6-1-2015

Transcriptional regulatory networks in Hepatitis C virus_induced hepatocellular carcinoma

Marwa Zahra

The American University in Cairo

School of Sciences and Engineering

Transcriptional Regulatory Networks in Hepatitis C Virus-induced Hepatocellular Carcinoma

Thesis submitted to:

The Biotechnology Graduate Program

In partial fulfillment of the requirements for

The degree of Master of Science in Biotechnology

By: Marwa Atef Zahra

Bachelors of Science in Biology, University of North Florida

Under the supervision of:

Dr. / **Hassan Azzazy**

Professor, Chemistry Department, The American University in Cairo

Dr. / **Ahmed Moustafa**

Associate Professor, Biology Department, The American University in Cairo

January 2016

The American University in Cairo

Transcriptional Regulatory Networks for differentially expressed genes in HCV-induced HCC

A Thesis Submitted by: Marwa Atef Zahra

To the Biotechnology Graduate Program

Jan 2016

In partial fulfillment of the requirements for the degree of

Master of Science in Biotechnology

Has been approved by:

Thesis Committee Supervisor/Chair _____

Affiliation_____

Thesis Committee Supervisor _____

Affiliation_____

Thesis Committee Reader/Examiner _____

Affiliation_____

Thesis Committee Reader/Examiner _____

Affiliation_____

_____	_____	_____	_____
Dept. Chair/Director	Date	Dean	Date

DEDICATION

I dedicate this work and any success that may come to my father, who taught me that life is lived in stages and to always make the best of each stage in my life. He taught me to be patient, wise, and to be strong so that even when I fall I can bring myself back up. So this is to the man who taught me everything and I owe him everything.

ACKNOWLEDGEMENTS

I would like to thank Dr. Hassan Azzazy for he is an inspiration every aspiring scientist. He is an icon and a role model for hard-work, great intelligence, and devotion to science. Also, Dr. Ahmed Moustafa as he never ceases to amaze me by his brilliance in what he does, I would like to thank him for all the skills he taught me in my honing of scientific knowledge and all his hard work with me in my thesis. I also would like to thank my husband, and children Mohamed and Malak for all the joy they bring to my life. As well as my second mother and father Somaya and Mohamed for all their support. Eman Rabie, Yasmeen Howeedy, and Moustafa Abo-Hawaya you are all amazing, intelligent, and beautiful people I would like to thank you for all your support.

ABSTRACT

The American University in Cairo

Transcriptional Regulatory Networks in Hepatitis C Virus-induced Hepatocellular Carcinoma

By: Marwa Zahra

Under the Supervision of: Dr. Hassan Azzazy & Dr. Ahmed Moustafa

HCV is an epidemic affecting an estimated 160 million individuals worldwide or approximately 2.35% of the world's population.(1) This is partly because HCV exhibits high genetic variation which thereby characterizes each region with its own genetic prevalence. Therefore, understanding the transcriptional regulatory elements that influence the progression of liver disease in the presence of HCV infection is thereby crucial for diagnostic and therapeutic purposes. Systems biology provides a road map by which these elements may be easily identified. In this study 124 microarray samples were assessed in order to determine differentially expressed genes for 4 tissue types/conditions (normal, cirrhosis, cirrhosis HCC, and HCC). Differentially expressed genes were assessed for their functional clustering and those genes were annotated with their potential transcription factors and miRNAs. Transcriptional regulatory networks were constructed to visualize each pairwise comparison between the 4 tissue types/conditions. In this study that 12 transcription factors were found to have high expression patterns amongst all 6 pairwise comparisons and these transcription factors also provide insight the conditions of the liver as it progresses through hepatic cirrhosis, hepatic steatosis, and the induction of cancer. With the plethora of miRNAs that are found in the liver, each liver condition was found to have its own signature miRNA expression pattern. In the 6 pairwise comparisons 14 miRNAs were found to have high expression patterns in all 6 pairwise comparisons and their regulation in HCC was determined as well as their impact on cellular homeostasis. Based on the findings of this study and a systematic analysis of many studies it can be concluded that as the liver progresses from cirrhosis to steatosis and eventually becoming carcinomic there are specific transcription factors regulating this transition through each stage. Whereas the condition of the liver digresses, the downregulation of miRNAs' expression makes the transition of the liver through each pathological stage more apparent. Therefore, an understanding of the transcriptional regulatory attributes acts as a road map to provide interference strategies in order to target the stages in the progression of HCV induced HCC.

Table of Contents

1. Thesis Approval Form……p.2
2. Dedication……p.3
3. Acknowledgments……p.4
4. Abstract……p.5
5. List of Figures……p.7
6. List of Tables……p.8
7. List of Abbreviations……p.9
8. Introduction……p.11
    a. HCV……p.11
    b. Liver Disease and HCV……p.19
    c. HCV induced HCC……p.21
    d. Transcription Regulation……p.23
    e. Microarray analysis……p.25
    f. Systems Approach……p.27
9. Hypothesis and Objectives……p.30
10. Materials and Methods……p.31
11. Results……p.37
    a. Group 1: Normal_Cirrhosis……p.39
    b. Group 2: Normal_CirrhosisHCC……p.43
    c. Group 3: Normal_HCC……p.46
    d. Group 4: Cirrhosis_CirrhosisHCC……p.49
    e. Group 5: Cirrhosis_HCC……p.52
    f. Group 6: CirrhosisHCC_HCC……p.55
12. Discussion and Conclusion……p.61
    a. Differentially Expressed Genes……p.61
    b. Transcription Factors……p.67
    c. MiRNAs……p.73
    d. Conclusion……p.76
13. Future Directions……p.78
14. References……p.80

FIGURES

Figure 1: HCV infection and the progression of liver disease…….…... 19
Figure 2: Schematic Display of Study Design……... 35
Figure 3: Diagram of the samples……..... 36
Figure 4: Column dendogram…… 37
Figure 5: (**A)** Scatter plot. **(B)**Volcano Plot Normal_Cirrhosis……. 38
Figure 6: Network visualization of TFs in group Normal_ Cirrhosis ……... 40
Figure 7: Network visualization of miRNAs group Normal_ Cirrhosis……..42
Figure 8: (A) Scatter plot **(B)** Volcano Plot in Normal_ Cirrhosis HCC……... 43
Figure 9: Network visualization of TFs Group Normal_ Cirrhosis HCC…….....44
Figure 10 Network visualization of miRNAs group Normal_ Cirrhosis HCC……....45
Figure 11: (A) Scatter plot **(B)** Volcano Plot in Normal_ HCC……. 46
Figure 12: Network visualization of TFs in group Normal_ HCC…….…47
Figure 13: Network visualization of miRNAs group Normal_ HCC……....48
Figure 14: (A) Scatter plot **(B)** Volcano Plot for cirrhosis_ cirrhosis HCC……....49
Figure 15: Network visualization of TFs in group cirrhosis_ cirrhosis HCC……..50
Figure 16: Network visualization of miRNAs in group cirrhosis_cirrhosis HCC…..……....51
Figure 17: (A) Scatter plot **(B)** Volcano Plot in cirrhosis_HCC……..……...52
Figure 18: Network visualization of TFs in group cirrhosis_ HCC……..53
Figure 19: Network visualization of miRNAs in group cirrhosis_ HCC……..…….....54
Figure 20: (A) Scatter plot. **(B)** Volcano Plot in group cirrhosis HCC and HCC…….....55
Figure 21: Network visualization of TFs in group cirrhosis HCC_ HCC……....56
Figure 22: Network visualization of miRNAs in group (cirrhosis_ HCC)……57
Figure 23: Overview of the TFs and miRNAs HCV induced HCC……..58
Figure 24: Illustration of Transcription Regulation HCV-induced HCC……..……59
Figure 25: DAVID analysis, Normal and HCC……....61
Figure 26: DAVID analysis Normal and cirrhosis……....62
Figure 27: DAVID analysis Normal and cirrhosis HCC……....62
Figure 28: DAVID analysis, cirrhosis and cirrhosis HCC……..63
Figure 29: DAVID analysis, cirrhosis and HCC…….…... 64
Figure 30: DAVID analysis, HCC and cirrhosis HCC samples……... 64

TABLES

Table 1: Summary of Databases…….. 34

Table 2: Regulation of miRNAs in HCV-induced HCC…… 41

Table 3: MiRNAs annotated with DE target genes…… 73

LIST OF ABBREVIATIONS

Ab: Antibody

ALT: Alanine aminotransferase

Apo: Apolipoprotein

cDNA: Complementary DNA

DAVID: Database for Annotation, Visualization and Integrated Discovery

DDX3X: DEAD box helicase

DE: Differentially expressed

dsRNA: Double-stranded RNA

ER: Endoplasmic reticulum

EGFR: Epidermal Growth Factor

GEO: Gene Expression Omnibus

HCV: Hepatitis C Virus

HCC: Hepatocellular Carcinoma

IFN: Interferon

IRF: Interferon regulatory factor

ISG: IFN stimulated genes

LD: Lipid Droplet

LDL: low-density lipoprotein

LVP: Lipo-viro-particle

MW: Membranous web

MDA5: Melanoma Differentiation-Associated protein 5

MICA: MHC class I polypeptide-related sequence A

NPC1L1: Niemann-Pick-CI-like cholesterol absorption

OCLN: Occludin

PAMP: Pathogen-associated molecular patterns

PNPLA3: Protein-like phospholipase domain containing protein -3

PKR: Protein Kinase R

RIG-1: retinoic acid-inducible gene 1

RMA: Robust multiarray average

SNP: Single nucleotide polymorphism

SRB1: Scavenger receptor class B type I

TF: Transcription factor

UTR: Untranslated region

VLDL: very low density lipoprotein

Chapter 1. Introduction

I. HCV

a. Epidemiology

HCV is an epidemic affecting an estimated 160 million individuals worldwide or approximately 2.35% of the world's population. (1) This is partly because HCV exhibits high genetic variation which thereby characterizes each region with its own genetic prevalence. How HCV affects every area in the world is specific to the genotype of HCV that is prevalent in that area. HCV strains are classified into seven genotypes (1-7) based on phylogenetic and sequence analyses of whole viral genomes.(2)Within each genotype, HCV is classified into 67 confirmed and 20 provisional subtypes. Strains that belong to the same subtype differ at <15% of nucleotide sites.(3)

The geographic distribution of HCV genotypes is very complex. HCV genotype 1 is the most prevalent worldwide, comprising 83.4 million cases (46.2% of all HCV cases). Approximately one-third of those cases are in East Asia alone. Genotype 3 is the next most prevalent globally comprising 30.1% of all cases and 54.3 million cases. While genotypes 2, 4, and 6, are responsible for a total 22.8% of all cases of HCV globally. Finally, genotype 5 comprises the remaining <1% of cases globally. Although genotypes 1 and 3 dominate in most countries irrespective of economic status, the largest proportions of genotypes 4 and 5 are in lower-income countries. (4)

While HCV exhibits high genetic variation as well as a rapid reproduction rate the reason as to how this virus has become an epidemic lies not in the lethality of the virus rather in the manner of its transmission. Blood transfusions, injection drug use, use of contaminated glass syringes, as

well as nosocomial infections are the main ways behind HCV transmission. Latin American countries have been very active in screening blood supplies, thus minimizing the risk of transmission through transfusion. (1) However, as in most western countries, other risk factors are currently playing a major role in accounting for new infections. The main risk for HCV transmission in countries with well-established HCV screening programs and lower HCV prevalence was found to be injection drug use.

Also, immigration from endemic countries was sometimes a major factor impacting the total number of HCV-infected persons: approximately 70% of cases in Israel, 37% in Germany and 33% in Switzerland were not born in the country. These data express the high heterogeneity of HCV across Europe, Canada and Israel. While most countries had prevalence rates of HCV cases from 1 to 2%, several other countries presented relatively high prevalence rates, such as Egypt (15%), Pakistan (4.7%) and Taiwan (4.4%). (1) Nosocomial infections and injection drug use are major risk factors in the region, and in some countries blood donors are still not universally screened. These regions are in desperate need of effective health policies that are watched over by stringent surveillance systems in order to control the spread of HCV.

Due to aging of the current HCV-infected population, the problem of hepatitis C is expected to increase. A study from USA has shown that aging of the HCV infected population has already resulted in a significant increase in the prevalence of cirrhosis and hepatocellular carcinoma (HCC) cases reported in that country during the period 1996–2006. According to another work from the same country, the number of HCV-related cirrhosis cases is estimated to increase by 24% and that of decompensated cirrhosis cases by 50%. (5) Also, a synergistic effect on morbidity and mortality is anticipated due to the overlapping worldwide epidemics of HCV, therefore the above estimates should be considered conservative.

b. HCV in Egypt

In Egypt, the spread of HCV stems from the outbreak of blood flukes that cause schistosomiasis. In the 1920s, in an attempt to treat the schistosomiasis outbreak with injection of potassium antimony tartrate through a mass inoculation campaign; the world's largest iatrogenic spread of blood-borne pathogen was created. Millions of intravenous injections were given with inadequately sterilized equipment which caused the spread of HCV. (6)

Today Egypt has a HCV epidemic, which is the largest HCV epidemic in the world. The most common variant of HCV in Egypt, genotype 4, has the highest pravelence and accounts for more than 90% of infections. (7) Given a national population of about 80 million persons, 7.8 million were estimated to be asymptomatically infected with HCV comprising a large reservoir of HCV in the population. The Egyptian demographic health survey in 2008, estimated HCV prevalence among the 15–59 years age group to be 14.7%. (8)

In order to evaluate the magnitude and patterns of exposure to HCV transmission epidemiological tools are needed. Also, for intervention to reduce HCV infection in Egypt, factors related to the distribution and determinates of HCV exposure must be evaluated.(9)

c. HCV Infection

HCV is a member of the Flaviviridae family, and is comprised of a single-stranded RNA genome 9.6 Kb in size and positive polarity. HCV has an error-prone RNA-dependent RNA polymerase that produces a highly variable progeny. This error-prone replication has allowed for an original HCV ancestor to evolve into seven genotypes and more than 80 subtypes due to the nucleotide variation that is among different HCV isolates (10). Also, due to the high mutation rate each infected individual has a collection of distinct but related HCV genomes generated by replication, known as quasi-species (11).

HCV viral particle is 40-80 nm in size. Viral particles are enveloped and contain their genetic material bound to the core protein (12). The viral genome is a positive stranded RNA that serves as a messenger RNA. The open reading frame is flanked by highly structured untranslated regions (UTR) that contain several domains important for viral replication. These domains are fairly well conserved compared to the rest of the genome, thus supporting their importance in viral fitness (13).The open reading frame encodes a polyprotein cleaved co- and post-transcriptionally by viral and cellular proteases into the three major structural proteins (core, E1 and E2) and seven non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B).

The viral particle is formed by the two envelope glycoproteins, E1 and E2, required for cell attachment and entry, and by the core capsid protein. P7, forms an ion-channel and participates in virus assembly and in the release of infectious virions (14). NS2 autoprotease also participates in virus assembly and might affect replication indirectly, as the cleavage at the NS2/NS3 junction seems to be a rate limiting step and a fully processed NS3 protein is required for replication. NS3, NS4A, NS4B, NS5A and NS5B form the replicase complex. The viral helicase/protease NS3, has an essential role in the processing of the non-structural proteins and of some host cellular proteins (14). NS4A is a cofactor that leads NS3 to the endoplasmic reticulum (ER) it also increases NS3 stability and protease and helicase activity. NS4B is involved in packaging the virus and also participates in membrane remodeling. NS5A is multifunctional in that it induces the formation of a membranous web, is involved in replication and in viral assembly, and affects the antiviral response (15). NS5B is the RNA dependent RNA polymerase required for replication and HCV assembly (16).

d. HCV Replication

In the blood, HCV virions circulate as lipo-viro-particles (LVPs) which are found to be associated with very low or low-density lipoproteins (VLDLs and LDLs). These LVPs may also associate with high levels of triglycerides, Apolipoprotein (Apo) E, ApoC, ApoB and low levels of cholesterol or phospholipids (17). This HCV virion association with lipids may facilitate its infection of hepatocytes which are the primary target for infection while at the same time aiding in its escape from neutralizing antibodies.

The exact mechanism of viral entry is still unclear entry occurs very likely by clathrin-mediated endocytosis (18). It has been suggested that the viral envelope proteins E1 and E2 bind different cellular proteins identified as receptors for viral entry such as scavenger receptor class B type 1 (SRB1) and CD81 (19). Also required for entry are the tight junction proteins claudin-1 (CLDN1) and occludin (OCLN), the epidermal growth factor receptor (EGFR) and the Niemann-Pick C1-like cholesterol absorption protein (NPC1L1), which may bind to virion-associated cholesterol at a late stage of HCV entry (20).

Once inside acidification of the endosome induces fusion of the viral envelope so that the virion undergoes un-coating. This fusion between the endosome and virion envelope leads to the release of the viral RNA into the cytoplasm. In the cytoplasm the viral RNA is transported to the rough ER for translation of the viral polyprotein; which is then further processed into the structural and non-structural proteins. HCV replicates in association with a specialized membrane structure named membranous web (MW), formed by double-membrane vesicles (21). The remaining components of the replication complex alongside the actions of NS4B and NS5A are the key players in the formation of the MW. NS5B which is the viral RNA dependent RNA polymerase, uses the viral genome as a template to produce negative-strand RNA intermediates.

These intermediates are then amplified to large amounts of new positive-strand RNA viral genomes.

Assembly of HCV virion is brought about by the synchronized actions of HCV structural proteins along with the viral replication complex, and takes place in close proximity to membrane bound lipid droplets (LD). Lipid droplets (LDs) are intracellular organelles storing cholesterol and triglycerides, and play an essential role in viral packaging. Inhibition of the synthesis of their lipid components blocks viral assembly. NS5A, as part of the replication complex, connects virus replication and packaging. New virus particles are bound to VLDL or LDL components and use the secretory pathway to be released as LVPs (22). Thus, HCV release is intimately associated with the synthesis of VLDLs and the lipid secretion pathway (17). Once released, LVPs can infect new host cells.

e. *HCV and Interferon*

The IFN-mediated innate immune response protects the cell against infection. Upon viral infection, several viral features or pathogen-associated molecular patterns (PAMPs) are detected by canonical or non-canonical receptors located on the cell membrane or intracellularly. Canonical receptors are receptors such as the retinoic acid-inducible gene I (RIG-I) or the Melanoma Differentiation-Associated protein 5 (MDA5) (23). In HCV for example, RIG-I can recognize the RNA genome within hours of infection and may initiate signaling before extensive viral protein synthesis takes place (24).

HCV can also be sensed by non-canonical receptor, protein kinase R (PKR) and DEAD box helicase DDX3X which are able to induce expression of antiviral genes. The first HCV sensor protein kinase R (PKR) is a kinase that is phosphorylated and activated by dsRNA binding and recognizes the 5´UTR of the HCV genome (25). The DEAD box helicase DDX3X, recognizes

3′UTR of HCV genome. DDX3X is a multifunctional protein involved in immune response activation (25). Although both PKR and DDX3X are antiviral, both are required for efficient HCV replication (26).

The activation of the Interferon regulatory factor (IRF) 3 and nuclear factor-κB (NF-κB) are independent of the sensors that initiate cascades for canonical and non-canonical receptors (25). IRF3 induces the expression of some IFN stimulated genes (ISGs), NF-κB induces pro-inflammatory cytokines and together, IRF3 and NF-κB serve to induce transcription of IFNβ. Then, the signaling that is initiated by IFNβ leads to upregulation of IRF7, which, together with IRF3, can lead to the synthesis of several IFN subtypes (25).

IFNλ signaling has a special impact on HCV infection in that several single nucleotide polymorphisms within the IFNλ loci associate strongly with the natural clearance of HCV infection and with response to IFNα-based antiviral therapy (27). Several ISGs function to induce an antiviral state that affects HCV infection; such as STAT1, STAT2, IRF1, 3, 7 and 9, PKR, OAS or RNase L which work to increase cell sensitivity to PAMPs. These in turn function as positive regulators to reinforce IFN signaling and prime cells for enhanced pathogen detection. For example IRF1, a transcription factor that induces a secondary wave of ISG transcription probably evolved to maintain an antiviral state after pathogen-mediated inhibition of the first wave of IFN response (28).

It becomes apparent as the role of ISGs as antivirals and positive regulators of the IFN response, that IFN is able to eliminate HCV infection. IFN therapy has been used in the treatment of HCV infected patients until very recently as treatment by overexpression of specific ISGs is not efficient.

This combined action of several ISGs causes a strong inhibition of HCV, rather than the single action of any one of several ISGs has lead HCV to counteract the antiviral role of the IFN response, and has evolved to block IFN induction and interfere with the action of several antiviral factors. The viral NS3-NS4A protease, which is a key component in the HCV innate immune evasion cleaves and inactivates MAVS and TRIF, interfering with the signaling of the major HCV sensors and attenuating IFN production (29). NS4A/B precursor then inhibits the transport of MHC Class I molecules to the cell surface. NS5A and the core proteins block the STAT pathway by several means (25). .

HCV has also evolved to exploit the expression of ISGs. ISGs function as negative regulators of the IFN response; since, viruses use many cellular systems, when the antiviral response affects the machineries required for viral viability; it compromises its own integrity. Yet, once IFN responses wear out it leads to autoimmune disorders (30). Therefore, the cell has evolved to express some ISGs that function as negative regulators of the IFN pathway which are essential to limiting the duration and strength of the IFN response and aid the cell back to homeostasis.

II. Liver Disease and HCV

At the onset of HCV infection the liver begins to react to the virus in many ways. As the virus begins to replicate and the liver sends signals to ISGs to react to these new particles in the body; the stage begins to set for changes to in the liver. HCV may have no effect on some livers, however for the majority of cases HCV digresses the condition of the liver in stages. These stages include the digression from cirrhosis, to steatosis, and eventually carcinoma. Clinical symptoms or signs of liver disease may or may not appear and a minority of patients experience

clinical symptoms. However, symptoms that do accompany liver disease include fatigue nausea, poor appetite, muscle aches, arthralgias, feverishness, weakness, and weight loss. Aside from symptoms, the majority of patients with HCV are recognized by having raised ALT serum levels. Again, a minority of patients may not have elevated ALT levels and these patients may be described as "healthy carriers". Studies have shown however that prolonged elevated ALT levels are associated with the development of cirrhosis in the liver. (31)

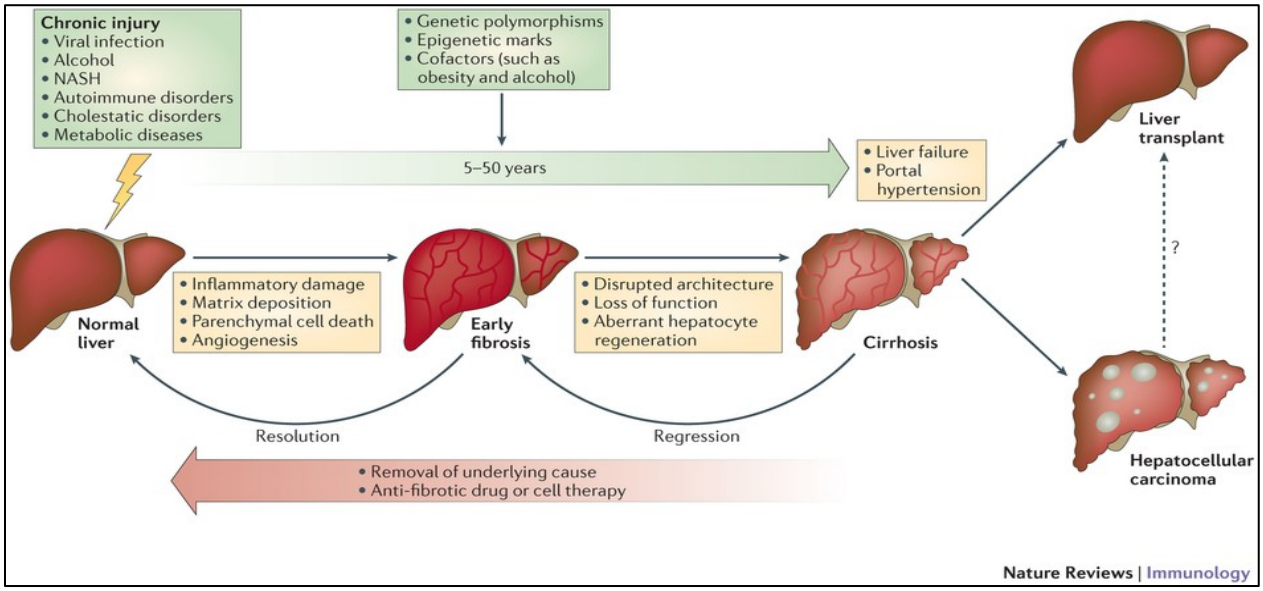

Figure 1: HCV infection and the progression of liver disease. Here it becomes apparent that with the right conditions liver disease progresses to HCC over a period of time. Although this is not a central dogma of liver disease, progression to HCC may proceed in this manner if all of the conditions shown are present.(32)

Chronic HCV can lead to cirrhosis which can develop rapidly; within 1–2 years of exposure, or slowly within 2–3 decades. In studies with 10–20 years of follow-up, cirrhosis develops in 20–30 percent of patients. Once cirrhosis develops symptoms of end-stage liver disease can appear. These symptoms are marked by fatigue, muscle weakness and wasting, fluid retention, easy bruisability, upper intestinal hemorrhage, jaundice, dark urine, and itching. Yet, some patients with cirrhosis remain asymptomatic of liver disease until they have major complications of cirrhosis, such as variceal hemorrhage or ascites or they die of an unrelated cause. (31)

Another condition that can hasten the development of cirrhosis in the liver and lead to the progression of end-stage liver disease in the presence of HCV infection is hepatic steatosis. Hepatic steatosis is defined as excessive lipid accumulation within the hepatocyte cytoplasm and has been recognized as the sole route for a direct cytopathic effect by the HCV. Several mechanisms have been proposed as to the process of lipid accumulation in the liver such as; HCV core protein may be interacting with apolipoprotein. AII which is a major component of high density lipoproteins and this interaction may be causing hepatocellular steatosis.(33) Also, it has been suggested that the HCV core protein induces oxidative stress within the mitochondria which leads to or contributes to lipid accumulation.(34) Yet, in the retrospect it is important to note that hepatic steatosis leads to the progression of end-stage liver disease, whether steatosis achieves this by aggravating fibrosis itself or the factors that are causing steatosis are also aggravating fibrosis is still unknown.

It is important to note that while IFN treatment may reduce the incidence of HCC. HCC is an inevitable event as the liver digresses in its condition by steatosis and cirrhosis. The mechanism by which this eventual digression leads to carcinoma is still unknown. However, through analysis of the key components that lead to the progression of liver disease to carcinoma it is apparent that the condition of the liver is at large the reason behind the incapability of the liver to maintain cellular homeostasis and in turn eventually digresses to carcinoma.

III. HCV induced Hepatocellular Carcinoma

a. Pathogenesis

The development of HCV-induced HCC is progressive and may occur over a period of 20-40 years. This multi-step progression involves establishment of chronic HCV infection, chronic hepatic inflammation, progressive liver fibrosis, initiation of neoplastic clones accompanied by

irreversible somatic genetic/epigenetic alterations, and progression of the malignant clones in a carcinogenic tissue microenvironment. Yet, each step of HCV-induced hepato-carcinogenesis is a potential target for therapeutic intervention or chemoprevention. Since, HCV is a RNA virus with limited integration of its genetic material into the host's genome, the carcinogenic potential of HCV is generally assumed to be linked to indirect mechanisms. A major obstacle for the understanding of the mechanisms linking HCV infection, inflammation and carcinogenesis is the lack of an *in vitro* and *in vivo* model systems. (35)

b. Host factors affecting Susceptibility

There have been associations made between various host polymorphisms associated with the immune system and HCV-induced HCC. A Japanese genome-wide association study comparing HCV-related HCC patients with chronic hepatitis C patients identified a single nucleotide polymorphism (SNP) in MHC class I polypeptide-related sequence A (MICA - rs2596542), which is involved in response of dendritic cells to type-I interferon in chronic hepatitis C (36). Another SNP in the MICA promoter (rs2596538) was associated with increased serum soluble MICA protein. In another study, in Caucasian HCV patients in Switzerland did not replicate the association with HCC for this locus, but for a nearby locus in HCP5 (rs2244546), suggesting that the MICA/HCP5 region contains a potential susceptibility locus.(37) An IL28B variant (rs12979860), initially identified as an interferon response predictor, may be associated with increased risk of HCV-related HCC (38).

Also, SNPs in genes associated with metabolic functions have also been weakly associated with HCV-induced HCC. In patients with chronic HCV with advanced fibrosis a positive association has been found between liver iron deposition and higher incidence of HCC (35). Also, a SNP in

the patatin-like phospholipase domain-containing protein 3 (PNPLA3) gene (rs738409) associated with alcoholic and non-alcoholic steatohepatitis may have weak association with HCV-related HCC (35, 39).

c. Assessment of Risk: HCV-induced HCC

Due to the high prevalence of HCV-induced cirrhosis and the population that is at risk of developing HCC, the need for surveillance is essential, and yet probably unmanageable. For instance, despite guidelines recommending HCC surveillance in subjects with cirrhosis, most patients at risk of HCC in the US do not receive recommended regular surveillance. Only 12% of cirrhotic HCV patients had routine annual surveillance in one US Veterans Affairs series and only 2% of HCV patients who developed HCC had previous appropriate screening in another series. Also, in a population-based US study, less than 20% of patients with cirrhosis who developed HCC received regular surveillance. (35)

This thereby expresses the need for prognostic indicators that would actively stratify HCV subjects, as well as patients with other liver diseases, into clearly defined risk groups to enable effective clinical management of patients (40).

IV. Transcriptional Regulation

The regulation of transcription is the means by which the cell facilitates the regulation of DNA to RNA in order to orchestrate gene activity. The attributes that aid in this facilitation are transcription factors and miRNAs. These attributes fine tune the rate of gene regulation and in turn affect overall transcription regulation.

a. Transcription Factors

Transcription factors include a wide number of proteins that regulate the transcription of genes. These, transcription factors, are proteins that are involved in the converting or transcribing DNA into RNA. Transcription factors have DNA-binding domains that allow them to bind to specific sequences of DNA.(41) Those transcription factors with promoter sequences bind to a DNA promoter sequence near the transcription start site and help form the transcription initiation complex. (42)While, other transcription factors bind to regulatory sequences, such as enhancer sequences and can stimulate or repress transcription of the related gene. (42) The action of transcription factors allows for unique expression of each gene in different cell types and during development.

b. ***miRNAs***

MicroRNAs act to fine-tune the expression of as much as 30% of all mammalian protein-encoding genes. Mature miRNAs are short, approximately 22 nucleotides in length, and single-stranded RNA molecules.(43) Their diverse expression patterns allow them to regulate many aspects of development and physiology.(44)

While the majority of the characterized miRNA genes are intergenic, or oriented antisense to neighboring genes and are therefore transcribed as independent units. A miRNA gene may also be transcribed together with its host gene, which provides a means for coupled regulation of miRNA and protein-coding gene.(45)

As much as 40% of miRNA genes may lie in the introns of protein and non-protein coding genes or even in exons of long nonprotein-coding transcripts.(46) These are found in a sense orientation, and thus usually are regulated together with their host genes. Other miRNA genes

with a common promoter include the 42-48% of all miRNAs originating from polycistronic units. (46)

The DNA template is not the final word on mature miRNA production: 6% of human miRNAs show RNA editing (IsomiRs), the site-specific modification of RNA sequences to yield products different from those encoded by their DNA.(47) This increases the diversity and scope of miRNA action beyond that implicated from the genome alone.

V. Microarray Analysis

Genes in biology are studied in many ways and analyzed through many perspectives, one of the most indispensable tools that biologists use to monitor gene levels in a given organism. A microarray is a glass slide on to which DNA molecules are fixed at specific locations called spots; each spot may contain copies of identical DNA molecules that uniquely correspond to a gene(48). This DNA may either be genomic or a short stretch of oligo-nucleotide strands that correspond to a gene.

Microarrays are manipulated in a variety of ways in order to measure gene expression. One of the most popular methods used is a means of comparison between a set of genes from a cell in a particular condition to the same set of genes in a reference cell under normal conditions. This is done by extracting RNA from these cells and converting it into cDNA using reverse transcriptase and nucleotides labelled with different fluorescent dyes. These dyes allow for samples to be differentially labeled. These samples then are allowed to hybridize onto the same glass slide where they will bind to specific spots on the glass slide containing its complementary sequences. Spots on the hybridized microarray are then excited by a laser and scanned at wavelengths suitable to the dyes used. The fluorescence that is emitted corresponds to the concentration of

bound nucleic acid, and the final image of the microarray is a spot that corresponds to a gene with an associated fluorescence value expressing the relative expression level of that gene. (48)

a. Data Analysis

There are many sources of systematic variations that may affect gene expression levels. Variation may arise from differential labeling efficiency of fluorescent dyes, or different amount of starting mRNA in the two samples. In order to eliminate variations and allow for a honed comparison of data from two samples the data that is obtained is normalized. Normalization is performed by choosing a gene-set for which expression levels should not change under the conditions studied. That gene set is expected to be 1, and any variability that is seen in other gene- sets is calculated based on its deviation. This normalization thus preserves biological variations and minimizes experimental variations. Several algorithms may also be called upon in order to achieve normalization; these include robust microarray analysis (RMA), GC-RMA as well as Microarray Analysis Suite (MAS5.0). These algorithms only differ in their way of background correction, scaling and calculation of the normalization factor.(49)

b. Determining Biological Relevance

Once data has been normalized it may be inspected for biologically significant variations that are statistically significant. A fold change value is the means by which biological significance is displayed. This fold change value is calculated as the ratio between the log expression intensity of a certain gene compared to its normal condition expression value. The higher this value, is the greater the biological significance. However, statistical significance for a data set is determined by another value the p-value; which is defined as the probability of accepting the NULL hypothesis when it should be rejected. The higher this value the less the data is statistically significant. These statistical tests when applied provide a set of genes that are differentially expressed (DE). These genes are of particular interest in that they show biological and statistical

significance across all experimental conditions and in turn express a strong over expression or down expression compared to their controls.(50) From this inferences can be made as to relationships between genes and their regulation in general.

Data from microarray analysis or gene expression data can be used to infer regulatory relationships. Studies have showed that expression data may be used as a tool to make predictions about transcriptional regulators for a given gene or for a gene set as well as identifying modules of co-regulated genes and their transcriptional regulators by means of a probabilistic model. This allows for further hypothesis to be made which are experimentally testable. (51)

VI. Systems Approach

The systems approach to biology is centralized around one particular theme which is the sum of the whole is greater than the sum of its parts. Since, the central dogma in biology is DNA to RNA to protein; it makes sense that one would approach a biological question in this manner, since no one of the pieces in the central dogma is enough to answer a biological question fully. Systems biology or the systems approach, studies biological systems by systematically breaking them down and looking at the gene, protein, and informational pathway responses. The integration of this data by means of mathematical models ultimately describes the structure of the system, and in turn answers questions that may arise to any disruptions that may happen to that system. (52)

New technologies for systematically characterizing cellular responses have brought about this new approach to understanding biological systems. Tools such as DNA sequencers, microarrays, and high-throughput proteomics have become the methods of choice for rapid and comprehensive assessment of biological system properties and dynamics. Microarrays are a

powerful tool that allows interrogation of complete human transcriptomes. Microarrays make it possible to distinguish single nucleotide differences and in turn add even the most miniscule details to the whole system. This increases the capacity for a suitable diagnostic approach to many diseases and a more refined method for approaching suitable treatment.

Approaching a disease such as HCV using a systems approach is most effective, in that it is a very robust disease that attacks biological systems through many attributes. Especially when considering HCV as it induces HCC, this approach becomes most suitable as it looks, not only at the gene-gene interactions involved but also at the regulatory aspects of those genes. This allows for the construction of transcription regulatory networks which takes into account all the regulatory elements that impress on a particular gene of interest. In these networks, transcription factors' (TFs) databases such as TRANSFAC, JASPAR, and SWISSREGULON are utilized to obtain information about the binding motif sites related to these TFs, as well as the potential interacting genes. Also, understanding how HCV affects biological systems would not be complete without considering miRNAs. miRNAs are known to regulate gene transcription via the interaction with mRNA. Over 2500 mature human miRNA have been isolated and described in a public database known as miRBase.(52) Therefore, understanding the changes in gene transcription regulated by miRNAs and provoked by HCV may explain causes for fibrosis and carcinoma in liver tissues.

In order to visualize all of the aspects that contribute to the regulation of a biological system Cytoscape 3.2.1 is utilized. Cytoscape is an open source software platform that allows for visualization of complex networks while at the same time integrating any data attributes to that network. This in turn allows for the visualization of a network as a whole system with all of the parts that may be involved in its regulation.

Since, HCV attacks the liver in a manner that disrupts its normal functioning, it is only feasible that in order to attack back at such a disease, that a clear picture must be drawn as to where, in the regulation of normal functioning, this disruption takes place. Accordingly, this thesis places an emphasis on the study of the interactions that take place in the transcription regulation that alters gene expression in HCV induced HCC.

Chapter 2. Hypothesis and Objectives

HCV has always been studied in a manner that examines a specific attribute that is affected by infection. In this study, HCV is analyzed based on how it affects transcription regulation as a whole and eventually leads to Hepatocellular Carcinoma. In this study it is hypothesized that:

Each tissue type or condition has a variable gene expression profile that determines its signature transcriptional regulatory network and therefore infers possible factors that influence HCV-induced HCC.

OBJECTIVE 1

In this study our aim is to be able to determine first differentially expressed genes in 6 pairwise comparisons in order to be able to determine those genes that are influenced by HCV induced HCC.

OBJECTIVE 2

After determining those genes that are differentially expressed in each of the 6 groups, transcription factors and miRNAs associated with the regulation of those genes are to be determined and analyzed for their association with those genes.

OBJECTIVE 3

Once, the transcription regulators for the differentially expressed genes in each of the 6 groups have been determined, a network will be constructed in order to visualize the attributes that transcriptionally regulate the differentially expressed genes in each group.

Chapter 3. Materials and Methods

Study design

In order to study transcription regulation, first, the differential expression in each of the four sample clusters (normal, cirrhosis, cirrhosis with HCC, and HCC) were compared against one another, using R studio, and 6 groups were obtained:

Group 1: Normal_ Cirrhosis; which compares normal mean expression values for differentially expressed genes against cirrhosis mean expression values for differentially expressed genes.

Group 2: Normal_Cirrhosis HCC; which compares normal mean expression values for differentially expressed genes against cirrhosis HCC mean expression values for differentially expressed genes

Group 3: Normal_HCC; which compares normal mean expression values for differentially expressed genes against HCC mean expression values for differentially expressed genes.

Group 4: Cirrhosis _Cirrhosis HCC; which compares cirrhosis mean expression values for differentially expressed genes against cirrhosis with HCC mean expression values for differentially expressed genes.

Group 5: Cirrhosis _ HCC; which compares cirrhosis mean expression values for differentially expressed genes against HCC mean expression values for differentially expressed genes.

Group 6: Cirrhosis HCC _ HCC; which compares cirrhosis with HCC mean expression values for differentially expressed genes against HCC mean expression values for differentially expressed genes.

After the differentially expressed genes were determined, these genes were annotated with their transcription factors and miRNAs in tables accessible to Cytoscape. The transcription regulatory networks obtained from Cytoscape were studied with reference to the literature.

Softwares employed:

In this study several softwares were employed. In order to attain microarray samples GEO which is a public functional genomics data repository supporting MIAME-compliant data submissions was employed. Within GEO array- and sequence-based data are accepted and tools are provided to help users query and download experiments and curated gene expression profiles. Once microarray data was provided R studio was employed in order to process microarray data for

further analysis. R studio is a free open source integrated development environment which utilizes a programming language for statistical computing and graphics**.** Visualization of transcriptional networks was performed using the software Cytoscape 3.2.1. which is an open source software platform for visualizing complex networks and integrating them with any type of attribute data. Finally, database for annotation visualization and integrated discovery (DAVID) was employed. DAVID is a bioinformatics tool that provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

Samples

The samples that were used in this study were obtained from a previous study, "Genes involved in viral carcinogenesis and tumor initiation in Hepatitis C virus- induced Hepatocellular Carcinoma" performed by the group of Mas V., etal. 2009. The samples corresponding to this study were obtained from the Gene Expression Omnibus (GEO) with the identification number GSE14323. From this database 124 samples were obtained and downloaded in the form of a compressed file 500MB in size. These samples were then extracted and prepared for further processing in the form of .cel files.

Sample Characteristics

The study that was used was restricted to patients with HCV infection. These samples included 124 liver tissue samples obtained from 88 distinct patients, 41 HCV-cirrhotic from patients without HCC, 17 cirrhotic tissues from patients with HCC, and 47 HCV-HCC tissues. Also, 13 patients with HCC-cirrhosis provided both tumor and cirrhotic tissue, 3 patients provided both cirrhotic or tumor tissue for array processing, and 19 normal liver tissues were

included. These patients were found to have normal liver function and histopathology, as well as the normal liver samples used which were also seronegative for HCV Ab.

Sample Processing: R studio

The samples that were obtained were further processed using R version 3.2.0. Samples in the form of .cel files were found to be in two different affymetrix platforms because of the difference in probes used for the samples. Therefore, in R studio an affy bioconductor package was downloaded to read the samples in the form of .cel files and samples were separated into DATA1 and DATA2 based on the Affymetrix platform corresponding to the samples. DATA1 in R corresponded to samples GSM358107- GSM358111, while DATA2 corresponded to samples GSM358112- GSM358227. These samples were converted from .cel files which are files that correspond to light intensities into .cdf files that converted light intensities into probe concentrations using the affy Bioconductor package in R studio. After which these values were then converted into expression values using the Robust Multiarray Average (RMA). These expression values were then saved for further processing.

Sample Processing: Determination of Differential Expression

In order to be able to analyze the dataset for differentially expressed genes, the samples were further analyzed using R version 3.2.0. Since this study contained 124 samples, these samples were grouped into 4 matrices, normal, cirrhosis, cirrhosis HCC, and carcinoma. The mean of the expression values was determined for the samples in each group and written into a new table with the probe ID beside it. A scatter plot was then created for groups against one another to show if there was differential expression. In order to test samples for biological and

statistical significance, the samples were filtered by a fold cut off of 2 and a p-value of 0.01. A volcano plot was then created in order to show biological and statistical significance.

Genes that were found to be both biologically and statistically significant from the volcano plot were filtered into a table by themselves based on the fold cutoff and p-value that were used as a criterion for filtration case by case. A logical vector was created in order to filter these genes and the fold cut off and p-value that was applicable for each case was setup in this vector in order for genes that were found to be true and true for both values were placed into this vector and written into a table.

In order to show the distance or similarity between groups after the filtered differentially expressed genes were written into each group a column dendogram was created. Row dendograms were also created in order to show similarity between differentially expressed genes in each case. A heat map was also created in order to visualize the differentially expressed genes across each group with yellow, corresponding to high expression, and red, corresponding to low expression.

Once the differentially expressed genes were determined for each case, they were manually paired with their corresponding transcription factors and miRNAs. Using various databases such as TRANSFAC, miRBASE, and GENECARDS, transcription factors and miRNAs were searched for in order to correspond to genes that were determined to be differentially expressed. These results were then written into a table in the form of a text file and were further processed.

Systems Approach: Networks

Tables that were created with differentially expressed genes and their corresponding transcription factors and miRNAs were evaluated using Cytoscape version 3.2.1. Tables were imported in the form of text files and a network was created in order to visualize the structure of how the genes are connected to their transcription factors and miRNAs. In each network genes are colored in blue, transcription factors are in green, and miRNAs are in red. The size of genes corresponded to their fold change, and each network was analyzed and clusters were determined if any.

Differentially expressed genes were also functionally annotated using DAVID database. This database clustered genes in each case based on their function. Genes were delivered to the database in the form of a list, case by case, and the database clustered the list of genes based on the similarity of the genes functionally.

Table 1: Summary of databases employed for analysis in this study. For each database that was employed a reference is given as a hyperlink by which this database may be accessed.

Database	Reference
Gene Expression Omnibus (GEO)	www.ncbi.nlm.nih.gov/**geo**
DAVID	https://**david**.ncifcrf.gov/
TRANSFAC	www.gene-regulation.com/pub/databases.html
MirBASE	www.**mirbase**.org
GENECARDS	www.**genecards**.org

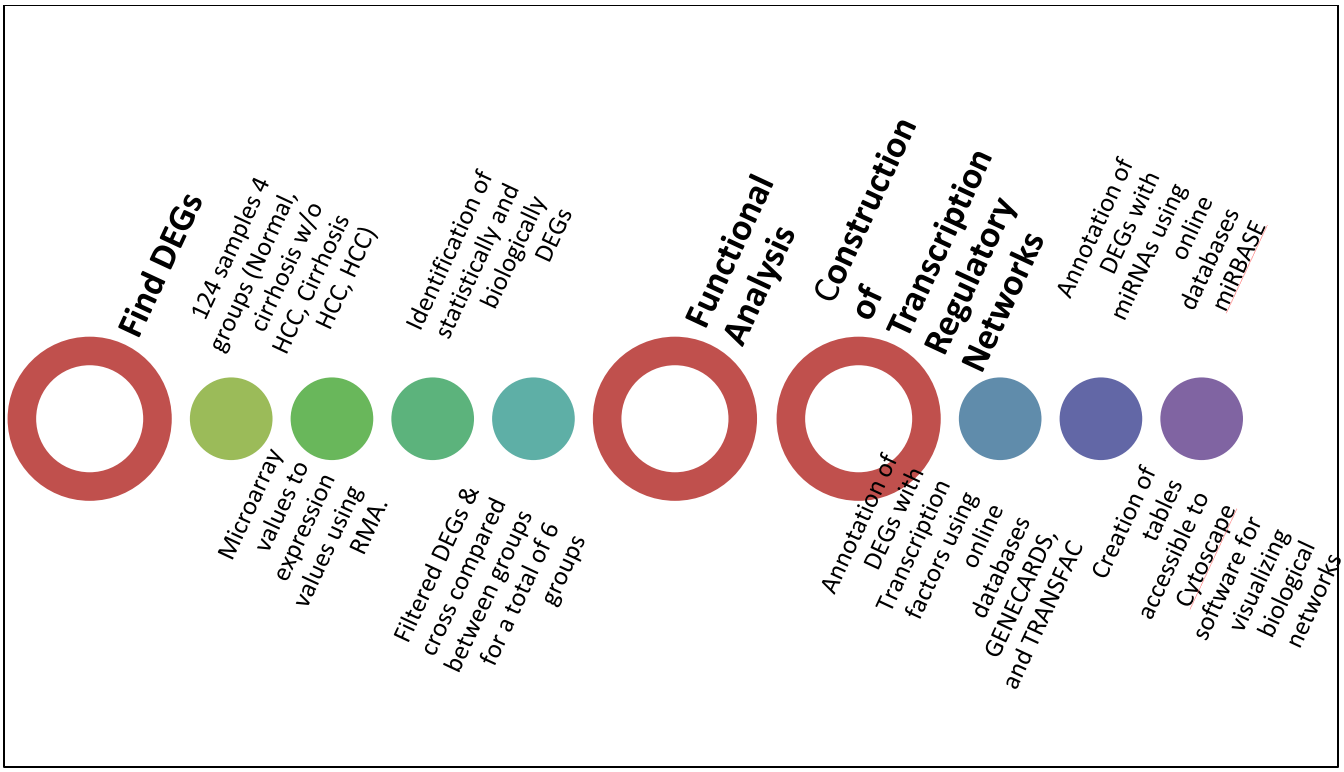

Figure 2: Schematic display of study design. As an overview of how this study was performed this diagram shows the process that the study proceeded in to create transcriptional regulatory networks for HCV-induced HCC.

Chapter 4. Results

The samples in this study were obtained from the Gene Expression Omnibus (GEO) with the identification number GSE14323. The samples were broken down as shown in Figure 3, 124 samples were obtained and analyzed using R studio in order to determine differential expression of genes in each of the 4 distinct samples' clusters.

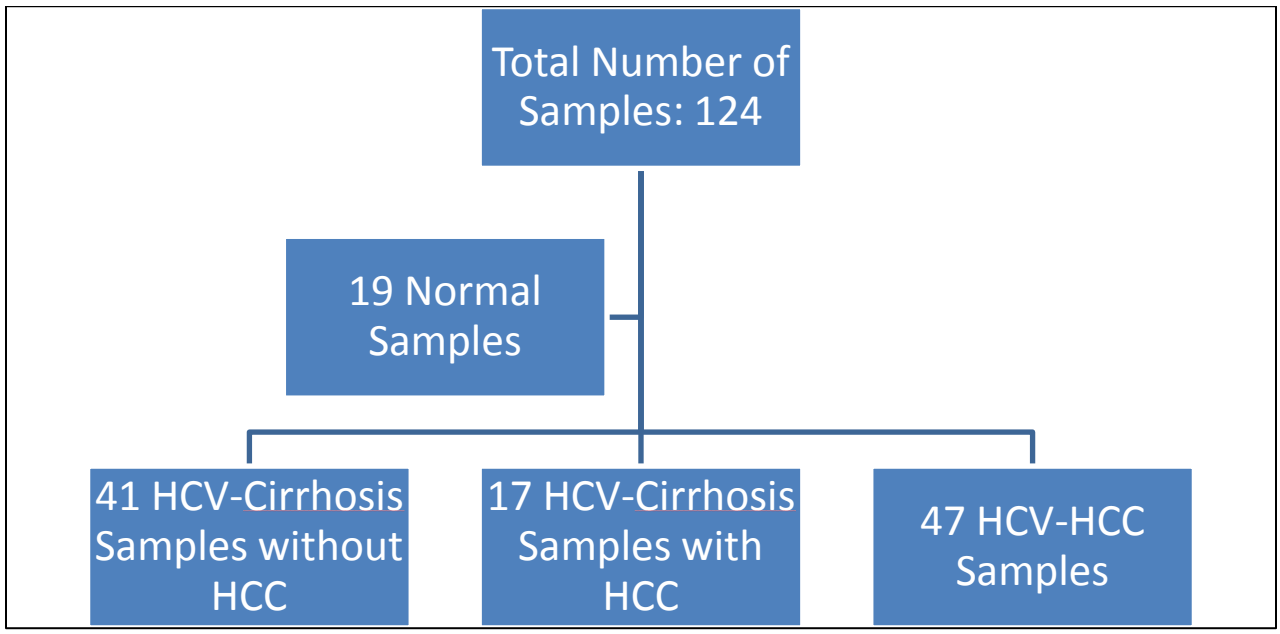

Figure 3: Diagram of the samples obtained for this study from GEO (GSE14323). A total of 124 samples were obtained; as 19 normal samples, 41 HCV-cirrhosis samples without HCC, 17 HCV-cirrhosis samples with HCC, and 47 HCV-HCC samples. These samples were further analyzed using Affymetrix microarray analysis.

The means of the expression values for differentially expressed genes in each of the four sample clusters were placed into matrices so that they may be correlated against one another as shown in figure 3, the column dendogram.

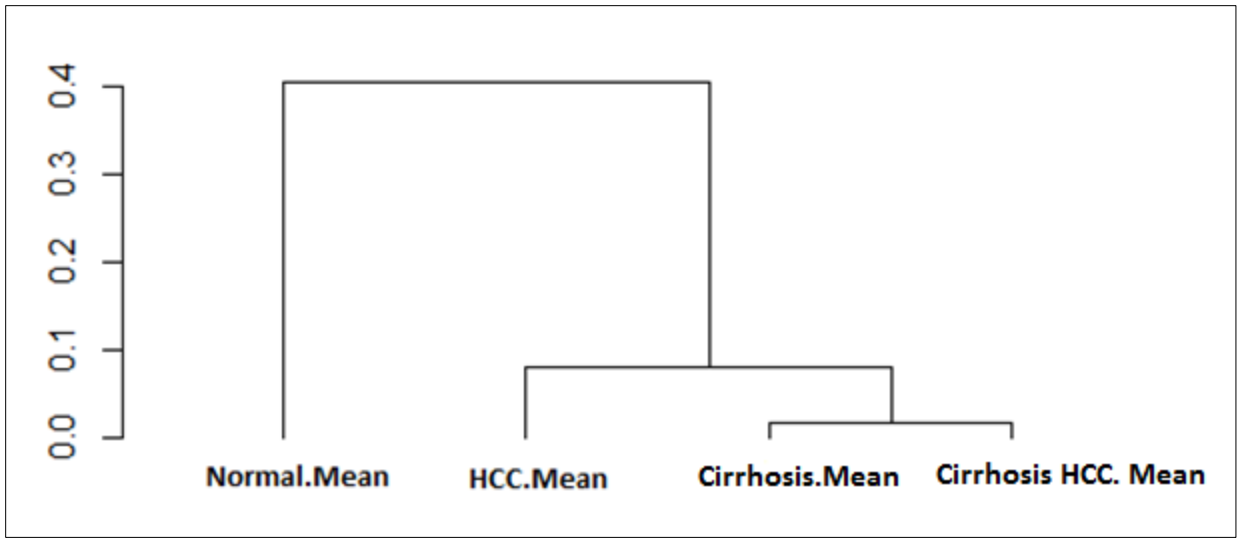

Figure 4: Column dendogram showing the correlation between the mean expression values of the genes in each of the 4 samples' clusters. The greater the value on the y-axis the lower the correlation between the groups, as is apparent the highest correlation is between cirrhosis and cirrhosis HCC.

As the value in the y-axis the lower the correlation between the groups, in figure 4 there is an obvious high correlation between the means of expression values for genes in the cases of Cirrhosis and Cirrhosis with HCC. Yet, the HCC samples showed to be more closely correlated to Cirrhosis Samples and not Cirrhosis HCC. Normal tissue samples showed the least correlation between Cirrhosis samples and Cirrhosis HCC samples, and showed more of a correlation with HCC samples in the mean expression values of genes.

In order to further analyze the differential expression in each of the four sample clusters, they were compared against one another in the fore mentioned 6 groups; (normal_cirrhosis, normal_cirrhosis HCC, normal_HCC, cirrhosis_cirrhosisHCC, cirrhosis_HCC, cirrhosisHCC_HCC). For each group, a scattered plot and a volcano plot were generated using R studio. Afterwards, the differentially expressed genes were used to create transcription regulatory networks with their corresponding transcription factors and miRNAs.

Group 1 (Normal_ Cirrhosis)

Since, each matrix corresponds to the expression values in RMA format obtained from each sample cluster, each matrix was analyzed against normal samples and a scatter plot was produced. In figure 5A each gene has an expression value which corresponds to a specific circle in the plot; those circles that are away from the red line correspond to genes that are differentially expressed. In order to further analyze these samples for biological and statistical significance a volcano plot was produced.

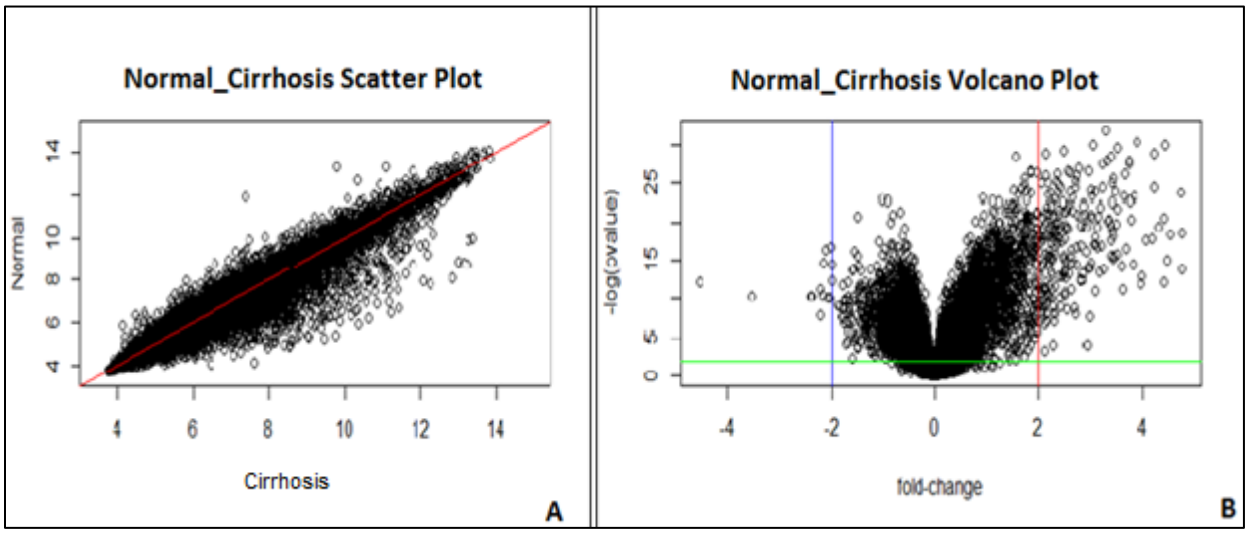

Figure 5: **(A)** Scatter plot of means of expression values for each gene in Normal group and Cirrhosis group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)**Volcano Plot expresses the means of expression values for genes in the Normal_Cirrhosis group. The fold cut off was at value of 2 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (normal and cirrhosis) that are found away from the volcano represent both statistical and biological significance and are further filtered out.

In figure 5B the expression values of genes are plotted against –log p-value and fold change. In order to analyze for those genes whose expression values are not only biologically relevant but statistically relevant as well. In figure 3B those circles on the right of the red line and above the green line have statistical and biological significance and are upregulated as well;

while those genes whose expression values lie to the left of the blue line and above the green line show to be both statistically and biologically significant and downregulated as well.

These filtered genes that are shown to have both biological and statistical significance were determined and further analyzed for their corresponding regulating transcription factors as well as miRNAs. In order to further analyze the factors that affect differentially expressed genes between normal tissue samples and cirrhotic tissue samples with HCV infection transcription factors for differentially expressed genes were determined using a series of databases and online resources. These transcription factors were further assessed for their frequency in regulating all of the differentially expressed genes in the group. In the Normal_Cirrhosis group 12 transcription factors were found to regulate differentially expressed genes most frequently. These transcription factors include AML1a, AP-1, ATF-2, c-Jun, CREB, C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure 6 a network was constructed showing those transcription factors and the differentially expressed genes that they regulate.

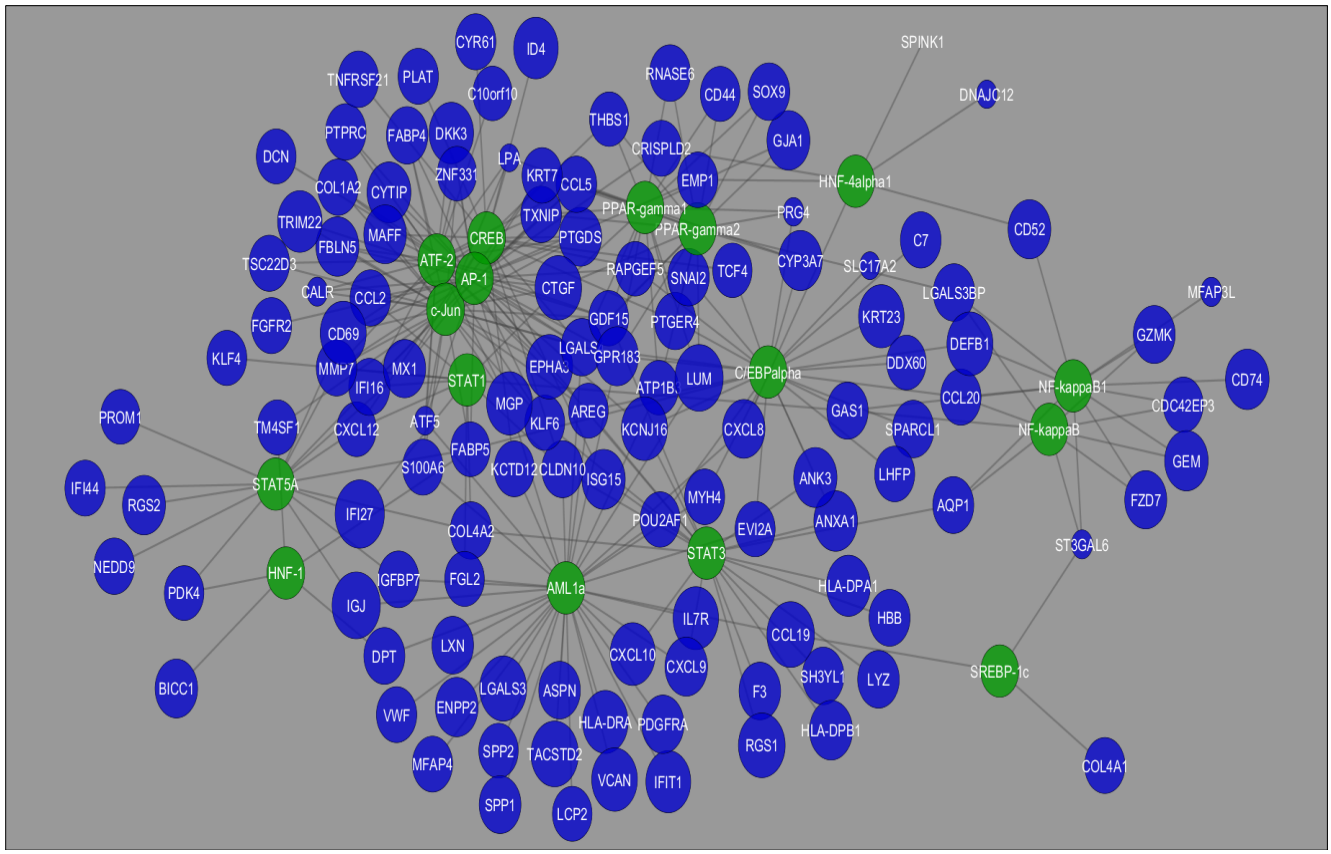

Figure 6: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in Group Normal_Cirrhosis. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. A plethora of potential miRNAs were found to regulate genes in this group and due to the plethora of miRNA that are naturally found in the liver, miRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. Also, a review of the literature made it possible to determine the type of regulation of these miRNAs observed in the presence of HCC. In the table 2 it becomes apparent that the majority of the miRNAs that are involved in regulation of differentially expressed genes are downregulated in the presence of HCC. The miRNAs in the table 2 were constructed in a

network as a visualization of the genes that are involved with these miRNAs. In figure 7 it becomes apparent that miRNAs are involved in the regulation of a vast array of genes.

Table 2: Differentially expressed genes annotated with their potential miRNAs were analyzed for expression patterns of miRNAs. The miRNAs reported in this table showed the highest expression patterns for differentially expressed genes in all 6 groups reported previously. The majority of these miRNAs were found to be down regulated in the presence of HCC.

miRNAs with highest Expression Patterns		
miRNA	**Regulation in HCV-HCC**	**Reference**
miRNA-335-5p	Up	(53)
miRNA-128	UP	(54)
miRNA-27 a&b	Down	(55)
miRNA-106a	Down	(55)
miRNA-15a	Down	(55)
miRNA-181 a&c	UP	(55)
miRNA-93	Up	(55)
Let-7	Down	(56)
miRNA-199a-3p	Down	(56)
miRNA-124	Down	(57)
miRNA-124-3p	Down	(57)
miRNA-29 a&c	Down	(57)
miRNA-26b-5p	Down	(58)
miRNA-200 a&b	Down	(58)
miRNA-607	?	none

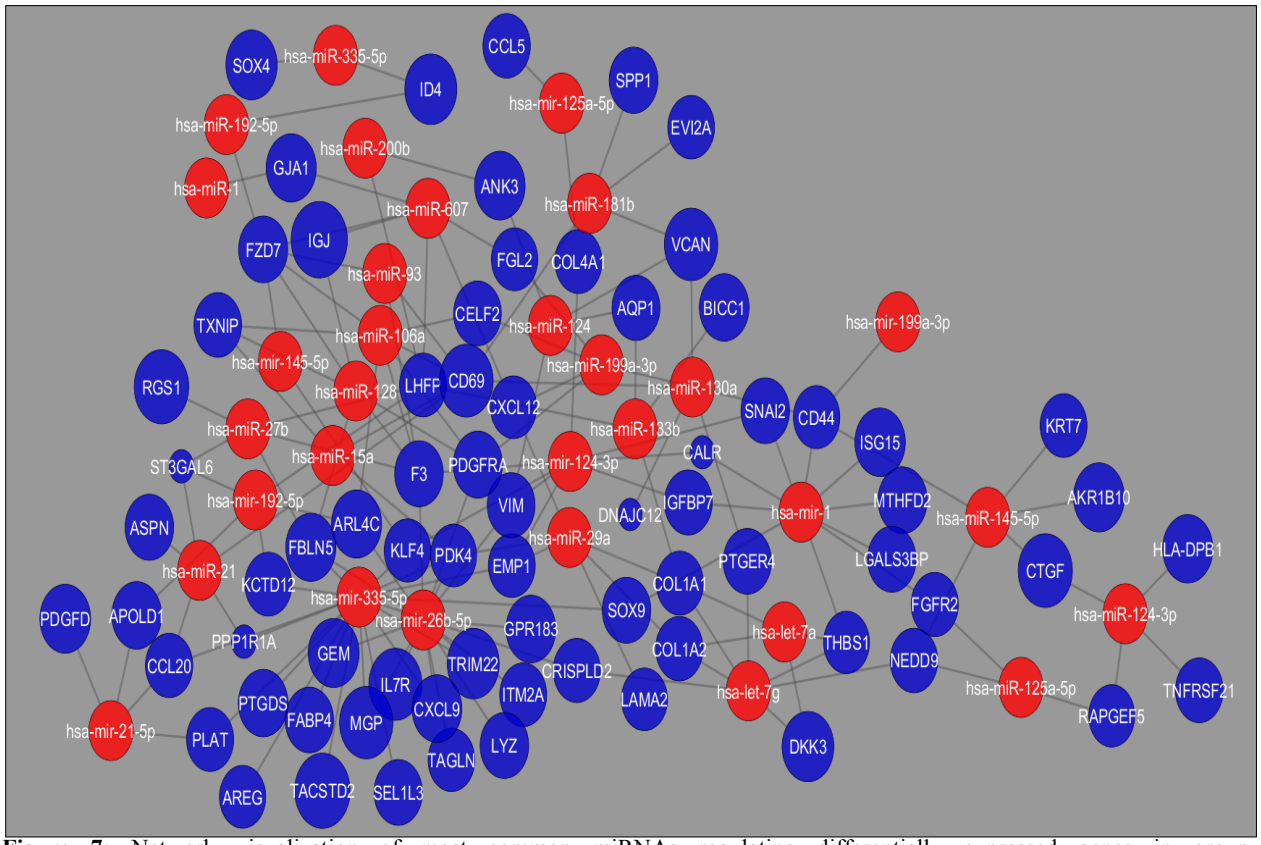

Figure 7: Network visualization of most common miRNAs regulating differentially expressed genes in group (Normal_Cirrhosis). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

Group 2 Normal_Cirrhosis HCC

The expression values in RMA format obtained from Normal sample cluster were analyzed against Cirrhosis HCC samples and a scatter plot was produced. In figure 8A, each gene has an expression value which corresponds to a specific circle in the plot. Those circles that are away from the red line correspond to genes that are differentially expressed. In the case of normal and cirrhosis HCC the majority of the differentially expressed genes are found below the red line. The scatter plot only indicates that there are differentially expressed genes, yet to test whether these genes are statistically and biologically significant a volcano plot was employed.

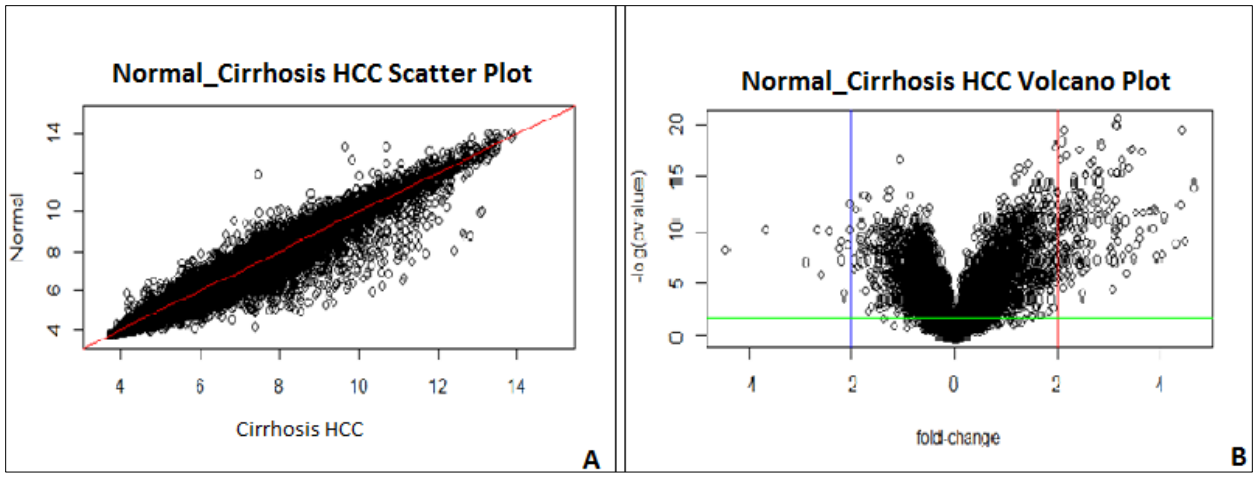

Figure 8: (A) Scatter plot of means of expression values for each gene in Normal group and Cirrhosis HCC group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)** Volcano Plot expresses the means of expression values for genes in the Normal_Cirrhosis HCC group. The fold cut off was at value of 2 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (normal and cirrhosis HCC) that are found away from the volcano represent both statistical and biological significance and are further filtered out

The volcano plot in figure 8B, displays the mean expression values for genes in both normal and cirrhosis HCC plotted against –log(p-value) and fold change. The fold cutoff for this analysis was 2 while the p-value was 0.01, however the –log(p-value) was plotted so that positive whole numbers were obtained and plotted. Here the majority of the genes with both biological and statistical significance lie to the right and left of the red and blue line, respectively. These genes were further filtered out in order to determine the transcription factors that facilitate their regulation.

Differentially expressed genes that were found to have both biological and statistical significance that had been filtered out, were further annotated with their respective transcription factors. Using different databases transcription factors were analyzed for a common expression pattern in their regulation of transcription amongst differentially expressed genes in Normal_CirrhosisHCC. Twelve transcription factors were found; AML1a, AP-1, ATF-2, c-Jun, CREB, C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure

9 a network was constructed to show the transcription factors and the differentially expressed genes that they regulate.

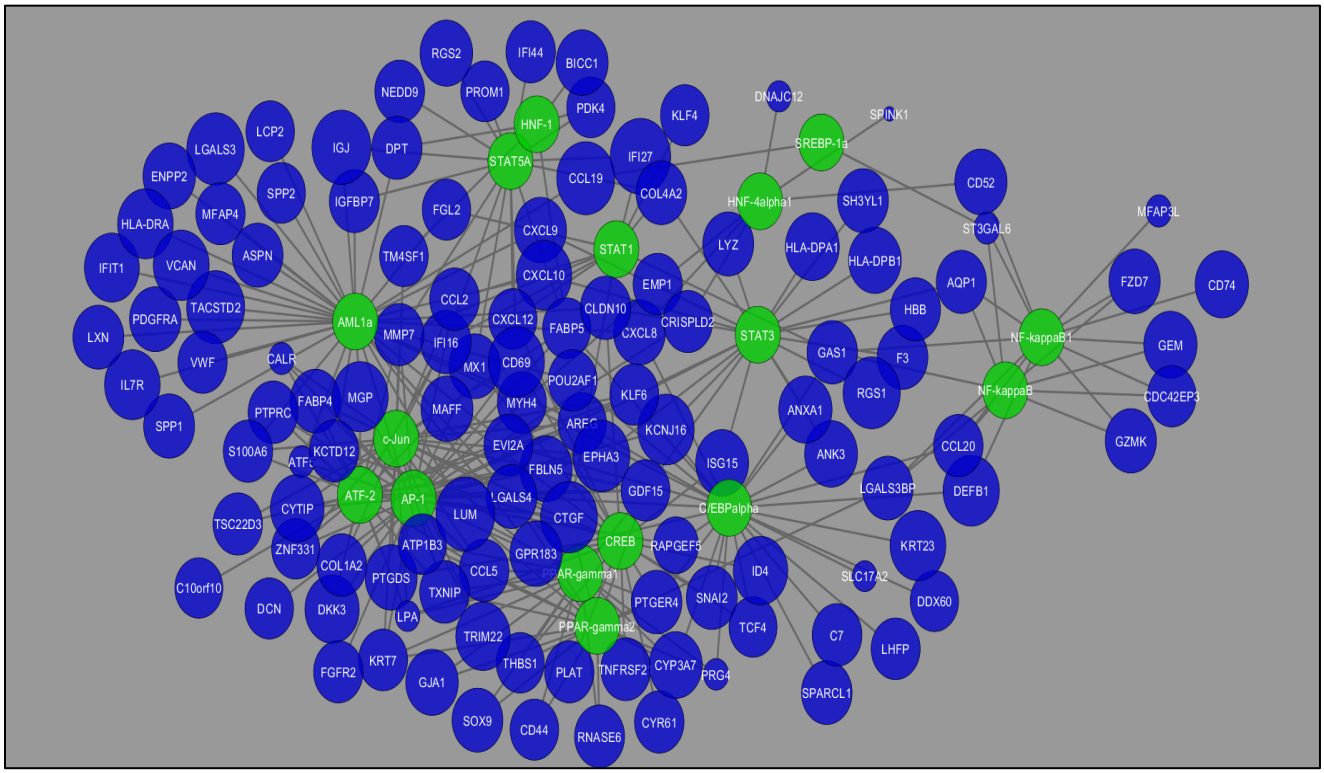

Figure 9: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in Group Normal_Cirrhosis HCC. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. A plethora of potential miRNAs were found to regulate genes in this group and due to the plethora of miRNA that are naturally found in the liver, miRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. Also, a review of the literature made it possible to determine the type of regulation

of these miRNAs observed in the presence of HCC. The network visualization in figure 10 shows the miRNAs that had the greatest expression pattern of regulation of genes in the group.

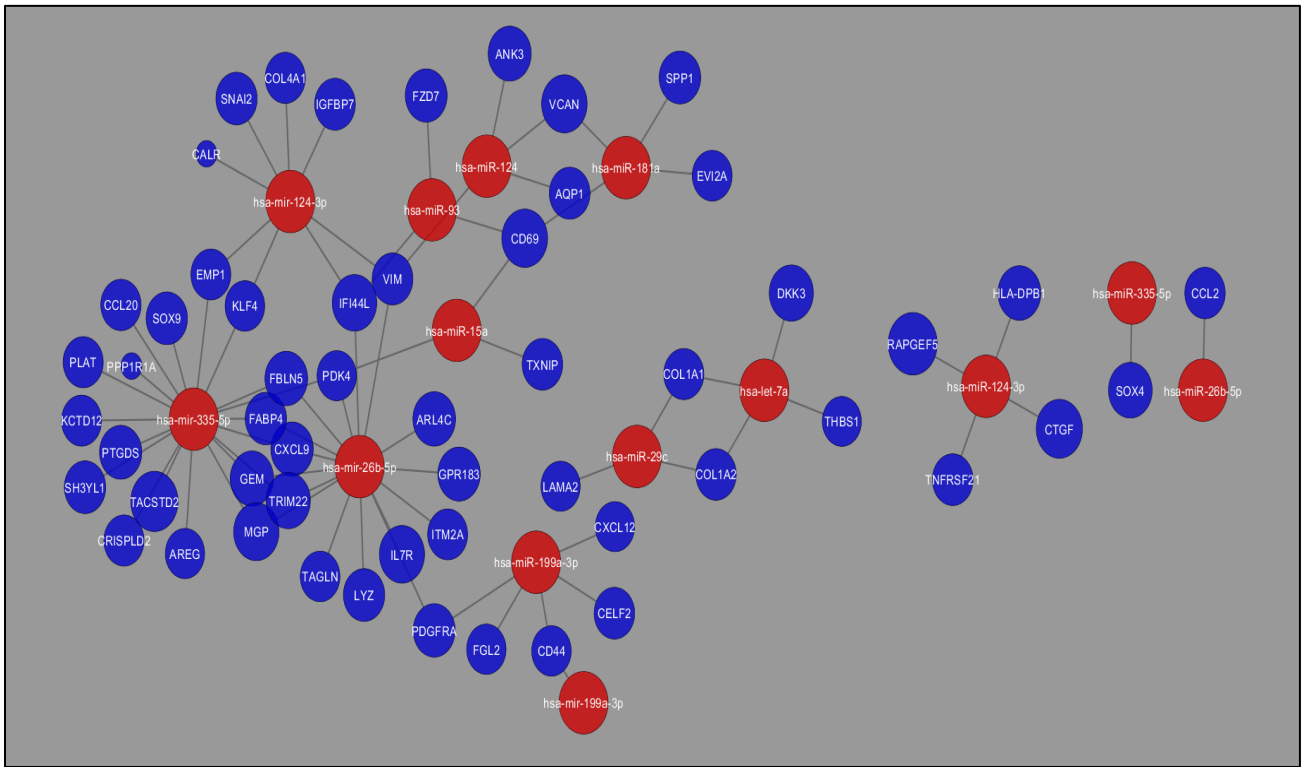

Figure 10: Network visualization of most common miRNAs regulating differentially expressed genes in group (Normal_Cirrhosis HCC). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

Group 3 Normal_ HCC

For group 3 a pairwise comparison for normal sample cluster was analyzed against HCC samples and a scatter plot was produced. In figure 11A, each gene has an expression value which corresponds to a specific circle in the plot. Those circles that are away from the red line correspond to genes that are differentially expressed. In the case of normal and HCC the majority of the differentially expressed genes are found below the red line. The scatter plot only indicates that there are differentially expressed genes, yet to test whether these genes are statistically and biologically significant a volcano plot was employed.

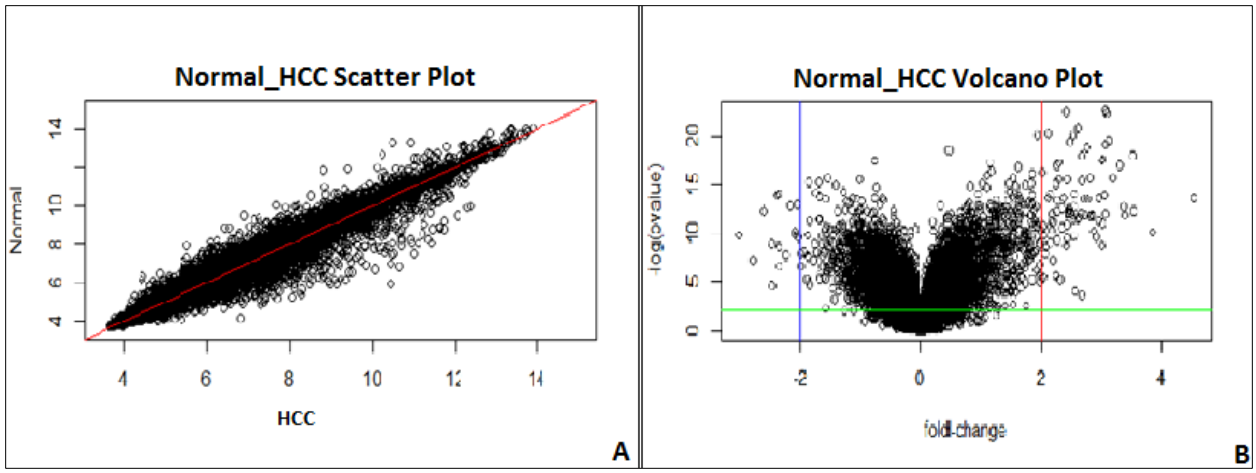

Figure 11: (A) Scatter plot of means of expression values for each gene in Normal group and HCC group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)** Volcano Plot expresses the means of expression values for genes in the Normal_ HCC group. The fold cut off was at value of 2 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (normal and HCC) that are found away from the volcano represent both statistical and biological significance and are further filtered out.

The volcano plot in figure 11B displays the mean expression values for genes in both normal and HCC plotted against –log (p-value) and fold change. The fold cutoff for this analysis was 2 while the p-value was 0.01, however the –log (p-value) was plotted so that positive whole numbers were obtained and plotted. Here the majority of the genes with both biological and statistical significance lie to the right and left of the red and blue line, respectively. These genes were further filtered out in order to determine the transcription factors that facilitate their regulation.

Differentially expressed genes that were found to have both biological and statistical significance that had been filtered out, were further annotated with their respective transcription factors. Using different databases these transcription factors were analyzed for a common expression pattern in their regulation of transcription amongst differentially expressed genes in Normal_HCC. Twelve transcription factors were found; AML1a, AP-1, ATF-2, c-Jun, CREB, C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure 12 a network was constructed to show those transcription factors and the differentially expressed genes that they regulate.

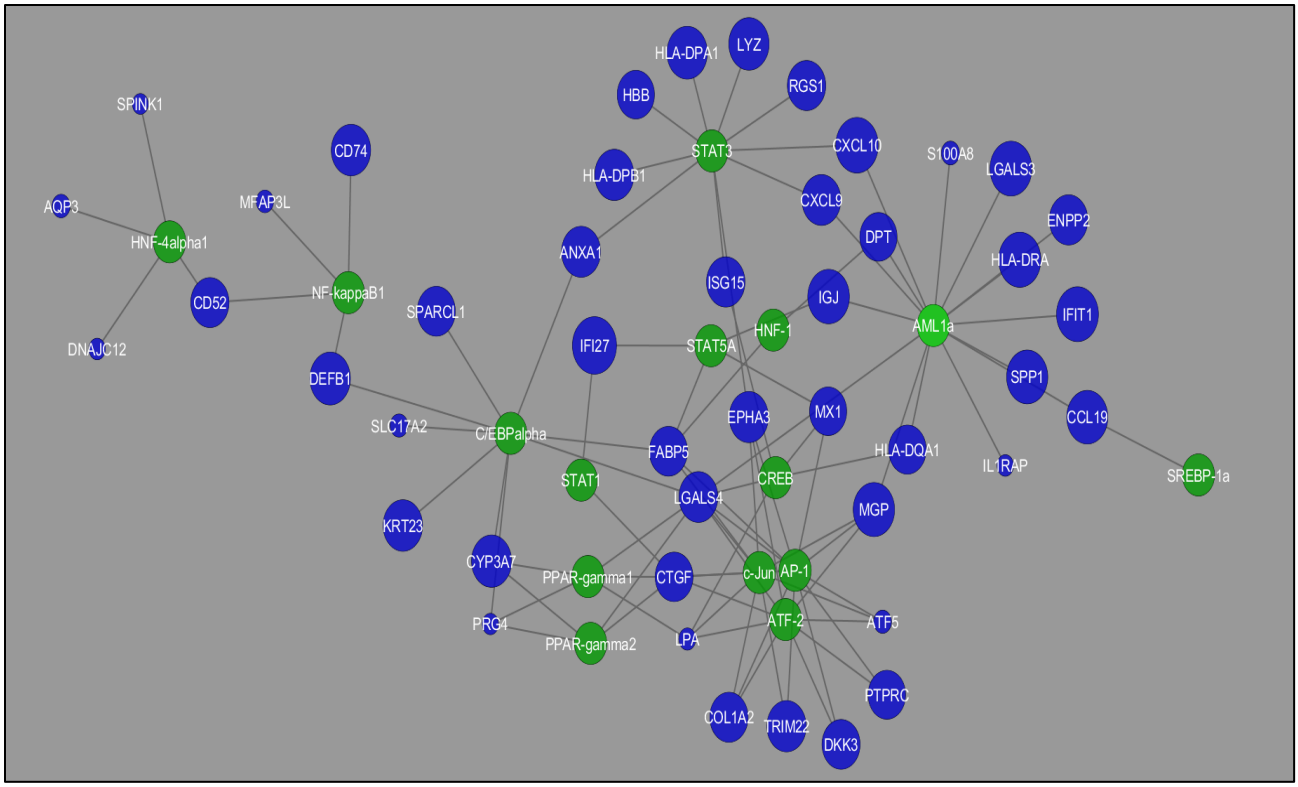

Figure 12: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in Group Normal_ HCC. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. MiRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. In figure 13, a network visualization shows the miRNAs that had the greatest expression pattern of regulation of genes in the group.

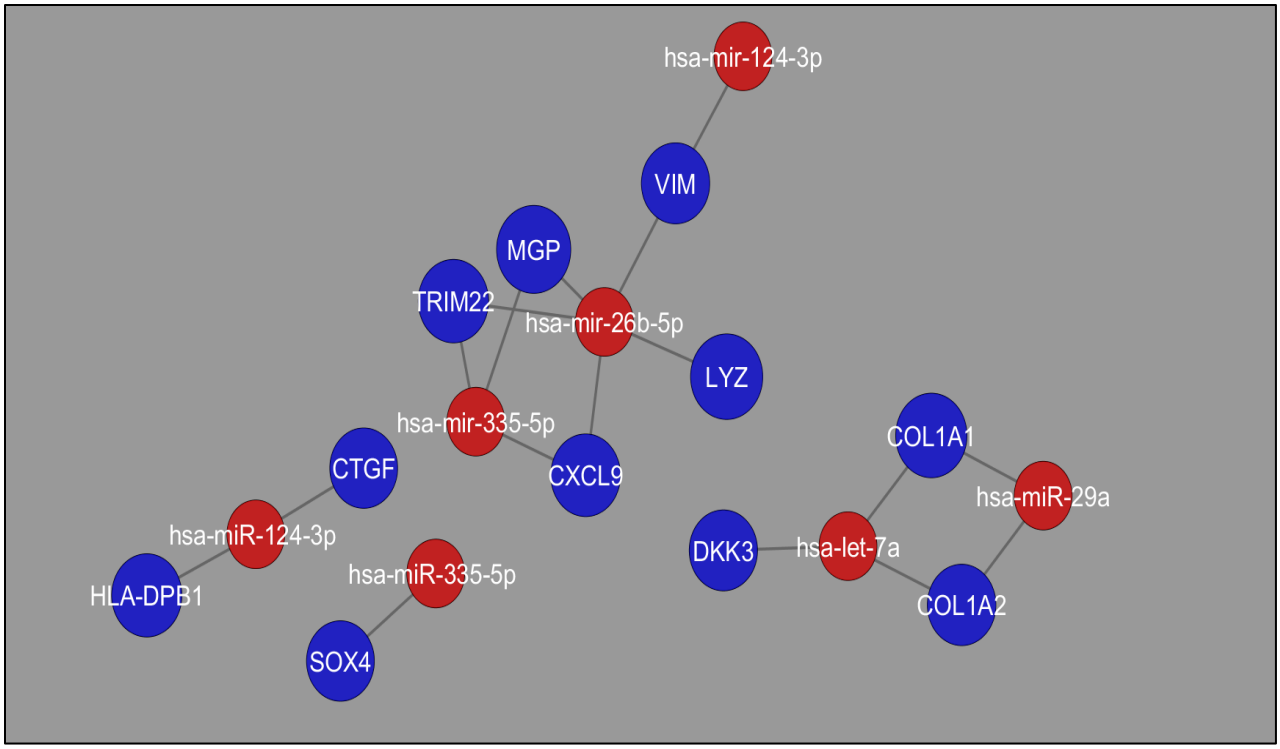

Figure 13: Network visualization of most common miRNAs regulating differentially expressed genes in group (Normal_ HCC). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

Group 4 Cirrhosis_ Cirrhosis HCC

In order to analyze how transcription regulation is influenced in stages to carcinoma each samples' clusters were analyzed in three groups the first of which cirrhosis_ cirrhosis HCC. Here, the expression values in RMA format obtained from cirrhosis sample cluster were analyzed against cirrhosis HCC samples and a scatter plot was produced. In figure 14A, each gene has an expression value which corresponds to a specific circle in the plot. Those circles that are away from the red line correspond to genes that are differentially expressed. In the case of cirrhosis and cirrhosis HCC the majority of the differentially expressed genes are found close to the red line with only a few scattered circles away from the red line. The scatter plot only indicates that there are differentially expressed genes, yet to test whether these genes are statistically and biologically significant a volcano plot was employed.

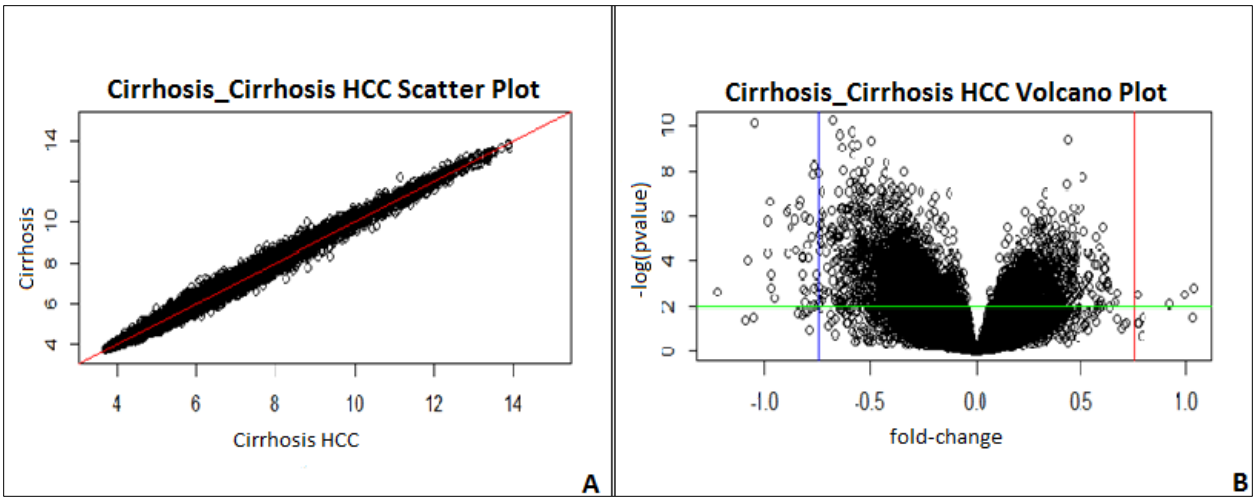

Figure 14: (A) Scatter plot of means of expression values for each gene in cirrhosis group and cirrhosis HCC group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)** Volcano Plot expresses the means of expression values for genes in the cirrhosis_ cirrhosis HCC group. The fold cut off was at value of 0.75 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (cirrhosis and cirrhosis HCC) that are found away from the volcano represent both statistical and biological significance and are further filtered out.

The volcano plot in figure 14B, displays the mean expression values for genes in both cirrhosis and cirrhosis HCC plotted against –log (p-value) and fold change. The fold cutoff for this analysis was 0.75 while the p-value was 0.01, however the –log (p-value) was plotted so that positive whole numbers were obtained and plotted. Here the majority of the genes with both biological and statistical significance are upregulated and lie to the right and left of the red and blue line, respectively and above the green line. Those circles that lie to the right and left of the red and blue line and below the green line are down regulated. These genes were further filtered out in order to determine the transcription factors that facilitate their regulation.

Differentially expressed genes that were found to have both biological and statistical significance that had been filtered out, were further annotated with their respective transcription factors. Using different databases these transcription factors were analyzed for a common expression pattern in their regulation of transcription amongst differentially expressed genes in cirrhosis_ cirrhosis HCC. Twelve transcription factors were found; AML1a, AP-1, ATF-2, c-Jun, CREB,

C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure 15, is a network showing those transcription factors and the differentially expressed genes that they regulate.

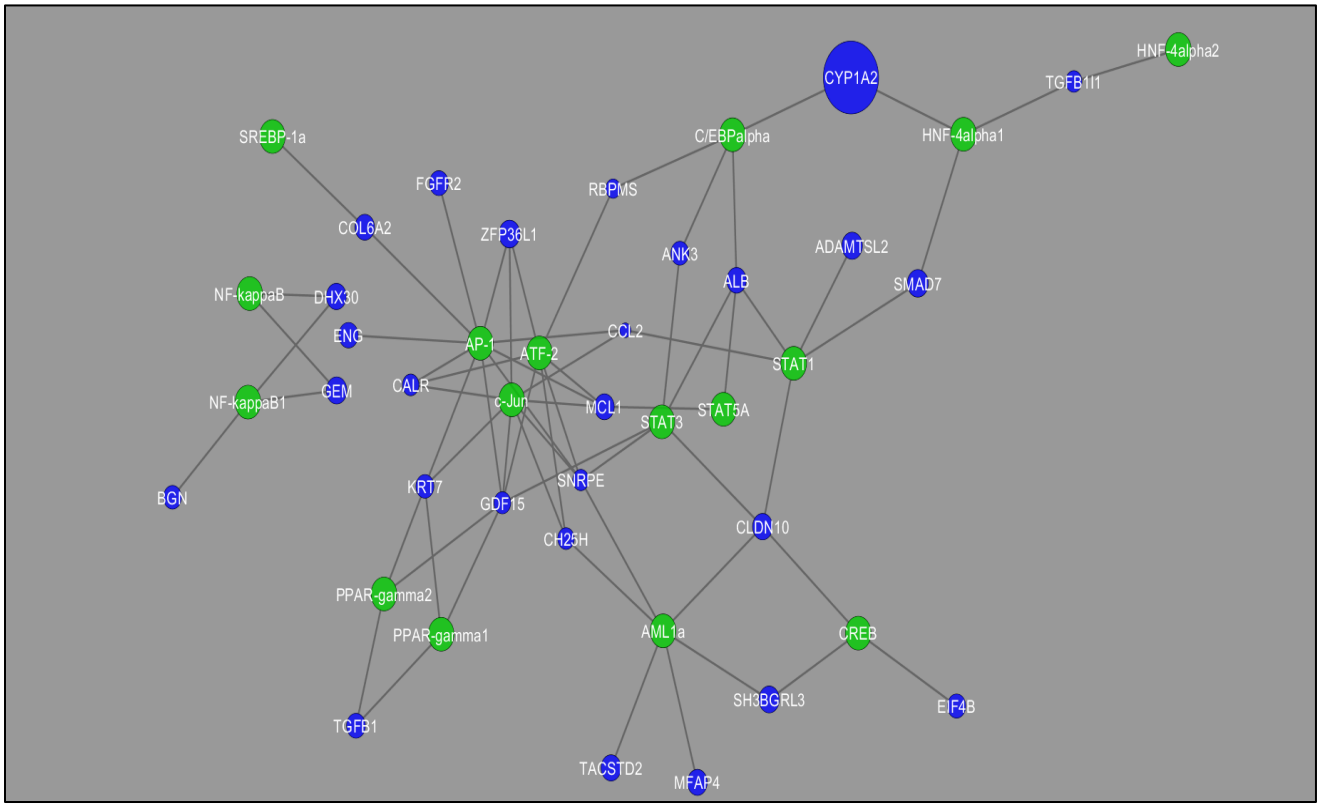

Figure 15: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in group cirrhosis_ cirrhosis HCC. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. MiRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. Also, a review of the literature made it possible to determine the type of regulation of these miRNAs observed in the presence of

HCC. The network visualization in figure 16, shows the miRNAs that had the greatest expression pattern of regulation of genes in the group.

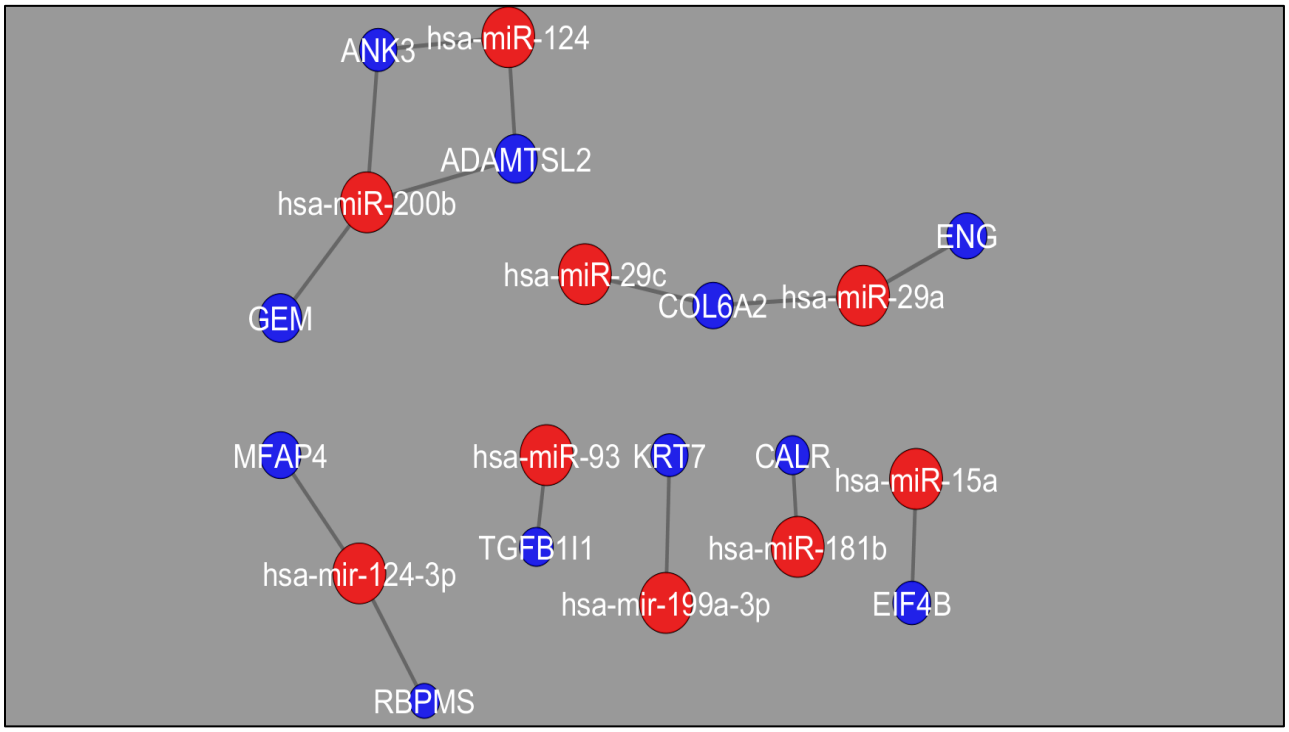

Figure 16: Network visualization of most common miRNAs regulating differentially expressed genes in group (cirrhosis_cirrhosis HCC). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

Group 5 Cirrhosis_HCC

Here, the expression values in RMA format obtained from cirrhosis sample cluster were analyzed against HCC samples and a scatter plot was produced. Each gene has an expression value which corresponds to a specific circle in the plot. Those circles that are away from the red line correspond to genes that are differentially expressed as shown in figure 17A. In the case of cirrhosis and HCC the majority of the differentially expressed genes are found close to the red line with only a few scattered circles away from the red line. The scatter plot in figure 17A, only indicates that there are differentially expressed genes, yet to test whether these genes are statistically and biologically significant a volcano plot was employed.

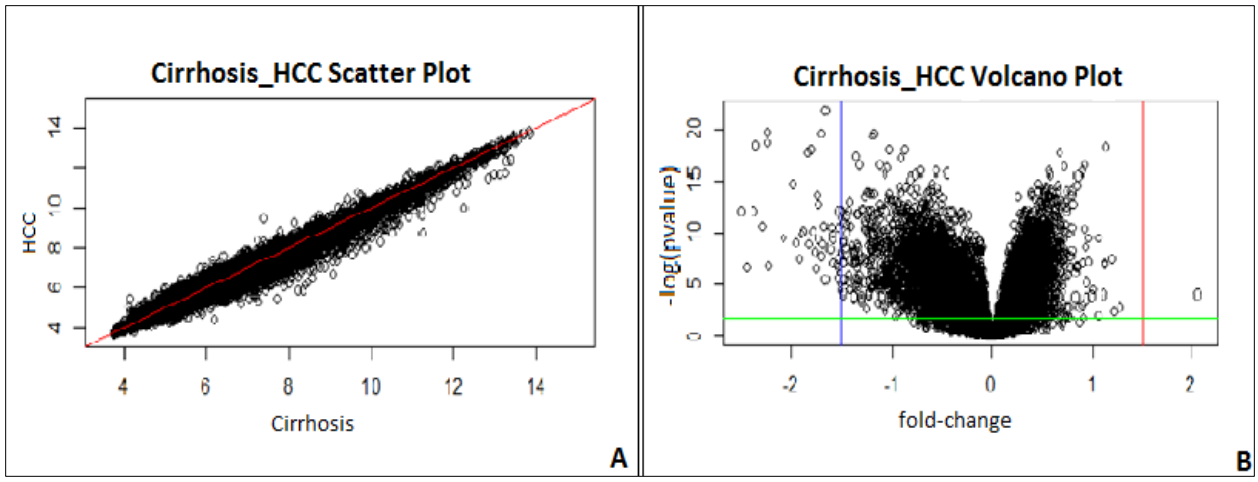

Figure 17: (A) Scatter plot of means of expression values for each gene in cirrhosis group and HCC group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)** Volcano Plot expresses the means of expression values for genes in the cirrhosis_HCC group. The fold cut off was at value of 0.75 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (cirrhosis and HCC) that are found away from the volcano represent both statistical and biological significance and are further filtered out.

The volcano plot in figure 17B displays the mean expression values for genes in both cirrhosis and HCC plotted against –log (p-value) and fold change. The fold cutoff for this analysis was 0.75 while the p-value was 0.01, however the –log (p-value) was plotted so that positive whole numbers were obtained and plotted. Here the majority of the genes with both biological and statistical significance are upregulated and lie to the right and left of the red and blue line, respectively and above the green line. Those circles that lie to the right and left of the red and blue line and below the green line are down regulated. These genes were further filtered out in order to determine the transcription factors that facilitate their regulation.

Differentially expressed genes that were found to have both biological and statistical significance that had been filtered out, were further annotated with their respective transcription factors. Using different databases these transcription factors were analyzed for a common expression pattern in their regulation of transcription amongst differentially expressed genes in Cirrhosis_HCC. Twelve transcription factors were found; AML1a, AP-1, ATF-2, c-Jun, CREB, C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure 18, is a

network showing those transcription factors and the differentially expressed genes that they regulate.

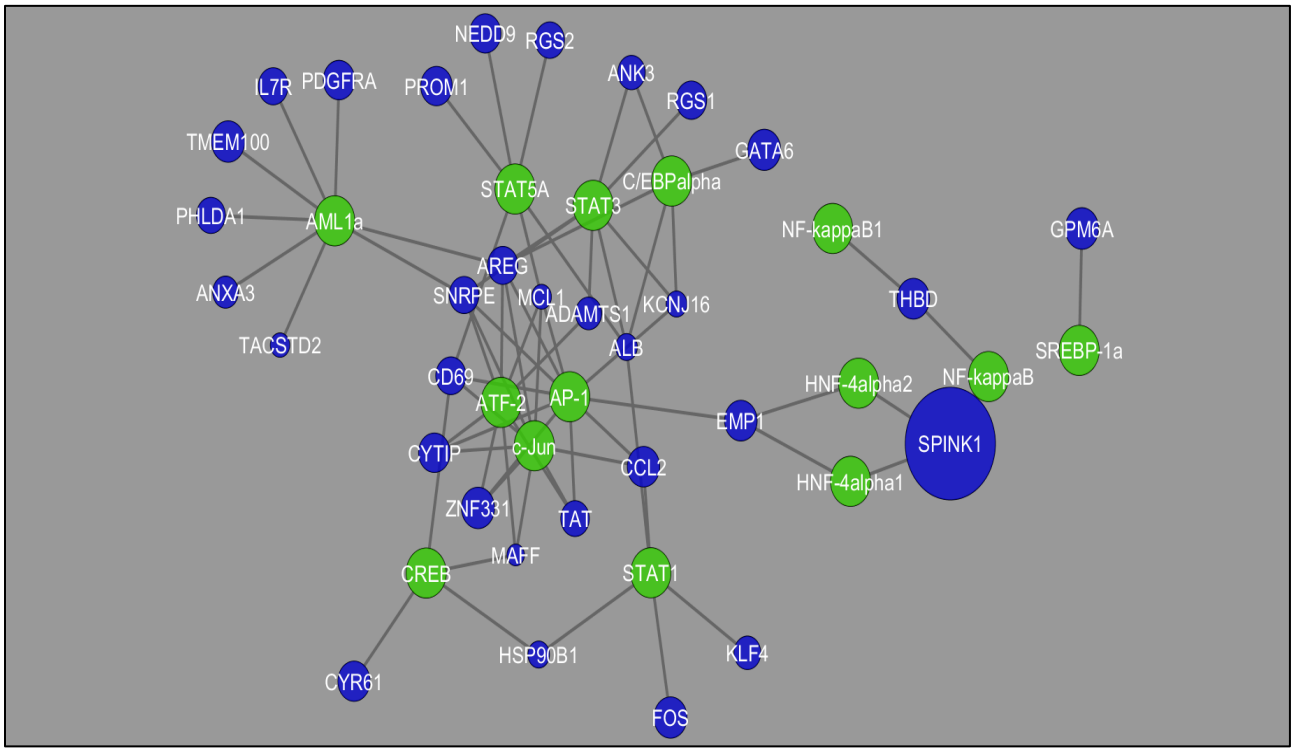

Figure 18: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in group cirrhosis_ HCC. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. A plethora of potential miRNAs were found to regulate genes in this group and due to the plethora of miRNA that are naturally found in the liver, miRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. Also, a review of the literature made it possible to determine the type of regulation of these miRNAs observed in the presence of HCC. The network visualization in figure 19, shows the miRNAs that had the greatest expression pattern of regulation of genes in the group.

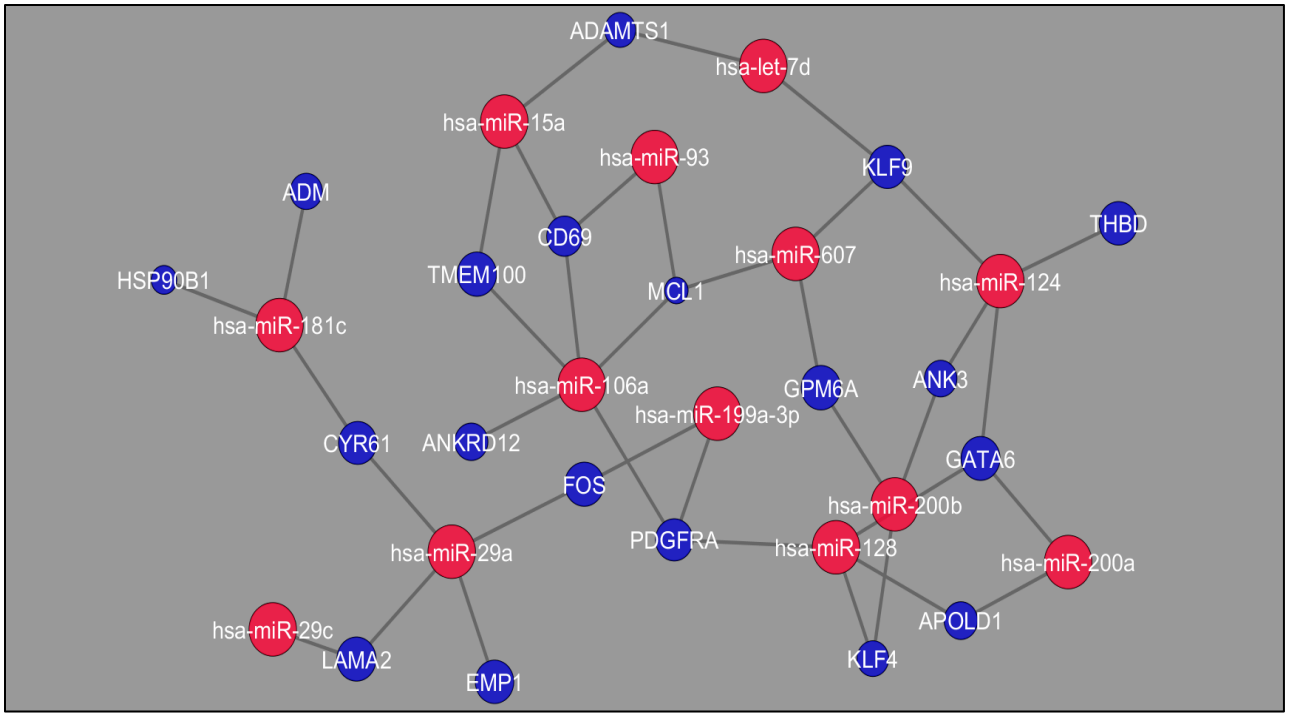

Figure 19: Network visualization of most common miRNAs regulating differentially expressed genes in group (cirrhosis_ HCC). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

Group 6 Cirrhosis HCC_ HCC

Expression values in RMA format obtained from cirrhosis HCC sample cluster was analyzed against HCC sample cluster and a scatter plot was produced. Each gene has an expression value which corresponds to a specific circle in the plot. Those circles that are away from the red line correspond to genes that are differentially expressed. In figure 20A, for cirrhosis HCC and HCC the majority of the differentially expressed genes are found close to the red line with only a few scattered circles away from the red line. The scatter plot only indicates that there are differentially expressed genes, yet to test whether these genes are statistically and biologically significant a volcano plot was employed.

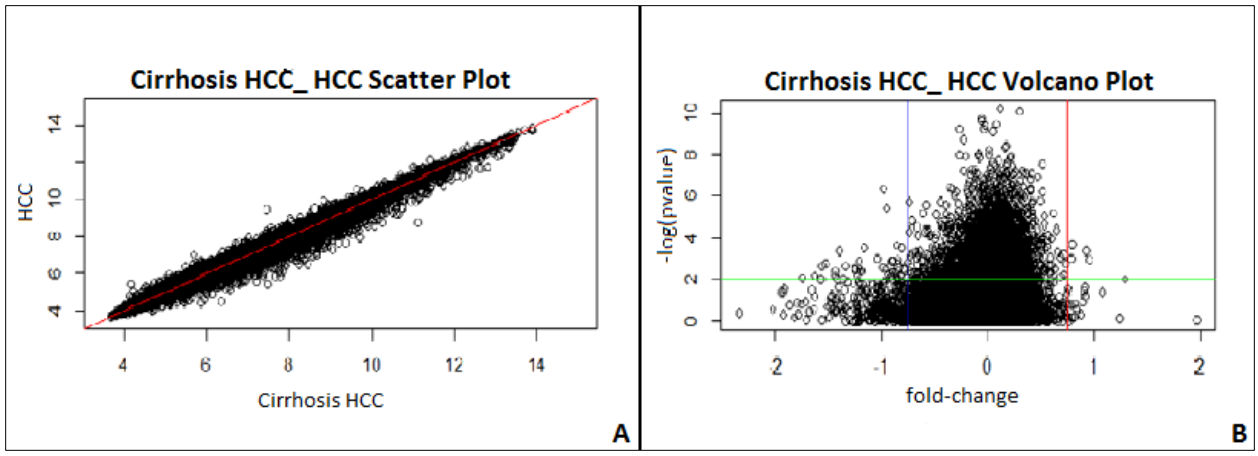

Figure 20: (A) Scatter plot of means of expression values for each gene in cirrhosis HCC group and HCC group. Each circle expresses an expression value for a gene; those circles that are found close to the red line are not differentially expressed genes while circles expressing values for genes away from the red line are differentially expressed genes. **(B)** Volcano Plot expresses the means of expression values for genes in the cirrhosis HCC_HCC group. The fold cut off was at value of 1.5 and the p-value that was used was 0.01. Circles representing expression values for genes in both groups (cirrhosis HCC and HCC) that are found away from the volcano represent both statistical and biological significance and are further filtered out.

The volcano plot in figure 20B displays the mean expression values for genes in both normal and HCC plotted against –log (p-value) and fold change. The fold cutoff for this analysis was 1.5 while the p-value was 0.01, however the –log (p-value) was plotted so that positive whole numbers were obtained and plotted. Here the genes with both biological and statistical significance are upregulated and lie to the right and left of the red and blue line, respectively and above the green line. The majority of the circles representing differentially expressed genes which lie to the right and left of the red and blue line and below the green line are down regulated. These genes were further filtered out in order to determine the transcription factors that facilitate their regulation.

Differentially expressed genes that were found to have both biological and statistical significance that had been filtered out, were further annotated with their respective transcription factors. Using different databases the transcription factors were analyzed for a common expression pattern in their regulation of transcription amongst differentially expressed genes in Cirrhosis HCC_HCC. Twelve transcription factors were found; AML1a, AP-1, ATF-2, c-Jun, CREB,

C/ebpalpha, HNF-1, HNF-4alpha, PPAR-gamma, STAT3, STAT5, and NF-Kß. In figure 21, is a network showing those transcription factors and the differentially expressed genes that they regulate.

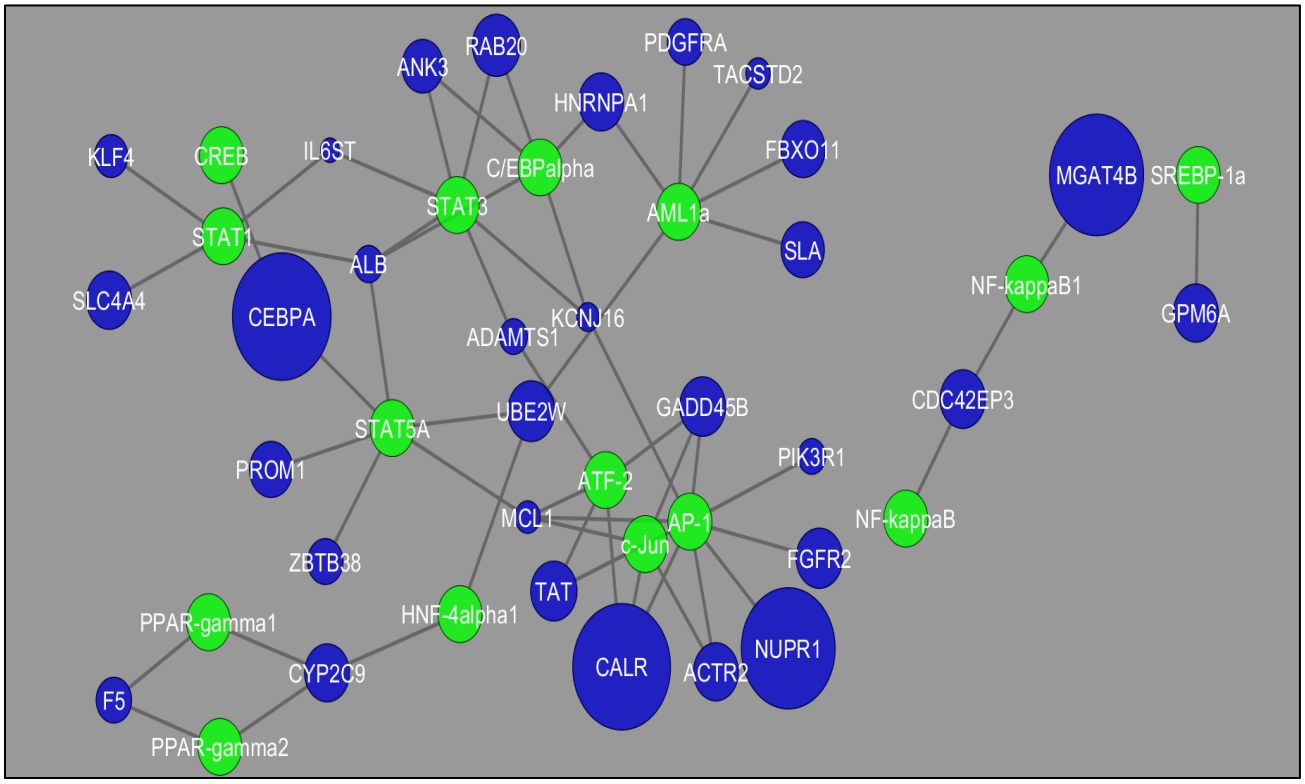

Figure 21: Network visualization of transcription factors that were found to be the most common in regulation of differentially expressed genes in group cirrhosis HCC_ HCC. The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The green circles represent the transcription factors that are regulating their respective differentially expressed genes.

Another database was used in order to determine miRNAs that regulate differentially expressed genes. This allowed for an annotation of differentially expressed genes with the miRNAs that regulate them. A plethora of potential miRNAs were found to regulate genes in this group and due to the plethora of miRNA that are naturally found in the liver, miRNAs that were chosen for further analysis were those that showed greatest expression pattern of genes that they regulated in the group. Also, a review of the literature made it possible to determine the type of regulation

of these miRNAs observed in the presence of HCC. The network visualization in figure 22, shows the miRNAs that had the greatest expression pattern of regulation of genes in the group.

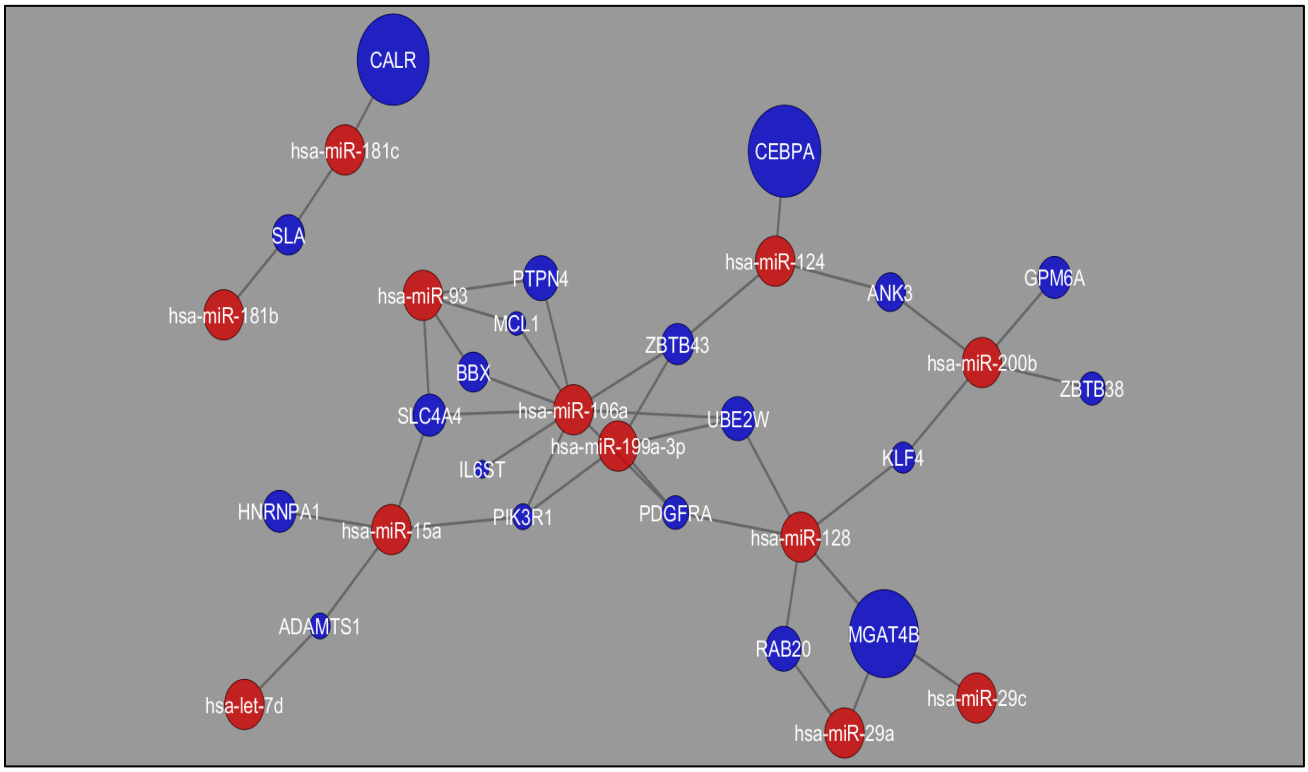

Figure 22: Network visualization of most common miRNAs regulating differentially expressed genes in group (cirrhosis_ HCC). The blue circles represent differentially expressed genes and their respective size is representative of their fold change value. The red circles represent the miRNAs that are regulating their respective differentially expressed genes.

In figure 23, the transcription factors with the highest expression in all 6 pairwise combinations were further analyzed through a systematic analysis of the literature that is present and it was determined that these transcription factors are involved in regulating HCV induced HCC into three stages angiogenesis, steatosis, and cancer induction. Also in figure 23, the miRNAs with the highest expression in all 6 pairwise combinations were analyzed similarly, and it was determined that the expression of miRNAs decreases as the liver digresses to carcinoma.

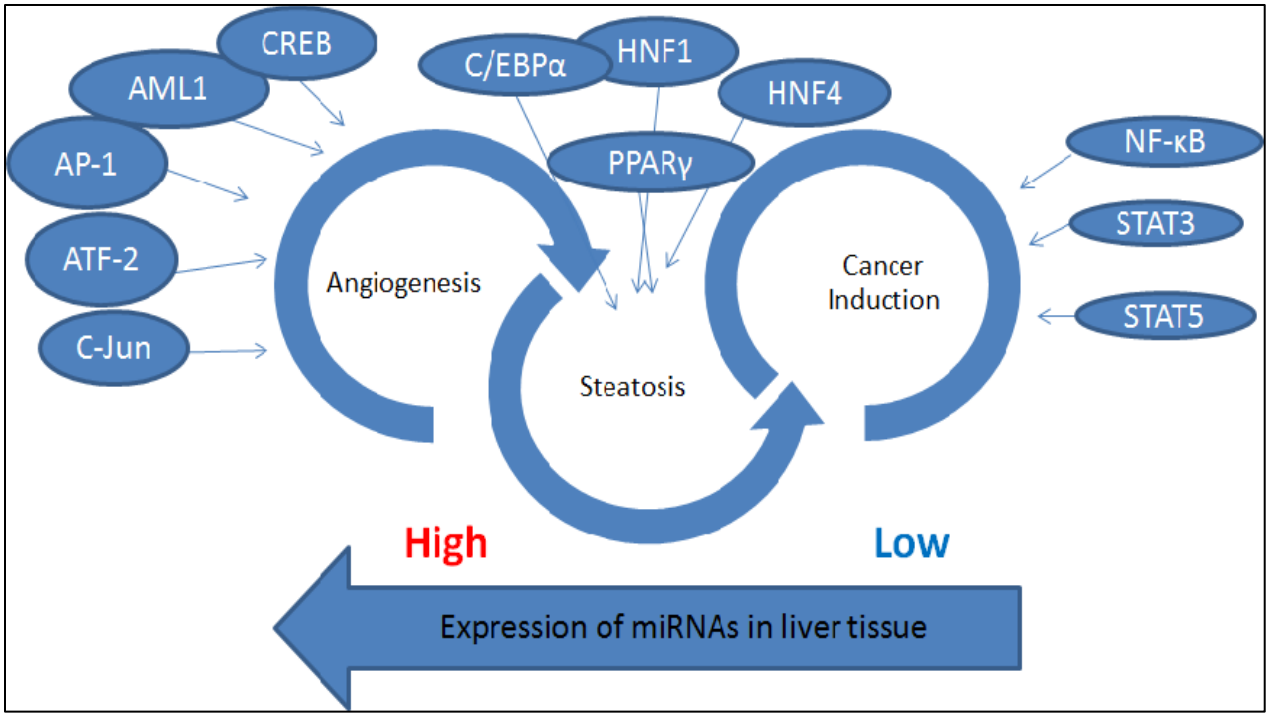

Figure 23: Overview of the transcription factors that regulate HCV induced HCC in each stage as the liver digresses to eventually becoming carcinomic; and the digression of miRNA expression in liver tissue as they digress to carcinoma.

However, in order to account for the differentially expressed genes that are regulated by the master regulators that were determined in this study and how they are biologically/ functionally organized another figure was produced. In figure 24, the differentially expressed regulated by their transcription factors in each stage in the digression of the liver is displayed so as to demonstrate the overall findings f this study and to illustrate the actions of all the key players in the transcription regulation of HCV-induced HCC.

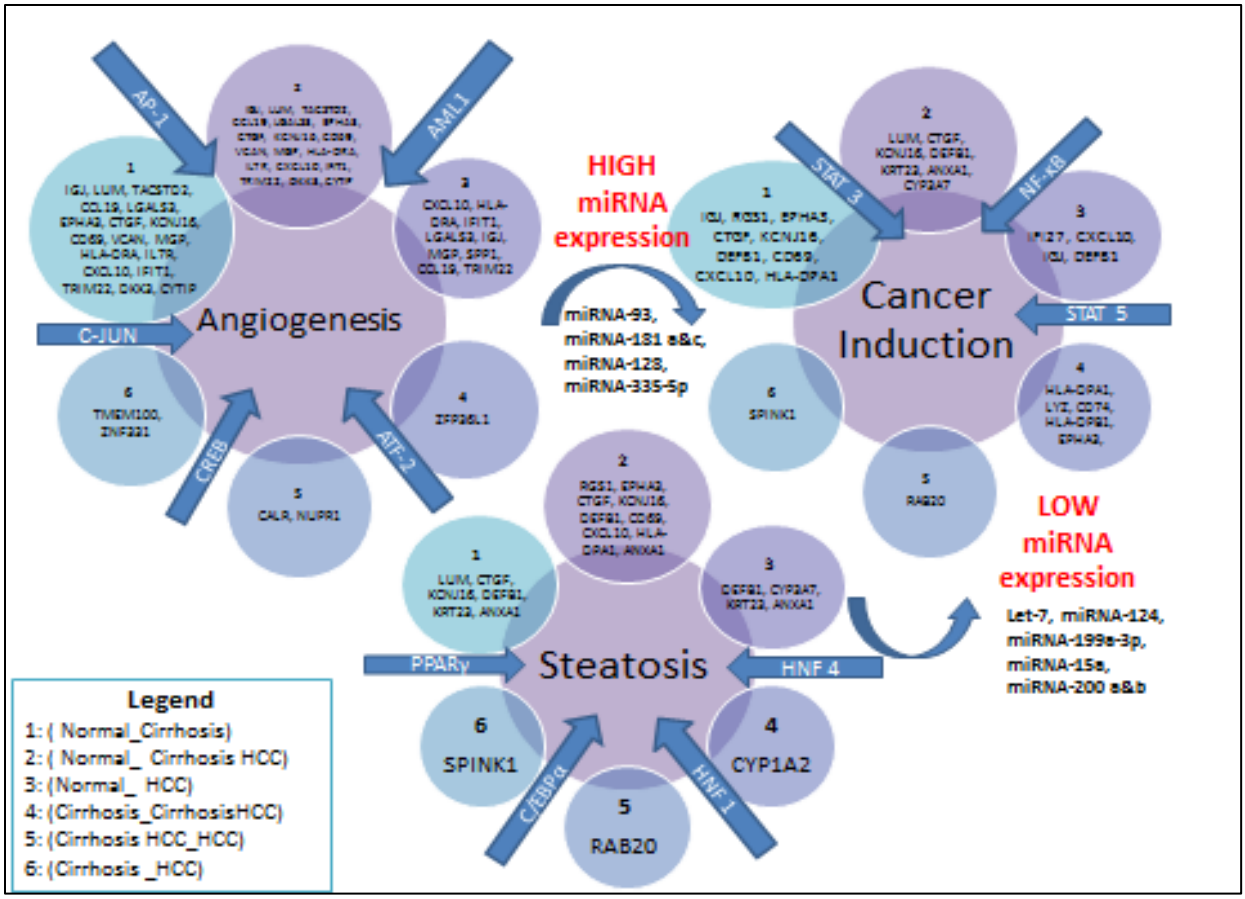

Figure 24: Transcription Regulation HCV-induced HCC. This is an illustration of the master regulators in HCV-induced HCC and their differentially expressed genes. The key genes and transcription factors associated with each condition of the liver is illustrated as they were found in each of the 6 pairwise comparisons, also with the general expression of miRNAs as the liver transitions from one condition to the next.

Chapter 5. Discussion and Conclusions

I. Differentially Expressed Genes

In the present study the aim was to identify the transcriptional regulation that gears differentially expressed genes in the progression of HCV induced HCC. This study highlights the molecular switches that may be involved in this progression and takes into consideration that transcriptional regulation is not only controlled by transcription factors but by other regulatory elements such as miRNAs that may influence the expression of genes. This analysis was conducted based on the findings of multiple studies, and therefore it must be noted that HCV induced HCC is a results of a combination of different regulatory elements both mRNA and miRNA in basis.

However, to begin it is vital to understand the differentially expressed genes that underlie this transcriptional regulation. In this study, after analyzing unique probe sets for every tissue type/ condition it was determined that the gene expression patterns varied significantly among the HCC and normal liver samples. These findings coincide with the study from which the microarray samples were attained Mas V.R. and colleagues, whose findings also show that genes associated with cell proliferation and mitosis had increased expression in HCC samples, while the genes that were expressed in lower levels were genes that are specifically expressed in differentiated hepatocytes. (Mas et al., 2009) Also, the findings in the present study coincide with the findings of Mas V.R. and colleagues for gene expression patterns amongst the 4 various tissue types/conditions. Yet, the DAVID analysis that was created for each pairwise comparison in the various tissue types/conditions show that between normal and HCC samples differentially expressed genes cluster functionally in signaling within the cell with an enrichment score of 9.13. (Figure 25) Again, here it becomes apparent that specificity in the functionality of genes is no longer present and the main function for differentially expressed genes in the late stage of

liver disease is signaling. Also, the functionality of the genes does not change amongst the various pairwise comparisons between normal liver and cirrhosis as well as cirrhosis HCC. In figure 26 and 27 below it becomes apparent that the primary function of differentially expressed genes in each pairwise comparison (cirrhosis and cirrhosis HCC) is signaling with an enrichment score of 13 and 18.31 respectively.

Figure 25: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between Normal and HCC samples**.** With an enrichment score of 9.13 the primary cluster is found to be with genes involved in signaling.

Figure 26: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between Normal and cirrhosis samples**.** With an enrichment score of 13 the primary cluster is found to be with genes involved in signaling.

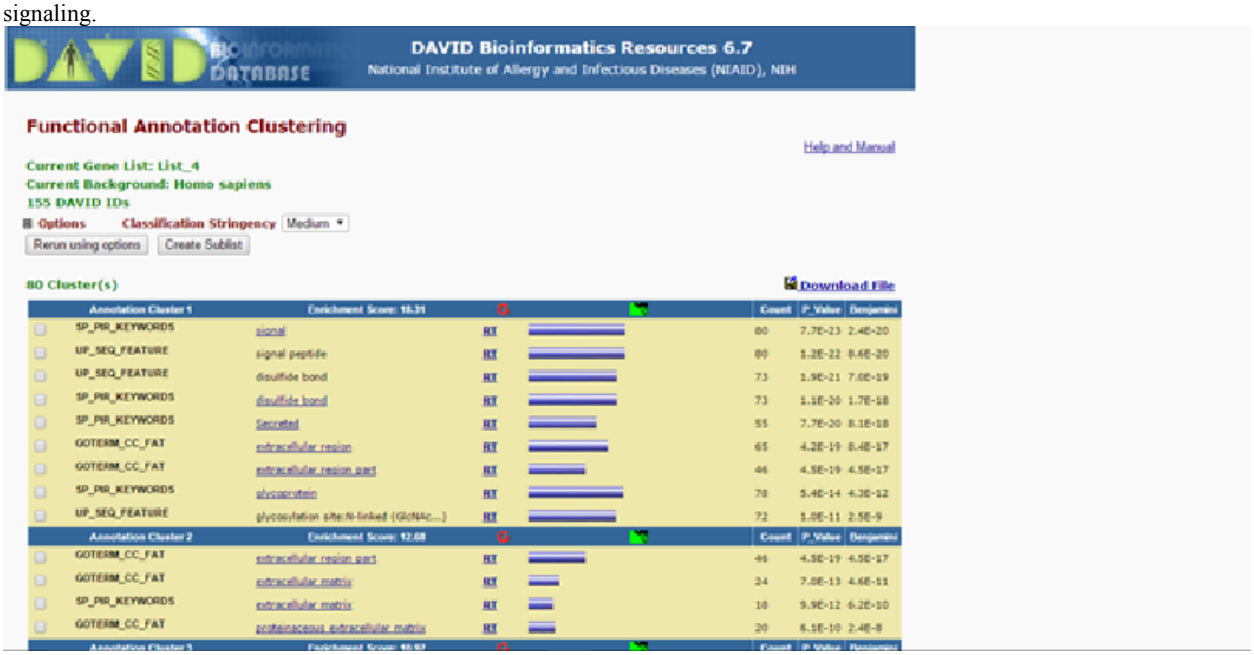

Figure 27: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between Normal and cirrhosis HCC samples**.** With an enrichment score of 18.31 the primary cluster is found to be with genes involved in signaling.

Since cirrhosis is an established risk factor for HCC the study by Mas V.R. and colleagues compared HCV-HCC samples with HCV cirrhosis samples they show that cell adhesion, cell division, and apoptosis were the more important cellular and molecular functions.(59) They also

found a lower number of differentially expressed genes in this pairwise comparison. These findings are similar to the findings of this study in that the DAVID analysis in figures 28, 29, and 30 that was used to compare cirrhosis samples to HCC in the different stages also was able to find that the primary functional clusters were involved in cell attachment, as well as key words involved in cell division and apoptosis.

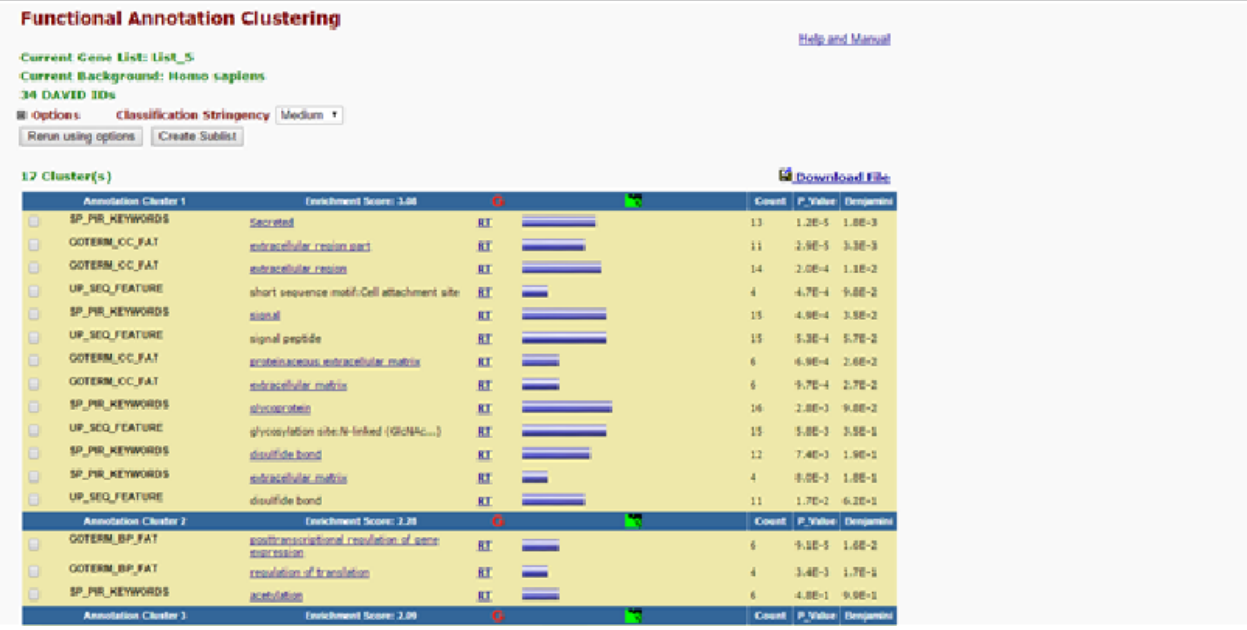

Figure 28: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between cirrhosis and cirrhosis HCC samples**.** With an enrichment score of 3.08 the primary cluster is found to be with genes involved in secretion and cell attachment.

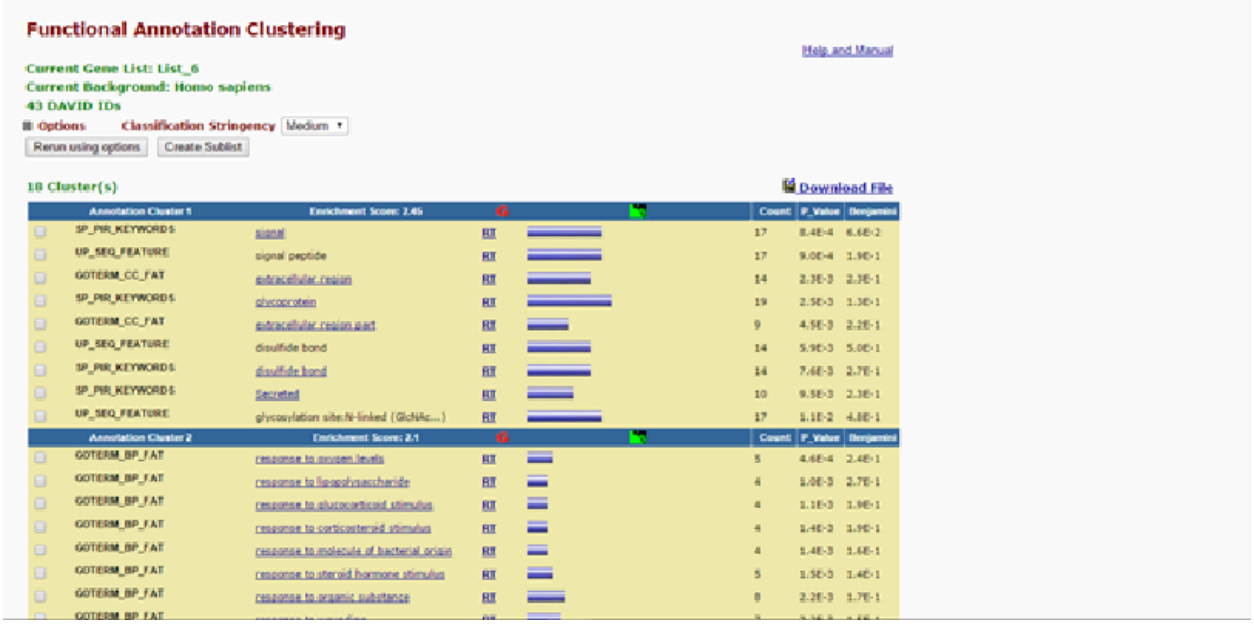

Figure 29: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between cirrhosis and HCC samples**.** With an enrichment score of 2.46 the primary cluster is found to be with genes involved in secretion and signaling.

Figure 30: DAVID analysis, functional annotation clustering of differentially expressed genes in the pairwise comparison between HCC and cirrhoisHCC samples**.** With an enrichment score of 1.23 the primary cluster is found to be with genes involved in response to stimulus and protein activity.

In this study it was also observed that gene expression related to normal liver function significantly decreased across tissue types from normal liver, to cirrhosis, to cirrhosis HCC, to

HCC; which coincides with the findings of (Mas et al., 2009). It was also observed that each pathological stage in the liver progression to HCC in the presence of HCV infection is marked by gene signatures; which are dictated by the master regulators that control their expression. Amongst the main signaling pathways that were found to be prevalent was the JAK/STAT signaling pathway, which Mas V.R. and his group noted to represent the more important downregulated canonical pathways. This pathway is one of the primary pathways that once downregulated tends to induce cancer. (Mas et al., 2009)

The David analysis above for each of the pairwise comparisons coincides with the findings of Mas V., etal in that there are gene signatures for each pathological tissue type in the progression to HCC. Gene signatures, are specific sets of genes with a similar biological function with high expression in each biological tissue type. Biologically, it can be observed that these gene signatures change from normal to cirrhotic tissue as genes struggle to maintain normal functioning such as cell adhesion and cell signaling. Then, as the condition of the liver digresses the gene signatures for cirrhosis HCC and HCC tissues changes as well as cells struggle to stay alive and therefore genes that are more prevalent are those involved in proliferation and apoptosis. It is also important to note that differentially expressed genes in each pairwise comparison show a progression of the condition of the liver. As the condition of the liver progresses to carcinoma the regulation of these genes is determined not only by the condition of the liver but also by the presence of regulatory elements underlying the command of that gene and in turn controlling the final state of the liver. These regulatory elements are transcription factors and miRNAs that underlie the switching on/off of these genes and in turn controlling the molecular biology of the condition of the liver as it progresses to HCC.

II. Transcription Factors

As HCV infection aggregates more and more within the liver, the condition of the liver worseness. As the liver digresses from a cirrhotic liver, to steatosis of the liver, and eventually becoming cancerous. Although, this sequence of events is not dogma and the liver may digress from any one of these conditions directly to carcinoma, if the attributes that satisfy the conditions for each event are present the liver will digress in this sequence and eventually become carcinomic. Each one of these conditions entails a different molecular key switch as different genes are turned on and off to substitute for the digressing functionality of the liver. As previously noted in each of the pairwise comparisons of the liver it was found that cell signaling as well as other important cellular functions were of the most common functionalities effected as the condition of the liver digresses from normal to cancerous in the presence of HCV infection. It is important to note that also of the keywords that arise in the functional annotation clustering tool (DAVID) is the keyword "fat"; this also is of value in that it shows that the functionality of the differentially expressed genes is also linked to lipid metabolism.

HCV infection relies heavily on lipid metabolism in the liver for viral entry and replication. Assembly of HCV virion is brought about by the synchronized actions of HCV structural proteins along with the viral replication complex, and takes place in close proximity to membrane bound lipid droplets (LD). Lipid droplets (LDs) are intracellular organelles storing cholesterol and triglycerides, and play an essential role in viral packaging. Inhibition of the synthesis of their lipid components blocks viral assembly. In the blood HCV virions circulate as lipo-viro-particles (LVPs) which are found to be associated with very low or low-density lipoproteins (VLDLs and LDLs). These LVPs may also associate with high levels of

triglycerides, Apolipoprotein (Apo) E, ApoC, ApoB and low levels of cholesterol or phospholipids (17). This HCV virion association with lipids may facilitate its infection of hepatocytes which are the primary target for infection while at the same time aiding in its escape from neutralizing antibodies. (18) This in turn would explain the association between differentially expressed genes and the keyword FAT in the functional annotation clustering of genes.

Annotation of the differentially expressed genes with their respective transcription factors coincides with the functional annotation of the differentially expressed genes in that it was found that main transcription factors involved in regulation of the differentially expressed genes were marked by 3 hepatic conditions; hepatic angiogenesis, hepatic steatosis, and hepatic carcinoma. In figure 4 above it becomes apparent that many genes involved in normal liver function are being regulated by the transcription factors involved in each hepatic condition. However as the number of genes in normal liver function decreases the master transcriptional regulators dictate the expression of specific genes involved in other pathways such as cell proliferation this is observed in figure 7, 10, 13, 16, and 19. Each one of these hepatic conditions calls upon a different key switch of genes and transcription factors however, it explains the digression of the liver to carcinoma and foretells the possible points at which molecular intervention may be key to stop this digression and possibly save the liver or freeze it in its current state.

a. Hepatic Angiogenesis

Angiogenesis and the disruption of liver vascular architecture have been linked to progression to cirrhosis and liver cancer (HCC) in chronic liver diseases, which contributes both to increased hepatic vascular resistance and portal hypertension and to decreased hepatocyte perfusion. (60)

Angiogenesis was also found to be important in adult tissue repair. (61)In the present study, 5 main transcription factors were determined to have a common expression pattern amongst all 6 pairwise comparisons. These 5 transcription factors may be the key switches in the effect of transcriptional regulation of differentially expressed genes; AP-1, C-JUN, AML1, CREB, and ATF-2. From the literature it was apparent that these 5 transcription factors regulate genes that are involved in hepatic angiogenesis.

Angiogenesis is implicated in cancer development by favoring progression, growth, and metastasis of cancer and is regulated by growth factors, including transforming growth factor-beta (TGF-ß), basic fibroblast growth factor, or platelet-derived growth factor. In tumor cells induction of hepatic angiogenesis during chronic HCV infection may contribute to the early development of cancer cells. (62)

One of the 5 transcription factors acute myelogenous leukemia 1 (AML1)/runt-related transcription factor 1 (Runx1) belongs to a family of transcriptional regulators called Runx. Gene-targeting studies in mice have demonstrated that AML1 is essential for early development of definitive hematopoiesis. (63) A study by Takakura and colleagues was able to show that AMl1 deficient embryos lack definitive hematopoiesis and thus show defective angiogenesis in the head and pericardium. (61) In a study by Hassan M. and colleagues the mechanistic role of HCV core protein is better understood as they provide evidence for its role in the regulation of hepatic angiogenesis during the course of HCV infection and demonstrate its ability to trigger the production of both TGF-ß2 and VEGF proteins by multiple pathways including PKC, RB/E2F1, ASK1-JNK/p38, and ERK. These pathways include transcription factors that were found to be the main transcription regulators of differentially expressed genes. (62) Of these pathways, the PKC pathway was found to turn on CREB and ATF-2, while ASK1-JNK/p38 was found to

control AP-1. AP-1 is an important target for the JNK signaling pathway and is composed of Jun , Fos, and related bZIP subunit. (64)

As the liver vascular architecture changes and the liver proceeds in the direction of fibrosis another condition starts to develop as well; hepatic steatosis. This condition may be in part as a reaction to the changes that are already taking place in the liver as the liver may call upon lipids for the increasing viral particles and replication that is present due to HCV infection. Rutkowski D.T. and colleagues suggest that one of the possible reasons behind steatosis is the suppression of a subset of metabolic transcription factors that regulate lipid homeostasis as a direct response to ER homeostasis.(65) The exact mechanism of this digression from hepatic angiogenesis to hepatic steatosis is unknown, however it is apparent that each condition hampers the functionality of the liver leading to carcinoma.

b. Hepatic Steatosis

Hepatic steatosis is common amongst patients with HCV as HCV core protein play an important role in the development of hepatic steatosis in HCV infection. HCV core protein has an elevated effect on the transcriptional activity of PPARγ (peroxisome proliferators-activated receptor γ). This transcription factor, PPARγ as well as HNF1, HNF4, and C/EBPα were found to be amongst the transcription factors with the highest expression patterns regulating differentially expressed genes in each of the 6 pairwise comparisons involved in this study.

PPARγ is a master regulator for adipocyte differentiation and is important in the regulation of a number of genes involved in fatty acid and glucose metabolism. Thus down regulation of liver PPARγ which contributes to regulation of lipid synthesis, transport, and storage within hepatocytes, causes the development of hepatic steatosis.(66) CCAAT/enhancer-binding protein

alpha (C/ebpα) another transcriptional regulator found to have high expression pattern in differentially expressed genes has been shown to be involved in regulating gluconeogenesis and lipogenesis while being downregulated in response to hepatic steatosis.

Hepatocyte nuclear factors 1 and 4 (HNF1 and HNF4) are a group of transcription factors that are involved with glucose, cholesterol, and fatty acid transport and metabolism. HCV induces HNF1 and HNF4 due to increased oxidative stress and direct protein-protein interactions between HCV non-structural component (NS) 5A and HNF1. Again, some HCV proteins, particularly the structural capsid protein, core, and the non-structural protein, NS5A, can induce hepatic steatosis by interfering with intracellular lipid metabolism. (67) Alteration of expression of these genes might be related to abnormal expression of metabolic enzymes again leading to hepatic steatosis. (68)

Univariate and multivariate analyses identified hepatic steatosis, together with aging, cirrhosis, and lack of IFN treatment, as significant independent risk factors for HCC. Several studies have revealed that hepatic steatosis, including steatosis induced by HCV core protein, predisposes to lipid peroxidation and excess free-radical activity with the potential risk of genomic mutations. Hepatic steatosis is an independent risk factor for HCC in patients with chronic HCV infection, although the factors responsible for steatosis could not be identified clearly. (69)

c. Cancer Induction

The production of a great number of cytokines, chemokines and growth factors, favoring increased cellular proliferation accompany the continuous cell death and inflammatory cell infiltration during cancer development. This continuing hepatocyte death triggers liver repair and

regeneration and eventually leads to severe liver fibrosis or cirrhosis. Multiple signaling pathways are involved in this injury-inflammation-regeneration response and in human HCC development. In the present study 3 transcription factors were found to have high expression patterns amongst all 6 pairwise comparisons as a result of prolonged HCV infection; these transcription factors include NF-κB, STAT3, and STAT 5.

Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) is a protein complex that regulates cell survival, immunity and inflammation, and is one of the more important pathways that is activated during liver injury and inflammation. One of the main functions of NF-κB in hepatocytes appears to be the production of cytokines that maintain the inflammatory microenvironment in which tumors develop.(70)

Signal transducer and activator of transcription 3 (STAT3) is another transcriptional factor involved in immune responses, inflammation and tumorigenesis, and was found to be critical for compensatory liver regeneration and chemically-induced HCC development. STAT3 belongs to the signal transducer and activator of transcription (STAT) family. Like its relatives, STAT3 is inactive in non-stimulated cells, but is rapidly activated by various cytokines and growth factors, such as IL-6 and EGF family members, as well as hepatocyte growth factor (HGF) STAT3 in cancer cells is activated by cytokines and growth factors that are produced within the tumor microenvironment. The expression of IL-6, one of the major STAT3-activating cytokines, and is elevated in human liver diseases and HCC. (71)

Signal transducer and activator of transcription 5 (STAT5) proteins are involved in cytosolic signaling and in mediating the expression of specific genes. While STAT5 activation plays an important role in promoting tumorigenesis via the upregulation of anti-apoptotic, cell proliferative, and invasion of metastasis-related genes. (71)

III. miRNAs

The molecular mechanisms underlying transcriptional regulation of miRNA genes in the liver remain largely unknown. Whether transcription factors are also involved in the transcriptional regulation of miRNAs in the liver is unclear since the functional expression of transcription factors can also be regulated by miRNAs. Yet, it is clear that miRNAs can potentially regulate every aspect of cellular activity and a diverse spectrum of liver functions. (54)

In the present study it was observed that each tissue type or condition had its own respective signature miRNA pattern, however there was a commonality amongst all 6 pairwise comparisons that were analyzed as shown in the table 1. As was observed for the expression of genes regulated by the master transcription factors involved in each hepatic condition, the number of genes that are regulated by miRNAs decrease as the number of genes involved in normal liver function decreased. This can be observed in the networks above in figures 8, 11, 14, 17 and 20. However unlike transcription factors miRNAs did not show a functional relationship between the genes that they regulate; again this expresses the specificity by which miRNAs regulate the expression of genes. In table 2 the miRNAs that had the highest expression were annotated with their predicted differentially expressed target genes in all 6 pairwise comparisons. After analysis using functional annotation clustering it becomes apparent that the differentially expressed genes

that miRNAs regulate are involved in maintaining cellular homeostasis and the up keeping of fat metabolism. Therefore, as cellular homeostasis becomes imbalanced and fat metabolism is disrupted the expression of miRNAs decreases as the liver digresses to carcinoma.

Table 3: The miRNAs with the highest expression in all 6 pairwise combinations were annotated with their differentially expressed target genes.

miRNAs	Predicted Target Genes
Let-7	THBS1, COL1A2, COL1A1, DKK3, ADAMTS1, KLF9
miR-124-3p	HLA-DPB1, COL4A1, TNFRSF21, CTGF, RAPGEF5, CALR, EMP1, IGFBP7, SNAI2, VIM, F3, KLF4, EMP1, IFI44L, MFAP4, RBPMS
miR-124	ANK3, AQP1, VCAN, VIM, THBD, KLF9, GATA6, ZBTB43, CEBPA
miR-29a	LAMA2, EMP1, PDK4, COL1A2, COL1A1, CXCL12, ENG, CYR61, RAB20, MGAT4B
miR-29c	LAMA2, COL1A2, COL1A1, COL6A2, LAMA2, MGAT4B
miR-26b-5p	GEM, ITM2A, MGP, FBLN5, CXCL9, GPR183, ARL4C, LYZ, PDK4, GPR183, F3, CCL2, PDGFRA, TRIM22, VIM, TAGLN, FABP4, IFI44L, VIM, ARL4C, IL7R
miR-335-5p	TRIM22, PDK4, SOX9, PTGDS, CRISPLD2, CXCL9, PPP1R1A,TACSTD2, MGP, CCL20, FABP4, GEM, FBLN5, AREG, PLAT, EMP1, SEL1L3, KCTD12, SOX4, ID4, SH3YL1,
miR-27a&b	ST3GAL6, FBLN5, RGS1, LHFP, F3
miR-106a	ARL4C, FZD7, PDGFRA, TXNIP, F3, CD69, CELF2, ANKRD12, TMEM100, CD69, MCL1, IL6ST,PIK3R1,PDGFRA, UBE2W, ZBTB43, PTPN4, MCL, BBX, SLC4A4
miR-128	PDGFRA, KLF4, APOLD1, UBE2W, PDGFRA, RAB20, MGAT4B
miR-15a	TXNIP, PDK4, CD69, EIF4B,TMEM100, ADAMTS1, HNRNPA1, ADAMTS1, SLC4A4, PIK3R1
miR-181a	CD69, VCAN, SPP1, EVI2A
miR-181c	CYR61, HSP90B1, ADM, CALR, SLA
miR-199a-3p	PDGFRA, CELF2, CXCL12, FGL2, CD44, KRT7, ZBTB43, PIK3R1, UBE2W, FOS
miR-200a	APOLD1, GATA6,
miR-200b	LHFP, ANK3, GEM, ADAMTSL2, KLF4, GPM6A, ZBTB38,
miR-93	F3, CD69, FZD7, TGFB1I1, VIM, MCL1, BBX, SLC4A4, PTPN4
miR-607	FZD7, IGJ, FGL2, CXCL12, LHFP, GJA1

Of the miRNAs that were found to be downregulated miR-27, miR-199, miR-200, and let-7 family members were amongst them and were found to target genes involved in cell cycle and

cell death regulation.(53) These miRNAs were also found in this study to have high expression in normal liver tissue and eventually no expression in carcinoma tissue with HCV infection. Downregulation of the miRNAs miR-26, miR-29, and miR-124 have been implicated in cell proliferation, apoptosis, angiogenesis, and poor prognosis.(58) In a previous study, a correlation between miR-106a and the degree of differentiation was shown to suggest an involvement of specific miRNAs in the progression of the disease.(54) Lee and colleagues provide data to support the down-regulation of miRNA miR15a and increased cellular proliferation. (58)

MiRNAs that were found to be upregulated were miRNA-335-5p, miR-93, miR-181a & c, and miR-128. MiRNA-335-5p was found to be associated with non-alcoholic fatty liver disease and also showed a connection with metabolic disorders. (53) MiR-93 is overexpressed in HCC and may have a critical role in cell proliferation by regulating the G1-to-S cell cycle transition. (53) MiR-181 is upregulated and promotes migration and invasion of HCC cells.(56) The up-regulated miR-128, has nearly perfect complementarity in its seed sequences with HCV RNA genomes and is capable of inhibiting HCV replication and infection.(54)

As for miR-607 while it was of the miRNAs with high expression patterns for differentially expressed genes there was no information in the literature to date as to how it influences the digression in the condition of the liver or even how it regulates genes in the development of HCC in the liver. However, it can be noted that mIR-607 does act as a tumor suppressor in other cancers, such as male breast cancer.

IV. Conclusions

Transcriptional regulation of HCV-induced HCC is very complex in that it encompasses a wide variety of aspects as the liver travels from condition to condition in its progression to HCC. In

this study the aim was to focus on determining high expression patterns in transcriptional regulation of differentially expressed genes that are common in all 6 pairwise comparisons so as to outline the transcriptional regulation that underlies each condition in HCV-induced HCC. These master regulators in each condition may explain the underlying molecular mechanisms and signaling pathways that are influenced by the progression of liver disease. As was discussed above, the PKC pathway was found to turn on CREB and ATF-2, while ASK1-JNK/p38 was found to control AP-1 which is composed of Jun, Fos in hepatic angiogenesis. While in hepatic steatosis, master transcriptional regulators were found to manipulate metabolic enzymes that drastically change in the condition of the liver. In the progression to cancer, it seems that IKK dependent signaling as well as JAK-STAT pathway are the major contributors to the final stages of liver disease as cell proliferation and apoptosis provide the means for complete progression to HCC.

Also, one of the major contributors in the progression of liver disease is the impact of miRNAs on transcriptional regulation again as was discussed above it is unclear whether transcription factors are also involved in the transcriptional regulation of miRNAs in the liver since the functional expression of transcription factors can also be regulated by miRNAs. Yet, it is clear that miRNAs can potentially regulate every aspect of cellular activity and a diverse spectrum of liver functions.(54) In this study it was apparent that there are a plethora of miRNAs that underlie the control of cellular homeostasis in the liver. It was also apparent that each liver condition had its signature expression pattern of miRNAs. Yet, the miRNAs that were found to have high expression patterns in all 6 pairwise comparisons were shown to be master regulators in transcriptional regulation and in turn having an impact on the conditions of the liver and the progression to HCC.

Chapter 6. Future Directions

Understanding the molecular mechanisms that underlie the influence HCV has on the progression to HCC is of vital importance in that it aids in the ability to create a more precise treatment regime with the least amount of negative impact on the patient. In the present study our aim was to determine transcriptional regulatory networks in HCV-induced HCC. Here we were able to determine master transcriptional regulators that dictate each stage in the progression to HCC with HCV infection. In the future we strive to accomplish 4 major objectives.

Objective 1

Determine the signature transcription factors with high expression patterns in each condition and the genes that they influence for a more honed understanding of the underlying molecular mechanism that influences cellular homeostasis in each stage of liver disease. This would also provide insight as to the possible signaling pathways that may be manipulated for stopping the progression of liver disease.

Objective 2

Determine the signature miRNAs with high expression patterns in each condition and the genes that they influence for a more honed understanding of the underlying molecular mechanism that influences cellular homeostasis in each stage of liver disease. This would also provide insight as to the possible signaling pathways that also may be manipulated for stopping the progression of liver disease.

Objective 3

Determine whether transcription factors and miRNAs effect one another transcriptionally; and the influence this may have on differentially expressed genes in HCV-induced HCC and the progression of liver disease.

Objective 4

Determine specific transcription factors in conjunction with its specific miRNAs that influence transcription regulation together in order to provide a mechanistic tool that may be produced in order to give a better treatment option for HCV patients and stop the progression of the liver to HCC.

References

1. Negro F, Alberti A. The global health burden of hepatitis C virus infection. Liver International. 2011;31(s2):1-3.
2. Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, et al. A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology. 1994;19(5):1321-4.
3. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014;59(1):318-27.
4. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology. 2015;61(1):77-87.
5. Davis GL, Alter MJ, El–Serag H, Poynard T, Jennings LW. Aging of hepatitis C virus (HCV)-infected persons in the United States: a multiple cohort model of HCV prevalence and disease progression. Gastroenterology. 2010;138(2):513-21. e6.
6. Voelker R. HCV in Egypt. JAMA. 2000;283(14):1816-.
7. Kamal SM, Nasser IA. Hepatitis C genotype 4: What we know and what we don't yet know. Hepatology. 2008;47(4):1371-83.
8. Mohamoud YA, Mumtaz GR, Riome S, Miller D, Abu-Raddad LJ. The epidemiology of hepatitis C virus in Egypt: a systematic review and data synthesis. BMC infectious diseases. 2013;13:288.
9. Guerra J, Garenne M, Mohamed MK, Fontanet A. HCV burden of infection in Egypt: results from a nationwide survey. Journal of viral hepatitis. 2012;19(8):560-7.
10. Jackowiak P, Kuls K, Budzko L, Mania A, Figlerowicz M, Figlerowicz M. Phylogeny and molecular evolution of the hepatitis C virus. Infection, Genetics and Evolution. 2014;21:67-82.
11. Martell M, Esteban JI, Quer J, Genescà J, Weiner A, Esteban R, et al. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. Journal of virology. 1992;66(5):3225-9.
12. Catanese MT, Uryu K, Kopp M, Edwards TJ, Andrus L, Rice WJ, et al. Ultrastructural analysis of hepatitis C virus particles. Proceedings of the National Academy of Sciences. 2013;110(23):9505-10.
13. Piñeiro D, Martinez-Salas E. RNA structural elements of hepatitis C virus controlling viral RNA translation and the implications for viral pathogenesis. Viruses. 2012;4(10):2233-50.
14. Moradpour D, Penin F. Hepatitis C virus proteins: from structure to function. Hepatitis C Virus: From Molecular Virology to Antiviral Therapy: Springer; 2013. p. 113-42.
15. Pflugheber J, Fredericksen B, Sumpter R, Wang C, Ware F, Sodora DL, et al. Regulation of PKR and IRF-1 during hepatitis C virus RNA replication. Proceedings of the National Academy of Sciences. 2002;99(7):4650-5.
16. Gouklani H, Bull RA, Beyer C, Coulibaly F, Gowans EJ, Drummer HE, et al. Hepatitis C virus nonstructural protein 5B is involved in virus morphogenesis. Journal of virology. 2012;86(9):5080-8.
17. Bassendine M, Sheridan D, Bridge S, Felmlee D, Neely R, editors. Lipids and HCV. Seminars in immunopathology; 2013: Springer.
18. Blanchard E, Belouzard S, Goueslain L, Wakita T, Dubuisson J, Wychowski C, et al. Hepatitis C virus entry depends on clathrin-mediated endocytosis. Journal of virology. 2006;80(14):6964-72.
19. Bartosch B, Vitelli A, Granier C, Goujon C, Dubuisson J, Pascale S, et al. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. Journal of Biological Chemistry. 2003;278(43):41624-30.
20. Sainz Jr B, Barretto N, Martin DN, Hiraga N, Imamura M, Hussain S, et al. Identification of the Niemann-Pick C1-like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor. Nature medicine. 2012;18(2):281-5.

21. Romero-Brey I, Merz A, Chiramel A, Lee J-Y, Chlanda P, Haselman U, et al. Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. PLoS Pathog. 2012;8(12):e1003056.
22. Lindenbach BD, Rice CM. Molecular biology of flaviviruses. Advances in virus research. 2003;59:23-61.
23. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nature immunology. 2004;5(7):730-7.
24. Saito T, Owen DM, Jiang F, Marcotrigiano J, Gale Jr M. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature. 2008;454(7203):523-7.
25. Carnero E, Fortes P. HCV infection, IFN response and the coding and non-coding host cell genome. Virus research. 2015.
26. Li Q, Pène V, Krishnamurthy S, Cha H, Liang TJ. Hepatitis C virus infection activates an innate pathway involving IKK-[alpha] in lipogenesis and viral assembly. Nature medicine. 2013;19(6):722-9.
27. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461(7262):399-401.
28. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. Nature. 2011;472(7344):481-5.
29. Brass V, Berke JM, Montserret R, Blum HE, Penin F, Moradpour D. Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. Proceedings of the National Academy of Sciences. 2008;105(38):14545-50.
30. Crow MK, Olferiev M, Kirou KA. Targeting of type I interferon in systemic autoimmune diseases. Translational Research. 2015;165(2):296-305.
31. Hoofnagle JH. Hepatitis C: the clinical spectrum of disease. Hepatology. 1997;26(s 3):15-20.
32. Pellicoro A, Ramachandran P, Iredale JP, Fallowfield JA. Liver fibrosis and repair: immune regulation of wound healing in a solid organ. Nature Reviews Immunology. 2014;14(3):181-94.
33. Yoon EJ, Hu K-Q. Hepatitis C virus (HCV) infection and hepatic steatosis. International journal of medical sciences. 2006;3(2):53.
34. Okuda M, Li K, Beard MR, Showalter LA, Scholle F, Lemon SM, et al. Mitochondrial injury, oxidative stress, and antioxidant gene expression are induced by hepatitis C virus core protein. Gastroenterology. 2002;122(2):366-75.
35. Goossens N, Hoshida Y. Hepatitis C virus-induced hepatocellular carcinoma. Clinical and molecular hepatology. 2015;21(2):105-14.
36. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nature genetics. 2011;43(5):455-8.
37. Lo P, Urabe Y, Kumar V, Tanikawa C, Koike K, Kato N, et al. Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk. PloS one. 2013;8(4):e61279.
38. Eurich D, Boas-Knoop S, Bahra M, Neuhaus R, Somasundaram R, Neuhaus P, et al. Role of IL28B polymorphism in the development of hepatitis C virus-induced hepatocellular carcinoma, graft fibrosis, and posttransplant antiviral therapy. Transplantation. 2012;93(6):644-9.
39. Valenti L, Rumi M, Galmozzi E, Aghemo A, Del Menico B, De Nicola S, et al. Patatin-Like phospholipase domain-containing 3 I148M polymorphism, steatosis, and liver damage in chronic hepatitis C. Hepatology. 2011;53(3):791-9.
40. Hoshida Y, Fuchs BC, Bardeesy N, Baumert TF, Chung RT. Pathogenesis and prevention of hepatitis C virus-induced hepatocellular carcinoma. Journal of hepatology. 2014;61(1):S79-S90.
41. Hobert O. Gene regulation by transcription factors and microRNAs. Science. 2008;319(5871):1785-6.

42. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science. 1989;245(4916):371-8.
43. Sevignani C, Calin GA, Siracusa LD, Croce CM. Mammalian microRNAs: a small world for fine-tuning gene expression. Mammalian Genome. 2006;17(3):189-202.
44. Faller M, Guo F. MicroRNA biogenesis: there's more than one way to skin a cat. Biochimica et biophysica acta. 2008;1779(11):663-7.
45. Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. The EMBO Journal. 2004;23(20):4051-60.
46. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Research. 2004;14(10a):1902-10.
47. Farid E A. Role of MicroRNA Molecules in Colon Cancer Etiology. Biology and Medicine. 2014;6(12):201.
48. Babu MM. Introduction to microarray data analysis. Computational Genomics: Theory and Application. 2004:225-49.
49. Harr B, Schlötterer C. Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. Nucleic Acids Research. 2006;34(2):e8-e.
50. Beißbarth T, Speed TP. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics. 2004;20(9):1464-5.
51. Segal E, Wang H, Koller D. Discovering molecular pathways from protein interaction and gene expression data. Bioinformatics. 2003;19(suppl 1):i264-i72.
52. Aderem A, Adkins JN, Ansong C, Galagan J, Kaiser S, Korth MJ, et al. A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. MBio. 2011;2(1):e00325-10.
53. Bala S, Marcos M, Szabo G. Emerging role of microRNAs in liver diseases. World journal of gastroenterology: WJG. 2009;15(45):5633.
54. Chen X-M. MicroRNA signatures in liver diseases. World journal of gastroenterology: WJG. 2009;15(14):1665.
55. Liu X, Wang T, Wakita T, Yang W. Systematic identification of microRNA and messenger RNA profiles in hepatitis C virus-infected human hepatoma cells. Virology. 2010;398(1):57-67.
56. Huang S, He X. The role of microRNAs in liver cancer progression. British journal of cancer. 2011;104(2):235-40.
57. Waldron PR, Holodniy M. MicroRNA and hepatitis C virus-challenges in investigation and translation: a review of the literature. Diagnostic microbiology and infectious disease. 2014;80(1):1-12.
58. Morishita A, Masaki T. miRNA in hepatocellular carcinoma. Hepatology Research. 2015;45(2):128-41.
59. Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, et al. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus–induced hepatocellular carcinoma. Molecular Medicine. 2009;15(3-4):85.
60. Fernández M, Semela D, Bruix J, Colle I, Pinzani M, Bosch J. Angiogenesis in liver disease. Journal of hepatology. 2009;50(3):604-20.
61. Takakura N, Watanabe T, Suenobu S, Yamada Y, Noda T, Ito Y, et al. A role for hematopoietic stem cells in promoting angiogenesis. Cell. 2000;102(2):199-209.
62. Hassan M, Selimovic D, Ghozlan H, Abdel-kader O. Hepatitis C virus core protein triggers hepatic angiogenesis by a mechanism including multiple pathways. Hepatology. 2009;49(5):1469-82.
63. Goyama S, Yamaguchi Y, Imai Y, Kawazu M, Nakagawa M, Asai T, et al. The transcriptionally active form of AML1 is required for hematopoietic rescue of the AML1-deficient embryonic para-aortic splanchnopleural (P-Sp) region. Blood. 2004;104(12):3558-64.

64. Sakurai T, Maeda S, Chang L, Karin M. Loss of hepatic NF-κB activity enhances chemical hepatocarcinogenesis through sustained c-Jun N-terminal kinase 1 activation. Proceedings of the National Academy of Sciences. 2006;103(28):10544-51.
65. Rutkowski DT, Wu J, Back S-H, Callaghan MU, Ferris SP, Iqbal J, et al. UPR pathways combine to prevent hepatic steatosis caused by ER stress-mediated suppression of transcriptional master regulators. Developmental cell. 2008;15(6):829-40.
66. Kim KH, Hong SP, Kim K, Park MJ, Kim KJ, Cheong J. HCV core protein induces hepatic lipid accumulation by activating SREBP1 and PPARγ. Biochemical and biophysical research communications. 2007;355(4):883-8.
67. Mirandola S, Bowman D, Hussain MM, Alberti A. Hepatic steatosis in hepatitis C is a storage disease due to HCV interaction with microsomal triglyceride transfer protein (MTP). Nutr Metab (Lond). 2010;7:13.
68. Takahara Y, Takahashi M, Zhang Q-W, Wagatsuma H, Mori M, Tamori A, et al. Serial changes in expression of functionally clustered genes in progression of liver fibrosis in hepatitis C patients. World journal of gastroenterology: WJG. 2008;14(13):2010.
69. Ohata K, Hamasaki K, Toriyama K, Matsumoto K, Saeki A, Yanagi K, et al. Hepatic steatosis is a risk factor for hepatocellular carcinoma in patients with chronic hepatitis C virus infection. Cancer. 2003;97(12):3036-43.
70. He G, Karin M. NF-κB and STAT3–key players in liver inflammation and cancer. Cell research. 2011;21(1):159-68.
71. Yoshida T, Hanada T, Tokuhisa T, Kosai K-i, Sata M, Kohara M, et al. Activation of STAT3 by the hepatitis C virus core protein leads to cellular transformation. The Journal of experimental medicine. 2002;196(5):641-53.