsyn, although there were 11 miRNAs that overlap. MiRsyn employs a filtering step based on Cox regression, thus eliminates transcripts that do not have prognostic value, and the methods used are different. These could be partly the reason why there is no overlap.

We built the regression model (Section 2.4) to constitute a baseline to assess whether miRCoop triplets could be identified with a more straightforward strategy. The distribution of the beta coefficients are provided in Supplementary File 1 Figure S3 and S4 and this simple method results in 24 potential triplets, which do not have overlap with miRCoop triplets or any of the tools we compare the miRCoop against (shown in the Venn diagram provided in Fig 6).

We also compare miRCoop with miRsyn and the regression method in terms of the genome and the TF relations. For the regression model, none of the synergistic miRNAs are located close by on the genome. We also find that none

(a) The largest connected component
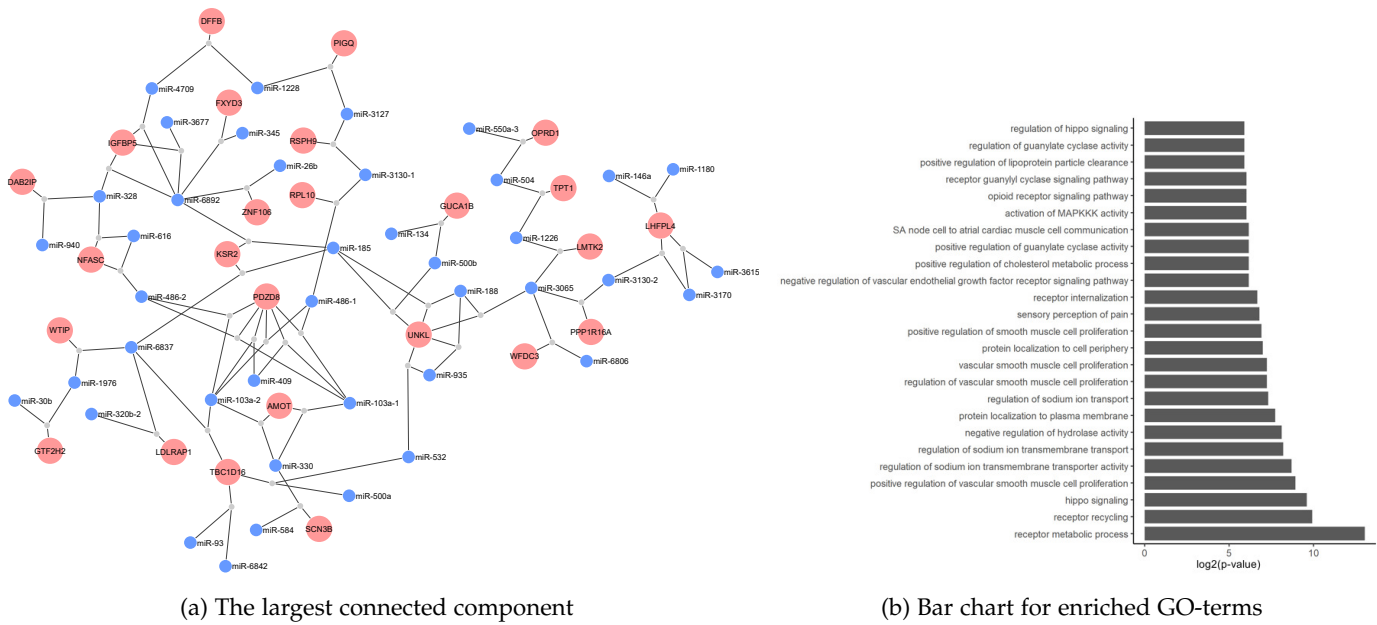

(b) Bar chart for enriched GO-terms

Fig. 5: a) The the largest connected component of miRCoop triplet network. Blue nodes are miRNAs and pink colors are mRNAs of the triples. The gray nodes are dummy nodes to connect the triplet miRNAs to the target mRNA. b) Bar chart for the top 25 enriched biological process GO-terms (p-value < 0.05) found over the mRNA set of the largest connected component in part a). The x-axis shows $\log 2$ transformed p-values.
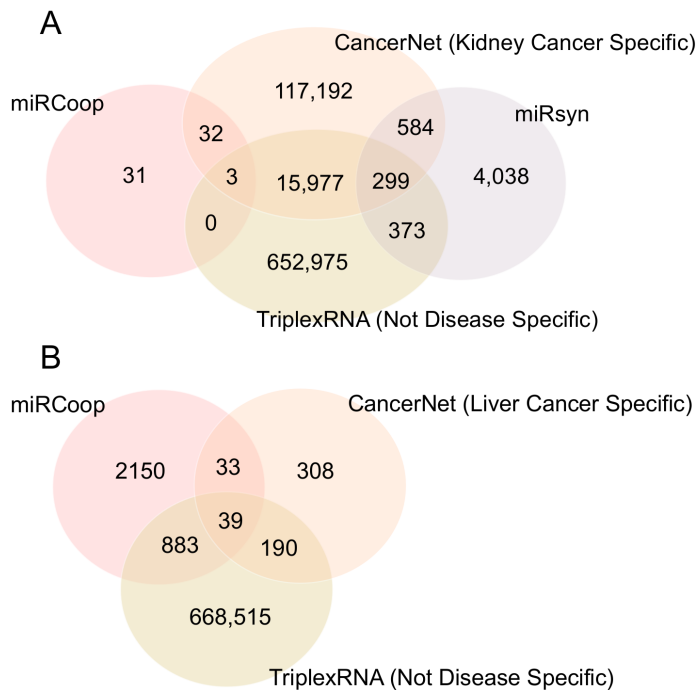

Fig. 6: The Venn diagrams for the number of synergistic miRNAs that are detected by different methods. A) Results for kidney cancer. For CancerNet predictions, kidney cancer-specific results are used. miRsyn results are obtained by running the same TCGA kidney data that we used for miRCoop. B) Results for liver cancer. Liver Hepatocellular Carcinoma miRNA-miRNA synergistic pairs are utilized for CancerNet. In the TriplexRNA database, we utilize TEC predictions for both kidney and liver cancer.

of the triplets are regulated with a common transcription factor. Thus, we conclude that this simple strategy is not sufficient to solve this problem. This method has various limitations, as well. It assumes that the nature of the statistical interaction of the individual miRNAs and mRNAs are linear. Also, the linear regression model assumes that the mRNA expression values are normally distributed (as the errors are assumed to be normally distributed in the standard regression). We also need to set arbitrary cutoffs on the coefficients to decide the strong and weak interactions.

We analyzed the synergistic miRNAs in miRsyn that are encoded close-by on the genome. We observe that 10 out of $1,079$ synergistic miRNA pairs suggested by miRsyn [76] are close-by on the genome. The number of 10 was not significantly high at confidence level 0.05 when tested with the permutation test. Next, we check if miRsyn triplets are co-regulated by common TFs. We conducted a statistical test exactly as we did in miRCoop to check if this number is interesting high. The synthetic interactions in the null model are constructed as described in Section 2.5. The test fails to reject the null hypothesis with p-value $= 1.00$ over 1000 samplings, and we observed that $3,810$ triplets are co-regulated by at least one common TF.

We further assess of the agreement of a method with other methods. For each method, we calculate the number of common miRNA pairs with the second method and divide this number by the number of all pairs that are detected by that method. We observe that among all the methods miRCoop has the highest agreement with other methods, Table 2. We also calculate a summary overall agreement of each method by dividing the number of miRNAs that common to two method compared divided by the number of total pairs of the method of interest. For example, TriplexRNA has $669,627$ total miRNA pairs, and among

those pairs, only 2.48% of them are identified by other methods, whereas this number is 53.03% for miRCoop.

TABLE 2: The agreement of a method in the row, method A, with a given method in the column, method B, is provided as a percentage. This number is calculated by dividing the number of common miRNA pairs between methods A and B with the total number of pairs identified by method B. We also provide an overall agreement score for each method, this is obtained by dividing the number of pairs that method B shares with at least one of the other methods with the number of pairs obtained by method B.

	miRCoop	CancerNet	TriplexRNA	miRsyn
miRCoop	100.00%	0.03%	0.00%	0.00%
CancerNet	53.03%	100.00%	2.43%	16.67%
TriplexRNA	4.55%	12.14%	100.00%	12.69%
miRSyn	0.00%	0.66%	0.10%	100.00%
Overall	53.03%	12.60%	2.48%	23.72%

Finally, we calculated the correlation between the expression values of the identified synergistic miRNA pairs. The histogram of these correlation values is provided in Supplementary File Figure S5 (A), mean values with blue dashed lines $(\mu(\rho_{kidney}) = 0.15$). When we calculate the correlation between synergistic CancerNet miRNA-miRNA pairs, we observe that the mean of the CancerNet kidney coefficient values $((\mu(\rho_{kidney}) = 0.0755)$ are smaller than the miRCoop correlation values (see Supplementary File Figure S5 (C)). Note that since our framework focus on the weak or low repression on the mRNA, this would also exclude some highly correlated miRNA pairs.

We also perform a randomization test to assess if these correlation values of the synergistic miRNA pairs by miRCoop are higher than expected. We sample miRNAs pairs set with the same size and calculated the correlation values among each pair and calculated the mean correlation coefficient values for this random set. We computed the empirical p-value as the fraction of the cases where the random set's mean value is higher than the correlation mean value of the actual synergistic miRNA set. In creating the synthetic pairs, we take the same approach as the other experiments, as described in Section 2.5. We use a random sampling-based approach to obtain a random set of triplets that constitutes the null model. Let (miRNA1, miRNA2, mRNA) be a triplet identified by miRCoop. For this triplet, in the first step, we randomly select one of the miRNAs: miRNA1 or miRNA2. A miRNA is then randomly chosen among the set of miRNA targets for mRNA and paired with the chosen miRNA in the first step. We repeat this procedure to generate synthetic pairs with the same size as the observed triplets; in our case, the set size is 66 for kidney cancer. When we conduct this experiment, we find that the empirical p-value is 0.018, which rejects the null hypothesis that the mean correlation coefficient values for the random synergistic miRNA are higher than the mean correlation than the observed miRCoop identified pair set.

3.7 Testing miRCoop on Liver Primary Site

To exhibit the miRCoop applicability to other cancers, we apply miRCoop on liver cancer. After the pre-processing step, $8,588$ mRNAs and 205 miRNAs remain, which we consider for further analysis. As described in Section 2.1, we use TargetScan and miRanda to find miRNA-mRNA target interactions. To decrease the false-positive predictions, we remove the 6-mer sites' predictions in TargetScan, and we apply stringent cutoffs for miRanda, a pairing score of > 155, and an energy score of < -20.

In the final set, $39,856$ potential pairwise miRNA-mRNA interactions are identified. We construct the potential miRNA-miRNA-mRNA set, and we filter these triplets according to their expressions, as described in Step 2, Section 2.2. The final set include $233,186$ candidate miRNA-miRNA-mRNA interactions. We test pairwise miRNA-mRNA interactions with kernel-based two-variable test and miRNA-miRNA-mRNA interaction with kernel-based Lancaster interactions. We detect $3,105$ potential synergistic miRNA-miRNA-mRNA interactions when we use 0.01 p-value cut-off. This number is much higher than the number of synergistic kidney triplets, probably because we have more candidate triplets after Step 2 for the liver in comparison to the kidney. The complete list of the identified triplets for liver cancer is listed in Supplementary File 5.

We evaluate liver miRCoop results with other methods, CancerNet, and TriplexRNA. When we consider the triplets computed in Liver Hepatocellular Carcinoma provided in CancerNet and we identify 72 common synergistic miRNA pairs with miRCoop. The shared triplets of miRCoop and TriplexRNA amount to 922. Importantly, this overlap is highly significant based on the sampling-based null model with a sampling size of $1,000$ (p-value < 0.0001). The Venn diagram for different methods are provided in Fig 6 (B). We also calculate the Pearson correlation coefficient between the miRNAs; the mean correlation was $\mu(\rho_{liver}) = 0.19$. The histogram of these correlation values is shown in Supplementary File Figure S5 (B). When we check against the correlation coefficients for CancerNet miRNA-miRNA pairs, we again detect that the mean of the CancerNet liver coefficient values is $\mu(\rho_{liver}) = 0.14$ Supplementary File Figure S5 (D) for the histogram of CancerNet coefficient values.)

We also apply the regression based method to detect triplets described in Section 2.2 on liver cancer data. To detect those triplets, we build regression models on candidate miRNA-miRNA-mRNA triplets identified after Expression Filter Step 2, Section 2.2. We apply the same thresholds for coefficients beta and p-values. We detect that 2 of the 22 triplets are common with miRCoop identified triplets. Moreover, 12 out of 21 cooperative miRNA pairs are mutual by miRCoop liver miRNA pairs. Finally, we associate liver and kidney cancer results. We observe that miRCoop recognizes 8 the same synergistic miRNA pairs cross cancers.

4 Conclusion

Post-transcriptional regulation of gene expression mediated by miRNAs is a fundamental mechanism to fine-tune the protein levels in the cell. The dysregulation of miRNA expression is also associated with many diseases, including cancer. miRNAs are not only potential biomarkers, but are also investigated as promising therapeutic agents; therefore, understanding the intricate ways they work is essential for

both expanding the knowledge on how cells work and translational purposes. In this work, we focus on the synergistic target regulation of miRNAs. To identify synergistic miRNAs pairs and their target mRNA, we propose a novel method: miRCoop. MiRCoop uses kernel-based statistical interactions and finds triplets where miRNA expression levels have no or only weak statistical dependencies with the mRNA levels but are jointly dependent. miRCoop makes use of both RNA sequence information for target predictions. We demonstrate the use of miRCoop in kidney liver cancer and find triplets with supporting biological evidence on their relevance. The work presented here can be extended in several directions.

Although it is possible to extend the Lancaster interaction test to detect interactions of higher-order, where more than two miRNAs regulate a mutual target, it would necessitate significantly higher computational resources. Therefore, at the moment, we limit our analysis to the three-variable interaction case. Finding higher-order interactions with ways to reduce the computation time can be a future line of research.

In this work, we focus on miRNAs that regulate a target gene directly which is those that simultaneously bind to the target. There are other ways in which pairs of miRNAs regulate a common target. For example, while one miRNA targets the gene of interest directly, another miRNA can target the transcription factor that regulates this gene ([14], [70]). In this case, one miRNA is acting directly on the target, and the partner miRNA is acting indirectly through the transcription factor. An additional line of future work would be to extend miRCoop to discover these multi-level regulations.

Since there are no synergistic interaction datasets that are readily available, we devise various strategies to check if the identified triplets and the individual RNAs are implicated in kidney cancer and have any biological evidence supporting a functional theme or regulatory interaction among themselves. When experimentally validated synergistic miRNA pairs become available, the analysis can be further improved in the light of the ground truth interactions. For example, we set the kernel width using the median heuristic, but the width could be tuned when an extensive set of ground truth interactions become available.

In the present study, to establish target relations between the miRNA and the mRNAs, we use two target prediction algorithms. We resort to prediction algorithms because newly cataloged miRNAs – which is valid for the TCGA profile miRNAs - are not incorporated in the experimentally validated miRNA target interaction datasets. To reduce the number of false positives, we only consider target interactions that are supported by both algorithms concurrently. The availability of experimentally validated miRNA-mRNA interactions should strengthen the outlined methodology. The number of synergistic triplets can be adjusted by changing the significance levels applied in the statistical tests. For example, in a scenario where a large number of interactions can be verified experimentally at scale, these settings can be adjusted to output more triplets.

In this study, we use all kidney and liver cancer patients to have a large sample size. However, the set of synergistic miRNAs could be different for different subpopulations of the same cancer type. As more data accumulate, the analysis can be repeated over subtypes of cancer.

Although we showcase results for kidney and liver cancer, the method can be applied to other diseases as well as healthy cells. The method can be extended to work with other species other than human whenever the expression profiles of matched miRNAs and mRNAs are available for over a set of samples, and miRNA target predictions are available.

REFERENCES

[1] Agarwal, V. *et al.* (2015). Predicting effective microrna target sites in mammalian mrnas. *elife*, **4**, e05005.
[2] An, J. *et al.* (2010). Identifying co-regulating microrna groups. *Journal of bioinformatics and computational biology*, **8**(01), 99–115.
[3] Antonov, A. V. *et al.* (2009). Geneset2mirna: finding the signature of cooperative mirna activities in the gene lists. *Nucleic acids research*, **37**(suppl_2), W323–W328.
[4] Cherone, J. M., Jorgji, V, and Burge C. B. Cotargeting among microRNAs in the brain *Genome Research*, **29**(11), 1791–1804.
[5] Briskin, D., Wang, P. Y., and Bartel, D. P. (2020). The biochemical basis for the cooperative action of microRNAs. *Proceedings of the National Academy of Sciences*, **117(30)** 17764–17774.
[6] Cantini, L.*et al.* (2019). Identification of microRNA clusters cooperatively acting on epithelial to mesenchymal transition in triple negative breast cancer. *Nucleic acids research*, **47.5** 2205-2215.
[7] Baek, D. *et al.* (2008). The impact of micrornas on protein output. *Nature*, **455**(7209), 64.
[8] Paul, S., and Singh, M. (2019) RFCM3: Computational Method for Identification of miRNA-mRNA Regulatory Modules in Cervical Cancer. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*
[9] Pan, C., Luo, J., and Zhang, J. (2019). Computational Identification of RNA-Seq Based miRNA-Mediated Prognostic Modules in Cancer. *IEEE Journal of Biomedical and Health Informatics*, **24**(2), 626–633.
[10] Bartel, D. P. (2009). Micrornas: target recognition and regulatory functions. *cell*, **136**(2), 215–233.
[11] Baskerville, S. and Bartel, D. P. (2005). Microarray profiling of micrornas reveals frequent coexpression with neighboring mirnas and host genes. *Rna*, **11**(3), 241–247.
[12] Betel, D. *et al.* (2010). Comprehensive modeling of microrna targets predicts functional non-conserved and non-canonical sites. *Genome biology*, **11**(8), R90.
[13] Boross, G. *et al.* (2009). Human micrornas co-silence in well-separated groups and have different predicted essentialities. *Bioinformatics*, **25**(8), 1063–1069.
[14] Bracken, C. P. *et al.* (2016). A network-biology perspective of microrna function and dysfunction in cancer. *Nature Reviews Genetics*, **17**(12), 719.
[15] Carlson, M. (2019). org.hs.eg.db: Genome wide annotation for human. r package version 3.8.2.
[16] Pearl, J. (2009). Causality. Cambridge university press.
[17] Chu, A. *et al.* (2016). Large-scale profiling of micrornas for the cancer genome atlas. *Nucleic acids research*, **44**(1), e3–e3.
[18] Ding, J. *et al.* (2014). Microrna modules prefer to bind weak and unconventional target sites. *Bioinformatics*, **31**(9), 1366–1374.
[19] Ding, J. *et al.* (2017). Ccmir: a computational approach for competitive and cooperative microrna binding prediction. *Bioinformatics*, **34**(2), 198–206.
[20] Durinck, S. *et al.* (2009). Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, **4**(8), 1184.
[21] Enright, A. J. *et al.* (2003). Microrna targets in drosophila. *Genome biology*, **5**(1), R1.

[22] Fedorko, M. *et al.* (2016). Micrornas in the pathogenesis of renal cell carcinoma and their diagnostic and prognostic utility as cancer biomarkers. *The International journal of biological markers*, **31**(1), 26–37.

[23] Friedman, R. C. *et al.* (2009). Most mammalian mrnas are conserved targets of micrornas. *Genome research*, **19**(1), 92–105.

[24] Fukumizu, K. *et al.* (2009). Kernel choice and classifiability for rkhs embeddings of probability distributions. In *Advances in neural information processing systems*, pages 1750–1758.

[25] Gam, J. J. *et al.* (2018). A mixed antagonistic/synergistic mirna repression model enables accurate predictions of multi-input mirna sensor activity. *Nature communications*, **9**(1), 2430.

[26] Gennarino, V. A. *et al.* (2012). Identification of microrna-regulated gene networks by expression analysis of target genes. *Genome research*, **22**(6), 1163–1172.

[27] Gonzalez-Perez, A. *et al.* (2013). Intogen-mutations identifies cancer drivers across tumor types. *Nature methods*, **10**(11), 1081.

[28] Hahne, F. and Ivanek, R. (2016). Statistical genomics: methods and protocols. *Chapter visualizing genomic data using Gviz and Bioconductor. New York: Springer New York*, pages 335–51.

[29] Harrow, J. *et al.* (2012). Gencode: the reference human genome annotation for the encode project. *Genome research*, **22**(9), 1760–1774.

[30] Hon, L. S. and Zhang, Z. (2007). The roles of binding site arrangement and combinatorial targeting in microrna repression of gene expression. *Genome biology*, **8**(8), R166.

[31] Consortium, E. P. *et al.* (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, **489**(7414), 57.

[32] Hua, X. *et al.* (2017). mirtrans: a resource of transcriptional regulation on micrornas for human cell lines. *Nucleic acids research*, **46**(D1), D168–D174.

[33] Bandyopadhyay, S. and Bhattacharyya, M. (2010). Putmir: a database for extracting neighboring transcription factors of human micrornas. *BMC bioinformatics*, **11**(1), 190.

[34] Wang, J. *et al.* (2009). Transmir: a transcription factor–microrna regulation database. *Nucleic acids research*, **38**(suppl_1), D119–D122.

[35] Sanchez, D. J. and Simon, M. C. (2018). Transcriptional control of kidney cancer. *Science*, **361**(6399), 226–227.

[36] Jeppesen, A. *et al.* (2010). Hyponatremia as a prognostic and predictive factor in metastatic renal cell carcinoma. *British journal of cancer*, **102**(5), 867.

[37] Krek, A. *et al.* (2005). Combinatorial microrna target predictions. *Nature genetics*, **37**(5), 495–500.

[38] Lai, X. *et al.* (2012). Computational analysis of target hub gene repression regulated by multiple and cooperative mirnas. *Nucleic acids research*, **40**(18), 8818–8834.

[39] Lai, X. *et al.* (2019). Systems biology-based investigation of cooperating micrornas as monotherapy or adjuvant therapy in cancer. *Nucleic acids research*.

[40] Lancaster, H. O. and Seneta, E. (1969). *Chi-square distribution*. Wiley Online Library.

[41] Li, Y. *et al.* (2014). Mirsynergy: detecting synergistic mirna regulatory modules by overlapping neighbourhood expansion. *Bioinformatics*, **30**(18), 2627–2635.

[42] Lv, M. *et al.* (2016). Angiomotin promotes renal epithelial and carcinoma cell proliferation by retaining the nuclear yap. *Oncotarget*, **7**(11), 12393.

[43] Lv, M. *et al.* (2017). Angiomotin family members: oncogenes or tumor suppressors? *International journal of biological sciences*, **13**(6), 772.

[44] Mangolini, A. *et al.* (2014). Differential expression of microrna501-5p affects the aggressiveness of clear cell renal carcinoma. *FEBS Open Bio*, **4**, 952–965.

[45] Meng, X. *et al.* (2015). Cancernet: a database for decoding multilevel molecular interactions across diverse cancer types. *Oncogenesis*, **4**(12), e177.

[46] Muandet, K. *et al.* (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, **10**(1-2), 1–141.

[47] Nagy, I. Z. *et al.* (1981). Intracellular na+: K+ ratios in human cancer cells as revealed by energy dispersive x-ray microanalysis. *The Journal of cell biology*, **90**(3), 769–777.

[48] Oosterwijk, E. and Gillies, R. (2014). Targeting ion transport in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1638), 20130107.

[49] Peng, Y. and Croce, C. M. (2016). The role of micrornas in human cancer. *Signal transduction and targeted therapy*, **1**, 15004.

[50] Schölkopf, B. *et al.* (1999). *Advances in Kernel Methods: Support Vector Learning*. MIT Press.

[51] Skala, S. L. *et al.* (2018). Detection of 6 tfeb-amplified renal cell carcinomas and 25 renal cell carcinomas with mitf translocations: systematic morphologic analysis of 85 cases evaluated by clinical tfe3 and tfeb fish assays. *Modern Pathology*, **31**(1), 179.

[52] Schmitz, U. *et al.* (2014). Cooperative gene regulation by microrna pairs and their identification using a computational workflow. *Nucleic acids research*, **42**(12), 7539–7552.

[53] Sejdinovic, D. *et al.* (2013). A kernel test for three-variable interactions. In *Advances in Neural Information Processing Systems*, pages 1124–1132.

[54] Selbach, M. *et al.* (2008). Widespread changes in protein synthesis induced by micrornas. *nature*, **455**(7209), 58.

[55] Shalgi, R. *et al.* (2007). Global and local architecture of the mammalian microrna–transcription factor regulatory network. *PLoS computational biology*, **3**(7), e131.

[56] Shu, X. *et al.* (2017). Microrna profiling in clear cell renal cell carcinoma tissues potentially links tumorigenesis and recurrence with obesity. *British journal of cancer*, **116**(1), 77.

[57] Smola, A. *et al.* (2007). A hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pages 13–31. Springer.

[58] Song, L. *et al.* (2013). Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Processing Magazine*, **30**(4), 98–111.

[59] Wahdan-Alaswad, R. and Liu, B. (2013). "sister" mirnas in cancers.

[60] Wang, S. *et al.* (2013). Functional cooperation of mir-125a, mir-125b, and mir-205 in entinostat-induced downregulation of erbb2/erbb3 and apoptosis in breast cancer cells. *Cell death & disease*, **4**(3), e556.

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

[62] Wu, S. *et al.* (2010). Multiple micrornas modulate p21cip1/waf1 expression by directly targeting its 3 untranslated region. *Oncogene*, **29**(15), 2302.

[63] Wyler, L. *et al.* (2014). Brain metastasis in renal cancer patients: metastatic pattern, tumour-associated macrophages and chemokine/chemoreceptor expression. *British journal of cancer*, **110**(3), 686.

[64] Xu, J. *et al.* (2010). Mirna–mirna synergistic network: construction via co-regulating functional modules and disease mirna topological features. *Nucleic acids research*, **39**(3), 825–836.

[65] Spirtes, Peter *et al.* (2000). Causation, prediction, and search. MIT press

[66] Yang, X. *et al.* (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.

[67] Yi, C. *et al.* (2013). The p130 isoform of angiomotin is required for yap-mediated hepatic epithelial cell proliferation and tumorigenesis. *Sci. Signal.*, **6**(291), ra77–ra77.

[68] Yu, G. *et al.* (2012). clusterprofiler: an r package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, **16**(5), 284–287.

[69] Yuan, X. *et al.* (2009). Clustered micrornas' coordination in regulating protein-protein interaction network. *BMC systems biology*, **3**(1), 65.

[70] Zhang, J. *et al.* (2016). Identifying mirna synergistic regulatory networks in heterogeneous human data via network motifs. *Molecular bioSystems*, **12**(2), 454–463.

[71] Lawrence, M. *et al.* (2013). Software for computing and annotating genomic ranges. *PLoS computational biology*, **9**(8), e1003118.

[72] Khan, A. *et al.* (2017). Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research*, **46**(D1), D260–D266.

[73] Kulakovskiy, I. V. *et al.* (2017). Hocomoco: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. *Nucleic acids research*, **46**(D1), D252–D259.

[74] Grant, C. E. *et al.* (2011). Fimo: scanning for occurrences of a given motif. *Bioinformatics*, **27**(7), 1017–1018.

[75] Otlu, B. *et al.* (2017). Glanet: genomic loci annotation and enrichment tool. *Bioinformatics*, **33**(18), 2818–2828.

[76] Zhang, J. *et al.* (2019). Identifying mirna synergism using multiple-intervention causal inference. *bioRxiv*, page 652180.

Gulden Olgun has received B.S., MSc, Ph.D. degrees from Department of Computer Engineering, Bilkent University. She is a Postdoctoral Visiting Fellow at Cancer Data Science Lab, NCI/NIH. Her research interests include machine learning, bioinformatics, and computational biology.

Oznur Tastan obtained her B.S. degree in Biological Sciences and Bioengineering from Sabanci University (2004) and her M.S. and Ph.D. degrees in Computer Science from Carnegie Mellon University (2007 and 2011), respectively. She worked as a postdoctoral researcher at Microsoft Research New England. She worked as an assistant professor at Computer Engineering Department of Bilkent University from 2012 to 2017. Since 2017, she has been an assistant professor at the Faculty of Engineering and Natural Sciences of Sabanci University, affiliated with the Computer Science and Engineering and Molecular Biology Genetics and Bioengineering programs.