



Article

Investigating Core Signaling Pathways of Hepatitis B Virus Pathogenesis for Biomarkers Identification and Drug Discovery via Systems Biology and Deep Learning Method

Shen Chang ¹, Lily Hui-Ching Wang ²  and Bor-Sen Chen ^{1,*}

¹ Laboratory of Automatic Control, Signal Processing and Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan; s107061595@m107.nthu.edu.tw

² Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu 30013, Taiwan; lilywang@life.nthu.edu.tw

* Correspondence: bschen@ee.nthu.edu.tw

Received: 6 July 2020; Accepted: 21 August 2020; Published: 31 August 2020



Abstract: Hepatitis B Virus (HBV) infection is a major cause of morbidity and mortality worldwide. However, poor understanding of its pathogenesis often gives rise to intractable immune escape and prognosis recurrence. Thus, a valid systematic approach based on big data mining and genome-wide RNA-seq data is imperative to further investigate the pathogenetic mechanism and identify biomarkers for drug design. In this study, systems biology method was applied to trim false positives from the host/pathogen genetic and epigenetic interaction network (HPI-GEN) under HBV infection by two-side RNA-seq data. Then, via the principal network projection (PNP) approach and the annotation of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, significant biomarkers related to cellular dysfunctions were identified from the core cross-talk signaling pathways as drug targets. Further, based on the pre-trained deep learning-based drug-target interaction (DTI) model and the validated pharmacological properties from databases, i.e., drug regulation ability, toxicity, and sensitivity, a combination of promising multi-target drugs was designed as a multiple-molecule drug to create more possibility for the treatment of HBV infection. Therefore, with the proposed systems medicine discovery and repositioning procedure, we not only shed light on the etiologic mechanism during HBV infection but also efficiently provided a potential drug combination for therapeutic treatment of Hepatitis B.

Keywords: hepatitis B virus infection; pathogenesis; host/pathogen interspecies genetic and epigenetic network (HPI-GEN); systems medicine discovery; drug-target interaction (DTI) model; deep learning; multiple-molecule drug

1. Introduction

Hepatitis B Virus (HBV) infection remains a prevalent health challenge worldwide, and is greatly associated with substantial morbidity and mortality due to serious liver diseases, such as hepatitis, fibrosis, cirrhosis, and even hepatocellular carcinoma [1]. According to the latest report of World Health Organization (WHO) [2], HBV affected an estimated 257 million people and causing 600,000 deaths per year. While current therapies seek to control the progression of the disease, life-long treatment and surveillance are still needed because resistance develops during treatment and reactivation often occurs after medication discontinuation [3,4]. In recent years, dedication to new drug design and immunotherapy development has been made by researchers to hopefully silence HBV in infected hepatocytes [5,6]. However, the host/pathogen cross-talk mechanism contributing

to the development and persistence of HBV infection requires further investigation, and its current therapies are still inadequate. Thus, new treatment options are needed to achieve a better therapy for HBV-infected patients.

Even so, *de novo* drug discovery is particularly an expensive, time-consuming, and high-risk process, let alone exploring a drug to eradicate HBV from patients, which is generally believed arduous [4,7,8]. Conventionally, the pipelines of drug development from laboratory to market take nearly average 13 to 15 years and cost approximately two to three billion U.S. dollars, in which about 10% of all drugs that start from pre-clinical trials ever make it to actual human testing [9,10]. Fortunately, as the advancement of biological science drives the progress of high-throughput functional genomics and proteomics technology, novel experimental and bioinformatics methods have not only accelerated the deciphering of protein interactions, but also given rise to the construction of networks for signaling pathways and protein complex identification in specific diseases [11]. Meanwhile, based on the manifold data types from wet-lab experiments, several computational approaches have been suggested to predict biological networks for expediting the modeling of functional pathways so as to exemplify the molecular mechanisms of cellular processes. Researchers are thereby being empowered to profoundly explore the complicated and mysterious mechanisms of biological systems for better investigating the potential drug targets [12]. On the other hand, with the development of molecular biology, drug repurposing, also known as drug repositioning, was proposed to reduce the time spent and the capital expenditure. The process of drug repositioning is to find new use outside the scope of the original medical indication for existing drugs [13]. Its prospect for new clinical indications is promising, since compared with other drug development strategies, a repositioned drug could be instantly put into preclinical testing and clinical trials for its safety and reliability proven in humans.

Nonetheless, the commonly used correlation networks in bioinformatics applications are often based on oversimplified algorithms. As inherent difficulties appear in assessing the impact of microRNA (miRNA), long non-coding RNA (lncRNA), epigenetic modification, microenvironment factors, etc., hardly can these networks mirror the complicated mechanisms of factual biological systems. In addition, while wide doses of the repositioned drugs can be tested using *in vitro* and *in vivo* experiments, it takes great time and effort. Therefore, to discover potential drugs for the treatment of HBV-infected patients, the development of a systematic procedure in computational framework based on potent models and existing data is indispensable.

In this study, we employed the systems biology technique to identify the significant biomarkers from the host/pathogen interspecies genetic and epigenetic interaction network (HPI-GEN) through two-side RNA-seq data during HBV infection. Subsequently, with the analysis of these biomarkers by the proposed drug discovery and repositioning procedure, potential combinations of multi-targets drugs were recommended to the clinical trials of Hepatitis B therapy. By and large, the process can be subdivided into a few steps: (1) Construct candidate HPI-GEN composed of host/pathogen interspecies protein-protein interaction network (HPI-PPIN) and host/pathogen interspecies gene regulation network (HPI-GRN) by big data mining. (2) Prune the false positives in candidate HPI-GEN by system identification analysis with the assistance of two-side profile RNA-seq data and Akaike Information Criterion (AIC) to identify the real HPI-GEN between host and pathogen. (3) Extract the core HPI-GEN from real HPI-GEN via principal network projection (PNP) approach. (4) Investigate the core signaling pathways related to manifold cellular function (especially host inflammatory and immune response, pathological anti-apoptosis, and cell survival) in the core HPI-GEN by the annotation of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways to gain insights into the etiologic mechanism of HBV infection. (5) Select the significant biomarkers related to pathological cellular dysfunctions. (6) Predict the candidate drugs from available drugs for the selected significant biomarkers by the pre-trained deep learning-based drug-target interaction (DTI) model. (7) Consider general criteria for drug design specifications including regulation ability, toxicity, and sensitivity to further sift out the promising drugs. (8) Combine these agreeable drugs as a potential multiple-molecule drug for possible therapeutic treatment of Hepatitis B. It is worthwhile to note that instead of a single drug,

a combination of promising drugs was proposed for the treatment to slow or even prevent inherent or adaptive drug resistance through the simultaneous blockade of multiple targets or pathways [14].

Nowadays, researches have been dedicated to drug discovery and design, while few of them propose an integral end-to-end systematic procedure from identifying significant biomarkers based on pathogenic mechanism to recommending promising drugs for one disease. Here, we dig into the pathogenic characters of viral proteins played on the pathogenesis to select the potential drug targets from two-side RNA-seq data and reinterpret the role of available drugs based on a pre-trained deep learning-based DTI model through big drug-target interaction data, which gives an alternative way for systems drug discovery and repositioning to overcome HBV infection.

2. Materials and Methods

2.1. The Construction of the HPI-GEN in HepaRG Cell Line during HBV Infection and the Application of Deep Learning-Based DTI Model for New Drug Design: An Overview

To gain new insight into the molecular mechanism of HBV infection, candidate cross-talk HPI-GEN, which consists of candidate intraspecies, candidate interspecies PPIN between host and pathogen, and candidate HPI-GRN encompassing regulatory networks between host gene, pathogen gene, host miRNA and host lncRNA, was constructed by big data mining and preprocessing of host/pathogen gene/miRNA/lncRNA expression from online database.

Given that the information for interspecies PPIN construction is insufficient, in addition to interactions attained from bio-databases, data amplification technique through nature-linguistic programming (NLP) and web scrapping was introduced to expand the dataset. Then, for identifying real HPI-GEN, a systematic identification approach was exploited based on the genome-wide RNA-seq data of HepaRG cells and Hepatitis B Virus under HBV infection. In addition, to deal with the highly complicate real HPI-GEN, we employed PNP method to extract core network structure consisted of the top 3000 nodes with highest energy. Such effective approach benefits us to investigate the core signaling pathways for establishing the crucial molecular pathogenic mechanism to identify significant biomarkers that contribute to the progression of HBV infection. Finally, following the drug discovery procedure, we built the deep learning-based DTI model to identify the potential multi-target drugs to target their significant biomarkers for therapeutic treatment of HBV infection.

2.2. Big Data Mining and Data Preprocessing of RNA-seq Data for Human and Pathogen

In this study, two-side sequencing-based dataset was obtained from the Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI). We obtained the data recorded by Mueller et al. [15], containing human and virus genomes with the time-course through GEO series with accession number GSE101575 [15]. The raw data of the dataset, which contains 27,267 human genes and 4 HBV genes, involves the gene expression profiles of human HepaRG cells from three patients during HBV infection at 6, 9, 12, and 24 h post reactivation with drug (RG7834) treatment and without. Only the data without medication administration was deployed to the ensuing system identification process.

To unravel the cellular mechanism during HBV infection, core signaling pathways identified from HPI-GEN are requisite. The candidate HPI-GEN, involving many experimental data and computational predictions from numerous databases, consists of multiple networks. Among them, the candidate human PPIs were obtained from the Biological General Repository for Interaction Datasets database (BioGRID) [16], DIP [17], BIND [18], IntAct [19], and MINT [20]; the candidate human miRNA and lncRNA regulations were individually obtained from Target Scan Human database [21], and CircuitDB [22]; the candidate regulations between human transcription factors and their targets were according to the Integrated Transcription Factor Platform database (ITFP) [23], the Human Transcriptional Regulation Interactions database (HTRIdb) [24], and the Transcription Factor database (TRANSFAC) [25]; the candidate host-pathogen interspecies PPIs were obtained from

VirusMentha [26], IMEx [27], PSICQUIC [28], the PPIN constructed by Zhong et al. [29], and based on the automatic pipeline method of NLP; the candidate host-pathogen interspecies regulations were collected from VIRmiRNA [30], ViRBase [31], miRecords [32], starBase v2.0 [33], miRTarBase [34], and based on the automatic pipeline method of NLP.

2.3. Text Mining of Human and Pathogen Protein Interactions

To reinforce the comprehension of the interaction relationships between host and pathogen, BioBERT (Bidirectional Encoder Representations from Transformer for Biomedical Text Mining) [35,36], an up-to-date BioNER (Biomedical Named Entity Recognition) approach is employed on the PubMed database to analyze binary interactions between host and pathogen. BioNER is a computerized procedure of identifying biomedical named entities such as genes, protein complexes, diseases, and chemicals in texts referring to information retrieval and data extraction. Amongst all the BioNER application, BioBERT, which has achieved state-of-the-art performance in terms of precision, is efficaciously utilized in biomedical text mining applications to address data.

In brief, we scraped and collected 26,732 existing PubMed articles associated with Hepatitis B Virus and Homo sapiens within 20 years and transformed it into XML format. Nonetheless, on account of Pubmed's property, the full texts from most articles in PMC aren't publicly accessible so that looking forward to automatically download integrated content is unrealistic. Consequently, we additionally crawled 2831 articles from PLOS ONE to strengthen our network complexity. Then, BioBERT model was exerted to discern known entities and identify the types of overlapping entities among each sentence separated from the relevant content with periods in the corpus.

In order to facilitate analysis and reconfirmation, we also gathered the synonym names of both host and pathogen proteins (compiled from BioThesaurus database [37]) to prevent omission of vital interactions owing to protein names mismatch. Afterwards, in each sentence, we extracted all present proteins and identified the correlations by determining the human proteins under the circumstance that Hepatitis B Virus proteins had existed. To abate computational complexity, in conformity with the subsequent system identification processing requirement, the verb recognition and identification were left out of account and binary interactions were built. For instance, providing that "HBx" and "BCL2" were simultaneously identified in the same sentence, we then set the corresponding value to 1 and so forth, vice versa to 0.

Furthermore, to verify that authentic interaction relationships derived from the statistical analysis of sentences, we scrutinized entire extracted entities and completed the construction of ultimate host and pathogen interaction-relationships matrix.

2.4. Dynamic Models of Candidate HPI-GEN for Human Cells and HBV During the Infection

Candidate HPI-GEN, constructed via big data mining techniques, is the epitome of an intricate network comprised of protein interactions and gene regulations between host and pathogen during HBV infection. As inevitable false-positive information extraction might incur system distraction from genuine consequence and discrepancy during the pathway analysis process, dynamic models were built to carry out systematic identification approach by two-side time-profile HBV infection RNA-seq data to prune false-positives in candidate HPI-GEN.

Protein abundances mirror a dynamic balance among a series of linked processes spanning the transcription, processing, and degradation of mRNAs to the translation, modification, and breakdown of the proteins themselves necessary for cellular protein production and maintenance [38]. As a consequence, for the protein interactive model of PPI pre-network in candidate HPI-GEN, the dynamic interaction model of the i th host protein can be described by the following equation:

$$p_i^H(t+1) = p_i^H(t) + \sum_{n=1}^{N_i} a_{in}^H p_i^H(t) p_n^H(t) + \sum_{v=1}^{V_i} b_{iv}^H p_i^H(t) p_v^P(t) + \alpha_i^H g_i^H(t) - \gamma_i^H p_i^H(t) + \beta_i^H + v_i^H(t), \text{ for } i = 1, 2, \dots, I, \alpha_i^H \geq 0 \text{ and } -\gamma_i^H \leq 0 \quad (1)$$

where $p_i^H(t)$, $p_n^H(t)$, $g_i^H(t)$ and $p_v^P(t)$ indicate the expression levels of the i th host protein, the n th host protein, the i th host gene, and the v th pathogen protein at time t , respectively; a_{in}^H and b_{iv}^H denote the interactive ability between the i th host protein and n th host protein and between the i th host protein and v th pathogen protein, respectively; N_i and V_i signify the number of host proteins and pathogen proteins interacting with the i th host protein; and I typifies the total number of the human proteins in candidate PPIN. α_i^H , $-\gamma_i^H$ and β_i^H indicate the translation rate from the corresponding mRNA, the degradation rate, and the basal expression level of the i th host protein, respectively; the basal level denotes the interactions with unknown factors, such as acetylation, methylation, and ubiquitination; and $v_i^H(t)$ represents the stochastic noise of the i th host protein at time t . Note that the intraspecies and interspecies biological mechanism of human proteins in candidate PPIN are represented in Equation (1) as the form of $\sum_{n=1}^{N_i} a_{in}^H p_i^H(t) p_n^H(t)$ and $\sum_{v=1}^{V_i} b_{iv}^H p_i^H(t) p_v^P(t)$, respectively.

Now that the interactions between the host and pathogen proteins are mutual and their biological characteristics involved in PPIs are similar, the dynamic PPIN of the q th pathogen protein in candidate HPI-GEN can be described by the analogous equation as follows:

$$p_q^P(t+1) = p_q^P(t) + \sum_{n=1}^{N_q} a_{qn}^P p_q^P(t) p_n^H(t) + \sum_{v=1}^{V_q} b_{qv}^P p_q^P(t) p_v^P(t) + \alpha_q^P p_q^P(t) - \gamma_q^P p_q^P(t) + \beta_q^P + v_q^P(t), \text{ for } q = 1, 2, \dots, Q, \alpha_q^P \geq 0 \text{ and } -\gamma_q^P \leq 0 \tag{2}$$

where $p_q^P(t)$, $p_n^H(t)$, $g_q^P(t)$ and $p_v^P(t)$ indicate the expression levels of the q th pathogen protein, the n th host protein, the q th pathogen gene, and the v th pathogen protein at time t , respectively; a_{qn}^P and b_{qv}^P denote the interactive ability between the q th pathogen protein and n th host protein and between the q th pathogen protein and v th pathogen protein, respectively; N_q and V_q signify the number of host proteins and pathogen proteins interacting with the q th pathogen protein; and Q typifies the total number of the pathogen proteins in candidate PPIN. α_q^P , $-\gamma_q^P$ and β_q^P indicate the translation rate from the corresponding mRNA, the degradation rate, and the basal expression level of the q th pathogen protein, respectively; the basal level β_q^P denotes the interactions with unknown factors such as acetylation and ubiquitination; and $v_q^P(t)$ represents the stochastic noise of the q th pathogen protein at time t . Similarly, the intraspecies and interspecies biological mechanism of pathogen proteins in candidate PPIN are represented in Equation (2) as the form of $\sum_{n=1}^{N_q} a_{qn}^P p_q^P(t) p_n^H(t)$ and $\sum_{v=1}^{V_q} b_{qv}^P p_q^P(t) p_v^P(t)$, respectively.

For the GRN of host genes in the candidate HPI-GEN, the dynamic regulatory model of the j th host gene can be described as follows:

$$g_j^H(t+1) = g_j^H(t) + \sum_{\tau=1}^{T_j} c_{j\tau}^H p_\tau^H(t) + \sum_{\ell=1}^{L_j} d_{j\ell}^H l_\ell^H(t) - \sum_{\mu=1}^{M_j} e_{j\mu}^H g_j^H(t) m_\mu^H(t) - \lambda_j^H g_j^H(t) + \delta_j^H + \omega_j^H(t), \text{ for } j = 1, 2, \dots, J, -e_{j\mu}^H \leq 0 \text{ and } -\lambda_j^H \leq 0 \tag{3}$$

where $g_j^H(t)$, $p_\tau^H(t)$, $m_\mu^H(t)$ and $l_\ell^H(t)$ indicate the expression levels of the j th host gene, the τ th host TF, the μ th host microRNA, and the ℓ th host lncRNA at time t , respectively; $c_{j\tau}^H$, $-e_{j\mu}^H$ and $d_{j\ell}^H$ represent the regulation ability of the τ th host TF, the μ th host microRNA, and the ℓ th host lncRNA on the j th host gene, respectively; T_j , M_j and L_j denote the number of host TFs, host miRNAs, and host lncRNAs, respectively, which regulate the expression level of the j th host gene; J indicates the total amount of genes in GRN; $-\lambda_j^H$ and δ_j^H indicate the degradation rate and the expression basal level of the j th host gene, respectively; and $\omega_j^H(t)$ denotes the stochastic noise due to modeling residue at time t . The basal level δ_j^H denotes the unknown regulations such as methylation. Note that since microRNAs silence gene expression by binding to target mRNAs [39], repression ability $-e_{j\mu}^H$ in candidate GRN model should be constrained negative. Likewise, $-\lambda_j^H$ degradation rate should be restricted non-positive due to the depression regulation of mRNA. Furthermore, though supporting evidence comes from

numerous research showing that HBV proteins play a vital role in the regulation of viral and host gene expression, different from other viruses such as HIV, whose viral protein directly binds to nucleic acids [40], HBV proteins are thought to affect host gene expression through transregulation, e.g., transactivation and transrepression [41–44]. Thus, it seems tenable to exclude the regulations between pathogen proteins and host gene in the Equation (3).

In the same way, for the GRN of pathogen genes in the candidate HPI-GEN, the dynamic regulatory model of the u th pathogen gene can be described as follows:

$$g_u^P(t+1) = g_u^P(t) + \sum_{\tau=1}^{T_u} c_{u\tau}^P p_{\tau}^H(t) + \sum_{\ell=1}^{L_u} d_{u\ell}^P l_{\ell}^H(t) - \sum_{\mu=1}^{M_u} e_{u\mu}^P g_u^P(t) m_{\mu}^H(t) - \lambda_u^P g_u^P(t) + \delta_u^P + \omega_u^P(t), \text{ for } u = 1, 2, \dots, U, -e_{u\mu}^P \leq 0 \text{ and } -\lambda_u^P \leq 0 \tag{4}$$

where $g_u^P(t)$, $p_{\tau}^H(t)$, $m_{\mu}^H(t)$ and $l_{\ell}^H(t)$ indicate the expression levels of the u th pathogen gene, the τ th host TF, the μ th host microRNA, and the ℓ th host lncRNA at time t , respectively; $c_{u\tau}^P$, $-e_{u\mu}^P$ and $d_{u\ell}^P$ represent the regulation ability of the τ th host TF, the μ th host microRNA, and the ℓ th host lncRNA on the u th pathogen, respectively; T_u , M_u and L_u denote the number of host TFs, host miRNAs, and host lncRNAs, respectively, which regulate the expression level of the u th pathogen; U indicates the total amount of pathogen genes in GRN; $-\lambda_u^P$ and δ_u^P indicate the degradation rate and the expression basal level of the u th pathogen, respectively; and $\omega_u^P(t)$ denotes the stochastic noise due to modeling residue at time t . It is worth noting that there is a consensus on the opinion about the indispensability of host protein synthesis machinery for pathogen protein production and gene transcription [45,46]. Walsh et al. [47] even suggested that viruses are fully reliant on the translation machinery of their host cells to produce the polypeptides that are essential for viral replication. As a result, only gene regulation conducted by host transcription factors $p_{\tau}^H(t)$ instead of pathogen proteins $p^P(t)$ is included in Equation (4).

Since the expression of the μ th host miRNA $m_{\mu}^H(t)$ and the ℓ th host lncRNA $l_{\ell}^H(t)$ in Equation (3) at time t might also be regulated by other regulators, the μ th miRNA can be described by the following stochastic dynamic regulatory equation:

$$m_{\mu}^H(t+1) = m_{\mu}^H(t) + \sum_{\tau=1}^{T_{\mu}} f_{\mu\tau}^H p_{\tau}^H(t) + \sum_{\ell=1}^{L_{\mu}} h_{\mu\ell}^H l_{\ell}^H(t) - \sum_{r=1}^{M_{\mu}} k_{\mu r}^H m_{\mu}^H(t) m_r^H(t) - \varphi_{\mu}^H m_{\mu}^H(t) + \eta_{\mu}^H + \varepsilon_{\mu}^H(t), \text{ for } \mu = 1, 2, \dots, M, -k_{\mu r}^H \leq 0 \text{ and } -\varphi_{\mu}^H \leq 0 \tag{5}$$

where $m_{\mu}^H(t)$, $p_{\tau}^H(t)$, $m_r^H(t)$ and $l_{\ell}^H(t)$ indicate the expression levels of the μ th host gene, the τ th host TF, the r th host miRNA and the ℓ th host lncRNA at time t , respectively; $f_{\mu\tau}^H$, $-k_{\mu r}^H$ and $h_{\mu\ell}^H$ represent the regulation ability of the τ th host TF, the r th host microRNA and the ℓ th host lncRNA on the μ th host miRNA, respectively; T_{μ} , M_{μ} and L_{μ} denote the number of host TFs, host miRNAs and host lncRNAs, respectively, which regulate the expression level of the μ th host miRNA; M indicates the total number of miRNAs in GRN; $-\varphi_{\mu}^H$ and η_{μ}^H indicate the degradation rate and the expression basal level of the μ th host miRNA, respectively; and $\varepsilon_{\mu}^H(t)$ denotes the stochastic noise due to modeling residue at time t .

Along the same line, the dynamic regulatory model of the host lncRNA in candidate GEIN can be described by the following dynamic equation:

$$l_{\ell}^H(t+1) = l_{\ell}^H(t) + \sum_{\tau=1}^{T_{\ell}} x_{\ell\tau}^H p_{\tau}^H(t) + \sum_{o=1}^{O_{\ell}} y_{\ell o}^H l_o^H(t) - \sum_{r=1}^{M_{\ell}} z_{\ell\mu}^H l_{\ell}^H(t) m_{\mu}^H(t) - \sigma_{\ell}^H l_{\ell}^H(t) + \phi_{\ell}^H + \psi_{\ell}^H(t), \text{ for } \ell = 1, 2, \dots, L, -z_{\ell\mu}^H \leq 0 \text{ and } -\sigma_{\ell}^H \leq 0 \tag{6}$$

where $l_{\ell}^H(t)$, $p_{\tau}^H(t)$, $m_{\mu}^H(t)$ and $l_o^H(t)$ indicate the expression levels of the ℓ th host lncRNA, the τ th host TF, the μ th host microRNA and the o th host lncRNA at time t , respectively; $x_{\ell\tau}^H$, $-z_{\ell\mu}^H$ and $y_{\ell o}^H$ represent the regulation ability of the τ th host TF, the μ th host microRNA and the o th host lncRNA on the ℓ th host

lncRNA, respectively; T_ℓ , M_ℓ and L_ℓ denote the number of host TFs, host miRNAs and host lncRNAs, respectively, which regulate the expression level of the ℓ th host lncRNA; L indicates the total number of lncRNA in GRN; $-\sigma_\ell^H$ and ϕ_ℓ^H indicate the degradation rate and the expression basal level of the ℓ th host lncRNA, respectively; and $\psi_\ell^H(t)$ denotes the stochastic noise due to modeling residue at time t .

2.5. Parameter Estimation of the Dynamic Models of Candidate HPI-GEN by System Identification Approach

In order to identify real HPI-GEN, the accurate model parameters training process by executing a system identification approach to the candidate HPI-GEN after the construction of dynamic model Equations (1)–(6) is requisite. Consequently, the stochastic Equations (1) and (2), which depict the relation of protein interactions of the i th host and q th pathogen protein, can be converted into the linear regression forms as below, respectively:

The i th host protein:

$$p_i^H(t+1) = [p_i^H p_1^H(t) \cdots p_i^H p_{N_i}^H(t) p_i^H p_1^P(t) \cdots p_i^H p_{V_i}^P(t) g_i^H(t) p_i^H(t) 1] \begin{bmatrix} a_{i1}^H \\ \vdots \\ a_{iN_i}^H \\ b_{i1}^H \\ \vdots \\ b_{iV_i}^H \\ \alpha_i^H \\ 1 - \gamma_i^H \\ \beta_i^H \end{bmatrix} + v_i^H(t) \tag{7}$$

$$\triangleq \phi_i^{HP}(t) \theta_i^{HP} + v_i^H(t), \text{ for } i = 1, 2, \dots, I$$

The q th host protein:

$$p_q^P(t+1) = [p_q^P p_1^H(t) \cdots p_q^P p_{N_q}^H(t) p_q^P p_1^P(t) \cdots p_q^P p_{V_q}^P(t) g_q^P(t) p_q^P(t) 1] \begin{bmatrix} a_{q1}^P \\ \vdots \\ a_{qN_q}^P \\ b_{q1}^H \\ \vdots \\ b_{qV_q}^P \\ \alpha_q^P \\ 1 - \gamma_q^P \\ \beta_q^P \end{bmatrix} + v_q^P(t) \tag{8}$$

$$\triangleq \phi_q^{PP}(t) \theta_q^{PP} + v_q^P(t), \text{ for } q = 1, 2, \dots, Q$$

where $\phi_i^{HP}(t)$ and $\phi_q^{PP}(t)$ represent the regression vectors that can be obtained from the expression data and θ_i^{HP} and θ_q^{PP} are the unknown parameter vectors to be estimated for the i th host protein, q th pathogen protein in host PPIN, respectively.

Thence, the Equations (7) and (8) of the i th host and q th pathogen protein can further be augmented for Y_i and Y_q time points as the following forms, respectively:

The i th host protein:

$$\begin{bmatrix} p_i^H(t_2) \\ p_i^H(t_3) \\ \vdots \\ p_i^H(t_{Y_i} + 1) \end{bmatrix} = \begin{bmatrix} \phi_i^{HP}(t_1) \\ \phi_i^{HP}(t_2) \\ \vdots \\ \phi_i^{HP}(t_{Y_i}) \end{bmatrix} \theta_i^{HP} + \begin{bmatrix} v_i^H(t_1) \\ v_i^H(t_2) \\ \vdots \\ v_i^H(t_{Y_i}) \end{bmatrix}, \text{ for } i = 1, 2, \dots, I \tag{9}$$

the Equation (9) can also be simply represented as

$$P_i^H = \Phi_i^{HP} \theta_i^{HP} + \Omega_i^{HP}, \text{ for } i = 1, 2, \dots, I \tag{10}$$

The q th host protein:

$$\begin{bmatrix} p_q^P(t_2) \\ p_q^P(t_3) \\ \vdots \\ p_q^P(t_{Y_q} + 1) \end{bmatrix} = \begin{bmatrix} \phi_q^{PP}(t_1) \\ \phi_q^{PP}(t_2) \\ \vdots \\ \phi_q^{PP}(t_{Y_q}) \end{bmatrix} \theta_q^{PP} + \begin{bmatrix} v_q^P(t_1) \\ v_q^P(t_2) \\ \vdots \\ v_q^P(t_{Y_q}) \end{bmatrix}, \text{ for } q = 1, 2, \dots, Q \tag{11}$$

the Equation (11) can also be simply represented as

$$P_q^P = \Phi_q^{PP} \theta_q^{PP} + \Omega_q^{PP}, \text{ for } q = 1, 2, \dots, Q \tag{12}$$

After a succession of equation transformation procedure in advance, the parameters in vectors θ_i^{HP} and θ_q^{PP} of Equations (10) and (12) can be estimated by individually employing the following constrained least-squares estimation problems, θ_i^{HP} parameters estimation:

$$\hat{\theta}_i^{HP} = \min_{\theta_i^{HP}} \|\Phi_i^{HP} \theta_i^{HP} - P_i^H\|_2^2, \text{ subject to } A_i^{HP} \hat{\theta}_i^{HP} \leq b_i^{HP} \tag{13}$$

where $A_i^{HP} = \begin{bmatrix} 0 & \dots & 0 & 0 & \dots & 0 & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{2 \times (I+Q+3)}$ and $b_i^{HP} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

θ_q^{PP} parameters estimation:

$$\hat{\theta}_q^{PP} = \min_{\theta_q^{PP}} \|\Phi_q^{PP} \theta_q^{PP} - P_q^P\|_2^2, \text{ subject to } A_q^{PP} \hat{\theta}_q^{PP} \leq b_q^{PP} \tag{14}$$

where $A_q^{PP} = \begin{bmatrix} 0 & \dots & 0 & 0 & \dots & 0 & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{2 \times (I+Q+3)}$ and $b_q^{PP} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Subsequently, we can acquire the interaction parameters in PPIN Equations (1) and (2) individually by resolving the least-squares problems in (13) and (14) with the help of the function *lsqlin* in MATLAB optimization toolbox and simultaneously ensure the protein translation rate α_i^H and α_q^P to be a non-negative value and the protein degradation rate $-\gamma_i^H$ and $-\gamma_q^P$ to be a non-positive value; that is to say $\alpha_i^H, \alpha_q^P \geq 0$ and $-\gamma_i^H, -\gamma_q^P \leq 0$.

In the same manner, the dynamic Equations (3)–(6) which depict the relationship of gene regulations for the j th host gene, u th pathogen gene, μ th host miRNA, and ℓ th host lncRNA can be rewritten into the linear regression forms below, respectively:

The j th host gene:

$$g_j^H(t+1) = [p_1^H(t) \cdots p_{T_j}^H(t) g_j^H(t) m_1^H(t) \cdots g_j^H(t) m_{M_j}^H(t) l_1^H(t) \cdots l_{L_j}^H(t) g_j^H(t) 1] \begin{bmatrix} c_{j1}^H \\ \vdots \\ c_{jT_j}^H \\ -e_{j1}^H \\ \vdots \\ -e_{jM_j}^H \\ d_{j1}^H \\ \vdots \\ d_{jL_j}^H \\ 1 - \lambda_j^H \\ \delta_j^H \end{bmatrix} + \omega_j^H(t) \tag{15}$$

$$\triangleq \phi_j^{HG}(t) \theta_j^{HG} + \omega_j^H(t), \text{ for } j = 1, 2, \dots, J$$

The u th pathogen gene:

$$g_u^P(t+1) = [p_1^H(t) \cdots p_{T_u}^H(t) g_u^P(t) m_1^H(t) \cdots g_u^P(t) m_{M_u}^H(t) l_1^H(t) \cdots l_{L_u}^H(t) g_u^P(t) 1] \begin{bmatrix} c_{u1}^P \\ \vdots \\ c_{uT_u}^P \\ -e_{u1}^P \\ \vdots \\ -e_{uM_u}^P \\ d_{u1}^P \\ \vdots \\ d_{uL_u}^P \\ 1 - \lambda_u^P \\ \delta_u^P \end{bmatrix} + \omega_u^P(t) \tag{16}$$

$$\triangleq \phi_u^{PG}(t) \theta_u^{PG} + \omega_u^P(t), \text{ for } u = 1, 2, \dots, U$$

The μ th host miRNA:

$$m_\mu^H(t+1) = [p_1^H(t) \cdots p_{T_\mu}^H(t) m_\mu^H(t) m_1^H(t) \cdots m_\mu^H(t) m_{R_\mu}^H(t) l_1^H(t) \cdots l_{L_\mu}^H(t) m_\mu^H(t) 1] \begin{bmatrix} f_{\mu 1}^H \\ \vdots \\ f_{\mu T_\mu}^H \\ -k_{\mu 1}^H \\ \vdots \\ -k_{\mu R_\mu}^H \\ h_{\mu 1}^H \\ \vdots \\ h_{\mu L_\mu}^H \\ 1 - \phi_\mu^H \\ \eta_\mu^H \end{bmatrix} + \varepsilon_\mu^H(t) \tag{17}$$

$$\triangleq \phi_\mu^{HM}(t) \theta_\mu^{HM} + \varepsilon_\mu^H(t), \text{ for } \mu = 1, 2, \dots, M$$

The ℓ th host lncRNA:

$$l_\ell^H(t+1) = [p_1^H(t) \cdots p_{T_\ell}^H(t) l_\ell^H(t) m_1^H(t) \cdots l_\ell^H(t) m_{M_\ell}^H(t) l_1^H(t) \cdots l_{L_\ell}^H(t) l_\ell^H(t) 1] \begin{bmatrix} x_{t_1}^H \\ \vdots \\ x_{t_{T_\ell}}^H \\ -z_{t_1}^H \\ \vdots \\ -z_{t_{M_\ell}}^H \\ y_{t_1}^H \\ \vdots \\ y_{t_{L_\ell}}^H \\ 1 - \sigma_\ell^H \\ \phi_\ell^H \end{bmatrix} + \psi_\ell^H(t) \tag{18}$$

$$\triangleq \phi_\ell^{HL}(t) \theta_\ell^{HL} + \psi_\ell^H(t), \text{ for } \ell = 1, 2, \dots, L$$

where $\phi_j^{HG}(t)$, $\phi_u^{PG}(t)$, $\phi_\mu^{HM}(t)$ and $\phi_\ell^{HL}(t)$ represent the regression vectors that can be obtained from the expression data and θ_j^{HG} , θ_u^{PG} , θ_μ^{HM} and θ_ℓ^{HL} are the unknown parameter vectors to be estimated for the j th host gene, u th pathogen gene, μ th host miRNA, and ℓ th host lncRNA in host GRN, respectively.

Then, Equations (15)–(18) of the j th host gene, u th pathogen gene, μ th host miRNA, and ℓ th host lncRNA can be further augmented for Y_j , Y_u , Y_μ and Y_ℓ time points as the following forms, respectively:

The j th host gene:

$$\begin{bmatrix} g_j^H(t_2) \\ g_j^H(t_3) \\ \vdots \\ g_j^H(t_{Y_j} + 1) \end{bmatrix} = \begin{bmatrix} \phi_j^{HG}(t_1) \\ \phi_j^{HG}(t_2) \\ \vdots \\ \phi_j^{HG}(t_{Y_j}) \end{bmatrix} \theta_j^{HG} + \begin{bmatrix} \omega_j^H(t_1) \\ \omega_j^H(t_2) \\ \vdots \\ \omega_j^H(t_{Y_j}) \end{bmatrix}, \text{ for } j = 1, 2, \dots, J \tag{19}$$

the Equation (19) can also be simply represented as:

$$G_j^H = \Phi_j^{HG} \theta_j^{HG} + \Omega_j^{HG}, \text{ for } j = 1, 2, \dots, J \tag{20}$$

The u th pathogen gene:

$$\begin{bmatrix} g_u^P(t_2) \\ g_u^P(t_3) \\ \vdots \\ g_u^P(t_{Y_u} + 1) \end{bmatrix} = \begin{bmatrix} \phi_u^{PG}(t_1) \\ \phi_u^{PG}(t_2) \\ \vdots \\ \phi_u^{PG}(t_{Y_u}) \end{bmatrix} \theta_u^{PG} + \begin{bmatrix} \omega_u^P(t_1) \\ \omega_u^P(t_2) \\ \vdots \\ \omega_u^P(t_{Y_u}) \end{bmatrix}, \text{ for } u = 1, 2, \dots, U \tag{21}$$

the Equation (21) can also be simply represented as:

$$G_u^P = \Phi_u^{PG} \theta_u^{PG} + \Omega_u^{PG}, \text{ for } u = 1, 2, \dots, U \tag{22}$$

The μ th host miRNA:

$$\begin{bmatrix} m_\mu^H(t_2) \\ m_\mu^H(t_3) \\ \vdots \\ m_\mu^H(t_{Y_\mu} + 1) \end{bmatrix} = \begin{bmatrix} \phi_\mu^{HM}(t_1) \\ \phi_\mu^{HM}(t_2) \\ \vdots \\ \phi_\mu^{HM}(t_{Y_\mu}) \end{bmatrix} \theta_\mu^{HM} + \begin{bmatrix} \varepsilon_j^H(t_1) \\ \varepsilon_j^H(t_2) \\ \vdots \\ \varepsilon_j^H(t_{Y_\mu}) \end{bmatrix}, \text{ for } \mu = 1, 2, \dots, M \tag{23}$$

the Equation (23) can also be simply represented as:

$$M_\mu^H = \Phi_\mu^{HM} \theta_\mu^{HM} + \Omega_\mu^{HM}, \text{ for } \mu = 1, 2, \dots, M \tag{24}$$

The ℓ th host lncRNA:

$$\begin{bmatrix} I_\ell^H(t_2) \\ I_\ell^H(t_3) \\ \vdots \\ I_\ell^H(t_{Y_\ell} + 1) \end{bmatrix} = \begin{bmatrix} \phi_\ell^{HL}(t_1) \\ \phi_\ell^{HL}(t_2) \\ \vdots \\ \phi_\ell^{HL}(t_{Y_\ell}) \end{bmatrix} \theta_j^{HG} + \begin{bmatrix} \psi_\ell^H(t_1) \\ \psi_\ell^H(t_2) \\ \vdots \\ \psi_\ell^H(t_{Y_\ell}) \end{bmatrix}, \text{ for } \ell = 1, 2, \dots, L \tag{25}$$

the Equation (25) can also be simply represented as:

$$L_\ell^H = \Phi_\ell^{HG} \theta_\ell^{HG} + \Omega_\ell^{HG}, \text{ for } \ell = 1, 2, \dots, L \tag{26}$$

After a succession of equation transformation procedure in advance, the parameters in vectors θ_j^{HG} , θ_u^{PG} , θ_μ^{HM} and θ_ℓ^{HL} of Equations (20), (22), (24) and (26) can be estimated by individually employing the following constrained least-square estimation problem,

$$\hat{\theta}_j^{HG} = \min_{\theta_j^{HG}} \frac{1}{2} \|\Phi_j^{HG} \theta_j^{HG} - G_j^H\|_2^2, \text{ subject to } A_j^{HG} \hat{\theta}_j^{HG} \leq b_j^{HG} \text{ where } b_j^{HG} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\text{and } A_j^{HG} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{(M_j+1) \times (T_j+M_j+L_j+2)} \tag{27}$$

$$\hat{\theta}_u^{PG} = \min_{\theta_u^{PG}} \frac{1}{2} \|\Phi_u^{PG} \theta_u^{PG} - G_u^P\|_2^2, \text{ subject to } A_u^{PG} \hat{\theta}_u^{PG} \leq b_u^{PG} \text{ where } b_u^{PG} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\text{and } A_u^{PG} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{(M_u+1) \times (T_u+M_u+L_u+2)} \tag{28}$$

$$\hat{\theta}_\mu^{HM} = \min_{\theta_\mu^{HM}} \frac{1}{2} \|\Phi_\mu^{HM} \theta_\mu^{HM} - M_\mu^H\|_2^2, \text{ subject to } A_\mu^{HM} \hat{\theta}_\mu^{HM} \leq b_\mu^{HM} \text{ where } b_\mu^{HM} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\text{and } A_\mu^{HM} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{(M_\mu+1) \times (T_\mu+M_\mu+L_\mu+2)} \tag{29}$$

$$\hat{\theta}_\ell^{HL} = \min_{\theta_\ell^{HL}} \frac{1}{2} \|\Phi_\ell^{HL} \theta_\ell^{HL} - L_\ell^H\|_2^2, \text{ subject to } A_\ell^{HL} \hat{\theta}_\ell^{HL} \leq b_\ell^{HL} \text{ where } b_\ell^{HL} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\text{and } A_\ell^{HL} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{(M_\ell+1) \times (T_\ell+M_\ell+L_\ell+2)} \tag{30}$$

Through applying the function *lsqlin* in MATLAB optimization toolbox to solve the constrained least-squares estimation problem in (27)–(30), we could obtain optimal regulatory parameters for GRN Equations in (3)–(6) and concurrently guarantee that the miRNA repression ability $-e_{j\mu}^H$, $-e_{u\mu}^P$, $-k_{\mu r}^H$ and $-z_{\ell\mu}^H$ as well as the degradation rate $-\lambda_j^H$, $-\lambda_u^P$, $-\varphi_\mu^H$ and $-\sigma_\ell^H$ values corresponding to the *j*th host gene, the *u*th pathogen gene, the *μ*th host miRNA, and the *ℓ*th host lncRNA to be non-positive, respectively.

So far, the structure of dynamic regression model construction is complete, yet if we substitute the expression data into the linear model right away, infinite solutions problems will emerge due to the lack of enough data points, in other words, information deficiency in comparison with parameters for evaluation, that is, overfitting. Thus, the cubic spline for extra numbers of time-profile data points interpolation is applied to prevent this trouble (For more details, readers can refer to Appendix A). Then, with the processed RNA-seq data, the accurate solutions of the constrained least-squares estimation problems in (13)–(14) and (27)–(30) can be attained.

Moreover, since the measurement technology of genome-wide protein expression in HepaRG cells and Hepatitis B Virus during infection hasn't been realized yet, and supporting evidence comes from research showing that the cellular concentrations of proteins correlate with the abundances of their corresponding mRNAs, which implied that the variance in protein abundance can be explained by that of mRNA [48–51] and a correlation coefficient of 48% was also obtained between the mRNA and protein abundances in human liver [49], i.e., the RNA-seq data of gene expressions can substitute protein expressions and contribute to sufficient information for resolving above constrained least-squares parameter estimation problems in (13) and (14) and (27)–(30). One could retain more information in our previous research [52].

2.6. Determination of Significant Interaction Pairs

On account of approaching the best combination to constitute the robust HPI-GEN in the parameter identification procedures, we exploited the cost function to obtain the optimal fit for the host/pathogen RNA-seq data, to be more precisely, estimation of the expected relative distance between the fitted model and the unrecognized authentic molecular pathogenic mechanism that actually generated the observed data. On the whole, as the cost function of the parameter estimation for linear regression model, mean squared error (MSE) is enough to calculate the residual variance. However, the cost from the model complexity that might also influences the performance should be taken into consideration as well. Akaike Information Criterion (AIC), taking the place of MSE, was thereby selected to assess the residual variance and model complexity at once. As the expected residual variance declines with rising parameter numbers for inadequate model complexities, there should be a minimum around the correct parameter number after repeated coordination [52]. Meanwhile, owing to computational efficiency, computing the AIC statistics for all possible combinations is impracticable. Stepwise methods are developed to decreasing the complexity of exhaustive searching [52]. Finally, while reaching the minimum of AIC value, we could acquire the real number of interactions or regulations for each protein (gene) one by one in the candidate HPI-GEN. Those insignificant regulations or interactions out of the real number determined by AIC should be pruned to obtain the real HPI-GEN.

For each model composing PPIN, the AIC values of the *i*th host and the *q*th pathogen protein can be defined individually as the following equations:

The *i*th host protein:

$$\begin{aligned} AIC_i^{HP}(N_i, V_i) &= \log\left(\frac{1}{Y_i}(P_i^H - \Phi_i^{HP} \hat{\theta}_i^{HP})^T (P_i^H - \Phi_i^{HP} \hat{\theta}_i^{HP})\right) + \frac{2(N_i+V_i)}{Y_i} \\ &= \log\left[(\sigma_i^{HP})^2\right] + \frac{2(N_i+V_i)}{Y_i} \end{aligned} \tag{31}$$

The q th pathogen protein:

$$\begin{aligned} AIC_q^{PP}(N_q, V_q) &= \log\left(\frac{1}{Y_q}(P_q^P - \Phi_q^{PP}\hat{\theta}_q^{PP})^T(P_q^P - \Phi_q^{PP}\hat{\theta}_q^{PP})\right) + \frac{2(N_q+V_q)}{Y_q} \\ &= \log\left[(\sigma_q^{PP})^2\right] + \frac{2(N_q+V_q)}{Y_q} \end{aligned} \tag{32}$$

for $i = 1, \dots, I$ and $q = 1, \dots, Q$, where $AIC_i^{HP}(N_i, V_i)$ with the model complexity $N_i + V_i$ and $AIC_q^{PP}(N_q, V_q)$ with the model complexity $N_q + V_q$ denote the i th host and the q th pathogen protein in PPI model, respectively; $(\sigma_i^{HP})^2$ and $(\sigma_q^{PP})^2$ signify the covariance of estimated residual error between P_i^H and $\Phi_i^{HP}\hat{\theta}_i^{HP}$ and between P_q^P and $\Phi_q^{PP}\hat{\theta}_q^{PP}$, respectively; and $\hat{\theta}_i^{HP}$ and $\hat{\theta}_q^{PP}$ represent the estimated parameters acquired from the solutions of the parameter estimation problems in (13) and (14), respectively. Suppose N_i^*, V_i^* and N_q^*, V_q^* could minimize $AIC_i^{HP}(N_i, V_i)$ and $AIC_q^{PP}(N_q, V_q)$ in (31) and (32), respectively. Then, N_i^*, V_i^* and N_q^*, V_q^* denote the real number of PPIs in the i th host protein and the q th pathogen protein. The insignificant PPIs out of these numbers should be pruned as false positives from candidate PPIs to obtain real PPIs in HPI-GEN.

For each model comprising GRN, the AIC values of the j th host gene, the u th pathogen gene, the μ th host miRNA, and the ℓ th host lncRNA can be defined individually as following equations:

The j th host gene:

$$\begin{aligned} AIC_j^{HG}(T_j + M_j + L_j) &= \log\left(\frac{1}{Y_j}(G_j^H - \Phi_j^{HG}\hat{\theta}_j^{HG})^T(G_j^H - \Phi_j^{HG}\hat{\theta}_j^{HG})\right) + \frac{2(T_j+M_j+L_j)}{Y_j} \\ &= \log\left[(\sigma_j^{HG})^2\right] + \frac{2(T_j+M_j+L_j)}{Y_j} \end{aligned} \tag{33}$$

The u th pathogen gene:

$$\begin{aligned} AIC_u^{PG}(T_u + M_u + L_u) &= \log\left(\frac{1}{Y_u}(G_u^P - \Phi_u^{PG}\hat{\theta}_u^{PG})^T(G_u^P - \Phi_u^{PG}\hat{\theta}_u^{PG})\right) + \frac{2(T_u+M_u+L_u)}{Y_u} \\ &= \log\left[(\sigma_u^{PG})^2\right] + \frac{2(T_u+M_u+L_u)}{Y_u} \end{aligned} \tag{34}$$

The μ th host miRNA:

$$\begin{aligned} AIC_\mu^{HM}(T_\mu + M_\mu + L_\mu) &= \log\left(\frac{1}{Y_\mu}(M_\mu^H - \Phi_\mu^{HM}\hat{\theta}_\mu^{HM})^T(M_\mu^H - \Phi_\mu^{HM}\hat{\theta}_\mu^{HM})\right) + \frac{2(T_\mu+M_\mu+L_\mu)}{Y_\mu} \\ &= \log\left[(\sigma_\mu^{HM})^2\right] + \frac{2(T_\mu+M_\mu+L_\mu)}{Y_\mu} \end{aligned} \tag{35}$$

The ℓ th host lncRNA:

$$\begin{aligned} AIC_\ell^{HL}(T_\ell + M_\ell + L_\ell) &= \log\left(\frac{1}{Y_\ell}(L_\ell^H - \Phi_\ell^{HL}\hat{\theta}_\ell^{HL})^T(L_\ell^H - \Phi_\ell^{HL}\hat{\theta}_\ell^{HL})\right) + \frac{2(T_\ell+M_\ell+L_\ell)}{Y_\ell} \\ &= \log\left[(\sigma_\ell^{HL})^2\right] + \frac{2(T_\ell+M_\ell+L_\ell)}{Y_\ell} \end{aligned} \tag{36}$$

for $j = 1, \dots, J$, $q = 1, \dots, Q$, $\mu = 1, \dots, M$, and $\ell = 1, \dots, L$ where $AIC_j^{HG}(T_j + M_j + L_j)$, $AIC_u^{PG}(T_u + M_u + L_u)$, $AIC_\mu^{HM}(T_\mu + M_\mu + L_\mu)$ and $AIC_\ell^{HL}(T_\ell + M_\ell + L_\ell)$ with the model complexity $T_j + M_j + L_j$, $T_u + M_u + L_u$, $T_\mu + M_\mu + L_\mu$ and $T_\ell + M_\ell + L_\ell$ denote the AIC values of the j th host gene, the u th pathogen gene, the μ th host miRNA, and the ℓ th host lncRNA in GRN model, respectively; $(\sigma_j^{HG})^2$, $(\sigma_u^{PG})^2$, $(\sigma_\mu^{HM})^2$ and $(\sigma_\ell^{HL})^2$ signify the covariance of estimated residual error between G_j^H and $\Phi_j^{HG}\hat{\theta}_j^{HG}$, G_u^P and $\Phi_u^{PG}\hat{\theta}_u^{PG}$, M_μ^H and $\Phi_\mu^{HM}\hat{\theta}_\mu^{HM}$, and L_ℓ^H and $\Phi_\ell^{HL}\hat{\theta}_\ell^{HL}$, respectively; and $\hat{\theta}_j^{HG}$, $\hat{\theta}_u^{PG}$, $\hat{\theta}_\mu^{HM}$ and $\hat{\theta}_\ell^{HL}$ represent the estimated parameters acquired from the solutions of the parameter estimation problems in (27)–(30), respectively. Similarly, the real numbers of regulation by TF, miRNA, lncRNA are obtained by minimizing the corresponding AICs in (33)–(36).

The insignificant regulations out of these real numbers should be pruned as false positives to obtain the real regulations by TF, miRNA, lncRNA in HPI-GEN.

So far, through a sequence of systematic identification processes, the construction for real HPI-GENs is broadly accomplished by pruning those insignificant interactions and regulations out of the corresponding AIC. Unfortunately, due to the excessively enormous network architecture, we could hardly extract the core pathways to investigate the pathogenesis during HBV infection, let alone screening of candidate proteins as new targets for drug development. We thereby extracted the core network from the real HPI-GEN via applying the principal network projection (PNP) method.

2.7. Extracting Core Network Structure from the Real HPI-GEN by Using PNP Approach

Prior to core network extraction from the real HPI-GEN through applying PNP method, it is essential to integrate the network estimated parameters from PPIN and GRN into a system matrix A as follows,

$$A = \begin{pmatrix} A_{HP,HP} & A_{HP,PP} & 0 & 0 \\ A_{PP,HP} & A_{PP,PP} & 0 & 0 \\ A_{HG,HP} & 0 & A_{HG,HM} & A_{HG,HL} \\ A_{HM,HP} & 0 & A_{HM,HM} & A_{HM,HL} \\ A_{HL,HP} & 0 & A_{HL,HM} & A_{HL,HL} \\ A_{PG,HP} & 0 & A_{PG,HM} & A_{PG,HL} \\ \hat{a}_{11}^H & \dots & \hat{a}_{1l}^H & \hat{b}_{11}^H & \dots & \hat{b}_{1Q}^H & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \hat{a}_{in}^H & \vdots & \vdots & \hat{b}_{iv}^H & \vdots & \vdots & 0 & \vdots & \vdots & \vdots & 0 & \vdots \\ \hat{a}_{1l}^H & \dots & \hat{a}_{ll}^H & \hat{b}_{1l}^H & \dots & \hat{b}_{lQ}^H & 0 & \dots & 0 & 0 & \dots & 0 \\ \hat{a}_{11}^P & \dots & \hat{a}_{1l}^P & \hat{b}_{11}^P & \dots & \hat{b}_{1Q}^P & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \hat{a}_{qn}^P & \vdots & \vdots & \hat{b}_{qv}^P & \vdots & \vdots & 0 & \vdots & \vdots & \vdots & 0 & \vdots \\ \hat{a}_{Q1}^P & \dots & \hat{a}_{Ql}^P & \hat{b}_{Q1}^P & \dots & \hat{b}_{QQ}^P & 0 & \dots & 0 & 0 & \dots & 0 \\ \hat{c}_{11}^H & \dots & \hat{c}_{1l}^H & 0 & \dots & 0 & -\hat{e}_{11}^H & \dots & -\hat{e}_{1M}^H & \hat{d}_{11}^H & \dots & \hat{d}_{1L}^H \\ \vdots & \hat{c}_{j\tau}^H & \vdots & \vdots & 0 & \vdots & \vdots & -\hat{e}_{j\mu}^H & \vdots & \vdots & \hat{d}_{j\ell}^H & \vdots \\ \hat{c}_{1l}^H & \dots & \hat{c}_{ll}^H & 0 & \dots & 0 & -\hat{e}_{1l}^H & \dots & -\hat{e}_{lM}^H & \hat{d}_{1l}^H & \dots & \hat{d}_{lL}^H \\ \hat{f}_{11}^H & \dots & \hat{f}_{1l}^H & 0 & \dots & 0 & -k_{11}^H & \dots & -k_{1M}^H & \hat{h}_{11}^H & \dots & \hat{h}_{1L}^H \\ \vdots & \hat{f}_{\mu\tau}^H & \vdots & \vdots & 0 & \vdots & \vdots & -k_{\mu r}^H & \vdots & \vdots & \hat{h}_{\mu\ell}^H & \vdots \\ \hat{f}_{M1}^H & \dots & \hat{f}_{Ml}^H & 0 & \dots & 0 & -k_{M1}^H & \dots & -k_{MM}^H & \hat{h}_{M1}^H & \dots & \hat{h}_{ML}^H \\ \hat{x}_{11}^H & \dots & \hat{x}_{1l}^H & 0 & \dots & 0 & -z_{11}^H & \dots & -z_{1M}^H & \hat{y}_{11}^H & \dots & \hat{y}_{1L}^H \\ \vdots & \hat{x}_{\ell\tau}^H & \vdots & \vdots & 0 & \vdots & \vdots & -z_{\ell\mu}^H & \vdots & \vdots & \hat{y}_{\ell o}^H & \vdots \\ \hat{x}_{l1}^H & \dots & \hat{x}_{ll}^H & 0 & \dots & 0 & -z_{l1}^H & \dots & -z_{lM}^H & \hat{y}_{l1}^H & \dots & \hat{y}_{lL}^H \\ \hat{c}_{11}^P & \dots & \hat{c}_{1l}^P & 0 & \dots & 0 & -\hat{e}_{11}^P & \dots & -\hat{e}_{1M}^P & \hat{d}_{11}^P & \dots & \hat{d}_{1L}^P \\ \vdots & \hat{c}_{u\tau}^P & \vdots & \vdots & 0 & \vdots & \vdots & -\hat{e}_{u\mu}^P & \vdots & \vdots & \hat{d}_{u\ell}^P & \vdots \\ \hat{c}_{Q1}^P & \dots & \hat{c}_{Ql}^P & 0 & \dots & 0 & -\hat{e}_{Q1}^P & \dots & -\hat{e}_{QM}^P & \hat{d}_{Q1}^P & \dots & \hat{d}_{QL}^P \end{pmatrix} \in \mathbb{R}^{(2I+2Q+L+M) \times (I+Q+L+M)}$$

where HP, PP, HG, HM, HL and PG denote the host protein, pathogen protein, host gene, host miRNA, host lncRNA and pathogen gene, respectively; \hat{a}_{in}^H , \hat{b}_{iv}^H and \hat{a}_{qn}^P , \hat{b}_{qv}^P mentioned in (1) and (2) could be acquired in $\hat{\theta}_i^{HP}$ and $\hat{\theta}_q^{PP}$ by resolving the parameter estimation problems in (13) and (14) and pruning false positives by AIC method in (31) and (32), respectively; $\{\hat{c}_{j\tau}^H, -\hat{e}_{j\mu}^H, \hat{d}_{j\ell}^H\}$, $\{\hat{f}_{\mu\tau}^H, -k_{\mu r}^H, \hat{h}_{\mu\ell}^H\}$, $\{\hat{x}_{\ell\tau}^H, -z_{\ell\mu}^H, \hat{y}_{\ell o}^H\}$, and $\{\hat{c}_{u\tau}^P, -\hat{e}_{u\mu}^P, \hat{d}_{u\ell}^P\}$ mentioned in (3)–(6) could be acquired in $\hat{\theta}_j^{HG}$, $\hat{\theta}_u^{PG}$, $\hat{\theta}_\mu^{HM}$ and $\hat{\theta}_\ell^{HL}$ by resolving the parameter estimation problems in (27)–(30) and pruning false positives by AIC method in (33)–(36) respectively. Note that since the regulations from pathogen proteins weren't taken into consideration, the corresponding parameters in matrix A are padded with zeros.

Thereafter, we extract the core components of HPI-GEN by PNP approach, a principal network analysis method for dimensionality reduction based on network structure projection technique, that is,

projecting matrix A to its principal singular vectors space. Accordingly, the combined network matrix A can be denoted by singular value decomposition form below,

$$A = USV^T \tag{37}$$

where $U \in \mathbb{R}^{(2I+2Q+L+M) \times (I+Q+L+M)}$, $V \in \mathbb{R}^{(I+Q+L+M) \times (I+Q+L+M)}$, and $S = \text{diag}(\sigma_1, \dots, \sigma_s, \dots, \sigma_{I+Q+L+M})$ is a diagonal matrix of $\sigma_1, \dots, \sigma_s, \dots, \sigma_{I+Q+L+M}$ which includes the $I + Q + L + M$ singular values of the matrix A in descending order, i.e., $\sigma_1 \geq \dots \geq \sigma_s \geq \dots \geq \sigma_{I+Q+L+M} \geq 0$.

$$S = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_s & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \sigma_{I+Q+L+M} \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}$$

Also, we define the expression fraction E_w of the interaction ability of the proteins and the transcriptional regulatory ability of the genes, lncRNAs, miRNAs as the normalization of singular value form,

$$E_w = \frac{\sigma_w^2}{\sum_{w=1}^{I+Q+L+M} \sigma_w^2} \tag{38}$$

According to the diminishing property of the singular values, under the basic premise of sustaining system energy from the whole network structure, we pick out the minimum X such that $\sum_{w=1}^X E_w \geq 0.85$, i.e., the vector space spanned by the orthonormal bases composed of the principal singular vectors corresponding to the top X singular values contains 85% energy of HPI-GEN structure.

Afterwards, we define and apply the projection value of each node, including gene, miRNA, lncRNA, and protein in the real HPI-GEN to the vector space spanned by the orthonormal basis composed of the principal singular vectors corresponding to the top X singular values as below,

$$\begin{aligned} \text{Proj}_R(A_{row,i}, V^{T*}) &= \left[\sum_{k=1}^X (A_{row,i} v_k)^2 \right]^{1/2} \\ \text{Proj}_C(A_{col,j}, U^*) &= \left[\sum_{k=1}^X (A_{col,j}^T u_k)^2 \right]^{1/2} \end{aligned} \tag{39}$$

where $A_{row,i}$ denotes the i th row vector of A for $i = 1, \dots, 2I + 2Q + L + M$; $A_{col,j}$ signifies j th column vector of A for $j = 1, \dots, I + Q + L + M$; V^{T*} ; and U^* represent the vector spaces spanned by the orthonormal basis $\{v_1, v_2, \dots, v_k\}$ and $\{u_1, u_2, \dots, u_k\}$ for $k = 1, \dots, X$, respectively; $\text{Proj}_R(A_{row,i}, V^{T*})$ and $\text{Proj}_C(A_{col,j}, U^*)$ indicate the projection value from $A_{row,i}$ and $A_{col,j}$ to V^{T*} and U^* vector spaces, respectively.

While the projection value $\text{Proj}_R(A_{row,i}, V^{T*})$ or $\text{Proj}_C(A_{col,j}, U^*)$ in (39) approaches zero, it intimates that the corresponding node i or j isn't the pivotal factor or independent to the core network extracted via the PNP procedure. Conversely, the larger projection value it gets, the greater the node contributes to the core network.

Eventually, we select the first 3000 components on the top of the list with higher energy in (39) as the core host-pathogen network from real HPI-GEN of HBV infection by ranking the projection value of each node and subsequently uploading those nodes into DAVID to obtain the KEGG pathways as

shown in Table 3. Taking KEGG pathways as reference, we systematically identify and investigate the molecular pathogenic mechanism of HBV infection from the core HPI-GEN as shown in Figure 4 and select the significant biomarkers for the further novel drug discovery procedure.

2.8. Deep Learning-Based Drug-Target Interaction Prediction

2.8.1. Data Preparation

Following the identification of promising pathogenic biomarkers, a potent drug-target interaction (DTI) model to predict the interactions between the identified biomarkers and their corresponding molecule drugs is crucial for discovering favorable molecular compounds for the therapeutic treatment. Thus, we constructed a deep learning-based DTI model to collect viable interactions to identify available drugs for the selected biomarkers. Meanwhile, since most patients in need of attention and treatment are infected with chronic hepatitis, the highest priority was assigned to efficacious drugs with lower toxicity that would not cause irreparable harm to health.

First, we integrated databases from UniProt [53], DrugBank [54], ChEMBL [55], Pubchem [56], and BindingDB [57] to assemble applicable drug-target interactions. As feature descriptors are simple numerical vectors designed to delineate the complicated information of objects, they have been widely utilized to describe the genomic sequences and chemical properties of molecules such as the characteristics from 2D, 3D spectrum of structure, molecular weight, predictive values of LogP, etc. of late [58,59]. In view of this property, we employed both the off-the-shelf build-in molecular and protein descriptor functions of python package pyBioMed to transform the features from each drug and its target among the previously collected drug-target interactions into descriptor under python 2.7 environment, respectively. These features include molecular dockings such as amino acid composition and dipeptide composition. For more details about the descriptor transformation, readers could access the documents of pyBioMed. Thereafter, we united the descriptors of both drugs and their relevant targets into a matrix as the training dataset to train the DTI model. The vector $v_{drug-target}$ corresponding to the descriptor of each drug-target pair is given below:

$$v_{drug-target} = [D, T] = [d_1, d_2, \dots, d_l, t_1, t_2, \dots, t_j] \quad (40)$$

where $v_{drug-target}$ denotes the vector of the descriptor for each the drug-target pair comprising two parts. The former part D is the descriptor of the drug and d_i represents the i th drug feature; the latter portion T is the descriptor of the target and t_j indicates the j th target characteristic; l is the total number of drug features; j is the total number of target features.

Before performing drug-target interaction prediction, to keep our DTI model from inferior performance arising from between-class imbalance and distracting variation, some data preprocessing procedures are requisite. The entire preprocessing procedure in our work encompasses down sampling, data partitioning, feature scaling, and dimensionality reduction.

- Down Sampling:

There are two categories of drug-target interaction in our samples: 16,000,000 samples for the unknown interactions (negative instances) and 60,000 samples for the known interactions (positive instances). Though deep learning methods scale well with the quantity of data and can often leverage extremely large datasets for good performance, imbalanced class distribution often causes the model to be overwhelmed by the large class and ignore the minority one. So as to mitigate the effect arising from class imbalance, we reduced the number of majority samples (negative instances) to 60,000 which will allow the model to learn from both classes equally.

- Data Partitioning:

After the down sampling procedure, we partition the data (the reconstructed matrix of drug-target pairs) into two sets, three-fourth for training and one-fourth for testing.

- Feature Scaling:

Since the variables of the features in each drug-target pair are measured at different scales, they do not contribute equally to the model fitting and might end up with creating a bias. To deal with this potential problem, a feature-wise scaling is usually implemented prior to model fitting. As powerful techniques of feature scaling, Min-max scaling (normalization) and Standardization are both widely used in data analysis. However, despite possessing the capability to shift and rescale the data into a limited range of values, normalization is sensitive to the outliers [60]; in contrast, standardization maintains useful information about outliers and makes the model less sensitive to them [61]. Thus, we apply Standardization on each feature and the corresponding mathematical formulation is shown as follows:

$$\begin{aligned} d_i^* &= \frac{d_i - \mu_i}{\sigma_i}, \forall i = 1, \dots, I \\ t_j^* &= \frac{t_j - \omega_j}{\delta_j}, \forall j = 1, \dots, J \end{aligned} \quad (41)$$

where d_i and d_i^* represent the i th feature of the drug before and after Standardization, respectively; μ_i and σ_i signify the mean and standard deviation of the i th drug feature; t_j and t_j^* indicate the j th feature of the target before and after Standardization, respectively; ω_j and δ_j denote the mean and standard deviation of the j th target feature; I is the total number of features in the drug while J is that of the target.

- Dimensionality Reduction:

Since the high-dimensional patterns in samples might augment the number of neurons in the network and elevate the computational complexity for training, we adopted principal components analysis (PCA) to distill 692 features from original 1014 features of the samples. One could retain as much information of interested [62,63].

Note that it is illegal and useless to carry out data preprocessing to the testing set solely relying on the characteristics in testing data in that we would never get any information from the features of testing data until the training phase is over. Hence, aside from the aforementioned transformation performed on the training data sets, we also provide the identical transformation on the testing data during the process of feature scaling and dimensionality reduction. More exactly, we extracted the mean and standard deviation variables of each feature in the training data set to standardize the corresponding feature in the testing data. Likewise, the transformation matrix for executing PCA of the features in training data was concurrently used to reduce the dimension of the features in testing data.

So far, the training data for tuning the network parameters of the deep learning-based DTI model and the testing data for evaluating the model performance are ready. Nonetheless, if learned merely on the features from the training set, the hyperparameters would always choose the maximum possible model capacity, leading to overfitting [63]. As a result, we randomly split out one-tenth of the training data to estimate the generalization error during training, allowing for the hyperparameters to be updated in every epoch. In addition, Early Stopping method was also employed to specify an arbitrarily large number of training epochs and stop training once improving the model's fit to the training data comes at the expense of increased generalization error. Eventually, through setting the optimizer as Adam [64] and learning rate = 0.001, we then trained our DTI model for 40 epochs with 100 samples in each mini-batch.

2.8.2. Parameters Tuning Process Based on Deep Learning Algorithm

In the network architecture, each layer can be simplified into a function below:

$$\hat{h}_n = \delta(\mathbf{w}^T x_n + \mathbf{b}) \quad (42)$$

where x_n and \hat{h}_n indicate the input and output vectors corresponding to the descriptor of the n th drug-target pair, respectively; δ denotes the activation function (ReLU for each hidden layer and Sigmoid for the output layer); and the vectors w and b are given as follows:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_e \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_e \end{bmatrix} \quad (43)$$

where the weights parameters w_1, w_2, \dots, w_e in w and bias parameters b_1, b_2, \dots, b_e in b are free variables that capture the model's representation of the data in each layer and are learned from each sample.

To evaluate the model performance, the predicted output is compared with the true label to compute a loss for the current set of model weights. Since drug-target interaction prediction issue is a binary classification problem, binary cross-entropy is chosen to calculate the loss of each iteration. The summation of the loss $L(y, \hat{y})$ for totally N samples is given by the expression as below:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^N (y_n * \log(\hat{y}_n) + (1 - y_n) * \log(1 - \hat{y}_n)) = \frac{1}{N} \sum_{n=1}^N C^n(y_n, \hat{y}_n) \quad (44)$$

where $C^n(y_n, \hat{y}_n)$ is loss of the n th sample calculated by binary cross-entropy, in which y_n is the n th actual label (1 or 0) and \hat{y}_n is the n th predicted probability distribution.

During the learning process, the parameter set θ (46) in each layer should achieve the minimization of the objective function (47) to obtain the optimal network parameters θ^* . Accordingly, backward propagation learning algorithm (48) is designed to update the network parameters of both weights (w_1, w_2, \dots, w_e) and bias (b_1, b_2, \dots, b_e).

$$\theta = \begin{bmatrix} w_1 \\ \vdots \\ w_e \\ b_1 \\ \vdots \\ b_e \end{bmatrix} \quad (45)$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} L(\theta) \quad (46)$$

$$\theta^i = \theta^{i-1} - \eta \nabla L(\theta^{i-1}) \quad (47)$$

where i is the i th iteration of the learning process; η is the learning rate (set as 0.001); and $\nabla L(\theta^{i-1})$ is the gradient of $L(\theta^{i-1})$ given as follows:

$$\nabla L(\theta^{i-1}) = \begin{bmatrix} \frac{\partial L(\theta^{i-1})}{\partial w_1} \\ \vdots \\ \frac{\partial L(\theta^{i-1})}{\partial w_e} \\ \frac{\partial L(\theta^{i-1})}{\partial b_1} \\ \vdots \\ \frac{\partial L(\theta^{i-1})}{\partial b_e} \end{bmatrix} \quad (48)$$

Consequently, by back-propagating a corrective error signal through the network, weighted connections between neurons in the DTI model are iteratively adjusted and assessed on the basis of the drug-target pairs in the training and the validation sets.

2.8.3. Measurement of Prediction Quality

Assessing a model is quite tricky. It isn't reliable to evaluate model performance (training accuracy of the final epoch) merely based on one specific validation set, in that the accuracy obtained from one validation set can be very different to that from another. Thus, we took advantage of a general technique named 10-fold cross validation to avoid the possible bias. That is, the original samples for training is randomly partitioned into 10 equal size subsamples; one is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. Through repeatedly training and validating the models 10 times with different partitions, we then compute an average score over the rounds to give an estimate of the model's predictive performance. Afterwards, the model parameters with different validation errors were applied to the testing data for generating the final test accuracy and loss.

Aside from that, to compare the effectiveness of our DTI model with that of other state-of-the-art ML-based approaches, we adopted AUC-ROC curve [65] to tell how good a specific model can distinguish two classes, that is, whether or not a drug interacts with a target. ROC represents the relationship between benefit (true positive rate) and cost (false positive rate), while AUC is a value in the range of 0 to 1 that indicates the degree of separability between classes. The higher AUC we obtain, the more accurate outcome we will get from the model. The formulas for AUC-ROC curve are shown as below:

$$\text{True Positive Rate} = \text{Recall/Sensitivity} = \frac{TP}{TP + FN} \quad (49)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (50)$$

$$\text{False Positive Rate} = 1 - \text{Specificity} = \frac{FP}{TN + FP} \quad (51)$$

where True Positive (TP) means that the actual value is positive and is judged correctly; False Positive (FP) shows that the actual value is positive but is judged by mistake; True Negative (TN) indicates the actual value is negative and is judged accurately; False Negative (FN) represents the actual value is negative but is judged in error.

3. Results

3.1. Overview of Systems Medicine Discovery Procedure

Though a great advance has been made to design and evaluate medicine based on the application of diversified drug repositioning approaches, it still takes a large amount of time and vigor if without a potent avenue to select the promising biomarkers and drugs for the clinical trials. In this study, we proposed a systematic biology method containing two sections including potential biomarkers identification and novel drugs discovery.

The overall flowchart of the proposed systems biology procedure is shown in Figure 1. On one hand, we applied a systematic identification analysis procedure to the constructed candidate HPI-GEN, considering both accuracy and model complexity with the help of the information from the two-side real-time profile RNA-seq. Note that the total nodes and edges respectively given in Tables 1 and 2 of the real HPI-GEN in Figure 2 abstracted from the candidate HPI-GEN has considerably shrunk in comparison with that of the candidate HPI-GEN, which indicates that the false-positives were eliminated during the identification process. Supported with the PNP approach, we then identified the core HPI-GEN in Figure 3 by selecting top-ranked 3000 nodes with significant projection values that could reflect 85% of the real HPI-GEN. The higher the projection value is, the stronger regulatory capacity it possesses in the host/pathogen cross-talk mechanism during HBV infection. In the meantime, given identified KEGG pathways derived through uploading nodes in core HPI-GEN into DAVID in Table 3, we selected the core signaling pathways so as to explore the underlying etiologic mechanism

in HBV infection as shown in Figure 4, involving microenvironmental factors by surveying relevant papers and research.

Table 1. The total number of nodes in candidate interspecies host/pathogen interspecies genetic and epigenetic interaction network (HPI-GEN) and identified HPI-GEN during Hepatitis B Virus (HBV) infection.

Node	Candidate HPI-GEN	Real HPI-GEN
LncRNA node	143	137
MiRNA node	922	922
TF node	2477	1325
Receptor node	2105	2105
Virus node	4	4
Protein node	13,465	12,258
Total node	19,116	16,751

Table 2. The total number of edges in candidate interspecies HPI-GEN and identified interspecies HPI-GEN during HBV infection.

Edge	Candidate HPI-GEN	Real HPI-GEN
LncRNA-LncRNA	0	0
LncRNA-miRNA	0	0
LncRNA-TF	156	9
LncRNA-Receptor	2	0
LncRNA-Protein	90	8
LncRNA-Virus	8	1
MiRNA-LncRNA	510	303
MiRNA-MiRNA	96	86
MiRNA-TF	42,913	7909
MiRNA-Receptor	26,888	9694
MiRNA-Protein	146,144	65,561
MiRNA-Virus	39	15
TF-LncRNA	231	157
TF-MiRNA	1439	944
TF-TF	33,426	8061
TF-Receptor	14,948	6500
TF-Protein	76,441	39,574
TF-Virus	54	38
Receptor-LncRNA	26	15
Receptor-MiRNA	143	102
Receptor-TF	2364	644
Receptor-Receptor	1490	693
Receptor-Protein	7755	4560
Virus-LncRNA	5	3
Virus-MiRNA	109	80
Virus-TF	509	77
Virus-Receptor	164	72
Virus-Protein	682	276
PPIs	4,039,657	906,428
Total edge	4,396,289	1,051,810

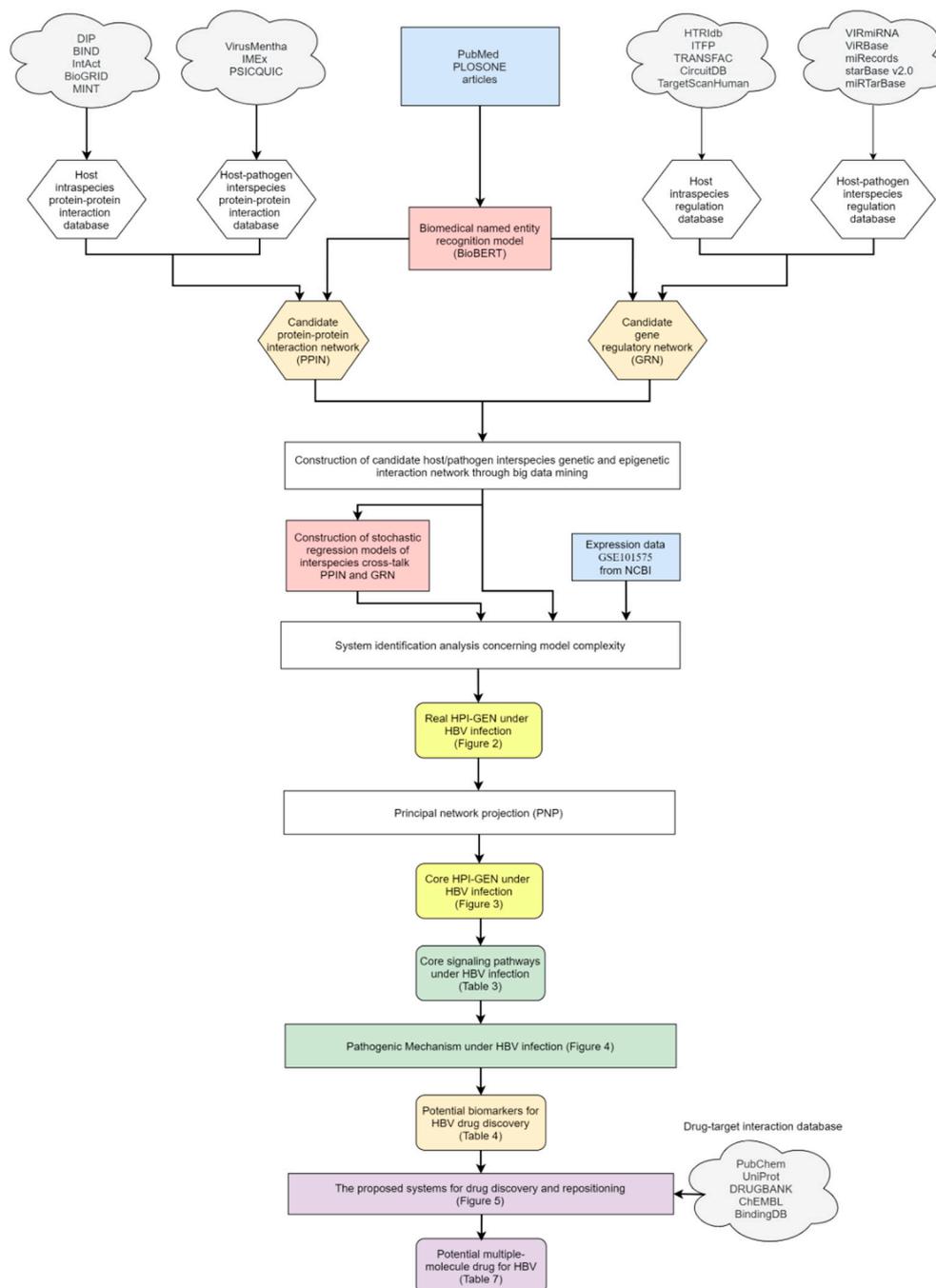


Figure 1. Flowchart of applying systems biology method to construct the candidate HPI-GEN, real HPI-GEN, core HPI-GEN, and core signaling pathways during the infection of HBV to identify the potential biomarker for later drug discovery repurposing. The grey blocks in the shape of a cloud indicate the online databases for constructing candidate HPI-GEN and later drug repositioning; the white diamond blocks are the corresponding data types of each database; the orange diamond blocks signify candidate protein-protein interaction network (PPIN) and gene regulation network (GRN); the blue blocks represent the input information including the raw RNA-seq data for system identification and the collected articles for BioBERT to augment the interspecies interactions and regulations in PPIN and GRN, respectively; the red blocks denote the models comprising BioBERT and stochastic regression models utilized in the systems biology approach; the white rectangle blocks indicate the systematic methods applied to constructing the candidate HPI-GEN and extracting core HPI-GEN; the purple blocks imply the drug discovery and repositioning procedure proposed afterwards; and the yellow rounded rectangular blocks are the real HPI-GEN and core HPI-GEN during the infection of HBV.

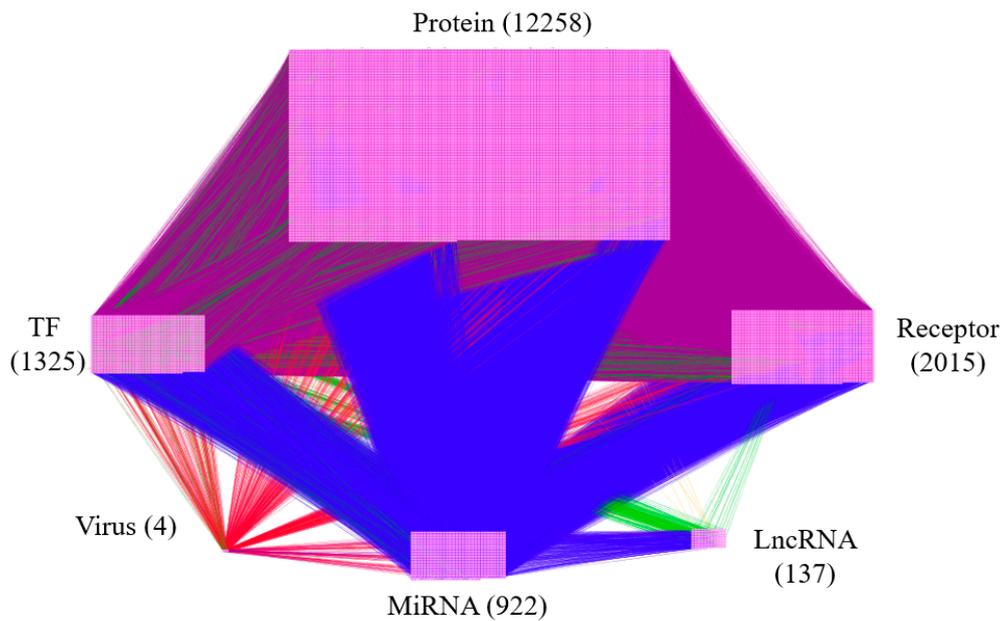


Figure 2. Real HPI-GEN between host and pathogen under infection of HBV. The real HPI-GEN describes the cross-talk interaction and regulation between host and pathogen after system identification analysis to prune false positives from candidate HPI-GEN. The purple lines denote PPIs; the green, yellow, blue, and red lines indicate the regulations by Transcriptional factors (TFs), lncRNAs, miRNAs, and Virus proteins respectively; and the total numbers of lncRNAs, miRNAs, proteins, receptors, Virus proteins, and TFs in the network are 137, 922, 12,258, 2015, 4, 1325, respectively.

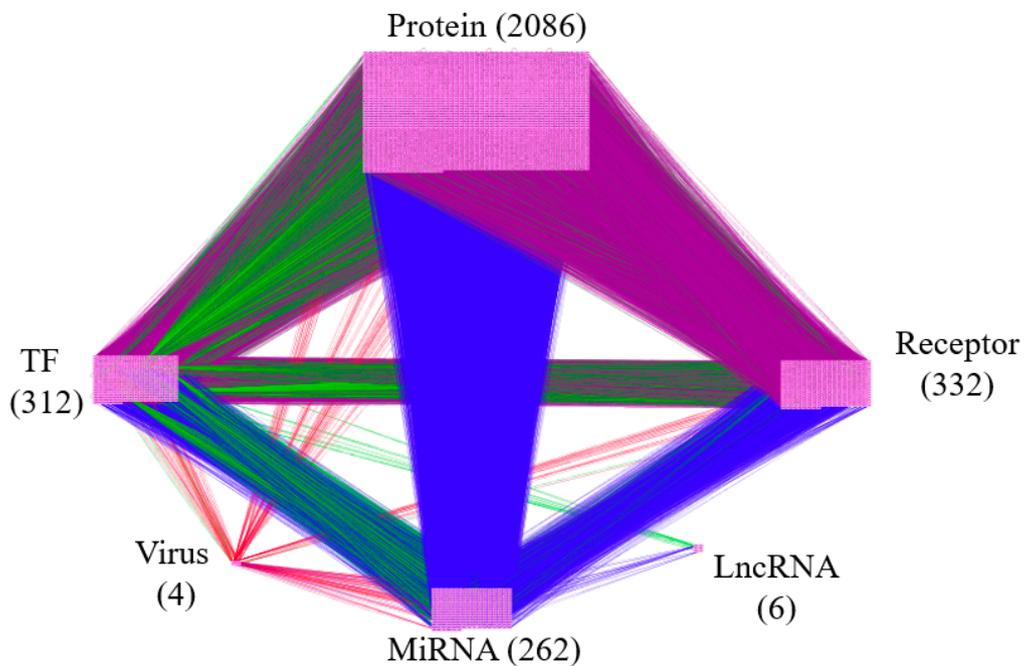


Figure 3. Core HPI-GEN under infection of HBV. The core HPI-GEN denotes the cross-talk interactions between host and pathogen after performing PNP to extract the top 3000 nodes from the real HPI-GEN. The purple lines denote PPIs; the green, yellow, blue, and red lines indicate the regulations by TFs, lncRNAs, miRNAs, and Virus proteins, respectively; and the total numbers of lncRNAs, miRNAs, proteins, receptors, Virus proteins, and TFs in the network are 6, 262, 2086, 332, 4, 312, respectively.

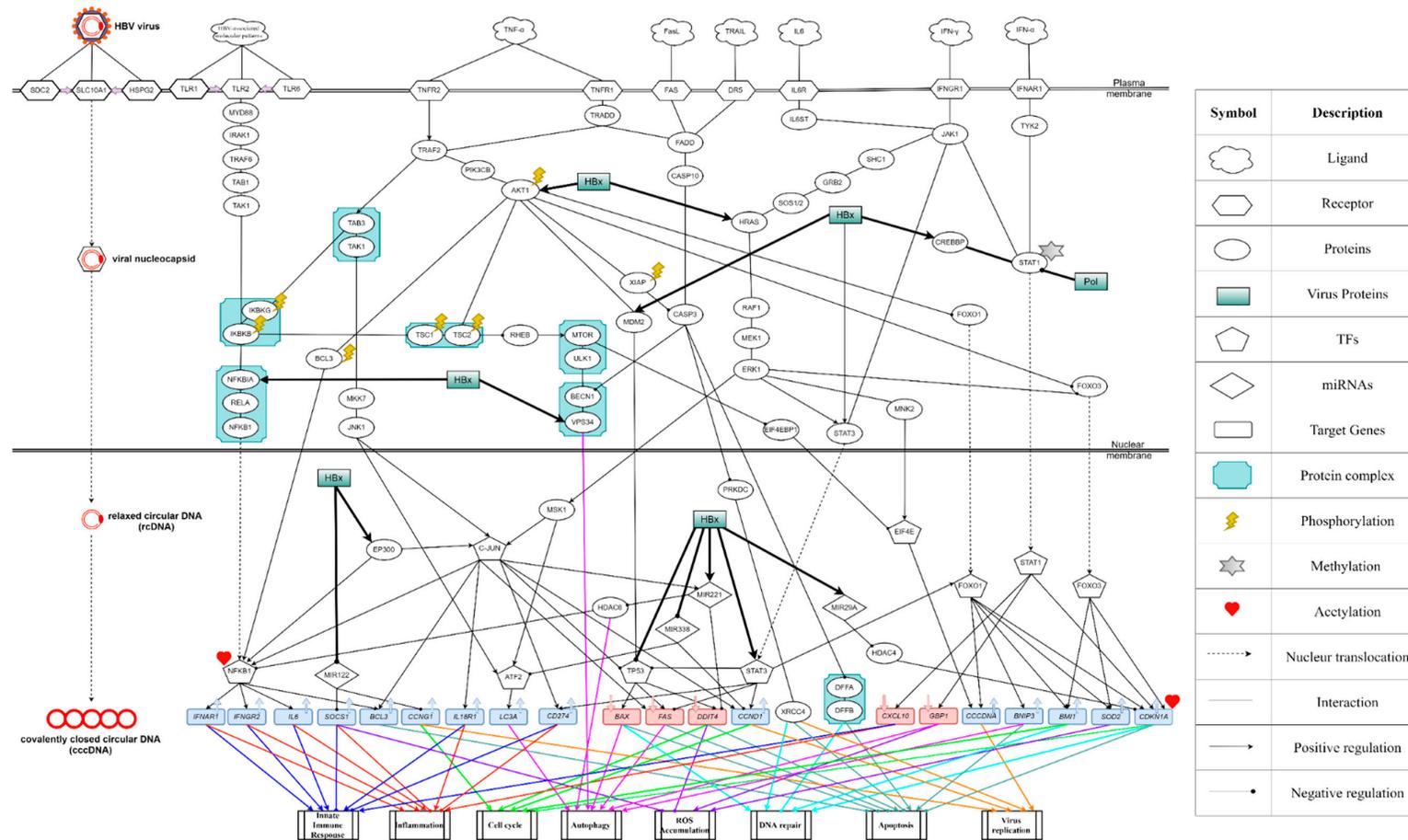


Figure 4. The core host/pathogen cross-talk signaling pathways during HBV infection. This is a network obtained from the core HPI-GEN by the annotation of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, which represents an intracellular pathogenic mechanism during the invasion of host hepatocyte by HBV. The solid black lines indicate the regulation (with an arrowhead) and interaction between nodes. The solid lines in different colors except black represent the regulations of the corresponding cellular functions; among them, the lines with an arrowhead denote the activation of cellular functions while the lines with a circular head indicate the inhibition of cellular functions. Furthermore, for the nodes of downstream target genes, the blocks with an upward pointing arrow signify a rising expression; comparatively, the blocks with a downward pointing arrow signify a dropping expression under HBV infection.

On the other hand, after determining the biomarkers from signaling pathways of the core HPI-GEN as shown in Table 4, we proposed a deep learning-based drug-target interaction (DTI) method to predict the potential interactions (dockings) between the identified biomarkers and their corresponding drugs for further candidate drugs discovery. The outcome of the DTI model is presented in a form of probability. The higher probability analyzed by the DTI model, the connection between the selected biomarker and the predicted drug is more likely to exist. However, given the enormous number of candidate drugs, the general drug specifications such as regulation ability, toxicity, and sensitivity from multiple databases were taken into consideration to reduce the search space and ensure the safety and efficacy in clinical trials. The pertinent flow-process diagram of the systems molecular drug design procedure is shown in Figure 5.

3.2. Extracting Core Signaling Pathways from Identified HPI-GEN and Core HPI-GEN in HBV Infection

For the purpose of analyzing the molecular mechanism under HBV infection, extracting core signaling pathways from candidate HPI-GEN is critical. The process to construct candidate HPI-GEN and identify real HPI-GEN as well as core HPI-GEN has been shown in the flowchart of Figure 1.

Thereafter, we identified the real HPI-GEN via parameter identification concerning the cost of model complexity under the auspices of corresponding RNA-seq data to prune false positives from candidate HPI-GEN (refer to Section 2.4–Section 2.6 for more details). The total numbers of the nodes of proteins, Transcription factors (TFs), miRNAs, lncRNAs, and Virus proteins as well as the edges of their interactions and regulations for the candidate HPI-GEN and identified real HPI-GEN are given in Tables 1 and 2, respectively. In addition, through employing the network visualizing software Cytoscape, the network of real HPI-GEN under HBV infection is shown in Figure 2.

Since the immense complexity of real HPI-GEN limits the possibility to efficiently investigate the pathogenic mechanism under HBV infection, applying principal network projection (PNP) and subsequently abstracting the top 3000 nodes with significant projection values on the 85% pivotal network structures to distill core HPI-GEN are essential. Meanwhile, the results of core HPI-GEN under the infection of HBV is shown in Figure 3.

Moreover, we employed DAVID Bioinformatics Resources version 6.8 to analyze the enrichments of relative pathways and specific cellular functions of core HPI-GEN as shown in Table 3.

Table 3. The pathway enrichment analysis of proteins by applying the DAVID in core HPI-GEN during HBV infection.

Term	Numbers	<i>p</i> -Value
MicroRNAs in cancer	103	8.6×10^{-18}
Pathways in cancer	108	8.7×10^{-10}
FoxO signaling pathway	49	5.6×10^{-9}
Viral carcinogenesis	63	6.3×10^{-8}
Hepatitis B	47	7.3×10^{-7}
Cell cycle	39	1.6×10^{-5}
MAPK signaling pathway	64	6.2×10^{-5}
TNF signaling pathway	33	1.3×10^{-4}
Ras signaling pathway	55	6.2×10^{-4}
PI3K-Akt signaling pathway	77	7.6×10^{-4}

Through the multifaceted assessment of the core HPI-GEN based on existing studies and the annotation of KEGG signaling pathways, we could attain the core signaling pathways associated with microenvironmental factors including cytokines, chemokines, etc., for the pathogenesis of HBV infection as represented in Figure 4. Then, on the basis of the investigating the consequence of the systematic pathogenic molecular mechanism, we could identify the potential biomarkers holding great promise for the development of therapeutic targets.

3.3. Analysis of Core Interspecies Cross-Talk Pathways to Investigate Host/Pathogen Offensive/Defensive Mechanism during HBV Infection

Based on the core host/pathogen cross-talk signaling pathways in Figure 4, several identified processes leading to dysfunction of antiviral machinery are investigated to afford a stepwise understanding of the host/pathogen offensive and defensive mechanism under HBV infection.

3.3.1. Inception of HBV Infection

Hepatitis B Virus (HBV) has a 3.2-kb circular and partially double-stranded DNA genome that contains four genes named S, C, P, and X genes coding for the envelope proteins, the core protein and a related protein termed precore protein, the viral DNA polymerase, and the multifunctional regulatory protein, respectively [66,67]. The journey of HBV virions starts from attaching to host hepatocytes through heparan sulfate proteoglycans (SDC2 and HSPG2) and promptly transfers to sodium acid cotransporter (SLC10A1), initiating the HBV life cycle. Once entering the cells, the viral nucleocapsid containing the partially double-stranded DNA, known as the relaxed circular DNA (rcDNA), will be released into the cytoplasm and transported into the nucleus. The plus strand of the rcDNA is then repaired and completed to generate the covalently closed circular (cccDNA), which later forms the HBV minichromosome and is transcribed into RNAs for viral replication [68].

3.3.2. Liver Microenvironment and Immune Pathogenesis under HBV Infection

Inflammatory and innate immune responses impact the pathogenesis of numerous diseases and participate in the activation of common inflammatory mediators and regulatory pathways. On top of the Toll pathways that could be directly stimulated by the viral particles, in response to the foreign invasions, leukocytes, e.g., neutrophils and macrophages that phagocytose and kill pathogens, might simultaneously coordinate additional host responses by synthesizing a wide range of inflammatory mediators, such as TNF- α , IL-6, IFN- α , and IFN- γ [69]. These cytokines and chemokines provoke the secretion of antiviral effectors and a series of immunoregulatory mechanisms against HBV infection. In general, cellular and molecular events and interactions would effectively mitigate impending injury or infection during inflammatory responses [70]. However, along with the frequent recurrence of inflammation, once imbalances of the microenvironment occur, it might lead to abnormal epigenetic variations and even immune dysregulation [71], which consequently incurs viral relapse.

3.3.3. Toll Pathway as the First Line of Defense against Infection

Toll-like receptors (TLRs) have long been recognized as the first line of antiviral immunity in that they initiate intracellular signaling pathways to induce antiviral mediators, such as interferons (IFNs) and other cytokines [72]. Upon the recognition of foreign invaders (HBV-associated molecular patterns), TLRs elicited phosphorylation of IKK β (IKKB) through several signaling transduction proteins, including MyD88, IRAK1, TRAF6, TAB1, and TAK1. I κ B α (NFKBIA) is a downstream protein of IKK β keeping NF- κ B sequestered in an inactive state in cytoplasm [73]. When phosphorylated by IKK β , I κ B α dissociated from NF- κ B, resulting in the nuclear translocation of NF- κ B and the induction of its several targets, in which the *Interleukin 6 (IL6)*, the *interferon-alpha receptor 1 (IFNAR1)*, the *interferon-gamma receptor 2 (IFNGR2)*, and *BCL3* were recognized as shown in Figure 4. Among them, IL-6 is the pleiotropic cytokine encoded by *IL6* gene that catalyzes immune reaction and inflammatory responses [74,75]; IFNAR1 is the receptor that binds type I interferons and activates the JAK-STAT signaling pathway, along with MAPK, PI3K, and Akt signaling pathways associated with multiple cellular functions, such as inflammation and immune response regulation [76]; IFNGR2 is a receptor that binds to interferon- γ (IFN- γ), which is the immune interferon produced predominantly by natural killer (NK) and natural killer T (NKT) cells in response to viral infection [77]; and BCL3 is a protein that inhibits DNA damage-induced apoptosis and acts as a transcriptional coregulator of NF- κ B [78,79].

In addition to stimulation of the downstream NF- κ B pathway, the phosphorylated IKK β also activated the pathway of mammalian target of rapamycin (mTOR) by triggering the disruption of TSC1/TSC2 complex. In general, TSC1 stabilizes TSC2 through direct binding, thereby preventing TSC2 from ubiquitination and degradation. Once IKK β phosphorylates TSC1 incurring collapse of its connection with TSC2, it obliquely leads to the stabilization of RhebGTP, which mediates signaling to mTOR [80,81]. As for the substrates of mTOR, among all the identified downstream proteins in the core host/pathogen signaling pathways as shown in Figure 4, the translational repressor 4EBP1 (EIF4EBP1) stood out to be a pivotal constituent. It is well acknowledged that when the 4EBP1 protein is hyperphosphorylated by mTOR kinase, it can't bind and hijack the eIF4E factor, i.e., a general initiator for translational regulation [82,83]. Concurrently, it is reported that the released eIF4E might form a complex with the 5' stem-loop structure of the HBV genome, thus ensuring viral replication [84].

3.3.4. TNF- α -Stimulated Signaling Pathways

TNF- α is a multifunctional cytokine that plays a principal role in cell proliferation, apoptosis, inflammation, and immune response. From the core host/pathogen cross-talk signaling pathways in Figure 4, as soon as TNF- α bound to TNF receptors (TNFRs), it contributed to the phosphorylation of Akt via both TNFR2/TRAF2/PIK3CB and TNFR1/TRADD/TRAF2/PIK3CB signaling transductions. Akt is an effector of Class I PI3K triggering multiple downstream pathways and has been shown to be a critical mediator of cell proliferation and survival in a variety of cell types [85]. In Figure 4, a few pathways of Akt were incorporated into the core host/pathogen cross-talk network. First of all, the activated Akt inhibited the induction of autophagy through transactivating mTOR, i.e., impeding GTPase-activating protein complex TSC1-TSC2 as previously described. It is worth noting that TSC2 rather than TSC1 was directly phosphorylated by Akt, which destabilized TSC2 and disrupted its affinity with TSC1 [86]. Secondly, Akt also enhanced MDM2-mediated ubiquitination leading to p53 (TP53) degradation [87] and catalyzed the nuclear localization and stabilization of BCL3 by phosphorylation [79], where BCL3 is the aforementioned transcription coregulator having been proven to interact with NF- κ B so as to facilitate its transcriptional ability. Moreover, it even caused nuclear exclusion and cellular dysfunctions of FoxO proteins through phosphorylation and repression [88]. Therefore, as the identified target genes of FoxO proteins in the core host/pathogen cross-talk pathways, BNIP3, BMI1, SOD2, and cyclin-dependent kinase inhibitor 1 (CDKN1A) encoding p21, might probably be down-regulated. Within, BNIP3 is a pro-apoptotic factor that induces cell death through interacting with anti-apoptotic proteins [89]; BMI1 is a critical epigenetic repressor that functions through chromatin remodeling and plays a central role in DNA damage repair [90,91]; and SOD2 is the antioxidant defense enzyme that participates in apoptotic signaling and oxidative stress regulation [92]. However, it is well documented that FoxO1/FoxO3 could rapidly relocate into the nucleus in response to STAT3 activation by IL-6, and the accumulation of FoxO1/FoxO3 in nuclei coincided with elevated expression of CDKN1A and other genes [93]. This phenomenon could interpret the high expression of the target genes BNIP3, BMI1, SOD2, and CDKN1A, as well as the promotion of viral replication, since activator FOXO1 has also been found to bind to HBV DNA and activate its transcription [94].

Aside from the activation of PI3K/Akt pathway, in response to cytokine signals, TRAF2 also recruited TAK1/TAB3 complex and IKK complex, which is composed of IKK α , IKK β , and IKK γ (IKBKG), bringing about the autophosphorylation-dependent activation of TAK1 [95] and the activation of its substrates, such as IKK β and MKK7. MKK7 then upregulated the transcriptional activity of c-JUN and ATF2 via activation of the JNK1. c-Jun is an AP-1 family transcription factor involved in the regulation of cell death and survival [96] as well as inflammation and innate immune response [97]. In Figure 4, the induction of c-Jun is discovered to be associated with the overexpression of the proinflammatory cytokine receptor IL18R1 essential for IL18 mediated signal transduction [98], the ligand PD-L1 (CD274) regulating the cellular immune responses to prevent inflammation [99], and the major cyclins CCND1 controlling the progression of a cell through the G1 phase by conjugating CDK4/6 enzymes required for the synthesis of cell cycle [100]. Besides, c-Jun not only modulated the expression of BCL3 by serving

as either the coactivator of NF- κ B or the activator of BCL3 gene to reinforce cell survival [101,102], but also induced miR-221 to interfere with the regulation ability of the histone deacetylase 6 (HDAC6), which has been reported to both suppress NF- κ B activation and control autophagosome maturation essential for ubiquitin-selective quality-control autophagy [103]. Apart from that, c-Jun could even impact the modulation of *Fas* expression, which sparked off the loss of Fas-ligand (FasL)-mediated apoptosis by means of down-regulating the activity of p53.

3.3.5. TAK-STAT Signaling Pathways

The Janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling pathway is a chain of interactions between proteins involved in various physiological processes, including immune function, cell growth, and cell death [104]. The activation of STAT3 was mainly provoked by JAK1 while JAK1 was brought in the proximity of receptors including the Glycoprotein 130 (IL6ST) and the gamma interferon receptor 1 (IFNGR1), which receive the signal from IL-6 and IFN- γ , respectively. However, the activation of STAT1 primarily stemmed from the phosphorylation of TYK2, a member of JAK family that was previously described by other research groups [105], by the alpha interferon receptor 1 (IFNAR1) residing in IFN- α stimulation. In spite of forming homodimers or heterodimers and translocating to the nucleus where they regulated the transcription of target genes, STAT3 and STAT1 played different even in part opposite roles in the process of pathogen infection. STAT3 constitutive activity is essential for cell survival and the extent of inflammation. Among the identified STAT3 targets in the host/pathogen cross-talk signaling pathway in Figure 4, an anti-inflammatory and anti-apoptosis factor referred to as *PD-L1* and the major cyclin *CCND1* involved in proliferation and cell cycle progression as mentioned previously were included. Additionally, STAT3 could also inhibit apoptosis by suppressing *Fas* transcription and p53 expression.

In contrast, STAT1 promoted apoptosis by inducing the expression of *CXCL10* and strengthened the antiviral ability through modulating the expression of *GBP1*. Amongst these two downstream targets, *CXCL10* is a ligand for the receptor *CXCR3* involved in chemotaxis, induction of apoptosis, and regulation of cell growth associated with a variety of human diseases including Hepatitis C [106] and Endotheliitis [107]. *GBP1* is a binding protein which has been shown to provide broad host protection against different pathogen classes through expediting oxidative killing and delivering antimicrobial peptides to autophagolysosomes [108]. Furthermore, STAT1 also negatively regulated the cell cycle through inducing the expression of the CDK inhibitors *CDKN1A*, a target of transcription factors FoxOs as stated previously.

Albeit STAT3 was directly activated by JAK1, it could also be indirectly triggered through the cytoplasmic signaling cascade of SHC/ERK pathway composed of signaling proteins, i.e., SHC1, GRB2, and SOS1, and the classical MAPK/ERK pathway comprised of signaling proteins, i.e., HRAS, RAF1, MEK1, and ERK1. Except inducing the activation of STAT3, MAPK/ERK pathway also contributed to the MSK1-mediated phosphorylation of c-Jun and ATF2, the inhibition of FoxO3 nuclear translocation, and most importantly, the regulation of eIF4E activity having been elucidated chiefly via phosphorylation on serine 209 of eIF4E by other research groups [83,109], which ultimately proliferated viral replication. More exactly, this was carried out by mitogen-activated kinase2 (MNK2), the downstream kinase of ERK1.

3.3.6. TNFs-Induced Apoptotic Pathways

As a component of the cell response to viral infection, apoptosis mediated by the sequential activation of caspases is indispensable for aborting the production and release of progeny virus. While diverse signaling pathways inducing apoptosis have been found, caspase-mediated proteolysis of downstream substrates is still a critical element of the execution pathway common to all forms of apoptosis [110]. And among all the avenues conducive to apoptosis progression, the pathways contributing to the modulation of caspase-3 (CASP3), a frequently activated death protease that

catalyzes the specific cleavage of many integral cellular proteins, occupied a key position and a majority of them began with the activation of FADD in Figure 4.

Induction of FADD, the upstream activating protein of caspase-10 (CASP10), was mediated by not only the bifurcated TNFR1-TRADD signaling cascades but also the interaction with the FasL receptor (FAS) and the TNF-related apoptosis-inducing ligand receptor (DR5). Upon stimulation, FADD in turn recruited the initiator caspase caspase-10 to form the death-inducing signaling complex, which leads to activation of a caspase cascade and eventual cell-death. Having been shown to activate the executioner caspase-3 through direct processing, caspase-10 is described to be functional-independent to caspase-8 (CASP8) in initiating Fas- and tumor necrosis factor-related ligand-mediated apoptosis [111,112]. Based on the cleavage potential of caspase-3 associated with the dismantling of the cell and the formation of apoptotic bodies, the identified critical targets including Beclin-1 (BECN1), DNA-PK (PRKDC), and the DNA fragmentation factor subunit alpha (DFFA) were specifically cleaved by caspase-3 with subsequent loss of its original capability. Among them, BECN1 is the component of the phosphatidylinositol-3-kinase (PI3K) complex that induces the appearance of autophagosome [113]; PRKDC is the nuclear serine/threonine protein kinase that takes part in degraded DNA repair [114]; and DFFA is a subunit of DFF. Further, according to previous study, whilst DFFA is cleaved, the DNA fragmentation factor subunit beta (DFFB) would dissociate from the cleaved fragments of DFFA and trigger both DNA fragmentation and chromatin condensation during apoptosis [115].

Additionally, it is worthwhile to note that Akt, including AKT1 and AKT2, has been reported to interact with X-linked inhibitor of apoptosis protein (XIAP) and protect it from ubiquitination and degradation [116]. Once phosphorylated by Akt, the XIAP would exploit its second baculovirus IAP repeat domain (BIR2) to inhibit the apoptotic executioner caspase-3 [117]. We believe that this identified episode carries a foreshadowing for the succeeding virus-induced malfunctions of programmed cell death.

3.3.7. Virus Proteins-Induced Pathogenic Mechanism

Proteins encoded by the viral genomes, especially HBx, a 154 amino acid gene product of the X gene, have emerged as promiscuous transactivators activating the transcription of host genes by interacting directly with nuclear transcription factors or by obliquely activating various signal transduction pathways in the cytoplasm and have been proven to be potent epigenetic modifying factors in hepatocytes.

HBx regulated the expression of *BCL3* to subvert apoptotic signaling pathways via both synergistically cooperating with p300 (EP300) to enhance c-Jun transcriptional activity and activating NF- κ B by either directly interacting with NF- κ B for augmenting its binding ability to its target DNA sequence or promoting phosphorylation of I κ B as shown in Figure 4. In addition, apoptosis inhibition could also be induced by dysfunction of caspase-3 through enhancing the ability of Akt, an upstream kinase of caspase-3 as stated earlier, by HBx. Once caspase-3 lost its function, in response to signal from damaged DNA, DNA-PKs, the substrate of caspase-3, would act in concert with XRCC4 and a number of tightly coupled proteins to repair DNA. Further, it was suggested that the intense activity or high expression of XRCC4 might be beneficial to the promotion of HBV cccRNA formation and thereby facilitate the viral replication [118]. Since Hepatitis B Virus supremely requires components from the host cell to complete its replication cycles, merely one developed avenue to hijack host signaling pathways can't meet its demand. In addition to the significant translation initiation factor eIF4E induced by HBx through Akt/mTOR pathway as just mentioned, eIF4E was activated via the MAPK/ERK pathway as HBx could also trigger Ras and its downstream signaling pathways. Hence, forming a node of convergence of the PI3K/Akt/mTOR and Ras/Raf/MAPK signaling pathways, the regulation of eIF4E is clearly consequential for HBx to promote viral replication. Additionally, overexpression of cyclin G1 (CCNG1) resulted from the loss of microRNA-122 (a liver-specific miRNA) induced by HBx could also augment virus proliferation. It was reported that CCNG1 specifically interacts with p53 and blocks its specific binding to HBV enhancer elements, which simultaneously abrogates p53-mediated inhibition of HBV transcription and releases cell proliferation from G1/S arrest [119].

p53 acts as a sequence-specific DNA binding protein that activates the transcription of substantial target genes associating with the regulation of cell cycle progression, apoptosis, DNA repair,

and senescence. In Figure 4, HBx interacted with p53 both directly and indirectly to inhibit its DNA consensus sequence binding and transcriptional activity, which has been reported to efficiently block p53-mediated cellular functions [120,121]. Apart from that, HBx also inhibited p53 by conjugating c-Jun and STAT3 as well as keeping MDM2, the principal cellular antagonist of p53 acting to limit the p53 growth-suppressive function [122], from ubiquitin-directed self-degradation.

As for the targets of p53, apoptosis-associated proteins, such as the proapoptotic protein Bax, which could permeabilize mitochondria and engage the apoptotic program giving rise to activation of caspase-3 [123,124], the apoptosis antigen Fas triggering the Fas signaling pathway as mentioned earlier, and the DDIT4 negatively regulating cell growth, proliferation, and survival via inhibition of the activity of mTOR [125] were identified in our core host/pathogen cross-talk signaling pathways in Figure 4. Further, having been reported to participate in the generation of ROS and p53-dependent DNA damage [126], DDIT4 was negative-regulated by the HBx-induced upregulation of microRNA-221 to promote aberrant proliferation. Such significant inverse correlation between DDIT4 mRNA levels and miR-221 expression has been observed in a mouse model of liver cancer as well [127]. Besides, the direct suppression of histone deacetylases 6 (HDAC6) by HBx-induced miR-221 would result in autophagosome maturation failure; nevertheless, HBx yet also bound to Vps34 (PI3KC3) to enhance its activity, inducing the autophagic response crucial for the formation of autophagosomes as well as the intensification of viral replication.

STAT1 and STAT3 are both critical elements in viral replication, despite acting as distinct roles from each other. Functional characterization of miR-338-3p indicated that miR-338-3p inhibited cell proliferation by inducing cell cycle arrest at the G1/S phase. The identified downregulation of miR-338-3p and tyrosine phosphorylation of STAT3 triggered by HBx might cause *CCND1* overexpression beneficial to virus cell cycle progression [128]. In addition, forced low expression of miR-338 by HBx also directly led to the increased level of autophagy. This process was mainly through the up-regulation of LC3 by the released transcription factor ATF2, which has been identified as a direct target of miR-338 [129]. Furthermore, instead of FoxO3, HBx appeared to assist FoxO1 in counteracting the reduction of its nuclear relocalization arising from Akt phosphorylation by inducing STAT3 activation, which is conducive to the transcription of cccDNA as previously described. In contrast, since interferons (IFNs) activated STATs via phosphorylation [130], once recruited by HBx, histone acetyltransferase (HAT) CBP (CEBBP) might dynamically regulate STAT1 acetylation to counteract IFN-catalyzed STAT1 phosphorylation, nuclear translocation, DNA binding, and target gene regulation. On the other hand, STAT1 nuclear translocation induced by IFN- α might also be deprived by HBV polymerase [131]. Previous studies demonstrated that miR-122 regulates type I IFN expression through both down-regulating SOCS1 and SOCS3 to enhance interferon-mediated suppression of HBV [132,133]. Nonetheless, it appears that merely the regulation of SOCS1 by miR-122 was identified in Figure 4. Finally, as a target gene of FoxOs and STAT1, *CDKN1A* overexpression was mediated by HBx-induced upregulation of miR-29a. Serving as a direct upstream of HDAC4, miR-29a could thereby alleviate the deacetylation of *CDKN1A* through HDAC4 inhibition.

3.4. Drugs Discovery and Repositioning Based on Selected Biomarkers for HBV

Through reviewing the cellular dysfunctions triggered by the core host/pathogen cross-talk interspecies signaling pathways under HBV infection shown in Figure 4, we selected the significant biomarkers including AKT1, NFKBIA, EIF4EBP1, HDAC6, STAT1, STAT3, and TP53 as drug targets shown in Table 4. However, there is still a long way to the clinical trials on account of the intrinsic difficulties for drug design and development. To this end, we proposed a systematic drug discovery and repositioning approach to uncover the multiple-molecule drug for therapeutic treatment based on the deep learning and data mining technique as shown in Figure 5.

Table 4. The drug targets identified for HBV infection.

Disease	Drug Targets
HBV infection	AKT1, NFKBIA, EIF4EBP1, HDAC6, STAT1, STAT3, TP53

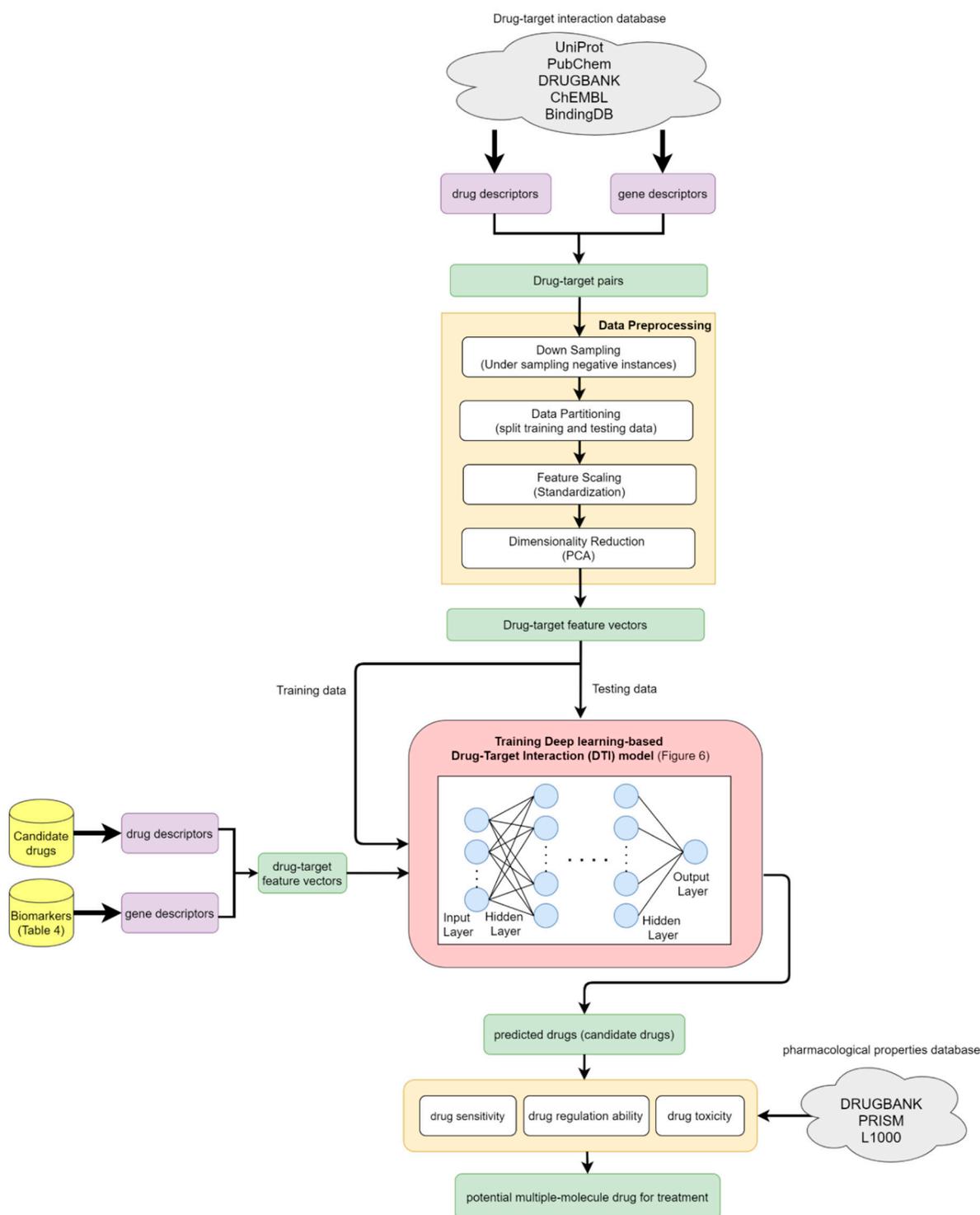


Figure 5. The flowchart of the proposed systems drug discovery and repositioning procedure. According to the potential biomarkers identified in Figure 4, we propose a systems drug discovery and repositioning method for figuring out promising multiple-molecule drug to alleviate HBV infection. First, we collected data in respect to drug-target interaction from various online databases. Then, we turned each drug and target into descriptors and assembled them into a matrix of drug-target pairs. To keep our model from poor performance, data preprocessing prior to training is indispensable. Accordingly, we adopted data preprocessing approaches including down sampling, data partitioning, feature scaling, and dimensionality reduction (For more details, readers can refer to Section 2.8.1 in Materials and Methods)

based on the property of the data. After training the Deep learning-based drug-target interaction (DTI) model shown in Figure 6 with the constructed training data, we evaluated the model performance (as shown in Table 5 and Figure 7) and compared the model efficacy of DTI with that of other traditional machine learning (ML)-based methods (as shown in Figure 8). Since there is only a neuron in the output layer which delivers the probability of whether the relation (docking) between a drug and a target exists, the DTI model enabled us to stress the importance on specific interactions (with output score approaching 1) and to identify candidate drugs. Next, concerning the availability of the predicted drugs (candidate drugs) via general criteria for drug design specifications such as regulation ability, toxicity, and sensitivity obtained from databases, we further filtered promising drugs that modulate multiple molecular targets and possess the efficacy with lower dosages. After all, low-dose drugs are more attainable and cause less harm to patients. In this regard, a set of multi-target drugs with appropriate toxicity for clinical investigations were selected as a multiple-molecule drug to overcome HBV infection.

Table 5. Model performance (10-fold cross validation).

	Validation Loss	Validation Accuracy (%)	Testing Loss	Testing Accuracy (%)
1	0.254831	92.09	0.236721	92.28
2	0.237422	91.89	0.2264	92.04
3	0.228304	92.35	0.226482	92.49
4	0.227555	92.62	0.228102	92.6
5	0.223151	92.55	0.218528	92.74
6	0.223817	93.04	0.227157	92.77
7	0.239133	92.39	0.225822	92.58
8	0.235536	92.93	0.226848	92.8
9	0.22345	93.1	0.218475	93.11
10	0.235327	92.21	0.219082	92.64
Average	0.232853	92.517	0.225362	92.605
Standard deviation	0.00983	0.409527	0.005559	0.294099

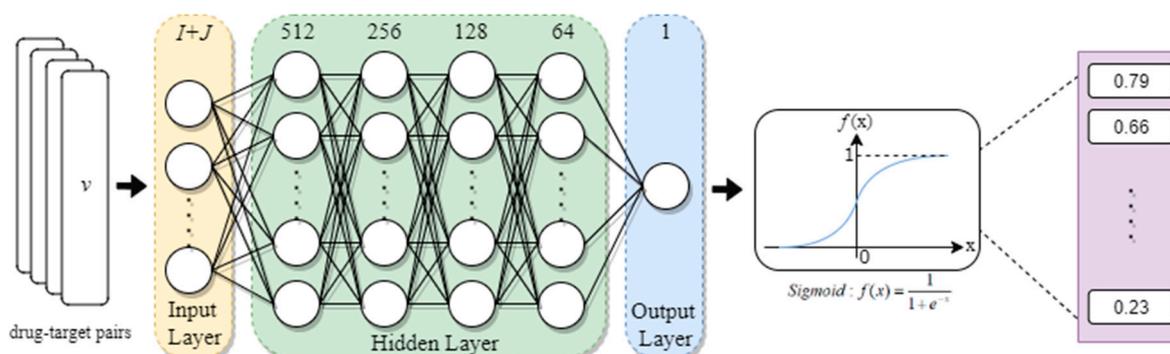


Figure 6. Configuration of the deep learning-based DTI model. v is the vector of each drug-target pair for the input layer; I is the total number of drug features; and J is the total number of target features.

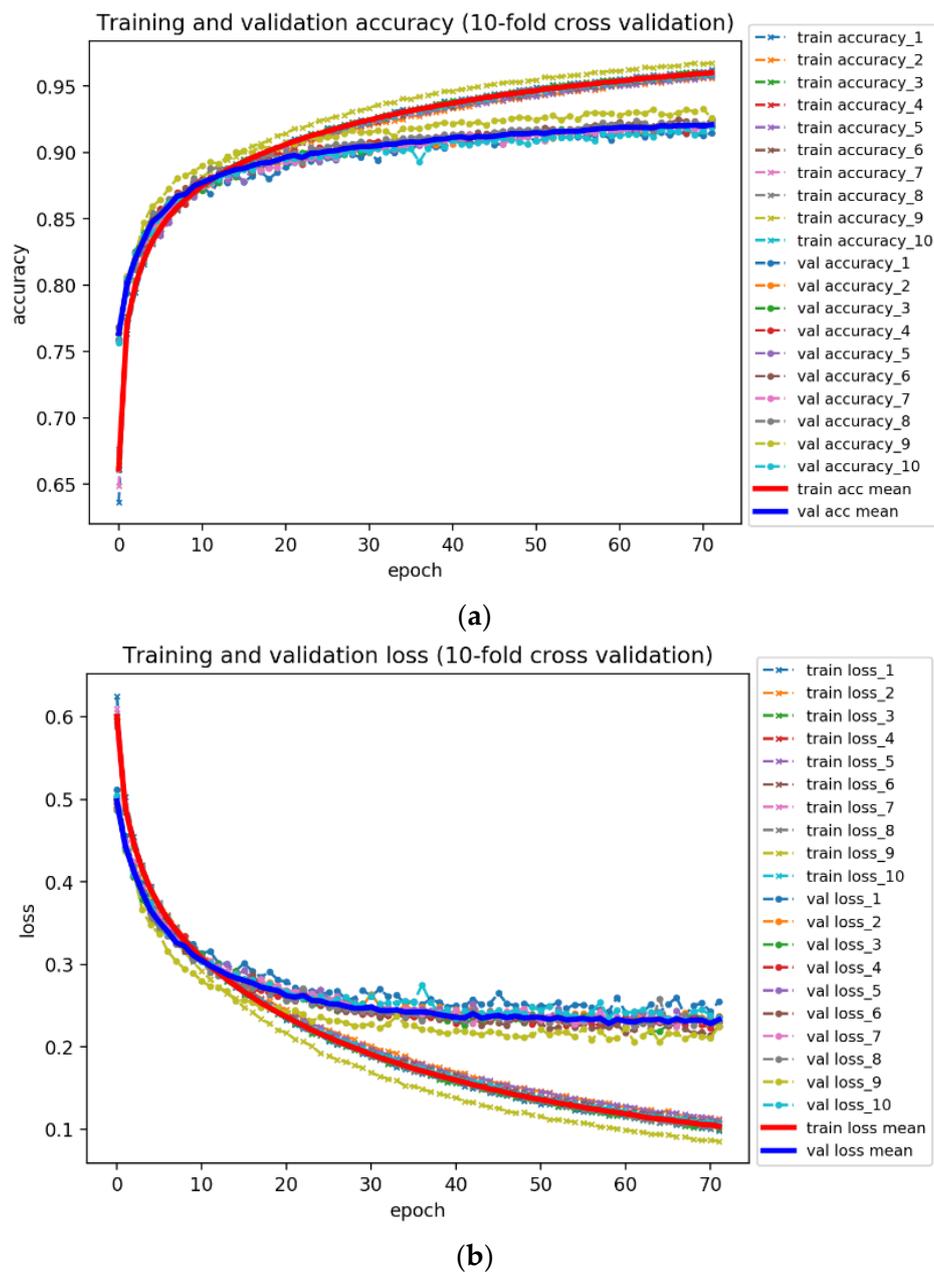


Figure 7. Training and Validation Learning curves (10-fold cross validation). “-x-” lines in different colors denote the training accuracy and loss, while “-o-” lines in different colors represent the validation accuracy and loss in (a,b), respectively. The bold lines in red and blue indicate the model’s average loss and accuracy of training and validation in (a,b), respectively. Moreover, since stopping the training of the neural network early before it has overfitted the training dataset can reduce overfitting and improve the generalization of the DTI model, Early Stopping approach is additionally employed to automatically stop the learning process at the epoch of 70.

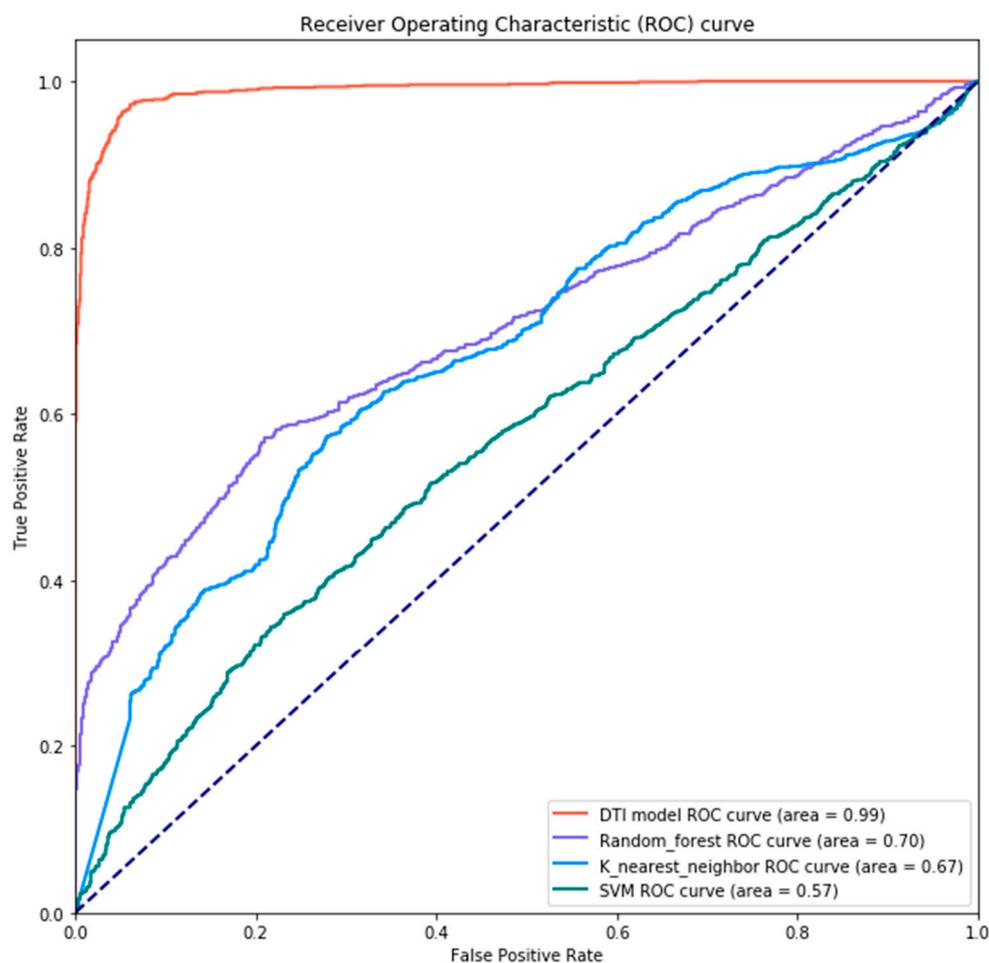


Figure 8. The Receiver Operating Characteristic (ROC) curve of models based on different methods. The accuracy of each method is calculated by the area under the corresponding ROC curve (AUC) (Readers can refer to Section 2.8.3 for further illustration), where the DTI model could achieve an accuracy of 0.99, which is higher than that of Random Forest (RF) (accuracy of 0.70), Nearest Neighbor (accuracy of 0.67), and Support Vector Machine (SVM) (accuracy of 0.57).

3.4.1. Deep Learning-Based Drug-Target Interaction Prediction Model

To effectively predict whether the identified targets interact with some existing drugs, we provide a deep learning-based DTI model as shown in Figure 6. It is a fully-connected neural network containing an input layer and five layers with weights; four hidden layers (512, 256, 128, and 64 neurons in each layer) and an output layer (1 neuron). The intermediate hidden layers are applied with rectified linear unit (ReLU) activation function that thresholds negative signals to 0 and passes through positive signal. This type of activation function allows faster learning in comparison with alternatives, e.g., sigmoid or tanh unit [134]. In contrast, the output of the last layer is fed to a sigmoid function which produces a likelihood probability (0,1) score of the corresponding predicted docking where a higher value symbolizes a more reliable drug-target interaction. In addition, to further reducing overfitting, we incorporated a dropout layer [135] with a probability of 0.5 following the output of each hidden layer.

For avoiding a poor performance due to assessing the model with a data set unfairly sampled (outliers or other anomalies are presented in one set), we introduced 10-fold cross validation measure to authenticate the learning effectiveness as shown in Table 5 and the corresponding learning curves that plot the learning performance of DTI model over experience and time are also presented in Figure 7. Note that since the model has been diagnosed well-fit and continuous training will likely lead to an

overfit, it automatically stops learning at the epoch of 70. Subsequently, via calculating the loss and accuracy of each DTI model yields on the test set, we received an average accuracy of 92.6 (%) with the standard deviation of 0.294 (%).

Moreover, with the increasement of experimental data, numerous machine learning approaches such as Random Forest (RF) [136], Support Vector Machine (SVM) [137], Nearest Neighbor [138], etc. have been applied to predict drug-target interactions. Therefore, to evaluate the model performance, we contrast our DTI model with a few of them, which are RF, Nearest Neighbor, and SVM. By adopting the visualization tool, Receiver Operating characteristic (ROC), we can clearly find that DTI outperforms other traditional binary classifiers as presented in Figure 8.

3.4.2. Systems Discovery and Design of the Multiple-Molecule Drug for HBV Infection

On the basis of the pre-trained DTI model, we could evaluate the docking ability between the identified biomarkers and the available drugs. The larger predicted result we obtain, the higher probability that the specific drug-target pair will have interaction. Namely, the candidate drugs would be sifted out owing to holding higher probability (with output score approaching 1) to interact with the identified biomarkers. However, the exploration of promising drugs from a large quantity of the candidates (predicted drugs) is still arduous. Accordingly, in order to narrow down the scope of the possible options and improve the reliability of the ultimate selected drug combination, general criteria for drug design specifications including regulation ability, toxicity, and sensitivity were applied to further filter out unavailable or inappropriate medications.

Under the criteria of the pharmacological properties, the regulation ability value for each drug was extracted from the LINCS L1000 level 5 dataset consisting of results from 978 genes of 75 cell lines treated with 19,811 small molecular compounds. By referring to LINCS L1000, we can examine whether a specific gene was up-regulated (positive values) or down-regulated (negative values) after being treated with a small molecular compound. Screening through a positive-regulation of NFKBIA, EIF4EBP1, HDAC6, STAT1, and TP53 and the negative-regulation of AKT1 and STAT3, the remaining drugs with agreeable regulation ability for the treatment of HBV-infected patients are shown in Table 6.

Table 6. The candidate multi-target drugs for the identified biomarkers.

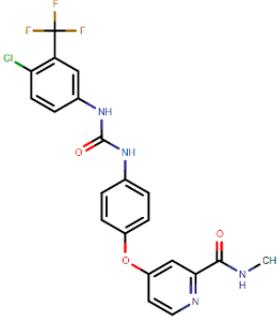
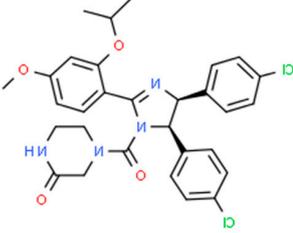
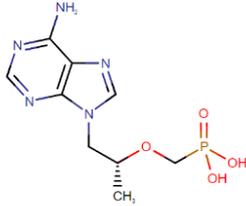
AKT1			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
ribavirin	−1.5283	1.9876	−1.18823
tacrolimus	−0.5897	2.7541	−1.50335
sorafenib	−0.3125	2.7885	−0.43586
tenofovir	−0.6152	2.955	−0.63588
meloxicam	−0.2726	3.4619	−0.68734
NFKBIA			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
chlortalidone	0.4056	1.8623	0.05198
busulfan	2.501	2.3207	−0.44631
rifaximin	0.681	2.6259	0.03563
sorafenib	3.5635	2.7885	−0.43586
meloxicam	3.2603	3.4619	−0.68734
EIF4EBP1			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
chlortalidone	0.4153	1.8623	0.05198
busulfan	0.7491	2.3207	−0.44631
rifaximin	1.7372	2.6259	0.03563
sorafenib	0.4733	2.7885	−0.43586
dactinomycin	0.2923	4.3767	0.00307

Table 6. Cont.

HDAC6			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
argatroban	0.5717	1.5976	−0.01405
chlortalidone	0.6495	1.8623	0.05198
entecavir	0.5388	2.3879	0.03615
nutlin-3	0.1749	2.529	−0.6695
meloxicam	0.1047	3.4619	−0.68734
STAT1			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
chlortalidone	0.234	1.8623	0.05198
ribavirin	0.982	1.9876	−1.18823
zafirlukast	0.2615	2.5723	−0.24097
sorafenib	0.5799	2.7885	−0.43586
hydroflumethiazide	1.5238	3.1299	0.3949
STAT3			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
argatroban	−1.2902	1.5976	−0.01405
ribavirin	−0.8934	1.9876	−1.18823
lamivudine	−0.3132	2.1348	−1.33248
zafirlukast	−0.5053	2.5723	−0.24097
tenofovir	−1.5088	2.955	−0.63588
hydroflumethiazide	−0.7348	3.1299	0.3949
meloxicam	−1.6727	3.4619	−0.68734
TP53			
Drug	Regulation ability (L1000)	Toxicity (LD50, mol/kg)	Sensitivity (EC50, nM)
acetylcysteine	1.0705	1.294	−1.2170
busulfan	2.2886	2.3207	−0.44631
nutlin-3	1.6295	2.529	−0.6695
sorafenib	0.5683	2.7885	−0.43586
calcitriol	0.4321	5.1352	0.28577

Subsequently, to further narrow down the possibilities for drug combinations, the drug toxicity specification (LD50) recorded in DrugBank as well as the drug sensitivity specification (EC50) obtained from the PRISM dataset, which is a dataset comprising sensitivity values of 4518 drugs tested across 578 human cell lines, were also adopted. Being the numeric index of lethality, median lethal dose (LD50) plays a pivotal role in drug safety evaluation [139]. The drug with lower LD50 possesses higher toxicity, and thus it is more detrimental. On the other hand, half maximal effective concentration (EC50) is used as a measurement of concentration, which is often adopted to evaluate the suitability and the efficacy of drugs, where lower EC50 indicates the stronger inducibility, i.e., higher sensitivity. Based on these concepts, the drug combination comprising Sorafenib, Nutlin-3, and Tenofovir as shown in Table 7 was then designed as the multiple-molecule drug to rejuvenate the dysfunctions in the pathogenic mechanism under HBV infection with minor collateral tissue damage.

Table 7. Potential multiple-molecule drug and the corresponding target genes for HBV infection therapy.

Targets Drugs	AKT1	NFKBIA	EIF4EBP1	HDAC6	STAT1	STAT3	TP53
Sorafenib	O	O	O		O		O
Nutlin-3				O			O
Tenofovir	O					O	
Chemical structures of multi-target drugs							
Sorafenib	Nutlin-3			Tenofovir			
							

O indicates the docking between each multi-target drug composing the multiple-molecule drug and their corresponding drug target.

Among them, Sorafenib is a multikinase inhibitor drug having been approved for several tumor types, including Hepatocellular Carcinoma (HCC) [140]. Although prominent effects of Sorafenib for apoptosis induction were recognized, several limitations have been indicated by numerous medical studies upon applying Sorafenib to various cancers as a single agent. Therefore, the use of Sorafenib in combination with other targeted agents appears to be the main strategy for combatting drug resistance and has been widely explored with promising results [141]. Falling under the category of small molecule inhibitor, Nutlin-3 is conducive to the restoration of p53 pro-apoptotic ability, leading to the activation of cell cycle arrest and apoptosis pathways [142]. Though it is a novel antitumor compound, which hasn't been clinically approved, the synergistic effect of Nutlin-3 combined with other agents contributes to diminishing the dose required in monotherapy and decreasing the occurrence of adverse drug events [143,144]. Being a potent nucleotide analogue reverse-transcriptase inhibitor, Tenofovir is highly active against human immunodeficiency virus (HIV). In spite of its clinical use associated with the risk of kidney injury [145], it has been shown highly effective to patients who have never undergone an antiretroviral therapy [146]. Owing to its low toxicity, trifle side effects, and subtle drug resistance compared to other antivirals, Tenofovir is thereby frequently used as a member of combined drugs against HBV to ensure stronger antiviral activity and a more favorable safety profile [147,148]. It is worth noting that while therapy with Tenofovir is commonly used to treat Hepatitis B, life-long therapy is needed, which often increases the risk of relapse. In this respect, the combination of other viable drugs with Tenofovir may have a chance to improve the therapeutic effect and shorten the course of therapy.

Taken together, by administering the proposed multiple-molecule drug, an up-regulation of NFKBIA, EIF4EBP1, HDAC6, STAT1, and TP53 accompanied by the down-regulation of AKT1 and STAT3 can plausibly be attained, yielding encouraging results for the treatment of HBV-infected patients.

4. Discussion

Traditional drugs commonly used for treatment in patients afflicted with HBV are interferon and nucleoside analogues. However, the low therapeutic response of patients to interferon and the inevitable development of drug resistance to nucleoside analogues render the clinical treatment challenging [149]. In addition, serious toxic accumulation of the nucleoside analogues during long-term therapy for HBV infection is nonnegligible as well. Therefore, great efforts have been placed into seeking highly specific

ligands affecting single target to alleviate HBV replication and to treat HBV-related complications. Yet, in view of the prohibitive development costs and timeline of a newly designed drug as well as the frequent failure of monotherapy due to drug resistance, growing interest has centered on discovering effective combinations of drugs for HBV treatment.

In this study, we develop a medicine discovery and repositioning procedure in terms of system identification approaches via two-side RNA-seq data and deep learning-based framework considering drug design specifications to find a possible drug combination for the treatment of HBV infection. Through investigation of core HPI-GEN annotated with KEGG pathways, the core signaling pathways responsible for the abnormally regulated cellular functions were extracted and unified in Figure 4. However, distinct cooperation between these pathways sparks diversified implication. Consequently, further discussion of the integrated effect is conducive to the selection of potential biomarkers in relation to the remedy of HBV infection.

4.1. Apoptosis in HBV Infection

As a fundamental and highly regulated biological process, apoptosis is a form of programmed cell death that participates in cell resistance to viral infection. Although mounting evidence shows that HBx is uniquely endowed with the capability to regulate apoptosis, its dual role as an apoptotic mediator raises the difficulty to elucidate the pathological mechanism [150].

Caspase-3 is the typical hallmark of apoptosis [110]. Numerous studies have shown that HBx abrogates the activity of caspase-3 [151–153], resulting in resistance to assorted apoptotic stimuli. In Figure 4, HBx-mediated inhibition activity of caspase-3 was chiefly through both activating XIAP, the inhibitor of caspase-3, and directly or obliquely regulating transcription factors such as NF- κ B and p53 to affect the expression of their downstream anti-apoptotic target, e.g., *BCL3* and *Bax*. Further, even if previous researches did not specifically indicate the paramount downstream anti-apoptotic genes of NF- κ B [154,155], *BCL3*, which hampers DNA damage-induced apoptosis and serves as a transcriptional coregulator of NF- κ B [78,79], counted for NF- κ B-mediated anti-apoptosis a lot in the pathogenic mechanism. Furthermore, inhibition of caspase-3 not only abrogated apoptotic killing triggered by the pathways in response to TNF- α , FasL, and TRAIL stimulation, but also contributed to the activation of autophagy, leading towards alleviated apoptosis coming from caspase-3 loss of its function to cleave Beclin-1.

Asides from that, *Fas*, indispensable for the FasL-induced pathway that gives rise to the activation of caspase-3, was downregulated during HBV infection. Negative-expression of *Fas* might impede the sustainable anti-viral response of death ligand stimulus-specific apoptotic pathway. In contrast, as a target gene relevant to execution and completion of apoptosis, the overexpressed *CDKN1A* triggered by HBx-induced HDAC4 inhibition was concordant with the earlier research that HBx can relieve a block on *CDKN1A* expression and prolong G1→S transition in human hepatoma cells [156,157]. Even though p21, the protein encoded by *CDKN1A*, is treated as a cyclin-dependent kinase inhibitor capable of inhibiting all cyclin/CDK complexes, it often promotes the assembly of type-D cyclins with CDK4 and CDK6 and arrests cell cycle progression in the G1 phase [156,158], which not only protects cells against apoptosis but also enhances the chance of tumorigenesis.

4.2. Autophagy in HBV Infection

Autophagy is a catabolic mechanism known to engulf long-lived proteins and damaged organelles via a lysosomal degradative pathway. Previous studies have suggested that HBV can hijack components in the autophagic regulatory pathways to promote its survival through augmentation of autophagosomes formation and inhibition of autolysosome maturation [159,160], which indicates the possibility of targeting the autophagic pathway for the treatment of Hepatitis B.

HBx-catalyzed autophagic response, which is crucial for the formation of autophagosomes as well as intensification of viral replication, resided mainly in the increased activity of the Beclin-1/Vps34 complex accompanied by the unchanged mTOR activity and positive regulation of LC3 during HBV

infection, which has also been reported in several research [159,160]. The phagophores undergo nucleation and elongation with Beclin-1/Vps34 complex and LC3 lipidation to form double-membraned autophagosomes [161], and the autophagic membranes generated during infection are used for viral envelopment [162]. However, rather than Beclin-1, only PI3KC3 was activated by the direct interconnection with HBx, which is in agreement with the study that the need for HBx to also induce the expression of Beclin-1 is redundant [163]. Moreover, although emerging study revealed that HBx efficiently impedes autophagic degradation by disturbing lysosomal acidification relevant to its degradative capacity without influencing on the fusion of autophagosomes and lysosomes [160], HBx-induced reduction of HDAC6 regulation ability might lead to autophagosome maturation failure [103]. Despite the necessity of further validation through biological experiments, this discovery provides possible mechanistic insight into the overall observation of HBV infection. Firstly, the intrinsic ability of HDAC6 to deacetylate NF- κ B, having been detected to hamper the invasion of cancer [164,165], was destructed by HBx through transrepression, which might gradually create a potent microenvironment favorable to the malignant transformation of cells. Meanwhile, the negative-regulation of HDAC6 in human HCCs and consequential loss of its tumor suppression ability supported by previous research [166] apparently advocate the hypothesis.

4.3. Inflammation and Innate Immune Response in HBV Infection

Inflammation and innate immune response represent a highly coordinated process aimed at fighting infection or tissue injury. When cells are subject to foreign invasion, the inflammatory response can ensure successful resolution of the condition and the restoration of tissue homeostasis. However, inappropriately controlled natural defense mechanisms will eventually lead to the progression of chronic diseases [167]. In general, innate immunity counts for a great deal in governing infection right after contact with the pathogen to limit the spread of the virus [168]. However, the exploitation of immune and inflammatory signaling pathways enables viruses to subvert antiviral immunity and replicate in the hostile environment [169].

Of all the proteins involved in the activation and execution of inflammation response, NF- κ B stands out as being crucial for this process in diverse metazoan organisms. During HBV infection, NF- κ B-dependent transcription accelerated by HBx not only triggered the anti-apoptosis mechanism in favor of viral persistence as mentioned earlier but also augmented the inflammatory response leading to hepatitis and cell transformation in accordance with the indication of previous study [170]. Concurrently, emerging research has also uncovered that the dysregulated continual synthesis of IL-6, an identified target of NF- κ B, plays a pathological effect on chronic inflammation and autoimmunity [167], which obliquely emphasizes the magnitude of the correlation between abnormally regulated NF- κ B and excessive inflammation. Besides, regarding viral immunoregulation, even though it is contradictory that HBx on one hand sensitized cells to inflammatory stimuli, but on the other positive-regulated PD-L1 through transactivation of c-Jun and activation of STAT3, the findings are consistent with the idea that production of PD-L1 transcripts pertained positively to the intensity of liver inflammation to prohibit and thereby evade the host immune response [171,172].

Moreover, unlike other proteins in their family which typically share common characteristics and functions, STAT1 and STAT3 poles apart and the mutual functional antagonism of them in T-cell-induced inflammation has also been elucidated [173]. Despite both serving as the target of miR-122, *SOCS1* was regulated by miR-122 and thus transactivated by HBx in the identified mechanism during HBV infection, while *SOCS3* did not. Therefore, we suggest that the abrogation of STATs activity by HBx to attenuate the interferon-mediated suppression of infection depends on STAT1 instead of STAT3. Meanwhile, blockades of IL-6- and IFNs-stimulated immune signaling pathways by viral proteins were through STAT1 rather than STAT3, which also supports the idea from another perspective. In addition, activation of STAT3 mediated by HBx also bolstered its unshakeable status on maintaining viral persistence. Likewise, inactivation of STAT1 mediated by HBV proteins counteracted its effects on the provocation of the innate antiviral system. Such phenomena have also been revealed in the

evidence of numerous studies [174]. It is worth noting that the reported capacity of HBx to block STAT1 nuclear import by means of affecting its methylation status was partially through the communication with CBP in the identified host/pathogen cross-talk. In such a way, that STAT1 lost its ability to block cell cycle progression and hinder hepatoma cell transformation in a course of prolonged inflammatory injury. This led to continuous destruction and regeneration of hepatocytes, increasing the chances of genetic alterations.

4.4. Discovery of Potential Drug Combination

Based on the significant biomarkers including AKT1, NFKBIA, EIF4EBP1, HDAC6, STAT1, STAT3, and TP53, the combination of agreeable multi-target drugs, i.e., Sorafenib, Nutlin-3, and Tenofovir, was eventually recommended as a potential treatment for later clinical researches of the multi-drug therapy treating HBV-infected patients. Meanwhile, to gain deeper insight into the discovery of promising drugs, *in silico* profiling using deep learning-based DTI model was performed to predict interactions (dockings) between available drugs and the identified biomarkers. Additionally, deciding whether a drug is ready for clinical trials greatly pertains to preclinical studies that yield preliminary efficacy, toxicity, safety, etc., information. Therefore, to further narrow down the scope and improve the reliability of the predicted drugs, the general pharmacological specifications for drug design including regulation ability, toxicity, and sensitivity were additionally adopted to evaluate the efficacy.

However, the entire process of moving a drug to clinical trials is actually a long way to go. Better preclinical preparation and assessment can be beneficial to the approval of a new medicine or drug combination. Considering the common reasons for withdrawal of an approved drug, e.g., hepatotoxicity and adverse events as well as the safety and tolerability assessment of a drug such as HLA (human leukocyte antigen) test, is conducive to better finding an agreeable option for drug combination. Repositioning drugs to treat with both common and rare disease is gradually becoming an attractive proposition. Taking advantage of drug repositioning avenues provides a faster discovery and translation of clinically relevant drug combinations. Furthermore, aided with the expediting development of computational technology and low-cost sequencing approaches, mounting publicly accessible data will certainly open up a broad spectrum of drug discovery and repositioning. Consequently, by systematically integrating more extensive data and information for both drugs and targets, a more appropriate drug combination out of the scope of original medical indication can be unveiled to bring new hope to HBV therapeutics.

5. Conclusions

The advent of the genomic eras has presented researchers with a myriad of high throughput biological data, which spark the interests in diverse biology applications. In this study, to investigate the pathogenetic mechanism under HBV infection, we constructed candidate host/pathogen interspecies genetic and epigenetic interaction network (HPI-GEN) by big data mining. Then, with the help of the two-side RNA-seq data, system identification strategies were applied to trim the false positives from the candidate HPI-GEN to obtain the real HPI-GEN. Thereafter, based on the extraction of PNP and the annotation of KEGG pathways, interspecies cross-talk signaling pathways are investigated from the real HPI-GEN for pathogenetic mechanism under HBV infection to identify significant biomarkers. Moreover, in order to discover promising drugs for the identified drug targets, we trained a deep learning-based DTI model to predict possible drug-target interactions. Eventually, with the consideration of general drug design specifications including regulation ability, toxicity, and sensitivity, a combination of multi-target drugs as a potential multiple-molecule drug was selected to abate HBV infection. It is worth pointing out that although only pathogenesis for HBV infection was investigated in this work, our unique workflow can also be utilized on a wide variety of disease in view of systems biology, holding utility to aid in diversified drug discovery and repurposing process. Meanwhile, along with the development of NGS (Next-Generation Sequencing) technology, additionally integrating analysis from available genomics data into our pipeline can reinforce better downstream analysis and

perform more exact biological interpretation. As more extensive genomic and pharmacological data being considered, the proposed pipeline could help reveal a more exhaustive host/pathogen offensive and defensive mechanism under HBV infection and facilitate the development of optimal therapy ensuring greater therapeutic outcome.

Author Contributions: Conceptualization, S.C. and B.-S.C.; Methodology, S.C. and B.-S.C.; Software, S.C. and B.-S.C.; Validation S.C. and B.-S.C.; Formal Analysis, S.C. and B.-S.C.; Investigation, S.C. and B.-S.C.; Data Curation, S.C.; Writing-original Draft Preparation, S.C.; Writing-Review and Editing, S.C., L.H.-C.W., and B.-S.C.; Visualization, S.C.; Supervision, B.-S.C.; Funding Acquisition, B.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Technology grant number MOST 107-2221-E-007-112-MY3.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In order to avoid overfitting in the parameter estimation process, the cubic spline interpolation method was employed to interpolate 4 time points into 1440 points. Even if the cubic interpolation couldn't increase the information for the parameter estimation, it could indeed prevent the parameters estimation in Equations (13) and (14) and (27)–(30) from overfitting during the identification process.

References

1. Chang, M.-H. Hepatitis B virus infection. *Semin. Fetal Neonatal Med.* **2007**, *12*, 160–167. [[CrossRef](#)] [[PubMed](#)]
2. *World Health Organization Global Hepatitis Report*; World Health Organization: Geneva, Switzerland, 2017; Available online: <http://apps.who.int/iris/bitstream/handle/10665/255016/9789241565455-eng.pdf?sequence=1> (accessed on 18 March 2020).
3. Hong, X.; Kim, E.S.; Guo, H. Epigenetic regulation of hepatitis B virus covalently closed circular DNA: Implications for epigenetic therapy against chronic hepatitis B. *Hepatology* **2017**, *66*, 2066–2077. [[CrossRef](#)] [[PubMed](#)]
4. Fanning, G.C.; Zoulim, F.; Hou, J.; Bertolotti, A. Therapeutic strategies for hepatitis B virus infection: Towards a cure. *Nat. Rev. Drug Discov.* **2019**, *18*, 827–844. [[CrossRef](#)]
5. Gehring, A.J.; Protzer, U. Targeting innate and adaptive immune responses to cure chronic HBV Infection. *Gastroenterology* **2019**, *156*, 325–337. [[CrossRef](#)]
6. Pol, S.; Nalpas, B.; Driss, F.; Michel, M.-L.; Tiollais, P.; Denis, J.; Bréchet, C. Efficacy and limitations of a specific immunotherapy in chronic hepatitis B. *J. Hepatol.* **2001**, *34*, 917–921. [[CrossRef](#)]
7. Shih, C.; Chou, S.-F.; Yang, C.-C.; Huang, J.-Y.; Chojijlsuren, G.; Jhou, R.-S. Control and eradication strategies of hepatitis B virus. *Trends Microbiol.* **2016**, *24*, 739–749. [[CrossRef](#)]
8. Nassal, M. HBV cccDNA: Viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut* **2015**, *64*, 1972. [[CrossRef](#)]
9. Nosengo, N. Can you teach old drugs new tricks? *Nature* **2016**, *534*, 314–316. [[CrossRef](#)]
10. Scannell, J.W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **2012**, *11*, 191–200.
11. Rao, V.S.; Srinivas, K.; Sujini, G.N.; Kumar, G.N.S. Protein-protein interaction detection: Methods and analysis. *Int J. Proteom.* **2014**, *2014*, 147648. [[CrossRef](#)] [[PubMed](#)]
12. Droit, A.; Poirier, G.; Hunter, J. Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. *J. Mol. Endocrinol.* **2005**, *34*, 263–280. [[CrossRef](#)] [[PubMed](#)]
13. Ashburn, T.T.; Thor, K.B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683. [[CrossRef](#)]
14. Antolin, A.A.; Workman, P.; Mestres, J.; Al-Lazikani, B. Polypharmacology in precision oncology: Current applications and future prospects. *Curr. Pharm. Des.* **2016**, *22*, 6935–6945. [[CrossRef](#)] [[PubMed](#)]
15. Mueller, H.; Wildum, S.; Luangsay, S.; Walther, J.; Lopez, A.; Tropberger, P.; Ottaviani, G.; Lu, W.; Parrott, N.; Zhang, J.D.; et al. A novel orally available small molecule that inhibits hepatitis B virus expression. *J. Hepatol.* **2017**, *68*, 412–420. [[CrossRef](#)] [[PubMed](#)]

16. Stark, C.; Breitkreutz, B.-J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [[CrossRef](#)]
17. Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The database of interacting proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32*, D449–D451. [[CrossRef](#)]
18. Bader, G.D.; Betel, D.; Hogue, C.W.V. BIND: The biomolecular interaction network database. *Nucleic Acids Res.* **2003**, *31*, 248–250. [[CrossRef](#)]
19. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [[CrossRef](#)]
20. Licata, L.; Briganti, L.; Peluso, D.; Perfetto, L.; Iannuccelli, M.; Galeota, E.; Sacco, F.; Palma, A.; Nardozza, A.P.; Santonico, E.; et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **2012**, *40*, D857–D861. [[CrossRef](#)]
21. Min, H.; Yoon, S. Got target? Computational methods for microRNA target prediction and their extension. *Exp. Mol. Med.* **2010**, *42*, 233–244. [[CrossRef](#)]
22. Friard, O.; Re, A.; Taverna, D.; De Bortoli, M.; Corá, D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinform.* **2010**, *11*, 435. [[CrossRef](#)] [[PubMed](#)]
23. Zheng, G.; Tu, K.; Yang, Q.; Xiong, Y.; Wei, C.; Xie, L.; Zhu, Y.; Li, Y. ITFP: An integrated platform of mammalian transcription factors. *Bioinformatics* **2008**, *24*, 2416–2417. [[CrossRef](#)] [[PubMed](#)]
24. Bovolenta, L.A.; Acencio, M.L.; Lemke, N. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genom.* **2012**, *13*, 405. [[CrossRef](#)] [[PubMed](#)]
25. Wingender, E.; Chen, X.; Hehl, R.; Karas, H.; Liebich, I.; Matys, V.; Meinhardt, T.; Prüss, M.; Reuter, I.; Schacherer, F. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res.* **2000**, *28*, 316–319. [[CrossRef](#)]
26. Calderone, A.; Licata, L.; Cesareni, G. VirusMentha: A new resource for virus-host protein interactions. *Nucleic Acids Res.* **2014**, *43*, D588–D592. [[CrossRef](#)]
27. Orchard, S.; Kerrien, S.; Abbani, S.; Aranda, B.; Bhate, J.; Bidwell, S.; Bridge, A.; Briganti, L.; Brinkman, F.S.L.; Cesareni, G.; et al. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat. Methods* **2012**, *9*, 345–350. [[CrossRef](#)]
28. del-Toro, N.; Dumousseau, M.; Orchard, S.; Jimenez, R.C.; Galeota, E.; Launay, G.; Goll, J.; Breuer, K.; Ono, K.; Salwinski, L.; et al. A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.* **2013**, *41*, W601–W606. [[CrossRef](#)]
29. Wu, Z.-J.; Zhu, Y.; Huang, D.-R.; Wang, Z.-Q. Constructing the HBV-human protein interaction network to understand the relationship between HBV and hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **2010**, *29*, 146. [[CrossRef](#)]
30. Qureshi, A.; Thakur, N.; Monga, I.; Thakur, A.; Kumar, M. VIRmiRNA: A comprehensive resource for experimentally validated viral miRNAs and their targets. *Database* **2014**. [[CrossRef](#)]
31. Li, C.-W.; Wang, W.-H.; Chen, B.-S. Investigating the specific core genetic-and-epigenetic networks of cellular mechanisms involved in human aging in peripheral blood mononuclear cells. *Oncotarget* **2016**, *7*, 8556–8579. [[CrossRef](#)]
32. Xiao, F.; Zuo, Z.; Cai, G.; Kang, S.; Gao, X.; Li, T. miRecords: An integrated resource for microRNA–target interactions. *Nucleic Acids Res.* **2008**, *37*, D105–D110. [[CrossRef](#)] [[PubMed](#)]
33. Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2013**, *42*, D92–D97. [[CrossRef](#)] [[PubMed](#)]
34. Hsu, S.-D.; Lin, F.-M.; Wu, W.-Y.; Liang, C.; Huang, W.-C.; Chan, W.-L.; Tsai, W.-T.; Chen, G.-Z.; Lee, C.-J.; Chiu, C.-M.; et al. miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **2010**, *39*, D163–D169. [[CrossRef](#)] [[PubMed](#)]
35. Kim, D.; Lee, J.; So, C.; Jeon, H.; Jeong, M.; Choi, Y.; Yoon, W.; Sung, M.; Kang, J. A neural named entity recognition and multi-type normalization tool for biomedical text mining. *IEEE Access* **2019**, *7*, 1. [[CrossRef](#)]
36. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2019**, *1*, 36. [[CrossRef](#)] [[PubMed](#)]

37. Liu, H.; Hu, Z.-Z.; Zhang, J.; Wu, C. BioThesaurus: A web-based thesaurus of protein and gene names. *Bioinformatics* **2005**, *22*, 103–105. [[CrossRef](#)] [[PubMed](#)]
38. Vogel, C.; Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **2012**, *13*, 227–232. [[CrossRef](#)]
39. Eulalio, A.; Huntzinger, E.; Izaurralde, E. Getting to the root of miRNA-Mediated gene silencing. *Cell* **2008**, *132*, 9–14. [[CrossRef](#)]
40. Zhang, S.; Pointer, D.; Singer, G.; Feng, Y.; Park, K.; Zhao, L.J. Direct binding to nucleic acids by Vpr of human immunodeficiency virus type 1. *Gene* **1998**, *212*, 157–166. [[CrossRef](#)]
41. Williams, J.S.; Andrisani, O.M. The hepatitis B virus X protein targets the basic region-leucine zipper domain of CREB. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3819. [[CrossRef](#)]
42. Cheong, J.H.; Yi, M.; Lin, Y.; Murakami, S. Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human hepatitis B virus X protein and may play a role in X transactivation. *EMBO J.* **1995**, *14*, 143–150. [[CrossRef](#)] [[PubMed](#)]
43. Maguire, H.F.; Hoeffler, J.P.; Siddiqui, A. HBV X protein alters the DNA binding specificity of CREB and ATF-2 by protein-protein interactions. *Science* **1991**, *252*, 842. [[CrossRef](#)] [[PubMed](#)]
44. Kuzhandaivelu, N.; Cong, Y.S.; Inouye, C.; Yang, W.M.; Seto, E. XAP2, a novel hepatitis B virus X-associated protein that inhibits X transactivation. *Nucleic Acids Res.* **1996**, *24*, 4741–4750. [[CrossRef](#)] [[PubMed](#)]
45. Ganem, D.; Prince, A.M. Hepatitis B virus infection—natural history and clinical consequences. *N. Engl. J. Med.* **2004**, *350*, 1118–1129. [[CrossRef](#)]
46. Mitra, B.; Thapa, R.J.; Guo, H.; Block, T.M. Host functions used by hepatitis B virus to complete its life cycle: Implications for developing host-targeting agents to treat chronic hepatitis B. *Antivir. Res.* **2018**, *158*, 185–198. [[CrossRef](#)]
47. Walsh, D.; Mohr, I. Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* **2011**, *9*, 860–875. [[CrossRef](#)]
48. Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **2007**, *25*, 117–124. [[CrossRef](#)]
49. Anderson, L.; Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **1997**, *18*, 533–537. [[CrossRef](#)]
50. de Sousa Abreu, R.; Penalva, L.O.; Marcotte, E.M.; Vogel, C. Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **2009**, *5*, 1512–1526. [[CrossRef](#)]
51. Maier, T.; Güell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583*, 3966–3973. [[CrossRef](#)]
52. Wang, Y.C.; Chen, B.S. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Syst. Biol.* **2010**, *4*, 20. [[CrossRef](#)] [[PubMed](#)]
53. The UniProt, C. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2018**, *47*, D506–D515. [[CrossRef](#)] [[PubMed](#)]
54. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; et al. DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035–D1041. [[CrossRef](#)] [[PubMed](#)]
55. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [[CrossRef](#)] [[PubMed](#)]
56. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)] [[PubMed](#)]
57. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **2006**, *35*, D198–D201. [[CrossRef](#)] [[PubMed](#)]
58. Khan, S.A.; Virtanen, S.; Kallioniemi, O.P.; Wennerberg, K.; Poso, A.; Kaski, S. Identification of structural features in chemicals associated with cancer drug response: A systematic data-driven analysis. *Bioinform. Oxf. Engl.* **2014**, *30*, i497–i504. [[CrossRef](#)]
59. Nandy, A.; Harle, M.; Basak, S. Mathematical descriptors of DNA sequences: Development and applications. *Gen. Pap. Ark.* **2006**, *9*, 211–238.

60. Gron, A. *Hands-On Machine Learning with Scikit-Learn, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Newton, MA, USA, 2017.
61. Raschka, S. *Python Machine Learning*; Packt Publishing: Birmingham, UK, 2015.
62. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [[CrossRef](#)]
63. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016.
64. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
65. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
66. Tiollais, P.; Pourcel, C.; Dejean, A. The hepatitis B virus. *Nature* **1985**, *317*, 489–495. [[CrossRef](#)] [[PubMed](#)]
67. Ganem, D.; Varmus, H.E. THE molecular biology of the hepatitis B viruses. *Annu. Rev. Biochem.* **1987**, *56*, 651–693. [[CrossRef](#)] [[PubMed](#)]
68. Tuttleman, J.S.; Pourcel, C.; Summers, J. Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. *Cell* **1986**, *47*, 451–460. [[CrossRef](#)]
69. Zakaria, M.K.; Sankhyani, A.; Ali, A.; Fatima, K.; Azhar, E.; Qadri, I. HBV/HCV Infection and Inflammation. *J. Genet. Syndr. Gene Ther.* **2014**, *5*, 241.
70. Chen, L.; Deng, H.; Cui, H.; Fang, J.; Zuo, Z.; Deng, J.; Li, Y.; Wang, X.; Zhao, L. Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget* **2017**, *9*, 7204–7218. [[CrossRef](#)]
71. Garcia-Gomez, A.; Rodríguez-Ubrevia, J.; Ballestar, E. Epigenetic interplay between immune, stromal and cancer cells in the tumor microenvironment. *Clin. Immunol.* **2018**, *196*, 64–71. [[CrossRef](#)]
72. Ma, Z.; Cao, Q.; Xiong, Y.; Zhang, E.; Lu, M. Interaction between hepatitis B virus and toll-like receptors: Current status and potential therapeutic use for chronic hepatitis B. *Vaccines* **2018**, *6*, 6. [[CrossRef](#)]
73. Jacobs, M.D.; Harrison, S.C. Structure of an I κ B α /NF- κ B Complex. *Cell* **1998**, *95*, 749–758. [[CrossRef](#)]
74. Tanaka, T.; Narazaki, M.; Kishimoto, T. IL-6 in inflammation, immunity, and disease. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016295. [[CrossRef](#)]
75. Lee, Y.; Park, U.S.; Choi, I.; Yoon, S.K.; Park, Y.M.; Lee, Y.I. Human interleukin 6 gene is activated by hepatitis B virus-X protein in human hepatoma cells. *Clin. Cancer Res.* **1998**, *4*, 1711. [[PubMed](#)]
76. Ivashkiv, L.B.; Donlin, L.T. Regulation of type I interferon responses. *Nat. Rev. Immunol.* **2014**, *14*, 36–49. [[CrossRef](#)]
77. Schoenborn, J.; Wilson, C. Regulation of Interferon- γ During Innate and Adaptive Immune Responses. *Adv. Immunol.* **2007**, *96*, 41–101. [[PubMed](#)]
78. Nolan, G.P.; Fujita, T.; Bhatia, K.; Huppi, C.; Liou, H.C.; Scott, M.L.; Baltimore, D. The bcl-3 proto-oncogene encodes a nuclear I kappa B-like molecule that preferentially interacts with NF-kappa B p50 and p52 in a phosphorylation-dependent manner. *Mol. Cell. Biol.* **1993**, *13*, 3557. [[CrossRef](#)] [[PubMed](#)]
79. Kashatus, D.; Cogswell, P.; Baldwin, A.S. Expression of the Bcl-3 proto-oncogene suppresses p53 activation. *Genes Dev.* **2006**, *20*, 225–235. [[CrossRef](#)] [[PubMed](#)]
80. Dann, S.G.; Thomas, G. The amino acid sensitive TOR pathway from yeast to mammals. *FEBS Lett.* **2006**, *580*, 2821–2829. [[CrossRef](#)]
81. Lee, D.-F.; Kuo, H.-P.; Chen, C.-T.; Hsu, J.-M.; Chou, C.-K.; Wei, Y.; Sun, H.-L.; Li, L.-Y.; Ping, B.; Huang, W.-C.; et al. IKK β suppression of TSC1 Links inflammation and tumor angiogenesis via the Mtor pathway. *Cell* **2007**, *130*, 440–455. [[CrossRef](#)]
82. Richter, J.D.; Sonenberg, N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **2005**, *433*, 477–480. [[CrossRef](#)]
83. Mohr, I. Phosphorylation and dephosphorylation events that regulate viral mRNA translation. *Virus Res.* **2006**, *119*, 89–99. [[CrossRef](#)]
84. Ryu, D.-K.; Ahn, B.-Y.; Ryu, W.-S. Proximity between the cap and 5' ϵ stem-loop structure is critical for the suppression of pgRNA translation by the hepatitis B viral polymerase. *Virology* **2010**, *406*, 56–64. [[CrossRef](#)]
85. Sugatani, T.; Hruska, K.A. Akt1/Akt2 and mammalian target of rapamycin/Bim play critical roles in osteoclast differentiation and survival, respectively, whereas Akt is dispensable for cell survival in isolated osteoclast precursors. *J. Biol. Chem.* **2005**, *280*, 3583–3589. [[CrossRef](#)] [[PubMed](#)]
86. Inoki, K.; Li, Y.; Zhu, T.; Wu, J.; Guan, K.-L. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat. Cell Biol.* **2002**, *4*, 648–657. [[CrossRef](#)] [[PubMed](#)]

87. Ogawara, Y.; Kishishita, S.; Obata, T.; Isazawa, Y.; Suzuki, T.; Tanaka, K.; Masuyama, N.; Gotoh, Y. Akt Enhances Mdm2-mediated Ubiquitination and Degradation of p53. *J. Biol. Chem.* **2002**, *277*, 21843–21850. [[CrossRef](#)] [[PubMed](#)]
88. Chiu, A.P.; Tschida, B.R.; Sham, T.-T.; Lo, L.H.; Moriarity, B.S.; Li, X.-X.; Lo, R.C.; Hinton, D.E.; Rowlands, D.K.; Chan, C.-O.; et al. HBx-K130M/V131I promotes liver cancer in transgenic mice via AKT/FOXO1 signaling pathway and arachidonic acid metabolism. *Mol. Cancer Res.* **2019**, *17*, 1582. [[CrossRef](#)]
89. Burton, T.R.; Gibson, S.B. The role of Bcl-2 family member BNIP3 in cell death and disease: NIPping at the heels of cell death. *Cell Death Differ.* **2009**, *16*, 515–523. [[CrossRef](#)]
90. Liu, J.; Cao, L.; Chen, J.; Song, S.; Lee, I.H.; Quijano, C.; Liu, H.; Keyvanfar, K.; Chen, H.; Cao, L.-Y.; et al. Bmi1 regulates mitochondrial function and the DNA damage response pathway. *Nature* **2009**, *459*, 387–392. [[CrossRef](#)]
91. Alkema, M.; Wiegant, J.; Raap, A.K.; Bems, A.; van Lohuizen, M. Characterization and chromosomal localization of the human proto-oncogene BMI-1. *Hum. Mol. Genet.* **1993**, *2*, 1597–1603. [[CrossRef](#)]
92. Danial, N.N.; Korsmeyer, S.J. Cell Death: Critical Control Points. *Cell* **2004**, *116*, 205–219. [[CrossRef](#)]
93. Oh, H.-M.; Yu, C.-R.; Dambuza, I.; Marrero, B.; Egwuagu, C.E. STAT3 protein interacts with Class O Forkhead transcription factors in the cytoplasm and regulates nuclear/cytoplasmic localization of FoxO1 and FoxO3a proteins in CD4(+) T cells. *J. Biol. Chem.* **2012**, *287*, 30436–30443. [[CrossRef](#)]
94. Shlomai, A.; Shaul, Y. The metabolic activator FOXO1 binds hepatitis B virus DNA and activates its transcription. *Biochem. Biophys. Res. Commun.* **2009**, *381*, 544–548. [[CrossRef](#)]
95. Hirata, Y.; Takahashi, M.; Morishita, T.; Noguchi, T.; Matsuzawa, A. Post-Translational modifications of the TAK1-TAB complex. *Int J. Mol. Sci.* **2017**, *18*, 205. [[CrossRef](#)] [[PubMed](#)]
96. Shaulian, E.; Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* **2002**, *4*, E131–E136. [[CrossRef](#)] [[PubMed](#)]
97. Schonhaler, H.B.; Guinea-Viniegra, J.; Wagner, E.F. Targeting inflammation by modulating the Jun/AP-1 pathway. *Ann. Rheum. Dis.* **2011**, *70*, i109. [[CrossRef](#)] [[PubMed](#)]
98. Akira, S. The role of IL-18 in innate immunity. *Curr. Opin. Immunol.* **2000**, *12*, 59–63. [[CrossRef](#)]
99. Qin, W.; Hu, L.; Zhang, X.; Jiang, S.; Li, J.; Zhang, Z.; Wang, X. The diverse function of PD-1/PD-L pathway beyond cancer. *Front. Immunol.* **2019**, *10*, 2298. [[CrossRef](#)] [[PubMed](#)]
100. Motokura, T.; Bloom, T.; Kim, H.G.; Jüppner, H.; Ruderman, J.V.; Kronenberg, H.M.; Arnold, A. A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature* **1991**, *350*, 512–515. [[CrossRef](#)] [[PubMed](#)]
101. Fujioka, S.; Niu, J.; Schmidt, C.; Sclabas, G.M.; Peng, B.; Uwagawa, T.; Li, Z.; Evans, D.B.; Abbruzzese, J.L.; Chiao, P.J. NF- κ B and AP-1 connection: Mechanism of NF- κ B-Dependent regulation of AP-1 activity. *Mol. Cell. Biol.* **2004**, *24*, 7806. [[CrossRef](#)] [[PubMed](#)]
102. Rebollo, A.; Dumoutier, L.; Renauld, J.C.; Zaballos, A.; Ayllón, V.; Martínez-A, C. Bcl-3 expression promotes cell survival following interleukin-4 deprivation and is controlled by AP1 and AP1-like transcription factors. *Mol. Cell. Biol.* **2000**, *20*, 3407–3416. [[CrossRef](#)]
103. Lee, J.-Y.; Koga, H.; Kawaguchi, Y.; Tang, W.; Wong, E.; Gao, Y.-S.; Pandey, U.B.; Kaushik, S.; Tresse, E.; Lu, J.; et al. HDAC6 controls autophagosome maturation essential for ubiquitin-selective quality-control autophagy. *EMBO J.* **2010**, *29*, 969–980. [[CrossRef](#)]
104. Aaronson, D.S.; Horvath, C.M. A road map for those who don't know JAK-STAT. *Science* **2002**, *296*, 1653–1655. [[CrossRef](#)]
105. Stark, G.R.; Kerr, I.M.; Williams, B.R.; Silverman, R.H.; Schreiber, R.D. How cells respond to interferons. *Annu. Rev. Biochem.* **1998**, *67*, 227–264. [[CrossRef](#)] [[PubMed](#)]
106. Ferrari, S.M.; Fallahi, P.; Ruffilli, I.; Elia, G.; Ragusa, F.; Paparo, S.R.; Patrizio, A.; Mazzi, V.; Colaci, M.; Giuggioli, D.; et al. Immunomodulation of CXCL10 Secretion by hepatitis C Virus: Could CXCL10 be a prognostic marker of chronic hepatitis C? *J. Immunol. Res.* **2019**, *2019*, 5878960. [[CrossRef](#)] [[PubMed](#)]
107. Park, J.-W.; Li, Z.; Choi, J.-S.; Oh, H.-J.; Park, S.-H.; Yoon, K.-C. Expression of CXCL9, -10, and -11 in the aqueous humor of patients with herpetic endotheliitis. *Cornea* **2012**, *31*, 1246–1250. [[CrossRef](#)] [[PubMed](#)]
108. Nordmann, A.; Wixler, L.; Boergeling, Y.; Wixler, V.; Ludwig, S. A new splice variant of the human guanylate-binding protein 3 mediates anti-influenza activity through inhibition of viral transcription and replication. *FASEB J.* **2011**, *26*, 1290–1300. [[CrossRef](#)]

109. Ueda, T.; Watanabe-Fukunaga, R.; Fukuyama, H.; Nagata, S.; Fukunaga, R. Mnk2 and Mnk1 are essential for constitutive and inducible phosphorylation of eukaryotic initiation factor 4E but not for cell growth or development. *Mol. Cell. Biol.* **2004**, *24*, 6539–6549. [[CrossRef](#)]
110. Porter, A.G.; Jänicke, R.U. Emerging roles of caspase-3 in apoptosis. *Cell Death Differ.* **1999**, *6*, 99–104. [[CrossRef](#)]
111. Stennicke, H.; Jürgensmeier, J.; Shin, H.; Deveraux, Q.; Wolf, B.; Yang, X.; Zhou, Q.; Ellerby, H.; Ellerby, L.; Bredesen, D.; et al. Pro-caspase-3 Is a Major Physiologic Target of Caspase8. *J. Biol. Chem.* **1998**, *273*, 27084–27090. [[CrossRef](#)]
112. Wang, J.; Chun, H.J.; Wong, W.; Spencer, D.M.; Lenardo, M.J. Caspase-10 is an initiator caspase in death receptor signaling. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13884. [[CrossRef](#)]
113. Dikic, I.; Elazar, Z. Mechanism and medical implications of mammalian autophagy. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 349–364. [[CrossRef](#)]
114. Gottlieb, T.M.; Jackson, S.P. The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell* **1993**, *72*, 131–142. [[CrossRef](#)]
115. Liu, X.; Zou, H.; Slaughter, C.; Wang, X. DFF, a Heterodimeric Protein That Functions Downstream of Caspase-3 to Trigger DNA Fragmentation during Apoptosis. *Cell* **1997**, *89*, 175–184. [[CrossRef](#)]
116. Dan, H.; Sun, M.; Kaneko, S.; Feldman, R.; Nicosia, S.; Wang, H.-G.; Tsang, B.; Cheng, J. Akt Phosphorylation and Stabilization of X-linked Inhibitor of Apoptosis Protein (XIAP). *J. Biol. Chem.* **2004**, *279*, 5405–5412. [[CrossRef](#)]
117. Scott, F.L.; Denault, J.-B.; Riedl, S.J.; Shin, H.; Ratus, M.; Salvesen, G.S. XIAP inhibits caspase-3 and -7 using two binding sites: Evolutionarily conserved mechanism of IAPs. *EMBO J.* **2005**, *24*, 645–655. [[CrossRef](#)] [[PubMed](#)]
118. Ahnesorg, P.; Smith, P.; Jackson, S.P. XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* **2006**, *124*, 301–313. [[CrossRef](#)] [[PubMed](#)]
119. Wang, S.; Qiu, L.; Yan, X.; Jin, W.; Wang, Y.; Chen, L.; Wu, E.; Ye, X.; Gao, G.F.; Wang, F.; et al. Loss of microRNA 122 expression in patients with hepatitis B enhances hepatitis B virus replication through cyclin G1-modulated P53 activity. *Hepatology* **2012**, *55*, 730–741. [[CrossRef](#)] [[PubMed](#)]
120. Wang, X.W.; Gibson, M.K.; Vermeulen, W.; Yeh, H.; Forrester, K.; Stürzbecher, H.W.; Hoeijmakers, J.H.; Harris, C.C. Abrogation of p53-induced apoptosis by the hepatitis B virus X gene. *Cancer Res.* **1995**, *55*, 6012–6016.
121. Lee, S.G.; Rho, H.M. Transcriptional repression of the human p53 gene by hepatitis B viral X protein. *Oncogene* **2000**, *19*, 468–471. [[CrossRef](#)]
122. Moll, U.M.; Petrenko, O. The MDM2-p53 Interaction. *Mol. Cancer Res.* **2003**, *1*, 1001.
123. Cregan, S.P.; MacLaurin, J.G.; Craig, C.G.; Robertson, G.S.; Nicholson, D.W.; Park, D.S.; Slack, R.S. Bax-Dependent Caspase-3 Activation Is a Key Determinant in p53-Induced Apoptosis in Neurons. *J. Neurosci.* **1999**, *19*, 7860. [[CrossRef](#)]
124. Chipuk, J.E.; Kuwana, T.; Bouchier-Hayes, L.; Droin, N.M.; Newmeyer, D.D.; Schuler, M.; Green, D.R. Direct activation of bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science* **2004**, *303*, 1010. [[CrossRef](#)]
125. Sofer, A.; Lei, K.; Johannessen, C.M.; Ellisen, L.W. Regulation of mTOR and cell growth in response to energy stress by REDD1. *Mol. Cell. Biol.* **2005**, *25*, 5834–5845. [[CrossRef](#)] [[PubMed](#)]
126. Ellisen, L.W.; Ramsayer, K.D.; Johannessen, C.M.; Yang, A.; Beppu, H.; Minda, K.; Oliner, J.D.; McKeon, F.; Haber, D.A. REDD1, a developmentally regulated transcriptional target of p63 and p53, links p63 to regulation of reactive oxygen species. *Mol. Cell* **2002**, *10*, 995–1005. [[CrossRef](#)]
127. Pineau, P.; Volinia, S.; McJunkin, K.; Marchio, A.; Battiston, C.; Terris, B.; Mazzaferro, V.; Lowe, S.W.; Croce, C.M.; Dejean, A. miR-221 overexpression contributes to liver tumorigenesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 264–269. [[CrossRef](#)] [[PubMed](#)]
128. Fu, X.; Tan, D.; Hou, Z.; Hu, Z.; Liu, G. miR-338-3p is down-regulated by hepatitis B virus X and inhibits cell proliferation by targeting the 3'-UTR region of CyclinD1. *Int. J. Mol. Sci.* **2012**, *13*, 8514–8539. [[CrossRef](#)] [[PubMed](#)]
129. Lu, R.; Yang, Z.; Xu, G.; Yu, S. miR-338 modulates proliferation and autophagy by PI3K/AKT/mTOR signaling pathway in cervical cancer. *Biomed. Pharmacother.* **2018**, *105*, 633–644. [[CrossRef](#)]

130. Darnell, J.E.; Kerr, I.M.; Stark, G.R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **1994**, *264*, 1415. [[CrossRef](#)]
131. Wu, M.; Xu, Y.; Lin, S.; Zhang, X.; Xiang, L.; Yuan, Z. Hepatitis B virus polymerase inhibits the interferon-inducible MyD88 promoter by blocking nuclear translocation of Stat 1. *J. Gen. Virol.* **2008**, *88*, 3260–3269. [[CrossRef](#)]
132. Gao, D.; Zhai, A.; Qian, J.; Li, A.; Li, Y.; Song, W.; Zhao, H.; Yu, X.; Wu, J.; Zhang, Q.; et al. Down-regulation of suppressor of cytokine signaling 3 by miR-122 enhances interferon-mediated suppression of hepatitis B virus. *Antivir. Res.* **2015**, *118*, 20–28. [[CrossRef](#)]
133. Li, A.; Song, W.; Qian, J.; Li, Y.; He, J.; Zhang, Q.; Li, W.; Zhai, A.; Kao, W.; Hu, Y.; et al. MiR-122 modulates type I interferon expression through blocking suppressor of cytokine signaling 1. *Int. J. Biochem. Cell Biol.* **2013**, *45*, 858–865. [[CrossRef](#)]
134. Xavier, G.; Antoine, B.; Yoshua, B. Deep sparse rectifier neural networks. *PMLR* **2011**, *15*, 315–323.
135. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
136. Cao, D.S.; Zhang, L.X.; Tan, G.S.; Xiang, Z.; Zeng, W.B.; Xu, Q.S.; Chen, A.F. Computational prediction of drug target interactions using chemical, biological, and network features. *Mol. Inform.* **2014**, *33*, 669–681. [[CrossRef](#)] [[PubMed](#)]
137. Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889. [[CrossRef](#)] [[PubMed](#)]
138. Liu, Y.; Wu, M.; Miao, C.; Zhao, P.; Li, X.-L. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.* **2016**, *12*, e1004760. [[CrossRef](#)] [[PubMed](#)]
139. LeBeau, J.E. The role of the LD50 determination in drug safety evaluation. *Regul. Toxicol. Pharmacol.* **1983**, *3*, 71–74. [[CrossRef](#)]
140. Wilhelm, S.; Carter, C.; Lynch, M.; Lowinger, T.; Dumas, J.; Smith, R.A.; Schwartz, B.; Simantov, R.; Kelley, S. Discovery and development of sorafenib: A multikinase inhibitor for treating cancer. *Nat. Rev. Drug Discov.* **2006**, *5*, 835–844. [[CrossRef](#)]
141. Ibrahim, N.; Yu, Y.; Walsh, W.R.; Yang, J.L. Molecular targeted therapies for cancer: Sorafenib mono-therapy and its combination with other therapies (review). *Oncol. Rep.* **2012**, *27*, 1303–1311.
142. Miao, R.; Xu, X.; Wang, Z.; Liu, S.; Qu, K.; Chen, W.; Liu, C. Synergistic effect of nutlin-3 combined with aspirin in hepatocellular carcinoma HepG2 cells through activation of Bcl-2/Bax signaling pathway. *Mol. Med. Rep.* **2018**, *17*, 3735–3743. [[CrossRef](#)]
143. Tokalov, S.V.; Abolmaali, N.D. Protection of p53 wild type cells from taxol by nutlin-3 in the combined lung cancer treatment. *BMC Cancer* **2010**, *10*, 57. [[CrossRef](#)]
144. Zheng, T.; Wang, J.; Song, X.; Meng, X.; Pan, S.; Jiang, H.; Liu, L. Nutlin-3 cooperates with doxorubicin to induce apoptosis of human hepatocellular carcinoma cells through p53 or p73 signaling pathways. *J. Cancer Res. Clin. Oncol* **2010**, *136*, 1597–1604. [[CrossRef](#)]
145. Fernandez-Fernandez, B.; Montoya-Ferrer, A.; Sanz, A.B.; Sanchez-Niño, M.D.; Izquierdo, M.C.; Poveda, J.; Sainz-Prestel, V.; Ortiz-Martin, N.; Parra-Rodriguez, A.; Selgas, R.; et al. Tenofovir nephrotoxicity: 2011 update. *AIDS Res. Treat.* **2011**, *2011*, 354908. [[CrossRef](#)] [[PubMed](#)]
146. Martin, P.; Lau, D.T.; Nguyen, M.H.; Janssen, H.L.; Dieterich, D.T.; Peters, M.G.; Jacobson, I.M. A Treatment algorithm for the management of chronic hepatitis b virus infection in the united states: 2015 update. *Clin. Gastroenterol. Hepatol.* **2015**, *13*, 2071–2087. [[CrossRef](#)]
147. Lee, Y.B.; Lee, J.-H.; Lee, D.H.; Cho, H.; Ahn, H.; Choi, W.-M.; Cho, Y.Y.; Lee, M.; Yoo, J.-J.; Cho, Y.; et al. Efficacy of entecavir-tenofovir combination therapy for chronic hepatitis b patients with multidrug-resistant strains. *Antimicrob. Agents Chemother.* **2014**, *58*, 6710. [[CrossRef](#)]
148. Marcellin, P.; Ahn, S.H.; Ma, X.; Caruntu, F.A.; Tak, W.Y.; Elkashab, M.; Chuang, W.-L.; Lim, S.-G.; Tabak, F.; Mehta, R.; et al. Combination of tenofovir disoproxil fumarate and peginterferon α -2a increases loss of hepatitis b surface antigen in patients with chronic hepatitis, B. *Gastroenterology* **2016**, *150*, 134–144. [[CrossRef](#)] [[PubMed](#)]
149. Fung, J.; Lai, C.L.; Seto, W.K.; Yuen, M.F. Nucleoside/nucleotide analogues in the treatment of chronic hepatitis B. *J. Antimicrob. Chemother.* **2011**, *66*, 2715–2725. [[CrossRef](#)] [[PubMed](#)]

150. Kew, M.C. Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **2011**, *26*, 144–152. [[CrossRef](#)]
151. Diao, J.; Khine, A.A.; Sarangi, F.; Hsu, E.; Iorio, C.; Tibbles, L.A.; Woodgett, J.R.; Penninger, J.; Richardson, C.D. X protein of hepatitis B virus inhibits Fas-mediated apoptosis and is associated with up-regulation of the SAPK/JNK pathway. *J. Biol. Chem.* **2001**, *276*, 8328–8340. [[CrossRef](#)]
152. Lee, Y.I.; Kang-Park, S.; Do, S.I.; Lee, Y.I. The hepatitis B virus-X protein activates a phosphatidylinositol 3-kinase-dependent survival signaling cascade. *J. Biol. Chem.* **2001**, *276*, 16969–16977. [[CrossRef](#)]
153. Gottlob, K.; Fulco, M.; Levrero, M.; Graessmann, A. The hepatitis B Virus HBx protein inhibits caspase 3 activity. *J. Biol. Chem.* **1999**, *273*, 33347–33353. [[CrossRef](#)]
154. Su, F.; Theodosis, C.N.; Schneider, R.J. Role of NF- κ B and myc proteins in apoptosis induced by hepatitis B virus HBx protein. *J. Virol.* **2001**, *75*, 215. [[CrossRef](#)]
155. Wang, T.; Wang, Y.; Wu, M.-C.; Guan, X.-Y.; Yin, Z.-F. Activating mechanism of transcription factor NF- κ B regulated by hepatitis B virus X protein in hepatocellular carcinoma. *World J. Gastroenterol.* **2004**, *10*, 356–360. [[CrossRef](#)] [[PubMed](#)]
156. Park, U.S.; Park, S.K.; Lee, Y.I.; Park, J.G.; Lee, Y.I. Hepatitis B virus-X protein upregulates the expression of p21waf1/cip1 and prolongs G1→S transition via a p53-independent pathway in human hepatoma cells. *Oncogene* **2000**, *19*, 3384–3394. [[CrossRef](#)]
157. Al-Anazi, M.R.; Nazir, N.; Colak, D.; Al-Ahdal, M.N.; Al-Qahtani, A.A. Deletion and functional analysis of hepatitis B virus X protein: Evidence for an effect on cell cycle regulators. *Cell. Physiol. Biochem.* **2018**, *49*, 1987–1998. [[CrossRef](#)] [[PubMed](#)]
158. Abbas, T.; Dutta, A. p21 in cancer: Intricate networks and multiple activities. *Nat. Rev. Cancer* **2009**, *9*, 400–414. [[CrossRef](#)]
159. Tang, H.; Da, L.; Mao, Y.; Li, Y.; Li, D.; Xu, Z.; Li, F.; Wang, Y.; Tiollais, P.; Li, T.; et al. Hepatitis B virus X protein sensitizes cells to starvation-induced autophagy via up-regulation of beclin 1 expression. *Hepatology* **2009**, *49*, 60–71. [[CrossRef](#)] [[PubMed](#)]
160. Liu, B.; Fang, M.; Hu, Y.; Huang, B.; Li, N.; Chang, C.; Huang, R.; Xu, X.; Yang, Z.; Chen, Z.; et al. Hepatitis B virus X protein inhibits autophagic degradation by impairing lysosomal maturation. *Autophagy* **2014**, *10*, 416–430. [[CrossRef](#)] [[PubMed](#)]
161. Rautou, P.-E.; Mansouri, A.; Lebrech, D.; Durand, F.; Valla, D.; Moreau, R. Autophagy in liver diseases. *J. Hepatol.* **2010**, *53*, 1123–1134. [[CrossRef](#)]
162. Li, J.; Liu, Y.; Wang, Z.; Liu, K.; Wang, Y.; Liu, J.; Ding, H.; Yuan, Z. Subversion of cellular autophagy machinery by hepatitis B virus for viral envelopment. *J. Virol.* **2011**, *85*, 6319. [[CrossRef](#)]
163. Sir, D.; Tian, Y.; Chen, W.-L.; Ann, D.K.; Yen, T.-S.B.; Ou, J.-H.J. The early autophagic pathway is activated by hepatitis B virus and required for viral DNA replication. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4383. [[CrossRef](#)]
164. Yang, C.-J.; Liu, Y.-P.; Dai, H.-Y.; Shiue, Y.-L.; Tsai, C.-J.; Huang, M.-S.; Yeh, Y.-T. Nuclear HDAC6 inhibits invasion by suppressing NF- κ B/MMP2 and is inversely correlated with metastasis of non-small cell lung cancer. *Oncotarget* **2015**, *6*, 30263–30276. [[CrossRef](#)]
165. Kim, M.; Lu, F.; Zhang, Y. Loss of HDAC-Mediated Repression and Gain of NF- κ B Activation Underlie Cytokine Induction in ARID1A- and PIK3CA-Mutation-Driven Ovarian Cancer. *Cell Rep.* **2016**, *17*, 275–288. [[CrossRef](#)] [[PubMed](#)]
166. Jung, K.H.; Noh, J.H.; Kim, J.K.; Eun, J.W.; Bae, H.J.; Chang, Y.G.; Kim, M.G.; Park, W.S.; Lee, J.Y.; Lee, S.Y.; et al. Histone deacetylase 6 functions as a tumor suppressor by activating c-Jun NH2-terminal kinase-mediated beclin 1-dependent autophagic cell death in liver cancer. *Hepatology* **2012**, *56*, 644–657. [[CrossRef](#)] [[PubMed](#)]
167. Jones, S.A. Directing transition from innate to acquired immunity: Defining a role for IL-6. *J. Immunol.* **2005**, *175*, 3463. [[CrossRef](#)] [[PubMed](#)]
168. Busca, A.; Kumar, A. Innate immune responses in hepatitis B virus (HBV) infection. *Virol. J.* **2014**, *11*, 22. [[CrossRef](#)] [[PubMed](#)]
169. Ong, E.Z.; Chan, K.R.; Ooi, E.E. viral manipulation of host inhibitory receptor signaling for immune evasion. *PLoS Pathog.* **2016**, *12*, e1005776. [[CrossRef](#)] [[PubMed](#)]
170. Luedde, T.; Schwabe, R.F. NF- κ B in the liver—linking injury, fibrosis and hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **2011**, *8*, 108–118. [[CrossRef](#)] [[PubMed](#)]
171. Schönrich, G.; Raftery, M.J. The PD-1/PD-L1 axis and virus infections: A delicate balance. *Front. Cell Infect. Microbiol.* **2019**, *9*, 207. [[CrossRef](#)]

172. Germanidis, G.; Argentou, N.; Hytioglou, P.; Vassiliadis, T.; Patsiaoura, K.; Germenis, A.; Speletas, M. Liver FOXP3 and PD1/PDL1 expression is down-regulated in chronic HBV hepatitis on maintained remission related to the degree of inflammation. *Front. Immunol.* **2013**, *4*, 207. [[CrossRef](#)]
173. Hong, F.; Jaruga, B.; Kim, W.H.; Radaeva, S.; El-Assal, O.N.; Tian, Z.; Nguyen, V.-A.; Gao, B. Opposing roles of STAT1 and STAT3 in T cell-mediated hepatitis: Regulation by SOCS. *J. Clin. Investig.* **2002**, *110*, 1503–1513. [[CrossRef](#)]
174. Dufour, J.-F.; Clavien, P.A. *Signaling Pathways in Liver Diseases*; Springer: New York, NY, USA, 2010; pp. 1–526.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).