



Article

# Identification of Time-Invariant Biomarkers for Non-Genotoxic Hepatocarcinogen Assessment

Shan-Han Huang <sup>1</sup>, Ying-Chi Lin <sup>1,2,3</sup> and Chun-Wei Tung <sup>4,5,\*</sup>

<sup>1</sup> Ph. D. Program in Toxicology, Kaohsiung Medical University, Kaohsiung 80708, Taiwan; u102832002@gmail.com (S.-H.H.); yclin@kmu.edu.tw (Y.-C.L.)

<sup>2</sup> School of Pharmacy, College of Pharmacy, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

<sup>3</sup> Research Center for Environmental Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

<sup>4</sup> Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei 11031, Taiwan

<sup>5</sup> National Institute of Environmental Health Sciences, National Health Research Institutes, Miaoli County 35053, Taiwan

\* Correspondence: cwtung@tmu.edu.tw

Received: 26 May 2020; Accepted: 14 June 2020; Published: 16 June 2020



**Abstract:** Non-genotoxic hepatocarcinogens (NGHCs) can only be confirmed by 2-year rodent studies. Toxicogenomics (TGx) approaches using gene expression profiles from short-term animal studies could enable early assessment of NGHCs. However, high variance in the modulation of the genes had been noted among exposure styles and datasets. Expanding from our previous strategy in identifying consensus biomarkers in multiple experiments, we aimed to identify time-invariant biomarkers for NGHCs in short-term exposure styles and validate their applicability to long-term exposure styles. In this study, nine time-invariant biomarkers, namely A2m, Akr7a3, Aqp7, Ca3, Cdc2a, Cdkn3, Cyp2c11, Ntf3, and Sds, were identified from four large-scale microarray datasets. Machine learning techniques were subsequently employed to assess the prediction performance of the biomarkers. The biomarker set along with the Random Forest models gave the highest median area under the receiver operating characteristic curve (AUC) of 0.824 and a low interquartile range (IQR) variance of 0.036 based on a leave-one-out cross-validation. The application of the models to the external validation datasets achieved high AUC values of greater than or equal to 0.857. Enrichment analysis of the biomarkers inferred the involvement of chronic inflammatory diseases such as liver cirrhosis, fibrosis, and hepatocellular carcinoma in NGHCs. The time-invariant biomarkers provided a robust alternative for NGHC prediction.

**Keywords:** time-invariant biomarkers; non-genotoxic hepatocarcinogens; toxicogenomics; machine learning

## 1. Introduction

Chemical exposure, including those from environmental, diet, and other sources, was estimated to account for about 45–50% of cancer formation [1]. The liver is the most vulnerable organ to chemicals capable of inducing cancers, and many chemicals are known to induce cancer in the liver [2,3]. Based on the pathogenic mechanism, these hepatocarcinogens can be categorized as either genotoxic or non-genotoxic hepatocarcinogens (NGHCs) [4,5]. In contrast to genotoxic hepatocarcinogens, which can be easily identified by in vitro bioassays [6], the assessment of NGHCs relies on long-term rodent bioassays [7]. Although the “gold standard” method provides the quantitative information on dose–response behavior for determining the carcinogenic potential of a chemical, it is hampered by high costs and inefficiency [7]. The development of novel well-validated short-term screening methods is therefore desirable for identifying potential NGHCs for further experimental validation.

Given the diverse mechanisms of action and organ-specificity of NGHCs, toxicogenomic (TGx) models are promising alternative approaches for assessing NGHCs and deciphering the underlying mechanism of the response [8]. Hepatic gene expression signatures derived from 5-day animal models were shown to be better biomarkers for hepatic tumor formation for NGHCs than traditional *in vivo* pathological and genomic biomarkers, such as liver histological changes, serum alanine aminotransferase activity, cytochrome P450 genes, and Tsc-22 or alpha2-macroglobulin messenger RNA [9]. A TGx-based model with 5-day animal data was also shown to have better predictive accuracy than quantitative structure–activity relationship (QSAR) models. To date, DrugMatrix [10], Gene Expression Omnibus accession no. 8858 (GSE8858) [11], and Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (TG-GATEs) [12] are three major large-scale datasets providing gene expression data under various NGHC exposure styles, including single or repeated low, medium, and high (maximum tolerated) doses treated for 1 day, 3 days, 5 days, 7 days, 14 days, and 28 days for TGx model development. A few TGx models were developed using biomarkers identified from single or multiple short-term NGHC microarray datasets [10,13–19]

Although the models performed well for NGHC assessment in their corresponding exposure styles, they derived very different biomarkers [10,13,16]. The results imply that the predictive performance of the reported biomarkers derived from one exposure style may not be useful for another exposure style. The expression of some biomarkers may vary dramatically over a short period of time, even be reversed. A biomarker whose expression varies in different timepoints may not be reliable for NGHCs prediction and mechanism interpretation. Utilizing only time-invariant biomarkers should derive a more reliable and applicable NGHC prediction model.

This study aimed to analyze the pattern of biomarkers and identify the time-invariant biomarkers for NGHCs prediction. Time-invariant biomarkers will be derived from short-term exposure styles and their prediction performance validated based on long-term exposure styles. A total of nine genes were identified as time-invariant biomarkers, including the upregulation of *Akr7a3*, *Aqp7*, *Cdc2a*, and *Cdkn3*, and downregulation of *A2m*, *Ca3*, *Cyp2c11*, *Ntf3*, and *Sds*. The comparison with published biomarkers showed that the time-invariant biomarkers achieved a reliable performance in various short-term exposure styles. The prediction results based on an independent test dataset further confirmed the usefulness of the time-invariant biomarkers. An enrichment analysis of the time-invariant biomarkers was conducted to provide a better inference of the underlying diseases associated with non-genotoxic hepatocarcinogenesis.

## 2. Materials and Methods

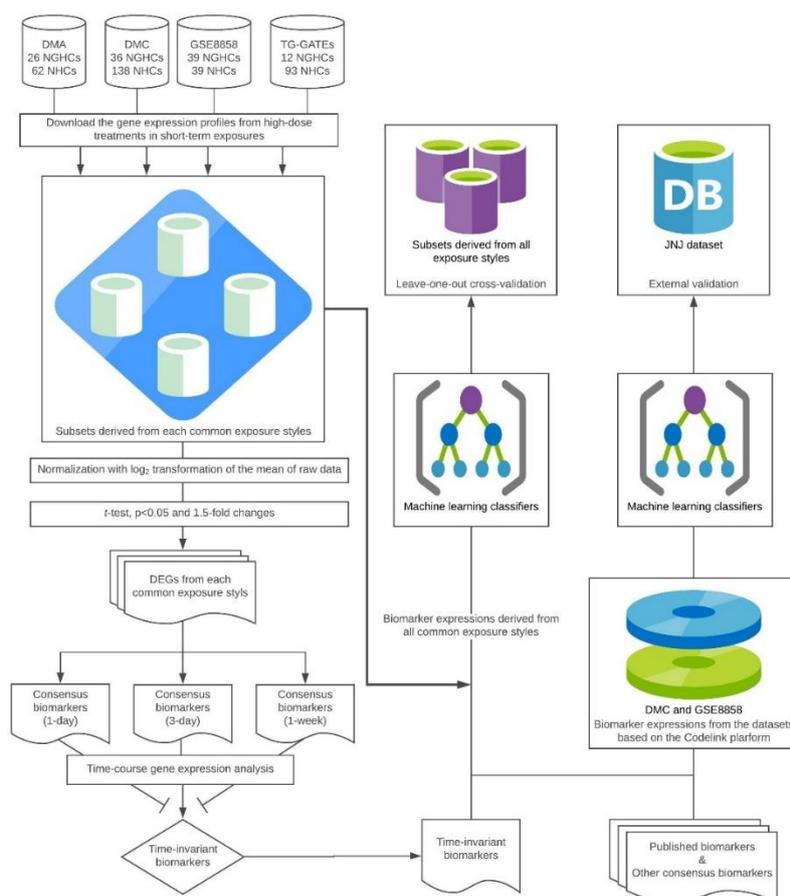
The systematic flow of the time-invariant biomarkers identification and model development analysis is shown in Figure 1

### 2.1. Chemical List and Microarray Datasets

The chemical list utilized in this study has been used in our previous work for consensus biomarkers for predicting NGHCs [16]. In brief, NGHCs and non-hepatocarcinogens (s) consistently classified by several published studies were compiled to get the largest list, and 274 chemicals (50 NGHCs and 224 NHCs) were identified.

Gene expression data from DrugMatrix, GSE8858, and TG-GATEs were analyzed. Based on the platforms utilized, DrugMatrix consists of DrugMatrix with the Affymetrix platform (DMA) and DrugMatrix with the Codelink platform (DMC). GSE8858 utilized the Codelink platform, and the TG-GATEs utilized the Affymetrix platform. The number of chemicals relevant to non-genotoxic hepatocarcinogenesis is 88 for DMA and 174 for DMC, respectively. GSE8858 is a subset of a large liver xenobiotic and pharmacological response database produced by Iconix Biosciences [20], which contains the gene expression profiles of 178 chemicals. A total of 105 chemicals from the TG-GATEs were identified as non-genotoxic hepatocarcinogens according to cytotoxic oxidative stress, one important

mechanism for NGHCs. The numbers of NGHCs:NHCs in the DMA, DMC, GSE8858, and TG-GATEs were 26:62, 36:138, 39:139, and 12:93, respectively.



**Figure 1.** System flow. Abbreviations: DMA, Drug Matrix with Affymetrix platform; DMC, DrugMatrix with Codelink platform; DEGs, differential expression genes. \* Consensus biomarkers (3 days) were identified as the time-invariant biomarkers; GSE8858, Gene Expression Omnibus accession no. 8858; TG-GATEs, Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System; JNJ dataset, Johnson and Johnson dataset; NGHc, non-genotoxic hepatocarcinogens; NHC, non-hepatocarcinogens.

The experimental protocols of all four datasets were similar in animal strain (Sprague–Dawley), sex (male), age (6–8 weeks old), and environmental conditions. Each dose–exposure style experiment (in vivo bioassay) was conducted in biological triplicates. The maximum tolerated dose (MTD) was defined as a 50% reduction in weight gain over the control after a 5-day repeated dose in DrugMatrix and GSE8858. In contrast, the highest dose was set as the dose that induces the minimum toxic effect over the course of a 4-week toxicity study in TG-GATEs.

GSE8858 consists of data from a 1-day single-dose as well as 3- and 5-day repeat once-daily dose experiments at the MTD. DrugMatrix consists of data from 6 h and 1-day single dose and 3- and 5-day repeated once-daily dose experiments at the MTD, 50% of the MTD (mid), and 25% of the MTD (low). TG-GATEs consists of data from 3-, 6-, 9- and 24-h single doses, and repeated once-daily doses of the 3-, 7-, 14- and 28-day experiments at high, middle, and low doses (dose ratio 10:3:1).

To maximize the number of data for subsequent analysis from the referenced databases, we defined the MTD treatments in DrugMatrix and GSE8858 and the highest dose treatment in TG-GATEs as high-dose, and the 5-day and the 7-day exposure styles were grouped as 1-week. The exposure styles of the 1-day, 3-day, and 1-week levels were grouped as short-term exposure, while the 14-day and 28-day levels were grouped as long-term exposure.

The metadata were downloaded from the websites of DrugMatrix (<ftp://anonftp.niehs.nih.gov/drugmatrix/>), GSE8858 (<ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE8858/>) and TG-GATEs (<ftp://ftp.dbcls.jp/archive/open-tggates/>) and imported into the RStudio software environment (RStudio, Boston, MA, USA). All the gene expression profiles were normalized and log<sub>2</sub>-transformed for subsequent analysis.

## 2.2. Identification of the Time-Invariant Biomarker Sets

Three common exposure styles of the referenced datasets, namely the 1-day, 3-day, and 1-week high-dose exposures, were considered for the identification of the time-invariant biomarkers. First, each consensus biomarker set was identified as the overlapped differential expressed genes (DEGs) based on a t-test ( $p < 0.05$ ), and a 1.5-fold change [16], which were derived from each common exposure style of these datasets. Subsequently, each consensus biomarker set was cross-checked with the other two exposure styles. The set of consistently up- or downregulated biomarkers in all three exposure styles was identified as the time-invariant biomarkers.

## 2.3. Model Development

Machine learning classifiers have been widely applied to model the complex relationships between biomarkers and toxicity. In this study, we employed seven well-known classifiers, including decision tree (J48) [21], bagging tree [22], boosting tree [23], k-nearest neighbor (kNN) [24], Naive Bayes (NB) [25], support vector machine (SVM) [26], and Random Forest (RF) [27], to evaluate the reliabilities of the consensus biomarkers. Published biomarker sets, including five genes from Eichner et al. 2014 (E5) [13], 19 genes from Fielden et al. 2011 (F19) [10], and nine genes from Uehara et al. 2011 (U9) [12], were utilized for comparison. E5 and U9 were obtained from the highest dose of the 14-day and 28-day exposure styles in TG-GATEs, respectively, while F19 was identified from the MTD of the 5-day exposure style in DrugMatrix.

The decision tree-based ensemble learning algorithm RF was found to perform best in our datasets. RF improves the prediction performances of decision trees and reduces variance to avoid overfitting based on a set of decision trees built on bootstrap samples from the training dataset and a fixed number of randomly selected features for tree splitting. The prediction of a given sample is based on a majority vote by the fully grown decision trees. The implementation of the RF algorithm was based on WEKA v3.8 (WEKA, Hamilton, New Zealand). The number of features for constructing a fully grown decision tree was set to the default value of the square root of the number of features and genes of each biomarker set. The optimal number of trees ranging from 10 to 100 was determined based on the AUC performance from the leave-one-out cross-validation (LOOCV). All the machine learning algorithms and LOOCV procedures were implemented using the package of WEKA. The variance interquartile range (IQR), as well as the coefficients of variances from the datasets (C.V.d) and exposures (C.V.e) were calculated to assist biomarker evaluation. IQR measured the overall variance based on the AUCs from the models with individual biomarker sets, while C.V.d and C.V.e indicated the source of the variances from the difference in the datasets and exposures, respectively. An IQR, C.V.d, or C.V.e value greater than 5% indicated that the performance of the models was not stable.

## 2.4. External Validation

The Johnson and Johnson dataset (JNJ) dataset [28], consisting of data from the single-dose 1-day experiments at the MTD for 9 NGHCs and 54 NHCs, was utilized for external validation of the developed models. For each chemical, the expression values measured based on the Codelink platform are available for analysis. The raw data were downloaded from the public database of Chemical Effects in Biological Systems [29] and were normalized and log<sub>2</sub>-transformed. Since JNJ utilized the Codelink platform, only models trained with the DMC and GSE8858 datasets, which also utilized the Codelink platform, were evaluated.

## 2.5. Enrichment Analysis

To better understand the roles of the time-invariant biomarkers, enrichment analysis of the Gene Ontology (GO), pathway, and disease terms were conducted based on the Comparative Toxicogenomics Database (CTD) [30]. In the version of August 2018, CTD includes over 2.3 million manually curated chemical–gene, chemical–phenotype, chemical–disease, gene–disease, and chemical–exposure interactions for 15,681 chemicals, 46,689 genes, 4340 phenotypes, and 7212 diseases. For further analysis and hypothesis development, CTD includes over 38 million toxicogenomic relationships, such as internal integration of these direct interactions generating over 24 million gene–disease sets that are statistically ranked and external integration with annotations from GO, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, and BioGRID. In the latest edition, the CTD has maintained and created MEDIC by merging the disease terms from the flat list of the Online Mendelian Inheritance in Man (OMIM) resource into the Medical Subject Headings (MESH) disease hierarchy. A corrected p-value less than 0.05 was considered as the criteria to identify the significantly enriched GO, pathway, and disease terms.

## 3. Results and Discussion

### 3.1. Time-Invariant Biomarkers and Machine Learning Classifiers

By analyzing the consensus biomarkers derived from the 1-day, 3-day, and 1-week experiments, we found that modulation of the biomarkers of the 1-day and 1-week levels were relatively less consistent than the 3-day level. The consensus biomarker set (all genes are consistently up- or downregulated) derived from the 3-day exposure style was found to be time-invariant in the short-term exposure (Table 1). In contrast, the modulation of E5, F19, and U9 varied in different exposure styles. While upregulated genes are the preferred biomarkers due to easy implementation of the diagnosis method, downregulated genes can be also useful as shown in the previous studies [31,32]

To evaluate the classification performance of the time-invariant biomarkers, we implemented seven machine learning algorithms and compared their LOOCV performance for choosing the best classifier. The RF-based models achieved the highest median AUC of 0.817 and the second-lowest variance IQR of 0.041 (Table 2). The Naive Bayes (NB) models yielded a median AUC of 0.800 with a variance IQR of 0.055. Although bagging tree (BaT) yielded the lowest variance IQR of 0.035 and its median AUC was 0.809, its C.V.d and C.V.e were both higher than 5%. Therefore, RF-based models were chosen for the following analysis.

To better understand whether or not the time-invariant biomarkers obtained from the short-term exposure datasets are robust, we further evaluated the prediction performance of the biomarkers on all exposure styles equal or longer than 1 day. The performance is shown in Table 3. The time-invariant biomarker set (consensus 3-day biomarkers) achieved the highest median AUC of 0.824 and a low IQR of 0.036 (Table 3). Consensus 1-week biomarkers achieved a median AUC of 0.810 for all exposure styles; however, its IQR (0.111), C.V.d (7.41%), and C.V.e (7.25%) were all higher than 5%. F19 achieved a median AUC of 0.809 for all exposure styles; however, its IQR (0.085) and C.V.d (6.47%) were both much higher than the time-invariant biomarker set. Please note that the time-invariant biomarker set further improved the median AUC value by 9% compared to our previously published consensus biomarkers obtained from the 1-day exposure style (median AUC of 0.733) [16]. The results indicated that the time-invariant biomarkers can also be applied to the long-term exposure style and still provide good prediction.

**Table 1.** Biomarkers and corresponding modulations for three short-term exposure styles.

Biomarker Set	Gene Symbol	Affymetrix ID	Codelink ID	Modulation			Time-Invariant	Reference
				1D	3D	1W		
Consensus Biomarkers (1-day)	A2m	1367794_at	NM_012488	–	–	–	Yes	Huang and Tung (2017) [16]
	Ca3	1386977_at	NM_019292	–	–	–	Yes	
	Cxcl1	1387316_at	NM_030845	–	–	–	Yes	
	Cyp8b1	1368435_at	NM_031241	–	–	+/-	No	
Consensus Biomarkers (3-day) *	A2m	1367794_at	NM_012488	–	–	–	Yes	This study
	Akr7a3	1368121_at	NM_013215	+	+	+	Yes	
	Aqp7	1368317_at	NM_019157	+	+	+	Yes	
	Ca3	1386977_at	NM_019292	–	–	–	Yes	
	Cdc2a	1367776_at	NM_019296	+	+	+	Yes	
	Cdkn3	1372685_at	BE113362	+	+	+	Yes	
	Cyp2c11	1387328_at	NM_019184	–	–	–	Yes	
	Ntf3	1387267_at	NM_031073	–	–	–	Yes	
	Sds	1369864_a_at	NM_053962	–	–	–	Yes	
	Akr7a3	1368121_at	NM_013215	+	+	+	Yes	
	Aqp7	1368317_at	NM_019157	+	+	+	Yes	
Consensus Biomarkers (1-week)	Atf3	1369268_at	NM_012912	+	+/-	+	No	This study
	beta-sarcoglycan	1374796_at	AI413058	+	+	+	Yes	
	Ca3	1386977_at	NM_019292	–	–	–	Yes	
	Cpt1b	1367742_at	NM_013200	+	+	+	Yes	
	Cyp2c11	1387328_at	NM_019184	–	–	–	Yes	
	Cyp17a1	1387123_at	NM_012753	–	+/-	+	No	
	Ntf3	1387267_at	NM_031073	–	–	–	Yes	
	RGD1562428_predicted	1376296_at	BF387347	+	+	+	Yes	
	Snx10	1383585_at	AI043753	+	+	+	Yes	
	Abcb4	1369161_at	NA	–	–	–	Yes	
E5	Akr7a3	1368121_at	NM_013215	+	+	+	Yes	Eichner et al. (2014) [13]
	Ccng1	1367764_at	NM_012923	+	+	+/-	No	
	Cdkn1a	1387391_at	NM_080782	+	+/-	+/-	No	
	Phlda3	1375224_at	AW520812	+/-	+/-	+	No	

Table 1. Cont.

Biomarker Set	Gene Symbol	Affymetrix ID	Codelink ID	Modulation			Time-Invariant	Reference
				1D	3D	1W		
F19	Akr7a3	1368121_at	NM_013215	+	+	+	Yes	Fielden et al.(2011) [10]
	Aldh1a1	1387022_at	CK222590	+	+	+	Yes	
	Anxa2	1367584_at	AA956299	+/-	+	+	No	
	Btg2	1386994_at	NM_017259	-	+/-	+/-	No	
	Cdkn1a	1387391_at	NM_080782	+	+/-	+/-	No	
	Cited4	1390008_at	NM_053699	+/-	+/-	+/-	No	
	ESTs	NA	BM388029	-	+	+	No	
	Gpr146	1373158_at	NA	-	-	-	Yes	
	Ica1	1367787_at	NM_030844	+	+	+	Yes	
	LitaF	1370928_at	U53184	+/-	+/-	+/-	No	
	Mat1a	1371031_at	X60822	-	-	-	Yes	
	Mgmt	1368311_at	NM_012861	+/-	+	+	No	
	Mt1a	1371237_at	CR458797	-	-	-	Yes	
	Ppia	1398850_at	BI303474	+/-	+/-	+/-	No	
	Prodh2	1389645_at	AI058310	-	-	-	Yes	
	Psmb9	1370186_at	NM_012708	+/-	+/-	+/-	No	
	Tap1	1388149_at	X57523	+/-	+/-	+/-	No	
	Trnt1	1383144_at	AI412002	+/-	+/-	+	No	
	Usp2	1387703_at	NM_053774	+/-	-	+/-	No	
	U9	Abcb1a	1370583_s_at	NA	+	+	+	
Acot9		1379262_at	NA	+	+	+	Yes	
Cd276_1		1395737_at	BF398424	+/-	+/-	+	No	
Cd276_2		1374198_at	NA	+/-	+	+	No	
Cdh13_1		1375719_s_at	NM_138889	+/-	+/-	+/-	No	
Cdh13_2		1373102_at	NA	+/-	+/-	+/-	No	
Ica1		1367787_at	NM_030844	+	+	+	Yes	
Tes		1383401_at	NM_173132	+	+/-	+/-	No	
Tmem184c		1379419_at	NA	+	+	+	Yes	

ID: identity; 1D: 1-day; 3D: 3-day; 1W: 1-week; +: Upregulation of the gene by the non-genotoxic hepatocarcinogens (NGHCs) compared with the non-hepatocarcinogens (NHCs); -: Downregulation of the gene by NGHCs compared with NHCs; +/-: Inconsistent modulations of the biomarkers in the referenced datasets, NA: Not available, \* Time-invariant biomarkers.

**Table 2.** Performance of time-invariant biomarkers using different machine learning algorithms.

Algorithm	Performance (Median AUC from LOOCV)	Variance		
		IQR	C.V.d	C.V.e
Bagging Tree (BaT)	0.809	0.035	5.33%	6.17%
Boosting Tree (BoT)	0.757	0.102	9.37%	9.21%
Decision Tree (J48)	0.598	0.197	22.98%	24.35%
k-Nearest Neighbor (kNN)	0.720	0.071	5.81%	7.09%
Naive Bayes (NB)	0.800	0.055	3.50%	4.00%
Random Forest (RF)	0.817	0.041	4.36%	4.74%
Support Vector Machine (SVM)	0.582	0.084	8.56%	3.07%

Abbreviations: LOOCV, leave-one-out cross-validation; IQR, interquartile range; C.V.d, coefficient of variation from datasets; C.V.e, coefficient of variation from exposures; AUC, area under the receiver operating characteristic curve.

**Table 3.** Performance of the time-invariant, consensus, and published biomarkers using Random Forest.

Signature	Dataset (Exposure Style)	Performance (Median AUC from LOOCV)		Variance (All Exposure)		
		Short-Term <sup>3</sup>	All Exposure <sup>4</sup>	IQR	C.V.d	C.V.e
Consensus biomarkers (1-day)	Multiple datasets <sup>1</sup> (1 day)	0.739	0.733	0.049 *	4.89%	4.02%
Time-invariant biomarkers /Consensus biomarkers (3-day)	Multiple datasets <sup>1</sup> (3 days)	0.817	0.824	0.036 *	4.34%	4.72%
Consensus biomarker (1-week)	Multiple datasets <sup>1</sup> (5 or 7 days <sup>2</sup> )	0.780	0.810	0.111	7.41%	7.25%
E5	TG-GATEs (14 days)	0.656	0.656	0.097	9.04%	9.47%
F19	DrugMatrix (5 days)	0.796	0.809	0.085	6.47%	3.88%
U9	TG-GATEs (28 days)	0.703	0.713	0.057	8.25%	7.09%

Note: <sup>1</sup> DrugMatrix, GSE8858, and TG-GATEs; <sup>2</sup> five days exposure in DrugMatrix and GSE8858, and seven days exposure in TG-GATEs; <sup>3</sup> 1-day, 3-day, and 1-week high-dose exposures; <sup>4</sup> common short-term merged 14 days and 28 days high-dose exposures in TG-GATEs. \* Significant difference ( $p < 0.05$ ). A model with IQR, C.V.d, and C.V.e values less than 0.05 is considered as with a stable performance.

### 3.2. External Validation

An independent dataset, the JNJ dataset, was applied for external validation of the time-invariant biomarkers. Table 4 presents the prediction performance of the different biomarker sets on the JNJ dataset. The models were trained with 1-day datasets of DMC or GSE8858, which are also based on the Codelink platform. For all the other published biomarkers, only the genes that can be identified in the Codelink platform were utilized. The time-invariant biomarkers achieved good performances with the highest AUC values of 0.862 and 0.857 for models based on DMC and GSE8858, respectively. The performance of the time-invariant biomarkers was better than other compared biomarker sets, even though the JNJ dataset was from 1-day experiments. Therefore, the time-invariant biomarkers are expected to be useful for identifying NGHCs regardless of exposure styles.

**Table 4.** Performance of the time-invariant, consensus, and published biomarkers during external validation.

Signature	Performance (AUC from the Training Datasets)	
	DMC	GSE8858
Consensus biomarkers (1-day)	0.753	0.852
Time-invariant biomarkers /Consensus biomarkers (3-day)	0.862	0.857
Consensus biomarker (1-week)	0.820	0.815
E5	0.632	0.562
F19	0.732	0.791
U9	0.338	0.465

Abbreviations: AUC, area under the receiver operating characteristic curve; DMC, DrugMatrix with Codelink platform; GSE8858, Gene Expression Omnibus accession no. 8858.

### 3.3. Analysis of the Time-Invariant Biomarkers

The identified nine time-invariant biomarkers are A2m, Akr7a3, Aqp7, Ca3, Cdc2a, Cdkn3, Cyp2c11, Ntf3, and Sds. A2m encodes a protease inhibitor and cytokine transporter that can inhibit a broad spectrum of proteases and inflammatory cytokines. Ca3 encodes a member of carbonic anhydrase. A2m and Ca3 had been identified as consensus NGHC biomarkers in our previous study [16] and their reduction have been associated with hepatocarcinogenicity of NGHCs [33,34].

Akr7a3 encodes an aldo-keto reductase that plays roles in the detoxification of aldehydes and ketones. It has been identified as an NGHC biomarker by published studies [10,13,17,35–37], in which it was upregulated by oxidative stress, a known tumor-promoting mechanism for NGHCs. The reductase level was also observed to be upregulated in rat hepatoma [38]. Aqp7 encodes a member of the aquaporin channel family, which facilitates the transport of glycerol from adipocytes to the liver. The encoded protein also allows the movement of water and urea across cell membranes. Aqp7 is a significant modulator of whole-body energy metabolism in a wide range of tissues, including in adipocytes and liver cells, in rats and humans [39]. The gene had been reported to be significantly elevated in malignant and borderline liver tumors compared to in benign tumors differentiated using rat liver slices [40], but the role of Aqp7 upregulation in liver tumor formation is still unknown.

Cyp2c11 encodes cytochrome P450 2C11 in rats, which is a functional counterpart of human Cyp2c9. The most abundant male-specific isoform of CYP in rats mediates the hydroxylation of some endogenous steroids, such as testosterone. Ntf3 encodes a neurotrophin protein that is closely associated with both nerve growth factor and brain-derived neurotrophic factor. Downregulation of Cyp2c11 and Ntf3 has been reported to play crucial roles in inflammation [41] and the AhR signaling pathway [42], respectively. These two modulations have been also observed activating following acute and subchronic exposures to NGHCs in previous studies [42]. Sds encodes the L-serine dehydratase, which is involved in the pathway gluconeogenesis; it is also a stress-associated gene that is downregulated after 24 h of treatment of hepatocarcinogens in vivo [43] and remains downregulated during the development of rat liver cancer [44].

Cdc2a, also known as Cdk1 (cyclin-dependent kinase 1), encodes a cell division control protein [45]. During liver regeneration, the essential cell cycle indicates the gene is sufficient to drive the proliferation of all cell types up to mid-gestation [45]. Cdc2a protein was reported frequently augmented in hepatocarcinoma (HCC) tissue, and such dysfunctional cell cycle regulation, which contributes to the generation of cancer stem cells, may promote tumorigenesis [46]. The signature is also one of the up-regulated DEGs promoting cirrhosis to HCC in published bioinformatics analysis [47] and is associated with the oxidative stress by exposure to diethylnitrosamine [15]. Diethylnitrosamine (DEN) is an environmental carcinogen as an initiator for hepatocarcinogenesis. After DEN short-term administration, lipid peroxidation can be detected, as well as overexpression of glutathione-S-transferase Pi (GSE-p); this is considered a marker of initiation in chemical-induced hepatocarcinogenesis. Cdkn3 encodes cyclin-dependent kinase inhibitor 3, which is involved in regulating the cell cycle. The protein acts as a cyclin-dependent kinase inhibitor that selectively binds to Cdk2 kinase to inhibit G1/S transition, as well as form a complex with Mdm2 and p53 to facilitate cell cycle progression [48]. Overexpression of Cdkn3 in HCC was correlated with poor tumor differentiation and advanced tumor stage. Cdkn3 had been reported as part of a vascular invasion signature [49]. In a previous study utilizing bioinformatics-based screening, the upregulation of Cdkn3 was also identified as a marker of transformation from cirrhosis into HCC and correlated with the occurrence, invasion, and recurrence of HCC [50]. Akr7a3 [51], Cdc2a [50,52], and Cdkn3 [50] have also been found as biomarkers for early diagnosis, staging, and prognosis in human liver cancer clinically.

To provide insights into the underlying mechanism, enrichment analysis was conducted to infer the GO, pathway, and disease terms associated with the identified biomarkers. Results show that seven disease terms were identified as significantly associated diseases with adjusted p-values < 0.05. For the GO and pathway terms, the analysis identified no significant GO or pathway terms. Table 5 lists only the inferred significant disease terms of which the corrected p-values were less than 0.05. Liver cirrhosis,

the end-stage of every chronic liver disease including fibrosis, is a major risk factor for primary liver cancer [53]. Chronic inflammation status associated with liver cirrhosis can induce oxidative stress and alter the functions of the oxidant-generating enzymes and oncogenic proteins of the cells and thereby promote liver cancers formation [54]. Chronic inflammation can also facilitate angiogenesis and the growth, invasion, and metastasis of tumor cells to promote cancer development [55]. Fibrosis, the accumulation of collagens in the hepatic extracellular matrix (ECM), could retard the turnover of ECM and results in the activation of growth factor signaling cascades and cell proliferation in the liver, which promote cancer development. More than 80% of HCC, the most common type of liver cancer, develops in fibrotic or cirrhotic livers, suggesting the importance of the two conditions in promoting liver cancer development [56]. Recently, many bioinformatics approaches have identified several key genes and pathways for transforming cirrhosis to HCC, and *Akr7a3*, *Cdc2a*, and *Cdkn3* were identified by these studies [42,50,51].

**Table 5.** Enriched Gene Ontology (GO) and disease terms of the time-invariant biomarkers.

Disease ID.	Disease Name	Involved Genes	Corrected <i>p</i> -Value *
MESH:D008106	Liver Cirrhosis (Experimental)	A2m, Aqp7, Ca3, Cdc2a, Cdkn3, Sds	$2.97 \times 10^{-7}$
MESH:D008103	Liver Cirrhosis	A2m, Aqp7, Ca3, Cdc2a, Cdkn3, Sds	$6.52 \times 10^{-7}$
MESH:D005355	Fibrosis	A2m, Aqp7, Ca3, Cdc2a, Cdkn3, Sds	$9.82 \times 10^{-7}$
MESH:D008107	Liver Diseases	A2m, Akr7a3, Aqp7, Ca3, Cdc2aCdkn3, Sds	$1.40 \times 10^{-6}$
MESH:D004066	Digestive System Diseases	A2m, Akr7a3, Aqp7, Ca3, Cdc2aCdkn3, Sds	$1.61 \times 10^{-5}$
MESH:D006528	Carcinoma, Hepatocellular	A2m Cdc2a, Cdkn3	0.015
MESH:D008113	Liver Neoplasms	A2m Cdc2a, Cdkn3	0.037

\*. The corrected significance of the enrichment was adjusted for multiple testing using the Bonferroni method.

There are some limitations to the study. First, despite the differences in the definitions of the high doses, the data were grouped together in this study. The 5-days and 7-days exposure styles were also grouped as 1-week. Furthermore, the modulation of the genes may be affected by different study designs. Especially, dose level is critical when evaluating chemicals, and a high-dose level may increase specificity compared with a lower-dose level. A previous study also concluded that the optimal exposure style for assessing NGHCs is a 3-day daily high dose [57]. However, given that we are looking for time-invariant biomarkers and the prediction model performed well in the external database, the effect of the differences in the experimental study design should not be an issue. Secondly, animal studies are still needed for the application of the biomarkers for NGHC assessment. The biomarkers were derived from animal experiments. Despite largely shortening the length of the study, animal studies are still needed for the NGHC assessment. As for the concern of species differences between mice and rats, published studies have shown that the biomarkers derived from mice were applicable to rats for classifying genotoxic hepatocarcinogens, NGHCs, and NHCs [58,59]. Future works may be the investigation of species differences in the time-invariant biomarkers identified from this study. Thirdly, caution should be taken when applying the biomarkers or the model for interpretation of human data. Rodents and humans have inherent species differences so that the mechanisms of actions identified from NGHC exposure may not be applied in humans [60].

#### 4. Conclusions

In summary, we have identified nine time-invariant biomarkers based on time-course gene expression data and further developed robust prediction models for NGHCs based on the time-invariant biomarkers. The analysis of the nine genes, namely *A2m*, *Akr7a3*, *Aqp7*, *Ca3*, *Cdc2a*, *Cdkn3*, *Cyp2c11*, *Ntf3*, and *Sds*, revealed the association between NGHCs and chronic inflammatory liver conditions,

including liver cirrhosis and fibrosis. The time-invariant biomarkers derived from the short-term exposure styles were found to be more stable than the other biomarkers. The time-invariant biomarkers and the developed models could be reliable screening methods to prioritize chemicals of potential non-genotoxic hepatocarcinogenesis prior to the traditional 2-year rodent bioassays. The time-invariant biomarkers and their linkage to chronic inflammatory diseases provide a better understanding of the mechanisms of action for chemical-induced carcinogenicity in rodents and their relevance in human risk. From a public health standpoint, the time-invariant biomarkers are expected to improve the accuracy of the NGHC predictions from short-term animal studies, shorten the time and expense associated with the evaluation, and thereby accelerate the safety assessment for potential environmental pollutants and drug candidates. Metabolomics [61,62] may also be potential methods for identifying biomarkers for NGHCs. The integration of biomarkers from genes and metabolites might further improve the accuracy for NGHC identification.

**Author Contributions:** C.-W.T. conceived of the idea and plan of work; S.-H.H. implement programs and performed experiments; S.-H.H., Y.-C.L., and C.-W.T. analyzed data and prepared the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Ministry of Science and Technology of Taiwan (MOST107-2221-E-038-020-MY3), National Health Research Institutes (NHRI-109A1-EMCO-0319204), and Taipei Medical University (TMU108-AE1-B36).

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

## References

1. Anand, P.; Kunnumakkara, A.B.; Sundaram, C.; Harikumar, K.B.; Tharakan, S.T.; Lai, O.S.; Sung, B.; Aggarwal, B.B. Cancer is a preventable disease that requires major lifestyle changes. *Pharm. Res.* **2008**, *25*, 2097–2116. [[CrossRef](#)] [[PubMed](#)]
2. Wogan, G.N. Impacts of chemicals on liver cancer risk. *Semin. Cancer Biol.* **2000**, *10*, 201–210. [[CrossRef](#)]
3. Santos, N.P.; Colaco, A.A.; Oliveira, P.A. Animal models as a tool in hepatocellular carcinoma research: A Review. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **2017**, *39*. [[CrossRef](#)] [[PubMed](#)]
4. Dieter, S. What is the meaning of 'A compound is carcinogenic'? *Toxicol. Rep.* **2018**, *5*, 504–511. [[CrossRef](#)]
5. Butterworth, B.E. Consideration of both genotoxic and nongenotoxic mechanisms in predicting carcinogenic potential. *Mutat. Res.* **1990**, *239*, 117–132. [[CrossRef](#)]
6. Plant, N. Can systems toxicology identify common biomarkers of non-genotoxic carcinogenesis? *Toxicology* **2008**, *254*, 164–169. [[CrossRef](#)] [[PubMed](#)]
7. Waters, M.D.; Jackson, M.; Lea, I. Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutat. Res.* **2010**, *705*, 184–200. [[CrossRef](#)]
8. Schaap, M.M.; Wackers, P.F.; Zwart, E.P.; Huijskens, I.; Jonker, M.J.; Hendriks, G.; Breit, T.M.; van Steeg, H.; van de Water, B.; Luijten, M. A novel toxicogenomics-based approach to categorize (non-)genotoxic carcinogens. *Arch. Toxicol.* **2015**, *89*, 2413–2427. [[CrossRef](#)]
9. Fielden, M.R.; Brennan, R.; Gollub, J. A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2007**, *99*, 90–100. [[CrossRef](#)]
10. Fielden, M.R.; Adai, A.; Dunn, R.T., II; Olaharski, A.; Searfoss, G.; Sina, J.; Aubrecht, J.; Boitier, E.; Nioi, P.; Auerbach, S.; et al. Development and evaluation of a genomic signature for the prediction and mechanistic assessment of nongenotoxic hepatocarcinogens in the rat. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2011**, *124*, 54–74. [[CrossRef](#)] [[PubMed](#)]
11. Liu, Z.; Kelly, R.; Fang, H.; Ding, D.; Tong, W. Comparative analysis of predictive models for nongenotoxic hepatocarcinogenicity using both toxicogenomics and quantitative structure-activity relationships. *Chem. Res. Toxicol.* **2011**, *24*, 1062–1070. [[CrossRef](#)] [[PubMed](#)]
12. Uehara, T.; Ono, A.; Maruyama, T.; Kato, I.; Yamada, H.; Ohno, Y.; Urushidani, T. The Japanese toxicogenomics project: Application of toxicogenomics. *Mol. Nutr. Food Res.* **2010**, *54*, 218–227. [[CrossRef](#)] [[PubMed](#)]

13. Eichner, J.; Wrzodek, C.; Romer, M.; Ellinger-Ziegelbauer, H.; Zell, A. Evaluation of toxicogenomics approaches for assessing the risk of nongenotoxic carcinogenicity in rat liver. *PLoS ONE* **2014**, *9*, e97678. [[CrossRef](#)] [[PubMed](#)]
14. Uehara, T.; Minowa, Y.; Morikawa, Y.; Kondo, C.; Maruyama, T.; Kato, I.; Nakatsu, N.; Igarashi, Y.; Ono, A.; Hayashi, H.; et al. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicol. Appl. Pharmacol.* **2011**, *255*, 297–306. [[CrossRef](#)]
15. Liu, Y.F.; Zha, B.S.; Zhang, H.L.; Zhu, X.J.; Li, Y.H.; Zhu, J.; Guan, X.H.; Feng, Z.Q.; Zhang, J.P. Characteristic gene expression profiles in the progression from liver cirrhosis to carcinoma induced by diethylnitrosamine in a rat model. *J. Exp. Clin. Cancer Res. CR* **2009**, *28*, 107. [[CrossRef](#)] [[PubMed](#)]
16. Huang, S.H.; Tung, C.W. Identification of consensus biomarkers for predicting non-genotoxic hepatocarcinogens. *Sci. Rep.* **2017**, *7*, 41176. [[CrossRef](#)] [[PubMed](#)]
17. Uehara, T.; Hirode, M.; Ono, A.; Kiyosawa, N.; Omura, K.; Shimizu, T.; Mizukawa, Y.; Miyagishima, T.; Nagao, T.; Urushidani, T. A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. *Toxicology* **2008**, *250*, 15–26. [[CrossRef](#)] [[PubMed](#)]
18. Nicolaidou, V.; Koufaris, C. Application of transcriptomic and microRNA profiling in the evaluation of potential liver carcinogens. *Toxicol. Ind. Health* **2020**. [[CrossRef](#)]
19. Tung, C.-W.; Jheng, J.-L. Interpretable prediction of non-genotoxic hepatocarcinogenic chemicals. *Neurocomputing* **2014**, *145*, 68–74. [[CrossRef](#)]
20. Ganter, B.; Tugendreich, S.; Pearson, C.I.; Ayanoglu, E.; Baumhueter, S.; Bostian, K.A.; Brady, L.; Browne, L.J.; Calvin, J.T.; Day, G.J.; et al. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* **2005**, *119*, 219–244. [[CrossRef](#)]
21. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers Inc.: Burlington, MA, USA, 1993.
22. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
23. Schapire Robert, E.; Freund, Y. Boosting: Foundations and Algorithms. *Kybernetes* **2013**, *42*, 164–166. [[CrossRef](#)]
24. Keller, J.M.; Gray, M.R.; Givens, J.A. A fuzzy K-nearest neighbor algorithm. *IEEE Trans. Syst. ManCybern.* **1985**, *SMC-15*, 580–585. [[CrossRef](#)]
25. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA, 4 August 2001; pp. 41–46.
26. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [[CrossRef](#)] [[PubMed](#)]
27. Breiman, L.J.M.I. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
28. Nie, A.Y.; McMillian, M.; Parker, J.B.; Leone, A.; Bryant, S.; Yieh, L.; Bittner, A.; Nelson, J.; Carmen, A.; Wan, J.; et al. Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol. Carcinog.* **2006**, *45*, 914–933. [[CrossRef](#)] [[PubMed](#)]
29. Lea, I.A.; Gong, H.; Paleja, A.; Rashid, A.; Fostel, J. CEBS: A comprehensive annotated database of toxicological data. *Nucleic Acids Res.* **2017**, *45*, D964–D971. [[CrossRef](#)]
30. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; McMorran, R.; Wieggers, J.; Wieggers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Res.* **2019**, *47*, D948–D954. [[CrossRef](#)] [[PubMed](#)]
31. Manica, G.C.; Ribeiro, C.F.; Oliveira, M.A.; Pereira, I.T.; Chequin, A.; Ramos, E.A.; Klassen, L.M.; Sebastião, A.P.; Alvarenga, L.M.; Zanata, S.M.; et al. Down regulation of ADAM33 as a Predictive Biomarker of Aggressive Breast Cancer. *Sci. Rep.* **2017**, *7*, 44414. [[CrossRef](#)] [[PubMed](#)]
32. Ghasemkhani, N.; Shadvar, S.; Masoudi, Y.; Talaei, A.J.; Yahaghi, E.; Goudarzi, P.K.; Shakiba, E. Down-regulated MicroRNA 148b expression as predictive biomarker and its prognostic significance associated with clinicopathological features in non-small-cell lung cancer patients. *Diagn. Pathol.* **2015**, *10*, 164. [[CrossRef](#)]
33. Kuhara, M.; Wang, J.; Flores, M.J.; Qiao, Z.; Koizumi, Y.; Koyota, S.; Taniguchi, N.; Sugiyama, T. Sexual dimorphism in LEC rat liver: Suppression of carbonic anhydrase III by copper accumulation during hepatocarcinogenesis. *Biomed. Res. (Tokyo, Japan)* **2011**, *32*, 111–117. [[CrossRef](#)] [[PubMed](#)]

34. Di Fiore, A.; Monti, D.M.; Scaloni, A.; De Simone, G.; Monti, S.M. Protective Role of Carbonic Anhydrases III and VII in Cellular Defense Mechanisms upon Redox Unbalance. *Oxidative Med. Cell. Longev.* **2018**, *2018*, 2018306. [[CrossRef](#)] [[PubMed](#)]
35. Ellinger-Ziegelbauer, H.; Gmuender, H.; Bandenburg, A.; Ahr, H.J. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat. Res.* **2008**, *637*, 23–39. [[CrossRef](#)] [[PubMed](#)]
36. Nakayama, K.; Kawano, Y.; Kawakami, Y.; Moriwaki, N.; Sekijima, M.; Otsuka, M.; Yakabe, Y.; Miyaura, H.; Saito, K.; Sumida, K.; et al. Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. *Toxicol. Appl. Pharmacol.* **2006**, *217*, 299–307. [[CrossRef](#)]
37. Romer, M.; Eichner, J.; Metzger, U.; Templin, M.F.; Plummer, S.; Ellinger-Ziegelbauer, H.; Zell, A. Cross-platform toxicogenomics for the prediction of non-genotoxic hepatocarcinogenesis in rat. *PLoS ONE* **2014**, *9*, e97640. [[CrossRef](#)]
38. Albrethsen, J.; Miller, L.M.; Novikoff, P.M.; Angeletti, R.H. Gel-based proteomics of liver cancer progression in rat. *Biochim. Biophys. Acta* **2011**, *1814*, 1367–1376. [[CrossRef](#)]
39. Hibuse, T.; Maeda, N.; Nagasawa, A.; Funahashi, T. Aquaporins and glycerol metabolism. *Biochim. Biophys. Acta* **2006**, *1758*, 1004–1011. [[CrossRef](#)]
40. Aikman, B.; de Almeida, A.; Meier-Menches, S.M.; Casini, A. Aquaporins in cancer development: Opportunities for bioinorganic chemistry to contribute novel chemical probes and therapeutic agents. *Met. Integr. Biometal Sci.* **2018**, *10*, 696–712. [[CrossRef](#)]
41. Beltran-Ramirez, O.; Sokol, S.; Le-Berre, V.; Francois, J.M.; Villa-Trevino, S. An approach to the study of gene expression in hepatocarcinogenesis initiation. *Transl. Oncol.* **2010**, *3*, 142–148. [[CrossRef](#)]
42. Ovando, B.J.; Vezina, C.M.; McGarrigle, B.P.; Olson, J.R. Hepatic gene downregulation following acute and subchronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2006**, *94*, 428–438. [[CrossRef](#)]
43. Heise, T.; Schug, M.; Storm, D.; Ellinger-Ziegelbauer, H.; Ahr, H.J.; Hellwig, B.; Rahnenfuhrer, J.; Ghallab, A.; Guenther, G.; Sissnaiske, J.; et al. In vitro-in vivo correlation of gene expression alterations induced by liver carcinogens. *Curr. Med. Chem.* **2012**, *19*, 1721–1730. [[CrossRef](#)] [[PubMed](#)]
44. Xu, C.S.; Wang, G.P.; Zhang, L.X.; Chang, C.F.; Zhi, J.; Hao, Y.P. Correlation between liver cancer occurrence and gene expression profiles in rat liver tissue. *Genet. Mol. Res. GMR* **2011**, *10*, 3480–3513. [[CrossRef](#)] [[PubMed](#)]
45. Malumbres, M.; Barbacid, M. Cell cycle, CDKs and cancer: A changing paradigm. *Nat. Rev. Cancer* **2009**, *9*, 153–166. [[CrossRef](#)] [[PubMed](#)]
46. Wu, C.X.; Wang, X.Q.; Chok, S.H.; Man, K.; Tsang, S.H.Y.; Chan, A.C.Y.; Ma, K.W.; Xia, W.; Cheung, T.T. Blocking CDK1/PDK1/ $\beta$ -Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma. *Theranostics* **2018**, *8*, 3737–3750. [[CrossRef](#)] [[PubMed](#)]
47. He, B.; Yin, J.; Gong, S.; Gu, J.; Xiao, J.; Shi, W.; Ding, W.; He, Y. Bioinformatics analysis of key genes and pathways for hepatocellular carcinoma transformed from cirrhosis. *Medicine* **2017**, *96*, e6938. [[CrossRef](#)]
48. Xing, C.; Xie, H.; Zhou, L.; Zhou, W.; Zhang, W.; Ding, S.; Wei, B.; Yu, X.; Su, R.; Zheng, S. Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. *Biochem. Biophys. Res. Commun.* **2012**, *420*, 29–35. [[CrossRef](#)]
49. Drozdov, I.; Bornschein, J.; Wex, T.; Valeyev, N.V.; Tsoka, S.; Malfetheriner, P. Functional and topological properties in hepatocellular carcinoma transcriptome. *PLoS ONE* **2012**, *7*, e35510. [[CrossRef](#)] [[PubMed](#)]
50. Jiang, C.H.; Yuan, X.; Li, J.F.; Xie, Y.F.; Zhang, A.Z.; Wang, X.L.; Yang, L.; Liu, C.X.; Liang, W.H.; Pang, L.J.; et al. Bioinformatics-based screening of key genes for transformation of liver cirrhosis to hepatocellular carcinoma. *J. Transl. Med.* **2020**, *18*, 40. [[CrossRef](#)]
51. Torres-Mena, J.E.; Salazar-Villegas, K.N.; Sánchez-Rodríguez, R.; López-Gabiño, B.; Del Pozo-Yauner, L.; Arellanes-Robledo, J.; Villa-Treviño, S.; Gutiérrez-Nava, M.A.; Pérez-Carreón, J.I. Aldo-Keto Reductases as Early Biomarkers of Hepatocellular Carcinoma: A Comparison Between Animal Models and Human HCC. *Dig. Dis. Sci.* **2018**, *63*, 934–944. [[CrossRef](#)]
52. Wu, M.; Liu, Z.; Li, X.; Zhang, A.; Lin, D.; Li, N. Analysis of potential key genes in very early hepatocellular carcinoma. *World J. Surg. Oncol.* **2019**, *17*, 77. [[CrossRef](#)]

53. Tarao, K.; Nozaki, A.; Ikeda, T.; Sato, A.; Komatsu, H.; Komatsu, T.; Taguri, M.; Tanaka, K. Real impact of liver cirrhosis on the development of hepatocellular carcinoma in various liver diseases-meta-analytic assessment. *Cancer Med.* **2019**, *8*, 1054–1065. [[CrossRef](#)] [[PubMed](#)]
54. Ohshima, H.; Tatemichi, M.; Sawa, T. Chemical basis of inflammation-induced carcinogenesis. *Arch. Biochem. Biophys.* **2003**, *417*, 3–11. [[CrossRef](#)]
55. Lu, H.; Ouyang, W.; Huang, C. Inflammation, a key event in cancer development. *Mol. Cancer Res. MCR* **2006**, *4*, 221–233. [[CrossRef](#)]
56. Affo, S.; Yu, L.X.; Schwabe, R.F. The Role of Cancer-Associated Fibroblasts and Fibrosis in Liver Cancer. *Annu. Rev. Pathol.* **2017**, *12*, 153–186. [[CrossRef](#)] [[PubMed](#)]
57. Pérez, L.O.; González-José, R.; García, P.P. Prediction of Non-Genotoxic Carcinogenicity Based on Genetic Profiles of Short Term Exposure Assays. *Toxicol. Res.* **2016**, *32*, 289–300. [[CrossRef](#)] [[PubMed](#)]
58. Furihata, C.; Suzuki, T. Evaluation of 12 mouse marker genes in rat toxicogenomics public data, Open TG-GATES: Discrimination of genotoxic from non-genotoxic hepatocarcinogens. *Mutat. Res. Genet. Toxicol. Environ. Mutagenesis* **2019**, *838*, 9–15. [[CrossRef](#)] [[PubMed](#)]
59. Kossler, N.; Matheis, K.A.; Ostefeldt, N.; Bach Toft, D.; Dhalluin, S.; Deschl, U.; Kalkuhl, A. Identification of specific mRNA signatures as fingerprints for carcinogenesis in mice induced by genotoxic and nongenotoxic hepatocarcinogens. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2015**, *143*, 277–295. [[CrossRef](#)] [[PubMed](#)]
60. Felter, S.P.; Foreman, J.E.; Boobis, A.; Corton, J.C.; Doi, A.M.; Flowers, L.; Goodman, J.; Haber, L.T.; Jacobs, A.; Klaunig, J.E.; et al. Human relevance of rodent liver tumors: Key insights from a Toxicology Forum workshop on nongenotoxic modes of action. *Regul. Toxicol. Pharmacol. RTP* **2018**, *92*, 1–7. [[CrossRef](#)]
61. Percival, B.C.; Gibson, M.; Wilson, P.B.; Platt, F.M.; Grootveld, M. Metabolomic Studies of Lipid Storage Disorders, with Special Reference to Niemann-Pick Type C Disease: A Critical Review with Future Perspectives. *Int. J. Mol. Sci.* **2020**, *21*, 2533. [[CrossRef](#)]
62. Leenders, J.; Grootveld, M.; Percival, B.; Gibson, M.; Casanova, F.; Wilson, P.B. Benchtop Low-Frequency 60 MHz NMR Analysis of Urine: A Comparative Metabolomics Investigation. *Metabolites* **2020**, *10*, 155. [[CrossRef](#)]



© 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/>).