Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in human renal cell carcinoma

R. Atallah

Department of Computer Science & Mathematics, School of Arts & Sciences, Lebanese American University, Byblos, email: roberto.atallah@lau.edu

Abstract

Background: Accumulated studies have revealed that long non-coding RNAs (lncRNAs) play critical roles in human diseases by acting as competing endogenous RNAs (ceRNAs). However, functional roles and regulatory mechanisms of lncRNA-mediated ceRNA in renal cell carcinoma (RCC) remain unknown. In the present study, we aimed to construct the lncRNA-miRNA-mRNA network based on ceRNA theory in RCC by using bioinformatic analyses of public datasets.

Method: This study collected the lncRNAs (GSE16441, GPL6480), mRNAs (GSE16441, GPL6480) and miRNAs (GSE16441, GPL8659) expression data within human tissue samples with RCC group and normal group based on Gene Expression Omnibus (GEO) database. Therefore, differentially expressed genes, and differentially expressed miRNAs were identified. Gene Ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried out. A ceRNA network was constructed.

Results: This study identified six differentially expressed lncRNAs (DELs), 87 differentially expressed miRNAs (DEMs) and 880 differentially expressed mRNAs (DEGs) identified regarding RCC. Furthermore, a lncRNA-miRNA-mRNA network was constructed through three lncRNAs (including two with down-regulation and one with up-regulation), three miRNAs (two with up-regulation and one with down-regulation), as well as 208 mRNAs (including 39 with up- regulation and 169 with down-regulation). Simultaneously, functional enrichment and pathway analyses first on genes of DEGs and second on genes within the as-constructed ceRNA network was conducted. According to our PPI/ceRNA network and functional enrichment analysis results, three key genes were found (BSND, RGMA and THS4). Each of these genes could be regulated by hsa-miR-140-5p, hsa-miR-28-5p , and hsa-miR-532-5p, with LINC00473 interacting with hsa-miR-140-5p, LINC00944 interacting with hsa-miR-532-5p, and LINC00982 interacting with hsa-miR-28-5p.

Conclusion: In conclusion, a ceRNA interaction sub-network was identified as a potential target for treating RCC. The study highlights the potential for targeting the lncRNAs, miRNAs, and mRNAs of this ceRNA network as a therapeutic strategy for RCC treatment.

Keywords: Renal Cell Carcinoma (RCC), competing endogenous RNA, long non-coding RNAs, network, bioinformatics analysis.

Citation:

Atallah, R. (2023). Integrated Analysis of lncRNA-miRNA-mRNA ceRNA network in human renal cell carcinoma.

1 INTRODUCTION

Renal cell carcinoma (RCC) is the most common type of kidney cancer, accounting for approximately 90% of all kidney cancer cases [1]. Despite advances in diagnosis and treatment, RCC still poses a significant health threat, with a 5-year survival rate of less than 10% for advanced stage disease [1].

In recent years, non-coding RNAs (ncRNAs) have emerged as important regulators of gene expression and have been implicated in the pathogenesis of various cancers, including RCC [2]. Among ncRNAs, long non-coding RNAs (lncRNAs) have garnered particular interest due to their involvement in numerous biological processes, such as cell differentiation, proliferation, and apoptosis [3]. Furthermore, lncRNAs are the ncRNAs that are over 200 bp in length and have been shown to play a critical role in the development and progression of RCC, thus they can be targeted for the treatment of the cancer [4].

Recently, competing endogenous RNA (ceRNA) networks have been proposed as a novel regulatory mechanism involving lncRNAs, microRNAs (miRNAs), and messenger RNAs (mRNAs) [5]. In ceRNA networks, lncRNAs can act as sponges for miRNAs, thereby regulating the expression of downstream target mRNAs [6]. Dysregulation of ceRNA networks has been implicated in various diseases, including cancer, like for example in the research conducted by [7] where integrated analysis of a ceRNA network has revealed potential prognostic lncRNAs in gastric cancer. However, the ceRNA mechanisms related to RCC remained unclear and require further investigation.

In this study, I aimed to comprehensively analyze the lncRNA-miRNA-mRNA ceRNA network in human RCC, by using bioinformatics analyses of public datasets. Firstly, this study made use of the Gene Expression Omnibus (GEO) database for obtaining the RCC-related lncRNA, mRNA and miRNA expression data. Therefore, the final DEGs (Differentially expressed genes), DEMs (Differentially expressed miRNAs) and DELs (Differentially expressed lncRNAs) were obtained. Then I used the DEGs to perform functional enrichment analysis to identify key biological pathways and processes involved in RCC. Cytoscape was further utilized to construct a lncRNA-miRNA-mRNA network, followed by the construction of a PPI network to finally come up with key genes expressed differently in RCC. The findings may provide insights into the molecular mechanisms underlying RCC and potential therapeutic targets for this deadly disease.

2 MATERIALS & METHODS

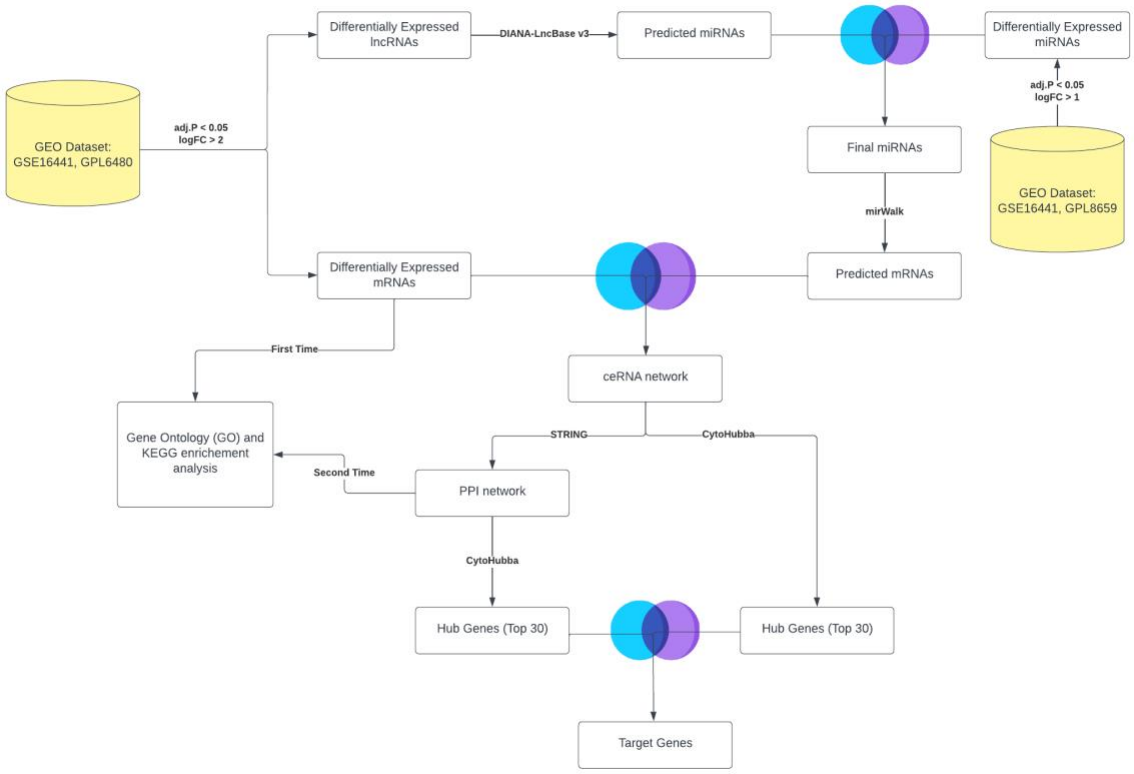

Figure 1: Flow chart of the work done in the paper inspired by the study [9]

2.1 Microarray Data

One dataset was chosen by setting the screening criteria for the species type as "Homo sapiens", from GEO database of National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/geo/), with such keywords being searched as "Renal Cell Carcinoma"**,** "miRNA", "mRNA". Then study type "Expression Profiling by Array" was selected. As a result, I included the dataset GSE16441**,** which is divided into two platforms. Platform GPL6480 (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F), which includes samples GSM413237 to GSM413270, contains the total RNA expression data from 17 RCC tumor samples and 17 corresponding non-tumor samples and was used to find differentially expressed lncRNA and mRNA. Platform GPL8659 (Agilent Human miRNA Microarray Rel12.0) which includes the samples GSM413271 to GSM413304, contains the miRNA microarray expression data from 17 RCC tumors and 17 corresponding non-tumor samples and was used to find differentially expressed microRNAs.

2.2 Differentially Expressed Genes, miRNA and lncRNA (DEGs, DEMs and DELs) analysis

DEGs are analyzed by R package "Linear Models for Microarray Data **(**limma**)**" [26] function for datasets and "GEOquery" R package was used to retrieve to GSE from GEO database. Let me note that the data was already LOWESS

normalized and log2 transformed as indicated in the matrix file, so no further data normalization was performed. For the GPL6480 platform (the total RNA one), log Foldchange > 2 and adjust. $P < 0.05$ were regarded as threshold values for selecting DEGs and DELs. For the GPL8659 platform samples (the micro-RNA one), log Foldchange > 1 and adjust. $P < 0.05$ were regarded as threshold values for selecting DEMs. Statistical significance for the selection of this threshold was found, and those genes that are up- and down- regulated can also be selected for performing the subsequent analysis. R software was also used to draw volcano map of DELs, DEMs and DEGs. This data would be used in the following ceRNA network construction and protein interaction network construction.

2.3 Prediction of miRNA-mRNA and lncRNA-miRNA pairs & Establishment of the ceRNA network of lncRNA-miRNA-mRNA

Interaction between lncRNAs and miRNAs were taken from DIANA-LncBase v3, miRNA-lncRNA targets (https://diana.e-ce.uth.gr/lncbasev3/interactions) [8]. It is a reference repository with experimentally supported miRNA targets on non-coding transcripts. It catalogues approximately ~500,000 entries, corresponding to ~240,000 unique tissue and cell-type specific miRNA-lncRNA interactions. I chose only miRNA with high confidence level. I then intersected the predicted miRNAs with the DEMs found.

Therefore, to find miRNA-mRNA interactions, first I tried the approach used in [9], by taking the mRNA targets of the 3 final miRNAs found by using three prediction databases, TargetScan (http://www.targetscan.org/vert_72/), miRDB (http://mirdb.org/) and mirTarBase (https://maayanlab.cloud/Harmonizome/resource/MiRTarBase), and then taking the targets that were found in 2 or more of these databases. By trying this approach, I only found 367 miRNA-mRNA interactions (in 2 or more databases), however after intersecting these mRNA with already found before differentially expressed genes, the number was reduced to 17 miRNA-mRNA interactions, which is a relatively low number compared to 211 in the [9] paper. Hence this number is not sufficient to do subsequent analysis of 30 hub genes and PPI network since the number of genes is less than 30. Therefore, I used a different approach used by many papers such as [10], by using a single database, the mirWalk database (http://mirwalk.umm.uniheidelberg.de/search_mirnas/) [11], which is a comprehensive database that provides predicted miRNA binding sites of known genes of human, rat, cow, mouse, and dog. The predicted target genes found in this database search were intersected with my DEGs to find final differentially expressed mRNAs, and enough was found to do the subsequent analysis.

The final ceRNA network was visualized using cytoscape.

2.4 PPI network analysis and key gene identification

The plug-in CytoHubba in Cytoscape software is a visualization software that obtains the dense relationship through the degree of connectivity. Those hub genes (top 30) in ceRNA network were identified by CytoHubba. The present study adopted the Search Tool for the Retrieval of interacting Genes/Proteins (STRING; version 11.5) for retrieving protein interactions between the 208 DEGs found earlier. The PPI network was visualized using cytoscape then CytoHubba was used to identify the hub genes (top 30) of PPI network. The key genes were determined by taking the intersection of hub genes in the ceRNA network and PPI network. They were used for follow-up analysis.

2.5 Gene Ontology & Pathway Analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/)) [12] is a resource that offers functional interpretation of plenty of genes derived from genomic research. In present study, DAVID database was used to perform Gene Ontology (GO) [13] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway

analysis [14]. The ontology contains three hierarchies: biological process (BP), cellular component (CC) and molecular function (MF). Pathway analysis is a functional analysis that maps genes to KEGG pathways. Herein, I performed GO analysis and KEGG pathway analysis twice, firstly using the differentially expressed genes that I obtained from the analysis performed on the GEO dataset, and second on the genes involved in the PPI network that was constructed (after performing the intersection with predicted mRNAs). The P value denoted the significance of the GO and pathway term enrichment in the DEGs. "P value <0.05" was set as the cut-off criterion. All of that was performed with an R script.

3 RESULTS

3.1 Differentially Expression Analysis

According to the cut-off criteria mentioned above, volcano plots were obtained. I finally found in the samples of total RNA, 880 differentially expressed mRNAs, having 213 upregulated and 667 downregulated. Among these total RNA, I have 6 DELs, 5 downregulated and 1 upregulated. In the miRNA sample I found 87 DEMs, 55 being upregulated and 32 downregulated.

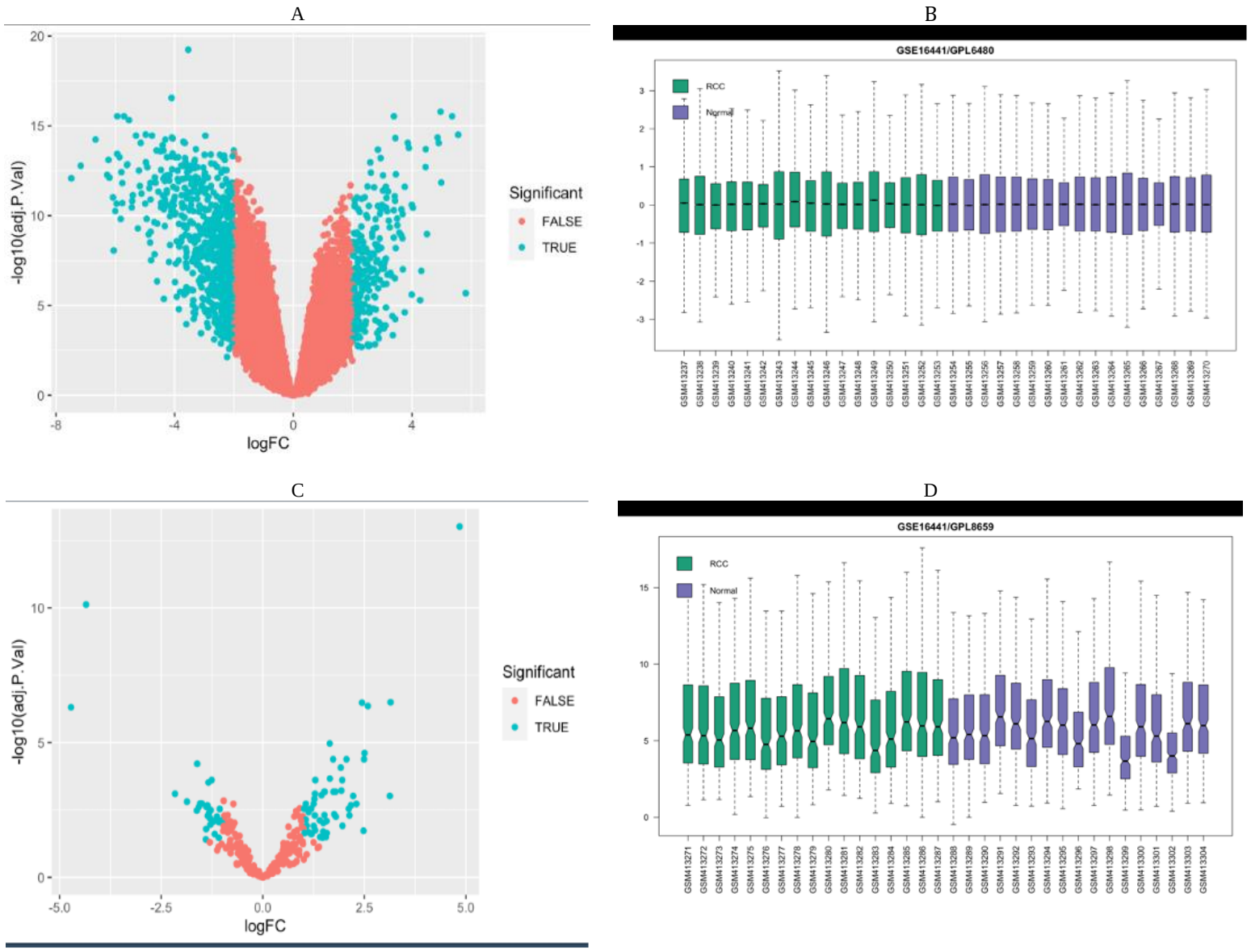

Figure 2: The figures of the differential expression analysis. (A) Volcano Plot for DEmRNAs sample, the blue points are the ones selected (B) Box Plot for the log2FC expression levels across the samples for the DEmRNAs samples (C) Volcano Plot for DEmiRNAs sample, the blue points are the ones selected (D) Box Plot for the log2FC expression levels across the samples for the DEmiRNAs samples.

Table 1: Top 20 significant DEGs

Gene Symbol	Adj. P Value	P Value	logFC
PTGER1	5.78479762139246E-20	1.70151115400684E-24	-3.54335620917647
RALYL	2.80241786002994E-17	1.64857806931581E-21	-4.10677647058824
EGLN3	1.64086249559428E-16	1.4479050199373E-20	4.96318627452941
TSPAN8	2.93122325328935E-16	4.13776805816822E-20	-5.93769281047059

Gene Symbol	Adj. P Value	P Value	logFC
S100A2	2.93122325328935E-16	5.29176575184151E-20	-5.709403268
TMCC1	2.93122325328935E-16	5.99046271465587E-20	3.39041176470588
SFRP1	4.80807493271403E-16	1.13137830053863E-19	-5.54035032682353
SERPINA5	3.05770532104436E-15	8.09440199111691E-19	-4.98003529411765
ANGPTL4	3.12217344171776E-15	9.18340326406778E-19	5.55695882352941
EPN3	3.51094704911591E-15	1.17412390145702E-18	-5.30582352941176
TMEM61	3.51094704911591E-15	1.23923067796314E-18	-4.76675032682353
ADH1A	3.54836417481639E-15	1.35680729079984E-18	-2.96694304388235
VEGFA	4.81313615956337E-15	2.26513849500012E-18	3.43758823529412
DNER	5.01211837211548E-15	2.50620660997597E-18	-4.06641045747059
KNG1	5.758430808406E-15	3.04876035505936E-18	-6.66922037817647
RASL11B	5.8137074558116E-15	3.24902763869699E-18	-4.31840378147059
C1orf116	7.79013203216713E-15	4.58270017775583E-18	-3.70823333335294
ENPP3	8.94669064409515E-15	5.56021677333002E-18	4.90058823529412
VTCN1	8.94669064409515E-15	5.82861845986701E-18	-5.14637647058824
CA10	8.94669064409515E-15	6.28288667607972E-18	-3.659082633

Table 2: All DELs

Gene Symbol	Adj. P Value	P Value	logFC
LINC01606	1.82363426395912E-12	7.02676888577697E-15	-3.30662980858824
LINC00982	3.20294848879567E-12	1.45083259978391E-14	-3.3453488562353
LINC00675	2.41879138270344E-11	1.66479552783283E-13	-4.27302941176471
LINC00864	1.23593668622082E-10	1.24328003614189E-12	-3.66775588235294
LINC00944	1.33653174552592E-06	7.83096428345088E-08	2.04624019611765
LINC00473	7.05529649627344E-06	5.66324611398618E-07	-2.77927478994118

Table 3: Top 20 significant DEMs

miRNA_ID	Adj. P Value	P Value	logFC
hsa-miR-210	9.5753081208223E-14	3.00166398771859E-16	4.84470588235294
hsa-miR-200c	7.54991049062034E-11	4.73348620101589E-13	-4.349281746
hsa-miR-155	3.17280250231403E-07	2.98382680468404E-09	3.1449019607647
hsa-miR-34b*	3.29915935283981E-07	4.13687693146058E-09	2.44196078429412
hsa-miR-34a	4.3713759192237E-07	6.85168639376756E-09	2.58705882352941
hsa-miR-141	4.90072663967156E-07	9.21766766082425E-09	-4.71941176470588
hsa-miR-130b	1.07471426439204E-05	2.35830716324273E-07	1.64801960782353
hsa-miR-142-3p	2.44563915708898E-05	6.13326434379682E-07	2.50502614376471
hsa-miR-886-3p	4.1498373795874E-05	1.31089055204926E-06	2.05882352941177
hsa-miR-93	4.1498373795874E-05	1.41318670258293E-06	1.73058823529412
hsa-miR-142-5p	4.1498373795874E-05	1.43097840675427E-06	2.49176470588235
hsa-miR-532-3p	6.10610814438312E-05	2.29696858095917E-06	-1.61470588235294
hsa-miR-146a	8.46410920676179E-05	3.44932350118819E-06	1.9185294117647
hsa-miR-34a*	0.000216706698320015	9.51063879774363E-06	1.65135834976471
hsa-miR-200a*	0.000245019417914596	1.1521289243633E-05	-1.25705042017647
hsa-miR-224	0.000245547977071059	1.29712913822915E-05	1.95515359476471
hsa-miR-342-5p	0.000245547977071059	1.30856288721254E-05	1.29205882352941
hsa-miR-564	0.000301444594162971	1.70094128367821E-05	-1.33764705882353
hsa-miR-21*	0.000603459627493111	3.5942736433759E-05	1.93379411764706
hsa-miR-126*	0.000669241811180675	4.19587342433025E-05	1.78785294117647

PS: The R code for finding the DEGs and DELs is found in the rscripts folder under the name deg.R

The R code for finding the DEMs is found in the rscripts folder under the name dem.R

3.2 Prediction of miRNA-mRNA and lncRNA-miRNA pairs & Establishment of the ceRNA network of lncRNA-miRNA-mRNA

From the six DELs, only four were found to interact with miRNAs by using the DIANA-LncBase database. These interactions which are a total 64 can be visualized in the figure 3.

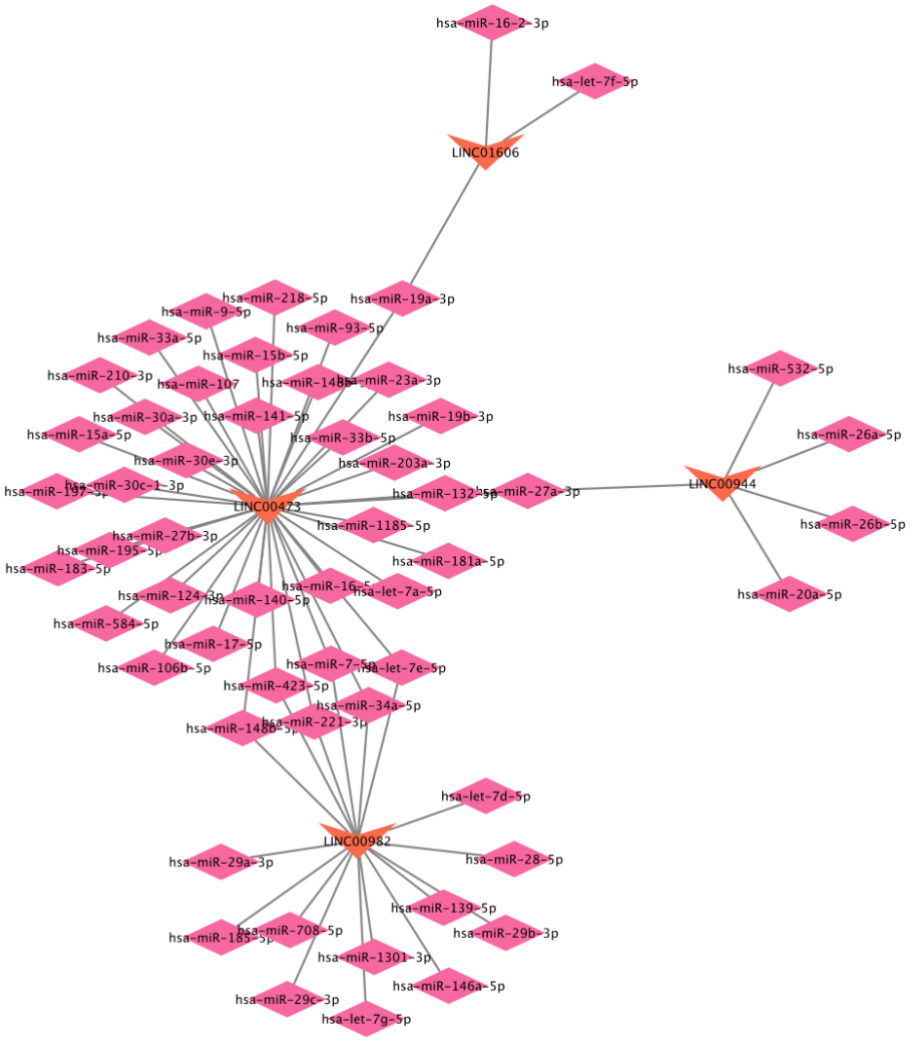

Figure 3: LncRNA-miRNA interactions predicted by DIANA lncBase. The triangle represents lncRNA and the quadrilateral represents miRNA.

However, 56 unique miRNAs were predicted (1 miRNA may interact with many lncRNA), and by taking the intersection of these 56 miRNAs with the DEmiRNAs already found before, we will find only three miRNAs interacting with three LncRNAs: LINC00982 interacts with hsa-miR-28-5p; hsa-miR-532-5p LINC00944 interacts with hsa-miR-532-5p and LINC00473 interacts with hsa-miR-140-5p.

By using the first approach of the three databases, I obtained (Table 4) which are the results of the 17 miRNA-mRNA interactions coming from the three miRNAs.

Table 4: Results of the miRNA-mRNA interactions using mirDB, targetScan and mirTarBase

lncRNA	miRNA	Target	mirDB	targetScan	mirTarBase	Sum
LINC00473	hsa-miR-140-5p	ERC2	1	1	0	2
LINC00473	hsa-miR-140-5p	FGF9	1	1	1	3
LINC00473	hsa-miR-140-5p	VEGFA	1	1	1	3
LINC00473	hsa-miR-140-5p	HDAC4	1	1	1	3
LINC00473	hsa-miR-140-5p	WNK4	1	1	0	2
LINC00473	hsa-miR-140-5p	STC2	1	1	0	2
LINC00473	hsa-miR-140-5p	RALGPS1	1	1	0	2
LINC00473	hsa-miR-140-5p	LMNB1	1	1	0	2
LINC00473	hsa-miR-140-5p	ACSL6	1	1	0	2
LINC00473	hsa-miR-140-5p	SHROOM3	1	1	0	2
LINC00473	hsa-miR-140-5p	RASSF2	1	1	0	2
LINC00473	hsa-miR-140-5p	PDGFRA	1	1	1	3
LINC00982	hsa-miR-28-5p	RNF150	1	1	0	2
LINC00982	hsa-miR-28-5p	SLC16A7	1	1	0	2
LINC00944	hsa-miR-532-5p	MPP7	1	1	0	2
LINC00944	hsa-miR-532-5p	CHL1	1	1	0	2
LINC00944	hsa-miR-532-5p	STC2	1	1	0	2

PS: The R code for extracting the miRNA-mRNA interactions from three databases is found in the rscripts folder under the name code2.R

However, these results were not used and I moved on by using another approach which is the mirWalk database. Then using mirWalk database, these three miRNAs have been found to interact with 4,997 target mRNAs. By doing the intersection of these 4,997 predicted mRNAs with the 880 differentially expressed mRNAs already found before, I found 208 mRNAs common between the two, these were the final DEmRNAs. 267 miRNA-mRNA interactions were selected

(because 1 mRNA can be targeted by many miRNA), and combined with the lncRNA-miRNA interaction, I obtained the ceRNA network in figure 4.

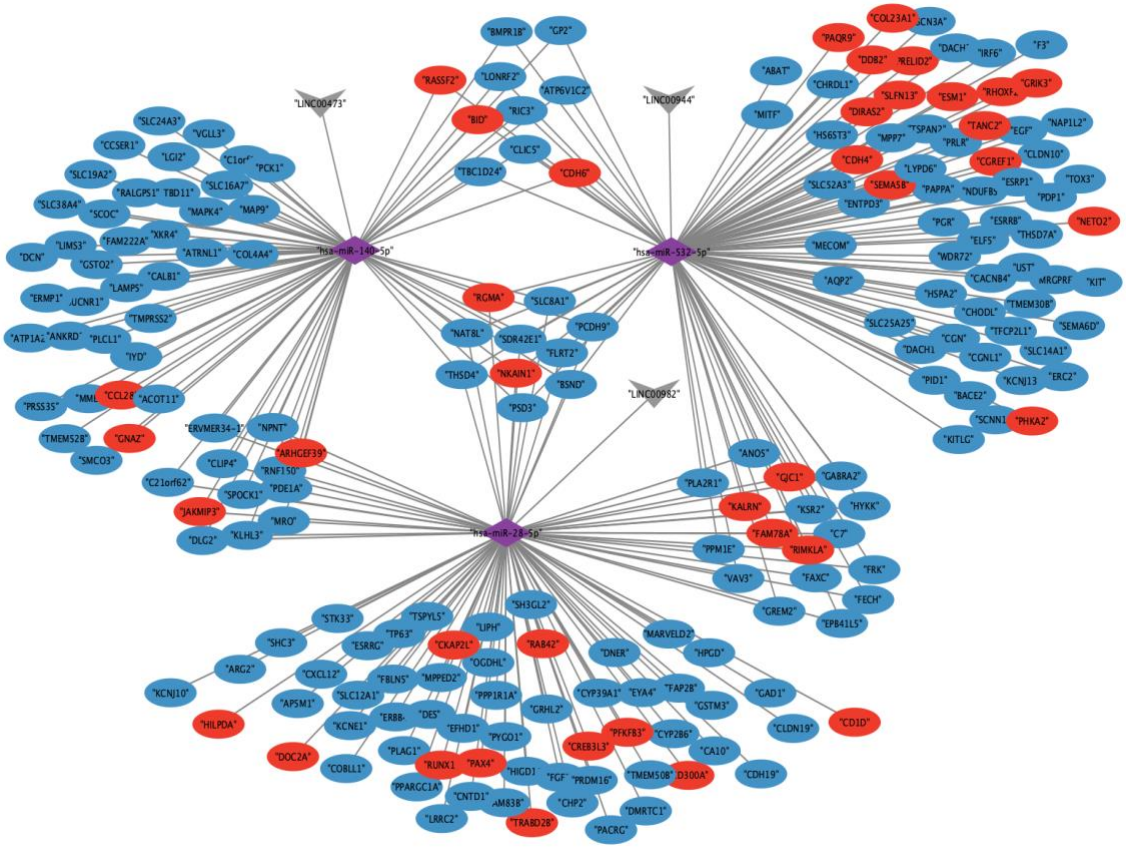

Figure 4: lncRNA-miRNA-mRNA ceRNA network. The triangle, lncRNA, the quadrilateral is miRNA, the oval is mRNA (red = up regulation, blue = down regulation)

PS: The R code for extracting the miRNA-mRNA interactions with mirWalk then using these results to build a table that would be used in cytoscape to build the network is found in the rscripts folder under the name code.R

3.3 PPI network analysis and key gene identification

The top 30 DEGs of ceRNA network acquired based on degree of connectivity using cytoHubba were visualized in figure 5. When uploading these DEGs identified in ceRNA network to STRING website, 188 edges were found and 208 nodes in constructing the PPI network (figure 6). The top 30 DEGs of PPI network obtained by cytoHubba were visualized in figure 7. There were three key genes (BSND, RGMA, THSD4) which were determined by Venn Diagram of ceRNA network and PPI network (figure 8). They key genes were then visualized with cytoscape (figure 9).

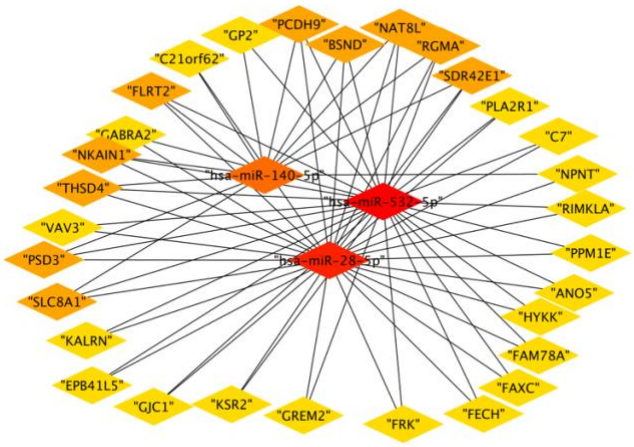

Figure 5: Relationship network diagram of hub genes of ceRNA network

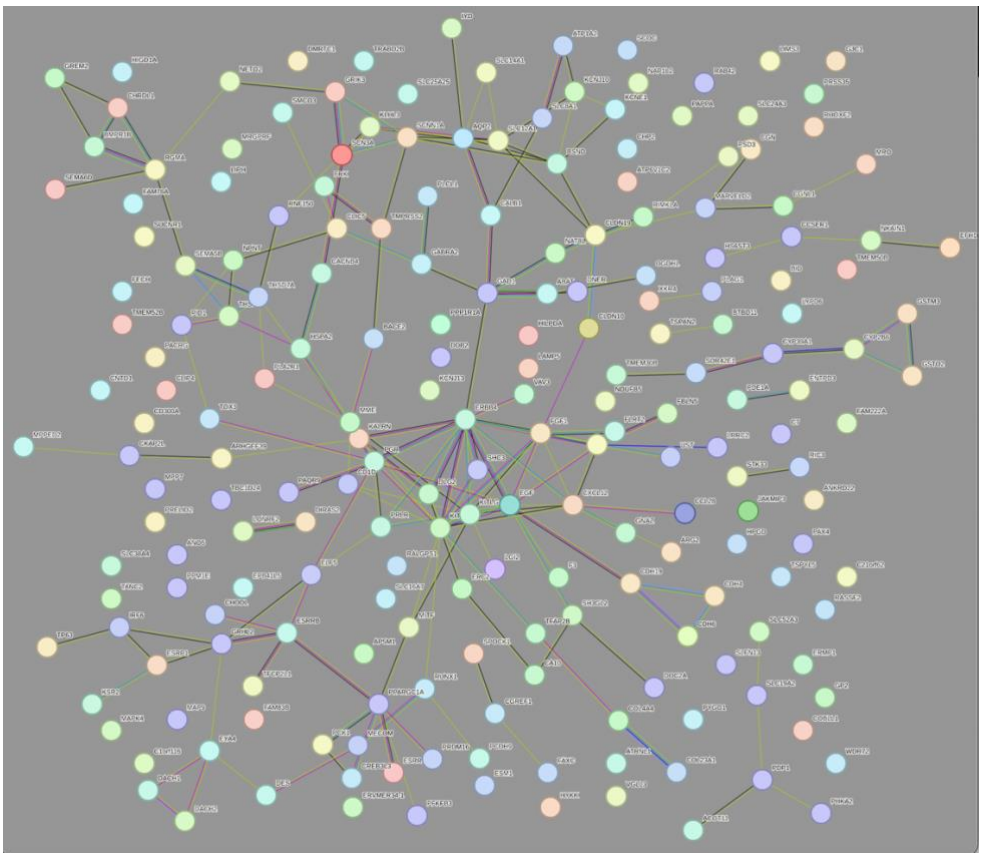

Figure 6: PPI network generated by STRING database

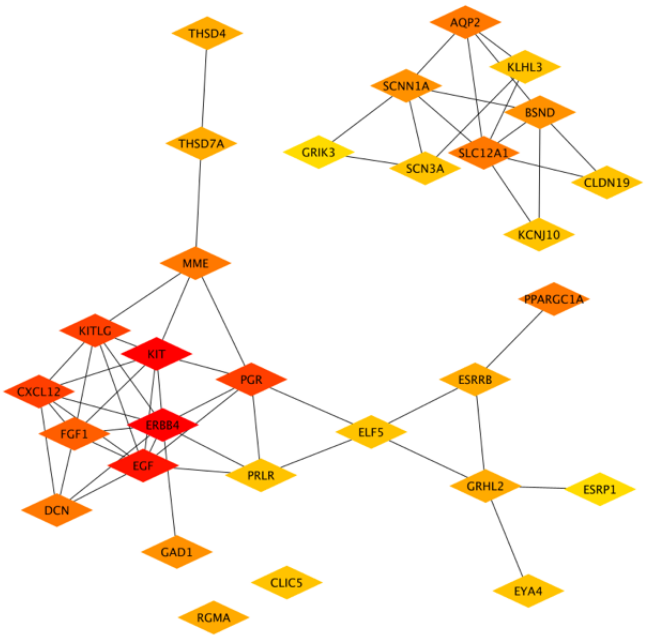

Figure 7: Hub genes from the STRING PPI network, top 30

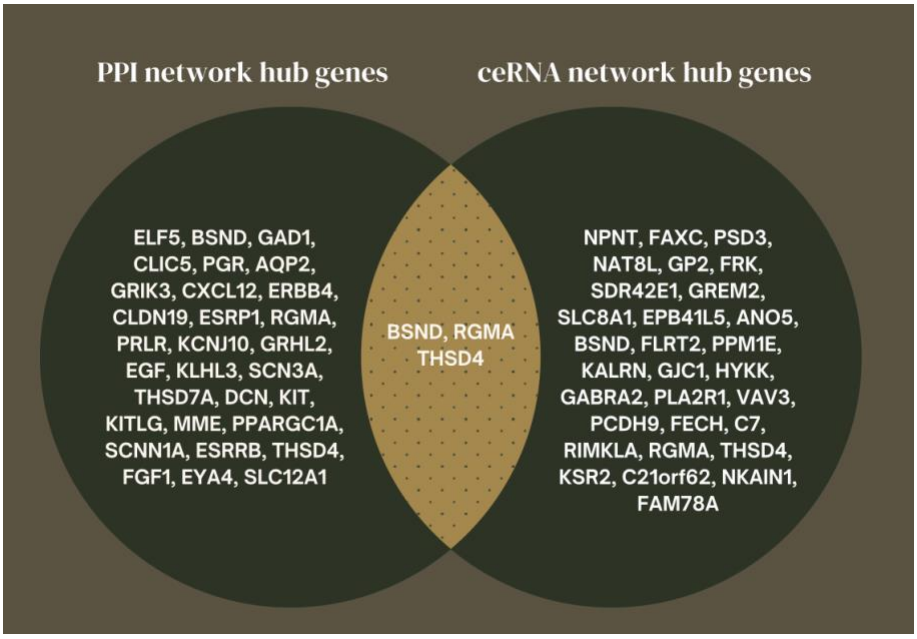

Figure 8: Venn diagram of key genes

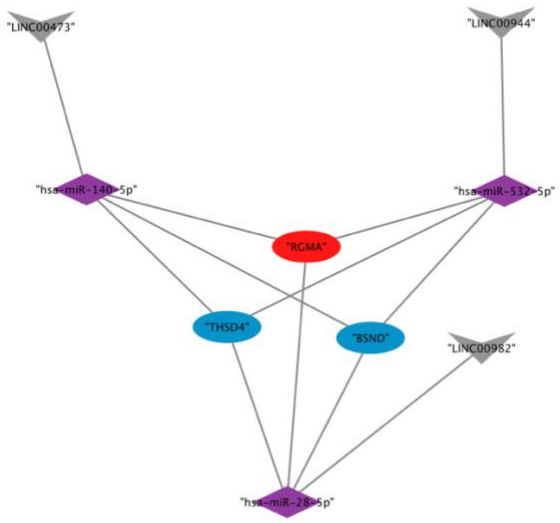

Figure 9: Key genes lncRNA-miRNA-mRNA interaction (sub-ceRNA network)

3.4 Gene Ontology and Pathway Analysis

In the first enrichment analysis that was done on the 880 differentially expressed genes, I found significant results in GO terms and KEGG pathways which will be displayed below in Table 5 and figure 10. Most of them are related to functions in the kidney that might be affected by Renal Cell Carcinoma, like excretion, membrane transports, mineral absorption, etc... Also, I see notice that 43 genes are related in pathways in cancer in the KEGG pathways, and others are related to other metabolic pathways in kidney. In the second enrichment analysis that was done on a reduced number, which is the 208 differentially expressed genes that were found in the PPI network, also results that may be associated with kidney functions and could be affected by renal cell carcinoma, such as plasma membrane components and functions, ions homeostasis and regulation, pathways in cancer, in many more displayed in (Table 6 and figure 11).

Table 5: Top 5 significant GO terms (by Categories) and top 5 KEGG pathways for the initial DEGs (880), ranked by P value

Category	Term	Count	PValue
CC	GO:0005886~plasma membrane	272	2.85940052440415E-13
CC	GO:0070062~extracellular exosome	139	3.50500144225035E-12
CC	GO:0016324~apical plasma membrane	43	2.82321549467074E-11
CC	GO:0016323~basolateral plasma membrane	33	1.10189805239432E-10
CC	GO:0005887~integral component of plasma membrane	97	3.22723797480229E-10
BP	GO:0006883~cellular sodium ion homeostasis	10	5.90998571526695E-08
BP	GO:0055075~potassium ion homeostasis	9	4.34846103479478E-07
BP	GO:0007588~excretion	10	7.31935526037974E-07
BP	GO:0009410~response to xenobiotic stimulus	27	8.45138468612154E-07
BP	GO:1902476~chloride transmembrane transport	16	3.15224234347677E-06
MF	GO:0005509~calcium ion binding	60	4.0016308863007E-08
MF	GO:0008201~heparin binding	24	1.33423262378997E-07
MF	GO:0005201~extracellular matrix structural constituent	20	8.88100755506465E-07
MF	GO:0004957~prostaglandin E receptor activity	4	0.000475402813899799
MF	GO:0020037~heme binding	16	0.000557613665676135
KEGG	hsa04978:Mineral absorption	12	3.07110547268301E-05
KEGG	hsa05200:Pathways in cancer	43	3.84016936401595E-05
KEGG	hsa03320:PPAR signaling pathway	12	0.000250879908947522
KEGG	hsa01100:Metabolic pathways	91	0.000290649640578363
KEGG	hsa04966:Collecting duct acid secretion	7	0.000705991979498007

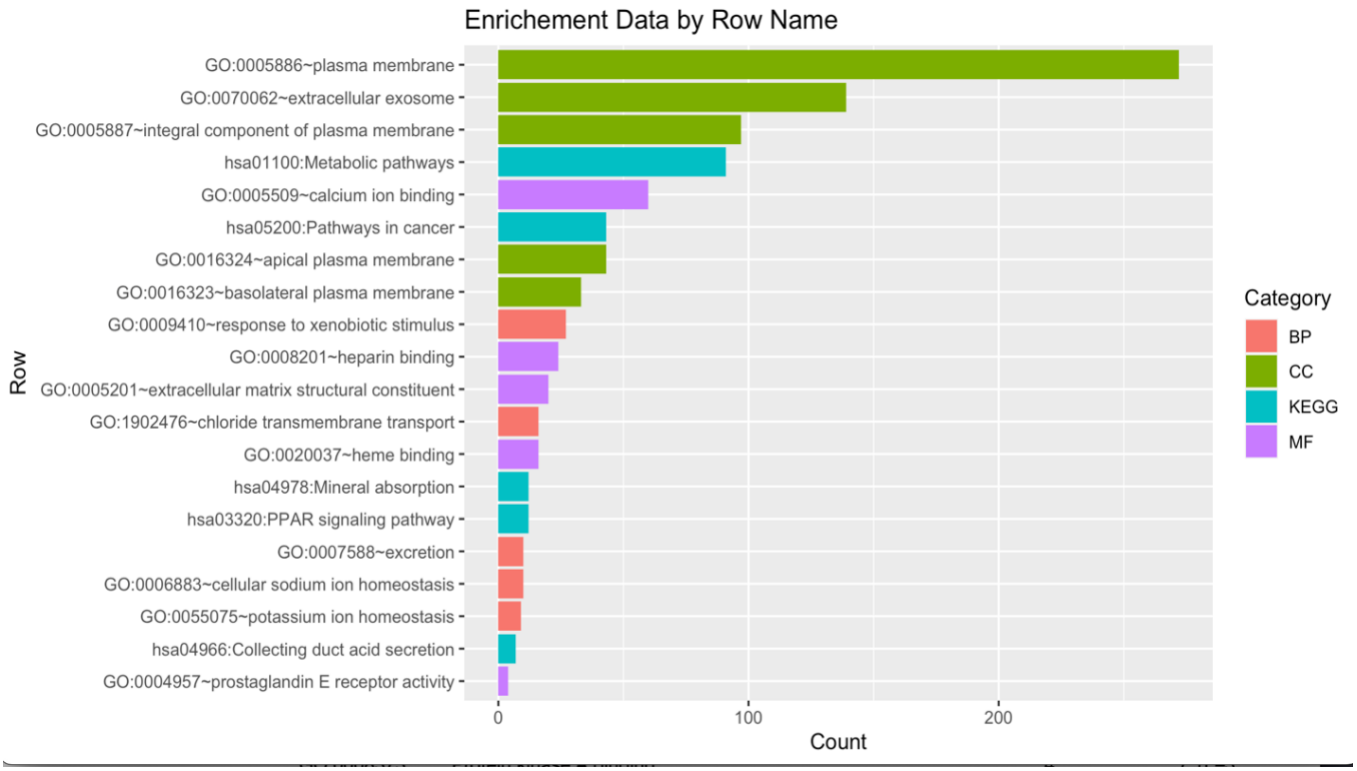

Figure 10: Display of the GO Terms and KEGG pathways by Count for the initial DEGs

Table 6: Top 5 significant GO terms (by Categories) and top 5 KEGG pathways for the PPI DEGs (208), ranked by P value

Category	Term	Count	PValue
CC	GO:0005886~plasma membrane	90	2.74595605282427E-08
CC	GO:0016021~integral component of membrane	81	3.06892450017348E-05
CC	GO:0016323~basolateral plasma membrane	11	0.000174057325935996
CC	GO:0030054~cell junction	10	0.000579422128706444
CC	GO:0005887~integral component of plasma membrane	26	0.00436113608124949
BP	GO:0055075~potassium ion homeostasis	4	0.00126833523478913
BP	GO:0070830~bicellular tight junction assembly	5	0.00159816331918595
BP	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	24	0.00181433897741802
BP	GO:0002070~epithelial cell maturation	3	0.0041372940462609
BP	GO:0035725~sodium ion transmembrane transport	6	0.0050839567526748
MF	GO:0005496~steroid binding	5	0.000229423780558543
MF	GO:0001228~transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	14	0.00118453341484719
MF	GO:0005509~calcium ion binding	18	0.00172893411595322
MF	GO:0005154~epidermal growth factor receptor binding	4	0.00488972829129382
MF	GO:0036122~BMP binding	3	0.0172749406488783
KEGG	hsa04530:Tight junction	8	0.00355806491234326
KEGG	hsa04974:Protein digestion and absorption	6	0.00701197814683411
KEGG	hsa00250:Alanine, aspartate and glutamate metabolism	4	0.00904442564690892
KEGG	hsa05200:Pathways in cancer	13	0.0205383536879084
KEGG	hsa04350:TGF-beta signaling pathway	5	0.024953509260323

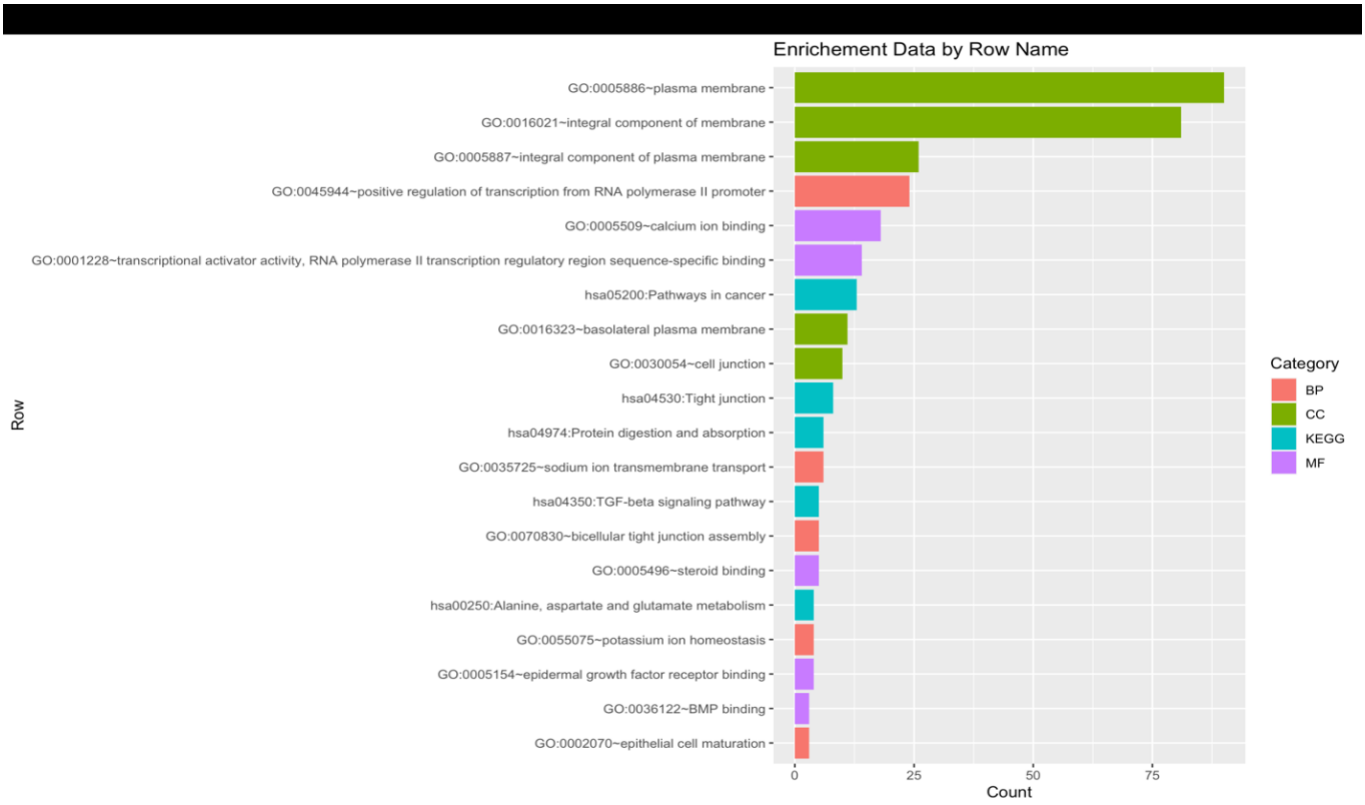

Figure 11: Display of the GO Terms and KEGG pathways by Count for the initial DEGs of the PPI network

PS: The R Code to perform enrichment analysis on the 880 DEGs is found in the rscripts folder under the name enrichment.R

PS: The R Code to perform enrichment analysis on the 208 DEGs extracted from the ceRNA network is found in the rscripts folder under the name enrichment2.R

4 DISCUSSION

Renal cell carcinoma (RCC) is a type of kidney cancer that is responsible for a significant number of cancer-related deaths worldwide [1]. The dysregulation of various genes, including those involved in the regulation of metabolic pathways, membrane transport, and apoptosis, is known to contribute to the development and progression of RCC. In recent years, long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression and have been implicated in the pathogenesis of various cancers, including RCC [2]. lncRNAs are suggested to play roles of ceRNAs for miRNAs; in other words, lncRNAs may play roles of miRNA “decoys” for regulating gene levels [27, 28]. As suggested by the theory, lncRNAs act as natural sponge for the competition adsorption of certain miRNAs and reduction of miRNA binding to corresponding target genes, thus resulting in alterations of miRNA target gene expression [29].

In this study, I investigated the ceRNA (competing endogenous RNA) network in RCC using bioinformatic analysis. Specifically, I focused on the interactions between lncRNAs, miRNAs, and mRNAs and how they contribute to the regulation of gene expression in RCC.

The study identified several lncRNAs that interacted with miRNAs and mRNAs in the RCC ceRNA network, including LINC00473 (down regulated), LINC00944 (up regulated), and LINC00982 (down regulated), as well as miRNAs hsa-miR-140-5p (up regulated), hsa-miR-28-5p (up regulated), and hsa-miR-532-5p (down regulated), which were found to interact with the genes RGMA (up regulated), BSND (down regulated), and THSD4 (down regulated).

This study is in-line with [15] that provided clinicopathological and experimental evidence that lncRNA LINC00944 acts as an oncogene in RCC when up regulated. Also, according to [16], LINC00473 down regulation has been found to play essential roles in the occurrence and development of multiple human cancers, such as lung cancer, gastric cancer, and colorectal cancer. Therefore, the downregulation of LINC00473 has been associated with poor patient outcomes, suggesting that LINC00473 may act as a tumor suppressor, hence the upregulation of LINC00473 may represent a potential therapeutic strategy. Finally, LINC00982 has been found to regulate cell proliferation in gastric cancer [17], breast cancer [18], renal cancer [19] and many others. The down regulation of LINC00982 has been associated with poor prognosis, suggesting that LINC00982 may act also as a tumor suppressor, hence being able to up regulate it will inhibit tumor progression and be also a potential therapeutic strategy.

Same for the miRNAs obtained was found to be dysregulated and involved in many types of cancers including RCC and others, for example [20] studied the upregulation of miR-140-5p in the case of lung cancer, and its expression has been linked to tumor progression, [21] found that miR-28-5p upregulation promoted the progression of ovarian cancer cell cycle, proliferation, migration, and invasion. Finally, [22] revealed that miR-532-5p was one of the most heavily downregulated miRNAs in renal cancer cell. And they proved that overexpression of miR-532-5p inhibited RCC cell proliferation, while knockdown of miR-532-5p promoted cell proliferation.

Considering the genes, The RGMA gene, also known as RGM domain family member A, is a gene that encodes the Repulsive Guidance Molecule A (RGMA) protein. RGMA is a protein that is involved in the regulation of neuronal growth and guidance during embryonic development. In addition to its role in neuronal development, RGMA has also been implicated in the regulation of angiogenesis, which is the formation of new blood vessels from pre-existing ones [23]. RGMA has been identified as a potential oncogene in various types of cancer. Regarding its association with cancer, some studies have suggested that RGMA upregulation may have protumorigenic effects. For example, in colorectal cancer, RGMA upregulation has been associated with increased tumor growth, invasion, and metastasis. It has been proposed that the upregulated RGMA may disrupt the repulsive guidance cues and promote tumor cell migration and invasion [24]. In the context of renal cell carcinoma, the upregulation of RGMA could suggest a potential role for this protein in promoting tumor growth and metastasis. It is possible that RGMA may be involved in the regulation of angiogenesis in the tumor microenvironment, which is an important process for the supply of nutrients and oxygen to the growing tumor. Alternatively, RGMA could be involved in the regulation of cell proliferation or migration, which are also important processes for cancer cell growth and invasion.

Regarding the BSND gene, the downregulation of BSND leads to a dysfunctional or absent barttin protein. This disrupts the normal function of chloride channels in the kidneys and inner ear, resulting in the characteristic symptoms of the syndrome, such as salt wasting, renal tubular defects, and hearing loss.

Currently, there is no strong evidence linking BSND downregulation to the development or progression of renal cell carcinoma (RCC) or other types of cancer. However, it is worth noting that downregulation or altered expression of various

genes, including those involved in ion transport and regulation, can be observed in cancer. These alterations can contribute to the dysregulation of cellular processes and potentially affect tumor development and progression.

THSD4, also known as Thrombospondin Type-1 Domain-Containing Protein 4, is a protein that is involved in various biological processes, including angiogenesis, cell adhesion, and extracellular matrix organization. [25] showed that the expression levels of THSD4 were downregulated in bladder cancer tissues compared with normal tissues.

These results provide insights into the regulatory mechanisms underlying RCC development and progression. The ceRNA network provides a novel perspective on the regulation of gene expression, particularly with the identification of lncRNAs as important players in the process. Additionally, the study highlights the potential for targeting these lncRNAs, miRNAs, and mRNAs as a therapeutic strategy for RCC treatment.

However, it is important to note that these findings are based on bioinformatic analysis, and further experimental validation is required to confirm the interactions between these RNAs and genes in RCC. Nonetheless, the results provide a valuable starting point for further investigation into the mechanisms underlying RCC pathogenesis and the development of new therapeutic strategies for this disease.

5 CONCLUSION

In conclusion, this work successfully constructed a ceRNA network by bioinformatics analysis based on the GEO database, providing a comprehensive resource for investigating the ceRNA regulation in RCC. Overall, these results suggest that the dysregulation of lncRNAs and miRNAs may play an important role in the development and progression of RCC by regulating the expression of target genes. Importantly, candidate prognostic biomarkers that are involved in the ceRNA network were screened (a sub-ceRNA network was created), involving 3 lncRNA (LINC00473, LINC00944, LINC00982), 3 miRNA (hsa-miR-140-5p, hsa-miR-28-5p, hsa-miR-532-5p) and 3 key/hub genes (RGMA, THSD4 and BSND) which may exhibit important roles in the therapeutic target and prognosis analysis in RCC patients.

REFERENCES

[1] Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., Heng, D. Y., Larkin, J., & Ficarra, V. (2017). Renal cell carcinoma. Nature Reviews Disease Primers, 3(1). https://doi.org/10.1038/nrdp.2017.9

[2] Slack, F. J., & Chinnaiyan, A. M. (2019). The Role of Non-coding RNAs in Oncology. Cell, 179(5), 1033–1055. https://doi.org/10.1016/j.cell.2019.10.017

[3] Kung, J. T., Colognori, D., & Lee, J. T. (2013). Long Noncoding RNAs: Past, Present, and Future. Genetics, 193(3), 651–669. https://doi.org/10.1534/genetics.112.146704

[4] Haiyan, S., Luo, G., & Chen, Q. (2021). Long noncoding RNAs as tumorigenic factors and therapeutic targets for renal cell carcinoma. Cancer Cell International, 21(1). https://doi.org/10.1186/s12935-021-01805-2

[5] Ala, U. (2020). Competing Endogenous RNAs, Non-Coding RNAs and Diseases: An Intertwined Story. Cells, 9(7), 1574. https://doi.org/10.3390/cells9071574

[6] Du Z, Sun T, Hacisuleyman E, Fei T, Wang X, Brown M, et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. Nat Commun. (2016) 7:10982. doi:10.1038/ncomms10982

[7] Qi, M., Yu, B., Yu, H., & Li, F. (2020). Integrated analysis of a ceRNA network reveals potential prognostic lncRNAs in gastric cancer. Cancer Medicine, 9(5), 1798–1817. https://doi.org/10.1002/cam4.2760

[8] Karagkouni, D., Paraskevopoulou, M. D., Tastsoglou, S., Skoufos, G., Karavangeli, A., Pierros, V., Zacharopoulou, E., & Hatzigeorgiou, A. G. (2019). DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz1036

[9] Zhang, H., Bian, C., Tu, S., Yin, F., Guo, P., Zhang, J., Song, X., Liu, Q., Chen, C., & Han, Y. (2021). Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in human aortic dissection. BMC Genomics, 22(1). https://doi.org/10.1186/s12864-021-08012-3

[10] Zhu, L., Yang, X., Yao, Z., Wang, Z., Lai, Y., Xu, S., Liu, K., & Zhao, B. (2023). Bioinformatic Analysis of lncRNA Mediated CeRNA Network in Intestinal Ischemia/Reperfusion Injury. Journal of Surgical Research, 284, 280–289. https://doi.org/10.1016/j.jss.2022.11.063

[11] Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. PLOS ONE, 13(10), e0206239. https://doi.org/10.1371/journal.pone.0206239

[12] Dennis, G., Sherman, B., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology, 4(9). https://doi.org/10.1186/gb-2003-4-9-r60

[13] Ashburner, M., Ball, C. G., Blake, J. B., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. P., Hill, D. J., Issel-Tarver, L., Kasarskis, A., Lewis, S. R., Matese, J. C., Richardson, J. D., Ringwald, M., Rubin, G. J., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556

[14] Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2019). New approach for understanding genome variations in KEGG. Nucleic Acids Research, 47(D1), D590–D595. https://doi.org/10.1093/nar/gky962

[15] Chen, C., & Zheng, H. (2021). LncRNA LINC00944 Promotes Tumorigenesis but Suppresses Akt Phosphorylation in Renal Cell Carcinoma. Frontiers in Molecular Biosciences, 8. https://doi.org/10.3389/fmolb.2021.697962

[16] Li, L., Zhang, X., Liu, N., Chen, X., & Peng, C. (2021). LINC00473: A novel oncogenic long noncoding RNA in human cancers. Journal of Cellular Physiology, 236(6), 4174–4183. https://doi.org/10.1002/jcp.30176

[17] Fei, Z., Yu, X., Zhou, M., Su, H., Zheng, Z., & Xie, C. (2016). Upregulated expression of long non-coding RNA LINC00982 regulates cell proliferation and its clinical relevance in patients with gastric cancer. Tumor Biology, 37(2), 1983–1993. https://doi.org/10.1007/s13277-015-3979-9

[18] Chi, F., Qiu, F., Jin, X., Chen, L. Q., He, G., & Han, S. (2022). LINC00982 Inhibits the Proliferation, Migration, and Invasion of Breast Cancer Cells Through the miR-765/DPF3 Axis. DNA And Cell Biology, 41(4), 424–436. https://doi.org/10.1089/dna.2021.0866

[19] Zhang, C., Li, X. Y., Luo, Z. Z., Wu, T. W., & Hu, H. (2019). Upregulation of LINC00982 inhibits cell proliferation and promotes cell apoptosis by regulating the activity of PI3K/AKT signaling pathway in renal cancer. Eur Rev Med Pharmacol Sci, 23(4), 1443-1450.

[20] Flamini, V., Jiang, W. G., & Cui, Y. (2017). Therapeutic Role of MiR-140-5p for the Treatment of Non-small Cell Lung Cancer. Anticancer Research, 37(8). https://doi.org/10.21873/anticanres.11825

[21] Xu, J., Jiang, N., Shi, H., Zhao, S., Yao, S., & Shen, H. (2017). miR-28-5p promotes the development and progression of ovarian cancer through inhibition of N4BP1. International Journal of Oncology, 50(4), 1383–1391. https://doi.org/10.3892/ijo.2017.3915

[22] Zhai, W., Ma, J., Zhu, R., Xu, C., Zhang, J. Z., Chen, Y., Chen, Z., Gong, D., Zheng, J., Chen, C., Li, S., Li, B., Huang, Y., Xue, W., & Zheng, J. (2018). MiR-532-5p suppresses renal cancer cell proliferation by disrupting the ETS1-mediated positive feedback loop with the KRAS-NAP1L1/P-ERK axis. British Journal of Cancer, 119(5), 591–604. https://doi.org/10.1038/s41416-018-0196-5

[23] Fujita, Y., & Yamashita, T. (2017). The roles of RGMa-neogenin signaling in inflammation and angiogenesis. Inflammation and Regeneration, 37(1). https://doi.org/10.1186/s41232-017-0037-6

[24] Chan, S. S. M., Griffith, O. L., Tai, I. T., & Jones, S. J. (2008). Meta-analysis of Colorectal Cancer Gene Expression Profiling Studies Identifies Consistently Reported Candidate Biomarkers. Cancer Epidemiology, Biomarkers & Prevention, 17(3), 543–552. https://doi.org/10.1158/1055-9965.epi-07-2615

[25] Tan, Z., Fu, S. H., Feng, R., Huang, Y., Li, N., Wang, H., & Wang, J. (2022). Identification of Potential Biomarkers for Progression and Prognosis of Bladder Cancer by Comprehensive Bioinformatics Analysis. Journal of Oncology, 2022, 1–17. https://doi.org/10.1155/2022/1802706

[26] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47. https://doi.org/10.1093/nar/gkv007

[27] Huang Y. The novel regulatory role of lncRNA-miRNA-mRNA axis in cardiovascular diseases. J Cell Mol Med. 2018;22(12):5768–75. https://doi.org/10.1111/jcmm.13866.

[28] Botti G, Marra L, Malzone MG, Anniciello A, Botti C, Franco R, et al. LncRNA HOTAIR as prognostic circulating marker and potential therapeutic target in patients with tumor diseases. Curr Drug Targets. 2017;18(1):27–34. https:// doi.org/10.2174/1389450117666151209122950.

[29] Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. Cell Mol Life Sci. 2016;73(13):2491–509. https://doi.org/10.1007/s00018-016-2174-5.