Research Article

Investigating Pathogenic and Hepatocarcinogenic Mechanisms from Normal Liver to HCC by Constructing Genetic and Epigenetic Networks via Big Genetic and Epigenetic Data Mining and Genome-Wide NGS Data Identification

Cheng-Wei Li ⓘ **, Yu-Kai Chiu, and Bor-Sen Chen** ⓘ

National Tsing Hua University, Kuang-Fu Road, Hsinchu 30013, Taiwan

Correspondence should be addressed to Bor-Sen Chen; bschen@ee.nthu.edu.tw

The prevalence of hepatocellular carcinoma (HCC) is still high worldwide because liver diseases could develop into HCC. Recent reports indicate nonalcoholic fatty liver disease and nonalcoholic steatohepatitis (NAFLD&NASH) and primary biliary cirrhosis and primary sclerosing cholangitis (PBC&PSC) are significant of HCC. Therefore, understanding the cellular mechanisms of the pathogenesis and hepatocarcinogenesis from normal liver cells to HCC through NAFLD&NASH or PBC&PSC is a priority to prevent the progression of liver damage and reduce the risk of further complications. By the genetic and epigenetic data mining and the system identification through next-generation sequencing data and its corresponding DNA methylation profiles of liver cells in normal, NAFLD&NASH, PBC&PSC, and HCC patients, we identified the genome-wide real genetic and epigenetic networks (GENs) of normal, NAFLD&NASH, PBC&PSC, and HCC patients. In order to get valuable insight into these identified genome-wide GENs, we then applied a principal network projection method to extract the corresponding core GENs for normal liver cells, NAFLD&NASH, PBC&PSC, and HCC. By comparing the signal transduction pathways involved in the identified core GENs, we found that the hepatocarcinogenesis through NAFLD&NASH was induced through DNA methylation of *HIST2H2BE*, *HSPB1*, *RPL30*, and *ALDOB* and the regulation of miR-21 and miR-122, and the hepatocarcinogenesis through PBC&PSC was induced through DNA methylation of *RPL23A*, *HIST2H2BE*, *TIMP1*, *IGF2*, *RPL30*, and *ALDOB* and the regulation of miR-29a, miR-21, and miR-122. The genetic and epigenetic changes in the pathogenesis and hepatocarcinogenesis potentially serve as potential diagnostic biomarkers and/or therapeutic targets.

1. Introduction

The liver is the largest internal organ of the human body and is involved in many important functions that require high harmonization to control biochemical processes [1]. However, alterations of molecular mechanisms in the liver have been linked to liver diseases such as hepatitis, steatosis, cirrhosis, and hepatocellular carcinoma (HCC). HCC is the end stage of general liver diseases that were classified as nonalcoholic fatty liver disease (NAFLD), autoimmune liver disease, viral hepatitis, alcoholic liver disease, and others. The prevalence of NAFLD, including the more aggressive

nonalcoholic steatohepatitis (NASH), is increasing with the growing epidemics of diabetes and obesity. The most frequent symptoms in autoimmune liver diseases (i.e., primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC)), which lead to cholestasis, are fatigue, jaundice, hyperpigmentation, or pruritus. Half a million patients are diagnosed with HCC worldwide each year. It has no significant symptoms in the early stage of HCC through NAFLD&NASH or PBC&PSC and tends to be observed in the advanced stage [2, 3]. Apart from regular surveillance for HCC, understanding the cellular mechanisms of the pathogenesis and hepatocarcinogenesis from normal liver cells to

HCC through NAFLD&NASH or PBC&PSC is a priority to slow the progression of liver damage and reduce the risk of further complications.

Recent reports indicate that NAFLD, NASH, PBC, and PSC are responsible for 13%, 4–27%, 3.8%, and 2% HCC cases, respectively. HCC is one of the most deadly malignant tumors, with five-year survival rates ranging from 3% to 28%. Although the reduction of triglycerides in the liver of NAFLD&NASH patients by regular exercise and liver transplantation in PBC&PSC patients can slow the progression of liver damage [4–6], these treatments have some limitations for the complete remedy of liver damage from the perspective of molecular mechanism.

The short noncoding RNA sequences (approximately 21-nucleotide long), microRNAs (miRNAs), were assumed to act by partial repression or degradation of targeted mRNAs. It has been observed that diet-induced obesity in mice results in the differential expression of 6% miRNAs [7]. These changes were similarly observed in human hepatocytes and immortalized liver cell lines exposed to various fatty acids [8]. miRNAs, mostly secreted from cells through active energy-dependent processes via storage in microvesicles, can also circulate freely in the blood [9]. miRNAs were also released by the increased cell death, as it occurs in NAFLD via ballooning degeneration, into the circulation to regulate diverse biological processes including the immune system, cell proliferation, differentiation, and development leading to major advances in the pathomechanism and the hepatogenesis mechanism [10]. Therefore, miRNAs have been proposed as attractive diagnostic biomarkers for investigating, noninvasively, the pathogenesis of NAFLD&NASH (or PBC&PSC) and the carcinogenesis of HCC [11].

Epigenetic regulation, including DNA methylation, histone modification, and chromatin remodelling, results in potentially reversible alterations in gene expression that do not involve permanent changes to the DNA sequence. It has been reported that maternal western diet increases the susceptibility of male offspring to NAFLD [12]. Additionally, by comparing liver biopsies before and after bariatric surgery, it has been observed that weight loss after bariatric surgery leads to NAFLD-associated methylation changes, partially reversible [13]. Therefore, abnormal DNA methylation is the hallmark of the pathogenesis of NAFLD&NASH (or PBC&PSC) and the carcinogenesis of HCC.

DNA methylation, including DNA methylation in the coding region (coding-region methylation) or in the promoter region (promoter-region methylation), regulates the process of mRNA transcription. Promoter region methylation directly affects the binding affinities of miRNAs and transcription factors (TFs) to vary transcriptional profiles [14, 15] while coding region methylation directly attenuates the basal gene expression. The characterization of aberrant DNA methylation is involved in the pathogenesis of NAFLD&NASH (or PBC/PSC) and the carcinogenesis of HCC.

In order to investigate the cellular mechanisms of NAFLD&NASH (or PBC&PSC) pathogenesis and NAFLD&NASH-associated (or PBC&PSC-associated) hepatocarcinogenesis in the liver cells of patients, we constructed

the stochastic models of the genome-wide genetic and epigenetic network (GEN) in human cells based on molecular mechanisms, including TF regulations, miRNA repressions, DNA methylation of genes, and protein-protein interactions (PPIs). By applying a system identification method and a system order detection scheme to prune the false positives from the candidate genome-wide GEN constructed by the genetic and epigenetic data mining, we identified the real genome-wide GENs of liver cells in patients with normal liver, NAFLD&NASH, PBC&PSC, and HCC using their corresponding genome-wide mRNA and miRNA expression data and its corresponding DNA methylation profiles. Although the nonsignificant network connections of a liver cell condition have been pruned out from a genome-wide candidate GEN based on a system order detection scheme using the genome-wide expression data, the real genome-wide GEN of a liver cell condition is still complex and complicated owing to the network involving multiple molecular mechanisms. We then applied the principal network projection (PNP) method to the real genome-wide GEN to extract the core network nodes (core proteins/genes/miRNAs), which could constitute signal transduction pathways; i.e., the pathways mediate the intracellular signaling cascade, from the significant network linking energy perspective. By comparing the core pathways with the major differences between normal liver and NAFLD&NASH and between NAFLD&NASH and HCC, we could unravel respectively the cellular mechanisms of NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis. Similarly, by comparing the core pathways with the major differences between normal liver and PBC&PSC and between PBC&PSC and HCC, we could unravel the cellular mechanisms of PBC&PSC pathogenesis and PBC&PSC-associated hepatocarcinogenesis, respectively. We finally proposed network biomarkers, i.e., a set of proteins and miRNAs, as potential diagnostic biomarkers and novel therapeutic drug targets in each liver condition for preventing NAFLD&NASH or PBC&PSC pathogenesis or NAFLD&NASH-associated or PBC&PSC-associated hepatocarcinogenesis. The proposed system biology method also has a potential for use in other liver disease screenings and treatments.

2. Results

This study focused on the construction of genome-wide GENs and their core GENs for different liver diseases, to further investigate the hepatocarcinogenic mechanisms, which were then used to design potential drugs for preventing hepatocarcinogenesis in NAFLD&NASH (or PBC&PSC) patients. The hepatocarcinogenic mechanisms for the progression from normal liver cells to HCC were divided into two progression paths with four progression stages, as shown in Figure 1(a). The upper progression path in Figure 1(a) comprises the pathogenesis and hepatocarcinogenesis from normal liver cells to HCC through NAFLD&NASH, represented as NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis, respectively. The lower progression path in Figure 1(a) comprises the pathogenesis and hepatocarcinogenesis from normal
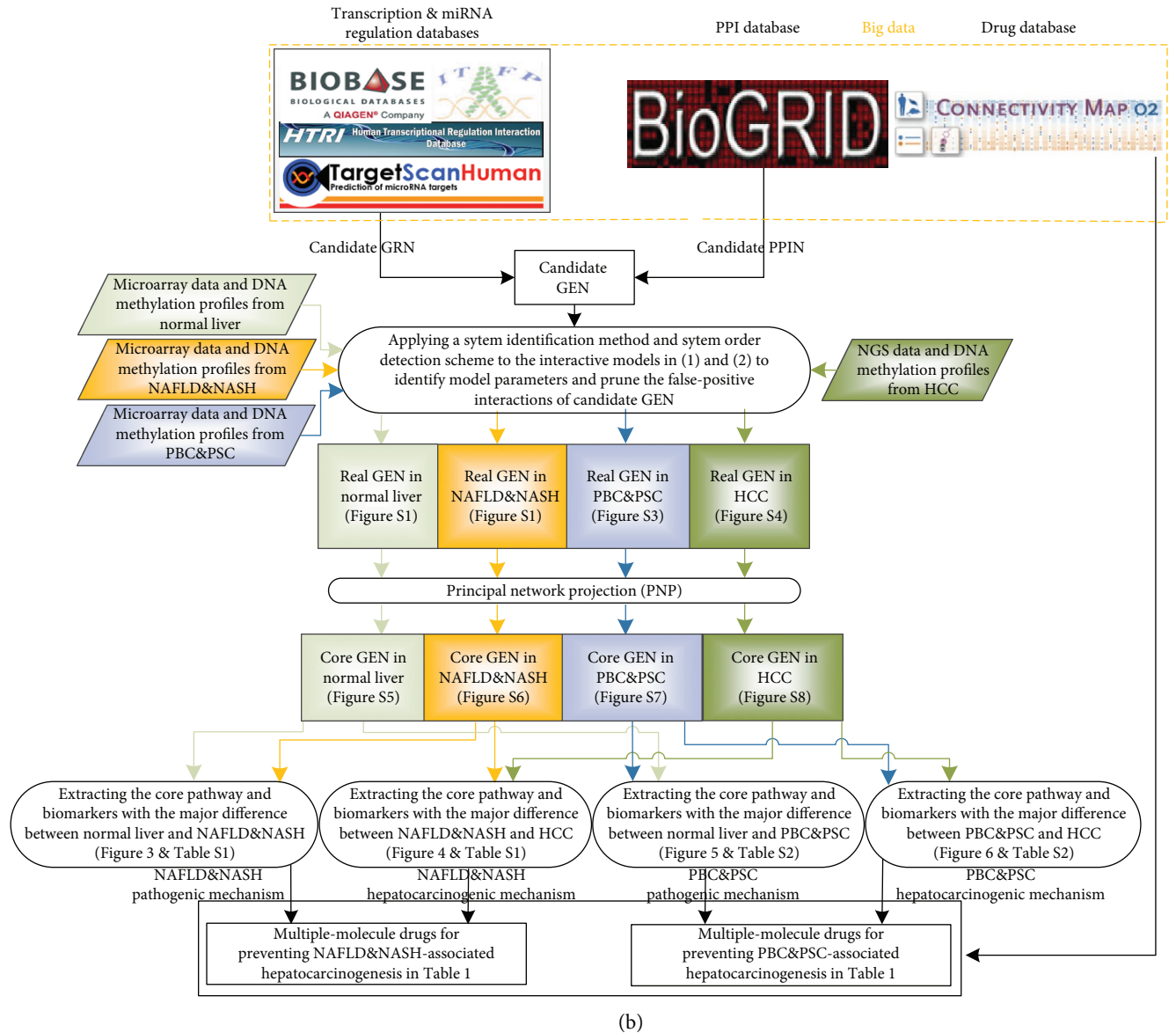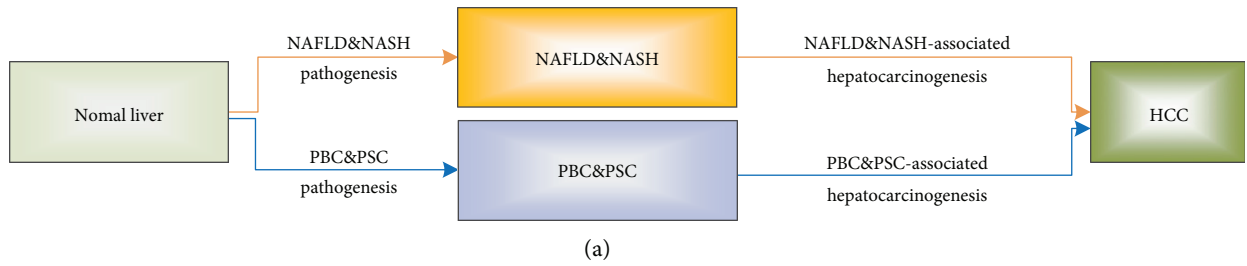
Figure 1: (a) Progression path in pathogenesis and hepatocarcinogenesis. (b) Flowchart for identifying core GENs and pathways.

liver cells to HCC through PBC&PSC, represented as PBC&PSC pathogenesis and PBC&PSC-associated hepato-carcinogenesis, respectively. Because previous studies have mostly investigated significant proteins or genes instead of using big database mining and genome-wide high-throughput data identification, they may have ignored the effects of the neglected proteins or genes in signaling pathways. In this study, a flowchart for constructing genome-wide GENs, core GENs, and core pathways with network biomarkers for the pathogenesis and hepatocarcinogenesis of HCC is shown in Figure 1(b). First, we constructed the genome-wide candidate GEN in human cells by big genetic and epigenetic database mining. We applied a system identification method and a system order detection scheme to the system models of a genome-wide candidate GEN in human cells to prune the false positives in the candidate GEN and to

finally identify the real genome-wide GENs in normal cells, NAFLD&NASH, PBC&PSC, and HCC using genome-wide microarray data and NGS data and their corresponding DNA methylation profiles. Moreover, we applied PNP to the real genome-wide GENs of the four cells to obtain their core GENs in normal cells, NAFLD&NASH, PBC&PSC, or HCC by comparing their core GENs. Furthermore, the signaling pathways of these core GENs were then compared to unravel the hepatocarcinogenic mechanisms from normal liver cells to HCC through NAFLD&NASH or PBC&PSC (see Materials and Methods). The pathogenic and hepatocarcinogenic mechanisms provide valuable insight into the potential drugs for treating patients with liver diseases or HCC.

We have identified genome-wide GENs for normal liver, NAFLD&NASH, PBC&PSC, and HCC using database mining, microarray data, NGS data, system modeling, and systematic analysis, as shown in Figures S1–S4, respectively (see Materials and Methods). Because the genome-wide real GENs are extremely complicated, we further unraveled the cellular mechanisms of pathogenesis and hepatocarcinogenesis in liver cells from the signal transduction pathway perspective. Therefore, we applied PNP to the genome-wide real GEN to obtain the projection value of each node based on the identified network parameters through the Akaike information criterion (AIC). The top nodes with the highest projection values constitute the core GENs for normal liver, NAFLD&NASH, PBC&PSC, and HCC (Figures S5–S8, respectively), which have complete connections in signal transduction, i.e., whole signaling cascade from receptors to TFs (in Figures 2–4, respectively). Finally, we further clarified the cellular mechanisms of pathogenesis and hepatocarcinogenesis based on the core signaling pathways of each progression stage by making use of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and the Gene Ontology (GO) database. Based on the core signaling pathways, we also proposed the network biomarkers for a potential drug design to prevent further liver damage of the patients with NAFLD&NASH or PBC&PSC.

2.1. Ten-Fold Cross-Validation to Test the Robustness of the Models and AIC.

In order to test the robustness of the proposed models in (1) and (2) and AIC in (10), we applied a ten-fold cross-validation test to the mRNA expressions in each cell. We randomly split each data set in normal liver, NAFLD&NASH, PBC&PSC, and HCC into ten sets (e.g., normal liver from 62 patients into 6, 6, 6, 6, 6, 6, 6, 6, 7, and 7 patients). We then repeated the system identification method ten times in normal liver, NAFLD&NASH, PBC&PSC, and HCC. In the protein-protein interaction network (PPIN) of the identified GEN, 88.48% (122,545/138,498) pairs in normal liver, 86.48% pairs (108,357/125,304) in NAFLD&NASH, 86.48% pairs (108,300/125,221) in PBC&PSC, and 14.51% pairs (15,506/106830) in HCC were selected independent of the ten split sets. Additionally, in the gene regulatory network (GRN) of the identified GEN, 19.62% (1212/6178) pairs in normal liver, 18.84% pairs (2298/12,197) in NAFLD&NASH, 19% pairs (2275/11,972) in PBC&PSC, and 33.74% pairs (10,752/31,869) in

HCC were selected independent of the ten split sets. Because the low independency of HCC in PPIN results from its largest sample size (369 patients) and a target gene has less candidate TFs (7.3 ± 9.2) in candidate GRN on average as compared to the candidate connecting partners of a protein (15.9 ± 81.9) in candidate PPIN on average, the proposed system identification method can identify a robust real GEN, especially for large sample sizes.

2.2. Analysis of Core Pathways to Investigate Underlying Cellular Mechanisms for NAFLD&NASH Pathogenesis and NAFLD&NASH-Associated Hepatocarcinogenesis.

In the upper progression path of Figure 1(a), the identified core pathways comprise 37 core protein-coding genes, i.e., network biomarkers, involved in at least one core GEN of normal liver cells, NAFLD&NASH, and HCC. The projection values of 37 network biomarkers $D(k)$ in (17) are also shown in Table S1. We projected these core genes to the KEGG and the GO database mining to get the relevant biological functions and then obtained the core pathways for NAFLD&NASH pathogenesis (Figure 5) and NAFLD&NASH-associated hepatocarcinogenesis (Figure 2).

In the pathogenesis of NAFLD&NASH in Figure 5, we identified eight genes having expression difference between normal liver cells and NAFLD&NASH, i.e., $HIST2H2BE$ (p value $\leq 1.06 \times 10^{-1}$), $RFC5$ (p value $\leq 1.5 \times 10^{-2}$), $HSPB1$ (p value $\leq 6.6 \times 10^{-2}$), $ZNF480$ (p value $\leq 1.0 \times 10^{-2}$), $TUBA1C$ (p value $\leq 3.48 \times 10^{-1}$), $RPL30$ (p value $\leq 1 \times 10^{-3}$), $FRAT2$ (p value $\leq 1.08 \times 10^{-1}$), and $TRMT1$ (p value $\leq 5.33 \times 10^{-1}$). Furthermore, three of those genes had basal level differences that might have been caused by DNA methylation on the corresponding genes. According to DNA methylation profiles in the normal liver cells and NAFLD&NASH, we found one gene that had undergone hypermethylation, i.e., $HIST2H2BE$ (p value $\leq 2.05 \times 10^{-1}$), and two genes that had undergone hypomethylation, i.e., $RPL30$ (p value $\leq 4.90 \times 10^{-1}$) and $TRMT1$ (p value $\leq 4.40 \times 10^{-1}$) in a comparison of normal liver cells with NAFLD&NASH cells.

In the hepatocarcinogenesis of NAFLD&NASH in Figure 2, we also identified eight genes with differences in expression between NAFLD&NASH and HCC, i.e., $HIST2H2BE$ (p value $\leq 1.00 \times 10^{-3}$), $RFC5$ (p value $\leq 1.00 \times 10^{-3}$), $HSPB1$ (p value $\leq 1.00 \times 10^{-3}$), $ZNF480$ (p value $\leq 1.00 \times 10^{-3}$), $TUBA1C$ (p value $\leq 1.00 \times 10^{-3}$), $RPL30$ (p value $\leq 1.00 \times 10^{-3}$), $FRAT2$ (p value $\leq 1.00 \times 10^{-3}$), and $ALDOB$ (p value $\leq 1.5 \times 10^{-2}$). Furthermore, five of those genes had basal level differences that might have been caused by DNA methylation on the corresponding genes. According to DNA methylation profiles of NAFLD&NASH and HCC, we found three genes that had undergone hypermethylation, i.e., $ALDOB$ (p value $\leq 8.50 \times 10^{-1}$), $HSPB1$ (p value $\leq 1.00 \times 10^{-3}$), and $RPL30$ (p value $\leq 1.00 \times 10^{-3}$), and two genes that had undergone hypomethylation, i.e., $TUBA1C$ (p value $\leq 1.00 \times 10^{-3}$) and $FRAT2$ (p value $\leq 1.00 \times 10^{-3}$) in a comparison of NAFLD&NASH with HCC cells.

After extracting two signaling pathways between normal cells and NAFLD&NASH and between NAFLD&NASH and HCC from the core GENs (Figures S5, S6, and S8) as shown
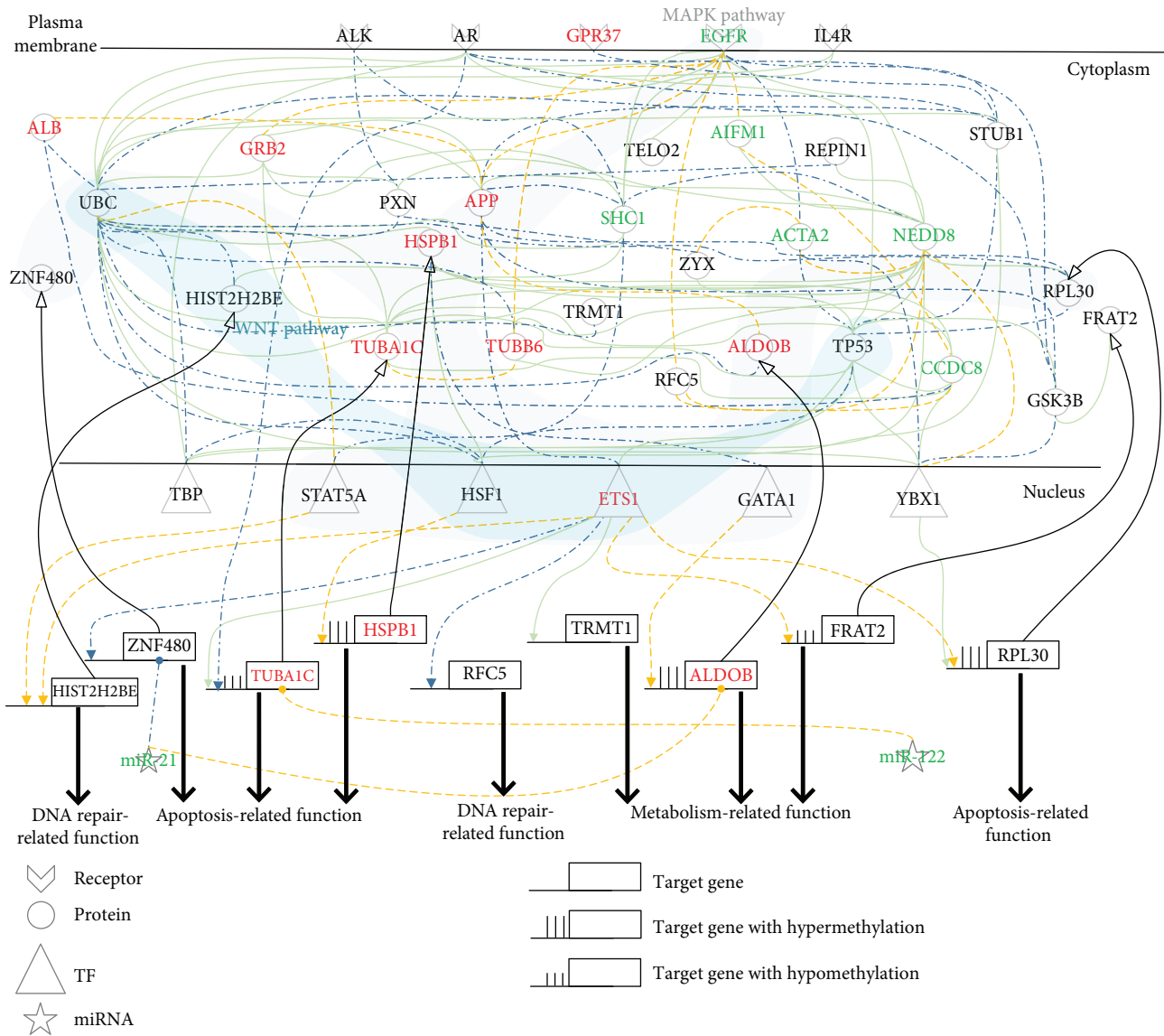
FIGURE 2: Signaling pathways for investigating cellular mechanisms of progression for pathogenesis from NAFLD&NASH to HCC. The blue dash-dotted lines and the yellow dotted lines represent the specific edges of the real GENs, including PPINs and GRNs, in NAFLD&NASH and HCC, respectively; the green solid lines denote the common edges between NAFLD&NASH and HCC; the red and black symbols indicate specific core members of the real GENs in NAFLD&NASH and HCC, respectively; and the green symbols are the common core members between NAFLD&NASH and HCC. In this figure, these core members are padded to complete the relevant signaling pathways for the convenience of analysis. The cytosolic proteins involved in the MAPK and the WNT pathways were highlighted by light-gray and light-blue bands, respectively. Dysfunctions of both metabolism and apoptosis via DNA hypermethylation and dysregulation of miR-21 contribute to tumorigenesis from NAFLD&NASH to HCC. Dysregulation of miR-21 and miR-122 contributes to tumor invasion and metastasis in HCC.

in Figures 5 and 2, respectively, we further investigated the cellular mechanisms underlying progression of NAFLD&NASH pathogenesis in Figure 5 and NAFLD&NASH-associated hepatocarcinogenesis in Figure 2. In Figure 5, we observed that in normal liver cells, the insulin receptor (ALK) and the prosaposin receptor (GPR37) were activated to allow ubiquitin C (UBC) to modulate two TFs (ETS1 and TBP) in the WNT and the MAPK signaling pathways to induce DNA repair through the mediation of *HIST2H2BE*, apoptosis through the mediation of *HSPB1*, and metabolism through the mediation of *FRAT2*. To maintain proper DNA

repair function, apoptosis and metabolism should be silenced by miR-21, miR-122, and miR-214. For example, miR-21 and miR-214 can silence *HIST2H2BE* and *HSPB1* in normal liver cells without significant liver damage, and miR-122 can silence *FRAT2* to avoid metabolism dysfunction. However, both excessive accumulation of hepatic triglycerides and abnormal DNA repair through the dysregulation of *HIST2H2BE* caused by hypermethylation caused normal liver cells to progress to NAFLD&NASH. In NAFLD&NASH, we observed that oxidative stress could activate epidermal growth factor receptor (EGFR) to facilitate (i) UBC
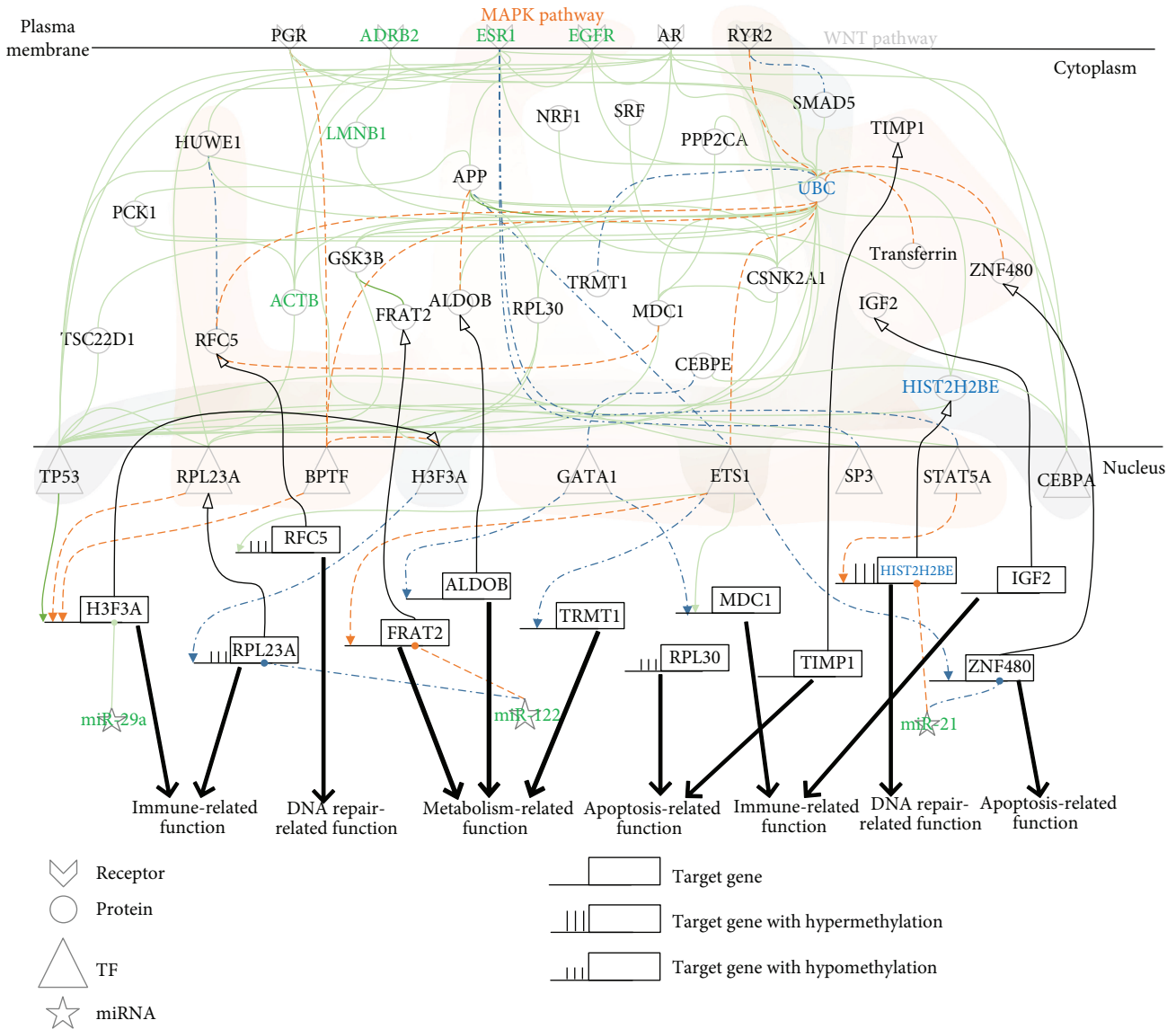
FIGURE 3: Signaling pathways for investigating cellular mechanisms of progression for pathogenesis from normal to PBC&PSC. The orange dotted lines and the blue dash-dotted lines represent the specific edges of the real GENs, including PPINs and GRNs, in normal and PBC&PSC, respectively; the green solid lines denote the common edges between normal and PBC&PSC; the blue and black symbols indicate specific core members of the real GENs in normal and PBC&PSC, respectively; and the green symbols are the common core members between normal and PBC&PSC. In this figure, these core members are padded to complete the relevant signaling pathways for the convenience of analysis. The cytosolic proteins involved in the MAPK and the WNT pathways were highlighted by light-green and light-blue bands, respectively. Defect of autoimmune response via DNA hypomethylation contributes to the dysfunction of DNA repair and gives rise to PBC&PSC. Dysregulation of miR-21 and miR-122 contributes to the dysfunction of apoptosis and autoimmune in PBC&PSC, respectively.

to repress HSF1 in the WNT and MAPK signaling pathways; (ii) APP to activate ETS1 in the MAPK signaling pathway to induce DNA repair function through the mediation of *RFC5*, to activate metabolism through the mediation of *TRMT1*, and to inhibit apoptosis through the mediation of *ZNF480*; and (iii) SHC1 to repress STAT5A in the MAPK signaling pathway. Furthermore, androgen receptor (AR) directly inhibited apoptosis through the mediation of *TUBA1C*. Intriguingly, *HSPB1* caused antiapoptosis without miR-214 silencing and *RFC5* caused DNA repair function. Thus,

antiapoptosis and DNA repair could overcome liver damage through the MAPK signaling pathway.

With a series of dysregulations and mutations, NAFLD&NASH developed into HCC through the cellular mechanisms of hepatocarcinogenesis progression, as shown in Figure 2. For example, *ZNF480* silenced by abnormal miR-21 can inhibit apoptosis to facilitate tumorigenesis. In HCC, liver damage and dysregulation due to mutations are still accumulated to facilitate EGFR to repress UBC. Then, UBC modulated STAT5A and HSF1 in the WNT and the
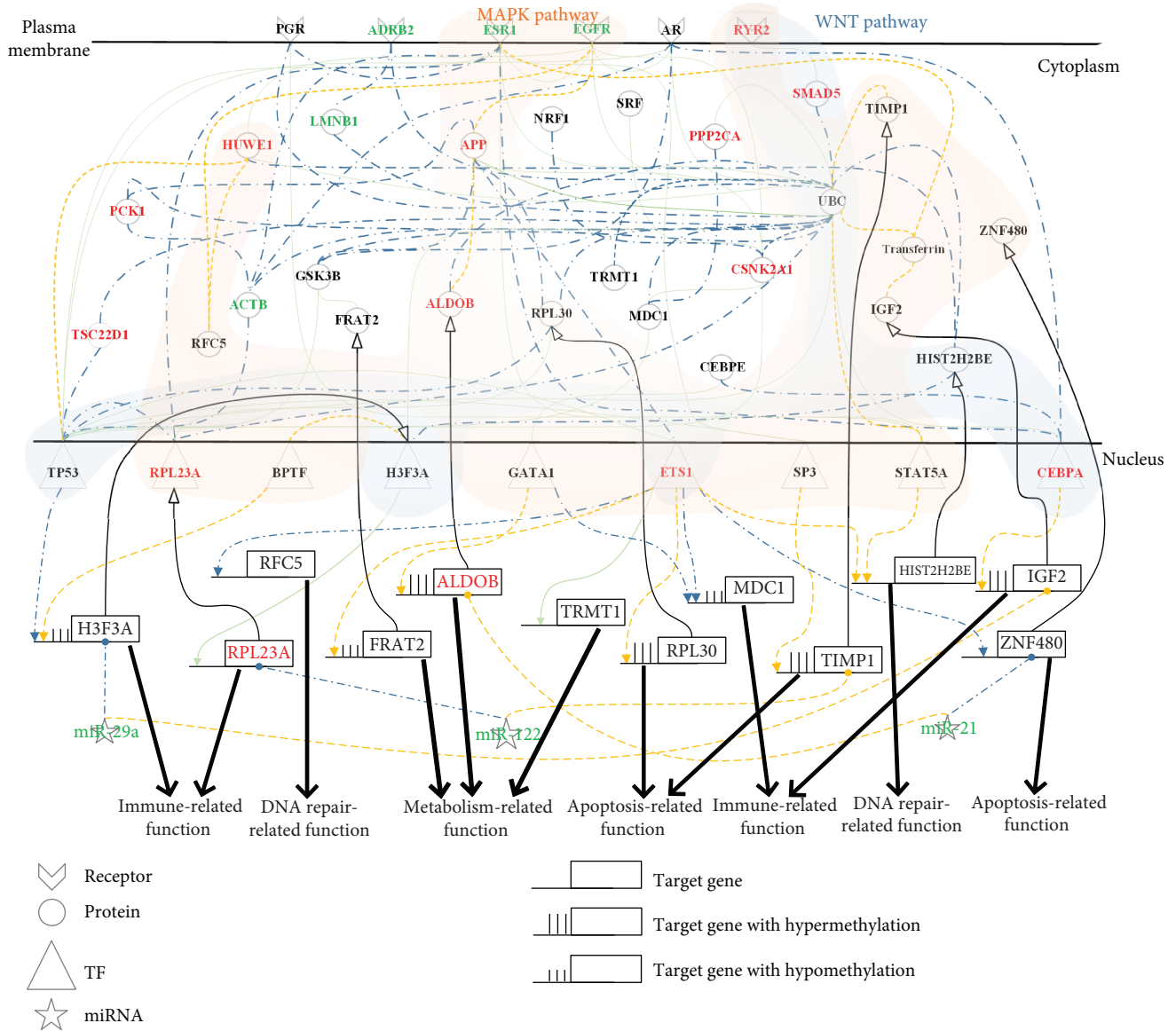
FIGURE 4: Signaling pathways for investigating cellular mechanisms of progression for pathogenesis from PBC&PSC to HCC. The blue dash-dotted lines and the yellow dotted lines represent the specific edges of the real GENs, including PPINs and GRNs, in PBC&PSC and HCC, respectively; the green solid lines denote the common edges between PBC&PSC and HCC; the red and black symbols indicate specific core members of the real GENs in PBC&PSC and HCC, respectively; and the green symbols are the common core members between PBC&PSC and HCC. In this figure, these core members are padded to complete the relevant signaling pathways for the convenience of analysis. The cytosolic proteins involved in the MAPK and the WNT pathways were highlighted by light-green and light-blue bands, respectively. Dysfunction of metabolism process, apoptosis, and autoimmune via DNA hypermethylation, and dysregulations of miR-21, miR-122, and miR-29a contribute to tumorigenesis from PBC&PSC to HCC. Dysregulation of miR-21, miR-122, and miR-29a contributes to tumor invasion and metastasis in HCC.

MAPK signaling pathways to inhibit DNA repair through the mediation of *HIST2H2BE* and to induce antiapoptosis through the mediation of *HSPB1*, which might finally facilitate tumor growth. Moreover, EGFR also represses TP53 to modulate ETS1 in the MAPK signaling pathway to promote dysfunction of metabolism and apoptosis through the mediation of *FRAT2* and *RPL30*. In addition, the dysregulation of *TUBA1C* caused by abnormal miR-122 silencing and the dysregulation of *ALDOB* caused by both hypermethylation and abnormal miR-21 silencing might facilitate tumor metastasis.

Based on two signaling pathways (Figures 5 and 2) involved in core GENs, we found five network biomarkers, UBC, amyloid precursor protein (APP), SHC-transforming protein 1 (SHC1), EGFR, and V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog 1 (ETS1), that play a central role in pathogenesis and hepatocarcinogenesis in the upper progression path of Figure 1(a). UBC facilitates a posttranslational modification where ubiquitin is attached to a substrate protein to control intracellular signaling and is involved in protein degradation, DNA repair, cell cycle regulation, kinase
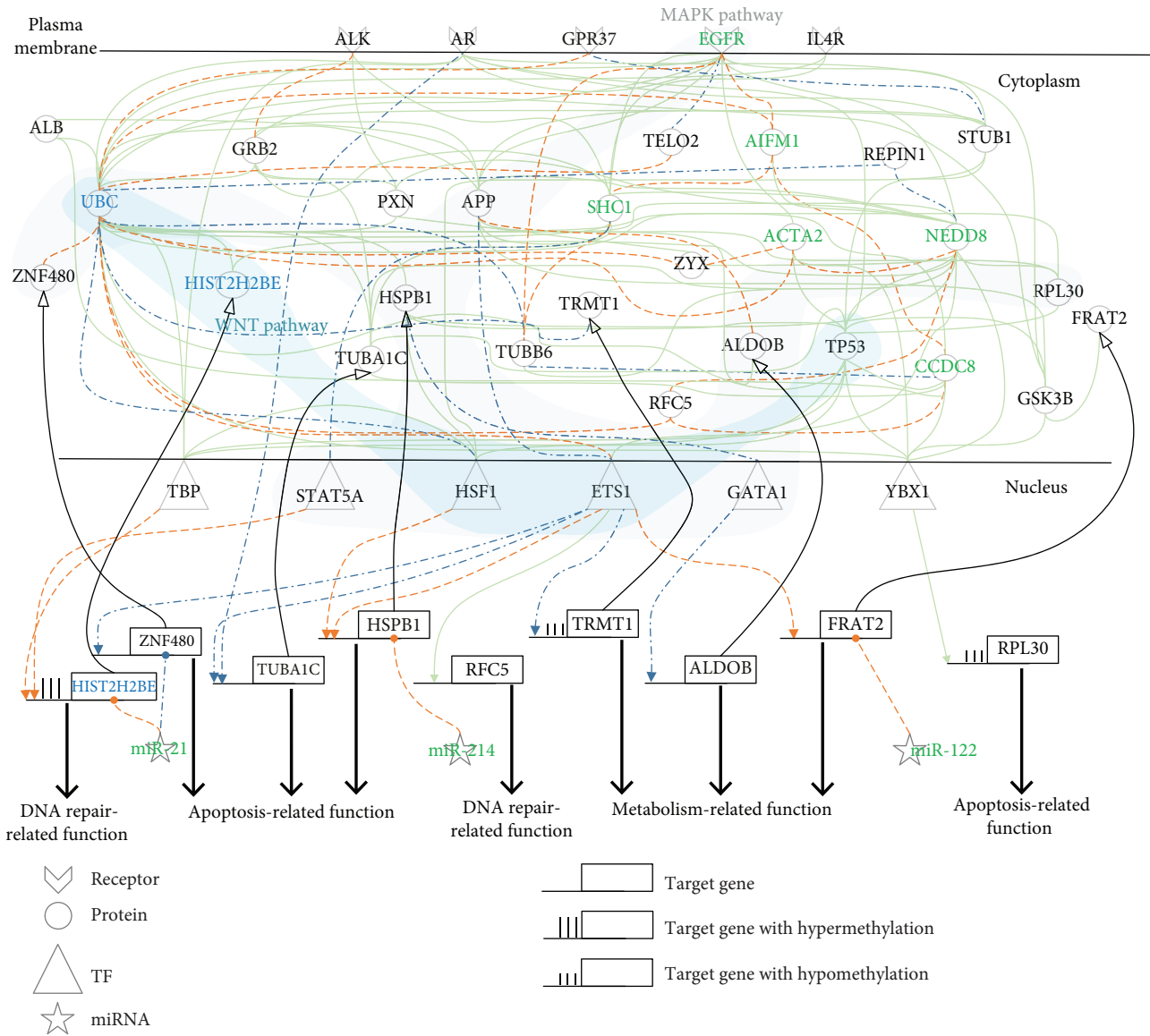
FIGURE 5: Signaling pathways for investigating cellular mechanisms of progression for pathogenesis from normal to NAFLD&NASH. The orange dotted lines and the blue dash-dotted lines represent the specific edges of the real GENs, including PPINs and GRNs, in normal and NAFLD&NASH, respectively; the green solid lines denote the common edges between normal and NAFLD&NASH; the blue and black symbols indicate specific core members of the real GENs in normal and NAFLD&NASH, respectively; and the green symbols are the common core members between normal and NAFLD&NASH. In this figure, these core members are padded to complete the relevant signaling pathways for the convenience of analysis. The cytosolic proteins involved in the MAPK and the WNT pathways were highlighted by light-gray and light-blue bands, respectively. Dysfunction of DNA repair via DNA hypermethylation and hepatic triglyceride excessive accumulation results in the pathogenesis of NAFLD&NASH. Dysregulation of miR-21 contributes to the dysfunction of apoptosis in NAFLD&NASH.

modification, endocytosis, and regulation of other cell signaling pathways. Therefore, UBC plays an important role in the appropriate regulation of the human system [16]. APP facilitates posttranslational modifications including glycosylation and phosphorylation. Its primary function is not yet known, although it has been implicated as a regulator of Alzheimer's disease, and may be associated with apoptosis, cell differentiation, and stress responses, and may play a critical role in intracellular signaling [17]. SHC1 has been reported to play an important role in the EGFR signaling pathway [18] and

is associated with carcinogenesis and metastasis [19]. It acts as a cell surface protein that binds to epidermal growth factor and a ligand to induce receptor dimerization and tyrosine autophosphorylation, which contributes to cell proliferation and differentiation [20]. Mutations in EGFR are associated with a number of cancers [21], but the precise role of EGFR in HCC is still unknown. ETS1 acts as a transcriptional activator or repressor of numerous genes and is involved in stem cell development, cell death, and tumorigenesis [22]. In addition, we also found nine genes that play a central role in the

upper progression path of Figure 1(a), i.e., two DNA repair-related genes (histone cluster 2 (H2be; *HIST2H2BE*) and replication factor C5 (*RFC5*)), four apoptosis-related genes (heat shock 27-kDa protein 1 (*HSPB1*), zinc finger protein 480 (*ZNF480*), tubulin alpha 1c (*TUBA1C*), and ribosomal protein L30 (*RPL30*)), and three metabolism-related genes (aldolase B (*ALDOB*), frequently rearranged in advanced T-cell lymphomas 2 (*FRAT2*), and tRNA methyltransferase 1 (*TRMT1*)).

2.3. Analysis of the Signaling Pathways Involved in the Core GENs to Investigate Underlying Cellular Mechanisms for PBC&PSC Pathogenesis and PBC&PSC-Associated Hepatocarcinogenesis. In the lower progression path of Figure 1(a), the identified core pathways comprise 40 core protein-coding genes, i.e., network biomarkers, involved in at least one core GEN of normal liver cells, PBC&PSC, and HCC. The projection values of the 40 network biomarkers $D(k)$ in (17) are also shown in Table S2. We projected these network biomarkers to KEGG and GO database mining to get the relevant biological functions and then obtained the core pathways for PBC&PSC pathogenesis (Figure 3) and PBC&PSC-associated hepatocarcinogenesis (Figure 4).

In the pathogenesis of PBC&PSC in Figure 3, we identified ten genes with expression differences between normal cells and PBC&PSC, i.e., *HIST2H2BE* (p value $\leq 5.62 \times 10^{-1}$), *RFC5* (p value $\leq 1.19 \times 10^{-1}$), *TIMP1* (p value $\leq 7.5 \times 10^{-2}$), *ZNF480* (p value $\leq 1.00 \times 10^{-3}$), *H3F3A* (p value $\leq 6.31 \times 10^{-1}$), *ALDOB* (p value $\leq 1.55 \times 10^{-1}$), *RPL30* (p value $\leq 2.62 \times 10^{-1}$), *FRAT2* (p value $\leq 7.16 \times 10^{-1}$), *TRMT1* (p value $\leq 1.00 \times 10^{-3}$), *IGF2* (p value $\leq 5.44 \times 10^{-1}$), and *MDC1* (p value $\leq 1.9 \times 10^{-2}$). Furthermore, four of those genes had basal level differences that might have been caused by DNA methylation of the corresponding genes. With DNA methylation profiles of normal and PBC&PSC for validation, we found one gene that had undergone hypermethylation change, i.e., *HIST2H2BE* (p value $\leq 1.2 \times 10^{-2}$), and three genes that had undergone hypomethylation, i.e., *RPL30* (p value $\leq 1.00 \times 10^{-3}$), *RPL23A* (p value $\leq 1.00 \times 10^{-3}$), and *RFC5* (p value $\leq 1.00 \times 10^{-3}$) in a comparison of normal cells with PBC&PSC cells.

In the hepatocarcinogenesis of PBC&PSC in Figure 4, we also identified ten genes having differences in expression between PBC&PSC and HCC, i.e., *HIST2H2BE* (p value $\leq 1.00 \times 10^{-3}$), *RFC5* (p value $\leq 1.00 \times 10^{-3}$), *TIMP1* (p value $\leq 1.00 \times 10^{-3}$), *ZNF480* (p value $\leq 1.00 \times 10^{-3}$), *H3F3A* (p value $\leq 1.00 \times 10^{-3}$), *RPL30* (p value $\leq 1.00 \times 10^{-3}$), *FRAT2* (p value $\leq 1.00 \times 10^{-3}$), *ALDOB* (p value $\leq 7.2 \times 10^{-2}$), *IGF2* (p value $\leq 1.07 \times 10^{-1}$), and *RPL23A* (p value $\leq 1.00 \times 10^{-3}$). Furthermore, six of those genes had basal level differences that might have been caused by DNA methylation on the corresponding genes. According to the DNA methylation profiles of PBC&PSC and HCC for validation, we found four genes with hypermethylation, i.e., *ALDOB* (p value $\leq 8.30 \times 10^{-1}$), *TIMP1* (p value $\leq 1.00 \times 10^{-3}$), *RPL30* (p value $\leq 1.00 \times 10^{-3}$), and *IGF2* (p value $\leq 1.00 \times 10^{-3}$), and two genes with hypomethylation, i.e.,

H3F3A (p value $\leq 1.00 \times 10^{-3}$) and *FRAT2* (p value $\leq 1.00 \times 10^{-3}$) in a comparison between PBC&PSC and HCC.

After extracting two signaling pathways between normal cells and PBC&PSC and between PBC&PSC and HCC from the core GENs (Figures S5, S7, and S8) as shown in Figures 3 and 4, respectively, we further investigated the cellular mechanisms underlying the progression of PBC&PSC pathogenesis in Figure 3 and PBC&PSC-associated hepatocarcinogenesis in Figure 4. In Figure 3, we observed that in normal liver cells, ryanodine receptor 2 (RYR2) is activated to facilitate UBC to modulate two TFs (ETS1 and BPTF) in the WNT and MAPK signaling pathways to induce DNA repair through the mediation of *RFC5*, metabolism through the mediation of *FRAT2*, and autoimmune function through the mediation of *MDC1* and *H3F3A*. Proper DNA repair, autoimmune function, and metabolism are maintained by silencing by miR-21, miR-122, and miR-29a. For example, miR-21 and miR-29a can silence *HIST2H2BE* and *H3F3A* in normal liver cells without significant liver damage, and miR-122 can silence *FRAT2* to avoid dysfunctions in metabolism. However, DNA hypomethylation of *RPL23A* contributes to defective autoimmune function causing liver damage, and DNA hypermethylation of *HIST2H2BE* contributes to dysfunction in DNA repair, which might cause normal liver cells to progress to PBC&PSC. In PBC&PSC, we observed that aberrant Ca^{2+} levels could activate RYR2 to facilitate SMAD5 to induce TP53 activated by UBC in the WNT and the MAPK signaling pathways, causing dysregulation of *H3F3A*. EGFR then facilitated APP to repress ETS1 in the MAPK signaling pathway to inhibit both DNA repair through the mediation of *RFC5* and immune function through the mediation of *MDC1*. In addition, ETS1 also induces both metabolism through the mediation of *TRMT1* and apoptosis through the mediation of *ZNF480* (see Figure 3). Furthermore, autoimmune defects through the mediation of *RPL23A* might still be accumulated by both abnormal miR-122 silencing and the dysregulation of H3F3A, and the dysfunction of DNA repair through the dysregulation of *HIST2H2BE* might also still be accumulated without miR-21 silencing, which might facilitate liver damage more seriously through the dysregulations of the MAPK and the WNT signaling pathways.

With a series of dysregulations and mutations, the cellular mechanism for hepatocarcinogenesis progression from PBC&PSC into HCC is shown in Figure 4. For example, *ZNF480* silenced by abnormal miR-21, *RPL23A* silenced by abnormal miR-122, and the dysregulation of *H3F3A* caused by both abnormal miR-29a silencing and hypomethylation might induce the accumulation of autoimmune defects and inhibit apoptosis to facilitate tumorigenesis. In HCC, liver damage and dysregulation due to mutations are still accumulated to facilitate ESR1 that could activate transferrin to repress UBC. Then, UBC could modulate STAT5A and CEBPA in the WNT and the MAPK signaling pathways to inhibit DNA repair through the dysregulation of *HIST2H2BE* and to induce the accumulation of autoimmune defects through the dysregulation of *IGF2*, which might ultimately facilitate tumor growth. In addition, UBC also repressed H3F3A and further suppressed BPTF in the

WNT and the MAPK signaling pathways to inhibit *H3F3A* by hypomethylation. Furthermore, EGFR repressed TP53 through the mediations of RFC5 and HUWE1 in the MAPK signaling pathway. Intriguingly, dysregulation of *TIMP1* caused by both abnormal miR-122 silencing and hypermethylation is involved in tumor proliferation. Additionally, the dysregulation of *ALDOB* and *IGF2* caused by hypermethylation and abnormal miRNA regulation through the mediation of miR-21 and miR-29a, respectively, might facilitate tumor metastasis.

Based on two signaling pathways (Figures 3 and 4) involved in core GENs, we found that five network biomarkers, including UBC, APP, EGFR, ETS1, and RYR2, play a central role in pathogenesis and hepatocarcinogenesis in the lower progression path of Figure 1(a). Intriguingly, UBC, EGFR, ETS1, and APP also played an important role in the upper progression path of Figure 1(a). RYR2 is a major mediator of the sarcoplasmic release of stored calcium ions and is involved in metabolism and cell survival [23]. Furthermore, we found seven genes in NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis that also played a central role in PBC&PSC pathogenesis and PBC&PSC-associated hepatocarcinogenesis, i.e., two DNA repair-related genes (*HIST2H2BE* and *RFC5*), two apoptosis-related genes (*ZNF480* and *RPL30*), and three metabolism-related genes (*ALDOB*, *FRAT2*, and *TRMT1*). Moreover, we also identified five genes that play a central role in the lower progression path of Figure 1(a), i.e., one apoptosis-related gene (TIMP metallopeptidase inhibitor 1 (*TIMP1*)), three immune-related genes (ribosomal protein L23a (*RPL23A*), H3 histone family 3A (*H3F3A*), and insulin-like growth factor 2 (*IGF2*)), and one DNA repair-related gene (mediator of DNA-damage checkpoint 1 (*MDC1*)).

2.4. Cellular Mechanisms of the Pathogenesis and Hepatocarcinogenesis from Normal Liver Cells to HCC through NAFLD&NASH or PBC&PSC Based on Core Signaling Pathways. Based on the system biology approach shown in Materials and Methods, we constructed the core GENs for the various liver diseases and HCC. We then further investigated the cellular mechanisms of progression from normal liver cells to HCC through the liver diseases by comparing core GENs of the upper or the lower progression paths to extract the core signaling pathways involved in the core GENs for each progression stage in Figure 1(a).

After the investigations of Figures 5 and 2 for NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis as shown in Figure 1(a), it can be seen that the general progression mechanism for the upper progression path is through the WNT and the MAPK signaling pathways, inducing DNA repair, apoptosis, and metabolism. Further, the liver homeostatic balance is maintained by miR-21, which silences the *HIST2H2BE* gene thereby promoting proper cellular functions such as the development, senescence, and proliferation of cells and the development and function of erythroblasts in normal liver cells; miR-214 silences the *HSPB1* gene to avoid unnecessary antiapoptosis,

and miR-122 silences the *FRAT2* gene to avoid dysfunction of metabolism causing liver damage. The accumulation of liver damage is caused by steatosis, and *HIST2H2BE* is regulated by hypermethylation through the MAPK signaling pathway, which facilitates abnormal DNA repair and the progression of normal liver cells to NAFLD&NASH, as shown in Figure 5. Furthermore, *ZNF480* can be inhibited and *HSPB1* can be activated to cause antiapoptosis, overcoming liver damage. However, the dysfunction of antiapoptosis caused oncogene proliferation and survival through the dysregulation of *ZNF480* by abnormal miR-21 silencing and through the dysregulation of *HSPB1* by hypermethylation, facilitating the dysregulation of the MAPK signaling pathway. Additionally, the dysfunction of metabolism caused by hypermethylation of *ALDOB* allows NAFLD&NASH to develop into HCC. Interestingly, abnormal miR-21 can still be accumulated to induce the dysregulation of *ALDOB*, and abnormal miR-122 can also be accumulated to induce the dysregulation of *TUBA1C*, which might facilitate tumor metastasis and invasion in HCC, as shown in Figure 2. To integrate the description given in Figures 5 and 2 of the NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis as shown in Figure 1(a) and to consider the impact of both miRNA regulation and DNA methylation on the mechanisms of progression in the upper progression path, we further investigated the role of miRNA regulation and DNA methylation in the mechanisms of progression in the upper progression path. Our objective was to discover the potential diagnostic biomarkers or drug targets that are summarized in Figure 6, with a more detailed description in Discussion.

We then observed that the general mechanism for the lower progression path is also through the WNT and the MAPK signaling pathways via different epigenetic regulations compared with the upper progression path that induces DNA repair, metabolism, and autoimmune function based on the investigations of Figures 3 and 4 for the PBC&PSC pathogenesis and PBC&PSC-associated hepatocarcinogenesis, as shown in Figure 1(a). To further maintain liver homeostatic balance by miR-29a silencing of *H3F3A* to promote proper gene regulation such as the regulation of *RPL23* to cause autoimmune function in normal liver cells, miR-21 can silence *HIST2H2BE* to avoid unnecessary DNA repair, and miR-122 can silence *FRAT2* to avoid dysfunction of metabolism causing liver damage. In contrast to normal liver cells that progress to NAFLD&NASH, *RPL23A* regulated by hypomethylation caused defects in the autoimmune response to promote more severe liver damage through the WNT and the MAPK signaling pathways, and *HIST2H2BE* was also regulated by hypermethylation to facilitate abnormal DNA repair through the MAPK signaling pathway in the progression of normal liver cells to PBC&PSC, as shown in Figure 3. Therefore, we realized that NAFLD&NASH, as a reversible liver disease, might allow repairing liver damage by an autoimmune response, but the dysfunction of the autoimmune response might cause more severe liver damage in PBC&PSC. Most cellular mechanisms share some similarity to NAFLD&NASH progressing to HCC other than the dysfunction of the autoimmune response caused by the
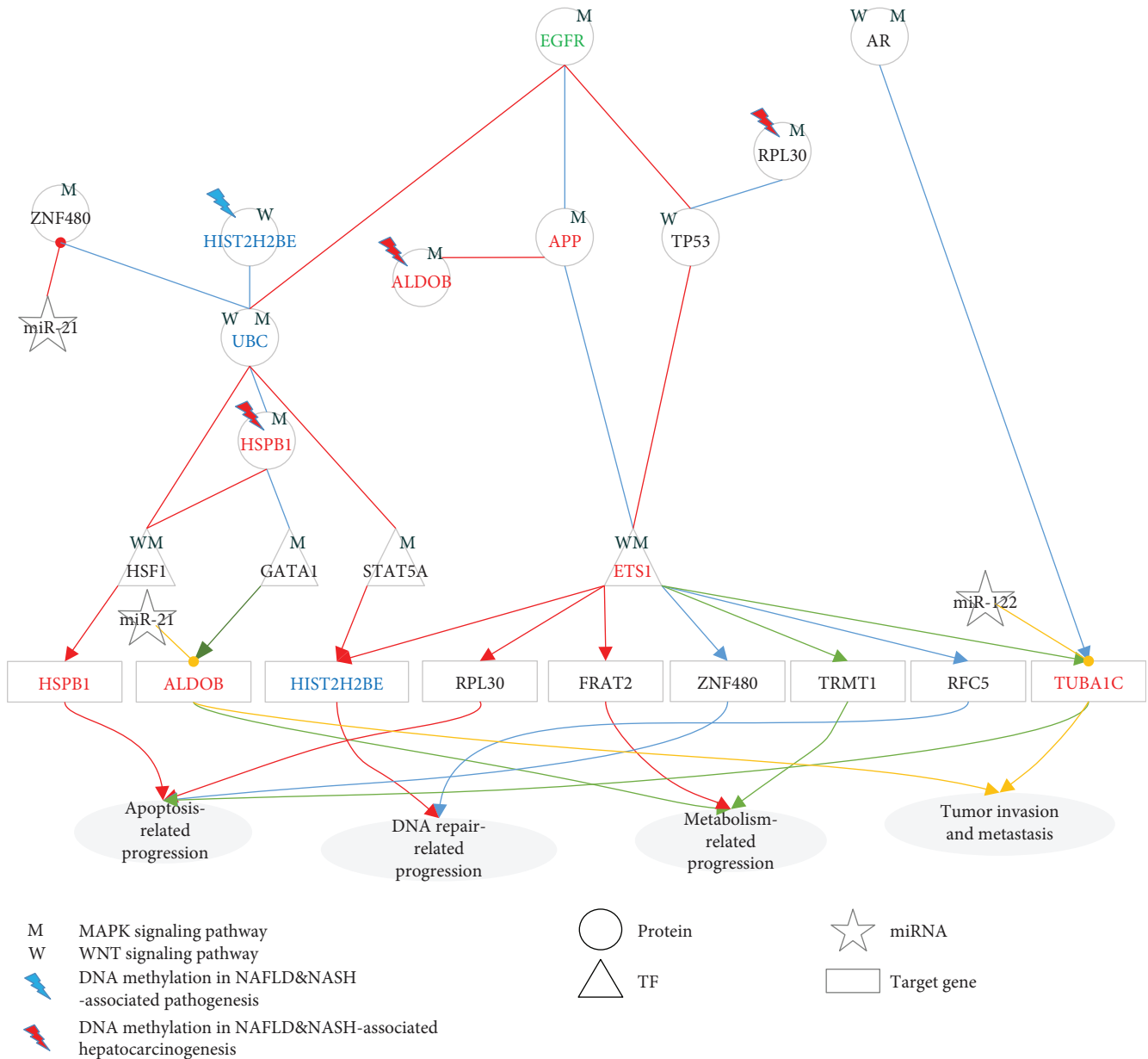
FIGURE 6: Progression mechanisms for the pathogenesis of NAFLD&NASH and the hepatocarcinogenesis through NAFLD&NASH. The blue, red, green, and yellow lines represent the interactions (or regulations) in NAFLD&NASH pathogenesis, NAFLD&NASH hepatocarcinogenesis, the hepatocarcinogenesis of normal liver through NAFLD&NASH, and the aggressive tumor progression, respectively, based on Figures 5 and 2. The epigenetic modification of HIST2H2BE could facilitate the dysfunction of metabolism-related progression through the dysregulations of the WNT and the MAPK signaling pathways resulting in NAFLD&NASH. The accumulated epigenetic modifications and dysregulations of miR-21 could facilitate the dysfunctions of metabolism-related, apoptosis-related, and DNA repair-related progression through dysregulations of the WNT and the MAPK signaling pathways resulting in HCC. Dysregulation of miR-21 and miR-122 could contribute to tumor invasion and metastasis to facilitate further aggressive tumor progression.

dysregulation of both *H3F3A* and *RPL23A* through abnormal miR-29a and miR-122 silencing, respectively, to facilitate the progression of PBC&PSC into HCC. Furthermore, abnormal miR-122 can still be accumulated to cause the dysfunction of apoptosis through the dysregulation of *TIMP1*, and abnormal miR-29a can still be accumulated to cause the dysfunction of autoimmune through the dysregulation of *IGF2*, and abnormal miR-21 can still be accumulated to cause the dysfunction of metabolism through the dysregulation of *ALDOB*, which

might facilitate tumor metastasis and invasion in HCC, as shown in Figure 4. To integrate the descriptions of Figures 3 and 4 of PBC&PSC pathogenesis and PBC&PSC-associated hepatocarcinogenesis as shown in Figure 1(a) and to consider the impact of both miRNA regulation and DNA methylation on the mechanisms of progression in the lower progression path, we further investigated the role of miRNA regulation and DNA methylation in the lower progression path. This provided insight into the development of the potential
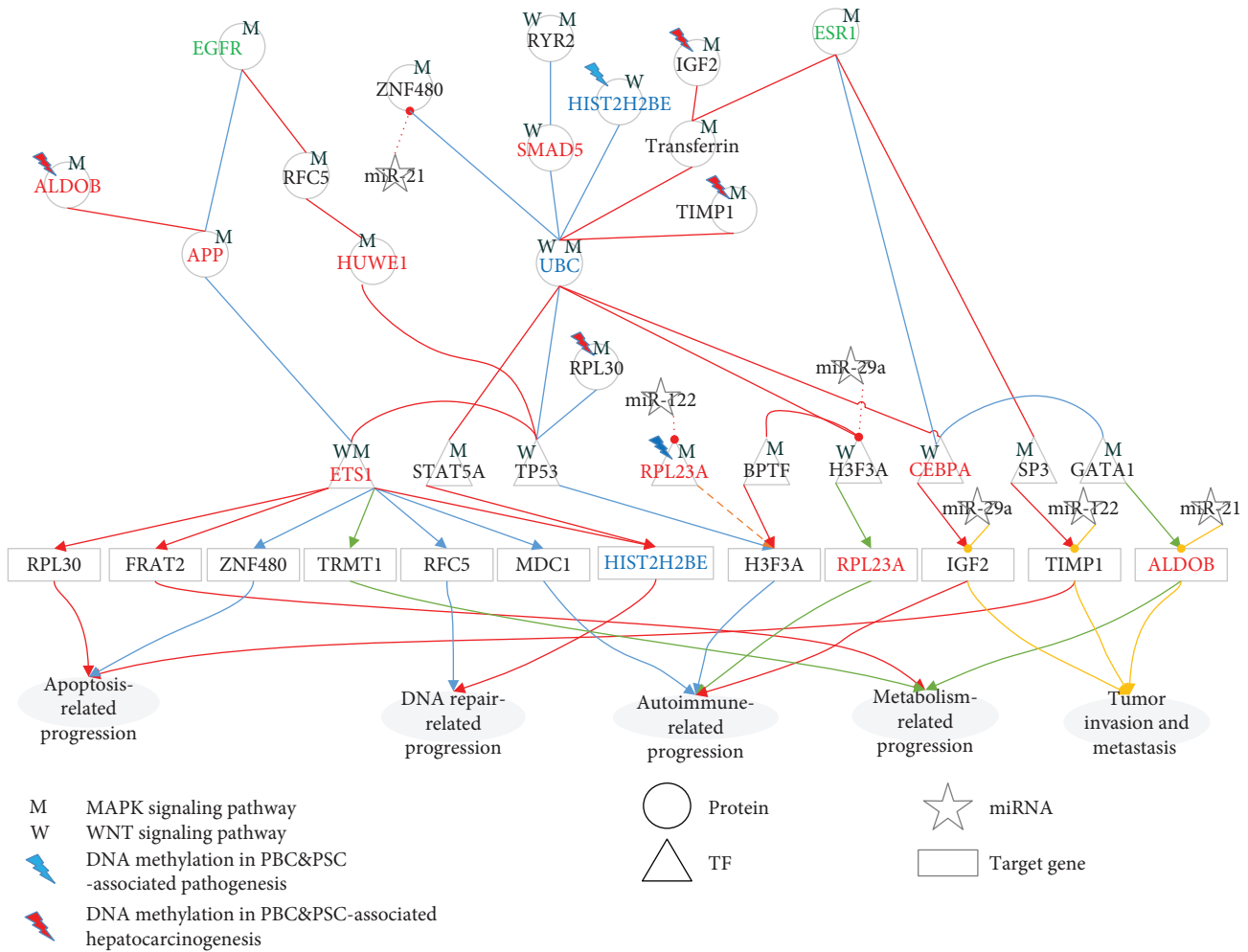
FIGURE 7: Progression mechanisms for the pathogenesis of PBC&PSC and the hepatocarcinogenesis through PBC&PSC. The blue, red, green, and yellow lines represent the interactions (or regulations) in PBC&PSC pathogenesis, PBC&PSC hepatocarcinogenesis, the hepatocarcinogenesis of normal liver through PBC&PSC, and the aggressive tumor progression, respectively, based on Figures 3 and 4. The epigenetic modifications of HIST2H2BE and RPL23A could facilitate the dysfunction of autoimmune-related progression through dysregulations of the WNT and the MAPK signaling pathways resulting in PBC&PSC. The accumulated epigenetic modifications and the dysregulation of miR-29a, miR-122, and miR-21 could facilitate the dysfunction of metabolism-related, apoptosis-related, autoimmune-related and DNA repair-related progression through the dysregulation of the WNT and the MAPK signaling pathways resulting in HCC. Dysregulation of miR-21, miR-122, and miR-29a could contribute to tumor invasion and metastasis to facilitate further aggressive tumor progression.

diagnostic biomarkers or drug targets summarized in Figure 7, with a more detailed description in Discussion.

3. Discussion

Recently, there has been an intensive investigation into epigenetic regulation and the role it plays in the progression of liver disease to HCC [24]. Epigenetic dysregulation promotes the pathogenesis of hepatocytes and facilitates the development of HCC [25]. In principle, epigenetic regulations include DNA methylation, histone modification, and miRNA transcription [26]. Histone modifications consist of acetylation, methylation, phosphorylation, and ubiquitination, and although they cannot be represented by an interactive model, they still caused significant variations in the genetic regulatory parameters in our method.

Hence, we considered the impact of both DNA methylation and miRNA regulation in GENs on pathogenesis and hepatocarcinogenesis from normal liver cells to HCC through different liver diseases. Because aberrant DNA methylation causes abnormal gene expression, DNA methylation is also regarded as a key process in the pathogenesis and hepatocarcinogenesis from normal liver cells to HCC through liver diseases. Recent studies have indicated that hypermethylation induced gene silencing, inhibit gene expression, and affect genomic stability [27]. miRNAs are small noncoding RNAs that modulate the transcription and translation of target genes and regulate different biological functions in the human liver. In addition, miRNAs are involved in posttranscriptional regulations that are associated with the proliferation, differentiation, and death of cells and with carcinogenesis [28]. Using the proposed system biology
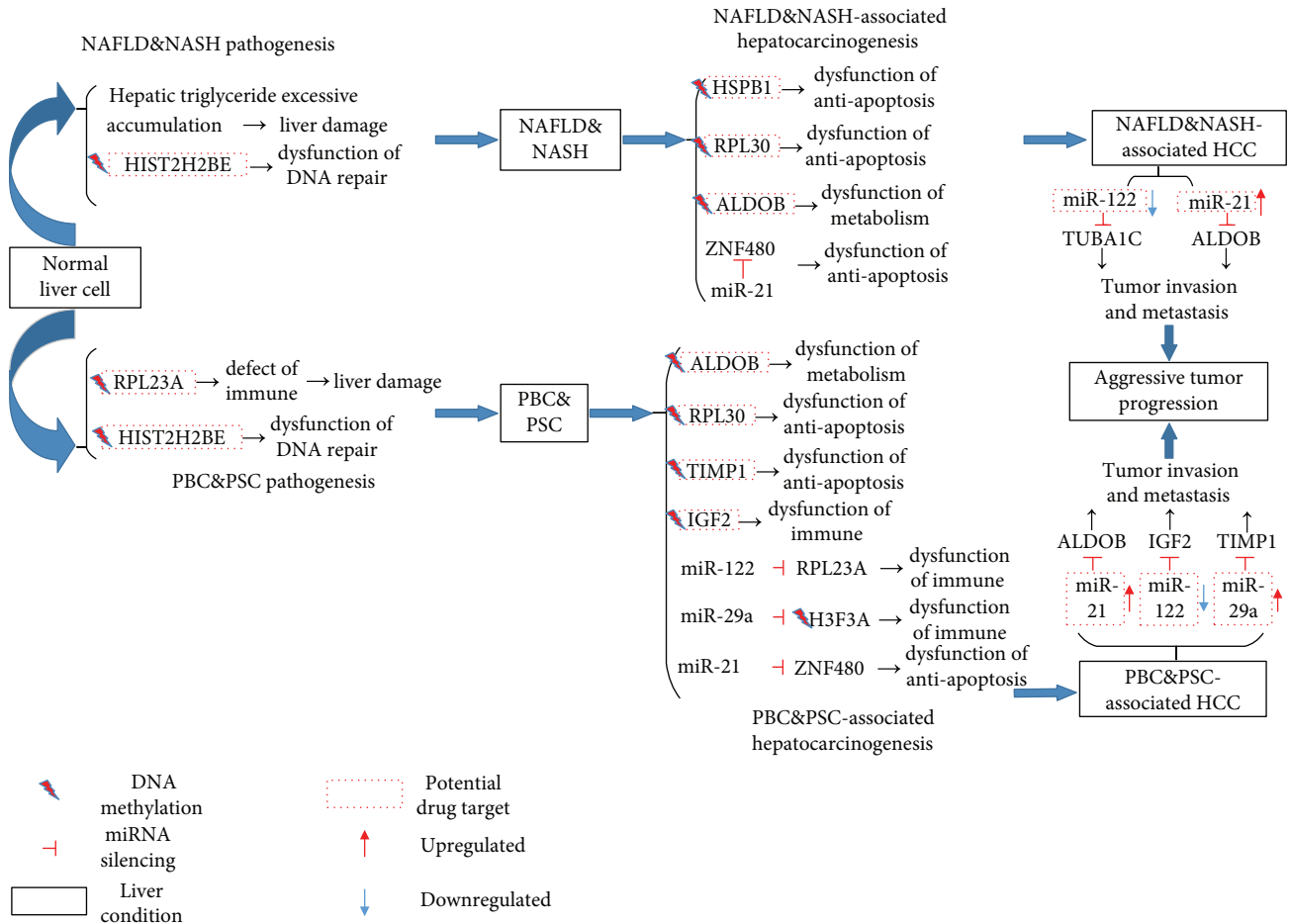
FIGURE 8: Schematic structure of the pathogenic and hepatocarcinogenic mechanisms of NAFLD&NASH and PBC&PSC.

approach, we identified core miRNAs and further extracted the core GENs and the involved signaling pathways that could realize cellular mechanisms for the progression of normal liver cells to HCC through liver diseases and could also provide drug targets for novel therapeutic schemes through alteration of methylation levels and miRNA expression.

Based on the proposed method, we identified not only the core miRNAs but also the core proteins and core target genes in signaling pathways by comparing different core GENs of the upper or lower progression paths in Figure 1(a). Normal liver cells provide a standard system for the toleration of intrinsic perturbations such as methylation accumulation and miRNA dysregulation to maintain liver homeostasis, but genetic mutations and epigenetic dysregulations are still accumulated and exceed the tolerable allowance of normal liver cells, causing liver diseases and eventually HCC. Therefore, the network biomarkers found to interact with other partners in the core signaling pathways could contribute to the genetic and epigenetic dysfunctions of the corresponding biological functions through genetic mutation, DNA methylation accumulation, and miRNA dysregulation.

According to our results, we shall further discuss the impact of miRNA regulation and DNA methylation on epigenetic mechanisms of pathogenesis and hepatocarcinogenesis from normal liver cells to HCC through NAFLD&NASH or PBC&PSC, as shown in Figures 6 and 7, respectively, as summarized in Figure 8.

3.1. Dysregulation of DNA Methylation and miRNA Regulation Contributes to Pathogenesis and Hepatocarcinogenesis from Normal Liver Cells to HCC through NAFLD&NASH or PBC&PSC

3.1.1. Dysregulation of DNA Methylation and miR-21 Contributes to NAFLD&NASH Pathogenesis and NAFLD& NASH-Associated Hepatocarcinogenesis. Recently, several studies have indicated that NAFLD&NASH pathogenesis is associated with obesity and insulin resistance, facilitating liver damage through steatosis caused by excessive accumulation of hepatic triglycerides and more severe liver damage through subsequently induced oxidative stress [24]. Our results support the theory that oxidative stress can activate the MAPK signaling pathway to facilitate a series of functions such as DNA repair, apoptosis, and metabolism. In addition, we also determined that aberrant methylation can affect HIST2H2BE, facilitating the dysfunction of metabolism to progress from normal liver cells to NAFLD&NASH, as shown in Figure 6. Dysfunction of metabolism contributes to the accumulation of genetic mutations, which might cause more severe liver damage.

It has been suggested that HIST2H2BE plays a central role in the progression of normal liver cell to NAFLD&NASH, so we suggested HIST2H2BE as a potential drug target that might overcome liver damage caused by excessive hepatic triglyceride accumulation.

After a brief discussion on NAFLD&NASH pathogenesis, we further discussed NAFLD&NASH hepatocarcinogenesis by abnormal miRNA regulation and aberrant methylation. In this study, we established that the change in the miRNA expression profile of miR-21 (p value $\leq 1.00 \times 10^{-3}$) results in a significant change in the gene expression profile of *ZNF480* (p value $\leq 1.00 \times 10^{-3}$) to progress from NAFLD&NASH to HCC, as shown in Figure 6. Recently, it has been reported that miR-21 is the first central oncomir to be associated with malignant cell proliferation, invasion, and metastasis in the carcinogenesis of a variety of cancers [29]. For instance, miR-21 can silence phosphatase and tensin homolog (PTEN), which is a tumor suppressor observed in HCC [29]. It not only reflects liver damage but also regulates other biological processes such as triglyceride and cholesterol metabolism by silencing target genes [30, 31] and the upregulation of ZNF480 resulting in tumor cell death [32]. Furthermore, our finding also demonstrated that epigenetic regulations, including DNA methylation of *HSPB1*, *ALDOB*, and *RPL30*, led to changes in the DNA methylation profiles of *HSPB1* (p value $\leq 1.00 \times 10^{-3}$), *ALDOB* (p value ≤ 0.85), and *RPL30* (p value $\leq 1.00 \times 10^{-3}$), which resulted in significant changes in the gene expression profiles of *HSPB1* (p value $\leq 1.00 \times 10^{-3}$), *ALDOB* (p value $\leq 1.5 \times 10^{-2}$), and *RPL30* (p value $\leq 1.00 \times 10^{-3}$) in the progression from NAFLD&NASH to HCC, as shown in Figure 6. In addition, it has been proposed that the dysregulations of HSPB1, ALDOB, and RPL30 are sufficient to trigger tumor progression [33, 34]. Therefore, we suggested that the epigenetic dysregulation of miR-21 and DNA methylation causes pathogenesis and hepatocarcinogenesis from the upper progression path to HCC, as shown in Figure 8 with more detailed discussion in the sequel.

3.1.2. Dysregulation of DNA Methylation, miR-21, miR-122, and miR-29a Contributes to PBC&PSC Pathogenesis and PBC&PSC-Associated Hepatocarcinogenesis. Based on our results from PBC&PSC pathogenesis, we identified that the DNA methylation profile changes for *RPL23A* and *HIST2H2BE* result in dysfunctions in the autoimmune and DNA repair response to progression from normal liver cells to PBC&PSC and give rise to the dysregulation of RYR2 through the WNT and the MAPK signaling pathways, to facilitate defects in autoimmunity as shown in Figure 7. Intriguingly, our results showed that PBC&PSC are associated with autoimmune dysfunction [35, 36]. It has been suggested that RPL23A and HIST2H2BE play a central role in the progression from normal liver cells to PBC&PSC, so we propose that RPL23A and HIST2H2BE are potential drug targets that might overcome the defects in autoimmunity and the subsequent dysfunction in metabolism.

After a brief discussion on PBC&PSC pathogenesis, we further discussed PBC&PSC hepatocarcinogenesis by abnormal miRNA regulation and aberrant methylation. We identified that the expression changes in miR-21 (p value $\leq$ 1.00 × 10^{-3}$), miR-122 ($p$ value $\leq 1.4 \times 10^{-2}$), and miR-29a (p value $\leq 1.00 \times 10^{-3}$) contribute to the significant gene expression changes in *ALDOB* (p value $\leq 0.72 \times 10^{-1}$), *RPL23A* (p value $\leq 1.00 \times 10^{-3}$), and *H3F3A* (p value $\leq 1.00 \times 10^{-3}$) between PBC&PSC and HCC, as shown in Figure 7. Recently, it has been reported that miR-122 is the most common miRNA; it is estimated to make up 70% of the total miRNA complement. It serves as an essential regulator and is involved in development, differentiation, and homeostasis, as well as the metabolism of glucose and lipids in the liver [1, 24]. Moreover, the loss of miR-122 has been associated with migration, invasion, and tumorigenesis, which could cause liver diseases and eventually HCC [1]. Consequently, the restoration of miR-122 represents antitumorigenesis functions that slow down the progression of both liver disease and HCC, which has been investigated and demonstrated in animal models [37]. Furthermore, a recent study has indicated that miR-29a might repress tumors formed by promoted apoptosis [38] and miR-29a is considered a potential target of therapy in autoimmune connective tissue disease [39]. Our findings demonstrated that epigenetic regulations, including DNA methylation of *RPL30*, *ALDOB*, *IGF2*, and *TIMP1*, led to changes in the DNA methylation profiles of *RPL30* (p value $\leq 1.00 \times 10^{-3}$), *ALDOB* (p value ≤ 0.83), *IGF2* (p value $\leq 1.00 \times 10^{-3}$), and *TIMP1* (p value $\leq 1.00 \times 10^{-3}$), which contributed to the significant changes in the gene expression profiles of *RPL30* (p value $\leq 1.00 \times 10^{-3}$), *ALDOB* (p value $\leq 7.2 \times 10^{-2}$), *IGF2* (p value $\leq 1.1 \times 10^{-2}$), and *TIMP1* (p value $\leq 1.00 \times 10^{-3}$) to progress from PBC&PSC to HCC, as shown in Figure 7. In addition, it has been proposed that the dysregulation of DNA methylation of *RPL30*, *ALDOB*, *IGF2*, and *TIMP1* is sufficient to trigger tumor progression [33, 34, 40]. Abnormal miR-29a and hypomethylation contribute to the dysregulation of H3F3A, which could lead to tumorigenesis [41]. In summary, we suggested that the epigenetic dysregulation of miR-21, miR-122, miR-29a, and DNA methylation could cause pathogenesis and hepatocarcinogenesis in the lower progression path to HCC, as shown in Figure 8 with more detailed discussion in the sequel.

3.1.3. Dysregulation of miR-21, miR-122, and miR-29a Contributes to Further Liver Damage in Patients with NAFLD&NASH or PBC&PSC. In this study, we confirmed that abnormal miR-21 silences ALDOB and abnormal miR-122 silences TUBA1C in the mechanisms of HCC progression, as shown in Figure 6. It has been proposed that TUBA1C is associated with tumor cell death and cell proliferation [42], and aberrant expression of ALDOB contributes to tumor metastasis, which is involved in the initiation and development of HCC [33, 43]. We also confirmed that abnormal miR-29a silences IGF2 and abnormal miR-122 silences TIMP1 in the mechanism of HCC progression, as shown in Figure 7. The dysregulation of IGF2 and TIMP1 was significantly realized in HCC. It has been reported that the dysregulation of IGF2 and TIMP1 results in tumor invasion and metastasis [40]. Therefore, we suggested that the dysregulation of miR-21, miR-29a, and miR-122 contributes

Table 1: Network biomarkers for preventing the hepatocarcinogenesis of NAFLD&NASH and PBC&PSC.

	NAFLD&NASH to HCC	PBC&PSC to HCC
The highly activated network biomarkers for potential inhibition strategy of drug design	ACTA2, AIFM1, APP, EGFR, GRB2, NEDD8, SHC1, TUBA1C, TUBB6, IL4R, GPR37, FRAT2, HIST2H2BE, HSPB1, ZNF480, TIMP1, ALDOB, ZYX, YBX1, TP53, STAT5A, AR, TELO2, PXN, TBP, REPIN1, HSF1, ETS1, STUB1, RFC5, RPL30, TRMT1, GSK3B, UBC, ALB, and MIR21	ACTB, ADRB2, APP, CSNK2A1, EGFR, HUWE1, LMNB1, PCK1, PPP2CA, SMAD5, TSC22D1, ESR1, PGR, RYR2, FRAT2, HIST2H2BE, ZNF480, H3F3A, TIMP1, ALDOB, IGF2, AR, TP53, SRF, NRF1, STAT5A, RPL23A, SP1, SP3, CEBPA, BPTF, ETS1, MDC1, RFC5, RPL30, TRMT1, UBC, GSK3B, TF, and MIR21
The repressed network biomarkers for potential activation strategy of drug design	ALK and GATA1	CEBPE and GATA1

to the dysregulation of TUBA1C, ALDOB, IGF2, and TIMP1, which might lead to tumor invasion and metastasis in the development of HCC, facilitating further aggressive tumor progression, as shown in Figure 8 with more detailed discussion in the sequel.

In brief, an increasing number of studies focus on epigenetic regulations that provide novel therapies from the perspective of epigenetic alterations. It has been proposed that the impact of miRNA regulation and methylation on different liver diseases and the mechanisms of progression from liver diseases to HCC are caused by epigenetic multiregulation. For instance, our results demonstrated that miR-122 and miR-21 might serve as epigenetic biomarkers in the upper progression path in Figure 8 and further provide potential drug targets for novel therapies through these two miRNAs and methylation of the target genes. Although the progression mechanism of the lower progression path is similar to that of the upper progression path in Figure 8, detailed progression mechanisms are differential. Hence, we identified miR-122, miR-21, and miR-29a as potential prognostic biomarkers for potential drug targets in the lower progression path in Figure 8. Interestingly, miR-122 and miR-21 may provide novel potential therapeutic targets for slowing down further liver damage in patients with NAFLD&NASH in Figure 8 and for predicting and treating HCC. PBC&PSC pathogenesis relates to autoimmune dysfunction and is different from NAFLD&NASH pathogenesis, which relates to metabolism dysfunction. Therefore, we suggested that miR-29a provides a novel potential therapeutic target for the restoration of aberrant immune response to treat PBC&PSC-developed HCC, as well as miR-122 and miR-21 in Figure 8. The results are supported by recent studies associated with miRNA regulation in the immune system [44].

Finally, the identified network biomarkers for preventing the hepatocarcinogenesis of NAFLD&NASH and PBC&PSC are shown in Table 1 (all potential drugs ranked in Table S3A).

4. Materials and Methods

The proposed methodology to identify the core signaling pathways and the network biomarkers of NAFLD&NASH pathogenesis, NAFLD&NASH-associated hepatocarcinogenesis, PBC&PSC pathogenesis, and PBC&PSC-associated

hepatocarcinogenesis for preventing the progression of NAFLD&NASH or PBC&PSC is summarized in the flowchart in Figure 1(b).

4.1. Big Data Mining and Preprocessing of mRNA Expression Data and Its Corresponding DNA Methylation Profiles for Liver Diseases and HCC. In order to identify the real GENs of liver cells from the four kinds of patients including normal, NAFLD&NASH, PBC&PSC, and HCC patients, the simultaneously measured genome-wide mRNA expression data and DNA methylation profiles in each sample of the patients with one of the liver conditions were needed. Also, the candidate GENs obtained from biological or computational experiments in human cells were required. In this study, we used the microarray data with its corresponding DNA methylation profiles of 134 samples in patients with normal (62 samples), NAFLD&NASH (47 samples), and PBC&PSC (25 samples) conditions from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (GSE61260) provided by Horvath et al. [45] and used the NGS data and its corresponding DNA methylation profiles of 369 samples in patients with HCC, obtained from the Cancer Genome Atlas (TCGA) (https://genome-cancer.ucsc.edu/). NGS data, microarray data, and DNA methylation profiles were measured using the Illumina HiSeq platform, the HuGen1.1ST platform, and the Illum450K platform, respectively.

Additionally, we considered that the candidate GEN in human cells includes the candidate PPIs and gene and miRNA regulations. According to the candidate PPIs from the Biological General Repository for Interaction Datasets (BioGRID) database [46], the gene regulations from the Integrated Transcription Factor Platform (ITFP) database, the Human Transcriptional Regulation Interactions (HTRI) database, and the TRANScription FACtor database (TRANSFAC) [47], and the miRNA regulations from TargetScanHuman database [48], we obtained 168,970 PPI pairs, 152,828 gene regulations, and 247,743 miRNA repressions in the candidate GEN of human cells. Because the candidate GEN of human cells contains all possible associations obtained from experimental and computational results, which contain many false positives, we need to construct the stochastic interactive/regulatory models of the candidate GEN in human cells to characterize the cellular mechanisms in the network. The

real GEN of each liver condition can then be obtained by pruning away false positives in the candidate GEN through a system identification method and a system order detection scheme in the stochastic models using the genome-wide mRNA expression data and its corresponding DNA methylation profiles.

To integrate the big data, including genome-wide mRNA expression data, its corresponding DNA methylation profiles, and the candidate interactions/regulations, from several databases, we used Matlab's text-file and string manipulation tools in text mining to unify the gene name based on the gene symbols in the NCBI Entrez Gene database.

4.2. Constructing the Stochastic Models of the GEN in Human Cells. We constructed the stochastic models of the candidate GEN in human cells to characterize the molecular mechanisms, including PPIs, transcriptional regulations, miRNA repressions, DNA methylation, and stochastic noises due to the modeling residue and the fluctuation of genes. The molecular mechanisms of the ith protein in the PPIN of the GEN in the cells of the nth patient sample can be described by the stochastic protein interactive model as follows:

$$y_i[n] = \sum_{j=1, j \neq i}^{N_i} a_{ij} y_i[n] y_j[n] + h_i + v_i[n],$$

$$\text{for } i = 1, \ldots, N, \ n = 1, \ldots, L, \ h_i \geq 0, \tag{1}$$

where $y_i[n]$, $y_j[n]$, and $v_i[n]$ represent the mRNA expression levels of the ith protein and its jth interactive protein and the stochastic noise of the ith protein due to model uncertainty or the fluctuation of expressions in patient n, respectively; L, N, and N_i indicate the number of patients, proteins, and proteins interacting with the ith protein in the candidate PPIN, respectively; and a_{ij} and h_i denote the interaction ability between the ith protein and its jth interactive protein and the basal level of the ith mRNA expression, respectively. The term $y_i[n] y_j[n]$, the product of the concentrations of two interactive proteins, can represent the probability of molecular interaction between two proteins of the complex protein interaction. The probability of molecular interaction between two proteins has been also modeled as the probability of molecule collision by a nonlinear multiplication scheme [49, 50]. The molecular mechanisms in (1) of the ith protein in the PPIN of human cells in the nth patient sample include PPIs ($\sum_{j=1, j \neq i}^{N_i} a_{ij} y_i[n] y_j[n]$), the stochastic noise ($v_i[n]$), and the basal levels of proteins (h_i).

The molecular mechanisms of the ith gene in the GRN of the GEN in the cells of the nth patient sample can be described by the stochastic gene regulatory model as follows:

$$x_i[n] = \sum_{j=1, j \neq i}^{K_i} b_{ij} y_j[n] M_i[n] - \sum_{v=1}^{V_i} c_{iv} x_i[n] x_v^{miR}[n] M_i[n]$$
$$+ M_i[n]\kappa_i + \omega_i[n], \tag{2}$$
$$\text{for } i = 1, \ldots, N, \ n = 1, \ldots, L, \ \kappa_i, c_{iv} \geq 0,$$

where $x_i[n]$, $y_j[n]$, $x_v^{miR}[n]$, and $\omega_i[n]$ indicate the expression levels of the ith target gene, its jth regulatory TF, its vth regulatory miRNA, and its stochastic noise, which results from model uncertainty or the fluctuation of expressions in GRN, in patient n, respectively; K_i and V_i represent the number of TFs and miRNAs binding to the ith gene based on candidate GEN in human cells, respectively; κ_i denotes the basal level of the ith gene; b_{ij} and c_{iv} indicate the transcriptional regulatory ability from the jth TF to the ith gene and the posttranscriptional regulatory ability of the vth miRNA to inhibit the ith gene ($-c_{iv} \leq 0$), respectively; and $M_i[n]$ represents methylation regulation of the ith gene, affecting the binding affinities of miRNAs, RNA polymerase, and TFs to the ith gene [51]. The effect of methylation on binding affinities of miRNAs, RNA polymerase, and TFs to the ith gene is expressed by the terms $c_{iv} x_i[n] x_v^{miR}[n] M_i[n]$, $M_i[n]\kappa_i$, and $b_{ij} y_j[n] M_i[n]$, respectively. The methylation regulation of the ith target gene $M_i[n]$ is defined as follows:

$$M_i[n] =: \frac{1}{1 + (m_i[n]/0.5)^2}, \tag{3}$$

where $m_i[n]$ denotes the DNA methylation profile of the ith gene in the nth patient sample. Because the numerical range of DNA methylation profiles $m_i[n]$ is between 0 and 1, the effect of DNA methylation on gene and miRNA regulations and basal levels is from 1 to 0.2. It means that if the DNA methylation profile of the ith gene increases, the binding affinities of miRNAs, RNA polymerase, and TFs to the ith gene decrease. The definition of methylation regulation in (3) avoids shutting down the bindings of miRNAs, RNA polymerase, and TFs to the ith gene by $m_i[n] = 1$. The basal level (κ_i) change between two liver conditions is also used to infer the effect of DNA methylation on the transcriptional regulation of the ith gene. The molecular mechanisms in (2) of the ith gene in the GRN of human cells in the nth patient sample include transcriptional regulations ($\sum_{j=1, j \neq i}^{K_i} b_{ij} y_j[n] M_i[n]$), miRNA repressions ($-\sum_{v=1}^{V_i} c_{iv} x_i[n] \cdot x_v^{miR}[n] M_i[n]$), the stochastic noise ($\omega_i[n]$), and the effects of RNA polymerase and DNA methylation ($M_i[n]\kappa_i$) on genes.

4.3. Identification of the Real GEN by Applying System Identification Method and System Order Detection Scheme to Prune False Positives in Candidate GEN Using Genome-Wide Expression Data and Its Corresponding DNA Methylation Profiles. After constructing the stochastic protein interactive model in (1) of PPIN and the stochastic gene regulatory model in (2) of GRN based on molecular mechanisms of human cells, we then applied the system identification method and system order detection scheme to the models using mRNA expression data and its corresponding DNA methylation profiles in normal liver, NAFLD&NASH, PBC&PSC, and HCC to identify their corresponding real GENs. In order to identify the parameters in (1) and (2) using mRNA expression data and its corresponding DNA methylation profiles, we

rewrite (1) and (2) as the following regression forms (4) and (5), respectively:

$$y_i[n] = \begin{bmatrix} y_i[n]y_1[n] & \cdots & y_i[n]y_{i-1}[n] & y_i[n]y_{i+1}[n] & \cdots & y_i[n]y_{N_i}[n] & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} a_{i1} \\ \vdots \\ a_{ii-1} \\ a_{ii+1} \\ \vdots \\ a_{iN_i} \\ h_i \end{bmatrix} + v_i[n] = \phi_{i,\text{PPIN}}[n]\theta_{i,\text{PPIN}} + v_i[n],$$

$$\text{for } i = 1, \ldots, N, \ n = 1, \ldots, L, \ h_i \geq 0,$$

$$(4)$$

$$x_i[n] = \begin{bmatrix} y_1[n]M_i[n] & \cdots & y_{i-1}[n]M_i[n] & y_{i+1}[n]M_i[n] \end{bmatrix}$$

$$\cdots \ y_{K_i}[n]M_i[n] \ x_i[n]x_1^{\text{miR}}[n]M_i[n]$$

$$\cdots \ x_i[n]x_{V_i}^{\text{miR}}[n]M_i[n] \ M_i[n] \Big]$$

$$\times \begin{bmatrix} b_{i1} \\ \vdots \\ b_{ii-1} \\ b_{ii+1} \\ \vdots \\ b_{iK_i} \\ -c_{i1} \\ \vdots \\ -c_{iV_i} \\ \kappa_i \end{bmatrix} + \omega_i[n] = \phi_{i,\text{GRN}}[n]\theta_{i,\text{GRN}} + \omega_i[n], \quad (5)$$

$$\text{for } i = 1, \ldots, N, \ n = 1, \ldots, L, \ \kappa_i, c_{iv} \geq 0,$$

where $\phi_{i,\text{PPIN}}[n]$ and $\phi_{i,\text{GRN}}[n]$ denote the regression vectors, which consist of mRNA expression data and its corresponding DNA methylation profiles, and $\theta_{i,\text{PPIN}}$ and $\theta_{i,\text{GRN}}$ indicate the parameter vectors, which consist of protein interaction abilities, transcriptional and posttranscriptional regulatory abilities and basal levels, i.e., linking weights of GEN.

Moreover, we rewrite the regression forms in (4) and (5) as the following matrix forms for all L patients, respectively:

$$Y_i = \Phi_{i,\text{PPIN}}\theta_{i,\text{PPIN}} + V_i, \quad \text{for } i = 1, \ldots, N, \ h_i \geq 0,$$

$$X_i = \Phi_{i,\text{GRN}}\theta_{i,\text{GRN}} + W_i, \quad \text{for } i = 1, \ldots, N, \ \kappa_i, c_{iv} \geq 0,$$

$$(6)$$

where

$$Y_i = \begin{bmatrix} y_i[1] \\ \vdots \\ y_i[L] \end{bmatrix},$$

$$X_i = \begin{bmatrix} x_i[1] \\ \vdots \\ x_i[L] \end{bmatrix},$$

$$\Phi_{i,\text{PPIN}} = \begin{bmatrix} \phi_{i,\text{PPIN}}[1] \\ \vdots \\ \phi_{i,\text{PPIN}}[L] \end{bmatrix},$$

$$\Phi_{i,\text{GRN}} = \begin{bmatrix} \phi_{i,\text{GRN}}[1] \\ \vdots \\ \phi_{i,\text{GRN}}[L] \end{bmatrix},$$

$$(7)$$

$$V_i = \begin{bmatrix} v_i[1] \\ \vdots \\ v_i[L] \end{bmatrix},$$

$$W_i = \begin{bmatrix} \omega_i[1] \\ \vdots \\ \omega_i[L] \end{bmatrix}.$$

Therefore, the parameter vectors $\theta_{i,\text{PPIN}}$ and $\theta_{i,\text{GRN}}$ can then be obtained by solving the following constrained least square parameter estimation problems, respectively [52]:

$$\min_{\theta_{i,\text{PPIN}}} \frac{1}{2}\left\|\Phi_{i,\text{PPIN}}\theta_{i,\text{PPIN}} - Y_i\right\|_2^2$$

$$\text{s.t.} \ \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 & -1 \end{bmatrix}, \quad \theta_{i,\text{PPIN}} \leq 0,$$

$$(8)$$

$$\min_{\theta_{i,\text{GRN}}} \frac{1}{2}\left\|\Phi_{i,\text{GRN}}\theta_{i,\text{GRN}} - X_i\right\|_2^2$$

$$\text{s.t.} \ \begin{bmatrix} 0 & \cdots & 0 & \cdots & \cdots & 0 & 1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & 0 & 0 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \theta_{i,\text{GRN}} \leq 0,$$

$$(9)$$

where the accuracies of the parameter estimations in (8) and (9) were given by the following error estimations $\widehat{\sigma}_i^2 = (Y_i - \Phi_{i,\text{PPIN}}\widehat{\theta}_{i,\text{PPIN}})^T(Y_i - \Phi_{i,\text{PPIN}}\widehat{\theta}_{i,\text{PPIN}})/L$ and $\widehat{\sigma}_i^2 = (X_i - \Phi_{i,\text{GRN}}\widehat{\theta}_{i,\text{GRN}})^T(X_i - \Phi_{i,\text{GRN}}\widehat{\theta}_{i,\text{GRN}})/L$, respectively, and $\widehat{\theta}_{i,\text{PPIN}}$ and $\widehat{\theta}_{i,\text{GRN}}$ represent the identified parameters. The solutions of the above problems, in which inequality constraints can guarantee the negative effect of miRNA

regulation on genes and the positive basal levels of genes/proteins, can be obtained by applying the system identification method in MATLAB optimization toolbox using genome-wide mRNA expression data and DNA methylation profiles based on a reflective Newton method for minimizing a quadratic function [53].

Since genome-wide expression measurement of protein behaviors in normal liver, NAFLD&NASH, PBC&PSC, and HCC has not been realized yet and gene expressions are proportional to their corresponding proteins, in which 73% variance of protein abundance can be explained by mRNA abundance [54], the mRNA expressions can replace protein expressions for the above constrained least square parameter estimation problems in (8) and (9). Because mRNA expression data (GSE61260) and its corresponding DNA methylation profiles (GSE61258) from normal liver, NAFLD&NASH, and PBC&PSC have been reported and NGS data and its corresponding DNA methylation profiles from HCC have been also proposed in TCGA, i.e., Y_i, X_i, $\Phi_{i,\text{PPIN}}$, and $\Phi_{i,\text{GRN}}$ in (8) and (9) are all available; protein interaction abilities, transcriptional and posttranscriptional regulatory abilities, and basal levels, i.e., linking weights of GEN, in $\theta_{i,\text{PPIN}}$ and $\theta_{i,\text{GRN}}$ can then be identified.

Because the candidate GEN obtained from biological or computational experiments in human cells contains all putative interactions in PPIN in (8) and regulations in GRN in (9), in order to prune the false-positive protein interaction abilities and transcriptional and posttranscriptional regulatory abilities, i.e., links of GEN, based on the genome-wide expression data in normal liver, NAFLD&NASH, PBC&PSC, and HCC, we applied a system order detection scheme, AIC, to the system identification method in solving (8) and (9) to detect the real system order (or real links) by minimizing the following AIC value [55, 56]:

$$\text{AIC}(\Delta_i) = \log\left(\widehat{\sigma}_i^2\right) + \frac{2(\Delta_i)}{L}, \tag{10}$$

where Δ_i denotes the number of parameters, i.e., $\Delta_i = N_i + 1$ in the estimation problem of the PPIN model (8) and $\Delta_i = K_i + V_i + 1$ in the estimation problem of the GRN model (9), and $\widehat{\sigma}_i^2$ is the estimated residual error obtained from the system identification method, i.e., $\widehat{\sigma}_i^2 = \left(Y_i - \Phi_{i,\text{PPIN}}\widehat{\theta}_{i,\text{PPIN}}\right)^T \left(Y_i - \Phi_{i,\text{PPIN}}\widehat{\theta}_{i,\text{PPIN}}\right)/L$ and $\widehat{\sigma}_i^2 = \left(X_i - \Phi_{i,\text{GRN}}\widehat{\theta}_{i,\text{GRN}}\right)^T \left(X_i - \Phi_{i,\text{GRN}}\widehat{\theta}_{i,\text{GRN}}\right)/L$ in the estimation problems of the PPIN model (8) and the GRN model (9), respectively. In this study, we applied stepwise selection, a combination of the forward selection and backward elimination techniques until none improves the minimization problem in (10), to the candidate GEN. According to the theory of system identification [55, 57], the true system order Δ_i^* (or the true number of links) of the real GEN could minimize $\text{AIC}(\Delta_i)$ in (10). After getting the true system order Δ_i^* by minimizing $\text{AIC}(\Delta_i)$ in (10), we could prune the false-positive interactions and regulations of the candidate GEN by deleting the insignificant links out of true system order Δ_i^* detected by AIC. Moreover, we applied Student's t-test

to the real GEN to calculate the p value of each interaction or regulatory ability (or each link of the real GEN) under the null hypothesis $H_0 : a_{im} = 0$ or $H_0 : b_{\text{in}} = 0$ [58]. In this study, we applied random permutation to the sample data to generate 1000 random sets of the data. For example, if none of the 1000 permutation values exceeds the test statistic, p value is less than or equal to 10^{-3}. The p value of each interaction or regulatory ability is used to support our findings.

According to the data in GSE61260 and GSE61258, we obtained genome-wide mRNA expression data and its corresponding DNA methylation profiles in normal liver from 62 patients, NAFLD&NASH from 47 patients, and PBC&PSC from 25 patients. Additionally, according to the data in TCGA, we obtained genome-wide mRNA expression data and its corresponding DNA methylation profiles from HCC in 369 patients. Therefore, we can identify four GENs of liver cells for normal liver, NAFLD&NASH, PBC&PSC, and HCC, respectively.

In brief, we first identified the interaction and regulatory abilities, including a_{ij}, b_{ij}, and c_{iv} in (1) and (2), of the candidate GEN in each of four liver conditions by applying the system identification method using the genome-wide expression data. We then applied AIC to prune the false-positive interactions and regulations in the candidate GEN and finally obtain the real GEN according to the gene expression data and its DNA methylation profiles in each of liver conditions. We applied Student's t-test to the real GEN to calculate the p value of each interaction or regulatory ability in each of four liver conditions to support our findings.

4.4. Extracting the Core GEN from the Genome-Wide Real GEN via PNP.

After identifying the genome-wide real GENs of four liver conditions using gene expression data and DNA methylation profiles to prune the false-positive interactions and regulations in the candidate GEN by applying a system order detection scheme, these genome-wide GENs are still very complex and not easy to get an insight into their pathogenic and hepatocarcinogenic mechanisms in four liver conditions. We then apply the PNP method to determine the core proteins/genes/miRNAs, which compose the core GENs from the genome-wide real GENs of four liver conditions. The PPIs, gene regulations, miRNA inhibitions, and DNA methylation of the real GENs are described by the following models:

$$y_i[n] = \sum_{j \in N_P} \widehat{a}_{ij} y_i[n] y_j[n] + \widehat{h}_i + v_i[n], \quad \text{for } i = 1, \ldots, N, \; n = 1, \ldots, L, \tag{11}$$

$$x_i[n] = \sum_{j \in K_G} \widehat{b}_{ij} y_j[n] M_i[n] - \sum_{v \in V_G} \widehat{c}_{iv} x_i[n] x_v^{\text{miR}}[n] M_i[n] + M_i[n]\widehat{\kappa}_i + \omega_i[n], \quad \text{for } i = 1, \ldots, N, \; n = 1, \ldots, L, \tag{12}$$

where N_P, K_G, and V_G indicate the real number of PPIs, gene regulations, and miRNA inhibitions, respectively, and $\widehat{a}_{ij}$, $\widehat{b}_{ij}$, and $\widehat{c}_{iv}$ represent the estimated interaction abilities of proteins, the estimated regulatory abilities of

TFs, and the estimated posttranscriptional regulatory abilities of miRNAs by applying a system identification method and a system order detection scheme to (8) and (9), respectively.

$$H = \begin{bmatrix} \widehat{a}_{11} & \cdots & \widehat{a}_{1j} & \cdots & \widehat{a}_{1N} & \widehat{b}_{11} & \cdots & \widehat{b}_{1j} & \cdots & \widehat{b}_{1N} & -\widehat{c}_{11} & \cdots & -\widehat{c}_{1v} & \cdots & -\widehat{c}_{1V} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{a}_{i1} & \cdots & \widehat{a}_{ij} & \cdots & \widehat{a}_{iN} & \widehat{b}_{i1} & \cdots & \widehat{b}_{ij} & \cdots & \widehat{b}_{iN} & -\widehat{c}_{i1} & \cdots & -\widehat{c}_{iv} & \cdots & -\widehat{c}_{iV} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{a}_{N1} & \cdots & \widehat{a}_{Nj} & \cdots & \widehat{a}_{NN} & \widehat{b}_{N1} & \cdots & \widehat{b}_{Nj} & \cdots & \widehat{b}_{NN} & -\widehat{c}_{N1} & \cdots & -\widehat{c}_{Nv} & \cdots & -\widehat{c}_{NV} \end{bmatrix}^{\mathrm{T}}. \tag{13}$$

If a link between any two members of genes, proteins, and miRNAs in the real GEN is disconnected, its corresponding element in H is set to zero. PNP is a network structure projection method based on singular value decomposition (SVD) as follows [56]:

$$H = QSR^{\mathrm{T}}, \tag{14}$$

where $Q \in \mathscr{R}^{(2N+V) \times N}$; $H = [h_1 \cdots h_k \cdots h_{2N+V}]^{\mathrm{T}} \in \mathscr{R}^{(2N+V) \times N}$; $h_k \in \mathscr{R}^{1 \times N}$; $R = [r_1 \cdots r_m \cdots r_N] \in \mathscr{R}^{N \times N}$; $r_m \in \mathscr{R}^{N}$; S is a diagonal matrix such as $S = \mathrm{diag}(s_1, \cdots, s_m, \cdots, s_N)$, which consists of N non-negative singular values of H with descending order $s_1 \geq \cdots \geq s_m \geq \cdots \geq s_N$; and $\mathrm{diag}(s_1, s_2)$ denotes the diagonal matrix of s_1 and s_2, i.e., $\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$. We then defined the eigenexpression fraction (E_m) as follows:

$$E_m = \frac{s_m^2}{\sum_{m=1}^{N} s_m^2}. \tag{15}$$

The principal components to satisfy $\sum_{m=1}^{M} E_m \geq 0.85$ with the minimal M are selected as the top M singular vectors of R in (14) to determine the core proteins/genes/miRNAs. Furthermore, the genome-wide real GEN in H is projected to the top M singular vector of R as follows:

$$p(k, m) = h_k r_m, \quad \text{for } k = 1, \dots, 2N + V, \ m = 1, \dots, M. \tag{16}$$

We can then obtain the projection value from each member of the real GEN onto the top M singular vectors (i.e., principal network structure) by applying the 2-norm to (16) as the dependent score $D(k)$ of each member as follows:

$$D(k) = \left(\sum_{m=1}^{M} [p(k, m)]^2 \right)^{1/2}, \quad \text{for } k = 1, \dots, 2N + V. \tag{17}$$

We generally integrated $\widehat{a}_{ij}$, $\widehat{b}_{ij}$, and $\widehat{c}_{iv}$, linking weights of the real GEN, from (11) and (12) as the following network structure matrix H:

According to the projection value, i.e., dependent score $D(k)$, from each member of the real GEN onto the top M singular vectors of H based on their respective interactions or regulatory abilities, $\widehat{a}_{ij}$, $\widehat{b}_{ij}$, and $\widehat{c}_{iv}$, we can finally calculate the dependent score $D(k)$ of proteins ($k = 1, \dots, N$), genes ($k = N + 1, \dots, 2N$) and miRNAs ($k = 2N + 1, \dots, 2N + V$) to the principal network structure of the genome-wide real GEN. If $D(k)$ is close to zero, it means that the kth member of the real GEN is independent of the top M singular vectors (or principal network structure). Otherwise, a member of the real GEN with higher $D(k)$ is more important for the principal network structure of the real GEN. In order to determine the core proteins/genes/miRNAs, which compose the core GEN from a genome-wide real GEN, we choose the top 5% proteins in $D(k)$ for $k = 1, \dots, N$ and their connecting genes and miRNAs not only to constitute the core GEN but also to constitute the complete connections in signal transduction; i.e., the top 5% proteins just could construct a complete signaling cascade from receptors to TFs. Therefore, we applied PNP to the real GENs of normal liver, NAFLD&NASH, PBC&PSC, and HCC to obtain the core GENs of them. By comparing the core GENs between two liver conditions, we further extracted signaling pathways and network biomarkers from the core GENs to investigate the cellular mechanisms of pathogenesis and hepatocarcinogenesis for preventing the progression of liver damage.

5. Conclusion

In this study, we focused on constructing the corresponding signaling pathways involved in the core GENs to further investigate cellular mechanisms of progression from normal liver cells to HCC through NAFLD&NASH or PBC&PSC in Figure 1(a), based on miRNA regulation and epigenetic regulation in GENs. To begin with, we proposed a novel approach to constructing the genome-wide candidate GEN using big database mining. We then pruned the false-positive regulatory interactions using a system order detection scheme to identify the genome-

wide real GEN. We also used PNP to construct core GENs for different liver diseases. Since core GENs are complicated, mechanism interpretation is difficult. Therefore, we extracted core networks for each progression stage in Figure 1(a) by comparing different core GENs and further submitted core networks to the KEGG database to analyze relevant signaling pathways and to investigate cellular mechanisms of progression from normal liver cells to HCC through NAFLD&NASH or PBC&PSC. Based on the core signaling pathways, we also realized genetic and epigenetic regulatory mechanisms of progression from normal liver cells to HCC through NAFLD&NASH or PBC&PSC using KEGG and GO database mining. Finally, we observed that some miRNA regulations and epigenetic regulations fulfilled essential roles in the mechanisms of progression from normal liver cells to HCC through the upper or the lower progression paths in Figure 8. Hence, our results show key network biomarkers of miRNA regulation and methylation in the mechanisms of progression. Therefore, network biomarkers could also be developed based on potential drug targets for the two progression paths of pathogenesis and hepatocarcinogenesis in Figure 8.

Abbreviations

AIC:	Akaike information criterion
aldolase B:	ALDOB
APP:	Amyloid precursor protein
AR:	Androgen receptor
ATZ:	Alpha-1 antitrypsin Z
BioGRID:	Biological General Repository for Interaction Datasets
EGFR:	Epidermal growth factor receptor
ETS1:	V-Ets avian erythroblastosis virus E26 oncogene homolog 1
FRAT2:	Frequently rearranged in advanced T-cell lymphomas 2
GEN:	Genome-wide real genetic and epigenetic network
GEO:	Gene expression omnibus
GO:	Gene ontology
GRN:	Gene regulatory network
H2be:	Histone cluster 2
H3F3A:	H3 histone family 3A
HCC:	Hepatocellular carcinoma
HSPB1:	Heat shock 27 kDa protein 1
HTRI:	Human transcriptional regulation interactions
IGF2:	Insulin-like growth factor 2
ITFP:	Integrated transcription factor platform
KEGG:	Kyoto encyclopedia of genes and genomes
NAFLD:	Nonalcoholic fatty liver disease
NASH:	Nonalcoholic steatohepatitis
NAFLD&NASH:	Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis
NCBI:	National Center for Biotechnology Information
miRNA:	MicroRNA
PBC:	Primary biliary cirrhosis
PBC&PSC:	Primary biliary cirrhosis and primary sclerosing cholangitis
PSC:	Primary sclerosing cholangitis
PNP:	Principal network projection
PPI:	Protein-protein interaction
PPIN:	Protein-protein interaction network
PTEN:	Phosphatase and tensin homolog
RFC5:	Replication factor C5
RPL30:	Ribosomal protein L30
RYR2:	Ryanodine receptor 2
SHC1:	SHC-transforming protein 1
SVD:	Singular value decomposition
TCGA:	The Cancer Genome Atlas
TF:	Transcription factor
TIMP1:	TIMP metallopeptidase inhibitor 1
TRANSFAC:	TRANScription FACtor database
TRMT1:	tRNA methyltransferase 1
TUBA1C:	Tubulin alpha 1c
UBC:	Ubiquitin C
ZNF480:	Zinc finger protein 480.

Data Availability

The data supporting the results of this article are included within the article and its additional files.

Disclosure

Cheng-Wei Li and Yu-Kai Chiu are the co-first author.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Figure S1: the genetic and epigenetic network (GEN) for normal liver cell. Figure S2: the genetic and epigenetic network (GEN) for NAFLD&NASH. Figure S3: the genetic and epigenetic network (GEN) for PBC&PSC. Figure S4: the genetic and epigenetic network (GEN) for HCC. Figure S5: the core genetic and epigenetic network (GEN) for normal liver cell 41. Figure S6: the core genetic and epigenetic network (GEN) for NAFLD&NASH. Figure S7: the core genetic and epigenetic network (GEN) for PBC&PSC. Figure S8: the core genetic and epigenetic network (GEN) for HCC. Table S1: the projection values of the 37 network biomarkers NAFLD&NASH pathogenesis and NAFLD&NASH-associated hepatocarcinogenesis. Table S2: the projection values of the 40 network biomarkers in PBC&PSC pathogenesis and PBC&PSC-associated hepatocarcinogenesis. Table S3: a list of ranked drugs to prevent

hepatocarcinogenesis of NAFLD&NASH (A) and PBC&PSC (B). *(Supplementary Materials)*

References

[1] S. Bandiera, S. Pfeffer, T. F. Baumert, and M. B. Zeisel, "miR-122–a key factor and therapeutic target in liver disease," *Journal of Hepatology*, vol. 62, no. 2, pp. 448–457, 2015.

[2] H. Nishikawa and Y. Osaki, "Non-B, non-C hepatocellular carcinoma," *International Journal of Oncology*, vol. 43, no. 5, pp. 1333–1342, 2013.

[3] H. Nishikawa, A. Arimoto, T. Wakasa, R. Kita, T. Kimura, and Y. Osaki, "Comparison of clinical characteristics and survival after surgery in patients with non-B and non-C hepatocellular carcinoma and hepatitis virus-related hepatocellular carcinoma," *Journal of Cancer*, vol. 4, no. 6, pp. 502–513, 2013.

[4] S. M. El-Haggar and T. M. Mostafa, "Comparative clinical study between the effect of fenofibrate alone and its combination with pentoxifylline on biochemical parameters and liver stiffness in patients with non-alcoholic fatty liver disease," *Hepatology International*, vol. 9, no. 3, pp. 471–479, 2015.

[5] M. J. Pollheimer and P. Fickert, "Animal models in primary biliary cirrhosis and primary sclerosing cholangitis," *Clinical Reviews in Allergy & Immunology*, vol. 48, no. 2-3, pp. 207–217, 2015.

[6] K. D. Williamson and R. W. Chapman, "Primary sclerosing cholangitis: a clinical update," *British Medical Bulletin*, vol. 114, no. 1, pp. 53–64, 2015.

[7] P. Arner and A. Kulyte, "MicroRNA regulatory networks in human adipose tissue and obesity," *Nature Reviews Endocrinology*, vol. 11, no. 5, pp. 276–288, 2015.

[8] Y. Y. Feng, X. Q. Xu, C. B. Ji, C. M. Shi, X. R. Guo, and J. F. Fu, "Aberrant hepatic microRNA expression in nonalcoholic fatty liver disease," *Cellular Physiology and Biochemistry*, vol. 34, no. 6, pp. 1983–1997, 2014.

[9] M. A. Javidi, A. H. Ahmadi, B. Bakhshinejad, N. Nouraee, S. Babashah, and M. Sadeghizadeh, "Cell-free microRNAs as cancer biomarkers: the odyssey of miRNAs through body fluids," *Medical Oncology*, vol. 31, no. 12, p. 295, 2014.

[10] S. Gilad, E. Meiri, Y. Yogev et al., "Serum microRNAs are promising novel biomarkers," *PLoS One*, vol. 3, no. 9, article e3148, 2008.

[11] J. K. DiStefano and G. S. Gerhard, "Circulating microRNAs in nonalcoholic fatty liver disease," *Expert Review of Gastroenterology & Hepatology*, vol. 10, no. 2, pp. 161–163, 2015.

[12] M. G. M. Pruis, Á. Lendvai, V. W. Bloks et al., "Maternal western diet primes non-alcoholic fatty liver disease in adult mouse offspring," *Acta Physiologica*, vol. 210, no. 1, pp. 215–227, 2014.

[13] M. Ahrens, O. Ammerpohl, W. von Schönfels et al., "DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery," *Cell Metabolism*, vol. 18, no. 2, pp. 296–302, 2013.

[14] M. Ehrlich and K. C. Ehrlich, "Effect of DNA methylation on the binding of vertebrate and plant proteins to DNA," *DNA Methylation*, vol. 64, pp. 145–168, 1993.

[15] M. Weber, I. Hellmann, M. B. Stadler et al., "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome," *Nature Genetics*, vol. 39, no. 4, pp. 457–466, 2007.

[16] D. Komander, "The emerging complexity of protein ubiquitination," *Biochemical Society Transactions*, vol. 37, no. 5, pp. 937–953, 2009.

[17] B. De Strooper and W. Annaert, "Proteolytic processing and cell biological functions of the amyloid precursor protein," *Journal of Cell Science*, vol. 113, Part 11, pp. 1857–1870, 2000.

[18] Y. Zheng, C. Zhang, D. R. Croucher et al., "Temporal regulation of EGF signalling networks by the scaffold protein Shc1," *Nature*, vol. 499, no. 7457, pp. 166–171, 2013.

[19] H.-J. Shih, H.-H. Chen, Y.-A. Chen et al., "Targeting MCT-1 oncogene inhibits Shc pathway and xenograft tumorigenicity," *Oncotarget*, vol. 3, no. 11, pp. 1401–1415, 2012.

[20] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano, "A comprehensive pathway map of epidermal growth factor receptor signaling," *Molecular Systems Biology*, vol. 1, no. 1, pp. E1–E17, 2005.

[21] H. Lanaya, A. Natarajan, K. Komposch et al., "EGFR has a tumour-promoting role in liver macrophages during hepatocellular carcinoma formation," *Nature Cell Biology*, vol. 16, no. 10, pp. 972–981, 2014.

[22] D. Xu, T. J. Wilson, D. Chan et al., "Ets1 is required for p53 transcriptional activity in UV-induced apoptosis in embryonic stem cells," *The EMBO Journal*, vol. 21, no. 15, pp. 4081–4093, 2002.

[23] M. J. Bround, R. Wambolt, D. S. Luciani et al., "Cardiomyocyte ATP production, metabolic flexibility, and survival require calcium flux through cardiac ryanodine receptors in vivo," *Journal of Biological Chemistry*, vol. 288, no. 26, pp. 18975–18986, 2013.

[24] C. Sun, J.-G. Fan, and L. Qiao, "Potential epigenetic mechanism in non-alcoholic fatty liver disease," *International Journal of Molecular Sciences*, vol. 16, no. 12, pp. 5161–5179, 2015.

[25] M. A. Dawson and T. Kouzarides, "Cancer epigenetics: from mechanism to therapy," *Cell*, vol. 150, no. 1, pp. 12–27, 2012.

[26] J. H. Lee, S. Friso, and S.-W. Choi, "Epigenetic mechanisms underlying the link between non-alcoholic fatty liver diseases and nutrition," *Nutrients*, vol. 6, no. 8, pp. 3303–3325, 2014.

[27] Y.-Y. Li, "Genetic and epigenetic variants influencing the development of nonalcoholic fatty liver disease," *World journal of gastroenterology: WJG*, vol. 18, no. 45, pp. 6546–6551, 2012.

[28] Y. Murakami, T. Yasuda, K. Saigo et al., "Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues," *Oncogene*, vol. 25, no. 17, pp. 2537–2545, 2006.

[29] F. Meng, R. Henson, H. Wehbe–Janek, K. Ghoshal, S. T. Jacob, and T. Patel, "MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer," *Gastroenterology*, vol. 133, no. 2, pp. 647–658, 2007.

[30] J. Xu, C. Wu, X. Che et al., "Circulating microRNAs, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis," *Molecular Carcinogenesis*, vol. 50, no. 2, pp. 136–142, 2011.

[31] C. Sun, F. Huang, X. Liu et al., "miR-21 regulates triglyceride and cholesterol metabolism in non-alcoholic fatty liver disease by targeting HMGCR," *International Journal of Molecular Medicine*, vol. 35, no. 3, pp. 847–853, 2015.

[32] E. K. Kim and E.-J. Choi, "Pathological roles of MAPK signaling pathways in human diseases," *Biochimica et Biophysica*

Acta (BBA) - Molecular Basis of Disease, vol. 1802, no. 4, pp. 396–405, 2010.

[33] S.-Y. Peng, P.-L. Lai, H.-W. Pan, L.-P. Hsiao, and H.-C. Hsu, "Aberrant expression of the glycolytic enzymes aldolase B and type II hexokinase in hepatocellular carcinoma are predictive markers for advanced stage, early recurrence and poor prognosis," *Oncology Reports*, vol. 19, no. 4, pp. 1045–1053, 2008.

[34] M. De Bortoli, R. C. Castellino, X.-Y. Lu et al., "Medulloblastoma outcome is adversely associated with overexpression of EEF1D, RPL30, and RPS20 on the long arm of chromosome 8," *BMC Cancer*, vol. 6, no. 1, p. 223, 2006.

[35] S. Hohenester, R. P. Oude-Elferink, and U. Beuers, "Primary biliary cirrhosis," in *Seminars in Immunopathology*, pp. 283–307, Springer, 2009.

[36] J. Baier and J. Mattner, "Mechanisms of autoimmune liver disease," *Discovery Medicine*, vol. 18, no. 100, pp. 255–263, 2014.

[37] S.-h. Hsu, B. Wang, J. Kota et al., "Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver," *Journal of Clinical Investigation*, vol. 122, no. 8, pp. 2871–2883, 2012.

[38] Y. Xiong, J. H. Fang, J. P. Yun et al., "Effects of MicroRNA-29 on apoptosis, tumorigenicity, and prognosis of hepatocellular carcinoma," *Hepatology*, vol. 51, no. 3, 2010.

[39] M. Ciechomska, S. O'Reilly, M. Suwara, K. Bogunia-Kubik, and J. M. van Laar, "MiR-29a reduces TIMP-1 production by dermal fibroblasts via targeting TGF-β activated kinase 1 binding protein 1, implications for systemic sclerosis," *PLoS One*, vol. 9, no. 12, article e115596, 2014.

[40] Y. Gong, E. Scott, R. Lu, Y. Xu, W. K. Oh, and Q. Yu, "TIMP-1 promotes accumulation of cancer associated fibroblasts and cancer progression," *PLoS One*, vol. 8, no. 10, article e77366, 2013.

[41] S. Kallappagoudar, R. K. Yadav, B. R. Lowe, and J. F. Partridge, "Histone H3 mutations—a special role for H3. 3 in tumorigenesis?," *Chromosoma*, vol. 124, no. 2, pp. 177–189, 2015.

[42] A. A. Nagle, F.-F. Gan, G. Jones, C.-L. So, G. Wells, and E.-H. Chew, "Induction of tumor cell death through targeting tubulin and evoking dysregulation of cell cycle regulatory proteins by multifunctional cinnamaldehydes," *PLoS One*, vol. 7, no. 11, article e50125, 2012.

[43] N. V. Chaika, F. Yu, V. Purohit et al., "Differential expression of metabolic genes in tumor and stromal components of primary and metastatic loci in pancreatic adenocarcinoma," *PLoS One*, vol. 7, no. 3, article e32996, 2012.

[44] Y. Bronevetsky and K. M. Ansel, "Regulation of miRNA biogenesis and turnover in the immune system," *Immunological Reviews*, vol. 253, no. 1, pp. 304–316, 2013.

[45] S. Horvath, W. Erhart, M. Brosch et al., "Obesity accelerates epigenetic aging of human liver," *Proceedings of the National Academy of Sciences*, vol. 111, no. 43, pp. 15538–15543, 2014.

[46] A. Chatr-aryamontri, B.-J. Breitkreutz, R. Oughtred et al., "The BioGRID interaction database: 2015 update," *Nucleic Acids Research*, vol. 43, no. D1, pp. D470–D478, 2015.

[47] L. A. Bovolenta, M. L. Acencio, and N. Lemke, "HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions," *BMC Genomics*, vol. 13, no. 1, p. 405, 2012.

[48] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, no. 1, pp. 15–20, 2005.

[49] A. E. N. Ferreira, A. M. J. Ponces Freire, and E. O. Voit, "A quantitative model of the generation of N(epsilon)-(carboxymethyl)lysine in the Maillard reaction between collagen and glucose," *Biochemical Journal*, vol. 376, no. 1, pp. 109–121, 2003.

[50] E. O. Voit, *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, 2000.

[51] Y. A. Medvedeva, A. M. Khamis, I. V. Kulakovskiy et al., "Effects of cytosine methylation on transcription factor binding sites," *BMC Genomics*, vol. 15, no. 1, p. 119, 2014.

[52] C. W. Li, W. H. Wang, and B. S. Chen, "Investigating the specific core genetic-and-epigenetic networks of cellular mechanisms involved in human aging in peripheral blood mononuclear cells," *Oncotarget*, vol. 7, no. 8, pp. 8556–8579, 2016.

[53] T. F. Coleman and Y. Y. Li, "A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.

[54] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation," *Nature Biotechnology*, vol. 25, no. 1, pp. 117–124, 2007.

[55] R. Johansson, *System Modeling & Identification*, Prentice-hall, 1993.

[56] C. W. Li and B. S. Chen, "Network biomarkers of bladder cancer based on a genome-wide genetic and epigenetic network derived from next-generation sequencing data," *Disease Markers*, vol. 2016, Article ID 4149608, 18 pages, 2016.

[57] L. Z. Chen, S. K. Nguang, and X. D. Chen, "On-line identification and optimization of feed rate profiles for high productivity fed-batch culture of hybridoma cells using genetic algorithms," *ISA Transactions*, vol. 41, no. 4, pp. 409–419, 2002.

[58] M. Pagano and K. Gauvreau, *Principles of Biostatistics*, Chapman and Hall/CRC, Pacific Grove, CA: Duxbury, 2nd edition, 2000.