

# Identification of miRNA-mRNA associations in hepatocellular carcinoma using hierarchical integrative model

**Rency S. Varghese**

Georgetown University Medical Center

**Yuan Zhou**

Georgetown University Medical Center

**Megan Barefoot**

Georgetown University Medical Center

**Cristina Di Poto**

Georgetown University Medical Center

**Abdalla Kara Balla**

MedStar Georgetown University Hospital

**Everett Oliver**

Georgetown University Medical Center

**Zaki A. Sherif**

Howard University

**Deepak Kumar**

North Carolina Central University

**Alexander H. Kroemer**

MedStar Georgetown University Hospital

**Mahlet G. Tadesse**

Georgetown University

**Habtom Ressom** (✉ [hwr@georgetown.edu](mailto:hwr@georgetown.edu))

<https://orcid.org/0000-0001-9296-2132>

---

## Research article

**Keywords:** miRNA, mRNA, hierarchical integrative model, pathway analysis, hepatocellular carcinoma, next generation sequencing.

**Posted Date:** October 17th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.16175/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on March 30th, 2020. See the published version at <https://doi.org/10.1186/s12920-020-0706-1>.

# Abstract

**Background :** The established role miRNA-mRNA regulation of gene expression has in oncogenesis highlights the importance of integrating miRNA with downstream mRNA targets. These findings call for investigations aimed at identifying disease-associated miRNA-mRNA pairs. Hierarchical integrative models (HIM) offer the opportunity to uncover the relationships between disease and the levels of different molecules measured in multiple omic studies.

**Results:** mRNA-seq and miRNA-seq data were acquired by analysis of tumor and normal liver tissues from 30 patients with hepatocellular carcinoma (HCC). We analyzed the data using HIM and identified 157 significant miRNA-mRNA pairs in HCC. The majority of these molecules have already been independently identified as being either diagnostic, prognostic, or therapeutic biomarker candidates for HCC. These pairs appear to be involved in processes contributing to the pathogenesis of HCC involving inflammation, regulation of cell cycle, apoptosis, and metabolism. For further evaluation of our method, we analyzed miRNA-seq and mRNA-seq data from TCGA network. While some of the miRNA-mRNA pairs we identified by analyzing both our and TCGA data are previously reported in the literature and overlap in regulation and function, new pairs have been identified that may contribute to the discovery of novel targets.

**Conclusion:** The results strongly support the hypothesis that miRNAs are important regulators of mRNAs in HCC. Furthermore, these results emphasize the biological relevance of studying miRNA-mRNA pairs.

## Abbreviations

Adj-N: Adjacent Normal Tissue; CNV: Copy Number Variation; EM: Expectation Maximization; FDR: False Discovery Rate (FDR); GESR: Genomics and Epigenomics Shared Resource; GU: Georgetown University; Georgetown University Medical Center (GUMC); HCC: hepatocellular carcinoma; HIM: Hierarchical integrative models; IPA: Ingenuity Pathway Analysis; MGUH: Medstar Georgetown University Hospital; STAR: Spliced Transcripts Alignment to a Reference; TCGA: The Cancer Genome Atlas; UMI: Unique Molecular Index.

## Background

MicroRNAs (miRNAs) play vital roles in many biological processes, including differentiation, cell signaling, and response to infection. Overwhelming evidence indicates that dysregulation of miRNA expression is a cause or indicator of several disease processes, including many cancers. MicroRNA dysregulation has been linked to cancer initiation and progression where miRNAs act as tumor suppressors or oncogenes, regulating multiple pathways including cell proliferation, differentiation, apoptosis, metastasis and angiogenesis [1]. Recent studies have highlighted several miRNAs that are differentially expressed in cancer stem cells establishing the role of miRNAs in targeting genes and pathways supporting cancer stemness [2]. Several miRNAs and their targets are involved in known risks

and hallmarks of liver diseases and HCC, such as steatosis, fibrosis, and cirrhosis [3]. For example, a series of experiments conducted in hepatoma cells have found miR-93-5p to be upregulated and act as a promoter of cell proliferation. Furthermore, the inverse correlation between miR-93-5p and PPARGC1A gene expression was validated in the TCGA liver cancer data [4]. Mature miRNAs are highly stable and have great utility as biomarkers of diagnosis/prognosis and disease progression.

The importance of miRNA in carcinogenesis can be better understood through linking miRNA with respective target genes. The impact that miRNA-mRNA pairs have in carcinogenesis also highlights the potential of miRNAs to become therapeutic targets [5]. Many studies demonstrate that disruption of miRNA-mRNA paired regulation contributes to tumorigenesis. For example, a study consisting of African American participants suggests that miRNA-mRNA pairs play a critical role in the activation of oncogenic pathways in prostate cancer [6]. A co-expression-based network approach has identified clustered regulatory circuits of miRNA-mRNA modules in hepatocyte, inflammatory-stress and proliferative process-activated subcategories of HCC [7]. Thus, the associations between miRNAs and their target genes, i.e., up or down regulated genes, is currently an area of significant research interest.

MicroRNAs either suppress an mRNA through translational repression or accelerated degradation, or activate an mRNA through stimulated stabilization or stimulated translation. Conversely, mRNAs either suppress miRNA through accelerated degradation, or activate miRNA by sequestering it from degradation. Typically, miRNA-mRNA associations exist as reciprocal pairs whereby cells select either high/low level of mRNA/miRNA or the opposite expression pattern [8]. However, dual-upregulation of paired miRNA-mRNA has been experimentally validated as an existing relationship observed in cancer [9]. As miRNA-mRNA regulation generally exists in various biological systems, understanding the functionality of miRNA-mRNA association is important for learning its underlying design principle.

Hierarchical integrative models (HIMs) offer the opportunity to uncover the relationships between disease and the levels of different molecules measured in multiple omic studies. In the past, they have been successfully implemented for integrating DNA methylation, copy number variation (CNV), and gene expression data in relation with survival outcomes [10]. Specifically, modeling approaches based on penalized likelihood methods and expectation-maximization algorithms were applied to various biological relationship scenarios between different molecular features to determine their effects on clinical outcome. In this paper, we formulated a HIM adapted for application to identify miRNA-mRNA associations related to HCC. We accomplished this through analysis of mRNA-seq and miRNA-seq data generated in house from HCC patients paired tumor and normal liver tissues. We demonstrate that our model offers insights into the molecular mechanisms of HCC. Through pathway and network analysis of selected miRNA-mRNA associations we uncover the biological relevance of these pairs in HCC.

### *Liver Tissues*

*Table 1* presents characteristics of 30 HCC patients whose liver tissues were analyzed to generate mRNA-seq and miRNA-seq data (GU dataset). The liver tissues were collected from 30 adult liver patients recruited at the hepatology clinics at Medstar Georgetown University Hospital (MGUH), Washington, DC.

The patients were diagnosed to have HCC based on well-established diagnostic imaging criteria and/or histology. Tissues were donated by patients scheduled for elective surgical procedures due to the cancerous or potentially cancerous conditions and/or other related benign conditions. These provide a source of both tumor and normal-control tissues. Prior to enrollment in the study, the patients were consented to participate using IRB-approved (2005–206) informed consent and HIPAA authorization forms. The samples from HCC patients include tumor tissues (HCC) and adjacent normal tissues (Adj-N). Thus, the samples used to generate the GU dataset consist of a total of 60 tissue samples from 30 HCC patients.

Furthermore, we analyzed miRNA-seq and mRNA-seq data obtained from the TCGA Research Network (TCGA dataset). The dataset we used in this study is a small subset of the TCGA data obtained by analysis of tumor and adjacent normal liver tissues from 49 patients (*Table 2*). These patients were matched using the TCGA patient identifiers to select their corresponding normal tissues.

## Results

### *Analysis of GU dataset*

We used a paired t-test to identify miRNAs and mRNAs that are differentially expressed between the HCC and adjacent normal tissues. Adjusted p-values were calculated using Benjamini-Hochberg (BH) method. Fold changes were calculated considering the median of raw intensity values. *Table 3* represents the number of features detected and the number of miRNAs and mRNAs selected with FDR < 0.05.

The miRNAs with FDR < 0.05 were matched to target pairs from databases using the miRNA target filter function that provides biological effects of miRNAs based on experimentally validated interactions from TarBase and miRecords, as well as predicted miRNA-mRNA interactions from TargetScan. The miRNA targets were then overlapped with the list of significant mRNAs from the mRNA-seq analysis to select reported targets of the significant miRNAs. A pathway analysis of this filtered set performed using IPA identified 469 pathways as significantly enriched with  $p < 0.05$ . FXR/RXR Activation, LXR/RXR Activation, Acute Phase Response Signaling, Hepatic Fibrosis / Hepatic Stellate Cell Activation, Sirtuin Signaling Pathway, and Valine Degradation I were selected as the top pathways enriched using the significant mRNAs.

The mechanistic submodel of HIM relating each of the significant (FDR < 0.05) 5,751 mRNAs to 238 miRNAs using glmnet identified 26,080 miRNA-mRNA associations. All of the 238 miRNAs were involved in a potential pair with 3,632 mRNAs as a result of the mechanistic submodel. All of the pairs and significant independent molecules that passed the univariate t-tests with FDR control were included in the clinical submodel for subsequent analysis. The clinical submodel selected 157 miRNA-mRNA pairs associated to HCC, which consist of 19 mRNAs and their related 90 miRNAs. *Figure 1* depicts the number of molecules that were selected at each level of the analysis in the HIM. The selected 157 miRNA-mRNA pairs are included in the *Supplementary Figure 1*.

MicroRNA target filter analysis was then performed in IPA to match associated pairs that are reported or already verified experimentally. Of the 157 significant pairs selected by HIM, there are 40 experimentally verified pairs involving the molecules selected in the pair. Of these, five of the miRNA-mRNA pairs are exact matched pairs. *Figure 2* presents these pairs with their confidence of prediction from the database and the dotplots of each pair.

The miRNA-mRNA associations identified by HIM were analyzed in IPA to generate association networks from the knowledge-base and to further select the pathways that are dysregulated in HCC involving these miRNA-mRNA pairs. The significant pathways selected by the molecules involved in these pairs can be found in *Figure 3*. A network generated by IPA showing the interactions and biological relevance of these selected molecules are shown in *Supplementary Figure 2*. The top biological processes depicted by this network are glomerular injury, inflammatory disease, and inflammatory response. The molecules drawn into this network from the database reflect the role of the selected molecules in pathogenesis of HCC as relevant to changes initiating inflammation, organismal injury, and connective tissue abnormalities.

### *Analysis of TCGA dataset*

We analyzed the TCGA dataset following the same approach as the GU dataset. Although miRNA-seq data consisted of 2,588 miRNAs, there were many missing values for the 49 subjects considered and thus only 740 miRNAs were analyzed in this study. As shown in *Table 4*, there were 354 miRNAs and 11,033 mRNAs with  $FDR < 0.05$  related to tumor vs. normal tissue.

A pathway analysis of significant miRNAs and mRNAs performed using IPA identified 522 pathways as significantly enriched with  $p < 0.05$ . Of these, 85% of the pathways overlap with the 469 significantly enriched pathways from GU dataset. Axonal Guidance Signaling, LXR/RXR Activation, LPS/IL-1 Mediated Inhibition of RXR Function, Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis, Hepatic Fibrosis / Hepatic Stellate Cell Activation, FXR/RXR Activation, were selected as the top pathways enriched using the significant mRNAs.

Of the 263 significant miRNA-mRNA pairs selected by the mechanistic model, six pairs were found to be experimentally verified based on mRNA target filter in IPA. Similar to the GU dataset, the pairs selected from the TCGA dataset overlap in regulation and function. The 263 miRNA-mRNA pairs and IPA generated network of interactions of these selected molecules are shown in *Supplementary Figures 3 and 4*, respectively. The molecules that are directly connected to the selected miRNA and mRNA to form a network in IPA reflect the role of those selected molecules in pathogenesis of HCC as relevant to changes involving organismal injury and abnormalities progressing to cancer. The pathways generated using the molecules from the combined models can be found in *Supplementary Figure 5*. The list and dot plots of the six verified miRNA-mRNA pairs are shown in *Supplementary Figure 6*.

## **Discussion**

Recently discovered significance of miRNAs in homeostasis and disease have increased the importance of incorporating miRNAs into gene regulatory networks. HIMs offer the opportunity to associate molecules measured in multiple omic studies across several levels to discover novel relationships pertaining to disease status. The purpose of this paper is to integrate miRNA-seq and mRNA-seq data to select associated miRNA-mRNA pairs that are of biological relevance to HCC.

### *Dysregulation of miRNA-mRNA pairs in HCC*

A literature search was conducted to validate selected molecules, many of which were found to be extensively reported in relation to HCC. Several of these selected molecules are already identified as being either diagnostic (VSP45, TERT, EBF2, hsa-miR-324-5p, hsa-miR-130a-3p, hsa-miR-200a-3p, hsa-miR-106b-5p, hsa-21-5p, hsa-424-5p), prognostic (CAP2, TRMT6, MT1JP, hsa-miR-10b-5p, hsa-miR-139-3p, hsa-miR-101, hsa-miR-122-5p, hsa-miR-1285-3p, hsa-miR-203a-3p), recurrence-related (CD34, hsa-miR-183), or therapeutic (AKR1B10, SLC25A4, CYP2B7, FBXL18, hsa-miR-146a-5p, hsa-miR-34a, hsa-miR-122, hsa-miR-671-5p) biomarker candidates for HCC [11–15]. These molecules appear to be involved in molecular events relating to inflammation, tumor microenvironment, cell cycle, apoptosis, and metabolism. These events in combination with key oncocytic molecular events facilitate pathogenesis, progression, and metastasis of HCC. This is evident by literature search findings as well as our pathway and network analysis results.

### *MARCO as a potential target gene of hsa-miR-10b-5p*

Our hierarchical integrated model results indicate that, macrophage scavenger receptor (MARCO) is associated with hsa-miR-10b-5p in both our GU and the TCGA datasets. *Figure 4* shows the dot plots of the pairs selected from both datasets. The hsa-miR-10b-5p to MARCO pair is involved in HCC through increasing cellular proliferation, differentiation, and migration in HCC. According to miRecords, there are five experimentally verified targets of hsa-miR-10b, and MARCO has not been reported as a target. Decreased expression of MARCO is associated with poor prognosis in HCC, potentially due to its role in modulating inflammatory signaling [16]. Recently, MARCO has been identified as a marker of non-inflammatory macrophages in the liver [17]. Expression of MARCO has been used to define a population of suppressive tumor-associated macrophages. Thus, upregulation of MARCO in HCC suggests a phenotype of heightened immunosuppression. Interestingly, hsa-miR-10b was also found to be downregulated in macrophages and explored as a circulating biomarker of HCC recurrence in bone marrow cells, suggesting a possible link between hsa-miR-10b and MARCO [18]. Hsa-miR-10b has also been identified as a prognostic biomarker of HCC in a study that used 33 miRNA signatures to characterize HCC using TCGA data [14]. Overexpression of miR-10b in HCC has been reported to promote cell proliferation, migration and invasion [19]. Thus, MARCO can be a potential target of hsa-miR-10b that is worth verifying. Further, additional miRNAs were identified by our analysis as having an influence on MARCO expression. These include hsa-miR-125b-2-3p, hsa-miR-146a-5p, hsa-miR-200a-3p, and hsa-miR-221 that are all known diagnostic biomarkers of HCC.

### *miRNA-mRNA pairs involved in tumor inflammation*

Pairs involving regulation of MARCO, IFITM10, and CD34 appear to function in HCC through changes involving tumor microenvironment and inflammation. IFITM10 is an interferon-induced transmembrane protein known to play a role in anti-viral response in HCC caused by virally-induced etiology [20]. IFITM10 was found to be targeted by hsa-miR-106b-5p in our analysis, a miRNA known to activate immunoregulatory T cells and myeloid-derived suppressor cells [21]. Hsa-miR-146a-5p is thought to be down-regulated in HCC leading to immune suppression and HCC-induced NK cell dysfunction. MARCO is a target gene found to be regulated by hsa-miR-146a-5p in our analysis suggesting functional relevance [22]. Hsa-miR-122 and Hsa-miR-146a-5p are both thought to be involved in regulating STAT3 signaling, a key pathway involved in immune suppression in tumor microenvironment [22]. Hsa-miR-122 is the most abundant miRNA in healthy human liver and down-regulation in HCC is associated with poor prognosis [13]. Further, restoration of miR-122 levels is used to indicate therapeutic response and sensitivity with several drugs used to treat HCC. Similarly, hsa-miR-34a suppression in HCC has been linked to increased CCL22 levels known to recruit regulatory T cells involved in immune escape [23]. Lastly, CD34 expression on liver sinusoidal endothelial cells is linked to chronic liver inflammation and thought to play a role in the pathogenesis of HCC from underlying cirrhosis [24]. The role of miRNA-mRNA regulation in HCC could prove useful to assess the immune status of these tumors and predict response to these novel therapeutics.

#### *miRNA-mRNA pairs involving cell cycle and apoptosis*

Pairs involving TRMT6, SLC25A40, MT1JP, and EBF2 are involved in HCC through regulation of cell cycle and apoptosis contributing to disease progression and metastasis. [25, 26]. Hsa-miR-671-5p was associated with CD34 and EBF2 in our analysis and is known to be associated with metastasis, proliferation, and EMT in HCC [27]. Dysregulated expression of hsa-miR-7 and hsa-miR-532-3p in HCC leads to increased proliferation, invasion, and metastasis through the PI3K/AKT signaling pathway [28]. Further, hsa-miR-532-3p along with hsa-miR-139 are associated with CHRD, FLAD1, MT1JP, and EBF2 in our analysis and have been identified as tumor suppressors in HCC that inhibit cell proliferation, migration, and invasion [29]. MT1JP is also thought to act as a tumor suppressor through regulating a series of pathways involving p53, such as the cell cycle, apoptosis and proliferation [30]. Similarly, TRMT6 upregulation correlates with poor prognosis in HCC and is a MYC target gene involved in G2M checkpoint and cell-cycle regulation [31]. Hsa-miR-101-5p is another molecule regulated by MYC signaling that is repressed, leading to progression and metastasis of HCC [32]. Hsa-miR-200a is thought to suppress proliferation in HCC by induction of G1 arrest through CDK6. Similarly, hsa-miR-125b-2-3p is an upstream regulator of oncogenic sirtuin that also causes G1 arrest. Interestingly, Sirtuin signaling was one of the top pathways identified from this analysis.

#### *miRNA-mRNA pairs involved in metabolic regulation*

Many miRNA-mRNA pairs identified contribute to aberrant metabolism in HCC and some of these are further associated with distinct etiology. For instance, target genes (FLAD1, SLFN1, GNS, FBXL18, CYP2B7P) are commonly upregulated genes in metabolic-induced HCC and are associated with risk

factors, such as diabetes, obesity, and other metabolic syndromes. These genes are commonly upregulated in a subclass of malignant HCC tumors under metabolic stress and associated with hypoxic behavior, epithelial-to-mesenchymal transition, and increased fatty acid biosynthesis [33]. CYP2B7P was found to be a target gene of hsa-miR-122 that is a diagnostic biomarker of HCC as well as of non-alcoholic fatty liver disease (NAFLD) [13]. MiR-122 is also known to affect metabolism of HCC cells through pyruvate kinase [34]. CYP2B7P was also found to be a gene target of hsa-miR-183-5p that is a biomarker of HCC and diabetes. Many of the top pathways identified from these selected miRNA-mRNA pairs are involved in metabolic regulation, including FXR/RXR activation, LPS-IL-1 mediated inhibition of RXR function, LXR/RXR activation, valine degradation, fatty acid B-oxidation, hepatic cholestasis and tryptophan degradation. Tryptophan degradation was one of the pathways found to overlap between significant pairs identified by TCGA and our datasets. In addition, we previously identified several molecules from the FXR/RXR activation pathway as being differentially expressed in HCC patients involved in bile acid homeostasis, lipid and glucose metabolism [35].

#### *Upstream miRNA regulators of TERT*

Telomerase plays an important role in maintaining telomere length and chromosomal stability in hepatocytes that is often dysfunctional in cirrhotic liver and HCC [13]. Early-stage and virally-induced tumors are often associated with telomere dysfunction along with p53 mutations. The relevance of TERT and TP53 in these HCC cases can be visualized by the top network depicting interaction amongst our selected miRNA-mRNA pairs (*Supplemental Figure 2*). Over 68% of HCC cases are thought to have TERT mutations influencing telomere maintenance and susceptibility for viral-genome integration [36]. In addition, TERT is a commonly amplified gene in HCC patients and its presence has been used to characterize a subpopulation of HCC cases [37]. For example, HCC patients with HBV are more likely to have mutations in P53 than TERT because HBV often integrates at the location of the TERT gene. In contrast, patients with HCV and HCC patients of Black or African American race are more likely to have TERT promoter mutations than p53. A cluster of genes appears to be associated with this subpopulation, including CAP2 and CHRD, that are involved in BMP signaling [38]. BMP signaling was another pathway found to overlap between the selected pairs identified by TCGA and our datasets and is associated with this subpopulation. TERT was found to be regulated by several miRNAs including hsa-miR-21-5p, hsa-miR-130b-5p, hsa-miR-675-3p, and hsa-miR-488-3p. All of these miRNAs are thought to play a role in HCC [37]. This makes miRNA pairs involved with regulating TERT a significant focus worth additional validation.

#### *Hierarchical integrative model to identify miRNA-mRNA pairs*

Some of the miRNA-mRNA pairs identified in this study are reported to have involvement in inflammation, cell cycle regulation, apoptosis, and metabolism. In addition to the functional categories that most pairs are involved in, several pairs are involved in resistance mechanisms, splicing alterations, and epigenetic dysregulation. Several known pairs stemming from the selected 19 mRNAs and associated 90 miRNAs have been experimentally validated in the database with different partners. For instance, TRMT6 is a

known target gene of both miR-21 and miR-146a-5p. From our analysis, miR-21 and miR-221 were both selected as being involved in significant pairs functioning in HCC. These two miRNAs are known oncomiRs in HCC whose mRNA targets are often found to act as tumor suppressors influencing tumor proliferation and migration [13]. In addition to TRMT6, several other genes were identified as targets of miR-21 that have not yet been validated, including TERT and TPS45. However, we recognize that the actual role the above mentioned pairs may have in the pathogenesis of HCC is much more complex and not fully known. The ability of HIMs to identify novel miRNA-mRNA pairs that have not been validated in existing databases is a strength that can be used for discovery of novel targets. For example, our analysis identified four genes that are potential targets of miR-221 (MARCO, AKR1B10, CD34, and LARS) that could have similar importance in HCC. This shows the ability of HIMs to predict novel miRNA-mRNA pairs of biological significance in relation to HCC that warrant additional validation in the future.

## Conclusion

In this study, we used hierarchical integrative model to integrate miRNA-seq and mRNA-seq data acquired by analysis of paired tumor and adjacent normal liver tissues from the same set of patients with HCC. The model led to identification of key miRNA-mRNA pairs and associated pathways that are potentially involved in HCC pathogenesis. The results strongly support the hypothesis that miRNAs are important regulators of mRNA expressions in HCC. Furthermore, these results emphasize the biological relevance of studying miRNA-mRNA pairs when analyzing miRNA-seq and mRNA-seq datasets. As mRNA-miRNA regulation generally exists in various biological systems, understanding the functionality of mRNA-miRNA is critical for elucidating its underlying role in disease pathology. Combining regulators of disease across multiple -omics levels can aid in providing a more synergistic view of disease progression in general and a roadmap to addressing the imbalance between statistically and biologically relevant information.

## Methods

### *RNA Isolation and quality assessment*

Total RNA was isolated from 20 mg frozen liver tissue using the Zymo Direct-zol RNA Miniprep Plus Kit (Zymo Research) according to the manufacturer's instructions. The RNA quality and quantity were analyzed with UV-VIS using the NanoDrop spectrophotometer (Thermo Fisher), and fluorometry using the Qubit 2.0 Fluorometer (Thermo Fisher). RNA integrity was assessed using the Agilent RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer. This work was performed in the Genomics and Epigenomics Shared Resource (GESR) at GUMC.

### *mRNA expression profiling*

RNA samples extracted from liver tissues were analyzed by mRNA-seq. Briefly, indexed paired-end sequencing libraries were prepared from 10 ng total RNA using the TruSeq RNA Access Library Prep Kit

(Illumina). Paired-end sequencing was performed on the HiSeq 4000 System (Illumina) using 150 bp paired-end (PE150) read mode.

The mRNA-seq data contained an average of 33M reads per sample. The raw data recorded in FASTQ files were imported into Partek® Flow® software, (Partek Inc., St. Louis, MO, USA) for quality assessment and mRNA-seq data analysis. Quality assessment was performed at each step of the procedure including RNA isolation, library preparation and sequencing. Reads containing the adapters or low quality (Qscore < = 5) bases were removed prior to alignment. For the mRNA-seq data, the average Q30 (base count of Phred value >30/ total base count) was 93.99% and the average GC content was 52.1%.

Alignment to the hg38 reference genome was performed using the spliced transcripts alignment to a reference (STAR) algorithm, with 89.6% alignment. The aligned reads were then filtered to remove duplicate reads and then quantified to the transcriptome through an Expectation Maximization (EM) method. To compare the gene expression between two conditions and eliminate systematic effects that are not associated with the biological differences of interest, we needed to normalize the samples. TMM (Trimmed mean of M-values) method was applied to estimate appropriate scaling factors for normalization. The normalized count data were exported from the Partek Flow software for further data analysis. After conducting the TMM normalization, Principal Components Analysis was used for outlier screening. One Adj-N tissue sample was detected as an outlier. Therefore, we deleted this Adj-N tissue sample and its corresponding HCC tissue sample from the further analysis.

During the sample preparation of the mRNA-seq experiment, four samples were randomly selected and were included as blinded samples for subsequent quality assessment. During data acquisition, these samples were marked as unknowns. Using cluster analysis of the result mRNA-seq data, we successfully clustered the 'Unknown' samples into their correct groups. This confirms the quality of the mRNA-seq data we acquired and preprocessed, prior to analyzing the data using hierarchical integrative model to identify mRNA-miRNA associations in HCC.

### *miRNA expression profiling*

Aliquots of the same RNA samples used for the mRNA-seq analysis described above were analyzed by miRNA-seq. Indexed sequencing libraries were prepared from 200 ng total RNA using the QIAseq miRNA Library Kit (Qiagen) in the GESR according to the manufacturer's instructions. Sequencing was performed on the NextSeq 550 System (Illumina) using 1×75 bp paired-end (SE75) read mode. The miRNA-Seq data generated were then analyzed using the QIAseq miRNA quantification data analysis software. After quality assessment of the samples, FASTQ files were uploaded to the data analysis center at Qiagen for further analysis. In the first step, the unique molecular index (UMI) counts were calculated and primary miRNA mapping was performed. In the second step, the UMI counts were analyzed to calculate the changes in miRNA expression. The quantified data were then normalized using TMM method.

### *Evaluation of dependence between tumor and adjacent normal tissues*

We investigated the dependence between the tumor and adjacent normal tissue pairs across all miRNA-seq and mRNA-seq data. A correlation analysis between the tumor and adjacent normal tissues revealed that 87% of the correlations for the mRNAs and 70% of the correlations for the miRNAs have magnitudes lower than 0.3. In addition, a comparison of fixed effects and mixed effects models using AIC and BIC suggested that including random effects for individuals to account for the matching is not necessary. Thus, in our formulation of hierarchical integrative model, we assumed that the sample groups were independent.

### *Hierarchical integrative model (HIM)*

Hierarchical integrative model offers the opportunity to associate molecules measured in multiple omic studies across several levels to uncover novel relationships pertaining to disease status. The HIM model we formulated for analysis of mRNA-seq and miRNA-seq data can be specified with two levels: (1) a mechanistic submodel relating mRNA to miRNA markers, and (2) a clinical submodel relating disease status to mRNA and miRNA, while accounting for the mechanistic relationships in the first level (*Figure 5*).

The mRNAs and miRNAs that go into the HIM are pre-selected using paired t-test considering only those mRNAs and miRNAs associated to the binary disease status  $Y$  (tumor or normal liver tissue). The mechanistic submodel fits multivariable models relating each mRNA to miRNAs. This is accomplished via lasso-regularized linear models using the R package *glmnet* [39]. The clinical submodel uses a penalized logistic regression model to relate the disease status  $Y$  to the linear predictors from the mechanistic submodel as well as mRNA and miRNA levels that were not selected in the mechanistic submodel. *Figure 6* presents the steps involved in using HIM to identify miRNA-mRNA pairs associated to HCC.

### *Mechanistic submodel*

The mechanistic submodel is formulated as in Eq. (1), in which  $\eta_j$  corresponds to the linear predictor from the penalized regression model, relating  $mRNA_j$  to its associated miRNAs and  $e_j$  is the residual.

$$mRNA_j = \alpha_0 + \alpha_1 + \alpha_2 + \dots + \alpha_p + e_j \quad \text{Eq. 1-mechanistic submodel}$$

00  $\eta_j$

### *Clinical submodel*

The clinical submodel is formulated as in Eq. (2), in which  $mRNA_{\bar{A}}$  corresponds to the set of mRNAs that passed the t-test ( $FDR < 0.05$ ) but were not picked in the mechanistic submodel (i.e., mRNAs related to  $Y$  but that do not have miRNAs associated with them);  $miRNA_{\bar{A}}$  corresponds to the set of miRNAs that passed the t-test ( $FDR < 0.05$ ) but were not picked in the mechanistic submodel (i.e., miRNAs not related to mRNAs but that could potentially affect the response  $Y$ ). The logistic regression model in the clinical submodel is fit using a lasso penalty to identify relevant markers.

$\text{logit PY} = 1 = \beta_0 + \sum \beta_{nj} + \sum \beta_{e_j} + \sum \beta_{kmRNA_k} + \sum \beta_{lmiRNA_l}$  Eq.2-clinical submodel

The selection of  $n_j$  would imply that the effects of  $mRNA_j$  on the outcome are modulated by the miRNAs associated to it, while the selection of  $e_j$  would imply that the effects of  $mRNA_j$  on the outcome result from factors other than those accounted for by its associated miRNAs. With this formulation, if  $n_j$  is selected,  $mRNA_j$  and all its associated miRNAs would be picked, although some of the miRNAs may not be relevant.

## Declarations

### *Ethics approval and consent to participate*

In this study, liver tissues from 30 adult patients recruited at Medstar Georgetown University Hospital (MGUH), Washington, DC were used. The study is conducted through a protocol approved by the Georgetown-Howard Universities Center for Clinical and Translational Science (GHUCCTS) Institutional Review Board (IRB), Washington, DC, under Protocol # 2014–0059 “Multi-omic Approaches for Liver Cancer Biomarker Discovery.” The collection and use of the tissues was approved by the IRB of Georgetown University, Washington, DC, under Protocol #2007–345 “Establishment of the High Quality Tumor Biobank and Clinical Database”. All patients signed a consent form permitting the use of donated tissue. The consent forms and their content were reviewed and approved by the IRB.

### *Consent for publication*

All authors consent to publish the material presented in this manuscript.

### *Availability data and material*

R code implementation of HIM and pre-processed miRNA-seq and mRNA-seq data are available at <https://github.com/ressomlab/HIM>. The TCGA dataset were obtained from the TCGA Research Network (<https://cancergenome.nih.gov>), using the Broad Institute firehose download link, <http://gdac.broadinstitute.org/> (accessed on January 20, 2019). Level 3 mRNA-seq and miRNA-seq data were obtained. The clinical information was available from the NCI genomic data commons data portal at (<https://portal.gdc.cancer.gov/projects/TCGA-LIHC>).

### *Competing interests*

All authors declared that they have no competing interests.

### *Funding*

The work presented in this paper is conducted with funding provided by NIH grants. U01CA185188 and R01GM123766 awarded to HWR provided support for omic data generation and analysis. U01CA194730 and R01MD012392 to DK provided support for miRNA-seq data acquisition.

### *Authors' contributions*

RSV and YZ performed the statistical analysis and generated the results and figures. MT provided guidance on the development of statistical models. MB performed the pathway analysis. CDP, AKB, and EO coordinated the sample collection, annotation of samples, and prepared the samples for generating the data using mRNA-Seq and miRNA-Seq. HWR, ZAS, DK, and AHL conceived the study design and provided interpretation of results. RSV and MB wrote the manuscript. All authors critically reviewed and approved the manuscript.

### *Acknowledgements*

The authors thank the Genomics and Epigenomics Shared Resource (GESR) for RNA isolation and library preparation prior to data generation.

## References

- [1] S. Oliveto, M. Mancino, N. Manfrini and S. Biffo. Role of microRNAs in translation regulation and cancer. *World J. Biol. Chem.* 8(1), pp. 45–56. 2017. DOI: 10.4331/wjbc.v8.i1.45.
- [2] M. Garofalo and C. M. Croce. Role of microRNAs in maintaining cancer stem cells. *Adv. Drug Deliv. Rev.* 87pp. 53–61. 2015. DOI: 10.1016/j.addr.2014.11.014.
- [3] C. H. Liu, J. Ampuero, A. Gil-Gomez, R. Montero-Vallejo, A. Rojas, R. Munoz-Hernandez, R. Gallego-Duran and M. Romero-Gomez. miRNAs in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis. *J. Hepatol.* 69(6), pp. 1335–1348. 2018. DOI: S0168–8278(18)32289-X.
- [4] X. Wang, Z. Liao, Z. Bai, Y. He, J. Duan and L. Wei. MiR–93–5p promotes cell proliferation through down-regulating PPARGC1A in hepatocellular carcinoma cells by bioinformatics analysis and experimental verification. *Genes (Basel)* 9(1), pp. 10.3390/genes9010051. 2018. DOI: E51.
- [5] E. Enerly, I. Steinfeld, K. Kleivi, S. K. Leivonen, M. R. Aure, H. G. Russnes, J. A. Ronneberg, H. Johnsen, R. Navon, E. Rodland, R. Makela, B. Naume, M. Perala, O. Kallioniemi, V. N. Kristensen, Z. Yakhini and A. L. Borresen-Dale. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* 6(2), pp. e16915. 2011. DOI: 10.1371/journal.pone.0016915.
- [6] B. D. Wang, K. Ceniccola, Q. Yang, R. Andrawis, V. Patel, Y. Ji, J. Rhim, J. Olender, A. Popratiloff, P. Latham, Y. Lai, S. R. Patierno and N. H. Lee. Identification and functional validation of reciprocal microRNA-mRNA pairings in african american prostate cancer disparities. *Clin. Cancer Res.* 21(21), pp. 4970–4984. 2015. DOI: 10.1158/1078–0432.CCR–14–1566.
- [7] V. Ramesh and K. Ganesan. Integrative analysis of transcriptome and miRNome unveils the key regulatory connections involved in different stages of hepatocellular carcinoma. *Genes Cells* 21(9), pp. 949–965. 2016. DOI: 10.1111/gtc.12396.

- [8] X. J. Tian, H. Zhang, J. Zhang and J. Xing. Reciprocal regulation between mRNA and microRNA enables a bistable switch that directs cell fate decisions. *FEBS Lett.* 590(19), pp. 3443–3455. 2016. DOI: 10.1002/1873–3468.12379.
- [9] A. Valinezhad Orang, R. Safaralizadeh and M. Kazemzadeh-Bavili. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int. J. Genomics* 2014pp. 970607. 2014. DOI: 10.1155/2014/970607.
- [10] M. Denis and M. G. Tadesse. Evaluation of hierarchical models for integrative genomic analyses. *Bioinformatics* 32(5), pp. 738–746. 2016. DOI: 10.1093/bioinformatics/btv653.
- [11] K. Kitagawa, G. Nakajima, H. Kuramochi, S. I. Ariizumi and M. Yamamoto. Lymphatic vessel endothelial hyaluronan receptor–1 is a novel prognostic indicator for human hepatocellular carcinoma. *Mol. Clin. Oncol.* 1(6), pp. 1039–1048. 2013. DOI: 10.3892/mco.2013.167.
- [12] C. Meng, X. Shen and W. Jiang. Potential biomarkers of HCC based on gene expression and DNA methylation profiles. *Oncol. Lett.* 16(3), pp. 3183–3192. 2018. DOI: 10.3892/ol.2018.9020.
- [13] R. Chauhan and N. Lahiri. Tissue- and serum-associated biomarkers of hepatocellular carcinoma. *Biomark Cancer.* 8(Suppl 1), pp. 37–55. 2016. DOI: 10.4137/BIC.S34413.
- [14] M. Lu, X. Kong, H. Wang, G. Huang, C. Ye and Z. He. A novel microRNAs expression signature for hepatocellular carcinoma diagnosis and prognosis. *Oncotarget* 8(5), pp. 8775–8784. 2017. DOI: 10.18632/oncotarget.14452.
- [15] A. Nagy, A. Lanczky, O. Menyhart and B. Gyorffy. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* 8(1), pp. 9227–018–27521-y. 2018. DOI: 10.1038/s41598–018–27521-y.
- [16] H. Sun, J. Song, C. Weng, J. Xu, M. Huang, Q. Huang, R. Sun, W. Xiao and C. Sun. Association of decreased expression of the macrophage scavenger receptor MARCO with tumor progression and poor prognosis in human hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* 32(5), pp. 1107–1114. 2017. DOI: 10.1111/jgh.13633.
- [17] S. A. MacParland, J. C. Liu, X. Z. Ma, B. T. Innes, A.M. Bartczak, B. K. Gage, J. Manuel, N. Khuu, J. Echeverri, I. Linares, R. Gupta, M. L. Cheng, L. Y. Liu, D. Camat, S. W. Chung, R. K. Seliga, Z. Shao, E. Lee, S. Ogawa, M. Ogawa, M. D. Wilson, J. E. Fish, M. Selzner, A. Ghanekar, D. Grant, P. Greig, G. Sapisochin, N. Selzner, N. Winegarden, O. Adeyi, G. Keller, G. D. Bader and I. D. McGilvray. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* 9(1), pp. 4383–018–06318–7. 2018. DOI: 10.1038/s41467–018–06318–7.
- [18] Q. Zhu, L. Gong, J. Wang, Q. Tu, L. Yao, J. R. Zhang, X. J. Han, S. J. Zhu, S. M. Wang, Y. H. Li and W. Zhang. miR–10b exerts oncogenic activity in human hepatocellular carcinoma cells by targeting

expression of CUB and sushi multiple domains 1 (CSMD1). *BMC Cancer* 16(1), pp. 806–016–2801–4. 2016. DOI: 10.1186/s12885-016-2801-4.

[19] C. G. Liao, L. M. Kong, P. Zhou, X. L. Yang, J. G. Huang, H. L. Zhang and N. Lu. miR-10b is overexpressed in hepatocellular carcinoma and promotes cell proliferation, migration and invasion through RhoC, uPAR and MMPs. *J. Transl. Med.* 12pp. 234–014–0234-x. 2014. DOI: 10.1186/s12967-014-0234-x.

[20] S. K. Narayana, K. J. Helbig, E. M. McCartney, N. S. Eyre, R. A. Bull, A. Eltahla, A. R. Lloyd and M. R. Beard. The interferon-induced transmembrane proteins, IFITM1, IFITM2, and IFITM3 inhibit hepatitis C virus entry. *J. Biol. Chem.* 290(43), pp. 25946–25959. 2015. DOI: 10.1074/jbc.M115.657346.

[21] M. Zhang, Q. Liu, S. Mi, X. Liang, Z. Zhang, X. Su, J. Liu, Y. Chen, M. Wang, Y. Zhang, F. Guo, Z. Zhang and R. Yang. Both miR-17-5p and miR-20a alleviate suppressive potential of myeloid-derived suppressor cells by modulating STAT3 expression. *J. Immunol.* 186(8), pp. 4716–4724. 2011. DOI: 10.4049/jimmunol.1002989.

[22] Y. Mano, S. Aishima, N. Fujita, Y. Tanaka, Y. Kubo, T. Motomura, A. Taketomi, K. Shirabe, Y. Maehara and Y. Oda. Tumor-associated macrophage promotes tumor progression via STAT3 signaling in hepatocellular carcinoma. *Pathobiology* 80(3), pp. 146–154. 2013. DOI: 10.1159/000346196.

[23] P. Yang, Q. J. Li, Y. Feng, Y. Zhang, G. J. Markowitz, S. Ning, Y. Deng, J. Zhao, S. Jiang, Y. Yuan, H. Y. Wang, S. Q. Cheng, D. Xie and X. F. Wang. TGF-beta-miR-34a-CCL22 signaling-induced treg cell recruitment promotes venous metastases of HBV-positive hepatocellular carcinoma. *Cancer. Cell.* 22(3), pp. 291–303. 2012. DOI: 10.1016/j.ccr.2012.07.023.

[24] N. Tsuji, S. Ishiguro, Y. Sasaki and M. Kudo. CD34 expression in noncancerous liver tissue predicts multicentric recurrence of hepatocellular carcinoma. *Dig. Dis.* 31(5–6), pp. 467–471. 2013. DOI: 10.1159/000355246.

[25] J. S. Jing, H. Li, S. C. Wang, J. M. Ma, L. Q. Yu and H. Zhou. NDRG3 overexpression is associated with a poor prognosis in patients with hepatocellular carcinoma. *Biosci. Rep.* 38(6), pp. 10.1042/BSR20180907. Print 2018 Dec 21. 2018. DOI: BSR20180907.

[26] X. Zhang, J. Sheng, Y. Zhang, Y. Tian, J. Zhu, N. Luo, C. Xiao and R. Li. Overexpression of SCAMP3 is an indicator of poor prognosis in hepatocellular carcinoma. *Oncotarget* 8(65), pp. 109247–109257. 2017. DOI: 10.18632/oncotarget.22665.

[27] M. F. Li, J. J. Zeng, A. P. Pan, Y. H. Lin, H. S. Lin, R. Z. Zhang, L. Yang, Y. Zhang, Y. W. Dang and G. Chen. Investigation of miR-490-3p expression in hepatocellular carcinoma based on reverse transcription-polymerase chain reaction (RT-qPCR) and a meta-analysis of 749 cases. *Med. Sci. Monit.* 24pp. 4914–4925. 2018. DOI: 10.12659/MSM.908492.

- [28] Y. Fang, J. L. Xue, Q. Shen, J. Chen and L. Tian. MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/akt pathway in hepatocellular carcinoma. *Hepatology* 55(6), pp. 1852–1862. 2012. DOI: 10.1002/hep.25576.
- [29] X. Liao, G. Zhu, R. Huang, C. Yang, X. Wang, K. Huang, T. Yu, C. Han, H. Su and T. Peng. Identification of potential prognostic microRNA biomarkers for predicting survival in patients with hepatocellular carcinoma. *Cancer Manag. Res.* 10pp. 787–803. 2018. DOI: 10.2147/CMAR.S161334.
- [30] L. Liu, H. Yue, Q. Liu, J. Yuan, J. Li, G. Wei, X. Chen, Y. Lu, M. Guo, J. Luo and R. Chen. LncRNA MT1JP functions as a tumor suppressor by interacting with TIAR to modulate the p53 pathway. *Oncotarget* 7(13), pp. 15787–15800. 2016. DOI: 10.18632/oncotarget.7487.
- [31] Y. Wang, Q. Huang, T. Deng, B. H. Li and X. Q. Ren. Clinical significance of TRMT6 in hepatocellular carcinoma: A bioinformatics-based study. *Med. Sci. Monit.* 25pp. 3894–3901. 2019. DOI: 10.12659/MSM.913556.
- [32] E. Chen, X. Xu, R. Liu and T. Liu. Small but heavy role: MicroRNAs in hepatocellular carcinoma progression. *Biomed. Res. Int.* 2018pp. 6784607. 2018. DOI: 10.1155/2018/6784607.
- [33] G. Bidkhori, R. Benfeitas, M. Klevstig, C. Zhang, J. Nielsen, M. Uhlen, J. Boren and A. Mardinoglu. Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. *Proc. Natl. Acad. Sci. U.S. A.* 115(50), pp. E11874-E11883. 2018. DOI: 10.1073/pnas.1807305115.
- [34] C. J. Lin, H. Y. Gong, H. C. Tseng, W. L. Wang and J. L. Wu. miR-122 targets an anti-apoptotic gene, bcl-w, in human hepatocellular carcinoma cell lines. *Biochem Biophys Res Commun* 375(3), pp. 315–20. 2008.
- [35] S. Rudraiah, X. Zhang and L. Wang. Nuclear receptors as therapeutic targets in liver disease: Are we there yet? *Annu. Rev. Pharmacol. Toxicol.* 56pp. 605–626. 2016. DOI: 10.1146/annurev-pharmtox-010715-103209.
- [36] M. L. Tornesello, L. Buonaguro, F. Izzo and F. M. Buonaguro. Molecular alterations in hepatocellular carcinoma associated with hepatitis B and hepatitis C infections. *Oncotarget* 7(18), pp. 25087–25102. 2016. DOI: 10.18632/oncotarget.7837.
- [37] Cancer Genome Atlas Research Network. Electronic address: [wheeler@bcm.edu](mailto:wheeler@bcm.edu) and Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 169(7), pp. 1327–1341.e23. 2017. DOI: S0092-8674(17)30639-6.
- [38] C. V. Rao, A. S. Asch and H. Y. Yamada. Frequently mutated genes/pathways and genomic instability as prevention targets in liver cancer. *Carcinogenesis* 38(1), pp. 2–11. 2017. DOI: 10.1093/carcin/bgw118.
- [39] J. Friedman, T. Hastie and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw* 33(1), pp. 1–22. 2010.

# Tables

**Table 1.** Characteristics of subjects whose liver tissues are analyzed by mRNA-seq and miRNA-seq in GU dataset

Characteristics		HCC (*N=30)
<b>Age</b>	<i>Mean (SD)</i>	61.67 (12.2)
<b>Gender</b>	<i>Male</i>	83.33%
<b>Race</b>	<i>AA</i>	30%
	<i>EA</i>	37%
	<i>Asian</i>	33%
<b>BMI</b>	<i>Mean (SD)</i>	26.17 (5.4)
<b>Smoking</b>	<i>Yes</i>	66.67%
	<i>No</i>	33.33%
<b>Alcohol</b>	<i>Yes</i>	47%
	<i>No</i>	53%
<b>HCV Serology</b>	<i>HCV Ab+ Reactive</i>	37%
	<i>HCV Ab+ Non-Reactive</i>	60%
	<i>HCV Ab+ Unavailable</i>	3.33%
	<i>HCV RNA+ Reactive</i>	16.67%
	<i>HCV RNA+ Non-Reactive</i>	17%
	<i>HCV RNA+ Unavailable</i>	63%
<b>HBV Serology</b>	<i>anti HBC+ Reactive</i>	10%
	<i>anti HBC+ Non-Reactive</i>	3.33%
	<i>anti HBC+ Unavailable</i>	87%
	<i>HBs Ag+ Reactive</i>	17%
	<i>HBs Ag+ Non-Reactive</i>	70%
	<i>HBs Ag+ Unavailable</i>	13%
<b>MELD</b>	<i>Mean (SD)</i>	10.04 (4.5)
	<i>MELD &lt;10</i>	56.67%
<b>AFP</b>	<i>Median (IQR)</i>	4.4 (14.05)
<b>Child Pugh Score</b>	<i>Mean (SD)</i>	6.14 (1.5)
	<i>Median (IQR)</i>	6 (1)
<b>Child Pugh Class</b>	<i>A</i>	73%
	<i>B</i>	13.33%
	<i>C</i>	3.33%
<b>AST</b>	<i>Median (IQR)</i>	123 (158)
<b>ALT</b>	<i>Median (IQR)</i>	117 (152)
<b>HCC Stage</b>	<i>Stage I</i>	43%
	<i>Stage II</i>	24%
	<i>Stage III &amp; IV</i>	33%

\* Adjacent normal tissue for the HCC patients are available for this study

**Table 2.** Characteristics of subjects from TCGA-LIHC cohort

Characteristics		HCC (*N=49)
Age	Mean (SD)	60.77 (16.2)
Gender	Male	55.10%
Race	AA	14%
	EA	67%
	Asian	12%
	not report	6.12%
HCC Stage	Stage I	34.69%
	Stage II	22.45%
	Stage III & IV	26.53%
	Unknown	16.33%

\* Adjacent normal tissue for 49 HCC patients are available for this study

**Table 3:** Number of mRNAs, miRNAs, and miRNA-mRNA pairs selected by HIM using the GU dataset.

	Statistical Analysis			Mechanistic Model	Clinical Model
	# of detected features	# of selected features with FDR < 0.05	total # of pairs	selected pairs (#mRNA) (#miRNA)	selected pairs (#mRNA) (#miRNA)
miRNA	2,195	238	1,368,738	26,080 (3,632) (238)	157 (19) (90)
mRNA	20,354	5,751			

**Table 4:** Number of molecules and miRNA-mRNA associations selected by HIM using the TCGA dataset.

				Mechanistic Model	Clinical Model
	# of detected features	# of selected features with FDR < 0.05	total # of pairs	selected pairs (#mRNA) (#miRNA)	selected pairs (#mRNA) (#miRNA)
miRNA	740	354	3,905,682	75,668 (8,098) (354)	263 (14) (136)
mRNA	18,407	11,033			

## Figures

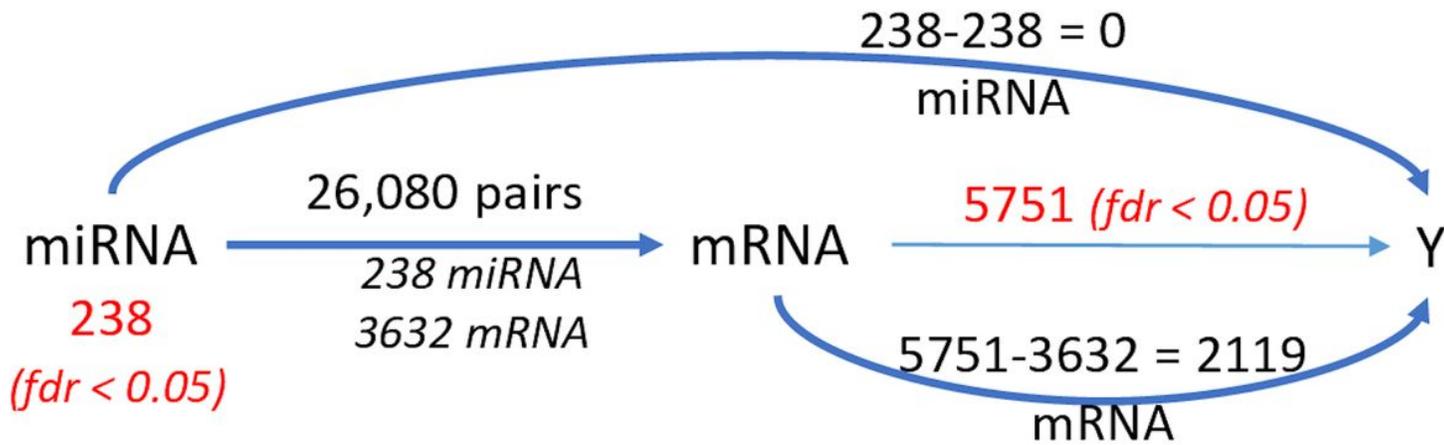


Figure 1

### Verified pairs

miRNA	FC	mRNA	FC	Confidence
hsa-miR-671-5p	↑	CD34	↑	Moderate (predicted)
hsa-miR-10b-5p	↑	EBF2	↑	Moderate (predicted)
hsa-miR-106b-5p	↑	GNS	↑	Moderate (predicted)
hsa-miR-532-3p	↑	IFITM10	↓	Moderate (predicted)
hsa-miR-197-3p	↑	TERT	↑	High (predicted)

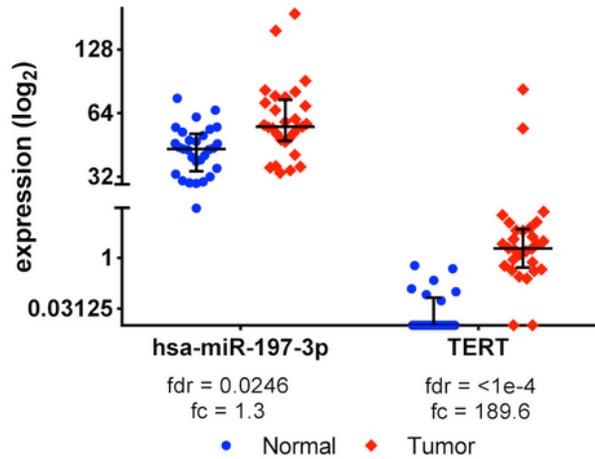
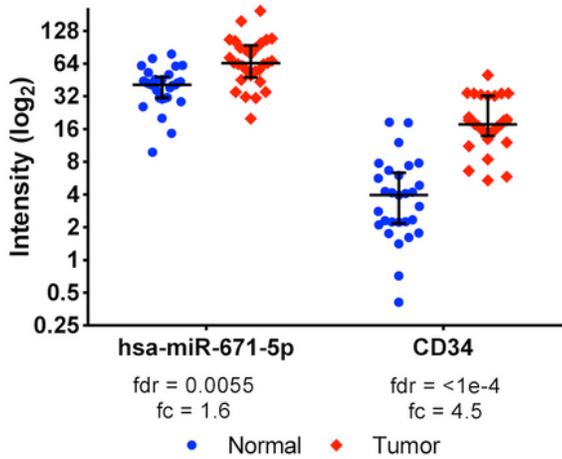
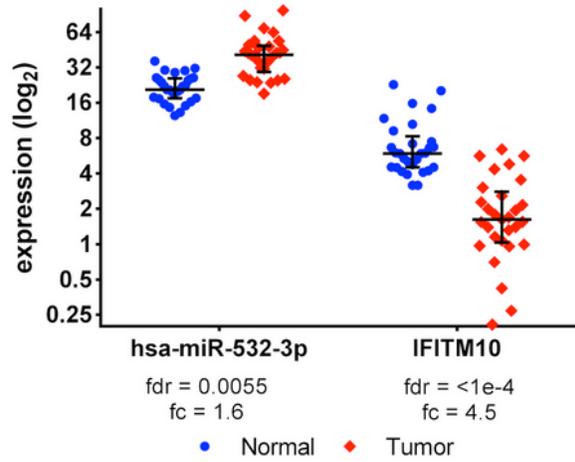
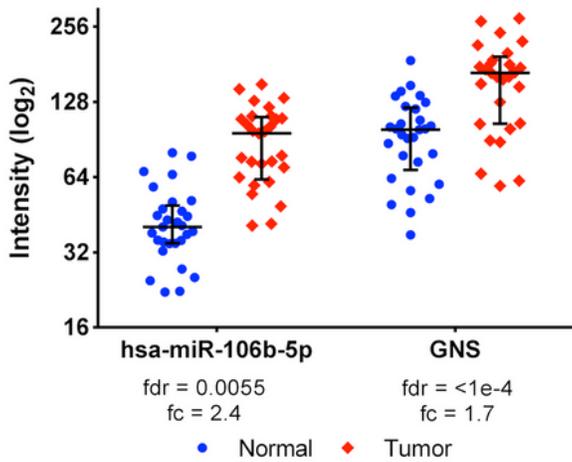
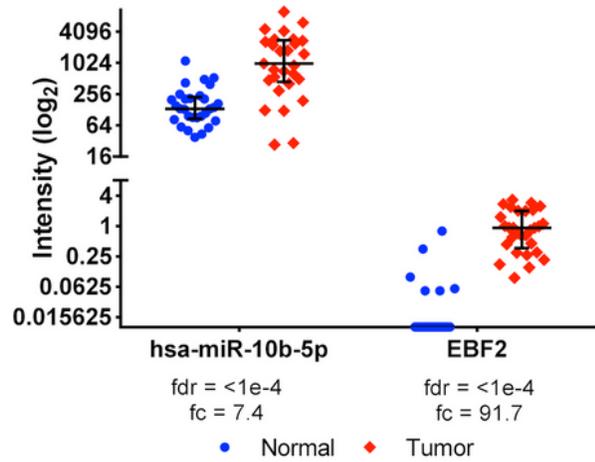


Figure 2

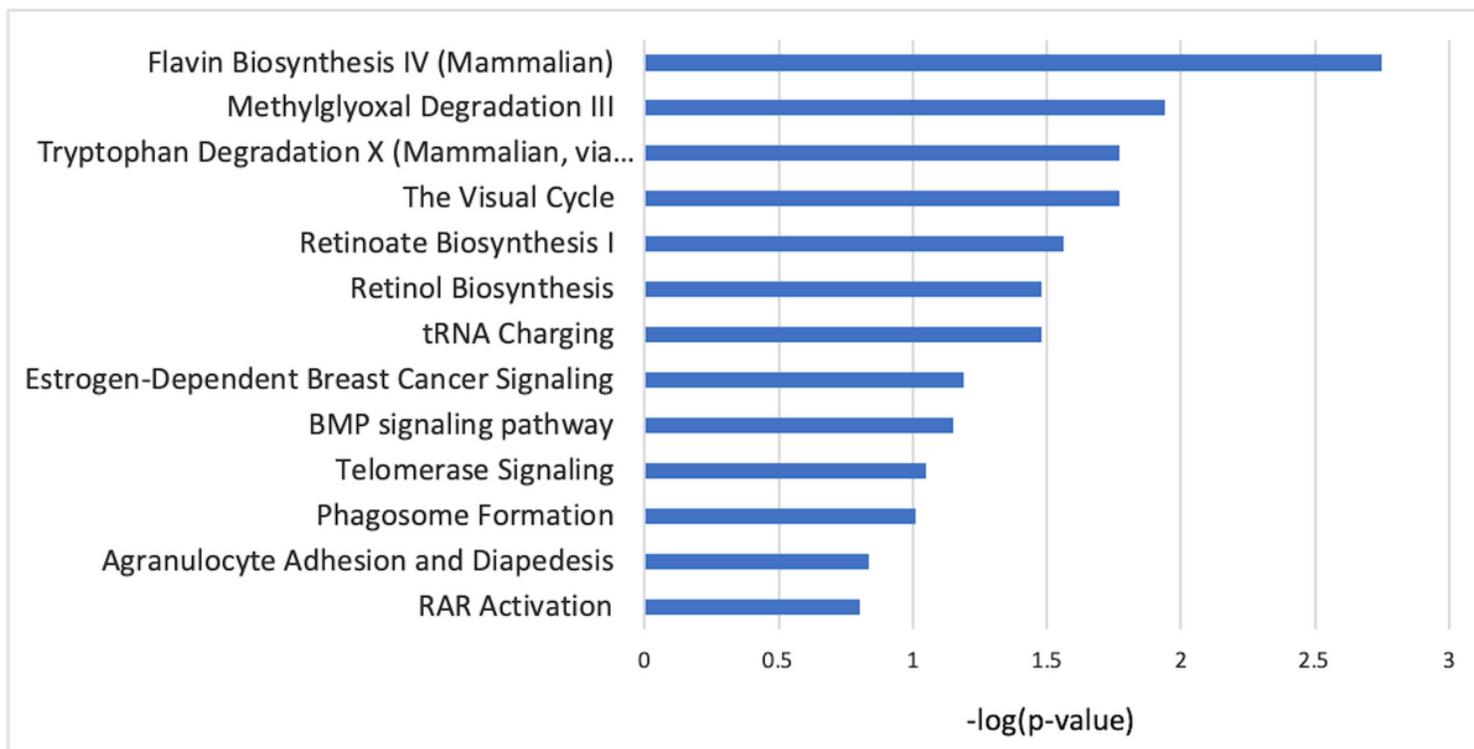


Figure 3

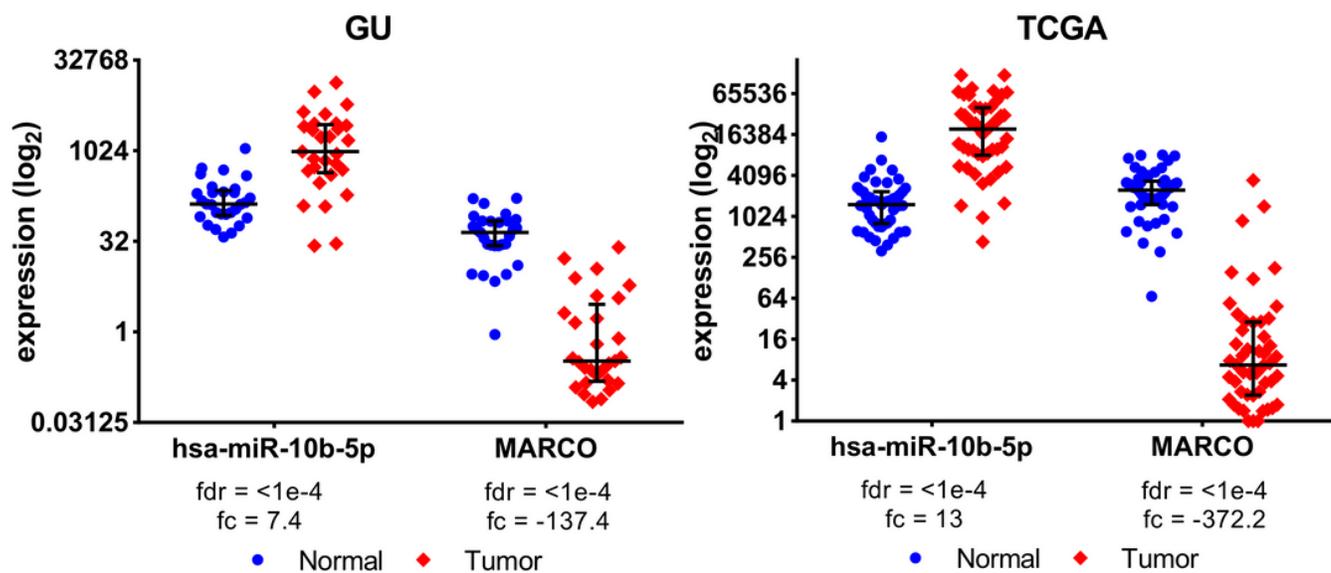


Figure 4

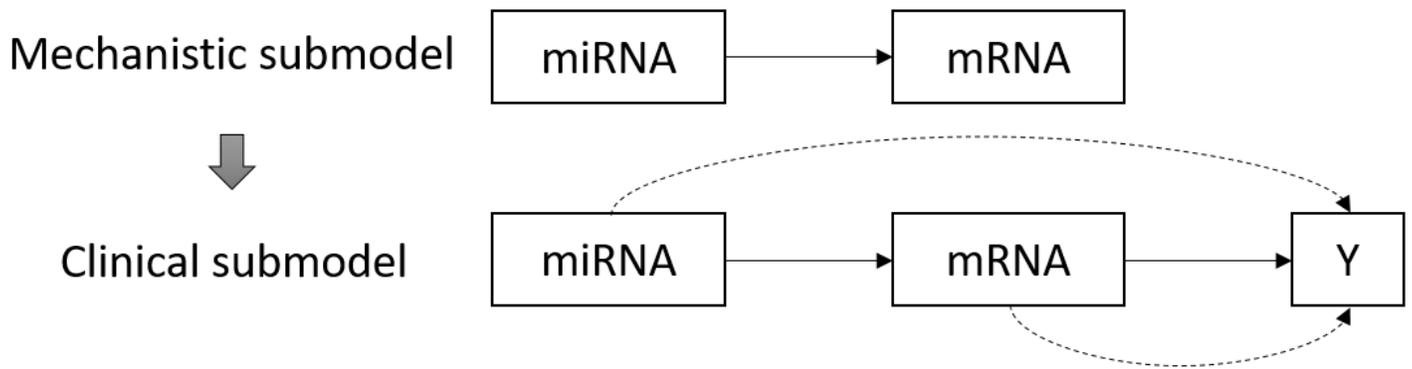


Figure 5

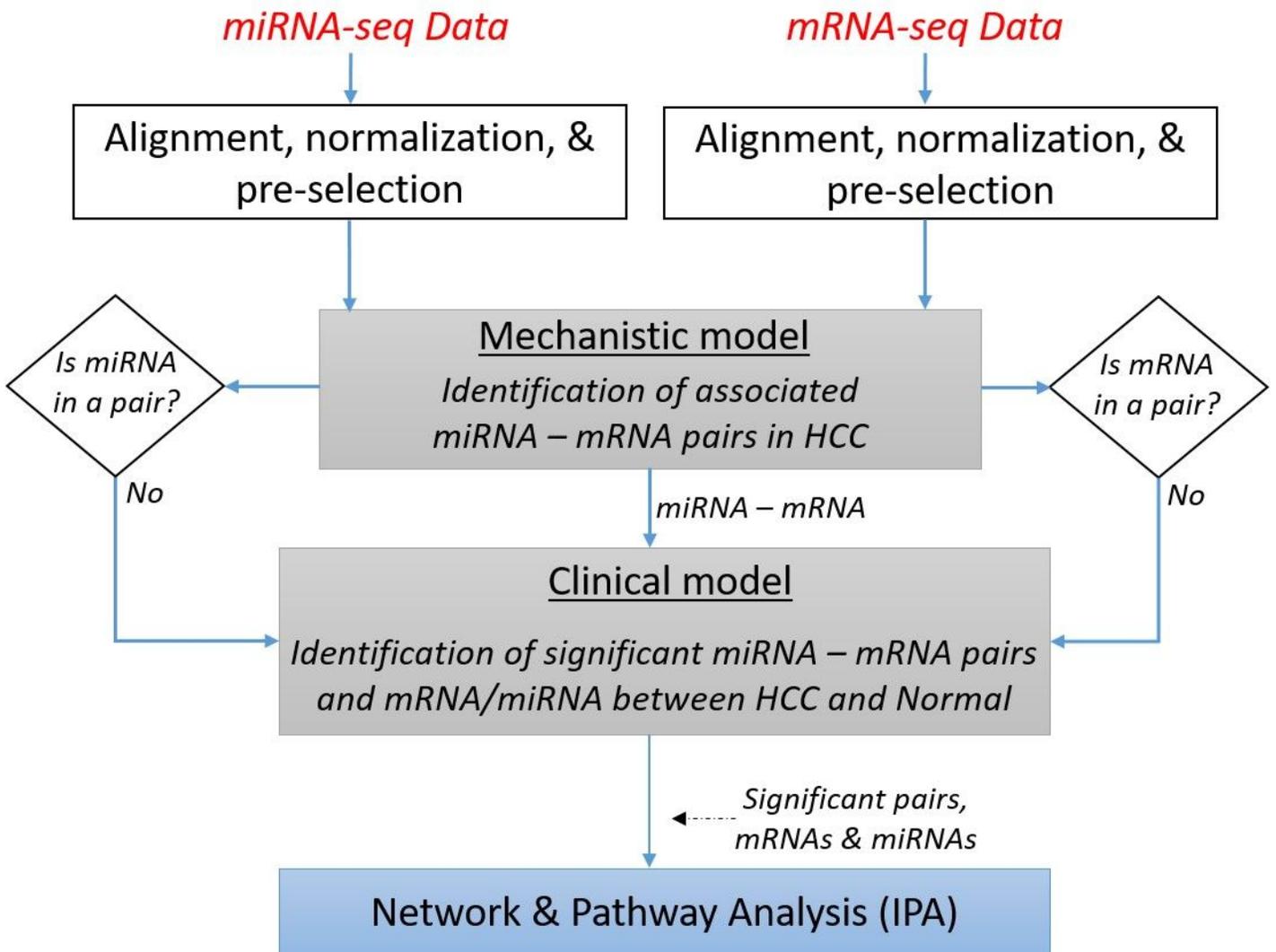


Figure 6

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure6.png](#)
- [SupplementaryFigure4.png](#)
- [SupplementaryFigure5.png](#)
- [SupplementaryFigure2.png](#)
- [SupplementaryFigure1.pdf](#)
- [SupplementaryFigure3.pdf](#)