Towards unravelling key genetic differences between Male and Female Clear Cell Renal Cell Carcinoma

Introduction

Kidney cancer is a known cancer type of the urinary system, it is usually originating from renal tubular epithelium, the rate of kidney cancer incidence was largely increased in recent years as the number of newly identified cases in the United States has been increased to 65,000 cases per year which lead to an average of 15,000 deaths annually (according to the most recent statistics report). Clear cell renal Cell Carcinoma (ccRCC) is the most common type of kidney cancers it is a cortical tumor, it represents about 80% of kidney cancer cases besides it associated with worst survival rates when compared with other subtypes of kidney cancer as chromophobe renal cell carcinoma, papillary renal cell carcinoma & collecting duct carcinoma[1]. It is defined by its malignant epithelial cells, clear cytoplasm & complicated growth pattern as well as its arborizing vasculature (arranged in nests).

The name “Clear Cell” indicate the appearance of tumor cells under the microscope (cancer cells look clear). Although some cancer cells have granular eosinophilic cytoplasm. This type of cancer occurs when kidney cells increase in number in uncontrolled manner forming a mass. Up till now, the principal cause of this kind of cancer is unknown, however, certain aggravating factors associated with its incidence such as smoking, obesity, hypertension, the excessive use of certain medication, the inherited disorders and germline mutations (the main inherited disorder is Von Hippel-Lindau “VHL” which happen as a result of a germline mutation in VHL gene in chromosome 3p25)

Treatment usually starts with surgery (commonly considered for tumors of size less than 4cm diameter to remove as much as possible of the tumor mass then it may be followed by either chemotherapy, biological therapy, radiation or targeted therapy. Recently combination strategies are largely recommended for the optimization of ccRCC control although the clinical improvement still with no booming results & about 30% of patients have been suffered from disease recurrence. Many oncogenic signalling pathways involved in this type of cancer (such as MAPK, VEGF& mTOR) have been considered as treatment targets but drug resistance & limited Progression Free Survival (PFS) still challenging, that’s why researchers are still working for more deep understanding of underlying molecular mechanisms of ccRCC progression & metastasis & still novel targets for treatment are urgently needed[1].

Gender-specific analysis done on group of ccRCC patients by different research groups, it was concluded that the incidence rate of ccRCC in male is greater than female, also, it was found that males have larger tumor size, higher tumor grade & lower overall survival than female. Identification of the complete genetic network, hub genes & enrichment analysis of DEGs & DEMS responsible for this gender-based difference will not only expand our comprehensive understanding of the differential genomic changes between male & female but also it will help researchers to choose the optimum therapeutic strategies to be used & to predict the upcoming progressive events.

Methodology

Datasets and pre-processing:

The level 3 RNAseq and miRNASeq data of clear-cell renal cell carcinoma (ccRCC) male and female samples were collected from the TCGA database. Raw count data and read count data were used to represent the mRNA and miRNA expression levels, respectively. There were 611 samples in total, these samples were classified into two cohorts: a male group and a female group, including 352 renal cell carcinomas and 52 normal tissue samples for male cohort, and 186 renal cell carcinomas and 20 normal tissue samples for female cohort. Each sample included the corresponding miRNA-seq and RNAseq data. The data was obtained from TCGA, so there was no need for approval by an Ethics Committee.

Identification of differentially expressed genes and miRNAs

Differentially expressed genes and miRNAs were selected based on their log fold change and adjusted p-values, which were generated by the DESeq package. The inclusion criteria were set as follows: 1) FDRPadj< 0.05 and | log2(fold change)| > 1; and 2) gene and miRNA expression level > 0

Construction of gene co-expression network

Hierarchical clustering was done by flashClust package in R, It is the same as hclust but faster. Because we applied WGCNA to the gene expression data in our analysis, the soft power threshold of β = 16 was selected, which is the smallest value that reaches level 0.9 on the independence scale free topology for female cohort.

Tfmir2 miRNA-gene network

TFmiR2 was used to identify the differentially expressed genes and miRNAs. TFmir2 is a web server used for constructing and analyzing disease-specific co-regulatory networks. The enrichment p-value threshold and ORA p-value threshold was set to 0.05 to test the regulation between the differentially expressed miRNAs and their target genes. Visualization of all networks was performed using Cytoscape 3.7.2

Gene Enrichment analysis:

For gene set enrichment analysis, KEGG pathways and GO functional categories were identified using the DAVID tool. Briefly, we determined which pathways/functional terms were annotated to at least one or two genes and were statistically
overrepresented in the study gene set. GO plot package was used to visualize and annotate modules with gene ontology GO terms

miRNA Enrichment Analysis

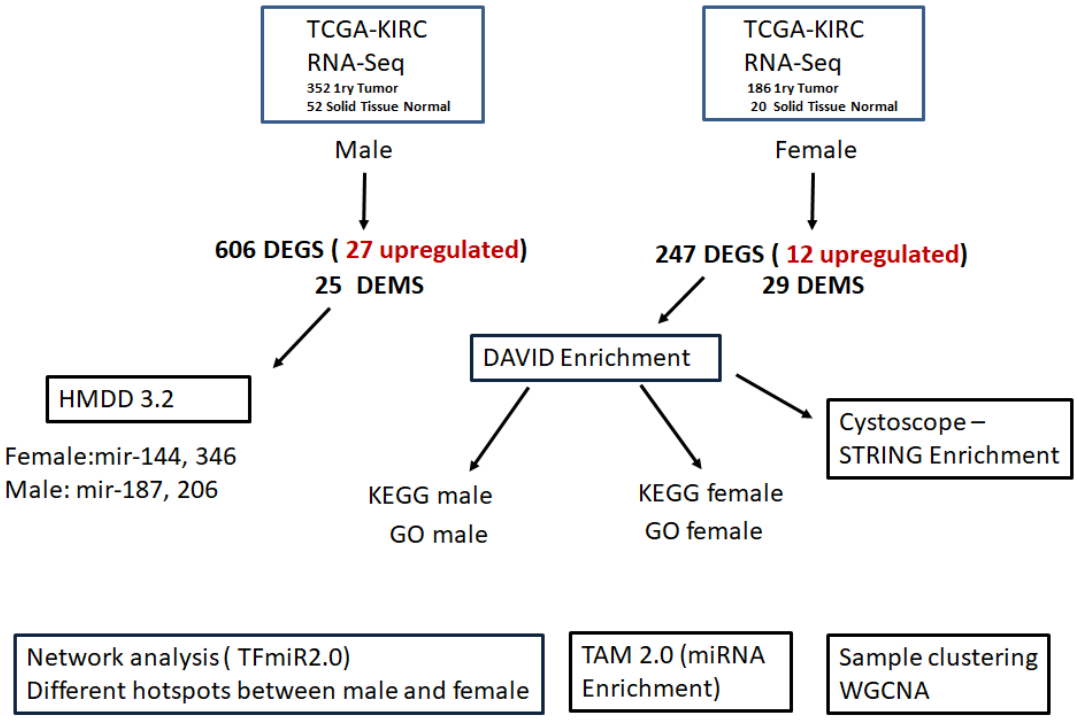
For the enrichment analysis of the miRNAs set, TAM online tool was used. HMDD (human miRNA disease enrichment analysis) was used for miRNA disease target network. For Gene-

Disease Enrichment, DisGenet (database for gene-disease association) and Clinvar Database was used as sources for disease-associated miRNAs and genes

Network Functional Enrichment Analysis: for performing networks functional enrichment analysis, integration and visualization from public databases using Cytoscape Stringapp. The resulting network recognized by STRING, interaction evidence score was 0.4 or greater, functional enrichment was retrieved at p-value 0.05

Results:

Pipeline Summary



Screening of DEGs & DEMs:

A total 404 male patient samples (352 primary tumor samples & 52 solid tissue normal) were obtained from TCGA & total 206 female samples (186 primary tumor & 20 solid tissue normal) were obtained from TCGA as well. After doing integrated analysis from genes of male & genes in female in the RNAseq. data, a total of 606 DEGs from male & 247 DEGS from female where obtained (using |log10 FC ≥1 and adjusted P value <0.05), including 27 upregulated genes and 579 downregulated genes in male & 12 upregulated genes in female & 235 downregulated genes in female in ccRCC samples compared to normal samples**.** Volcano plots were generated to identify the correlation between DEGs (**Figure1)**, also we get 25 DEMs from male & 29 DEMs from female**. (Figure 2)**

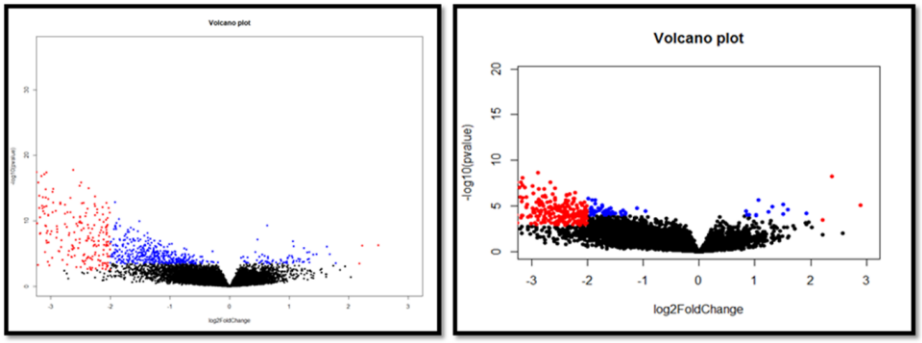


Figure 1: Volcano Plot: Tumor vs Normal (DEGs)

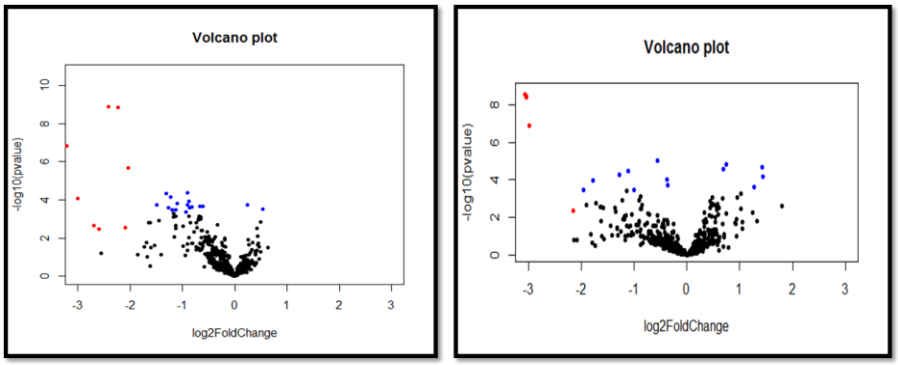


Figure 2: Volcano Plot: Tumor vs Normal (DEMs)

GO term enrichment analysis of DEGs:

DAVID website was used to perform GO enrichment analysis to DEGS of male & female (the analysis includes: BP, CC & MF group).

For male DEGS: the male DEGs were mainly enriched in multicellular organismal process, single-multicellular organism process, regulation of biological quality, single-organism transport, response to stimulus, single-organism transport, single-organism localization, single-organism process, establishment of localization, multicellular organism development, single-organism cellular process. single-organism, developmental process, biological regulation. Down regulated DEGs are mainly enriched in plasma membrane, integral to membrane, endoplasmic reticulum & cytoskeleton, cell fraction, extracellular region & cytosol while up regulated genes were mainly enriched in plasma membrane, nucleus, Golgi membrane, cytosol, extracellular region & extracellular space. No significant differences were remarked between up & down regulated DEGs in their enrichment at MF, as both of them were mainly enriched in protease binding, nicotinic acid receptor activity, L-glutamate transmembrane transporter activity, protein binding, actin binding, voltage-gated potassium channel activity, signal transducer activity & calcium ion binding & voltage gated K channel activity.

For female DEGs: there is no marked differences between up regulated & down regulated female genes in GO enrichment analysis , female DEGs were mainly enriched in fatty acid metabolic process, microtubule-based process, carbohydrate transport, aromatic amino acid family metabolic process, intracellular signaling cascade, ectoderm development, M phase of mitotic cell cycle, ion transport, response to hypoxia, immune system development, cell adhesion, transmembrane transport & regarding for CC female DEGs were mainly enriched in lysosomal membrane, plasma membrane, nucleus, extracellular region, extracellular space, integral component of membrane & endoplasmic reticulum membrane & regarding for MF analysis of female DEGS it was found that they are enriched in transporter activity, protein binding, ATP binding, hydrogen-exporting ATPase activity, RNA binding, DNA binding & GTPase activity.

Pathway enrichment analysis:

The screened DEGs analyzed using the KEGG pathway- for the full interaction network - (Considering only P values less than 0.05). Male DEGs were enriched in Neuroactive ligand-receptor interaction, Retinol metabolism, complement & coagulation cascade, Retinol metabolism,Xenobiotics metabolism - cytochrome P450, Wnt signaling pathway, cell adhesion molecule, chemical carcinogenesis & Calcium signaling pathway as shown in Figure 3. While in female the most enriched pathways were Collecting duct acid secretion, Retinol metabolism, pentose & glucuronate interconversions Xenobiotics metabolism - cytochrome P450, epithelial cell .signaling, receptor interaction & chemical carcinogenesis as shown in (Figure 3), From the results obtained it can be deduced that the most important pathways in male were Wnt signaling pathway, complement & coagulation cascade & metabolism of xenobiotics. Whereas the most important pathways in female were Collecting duct acid secretion & Retinol metabolism.

Figure 3: left: male: right female KEGG pathways

By performing enrichment analysis of the resulted pathways using Cytoscape STRING to figure out the correlations between genes & the corresponding pathways, it was identified that ADH1C, VTN, PROC, CYP2B6 , ADH4 & KNG1 genes were involved in the most important KEGG pathways in male (**Figure 4**) & ADH1C, ACPP, KCN13, SFRP1, CYP2C9, SLC13A1genes were involved in more than one pathway in female **(Figure 5)**

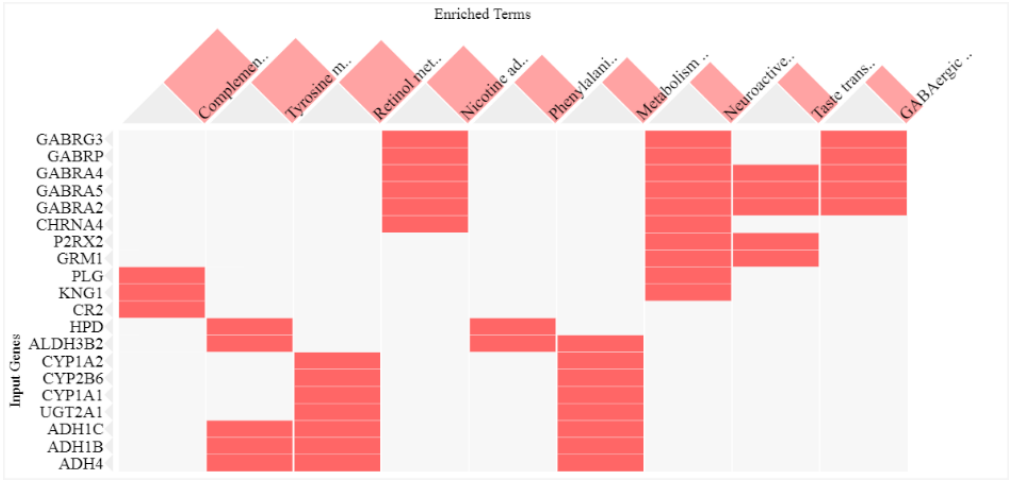


Figure 4

KEGG pathway Enrichment in Male

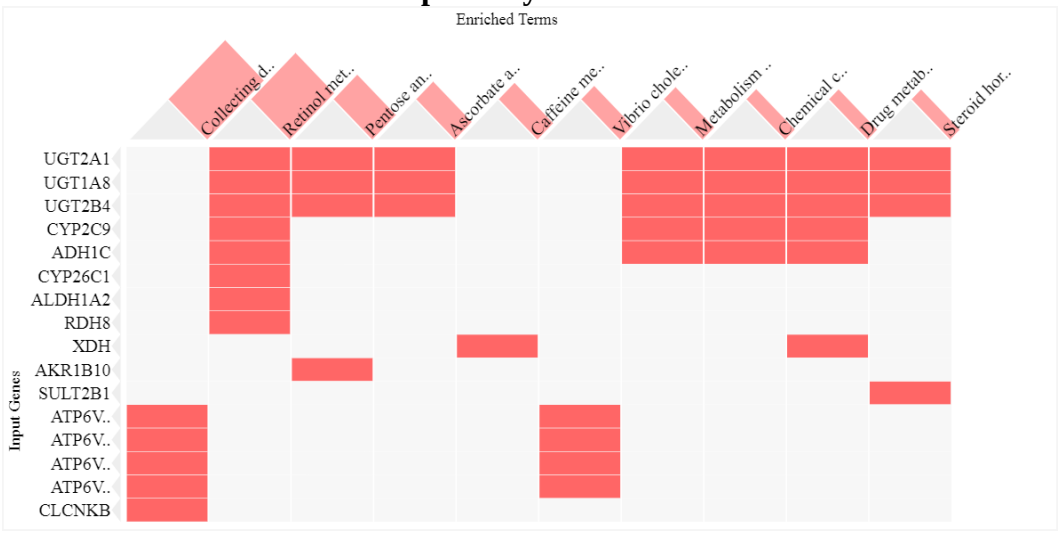
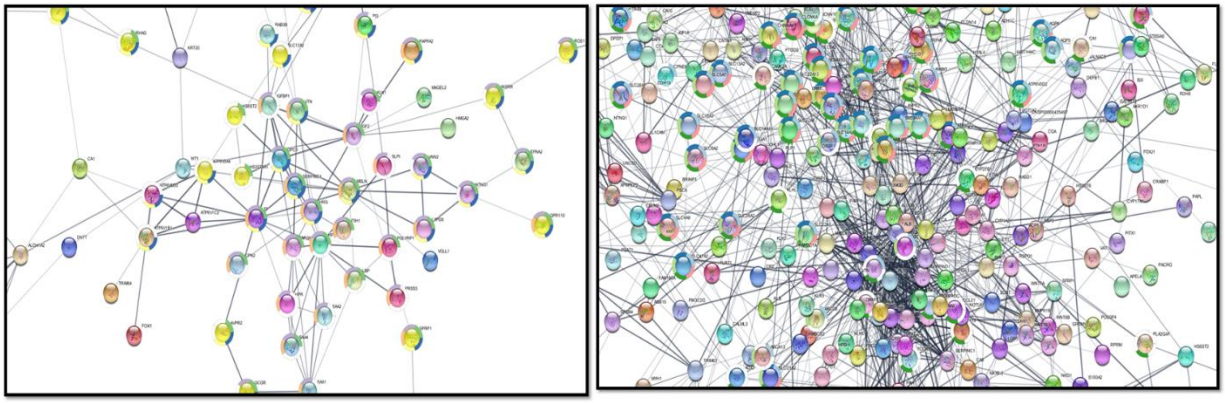


Figure 5 KEGG pathway Enrichment in Female

Network construction of DEGs: The Key genes of DEGs (either male or female) & gene interactions can be identified by using STRING online database, then the results are filtered using Cytoscape software **(Figure 6).**



Female Cytoscape STRING Enrichment

Male Cytoscape STRING Enrichment

Figure 6: Cytoscape STRING Enrichment

Enrichment analysis of DEGs & DEMs Using TFmir2: Enrichment analysis for male & female DEMs against DEGs of each one was performed to combine all TF, miRNA & gene co-regulatory interaction. We get 229 nodes & 862 edges from the full interaction network & for Disease interaction network we get 74 nodes & 150 edges from male DEGs &DEMs analysis **(Figure7)** after using ORA analysis (of the disease interaction network) we get KEGG pathways in which male (DEGs & DEMs) are enriched in, downloading the data using DAVID website & visualizing it in R Studio we get the most important KEGG pathways (with P value <0.05)

These pathways are: Cell adhesion molecules (CAMs),**Hippo signaling pathway**, **Proteoglycans in cancer**,Complement and coagulation cascades,: **Pathways in cancer**, Melanogenesis, **Basal cell carcinoma**, Hepatitis C, **Wnt signaling pathway**, PIK3 -AKT signaling pathway & Signaling pathways regulating pluripotency of stem cells as shown in **Table (1)**. By repeating the same steps for female we get only 6 nodes &4edges from the full interaction network **(Figure 8)** & **we didn't get disease interaction network in Female**

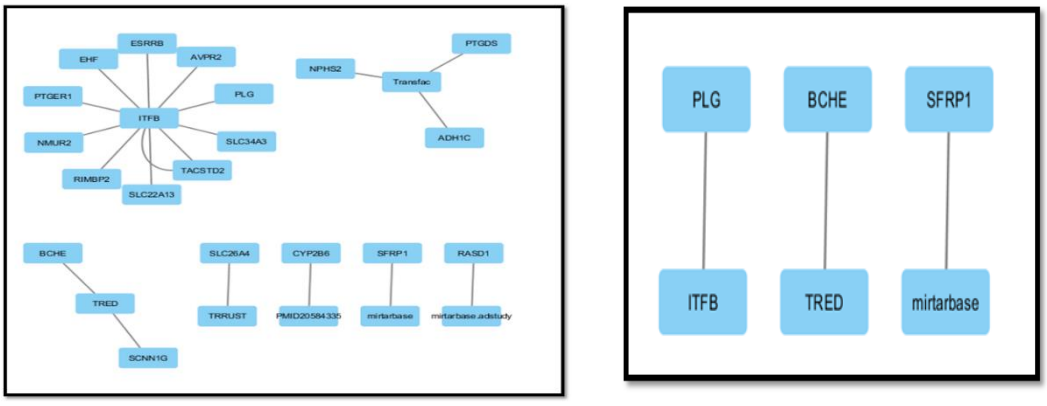


Figure 7: Male **(left)** Complete Interaction Network: **(Right)** Disease Interaction Network

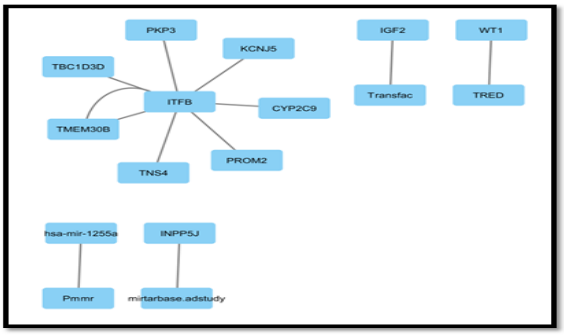


Figure 8: Female Complete Interaction Network

Category	Term	Count	P value
KEGG Pathway	**hsa05217:Basal cell carcinoma**	**4**	**3.68E-03**
KEGG Pathway	**hsa05205:Proteoglycans in cancer**	**6**	**5.68E-03**
KEGG Pathway	**hsa05200:Pathways in cancer**	**9**	**1.61E-03**
KEGG Pathway	**hsa05160:Hepatitis C**	**5**	**7.12E-03**
KEGG Pathway	**hsa04916:Melanogenesis**	**5**	**2.57E-03**
KEGG Pathway	**hsa04610:Complement and coagulation cascades**	**5**	**6.45E-04**
KEGG Pathway	**hsa04550:Signaling pathways regulating pluripotency of stem cells**	**5**	**8.51E-03**
KEGG Pathway	**hsa04530:Tight junction**	**4**	**1.38E-02**
KEGG Pathway	**hsa04514:Cell adhesion molecules (CAMs)**	**8**	**1.45E-05**
KEGG Pathway	**hsa04390:Hippo signaling pathway**	**7**	**2.10E-04**
KEGG Pathway	**hsa04310:Wnt signaling pathway**	**5**	**8.10E-03**
KEGG Pathway	**hsa04151:PI3K-Akt signaling pathway**	**6**	**4.80E-02**

Table (1)
(Top KEGG pathways enriched in male disease interaction network by tfmir2)

As well, hub genes were identified from degree hotspots in male & female from the **TFmiR 2.0,** it was found that the hub genes in male are: KNG1, ALB, AHSG, SERPINC1, APOA2, NOTUM, PROC, SERPIND1, CHRDL1, CHGB, ORM1, ITIH2, EGF, AMELX, NMUR2, CASR, PLG, GRM1, AVPR2, IGF2, HRG, PTGER1, IL6, EGF, PLG, LSAMP, KIT, SFRP1, BMPR1B, ASB15.

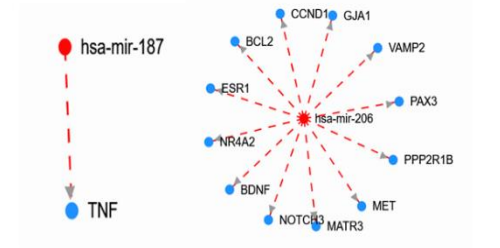
The only hub gene for female: TRIM63.

Hub genes: it was reported that male DEGs has many hub genes which are enriched in different types of cancers & those genes are absent in female, Also some of these hub genes can promote oncogenic pathways indicating its role in cancer. this is shown in details in **(Table 2)**

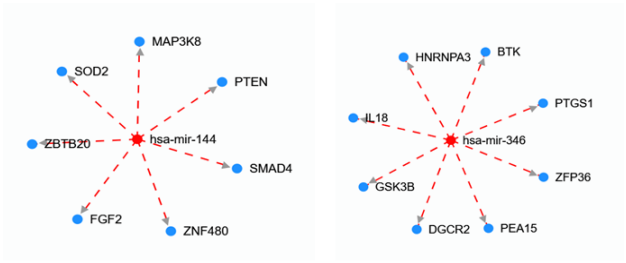
Hub gene	Expression in Cancer	Hub gene	Expression in Cancer
KNG1	Prognostic marker in liver cancer highly expressed in thyroid & breast cancer cells	**EGF**	Key gene in epithelial malignancies
BMPR1B	Enriched in breast and prostate cancers, prognostic in endometrial cancer	**AHSG**	Enriched in liver cancer
SFRP1	Prognostic marker in head and neck cancer	**SERPINC1**	Prognostic marker in liver cancer
KIT (proto oncogene tyrosine kinase)	**Prognostic marker in renal cancer**	**APOA2**	Enriched in liver cancer
LSAMP	Enriched in Glioma	**NOTUM**	Biomarker for Wnt driven tumors
IL6	uncommon **prognostic marker in renal cancer**	**SERPIND1**	Prognostic biomarker in endometrial cancer & enriched in liver cancer
PTGER1	Enriched in Ovarian Cancer	**CHGB**	Prognostic marker in head & neck cancer & pancreatic cancer
HRG	Enriched in Liver Cancer	**ORMI**	Prognostic marker in **renal cancer**
EGF	deregulation leads to progression of some cancers	**GRM1**	weak association with Breast Cancer
IGF2	Prognostic marker in Liver cancer	**ITIH2**	Prognostic marker in liver cancer
PROC	Enriched in Liver Cancer	**PLG**	**Prognostic marker** in **renal Cancer,** enriched in Liver Cancer

(Table 2)
Hub genes identified in male ccKIRC

DEMs analysis using HMDD:



Downregulated miRNA in male with tumor suppressor activity



Upregulated miRNA in Female with tumor suppressor activity

miRNA type	Gender	Target downregulated cancer genes/markers
mir-206	Male	VAMP2, MET, NOTCH3, PPP2R1B, ESR1
mir-187	Male	TNF
mir-144	Female	PTEN,SMAD4, MAP3K8
mir-346	Female	PTGS1, GSK3B

Table 3: Unique finding in miRNA analysis revealed by HMDD Database

Male cohort miRNA analysis:
Down regulation of mir-206 was found to play vital role in ccRCC development. Up regulation of mir-206 will inhibit renal cancer cell proliferation, invasion and migration. Consequently miR-206 is functioned as a novel cell cycle regulator and tumor suppressor in ccRCC and could be considered as a potential target for ccRCC therapy. In case of mir-187, it was found to be down-regulated in clear cell renal cell carcinoma and associated with lower survival, inhibition of cell growth and migration through targeting B7-H3.

Female cohort miRNA analysis:
Through HMDD, mir-144 was found to be enriched in construction and comprehensive analysis of dysregulated long non-coding RNA-associated competing endogenous RNA network in clear cell renal cell carcinoma. The upregulated mir-144 was identified in the differential analysis and used to be biomarker for ccrcc

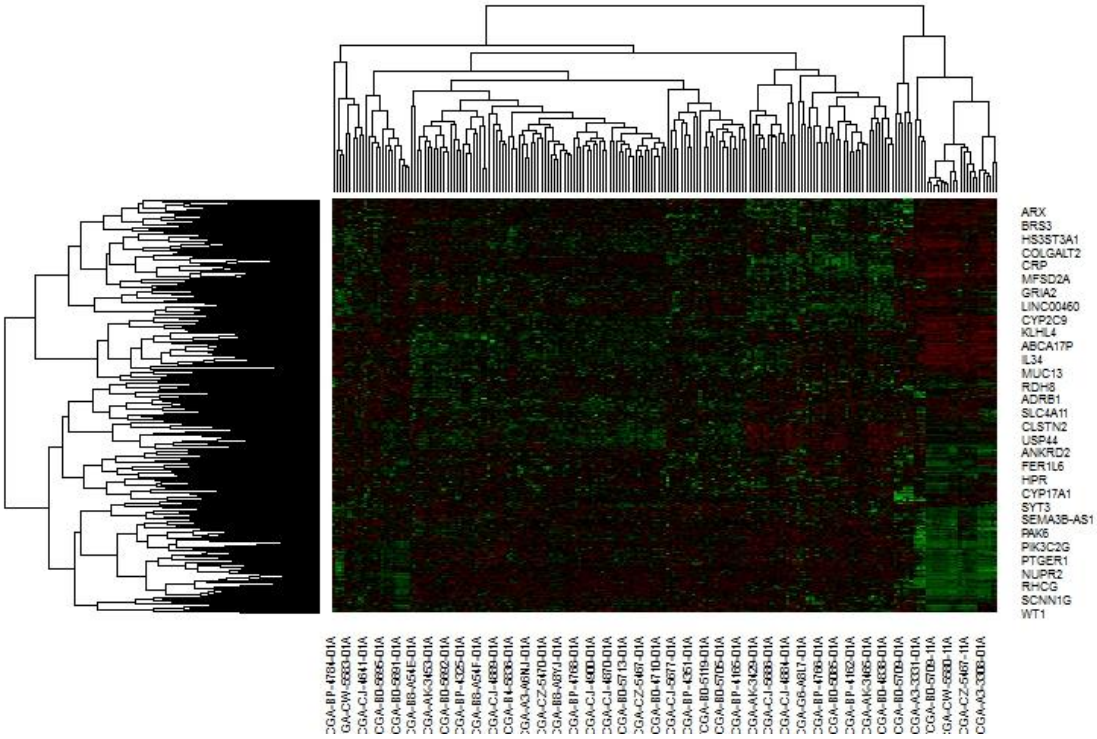
Keymark finding:

EVX1 has an interesting unique nature, it was found to be the only upregulated differentially expressed gene and at the same time was recognized as a hot spot according to TFmiR 2.0

network analysis. Expression of EVX1 gets intensified by the presence of BMP and WNT signaling pathways. It has been stated by other research groups that EVX1 is a crucial downstream effector for WNT and BMP signaling. This interesting key genetic element (EVX1) is considered a prominent finding in our study and represents a potential therapeutic target KIT and PLG genes are two interesting prognostic markers that have been identified as hub genes in case of male, this signifies the importance of personalized-gender based diagnostic kit between male and female patient kidney cancer.

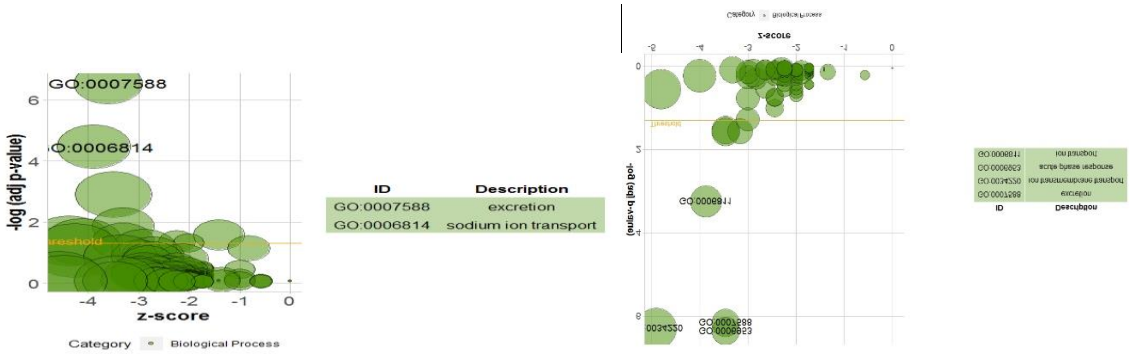
Discussion

ccRCC is known to be the most common subtype of ccRCC, it has complicated molecular background mechanisms & understanding of those mechanisms is crucial for diagnosis & treatment. it was identified that ccRCC incidence rate in male is greatly different than female, RNASeq.& differential gene & miRNAs expression analysis can provide efficient methods to understand human genome & to explore the molecular targets in male & female which can provide answers about the major differences between them. Results of enrichment analysis of male (using TFmir2) provide us with disease network interaction while we didn't get the same type of interaction in females. this disease network enhance our understanding with the disease pathways in which male DEGs are enriched in & from the results we saw that male DEGs were enriched in Hippo signaling pathway which has known role in regulation of cell proliferation & apoptosis, overexpression of genes involved in this pathway can lead to tissue overgrowth & hence developing of cancer , the enrichment of male DEGs in this pathway while this enrichment is not appeared in female may be one of the leading causes of increasing ccRCC incidence in male rather than female , as well the enrichment of male DEGS in Wnt signaling pathway which is identified in carcinogenesis as this pathway control cell proliferation & cell migration so it could be another reason for higher incidence rate of the disease in males rather than females, it was also reported that male DEGs were enriched in pathways in cancer & proteoglycans in cancer & both of them has a great role in enhancing tumor growth & both are not identified in female DEGs KEGG pathways.Also, it was identified that male DEGs were enriched in other type of cancer which is " Basal cell carcinoma" as well as in "Hepatitis c".

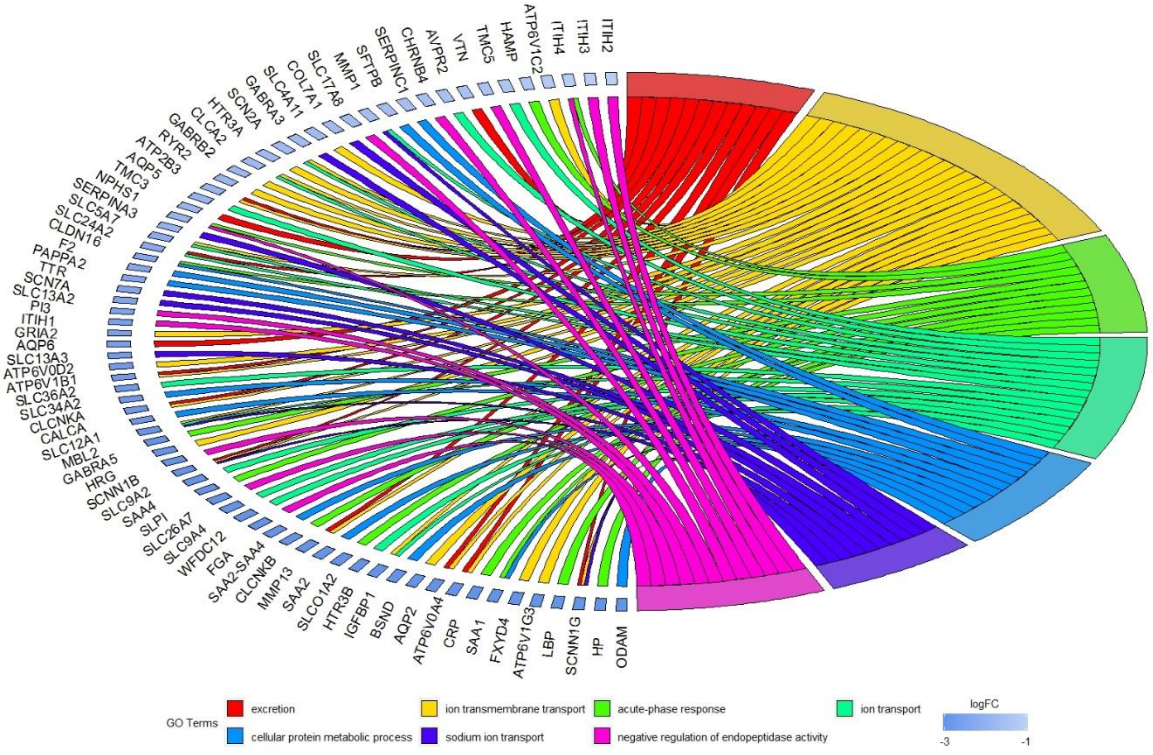


Heatmap for DEGs expression in Female cohort

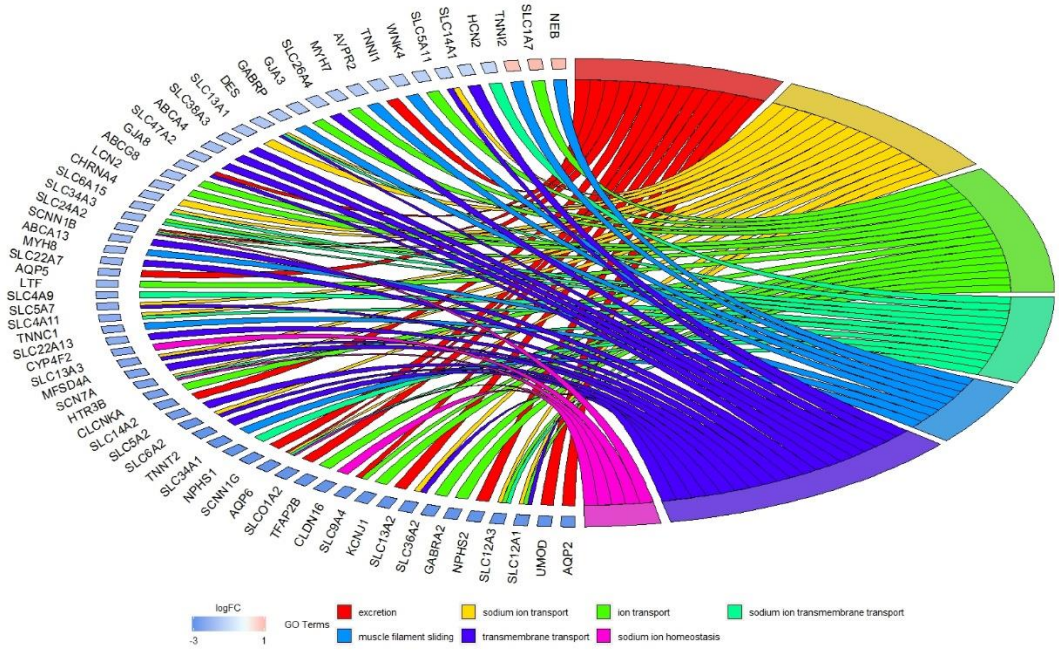
GO PLOT



Bubble GO Plot (BP): left –male / right -female

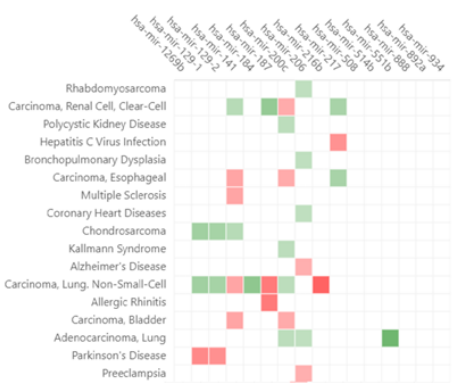


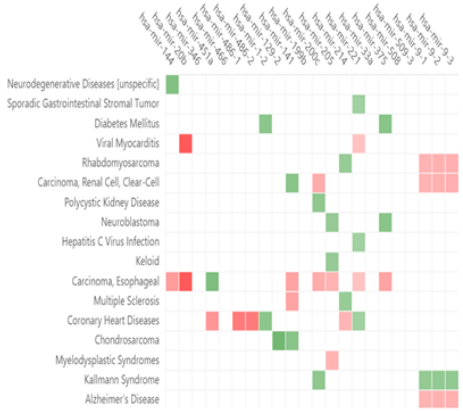
Chord Plot Biological Process: Female



Chord Plot Biological Process: Male

TAM 2.0 for miRNA set analysis





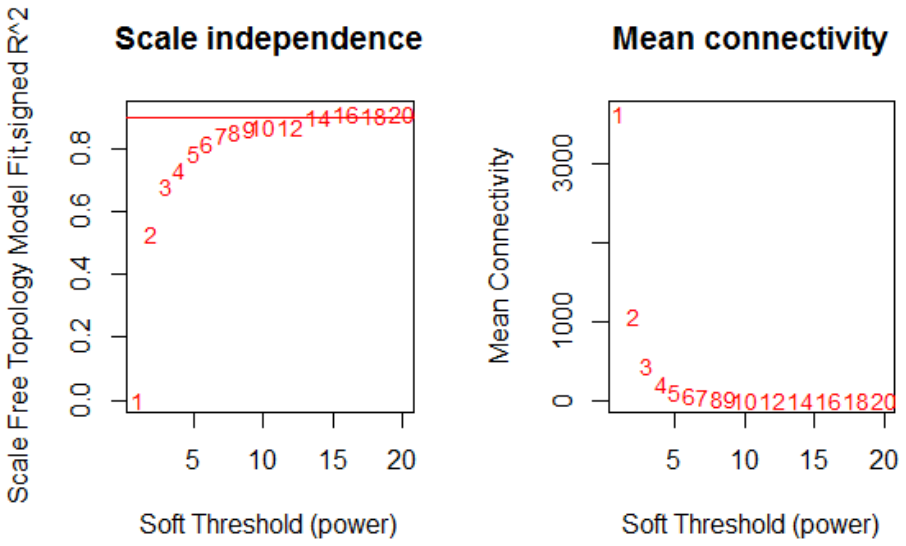
Male

Female

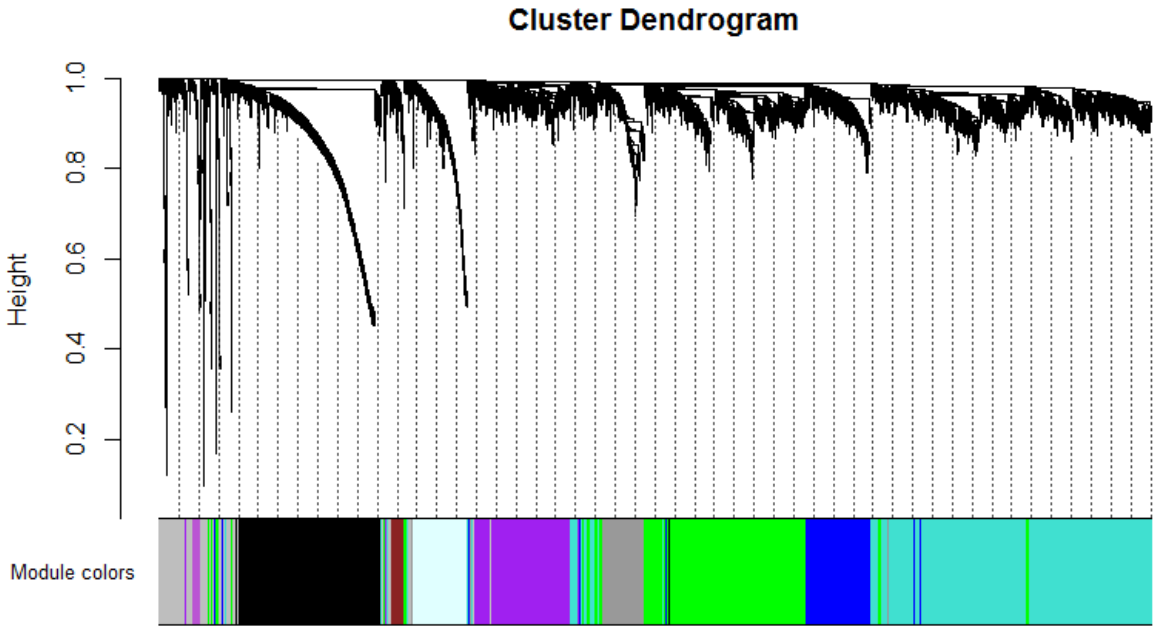
Heatmap depicts overall view of miRNA-disease association. Green represents down-regulated miRNAs, red represents upregulated miRNAs

GENE	MOLECULE	PHENOTYPE
KDR	sorafenib	Carcinoma, Renal Cell
ABCB1	sorafenib	Carcinoma, Renal Cell, Hypertension
IL13	sunitinib	Carcinoma, Renal Cell
VEGFA	sorafenib	Carcinoma, Renal Cell, hand-foot syndrome
UGT1A1	pazopanib	Carcinoma, Renal Cell
ABCB1	sunitinib	Carcinoma, Renal Cell
STAT3	interferons	Carcinoma, Renal Cell
ABCB1	sunitinib	Carcinoma, Renal Cell
FLT4	sunitinib	Carcinoma, Renal Cell

PharmGKB list of suggested molecules for renal related cancer genes



Analysis of network topology for various soft-thresholding powers for Female ccRCC



Clustering Dendogram in Female

References

1- Development and validation of a metastasis-associated prognostic signature based on single-cell RNA-seq in clear cell renal cell carcinoma,JO - Aging,JA - Aging (Albany NY), 2019

Contributions:

Sahar Mostafa

1. DAVID(Go enrichment analysis, KEGG pathway enrichment analysis)
2. Tfmir2 (enrichment analysis of DEGS & DEMS, Identification of hub genes in DEGS)
3. STRING (network construction of DEGS)
4. HMDD: MicroRNA enrichment analysis & visualization
5. GO plot(running part of R script)
6. Writing: Introduction, Report of results, Hub gene table, Discussion

Noha Ismail

1- TFmiR hotspot analysis and relating them to Renal cancer
2- Cytoscape visualisation for TFmir results
3- HMDD-miRNA analysis of unique target miRNAs and visualization
4- Cytoscape STRING Network Enrichment
5- TAM analysis and visualization
6- PharmGKB
7- Writing part of results (HMDD, keymark findings, Hub gene table), Methodology
8- Presentation ideation and PowerPoint
9- Getting Volcano Plot for Male

Hagar Elshora

1. Differential Expression Analysis ;
    a) Modifying mRNA and miRNA R script for performing male and female cohorts
    b) Performing the female mRNA analysis in R
    c) Performing the miRNA analysis for both male and female cohort in R
2. WGCNA : for female cohort till topological overlap and hierarchical clustering for male cohort
3. TAM miRNA enrichment analysis
4. Go plot : generating (circle, bubble charts & chords) for DAVID enrichment tool with cooperation of Aya abdulmonem; coursemate
5. Disgenet and ClinVar gene-disease Enrichment
6. KEGG pathway enrichment visualization
7. Heatmap for differential expressed genes in female cohort
8. Writing : part of Methodology and part of Results (miRNA enrichment)
9. Performing presentation for the course project

Enas Helmy

1- GO plot
2- DAVID (Tables required for Goplot)
3- Presentation (pathways)