

## Article

# Identification of Therapeutic Targets for the Selective Killing of HBV-Positive Hepatocytes

Chien-Jung Huang<sup>1</sup>, Lily Hui-Ching Wang<sup>2,3</sup> and Yu-Chao Wang<sup>1,\*</sup>

<sup>1</sup> Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei 112304, Taiwan; shana96300@gmail.com

<sup>2</sup> Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu 300044, Taiwan; lilywang@life.nthu.edu.tw

<sup>3</sup> Department of Medical Science, National Tsing Hua University, Hsinchu 300044, Taiwan

\* Correspondence: yuchao@ym.edu.tw

**Abstract:** The hepatitis B virus (HBV) infection is a major risk factor for cirrhosis and hepatocellular carcinoma. Most infected individuals become lifelong carriers of HBV as the drugs currently used to treat the patients can only control the disease, thereby achieving functional cure (loss of the hepatitis B surface antigen) but not complete cure (elimination of infected hepatocytes). Therefore, we aimed to identify the target genes for the selective killing of HBV-positive hepatocytes to develop a novel therapy for the treatment of HBV infection. Our strategy was to recognize the conditionally essential genes that are essential for the survival of HBV-positive hepatocytes, but non-essential for the HBV-negative hepatocytes. Using microarray gene expression data curated from the Gene Expression Omnibus database and the known essential genes from the Online GENE Essentiality database, we used two approaches, comprising the random walk with restart algorithm and the support vector machine approach, to determine the potential targets for the selective killing of HBV-positive hepatocytes. The final candidate genes list obtained using these two approaches consisted of 36 target genes, which may be conditionally essential for the cell survival of HBV-positive hepatocytes; however, this requires further experimental validation. Therefore, the genes identified in this study can be used as potential drug targets to develop novel therapeutic strategies for the treatment of HBV, and may ultimately help in achieving the elusive goal of a complete cure for hepatitis B.

**Keywords:** hepatitis B virus; hepatocytes; conditionally essential genes; weighted co-expression network; random walk with restart; support vector machine; therapeutic targets



**Citation:** Huang, C.-J.; Wang, L.H.-C.; Wang, Y.-C. Identification of Therapeutic Targets for the Selective Killing of HBV-Positive Hepatocytes. *J. Pers. Med.* **2021**, *11*, 649. <https://doi.org/10.3390/jpm11070649>

Academic Editor: Lorenza Rimassa

Received: 17 June 2021

Accepted: 8 July 2021

Published: 10 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hepatitis B virus (HBV) is a double-stranded DNA virus and a member of the family *Hepadnaviridae*. HBV infection causes hepatitis B, which is an infectious disease that leads to acute or chronic hepatitis. According to a report by the World Health Organization in 2015, an estimated 257 million people were living with chronic HBV infection globally, and the prevalence in adults in the sub-Saharan regions and East Asia was 5–10% [1,2]. In the case of HBV carriers, the HBV antigens, mainly hepatitis B surface antigen (HBsAg), can be detected in the blood of the patients. Based on the transmission of HBV, this infection can be divided into blood and body fluid infections, and the mode of infections can further be divided into vertical and horizontal infections. Vertical infection refers to the transfer of the infection from the HBV carrier mothers to their newborn babies via the placenta or birth canal, whereas horizontal infection refers to the infection caused by contact with the blood or body fluid of the carriers, for example, blood transfusion, injection, and the tattooing process. As vertical infection is the main source of infection, Taiwan has been administering the hepatitis B vaccine since 1984. The vaccination program has effectively blocked the vertical infection of hepatitis B, and the prevalence of hepatitis B among those

younger than 20 years old decreased from 9.8% in 1984 to 0.6% in 2004 [3]. However, the carrier rate of hepatitis B in adults is approximately 15% in Taiwan. Even if the hepatitis B vaccine is effective, most infected individuals will become lifelong carriers of HBV as there is no cure for HBV infection. Therefore, controlling or even curing the HBV infection is an essential public health issue that must be addressed. At present, there are two types of agents available for the treatment of HBV infection: interferons and nucleotide analogs. Interferons have antiviral, anti-proliferative, and immunomodulatory effects, whereas nucleotide analogs suppress the replication of HBV [4]. Nevertheless, the current drugs for treating HBV infection can only control the disease and achieve a functional cure (loss of HBsAg). In addition, regardless of the type of drug, drug resistance is eventually seen in the patients [5,6]. For HBV carriers, the virus always exists in the liver and the HBV DNA is permanently integrated into the host genome [7]. Although the virus may not harm the liver directly, the human immune cells will attack the liver cells when they recognize the virus in the liver cells, resulting in inflammation and impairment of liver functions, which may eventually lead to cirrhosis and hepatocellular carcinoma (HCC). Therefore, how to achieve complete cure (elimination of infected hepatocytes and the cells with integrated HBV DNA) for HBV carriers is still a problem to be solved [8]. In order to eliminate HBV-infected hepatocytes, our strategy was to identify the target genes that could be used for the selective killing of HBV-positive hepatocytes. The target genes must be crucial for survival of the HBV-positive hepatocytes but insignificant for the survival of HBV-negative hepatocytes.

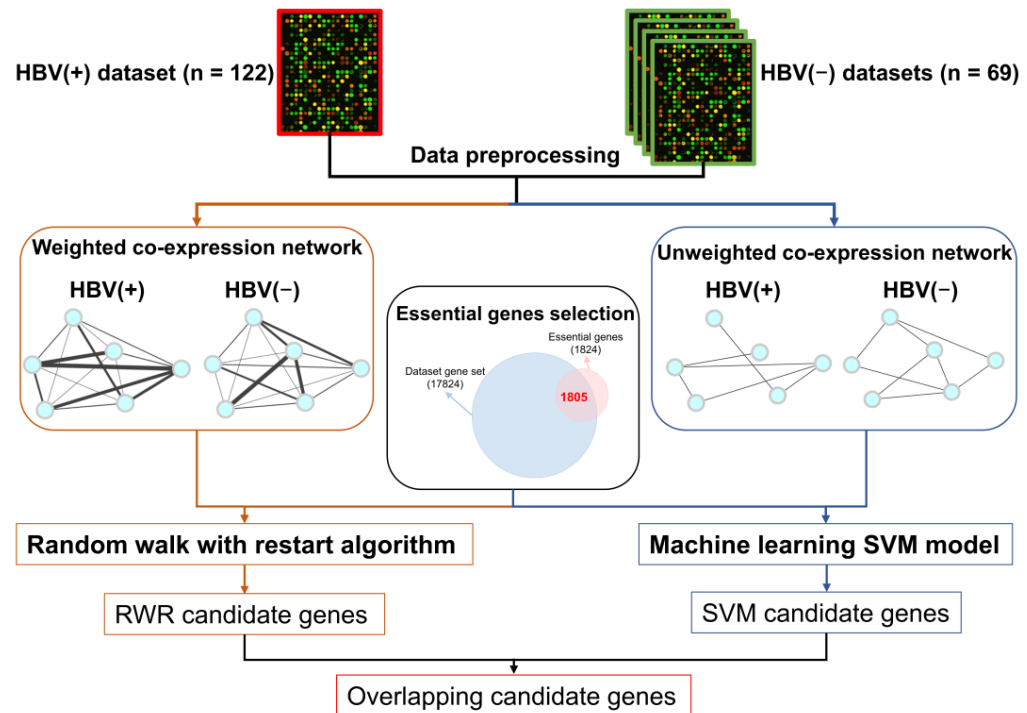
Genes that are critical for the survival of organisms are called essential genes. However, the essentiality of a gene is not an intrinsic property, indicating that the gene may be essential in some conditions but not essential in other circumstances. Essentiality is highly dependent on a variety of factors, such as the function and expression pattern of the gene, genetic context of the organism, and environment. [9] This property makes essential genes therapeutic targets of diseases. In the past, the method of identifying essential genes was to observe the survival of the cells with target gene knock-out or knock-down through experiments. However, different experimental methods may produce different results. For example, the CRISPR-based methods identify more essential genes than the siRNA-based methods [10]. Genes with different essentialities in different situations are called conditionally essential genes (CEGs) [11,12], which can be applied to develop more effective or more specific drugs. In cancer therapies, CEGs are essential for specific tumor cells but not for other normal cells [13]. For example, poly (ADP-ribose) polymerase (PARP) inhibitors have been approved for the treatment of ovarian cancer. As the PARP protein is only required for the *BRCA1*- or *BRCA2*-deficient cells, it can be used for selectively eliminating the cancer cells [14]. The concept of this treatment, which takes advantage of the different essentialities of different cells, is similar to our strategy to identify therapeutic targets for the selective killing of HBV-positive hepatocytes.

With advancements in technology, many high-throughput technologies have been developed in recent years to explore biomedicine, and many datasets have also been deposited in public databases. In this study, we collected the gene expression profiles of hepatitis B from a public database and developed a computational method to identify the CEGs in the context of HBV infection. The random walk with restart (RWR) algorithm and the machine learning approach with support vector machine (SVM) were used for the identification of CEGs. The identified CEGs were found to play key roles in the survival of the HBV-infected liver cells but were not related to the survival of the normal liver cells. Therefore, these CEGs can be used as therapeutic targets to design potent drugs for HBV infection, which can selectively kill the HBV-infected liver cells without affecting the normal liver cells. Moreover, by applying the method for the regeneration of normal liver cells, we hope to achieve the ultimate goal of a complete cure for hepatitis B.

## 2. Materials and Methods

### 2.1. Overview of the Method

The workflow of this study is shown in Figure 1. The gene expression data of HBV-positive (HBV(+)) and HBV-negative (HBV(-)) samples were collected and downloaded from a public database. After data preprocessing, the two types of networks, weighted co-expression networks and unweighted co-expression networks, were constructed. With the selected essential genes, two algorithms were used to identify the CEGs: the RWR algorithm and the SVM approach. The identified CEGs were the intersection of the candidate genes predicted by the two algorithms.



**Figure 1.** Schematic representation of the workflow of this study.

### 2.2. Datasets and Sample Characteristics

We curated the microarray gene expression data of the HBV(+) and HBV(-) samples from the Gene Expression Omnibus (GEO) database (GPL570 platform). Among the curated datasets, HBV(+) samples ( $n = 122$ ) are chronic hepatitis B liver samples obtained from GSE83148 [15] and HBV(-) samples were obtained from GSE83148 ( $n = 6$ ) [15], GSE6764 ( $n = 10$ ) [16], GSE14668 ( $n = 11$ ) [17], GSE38941 ( $n = 10$ ) [18], GSE23343 ( $n = 7$ ) [19], GSE28619 ( $n = 7$ ) [20], GSE62029 ( $n = 10$ ) [21], and GSE101685 ( $n = 8$ ). None of the HBV(-) samples showed evidence of infection with hepatitis B, C, and HCC. Information about these curated datasets is presented in Table 1.

### 2.3. Data Preprocessing

Due to the different methods for the normalization of these curated datasets, we downloaded the raw data (CEL file) of each sample for merging and used the Robust Multichip Average (RMA) algorithm [22] (R package—affy\_1.68.0) for normalization. Then, the batch effects were removed with the “removeBatchEffect” function (R package—limma\_3.46.0) [23]. As the curated datasets contained expression profiles for protein-coding genes and non-coding RNAs, 17,824 protein-coding genes were selected for further analysis.

**Table 1.** The information of GEO microarray datasets used in the study.

Type	GEO ID	Sample Number	Country	Year
HBV(+)	GSE83148	122	China	2017
HBV(−)	GSE83148	6	China	2017
	GSE6764	10	USA	2007
	GSE14668	11	USA	2010
	GSE38941	10	USA	2012
	GSE23343	7	Japan	2010
	GSE28619	7	Spain	2012
	GSE62029	10	Italy	2015
	GSE101685	8	Taiwan	2019

#### 2.4. Selection of the Known Essential Genes

Based on the objectives of this study, the essential genes were selected as the seed nodes of the RWR algorithm. We downloaded the Homo sapiens essential genes annotation file from the Online GENE Essentiality (OGEE) database (version 2), which lists the essentiality consensus of 21,556 genes that are experimentally verified. A total of 7168 genes were labeled as “Essential” (marked as essential in all studies) or “Conditional” (marked as essential in some studies and non-essential in other studies). However, because each gene was labeled with a different number of studies, the genes with at least four studies (and three out of four) marked as essential were selected as the known essential genes in this study. Based on this criterion, a total of 1805 known essential genes were selected for further analysis.

#### 2.5. RWR Algorithm

To execute the RWR algorithm, weighted co-expression networks for both HBV(+) and HBV(−) samples were constructed, respectively. First, the gene–gene Spearman correlation matrix was constructed by calculating the Spearman correlation coefficients between every gene pair. Then, the soft threshold picking method based on weighted gene co-expression network analysis (WGCNA) [24] was used to select the suitable power  $\beta$  for both HBV(+) and HBV(−) matrices. The value of power  $\beta$  was selected according to the scale-free topology criterion, which aims to simulate the natural biological network structure [25]. The steps for constructing weighted gene co-expression networks are shown in Supplementary Figure S1.

RWR is a ranking algorithm that simulates a random walker starting from the seed nodes, which represent the known targets with specific properties, thereby calculating the probability of each node as a novel target. It has been applied to networks of different structures (e.g., microRNA–target gene network [26] and protein–protein interaction network [27]) and has been successfully used to predict potential disease-associated microRNAs/genes. The RWR formula is as follows:

$$P_{t+1} = (1 - r)W^T P_t + rP_0 \quad (1)$$

where  $W$  is the row-normalized transition matrix, which can be converted from our network correlation matrix given that power  $\beta$  is selected,  $W_{ij}$  of the transition matrix  $W$  is the transition probability from node  $i$  to node  $j$ ,  $P_0$  is the  $N \times 1$  initial score vector ( $N$  is the number of genes in the networks), in which the seed nodes are assigned equal values and the sum of the values equals 1, where other non-seed nodes are assigned 0,  $P_t$  is the score vector after  $t$  iterations, and  $r$  is the restart probability of the walker returning to the seed nodes, which can ensure the importance of the seed nodes. The implication of the restart probability is that in every step, scores with  $r$  ratio remained in the seed nodes, whereas others with  $1 - r$  ratio were shared with neighbors. After  $t$  iterations of this algorithm, the loop stopped when  $\|P_{t+1} - P_t\| < 10^{-6}$  [28], indicating that the score of all of the nodes reached a steady state. In this case,  $P_{t+1}$  was the final score vector, where nodes (genes) with higher scores were considered more likely to be the seed nodes. With known essential

genes as the seed nodes, the scores determined by RWR were defined as the essential scores for each gene. The steps for calculating the essential difference scores are shown in Supplementary Figure S2. Genes with larger essential difference scores are more likely to be the CEGs in HBV-infected hepatocytes. To avoid false positives in the RWR results, we also performed a permutation test to calculate an empirical  $p$ -value for each gene [27,29–31].

Based on the essential difference score and the corresponding empirical  $p$ -values calculated for each gene, three criteria were used to screen RWR candidate genes: (1) genes with empirical  $p$ -values less than 0.05, (2) genes with the top 10% essential difference score, and (3) genes with the top 25% essential score ranking in the HBV(+) network and the bottom 25% essential score ranking in the HBV(−) network. Genes that met all three criteria were identified as RWR candidate genes.

### 2.6. SVM Approach

In addition to the RWR algorithm, a machine learning method using SVM was also used for the identification of CEGs. Machine learning is one of the main methods for predicting essential genes [32,33]. We referred to the study by Hwang et al. to predict essential genes using network features and SVM [34]. As some network features can only be computed in unweighted networks, unweighted gene co-expression networks for both HBV(+) and HBV(−) samples were constructed, respectively. Similarly, correlation matrices for both conditions were first constructed. Subsequently, the concept of choosing a hard threshold according to the scale-free topology criterion was employed to transform the weighted networks into unweighted gene co-expression networks. That is, we set the cut-off in the range [0.1–0.9] in increments of 0.1, and tested whether the resulting networks of each cut-off displayed a scale-free topological structure.

Seven network features were calculated to train the SVM classification model: degree (K), betweenness centrality (BC), closeness centrality (CC), clustering coefficient (CCo), neighbors' intra-degree (NID), essentiality index (EI), and common-function degree (CFK) [34]. Among these features, K, BC, CC, CCo, and NID were calculated using python package—NetwrokX\_1.11 [35], EI was calculated using the known essential genes we filtered, and CFK was calculated using the Gene Ontology (GO) annotation file downloaded from the National Center for Biotechnology Information (NCBI) website. The seven network features for each gene were calculated in HBV(+) and HBV(−) unweighted gene co-expression networks, respectively. The steps for calculating the essential difference probabilities are shown in Supplementary Figure S3. Genes with larger essential difference probabilities are more likely to be the CEGs in HBV-infected hepatocytes.

For the selection of SVM candidate genes, because there is no empirical  $p$ -value for each gene, the following two criteria as in RWR were used: (1) genes with the top 10% difference essential probability, and (2) genes with the top 25% essential probability ranking in the HBV(+) network and the bottom 25% essential probability ranking in the HBV(−) network.

### 2.7. Gene Set Enrichment Preranked Analysis (GSEAPreranked)

Based on the ranked essential difference scores from RWR and the ranked essential difference probabilities from SVM, we obtained the average ranking of each gene in the networks. Given the ranked list of genes, the GSEAPreranked analysis for enriched pathways and functions was performed with the “gseKEGG” and “gseGO” function (R package—clusterProfiler\_3.18.1) [36]. The enriched pathways and functions were analyzed with the Kyoto Encyclopedia of Genes and Genomes (KEGG) (release: 10 June 2021) and GO biological process (GOBP) gene sets (release: 1 February 2021) embedded in clusterProfiler. Those gene sets with Benjamini–Hochberg adjusted  $p$ -values less than 0.05 were selected as the enriched pathways/functions. The “simplify” function in clusterProfiler was further employed to remove redundancy of enriched GO terms. The enrichment map showing the correlation between enriched GO terms was displayed using the “emapplot” function (R package—enrichplot\_1.13.0.993).

### 3. Results

#### 3.1. Data Preparation

Using the microarray gene expression profiles for HBV(+) and HBV(−) samples downloaded from the GEO database, a computational framework was developed to predict the CEGs. After normalization of the raw data of each sample, the batch effect was removed, and 17,824 protein-coding genes were selected for further analysis. The gene expression profile distributions for all samples are shown in Supplementary Figure S4.

The known essential genes were curated before proceeding with the RWR and SVM analyses. The Homo sapiens essential genes annotation file was downloaded from the OGEE database, which labeled 7168 genes as essential genes. In addition, we also investigated the number of known human essential genes and found that only ~10% of ~20,000 genes in human cells are essential for cell survival [13,37]. Therefore, 7168 essential genes were further screened. Based on the data provided by OGEE, each gene was marked with different numbers of essentiality status—Essential (E) or Non-Essential (NE)—according to the experimental results of each study (dataset). In other words, a gene can be essential in some datasets but non-essential in other datasets. The distribution of marked datasets for 7168 essential genes indicated that most of the essential genes were marked in 8 datasets. The genes that were marked as ‘E’ in at least four datasets or marked as ‘E’ in three out of four datasets were selected as known essential genes with high confidence. Finally, 1805 known essential genes were selected for further analysis.

#### 3.2. Using RWR to Identify the CEGs

For the RWR analysis, the weighted gene co-expression networks for HBV(+) and HBV(−) samples were constructed, where the power  $\beta$  needed to be determined first. The different power  $\beta$ 's and their corresponding  $R^2$  in the scale-free topological model fitting are shown in Supplementary Figure S5. The results showed that  $R^2$  increased as the power  $\beta$  increased. Based on the results shown in Supplementary Figure S5 and the criteria that  $R^2 \geq 0.7$ , power  $\beta$  was selected as six for HBV(+) and four for HBV(−) weighted networks, and the corresponding  $R^2$  was 0.71 and 0.74 in HBV(+) and HBV(−), respectively.

The essential score for each gene in the HBV(+) and HBV(−) weighted gene co-expression networks was determined using the RWR algorithm, respectively. The quantile-quantile plot showed that there was only a small deviation in the essential score distribution of genes in HBV(+) and HBV(−) networks (Supplementary Figure S6A), demonstrating that the essential scores for HBV(+) and HBV(−) were comparable. The essential difference score was calculated by subtracting the essential score of HBV(−) from the essential score of HBV(+), and the distribution of essential difference scores is shown in Supplementary Figure S6B. Genes with larger essential scores indicated that they are more likely to be essential genes. Therefore, the genes with larger essential difference scores indicate that they are more likely to be essential in HBV(+) samples but less likely to be essential in HBV(−) samples, making them CEGs. Based on the three criteria mentioned above, that is, essential difference score ranking, essential score ranking, and empirical  $p$ -value, 309 candidate genes were identified using the RWR algorithm.

Referring to the literature [38,39], the restart probability  $r$  in Equation (1) was initially set as 0.7, and 309 candidate genes were identified. Subsequently, we tested the impact of different  $r$  values on the identified RWR candidate genes. As the  $r$  value used in the RWR algorithm were between 0 and 1, we set the  $r$  value in the range [0.1–0.9] in increments of 0.2, and tested whether the identified RWR candidate genes of each  $r$  value were significantly different. The number of candidate genes identified for different values of  $r$  is listed in Supplementary Table S1. The number of overlapped candidate genes and their Jaccard similarities are listed in Supplementary Table S2. The average Jaccard similarity for  $r = 0.7$  is 0.72, demonstrating that the  $r$  value has little impact on our final RWR candidate genes.

### 3.3. Using SVM to Identify the CEGs

For the SVM analysis, seven network topological features that could be used to predict essential genes were calculated. As only four features could be calculated in the unweighted network, the unweighted networks of HBV(+) and HBV(−) samples were constructed first. Similarly, the Spearman correlation coefficients for each gene pair were calculated, leading to correlation matrices for both conditions. Subsequently, the hard threshold was determined. Based on the preset cut-offs and the corresponding scale-free model fitting index  $R^2$  shown in Supplementary Figure S7, the hard threshold was selected as 0.5 for both HBV(+) and HBV(−), such that the corresponding  $R^2 \geq 0.7$ . The scale-free model fitting index  $R^2$  and the identified  $\gamma$  (slope parameter when fitting power law distribution) of each cut-off, as well as the degree statistical characteristics of the corresponding unweighted network, are listed in Supplementary Table S3. The results showed that in the HBV(+) network, when the cut-off was 0.1–0.2, it did not conform to the power law (the slope parameter should be positive), even with a very high  $R^2$ . Therefore, we did not choose 0.1 and 0.2 as the hard threshold for unweighted network construction. With 0.5 as the hard threshold, the weighted HBV(+) and HBV(−) networks can be transformed into unweighted networks. During the transformation, some genes had a degree of zero in the unweighted network and were excluded from further analysis, leading to 16,419 genes remaining for SVM analysis.

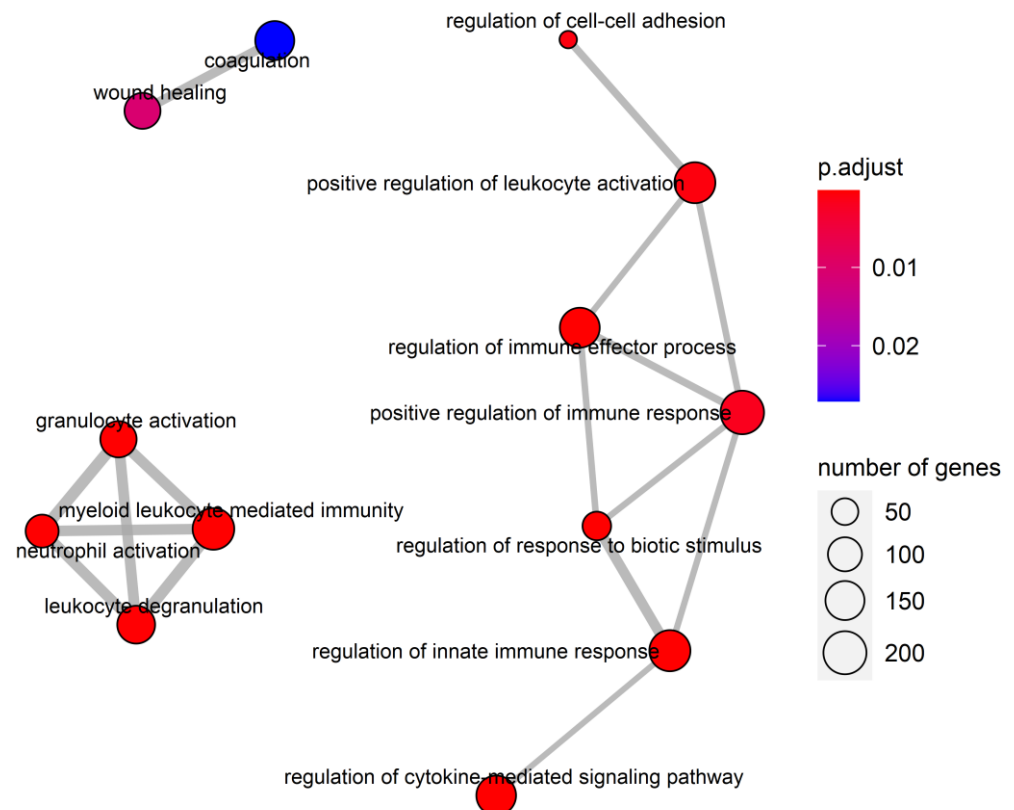
Based on the unweighted HBV(+) and HBV(−) networks, the SVM classification model was used to predict candidate genes. As SVM is a supervised method, both positive and negative data are required for model training. The positive group of training data consisted of the 1754 known essential genes (after mapping 1805 known essential genes to the unweighted networks, the number of known essential genes in the unweighted networks was 1754). However, based on the principle of the essentiality of genes, it is difficult to define non-essential genes. We referred to the method described by Yang et al., which used random sampling for non-essential gene selection [32]. In this study, the negative group of training data was 1754 genes that were randomly selected from the 14,665 genes other than the known essential genes, and the prediction data included 14,665 remaining genes (Supplementary Figure S3). The non-essential genes were randomly selected 1000 times, and thus, 1000 SVM models were trained to provide essential status for each gene in the network. After 1000 SVM classifications, the essential probability, which was calculated by dividing essential counts (the number of times to be predicted as essential genes) by 1000, was obtained for each gene in HBV(+) and HBV(−) networks, respectively. Furthermore, the essential difference probability was calculated by subtracting the essential probability of HBV(−) from the essential probability of HBV(+). Based on the two criteria mentioned above, that is, essential difference probability ranking and essential probability ranking, 688 candidate genes were identified using the SVM algorithm.

Cross-validation was performed to verify the prediction rate of this model. The 1754 known essential genes and 1754 randomly selected non-essential genes were divided into 70% training data and 30% testing data to perform cross-validation. The procedure was repeated 1000 times and the average precision, recall, and accuracy were calculated. The three indicators are between 0.70 and 0.72 in HBV(+) and HBV(−), indicating that this SVM model can be used to predict essential genes.

Although the network topological features were demonstrated to be useful for predicting essential genes, we wanted to investigate whether they can be used for prediction in both HBV(+) and HBV(−) networks. Based on the SVM model construction method, the essential genes were fixed, and non-essential genes were randomly selected. The Mann–Whitney U test was used to investigate whether these seven network features were significantly different between essential and non-essential genes. The results of statistical tests are listed in Supplementary Table S4, indicating that regardless of the HBV(+) or HBV(−) unweighted networks, essential and non-essential genes have significant differences in all seven network features, verifying that these features can be directly applied to our networks.

### 3.4. GSEAPreranked for the Average Ranking Gene List

Given the ranked gene lists based on essential difference scores from RWR and essential difference probabilities from SVM, GSEAPreranked analysis was performed to determine whether some pathways showed significant enrichment. The given ranked gene list was positively enriched in 26 KEGG pathways and 33 GOBP functions (Supplementary Tables S5 and S6). It is worth noting that one of the enriched KEGG gene sets was Fc gamma R-mediated phagocytosis (adjusted  $p$ -value = 0.047). It has been reported that Fc gamma receptors are involved in the emerging immunotherapy of chronic HBV infection treatment, which can durably suppress the levels of HBsAg and HBV DNA via Fc $\gamma$  receptor-dependent phagocytosis [40,41]. In addition, pathways related to cell survival such as apoptosis (adjusted  $p$ -value = 0.032) have also been identified, suggesting that the designed workflow may be used to identify CEGs. We calculated the Jaccard coefficient of enriched GOBP gene sets and further analyzed the correlation between the results. The threshold of the Jaccard coefficient was set to 0.25, and the nodes with degree less than one were filtered. The enrichment map showed that the enriched GO functions can be divided into three groups (Figure 2). The main group is related to the regulation of immune responses, including cell–cell adhesion, leukocyte activation, immune effector process, immune response, response to biotic stimulus, innate immune response, and cytokine-mediated signaling pathway, indicating that most of the enriched GOBP gene sets are related to immune regulation. The second group is related to the leukocyte, including granulocytes activation, myeloid leukocyte-mediated immunity, neutrophil activation, and leukocyte degranulation. The last group includes coagulation and wound healing.

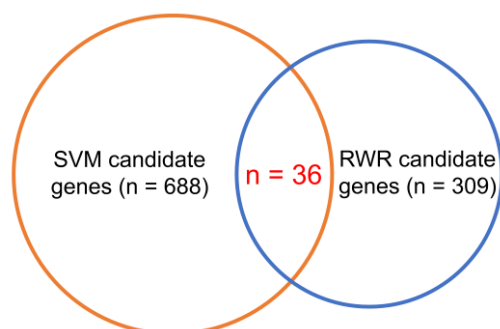


**Figure 2.** The enrichment map of the gene set enrichment analysis (GSEA) of Gene Ontology (GO) gene sets. The edges between the enriched GO terms indicate that the Jaccard similarity among the gene sets was larger than 0.25, in which the width represents the level of Jaccard similarity. The size of nodes shows the size of GOBP gene sets. The color of nodes indicates the level of adjusted  $p$ -values for enriched gene sets.



### 3.5. Candidate Genes Analysis

With 309 RWR and 688 SVM candidate genes identified, these two gene sets were intersected, resulting in 36 overlapping candidate genes, which are potentially the CEGs (Figure 3). The ranked gene list of these 36 overlapping candidate genes is listed in Table 2, where the genes with higher ranking were more likely to be CEGs in HBV-infected hepatocytes, according to our analyses.

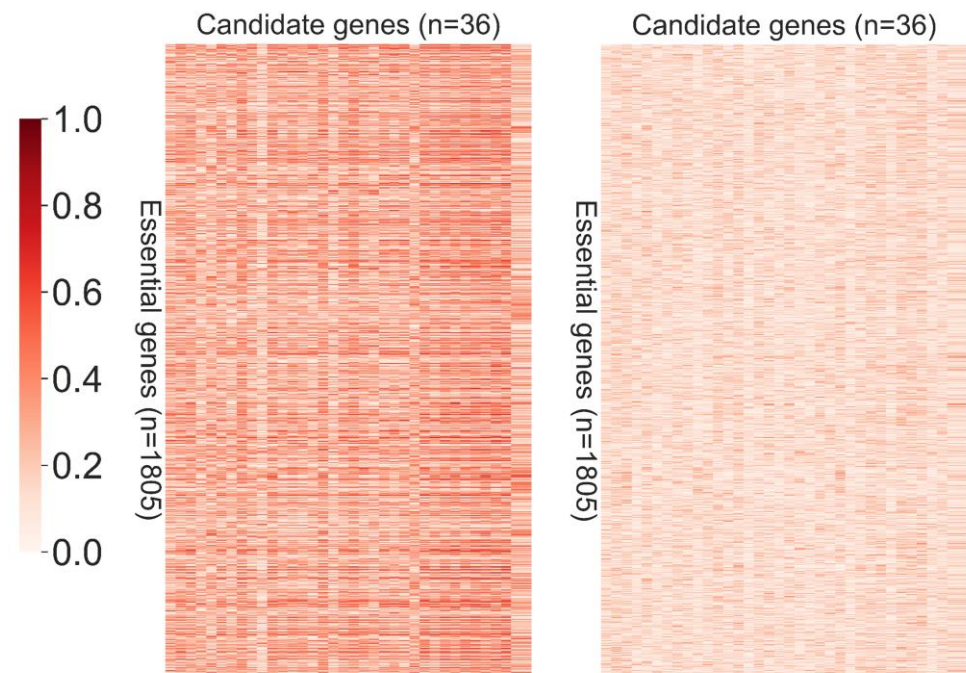


**Figure 3.** The 36 overlapping candidate genes are the intersection of the random walk with restart (RWR) and support vector machine (SVM) candidate genes.

**Table 2.** The ranked gene list of potential conditionally essential genes.

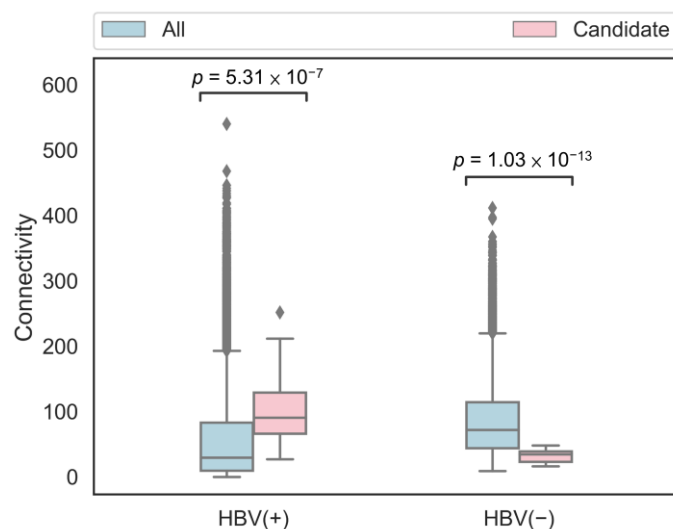
Ranking	Entrez ID	Gene Symbol	Association with HBV/HCC	Ranking	Entrez ID	Gene Symbol	Association with HBV/HCC
1	200030	<i>NBPF11</i>		19	27350	<i>APOBEC3C</i>	HBV [42,43]
2	284565	<i>NBPF15</i>		20	57798	<i>GATAD1</i>	HCC [44,45]
3	221937	<i>FOXK1</i>	HCC [46]	21	342979	<i>PALM3</i>	
4	148266	<i>ZNF569</i>		22	4335	<i>MNT</i>	
5	9883	<i>POM121</i>		23	57186	<i>RALGAPA2</i>	HCC [47]
6	2077	<i>ERF</i>		24	9743	<i>ARHGAP32</i>	
7	5141	<i>PDE4A</i>	HBV and HCC [48]	25	7205	<i>TRIP6</i>	HCC [49]
8	7328	<i>UBE2H</i>	HBV and HCC [50,51]	26	51479	<i>ANKFY1</i>	
9	84433	<i>CARD11</i>		27	6934	<i>TCF7L2</i>	HBV [52]
10	6711	<i>SPTBN1</i>	HCC [53]	28	7025	<i>NR2F1</i>	
11	11156	<i>PTP4A3</i>	HCC [54]	29	79719	<i>AAGAB</i>	
12	89122	<i>TRIM4</i>	HCC [55,56]	30	23365	<i>ARHGEF12</i>	HBV and HCC [57]
13	999	<i>CDH1</i>	HBV and HCC [58]	31	5269	<i>SERPINB6</i>	
14	23568	<i>ARL2BP</i>		32	56935	<i>SMCO4</i>	
15	60401	<i>EDA2R</i>		33	408	<i>ARRB1</i>	HBV and HCC [59]
16	7559	<i>ZNF12</i>		34	90268	<i>OTULIN</i>	
17	6310	<i>ATXN1</i>		35	81030	<i>ZBP1</i>	
18	4001	<i>LMNB1</i>	HCC [60]	36	342371	<i>ATXN1L</i>	

To validate the overlapping candidate genes, we first investigated whether the gene expression of these identified genes was more significantly correlated with known essential genes in HBV(+) samples as compared to in HBV(−) samples. The Spearman correlation coefficients of gene expression between 36 overlapping candidate genes and 1805 known essential genes were calculated, and the correlation heatmaps of HBV(+) and HBV(−) are shown in Figure 4. The spearman correlation coefficients are taken as absolute values, which were between 0 and 1, so the darkest color in the color bar showed the highest correlation. The results show that the correlation between the identified overlapping candidate genes and known essential genes is higher in HBV(+) samples, but lower in HBV(−) samples, which means that the 36 overlapping candidate genes are more likely to be essential genes in HBV(+) samples, but not in HBV(−) samples.



**Figure 4.** Correlation between the overlapping candidate genes and known essential genes. Left: HBV(+) samples, right: HBV(−) samples.

Many studies have indicated that essential genes are in the network center (that is, the connection with other genes was higher), but non-essential genes were the opposite [13,61]. Therefore, the connectivity of the overlapping candidate genes in both HBV(+) and HBV(−) networks was analyzed. The boxplots of the connectivity for all genes and 36 overlapping candidate genes in the HBV(+) and HBV(−) weighted networks are shown in Figure 5. Connectivity is a connection degree indicator for genes and other genes. The connectivity of these overlapping candidate genes is relatively large in the HBV(+) network, whereas the connectivity of the overlapping candidate genes in the HBV(−) network is relatively small. According to the literature, these 36 overlapping candidate genes are more likely to be essential genes in HBV(+) samples, but are less likely in HBV(−) samples, which are the CEGs we would like to identify.



**Figure 5.** The boxplots of connectivity for all/candidate genes in the HBV(+) and HBV(−) weighted networks. *p*-value was obtained using the Mann–Whitney U test.

#### 4. Discussion

The aim of this study was to identify the CEGs among HBV(+) and HBV(−) samples, which can be used as therapeutic targets for treating HBV infection. For the purpose of this study, HBV(+) and HBV(−) gene co-expression networks were constructed. In the first step of network construction, the GSE83148 dataset was downloaded from the GEO database, which contains 122 HBV(+) and 6 HBV(−) samples, and the Spearman correlation coefficients for every gene pair were calculated. As the calculation of Spearman correlation coefficients is limited by the sample size, only six HBV(−) samples could have caused deviations in the calculation of correlation. Therefore, we further collected all available HBV(−) samples ( $n = 63$ ) from the GPL570 platform (the same as the GSE83148 dataset) and demonstrated that these samples were not infected with hepatitis B, C, and HCC. As these HBV(−) samples were retrieved from different datasets and different normalization methods were applied, the CEL file of raw data for each sample was downloaded, and RMA normalization was performed for all HBV(+) and HBV(−) samples. In addition, batch effects were removed. Even with all the HBV(−) samples in the GPL570 platform, the sample sizes among 122 HBV(+) and 69 HBV(−) samples were still not balanced. Consequently, there may have been some deviations in calculating the correlations and comparing the two networks. We believe that if more HBV(−) samples can be collected in the future, the identified CEGs can be confidently identified. Moreover, HBV(−) gene expression profiles were retrieved from samples from different countries, which may have different genetic backgrounds (Table 1). Since different genetic backgrounds of the HBV(+) and HBV(−) samples may influence the results, further studies are needed to investigate the possible impact of various genotypes on the identification of CEGs.

Based on two different approaches, RWR and SVM, 36 overlapping candidate genes were identified in this study. We further searched the literature for the 36 candidate genes to investigate their associations with HBV infection. Several genes have been reported to be associated with HBV replication, integration, and other related functions. The *APOBEC3* family has antiviral activity against retroviruses and can also inhibit HBV replication [42,43]. *UBE2H* was a breakpoint of HBV integration in plasma DNA for HBV-related HCC samples, and was upregulated in HCC [50,51]. *TCF7L2* may participate in the upregulation of HBV core promoter activity via the interaction with the enhancer region and *PUF60* [52]. Multiple TRIM proteins with E3 ligase function can inhibit HBV transcription [55]. *ARHGEF12* was recurrently inserted by HBV in tumor-adjacent tissues of HBV-related HCC samples to increased its gene expression [57]. HBV could have a large effect on the concurrent DNA methylation of *CDH1*, *DNMT3b*, and *ESR1* in the serum of HBV-related HCC samples [58]. *ARRB1* has been indicated to promote HCC, and was upregulated by hepatitis B virus X protein (HBx) in mouse models [59]. Interestingly, among other candidate genes, some genes are related to the progression of HCC, including, *GATAD1* [44,45], *FOXK1* [46], *RALGAPA2* [47], *PDE4A* [48], *TRIP6* [49], *SPTBN1* [53], *PTP4A3* [54], *TRIM4* [56], and *LMNB1* [60]. These results suggest that the 36 candidate genes are associated with the functions of HBV and hepatocytes; however, this association needs to be further validated. Moreover, to investigate whether some enriched GOBP functions were shared among the 36 overlapping candidate genes, 273 RWR-specific genes, and 652 SVM-specific genes, functional annotation clustering was applied to identify the enriched GOBP functions. After removing redundancy, 37, 39, and 77 enriched GOBP functions were identified for overlapping genes, RWR-specific genes, and SVM-specific genes, respectively (Supplementary Figure S8). Only two GOBP functions were found to be shared among overlapping candidate genes and SVM-specific genes. Since RWR and SVM use different principles for identifying the CEGs, the result is not surprising. In addition, through literature surveys, we found that most of the 36 overlapping genes are related to HBV or hepatocytes (Table 2), demonstrating that application of these two methods rather than a single one can avoid the false positives. Therefore, the CEGs identified by both methods are more robust.

## 5. Conclusions

In contrast to conventional studies, we used the concept of CEGs to develop targeted drugs for the HBV-infected liver cells. We also developed a computational framework to predict the conditionally essential genes in the HBV-infected liver cells. This computational framework contains two algorithms: RWR and SVM. A total of 309 RWR and 688 SVM candidate genes were identified in this study. The RWR and SVM candidate genes were then intersected to identify the 36 overlapping candidate genes, which may be CEGs. The gene expression correlation heatmap of these 36 genes and 1805 known essential genes showed a higher similarity in the HBV(+) samples than the HBV(−) samples. The results of the connectivity analysis were consistent with the network characteristics of essential and non-essential genes mentioned in the literature. Essential genes exhibit higher connectivity, but non-essential genes do not. Most candidate genes are associated with the functions of HBV and hepatocytes, indicating that these 36 target genes could be used as potential drug targets to develop novel strategies for managing HBV infection and may ultimately help in achieving the elusive goal of a complete cure for hepatitis B.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/jpm11070649/s1>, Figure S1: Steps for the construction of the weighted gene co-expression networks; Figure S2: Steps for calculating the essential difference score using the random walk with restart (RWR) algorithm; Figure S3: Steps for calculating the essential difference probability using the support vector machine (SVM) approach; Figure S4: The gene expression profile distributions across HBV(+) and HBV(−) samples after normalization; Figure S5: Different power  $\beta$ 's and their corresponding  $R^2$  in the scale-free topological model fitting for the soft threshold selection; Figure S6: Distributions of the essential scores and essential difference scores; Figure S7: Different cut-offs and their corresponding  $R^2$  in the scale-free topological model fitting for the hard threshold selection; Figure S8: The Venn diagram showing the relation between the enriched GOBP functions for overlapping candidate genes, RWR-specific genes, and SVM-specific genes. Table S1: The number of identified candidate genes for different values of  $r$ ; Table S2: The number of overlapped candidate genes and their Jaccard similarities; Table S3: The scale-free model fitting index  $R^2$ , the identified  $\gamma$ , and the degree statistical characteristics of the corresponding unweighted network for each cut-off; Table S4: Statistical test results comparing the network features among essential and non-essential genes; Table S5: The enrichment results of the enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets based on the gene set enrichment analysis (GSEA); Table S6: The enrichment results of the enriched Gene Ontology (GO) gene sets based on the gene set enrichment analysis (GSEA).

**Author Contributions:** Conceptualization, L.H.-C.W. and Y.-C.W.; methodology, C.-J.H. and Y.-C.W.; software, C.-J.H.; formal analysis, C.-J.H.; investigation, C.-J.H.; data curation, C.-J.H.; writing—original draft preparation, C.-J.H. and Y.-C.W.; visualization, C.-J.H.; supervision, L.H.-C.W. and Y.-C.W.; funding acquisition, Y.-C.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology, Taiwan, grant number MOST 107-2221-E-010-019-MY3.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/>. Accession numbers were shown in Materials and Methods. Data was accessed on 3 June 2021.

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

## References

1. World Health Organization. *Global Hepatitis Report*; WHO: Geneva, Switzerland, 2017.
2. Stasi, C.; Silvestri, C.; Voller, F. Emerging Trends in Epidemiology of Hepatitis B Virus Infection. *J. Clin. Transl. Hepatol.* **2017**, *5*, 272–276. [[CrossRef](#)]

3. Chen, C.-Y.; Hsu, H.-Y.; Liu, C.-C.; Chang, M.-H.; Ni, Y.-H. Stable seroepidemiology of hepatitis B after universal immunization in Taiwan: A 3-year study of national surveillance of primary school students. *Vaccine* **2010**, *28*, 5605–5608. [[CrossRef](#)]
4. Alonso, S.; Guerra, A.-R.; Carreira, L.; Ferrer, J.-Á.; Gutiérrez, M.-L.; Fernandez-Rodriguez, C.M. Upcoming pharmacological developments in chronic hepatitis B: Can we glimpse a cure on the horizon? *BMC Gastroenterol.* **2017**, *17*, 168. [[CrossRef](#)]
5. Zoulim, F.; Locarnini, S. Optimal management of chronic hepatitis B patients with treatment failure and antiviral drug resistance. *Liver Int.* **2013**, *33*, 116–124. [[CrossRef](#)]
6. Lok, A.S.; Zoulim, F.; Dusheiko, G.; Ghany, M.G. Hepatitis B cure: From discovery to regulatory approval. *Hepatology* **2017**, *66*, 1296–1313. [[CrossRef](#)]
7. Shafritz, D.A.; Shouval, D.; Sherman, H.I.; Hadziyannis, S.J.; Kew, M.C. Integration of Hepatitis B Virus DNA into the Genome of Liver Cells in Chronic Liver Disease and Hepatocellular Carcinoma. *N. Engl. J. Med.* **1981**, *305*, 1067–1073. [[CrossRef](#)]
8. Levrero, M. Future therapies for HBV Cure. In Proceedings of the 9th Paris Hepatitis Conference, Paris, France, 11–12 January 2016.
9. Chen, W.-H.; Lu, G.; Chen, X.; Zhao, X.-M.; Bork, P. OGEE v2: An update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res.* **2016**, *45*, D940–D944. [[CrossRef](#)]
10. Hart, T.; Chandrashekar, M.; Aregger, M.; Steinhart, Z.; Brown, K.R.; MacLeod, G.; Mis, M.; Zimmermann, M.; Fradet-Turcotte, A.; Sun, S.; et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **2015**, *163*, 1515–1526. [[CrossRef](#)]
11. Wang, T.; Birsoy, K.; Hughes, N.W.; Krupczak, K.M.; Post, Y.; Wei, J.J.; Lander, E.S.; Sabatini, D.M. Identification and characterization of essential genes in the human genome. *Science* **2015**, *350*, 1096. [[CrossRef](#)]
12. Chen, W.-H.; Minguéz, P.; Lercher, M.J.; Bork, P. OGEE: An online gene essentiality database. *Nucleic Acids Res.* **2011**, *40*, D901–D906. [[CrossRef](#)]
13. Rancati, G.; Moffat, J.; Typas, A.; Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **2017**, *19*, 34. [[CrossRef](#)]
14. Farmer, H.; McCabe, N.; Lord, C.J.; Tutt, A.N.J.; Johnson, D.A.; Richardson, T.B.; Santarosa, M.; Dillon, K.J.; Hickson, I.; Knights, C.; et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **2005**, *434*, 917–921. [[CrossRef](#)]
15. Zhou, W.; Ma, Y.; Zhang, J.; Hu, J.; Zhang, M.; Wang, Y.; Li, Y.; Wu, L.; Pan, Y.; Zhang, Y.; et al. Predictive model for inflammation grades of chronic hepatitis B: Large-scale analysis of clinical parameters and gene expressions. *Liver Int.* **2017**, *37*, 1632–1641. [[CrossRef](#)]
16. Wurmbach, E.; Chen, Y.-B.; Khitrov, G.; Zhang, W.; Roayaie, S.; Schwartz, M.; Fiel, I.; Thung, S.; Mazzaferro, V.; Bruix, J.; et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **2007**, *45*, 938–947. [[CrossRef](#)]
17. Farci, P.; Diaz, G.; Chen, Z.; Govindarajan, S.; Tice, A.; Agulto, L.; Pittaluga, S.; Boon, D.; Yu, C.; Engle, R.E.; et al. B cell gene signature with massive intrahepatic production of antibodies to hepatitis B core antigen in hepatitis B virus-associated acute liver failure. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8766. [[CrossRef](#)]
18. Nissim, O.; Melis, M.; Diaz, G.; Kleiner, D.E.; Tice, A.; Fantola, G.; Zamboni, F.; Mishra, L.; Farci, P. Liver Regeneration Signature in Hepatitis B Virus (HBV)-Associated Acute Liver Failure Identified by Gene Expression Profiling. *PLoS ONE* **2012**, *7*, e49611. [[CrossRef](#)]
19. Misu, H.; Takamura, T.; Takayama, H.; Hayashi, H.; Matsuzawa-Nagata, N.; Kurita, S.; Ishikura, K.; Ando, H.; Takeshita, Y.; Ota, T.; et al. A Liver-Derived Secretory Protein, Selenoprotein P, Causes Insulin Resistance. *Cell Metab.* **2010**, *12*, 483–495. [[CrossRef](#)]
20. Affò, S.; Dominguez, M.; Lozano, J.J.; Sancho-Bru, P.; Rodrigo-Torres, D.; Morales-Ibanez, O.; Moreno, M.; Millán, C.; Loaeza-del-Castillo, A.; Altamirano, J.; et al. Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis. *Gut* **2013**, *62*, 452. [[CrossRef](#)] [[PubMed](#)]
21. Diaz, G.; Zamboni, F.; Tice, A.; Farci, P. Integrated ordination of miRNA and mRNA expression profiles. *BMC Genom.* **2015**, *16*, 767. [[CrossRef](#)]
22. Hobbs, B.; Collin, F.; Antonellis, K.J.; Irizarry, R.A.; Speed, T.P.; Scherf, U.; Beazer-Barclay, Y.D. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **2003**, *4*, 249–264. [[CrossRef](#)]
23. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)] [[PubMed](#)]
24. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
25. Zhang, B.; Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [[CrossRef](#)]
26. Le, D.H.; Verbeke, L.; Son, L.H.; Chu, D.T.; Pham, V.H. Random walks on mutual microRNA-target gene interaction network improve the prediction of disease-associated microRNAs. *BMC Bioinform.* **2017**, *18*, 479. [[CrossRef](#)]
27. Li, L.; Wang, Y.; An, L.; Kong, X.; Huang, T. A network-based method using a random walk with restart algorithm and screening tests to identify novel genes associated with Menière’s disease. *PLoS ONE* **2017**, *12*, e0182592. [[CrossRef](#)]
28. Köhler, S.; Bauer, S.; Horn, D.; Robinson, P.N. Walking the Interactome for Prioritization of Candidate Disease Genes. *Am. J. Hum. Genet.* **2008**, *82*, 949–958. [[CrossRef](#)]

29. Wang, S.; Huang, G.; Hu, Q.; Zou, Q. A network-based method for the identification of putative genes related to infertility. *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 2716–2724. [[CrossRef](#)]
30. Yuan, F.; Zhang, Y.-H.; Wan, S.; Wang, S.; Kong, X.-Y. Mining for Candidate Genes Related to Pancreatic Cancer Using Protein-Protein Interactions and a Shortest Path Approach. *Biomed. Res. Int.* **2015**, *2015*, 623121. [[CrossRef](#)]
31. Wang, B.; Yuan, F.; Kong, X.; Hu, L.-D.; Cai, Y.-D. Identifying Novel Candidate Genes Related to Apoptosis from a Protein-Protein Interaction Network. *Comput. Math. Methods Med.* **2015**, *2015*, 715639. [[CrossRef](#)]
32. Yang, L.; Wang, J.; Wang, H.; Lv, Y.; Zuo, Y.; Li, X.; Jiang, W. Analysis and identification of essential genes in humans using topological properties and biological information. *Gene* **2014**, *551*, 138–151. [[CrossRef](#)]
33. Tian, D.; Wenlock, S.; Kabir, M.; Tzotzos, G.; Doig, A.J.; Hentges, K.E. Identifying mouse developmental essential genes using machine learning. *Dis. Models Mech.* **2018**, *11*, dmm034546. [[CrossRef](#)] [[PubMed](#)]
34. Hwang, Y.-C.; Lin, C.-C.; Chang, J.-Y.; Mori, H.; Juan, H.-F.; Huang, H.-C. Predicting essential genes based on network and sequence analysis. *Mol. Biosyst.* **2009**, *5*, 1672–1678. [[CrossRef](#)] [[PubMed](#)]
35. Hagberg, A.; Swart, P.; Chult, D.S. Exploring Network Structure, Dynamics, and Function Using NetworkX. In Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 19–24 August 2008.
36. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
37. Chen, H.; Zhang, Z.; Jiang, S.; Li, R.; Li, W.; Zhao, C.; Hong, H.; Huang, X.; Li, H.; Bo, X. New insights on human essential genes based on integrated analysis and the construction of the HEGIAP web-based platform. *Brief Bioinform.* **2020**, *21*, 1397–1410. [[CrossRef](#)]
38. Shang, D.; Li, C.; Yao, Q.; Yang, H.; Xu, Y.; Han, J.; Li, J.; Su, F.; Zhang, Y.; Zhang, C.; et al. Prioritizing candidate disease metabolites based on global functional relationships between metabolites in the context of metabolic pathways. *PLoS ONE* **2014**, *9*, e104934. [[CrossRef](#)]
39. Li, Y.; Li, J. Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data. *BMC Genom.* **2012**, *13*, S27. [[CrossRef](#)]
40. Gao, Y.; Zhang, T.-Y.; Yuan, Q.; Xia, N.-S. Antibody-mediated immunotherapy against chronic hepatitis B virus infection. *Hum. Vaccin. Immunother.* **2017**, *13*, 1768–1773. [[CrossRef](#)] [[PubMed](#)]
41. Li, D.; He, W.; Liu, X.; Zheng, S.; Qi, Y.; Li, H.; Mao, F.; Liu, J.; Sun, Y.; Pan, L.; et al. A potent human neutralizing antibody Fc-dependently reduces established HBV infections. *eLife* **2017**, *6*, e26738. [[CrossRef](#)]
42. Rösler, C.; Köck, J.; Kann, M.; Malim, M.H.; Blum, H.E.; Baumert, T.F.; von Weizsäcker, F. APOBEC-mediated interference with hepadnavirus production. *Hepatology* **2005**, *42*, 301–309. [[CrossRef](#)]
43. Nguyen, D.H.; Gummuluru, S.; Hu, J. Deamination-Independent Inhibition of Hepatitis B Virus Reverse Transcription by APOBEC3G. *J. Virol.* **2007**, *81*, 4465. [[CrossRef](#)]
44. Yan, Y.; Huang, P.; Mao, K.; He, C.; Xu, Q.; Zhang, M.; Liu, H.; Zhou, Z.; Zhou, Q. Anti-oncogene PTPN13 inactivation by hepatitis B virus X protein counteracts IGF2BP1 to promote hepatocellular carcinoma progression. *Oncogene* **2021**, *40*, 28–45. [[CrossRef](#)] [[PubMed](#)]
45. Sun, W.; Zhang, Y.; Wong, K.C.; Liu, K.; Yang, Y.; Wu, B.; Tong, J.H.M.; Chan, A.W.H.; Chan, H.L.Y.; Yu, J. Increased expression of GATA zinc finger domain containing 1 through gene amplification promotes liver cancer by directly inducing phosphatase of regenerating liver 3. *Hepatology* **2018**, *67*, 2302–2319. [[CrossRef](#)] [[PubMed](#)]
46. Meng, F.; Liu, J.; Lu, T.; Zang, L.; Wang, J.; He, Q.; Zhou, A. SNHG1 knockdown upregulates miR-376a and downregulates FOXK1/Snai1 axis to prevent tumor growth and metastasis in HCC. *Mol. Ther. Oncolytics* **2021**, *21*, 264–277. [[CrossRef](#)]
47. Tian, L.; Chan, L.K.; Ho, D.W.H.; Zhao, L.; Ng, I.O.L. Abstract 2589: Dysregulation of RalGAP2 in hepatocellular carcinoma. *Cancer Res.* **2020**, *80*, 2589. [[CrossRef](#)]
48. Peng, Y.; Li, Y.; Tian, Y.; Ao, G. PDE4a predicts poor prognosis and promotes metastasis by inducing epithelial-mesenchymal transition in hepatocellular carcinoma. *J. Cancer* **2018**, *9*, 2389–2396. [[CrossRef](#)]
49. Zhao, W.; Dai, Y.; Dai, T.; Xie, T.; Su, X.; Li, J.; Zhou, X.; Meng, K.; Zhao, X. TRIP6 promotes cell proliferation in hepatocellular carcinoma via suppression of FOXO3a. *Biochem. Biophys. Res. Commun.* **2017**, *494*, 594–601. [[CrossRef](#)]
50. Chiu, A.P.; Tschida, B.R.; Lo, L.H.; Moriarty, B.S.; Rowlands, D.K.; Largaespada, D.A.; Keng, V.W. Transposon mouse models to elucidate the genetic mechanisms of hepatitis B viral induced hepatocellular carcinoma. *World J. Gastroenterol.* **2015**, *21*, 12157–12170. [[CrossRef](#)]
51. Li, W.; Cui, X.; Huo, Q.; Qi, Y.; Sun, Y.; Tan, M.; Kong, Q. Profile of HBV Integration in the Plasma DNA of Hepatocellular Carcinoma Patients. *Curr. Genomics* **2019**, *20*, 61–68. [[CrossRef](#)]
52. Sun, S.; Nakashima, K.; Ito, M.; Li, Y.; Chida, T.; Takahashi, H.; Watashi, K.; Sawasaki, T.; Wakita, T.; Suzuki, T. Involvement of PUF60 in Transcriptional and Post-transcriptional Regulation of Hepatitis B Virus Pregenomic RNA Expression. *Sci. Rep.* **2017**, *7*, 12874. [[CrossRef](#)]
53. Lin, L.; Chen, S.; Wang, H.; Gao, B.; Kallakury, B.; Bhuvaneshwar, K.; Cahn, K.; Gusev, Y.; Wang, X.; Wu, Y.; et al. SPTBN1 inhibits inflammatory responses and hepatocarcinogenesis via the stabilization of SOCS1 and downregulation of p65 in hepatocellular carcinoma. *Theranostics* **2021**, *11*, 4232–4250. [[CrossRef](#)]

54. Mayinuer, A.; Yassen, M.; Mogushi, K.; Obulhasim, G.; Xieraili, M.; Aihara, A.; Tanaka, S.; Mizushima, H.; Tanaka, H.; Arai, S. Upregulation of Protein Tyrosine Phosphatase Type IVA Member 3 (PTP4A3/PRL-3) is Associated with Tumor Differentiation and a Poor Prognosis in Human Hepatocellular Carcinoma. *Ann. Surg. Oncol.* **2013**, *20*, 305–317. [[CrossRef](#)]
55. Kong, F.; You, H.; Kong, D.; Zheng, K.; Tang, R. The interaction of hepatitis B virus with the ubiquitin proteasome system in viral replication and associated pathogenesis. *Viol. J.* **2019**, *16*, 73. [[CrossRef](#)]
56. Dong, Z.-R.; Zhou, W.; Sun, D.; Yan, Y.-C.; Yang, C.-C.; Yang, Y.-F.; Li, H.-C.; Zhi, X.-T.; Li, T. Role of the E3 Ubiquitin Ligase TRIM4 in Predicting the Prognosis of Hepatocellular Carcinoma. *J. Cancer* **2020**, *11*, 4007–4014. [[CrossRef](#)]
57. Ding, D.; Lou, X.; Hua, D.; Yu, W.; Li, L.; Wang, J.; Gao, F.; Zhao, N.; Ren, G.; Li, L.; et al. Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing-Based Approach. *PLoS Genet.* **2012**, *8*, e1003065. [[CrossRef](#)] [[PubMed](#)]
58. Dou, C.-Y.; Fan, Y.-C.; Cao, C.-J.; Yang, Y.; Wang, K. Sera DNA Methylation of CDH1, DNMT3b and ESR1 Promoters as Biomarker for the Early Diagnosis of Hepatitis B Virus-Related Hepatocellular Carcinoma. *Dig. Dis. Sci.* **2016**, *61*, 1130–1138. [[CrossRef](#)] [[PubMed](#)]
59. Lei, Y.; Xu, X.; Liu, H.; Chen, L.; Zhou, H.; Jiang, J.; Yang, Y.; Wu, B. HBx induces hepatocellular carcinogenesis through ARRB1-mediated autophagy to drive the G1/S cycle. *Autophagy* **2021**, 1–19. [[CrossRef](#)] [[PubMed](#)]
60. Sun, S.; Xu, M.Z.; Poon, R.T.; Day, P.J.; Luk, J.M. Circulating Lamin B1 (LMNB1) Biomarker Detects Early Stages of Liver Cancer in Patients. *J. Proteome Res.* **2010**, *9*, 70–78. [[CrossRef](#)] [[PubMed](#)]
61. Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **2010**, *12*, 56. [[CrossRef](#)] [[PubMed](#)]