miRCoop: Identifying Cooperating miRNAs via Kernel Based Interaction Tests

Gulden Olgun[1] and Oznur Tastan[2,*]
[1] Department of Computer Engineering, Bilkent University, Ankara, Turkey
[2] Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla, Istanbul, Turkey
* otastan@sabanciuniv.edu

Abstract

Although miRNAs can cause widespread changes on expression programs, single miRNAs typically induce mild repression on their targets. Cooperativity is reported as one strategy to overcome this constraint. Expanding the catalog of synergistic miRNAs is critical for understanding the regulation of various gene expression programs. In this study, we develop miRCoop to identify synergistic miRNA pairs which have weak or no repression on the target mRNA, but when bound together induce strong repression of their target's expression. To discover triplets of RNAs whose expression levels follow these statistical interaction patterns, miRCoop uses kernel-based interaction tests together with miRNA and mRNA target information. We apply our approach to kidney tumor and 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 triplets are enriched for biological processes, which are relevant to kidney cancer. Some of the synergistic miRNAs are very closely encoded in the genome hinting a functional association among them. We believe miRCoop can aid our understanding of the complex regulatory interactions in different health and disease states of the cell.

1 Introduction

MicroRNAs (miRNAs) are small non-coding RNAs that play extensive regulatory roles across different biological processes and pathologies such as cancer [1, 2]. By binding the complementary sequence of the target messenger RNAs (mRNAs), miRNAs repress their target mRNA gene expression [3]. Despite the widespread changes they have on the protein levels,single miRNAs typically induce modest repression on their targets [4]. 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 [5]). In this work, we focus on the latter case.

There is a substantial amount of experimental evidence on the cooperativity of miRNAs to target an individual gene where a stronger degree of regulation is achieved with multiple miRNAs than is possible by the single miRNAs. Enright et al. [6] are the first to show the cooperativity of lin–4 and let–7 in Drosophila experimentally. Krek et al. [7] 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. [8] reports synergistic interaction of miR–93 and miR–572 in repressing the p21 gene in human SK–Mel–147 melanoma cells. Yang et al. [9] shows that miR–17–5p and miR–17–3p can synergistically induce prostate tumor growth and invasion by repressing the same target TIMP3. Others also show similar lines of evidence on the synergistic interactions of miRNAs to regulate a mutual target [10, 14, 13, 12, 11].

Despite the accumulating evidence, our understanding of synergistic miRNAs is still rudimentary.

Determining which pairs of miRNAs can synergize and which gene 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 [11]. 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 [11].

Several computational methods exist in the literature to aid our understanding of the cooperative miRNAs. However, the majority of these inspect the cooperativity of miRNAs in regulating a set of genes which are part of a functional module [23, 25, 16, 17, 15, 24, 21, 20, 18, 22, 19]. On the other hand, there are only a few computational methods for discovering synergistic miRNAs that co-target the same mRNA. Among these methods, Hon et al. [26] analyze multiple miRNAs targeting a single gene, but their analysis solely relied on finding common targets of miRNAs by analyzing the miRNA–mRNA interactions derived from target prediction algorithms. TriplexRNA [27] is a computational framework that integrates miRNA target prediction, structure prediction, molecular dynamics simulations, and mathematical modeling methods to find synergistic miRNAs and their mutual targets. CancerNet [22] detects candidate cooperative miRNA pairs based on the functional similarity of the targets and the proximity of the miRNA targets in the protein-protein interactions for a given cancer. Neither of these methods takes into account the expression profiles of the RNAs. Finally, Zhang et al. [28] apply causal inference methods to find synergistic pairs where miRNA's knockout is simulated using gene expressipon data to discover synergistic miRNAs and in a more recent study [29] extend this work to incorporate the simulation of multiple knockouts of miRNAs.

In this study, we develop miRCoop for detecting synergistic miRNAs and their mutual targets. We focus on a specific type of synergistic relationship: the miRNA pairs which have weak or no repression on the target mRNA but when together induce strong repression of the expression. 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. We focus on this specific case as it is statistically more realistic to find. Additionally, it can uncover more novel RNAs that are important for a disease or condition which cannot be captured by merely checking the miRNAs individual effect.

We cast the problem of finding such triplets 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, yet when taken jointly they are dependent with the mRNA. This case is similar to the example wherein adding sugar to coffee or stirring the coffee alone have weak or no effect on the coffee's sweetness, but the joint presence of the two has a strong effect [30]. To test whether the triplets of RNAs have such relationships, we resort to the kernelized Lancaster three-variable interaction developed by [30]. This test relies on mapping the probability distributions into a reproducing kernel Hilbert space (RKHS), which is a powerful approach in machine learning (reviewed in [31]). The kernel trick helps us to deal with small samples sizes for the expression profiles available, and due to its non-parametric nature, it also avoids any distributional assumptions. In addition to the expression profiles, miRCoop make 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 reveal 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, and 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. Matlab code for identification of the synergistic triplets are provided in `https://github.com/guldenolgun/miRCoop`

2 METHODS

miRCoop relies on miRNA–mRNA target predictions and kernel statistical tests on the expression data. The overall pipeline of miRCoop is summarized in Figure 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 miRNA that can bind to a common target mRNA (Figure 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 [32] and miRanda [33]. When using TargetScan, the 6–mer sites predictions are filtered since they are classified as poorly conserved [32]. 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, at the end of this step, there is still an extensive list of candidate triplets. 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, Figure 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 [34]. 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 the 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.

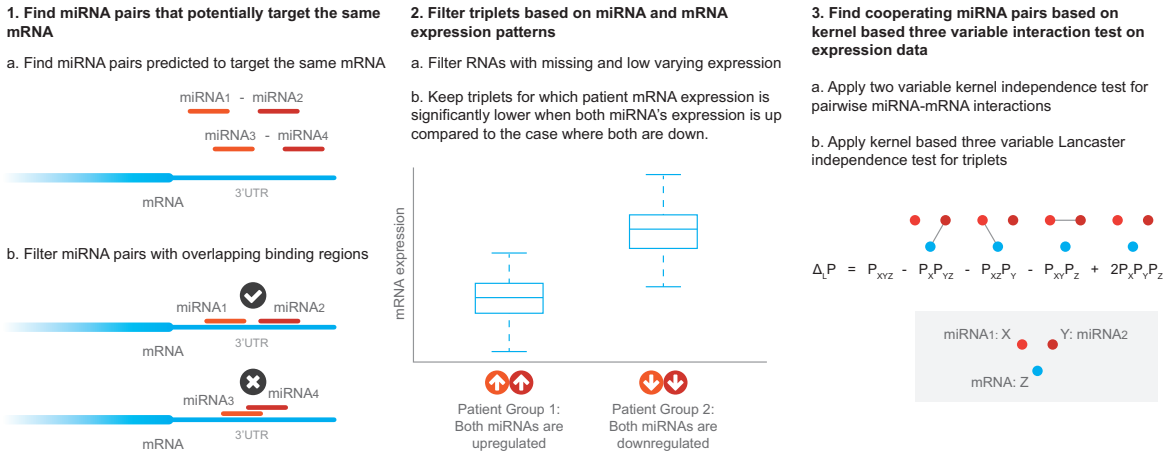


Figure 1: Overview of the methodology. Each box represents a step in the methodology.

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 interactions tests on RNA expression data to find putative miRCoop triplets, Figure 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, 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) [30].

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 [30] and considers only triplets which reject the null hypothesis. Here, we provide a brief description of these kernel-based tests and later detail on how we use this method for discovering miRCoop triplets.

2.3.1 Kernel Three-Variable Lancaster Interaction Test

An interaction measure associated to a multidimensional probability distribution P is a signed measure $\triangle P$ that vanishes whenever P can be factorized as a product of its marginal distributions [35]. Lancaster interaction [35, 30] is such a measure which is the difference between the joint probability distribution and the product of the marginal distributions. When there are two variables, X and Y, the Lancaster interaction is defined as:

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

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

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

A statistical dependency test can be constructed where the Lancaster interaction is the test statistic [35, 30]. The null hypothesis, in this case, is that the Lancaster interaction is zero. The null distribution of the test statistic can be estimated

using permutation-based approach where the indices of the variables are shuffled [30]. Sejdinovic et al. [30] formulates a kernel based three-variable Lancaster interaction test by embedding the probability distributions into reproducing kernel Hilbert spaces (RHKS). In their general formulation, 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. In our case $p = 1$.

The effectiveness of embedding probability distributions are demonstrated in different machine learning and inference task (reviewed in [36] and [31]). The key idea of is to implicitly map probability distributions into potentially infinite dimensional feature spaces using kernel functions [37]. Kernel embedding of the probability distributions avoids the restrictive assumptions about the form of the distributions such as, variables to be discrete-valued, or assumptions about the form of the relationships among the variables such as linear relationship between variables [36]. Using the kernel trick, instead of constructing the new feature space, computations on the probability distributions can be calculated implicitly via feature space operations, that is as Gram matrix operations. The probabilities can be expressed as certain expectations of kernel functions. Sejdinovic et al. [30] arrives at the following test statistic for the two variable and three-variable dependency test:

$$\| \triangle_L \hat{P} \|^2_{k\otimes l} = \| \hat{P}_{XY} - \hat{P}_X \hat{P}_Y \|^2_{k\otimes l} = \frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++} \tag{3}$$

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

Here, k denotes the kernel function defined in domain $\mathcal{X}$ that maps P_X to RHKS. K is a symmetric $n \times n$ Gram matrix for the kernel k computed over the observations' X dimension, n is the number of samples. Similarly, l and m are kernel functions defined in domains of $\mathcal{Y}$ and $\mathcal{Z}$, respectively, while L and M are the Gram matrices computed over observations sampled from domains $\mathcal{Y}$ and $\mathcal{Z}$, respectively. $\circ$ denotes an Hadamard product of matrices. $\otimes$ indicates the Kronecker product. $\mathbf{1}$ denotes an $n \times 1$ vector and $K_{++} = \Sigma_{i=1}^n \Sigma_{j=1}^n K_{ij}$. $\tilde{K}$ denotes the corresponding centered matrix of K, where $\tilde{K} = HKH$ and $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}$ 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 4, the norm of the embeddings can be computed using these centered matrices.

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-Y and X-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:

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|_2^2}{2\sigma^2}\right) \tag{5}$$

, where σ is a non-negative real number that determines the width of the kernel. We set σusing the median heuristic [38]. 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 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 (Figure 1). By repeating this procedure for all samplings $\forall b \in \{1, \ldots, B\}$, we obtain a collection of B many triplet sets, $S^{B}_{b=i}$ 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 asses the statistical significance of the statistic of interest (such as the number of TFs (See Results), 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.

2.5 Data Collection and Processing

We apply miRCoop on patient-derived kidney cancer samples. We retrieve RNASeq gene and miRNA expression data from GDC Data Portal [39] on Sept 2nd 2017. We only use primary tumor solid samples that concurrently include mRNA and miRNA expression data, which contain 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 [40]. 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 [41].

We utilize the available Reads Per Kilobase Million Reads (RPKM) values and Fragments per Kilobase of transcript per Million mapped reads upper quartile normalization (FPKM–UQ) for miRNA and mRNA expression data, respectively. 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 median absolute deviation (MAD) below 0.5. In the final set, we have 12,113 mRNAs and 349 miRNAs in kidney tumor samples.

3 RESULTS

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

We applied the methodology summarized in Figure 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 Figure 2. The target prediction step produce $31,184$ 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. These triplets include 41 different mRNAs, 27 of which are reported to be cancer driver genes for kidney cancer as reported in IntOGen [42]. The complete list of the identified triplets is listed in Supplementary File 1 and the network for the potential miRNA-miRNA pairs are provided in Figure 3.

3.2 Analysis of 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 The Null Model for miRCoop Triplets. Below we describe the biological evidence relevant to the identified triplet set.

3.3 Comparison with Other Methods

We compare our proposed methodology with two other computational methods: CancerNet [22] and TriplexRNA [27]. CancerNet provides cancer-specific pre-computed synergistic pairs of miRNAs.

Figure 2: The number of remaining candidate triplets at the end of each step of the pipeline. The number of available RNAs that participate in the analysis is provided at the beginning.

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; however, we do not find a significant overlap, Figure 4. One reason for the low overlap could be that TriplexRNA triplets are not cancer-specific while the triplets we identified are.

3.4 Triplets Coded Close on the Genome

Recent studies reveal that miRNAs, which regulate similar process, are located nearby on the genome [25]. Others report that proximal pairs of miRNAs are generally co-expressed [43] indicating a functional link. Thus, we investigate whether the identified synergistic pairs are close on the genome.

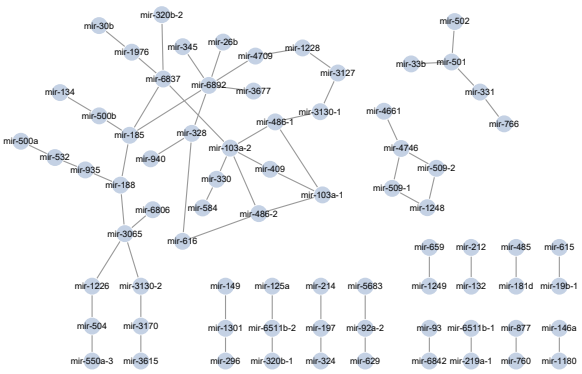


Figure 3: The network shows the miRNA synergistic pairs discovered for kidney cancer. Nodes are miRNAs and edges exist between them if they are participate in at least one miRCoop triplet.

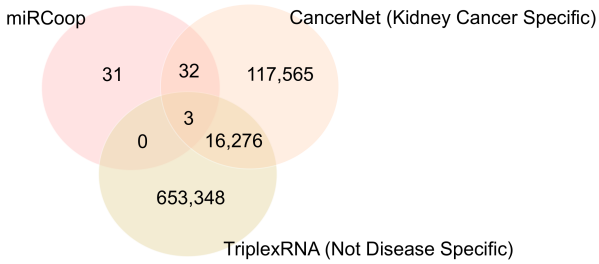


Figure 4: Venn diagram for the number of synergistic miRNAs that are reported by different methods for human. In CancerNet [22] we use the kidney specific triplets. The TriplexRNA [27] database predicts the RNA triplexes constitute of two synergistic miRNAs and their mutual target mRNAs. In TriplexRNA, we utilized predictions that are called as TEC. In TEC, predicted RNA triplexes are closeby within 13-35 nt and their predicted equilibrium concentration (TEC) > 50 nM.

Specifically, 10 kb of their transcription start and end sites within the genome are considered spatially close. We observe that three of such pairs are very close, and these pairs form two different clusters (Table 1). These clustered miRNAs are part of the miR–500 and miR–132 miRNA family. There is evidence reporting that miR–500 family take part in kidney cancer [45, 46, 44]. Genomic locations of these synergistic miR–500 family miRNAs are illustrated in Figure 5.

3.5 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

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.

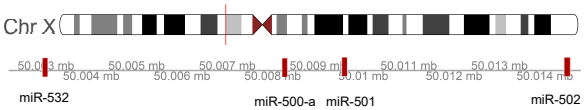


Figure 5: The red marks on the genome axis indicate the miRNA positions of the synergistic miRNAs that belong to the miR-500 family. The figure was created with Gviz package in R [47].

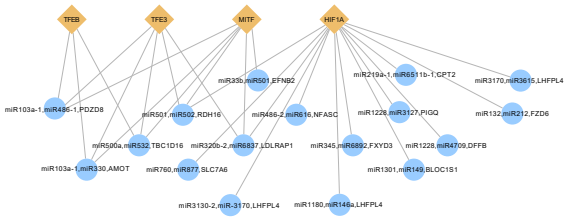


Figure 6: Network for the synergistic triplets regulated by common TFs. Each node denotes a triplet or a TF while an edge represents an interaction between them. TFs are represented by the diamond-shaped yellow symbol, and triplets are represented by the circle-shaped blue symbol.

identified triplets are potentially co-regulated by a common TF. To be able to conduct this analysis, we first curated a set of miRNA–TF and a set of mRNA–TF regulatory interactions.

Since none of the TF–miRNA databases are comprehensive, we use two strategies to curate the set of TF–miRNA regulatory interactions. We first use data from available databases; this includes TransmiR [48], PuTmiR [49] and miRTrans [50] databases. miRTrans database provides TF–miRNA interactions for different tissues and cell types; we only consider the renal cell lines HRE, human Renal 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 [51] and TF motifs. We retrieve the $10,000$ upstream regions of transcription start of the gene in question using the R package GenomicRanges [52] and check if this region overlaps with the ChIP-Seq peaks in three kidney-related cell lines (HRE, HEK293B, HEK293) using the GLANET tool [53]. We compile the TF–miRNA interactions from the databases and discovered through analyzing ChIP-Seq data. For the mRNA–TF regulatory interaction set, we conducted the same analysis on the ChIP-Seq data. Furthermore, we scan the promoter regions with the TFs' binding motifs with JASPAR [54] and Hocomoco v.11 [55] using the FIMO [56] 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 regulate 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 (p-value < 0.04). The list of these TF–triplet relations is provided in Supplementary File 2.

We inspect a subset of the TF–triplet regulations more in detail. Recent studies report that the transcription factors, HIF1A [57], TFEB, TFE3 and MITF [58] 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 Figure 6. Additionally, we observe that the mRNA pairs that are encoded very closely on the genome (see Section Triplets Coded Close on the Genome) 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) [58] regulate the same triplets: (miR103a.1, miR486.1, PDZD8) and (miR103a.1, miR330, AMOT).

3.6 Functional Enrichment Analysis

Synergistic miRNAs are likely to regulate genes with the same or similar functions [24]. 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 components analysis reveals that the network includes 13 connected components. 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 ($p-value0.032$) (Figure 7a) indicating that the cooperative miRNAs are directly or indirectly linked to each other through their targets.

A gene ontology (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 (Figure 7b). Among these, sodium and guanylate related processes are particularly reported as important in kidney cancer [59, 61, 60]. Also, the genes that are already known as the drivers of renal cancer are annotated with these processes. An example is Angiomotin (AMOT). Lv et al., [62], reports that AMOT is an oncogene in renal cancer. AMOT family members is also reported to promote cancer initiation in other cancers and act as a tumor suppressor in some others [64, 63].

Interestingly, the brain related biological processes such as layer formation in the cerebral cortex, synapse organization, embryonic brain development are found to be enriched (Figure 7b). It is reported that brain metastasis is common for renal cancer and in 15 % of cases, the tumor metastasizes to the brain [65]. Additionally, brain metastasis is the main cause of morbidity and mortality in kidney cancer [65]. These findings suggest that mRNAs regulated by the synergistic triplets may participate in the mechanisms that promote brain metastasis.

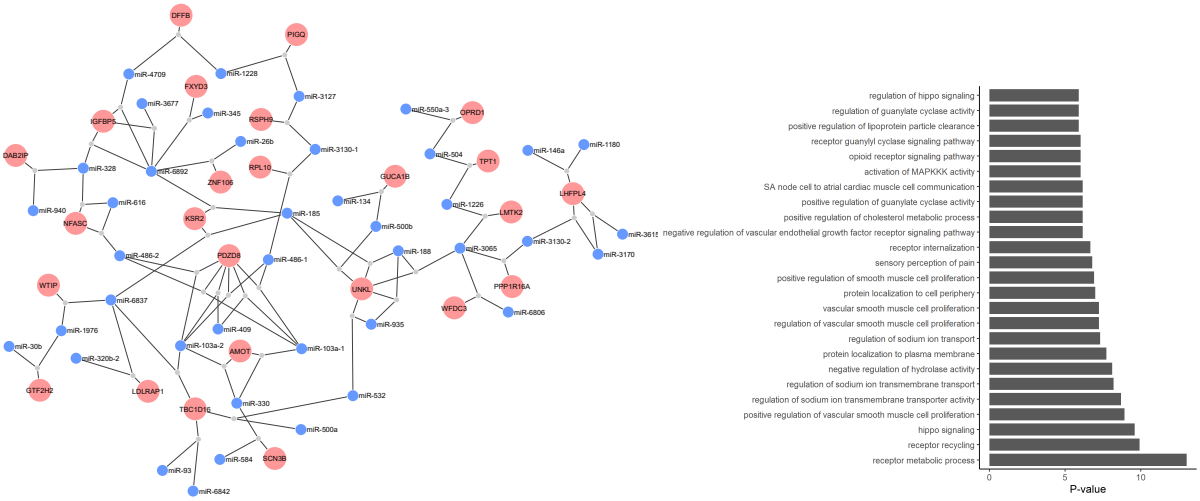
4 DISCUSSION

In this work, we focus on miRNAs that regulate a target gene directly that 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 [5, 28]. 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 regulatory interactions.

Since there are no synergistic interaction datasets which 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.

Although we present results for kidney 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



(a) The largest connected component

(b) Bar chart for enriched GO-terms

Figure 7: 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 found over the mRNA set of the largest connected component in part a).

and mRNAs are available for over a set of samples, and miRNA target predictions are available.

5 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 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 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.

6 ACKNOWLEDGEMENTS

The results shown here are in whole or part based upon data generated by the TCGA Research Network: `https://www.cancer.gov/tcga`. We thank Sabanci and I.D.Bilkent Universities for internal funding support.

References

[1] R. C. Friedman, K. K.-H. Farh, C. B. Burge, and D. P. Bartel, “Most mammalian mrnas

are conserved targets of micrornas," *Genome research*, vol. 19, no. 1, pp. 92–105, 2009.

[2] Y. Peng and C. M. Croce, "The role of micrornas in human cancer," *Signal transduction and targeted therapy*, vol. 1, p. 15004, 2016.

[3] D. P. Bartel, "Micrornas: target recognition and regulatory functions," *cell*, vol. 136, no. 2, pp. 215–233, 2009.

[4] M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, "Widespread changes in protein synthesis induced by micrornas," *nature*, vol. 455, no. 7209, p. 58, 2008.

[5] C. P. Bracken, H. S. Scott, and G. J. Goodall, "A network-biology perspective of microrna function and dysfunction in cancer," *Nature Reviews Genetics*, vol. 17, no. 12, p. 719, 2016.

[6] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks, "Microrna targets in drosophila," *Genome biology*, vol. 5, no. 1, p. R1, 2003.

[7] A. Krek, D. Grün, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. Da Piedade, K. C. Gunsalus, M. Stoffel, *et al.*, "Combinatorial microrna target predictions," *Nature genetics*, vol. 37, no. 5, pp. 495–500, 2005.

[8] X. Lai, U. Schmitz, S. K. Gupta, A. Bhattacharya, M. Kunz, O. Wolkenhauer, and J. Vera, "Computational analysis of target hub gene repression regulated by multiple and cooperative mirnas," *Nucleic acids research*, vol. 40, no. 18, pp. 8818–8834, 2012.

[9] X. Yang, W. W. Du, H. Li, F. Liu, A. Khorshidi, Z. J. Rutnam, and B. B. Yang, "Both mature mir-17-5p and passenger strand mir-17-3p target timp3 and induce prostate tumor growth and invasion," *Nucleic acids research*, vol. 41, no. 21, pp. 9688–9704, 2013.

[10] J. J. Gam, J. Babb, and R. Weiss, "A mixed antagonistic/synergistic mirna repression model enables accurate predictions of multi-input mirna sensor activity," *Nature communications*, vol. 9, no. 1, p. 2430, 2018.

[11] X. Lai, M. Eberhardt, U. Schmitz, and J. Vera, "Systems biology-based investigation of cooperating micrornas as monotherapy or adjuvant therapy in cancer," *Nucleic acids research*, 2019.

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

[13] S. Wang, J. Huang, H. Lyu, C. Lee, J. Tan, J. Wang, and B. Liu, "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*, vol. 4, no. 3, p. e556, 2013.

[14] S. Wu, S. Huang, J. Ding, Y. Zhao, L. Liang, T. Liu, R. Zhan, and X. He, "Multiple micrornas modulate p21cip1/waf1 expression by directly targeting its 3 untranslated region," *Oncogene*, vol. 29, no. 15, p. 2302, 2010.

[15] J. An, K. P. Choi, C. A. Wells, and Y.-P. P. Chen, "Identifying co-regulating microrna groups," *Journal of bioinformatics and computational biology*, vol. 8, no. 01, pp. 99–115, 2010.

[16] A. V. Antonov, S. Dietmann, P. Wong, D. Lutter, and H. W. Mewes, "Geneset2mirna: finding the signature of cooperative mirna activities in the gene lists," *Nucleic acids research*, vol. 37, no. suppl_2, pp. W323–W328, 2009.

[17] G. Boross, K. Orosz, and I. J. Farkas, "Human micrornas co-silence in well-separated groups and have different predicted essentialities," *Bioinformatics*, vol. 25, no. 8, pp. 1063–1069, 2009.

[18] J. Ding, X. Li, and H. Hu, "Microrna modules prefer to bind weak and unconventional target sites," *Bioinformatics*, vol. 31, no. 9, pp. 1366–1374, 2014.

[19] J. Ding, X. Li, and H. Hu, "Ccmir: a computational approach for competitive and cooperative microrna binding prediction," *Bioinformatics*, vol. 34, no. 2, pp. 198–206, 2017.

[20] V. A. Gennarino, G. d'Angelo, G. Dharmalingam, S. Fernandez, G. Russolillo, R. Sanges,

M. Mutarelli, V. Belcastro, A. Ballabio, P. Verde, *et al.*, "Identification of microrna-regulated gene networks by expression analysis of target genes," *Genome research*, vol. 22, no. 6, pp. 1163–1172, 2012.

[21] Y. Li, C. Liang, K.-C. Wong, J. Luo, and Z. Zhang, "Mirsynergy: detecting synergistic mirna regulatory modules by overlapping neighbourhood expansion," *Bioinformatics*, vol. 30, no. 18, pp. 2627–2635, 2014.

[22] X. Meng, J. Wang, C. Yuan, X. Li, Y. Zhou, R. Hofestädt, and M. Chen, "Cancernet: a database for decoding multilevel molecular interactions across diverse cancer types," *Oncogenesis*, vol. 4, no. 12, p. e177, 2015.

[23] R. Shalgi, D. Lieber, M. Oren, and Y. Pilpel, "Global and local architecture of the mammalian microrna–transcription factor regulatory network," *PLoS computational biology*, vol. 3, no. 7, p. e131, 2007.

[24] J. Xu, C.-X. Li, Y.-S. Li, J.-Y. Lv, Y. Ma, T.-T. Shao, L.-D. Xu, Y.-Y. Wang, L. Du, Y.-P. Zhang, *et al.*, "Mirna–mirna synergistic network: construction via co-regulating functional modules and disease mirna topological features," *Nucleic acids research*, vol. 39, no. 3, pp. 825–836, 2010.

[25] X. Yuan, C. Liu, P. Yang, S. He, Q. Liao, S. Kang, and Y. Zhao, "Clustered micrornas' coordination in regulating protein-protein interaction network," *BMC systems biology*, vol. 3, no. 1, p. 65, 2009.

[26] L. S. Hon and Z. Zhang, "The roles of binding site arrangement and combinatorial targeting in microrna repression of gene expression," *Genome biology*, vol. 8, no. 8, p. R166, 2007.

[27] U. Schmitz, X. Lai, F. Winter, O. Wolkenhauer, J. Vera, and S. K. Gupta, "Cooperative gene regulation by microrna pairs and their identification using a computational workflow," *Nucleic acids research*, vol. 42, no. 12, pp. 7539–7552, 2014.

[28] J. Zhang, T. D. Le, L. Liu, J. He, and J. Li, "Identifying mirna synergistic regulatory networks in heterogeneous human data via network motifs," *Molecular bioSystems*, vol. 12, no. 2, pp. 454–463, 2016.

[29] J. Zhang, V. V. H. Pham, L. Liu, T. Xu, B. Truong, J. Li, and T. D. Le, "Identifying mirna synergism using multiple-intervention causal inference," *bioRxiv*, p. 652180, 2019.

[30] D. Sejdinovic, A. Gretton, and W. Bergsma, "A kernel test for three-variable interactions," in *Advances in Neural Information Processing Systems*, pp. 1124–1132, 2013.

[31] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf, *et al.*, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends® in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.

[32] V. Agarwal, G. W. Bell, J.-W. Nam, and D. P. Bartel, "Predicting effective microrna target sites in mammalian mrnas," *elife*, vol. 4, p. e05005, 2015.

[33] D. Betel, A. Koppal, P. Agius, C. Sander, and C. Leslie, "Comprehensive modeling of microrna targets predicts functional non-conserved and non-canonical sites," *Genome biology*, vol. 11, no. 8, p. R90, 2010.

[34] F. Wilcoxon, "Individual comparisons by ranking methods. biom bull 1: 80–83," 1945.

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

[36] L. Song, K. Fukumizu, and A. Gretton, "Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 98–111, 2013.

[37] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*, pp. 13–31, Springer, 2007.

[38] K. Fukumizu, A. Gretton, G. R. Lanckriet, B. Schölkopf, and B. K. Sriperumbudur, "Kernel choice and classifiability for rkhs embeddings of probability distributions," in *Advances in neural information processing systems*, pp. 1750–1758, 2009.

[39] C. G. A. R. Network *et al.*, "Comprehensive molecular characterization of clear cell renal cell carcinoma," *Nature*, vol. 499, no. 7456, pp. 43–49, 2013.

[40] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, *et al.*, "Gencode: the reference human genome annotation for the encode project," *Genome research*, vol. 22, no. 9, pp. 1760–1774, 2012.

[41] S. Durinck, P. T. Spellman, E. Birney, and W. Huber, "Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart," *Nature protocols*, vol. 4, no. 8, p. 1184, 2009.

[42] A. Gonzalez-Perez, C. Perez-Llamas, J. Deu-Pons, D. Tamborero, M. P. Schroeder, A. Jene-Sanz, A. Santos, and N. Lopez-Bigas, "Intogen-mutations identifies cancer drivers across tumor types," *Nature methods*, vol. 10, no. 11, p. 1081, 2013.

[43] S. Baskerville and D. P. Bartel, "Microarray profiling of micrornas reveals frequent coexpression with neighboring mirnas and host genes," *Rna*, vol. 11, no. 3, pp. 241–247, 2005.

[44] M. Fedorko, D. Pacik, R. Wasserbauer, J. Juracek, G. Varga, M. Ghazal, and M. I. Nussir, "Micrornas in the pathogenesis of renal cell carcinoma and their diagnostic and prognostic utility as cancer biomarkers," *The International journal of biological markers*, vol. 31, no. 1, pp. 26–37, 2016.

[45] A. Mangolini, A. Bonon, S. Volinia, G. Lanza, R. Gambari, P. Pinton, G. R. Russo, L. del Senno, L. DellAtti, and G. Aguiari, "Differential expression of microrna501-5p affects the aggressiveness of clear cell renal carcinoma," *FEBS Open Bio*, vol. 4, pp. 952–965, 2014.

[46] X. Shu, M. Hildebrandt, J. Gu, N. M. Tannir, S. Matin, J. A. Karam, C. G. Wood, and X. Wu, "Microrna profiling in clear cell renal cell carcinoma tissues potentially links tumorigenesis and recurrence with obesity," *British journal of cancer*, vol. 116, no. 1, p. 77, 2017.

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

[48] J. Wang, M. Lu, C. Qiu, and Q. Cui, "Transmir: a transcription factor–microrna regulation database," *Nucleic acids research*, vol. 38, no. suppl_1, pp. D119–D122, 2009.

[49] S. Bandyopadhyay and M. Bhattacharyya, "Putmir: a database for extracting neighboring transcription factors of human micrornas," *BMC bioinformatics*, vol. 11, no. 1, p. 190, 2010.

[50] X. Hua, R. Tang, X. Xu, Z. Wang, Q. Xu, L. Chen, E. Wingender, J. Li, C. Zhang, and J. Wang, "mirtrans: a resource of transcriptional regulation on micrornas for human cell lines," *Nucleic acids research*, vol. 46, no. D1, pp. D168–D174, 2017.

[51] E. P. Consortium *et al.*, "An integrated encyclopedia of dna elements in the human genome," *Nature*, vol. 489, no. 7414, p. 57, 2012.

[52] M. Lawrence, W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey, "Software for computing and annotating genomic ranges," *PLoS computational biology*, vol. 9, no. 8, p. e1003118, 2013.

[53] B. Otlu, C. Firtina, S. Keleş, and O. Tastan, "Glanet: genomic loci annotation and enrichment tool," *Bioinformatics*, vol. 33, no. 18, pp. 2818–2828, 2017.

[54] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S. R. Kulkarni, G. Tan, *et al.*, "Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework," *Nucleic acids research*, vol. 46, no. D1, pp. D260–D266, 2017.

[55] I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, *et al.*,

"Hocomoco: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis," *Nucleic acids research*, vol. 46, no. D1, pp. D252–D259, 2017.

[56] C. E. Grant, T. L. Bailey, and W. S. Noble, "Fimo: scanning for occurrences of a given motif," *Bioinformatics*, vol. 27, no. 7, pp. 1017–1018, 2011.

[57] D. J. Sanchez and M. C. Simon, "Transcriptional control of kidney cancer," *Science*, vol. 361, no. 6399, pp. 226–227, 2018.

[58] S. L. Skala, H. Xiao, A. M. Udager, S. M. Dhanasekaran, S. Shukla, Y. Zhang, C. Landau, L. Shao, D. Roulston, L. Wang, *et al.*, "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*, vol. 31, no. 1, p. 179, 2018.

[59] A. Jeppesen, H. Jensen, F. Donskov, N. Marcussen, and H. Von Der Maase, "Hyponatremia as a prognostic and predictive factor in metastatic renal cell carcinoma," *British journal of cancer*, vol. 102, no. 5, p. 867, 2010.

[60] I. Z. Nagy, G. Lustyik, V. Z. Nagy, B. Zarandi, and C. Bertoni-Freddari, "Intracellular na+: K+ ratios in human cancer cells as revealed by energy dispersive x-ray microanalysis.," *The Journal of cell biology*, vol. 90, no. 3, pp. 769–777, 1981.

[61] E. Oosterwijk and R. Gillies, "Targeting ion transport in cancer," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1638, p. 20130107, 2014.

[62] M. Lv, S. Li, C. Luo, X. Zhang, Y. Shen, Y. Sui, F. Wang, X. Wang, J. Yang, P. Liu, *et al.*, "Angiomotin promotes renal epithelial and carcinoma cell proliferation by retaining the nuclear yap," *Oncotarget*, vol. 7, no. 11, p. 12393, 2016.

[63] M. Lv, Y. Shen, J. Yang, S. Li, B. Wang, Z. Chen, P. Li, P. Liu, and J. Yang, "Angiomotin family members: oncogenes or tumor suppressors?," *International journal of biological sciences*, vol. 13, no. 6, p. 772, 2017.

[64] C. Yi, Z. Shen, A. Stemmer-Rachamimov, N. Dawany, S. Troutman, L. C. Showe, Q. Liu, A. Shimono, M. Sudol, L. Holmgren, *et al.*, "The p130 isoform of angiomotin is required for yap-mediated hepatic epithelial cell proliferation and tumorigenesis," *Sci. Signal.*, vol. 6, no. 291, pp. ra77–ra77, 2013.

[65] L. Wyler, C. Napoli, B. Ingold, T. Sulser, M. Heikenwälder, P. Schraml, and H. Moch, "Brain metastasis in renal cancer patients: metastatic pattern, tumour-associated macrophages and chemokine/chemoreceptor expression," *British journal of cancer*, vol. 110, no. 3, p. 686, 2014.