Article

Identifying Drug Targets of Oral Squamous Cell Carcinoma through a Systems Biology Method and Genome-Wide Microarray Data for Drug Discovery by Deep Learning and Drug Design Specifications

Yi-Chung Lin and Bor-Sen Chen *

Laboratory of Automatic Control, Signaling Processing and Systems Biology, Department of
Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan
* Correspondence: bschen@ee.nthu.edu.tw

Abstract: In this study, we provide a systems biology method to investigate the carcinogenic mechanism of oral squamous cell carcinoma (OSCC) in order to identify some important biomarkers as drug targets. Further, a systematic drug discovery method with a deep neural network (DNN)-based drug–target interaction (DTI) model and drug design specifications is proposed to design a potential multiple-molecule drug for the medical treatment of OSCC before clinical trials. First, we use big database mining to construct the candidate genome-wide genetic and epigenetic network (GWGEN) including a protein–protein interaction network (PPIN) and a gene regulatory network (GRN) for OSCC and non-OSCC. In the next step, real GWGENs are identified for OSCC and non-OSCC by system identification and system order detection methods based on the OSCC and non-OSCC microarray data, respectively. Then, the principal network projection (PNP) method was used to extract core GWGENs of OSCC and non-OSCC from real GWGENs of OSCC and non-OSCC, respectively. Afterward, core signaling pathways were constructed through the annotation of KEGG pathways, and then the carcinogenic mechanism of OSCC was investigated by comparing the core signal pathways and their downstream abnormal cellular functions of OSCC and non-OSCC. Consequently, HES1, *TCF*, NF-κB and SP1 are identified as significant biomarkers of OSCC. In order to discover multiple molecular drugs for these significant biomarkers (drug targets) of the carcinogenic mechanism of OSCC, we trained a DNN-based drug–target interaction (DTI) model by DTI databases to predict candidate drugs for these significant biomarkers. Finally, drug design specifications such as adequate drug regulation ability, low toxicity and high sensitivity are employed to filter out the appropriate molecular drugs metformin, gefitinib and gallic-acid to combine as a potential multiple-molecule drug for the therapeutic treatment of OSCC.

Keywords: oral squamous cell carcinoma (OSCC); deep neural network-based drug–target interaction (DNN-based DTI) model; genome-wide genetic and epigenetic network (GWGEN); significant biomarkers; drug design specifications

1. Introduction

In recent years, oral cancer has been the eighth most common malignancy of the head and neck, with more than 145,500 deaths worldwide each year [1,2], and oral squamous cell carcinoma (OSCC) accounts for approximate 90% of all cancers in the oral cavity [3]. Despite many advancements in cancer treatment, the 5-year survival rate for OSCC patients is 50% [4], which has remained unchanged over the past decade. In recent years, various environmental factors, such as smoking and chewing betel nut, are the main causes of OSCC [5]. However, not everyone exposed to these triggers will develop oral cancer. Many studies have shown that the occurrence of oral cancer is the result of oncogene activation or tumor suppressor gene inactivation. Therefore, a better understanding of the regulatory

networks between molecular interactions and signaling pathways is crucial for identifying new prognostic markers or therapeutic targets for OSCC.

Chronic inflammation can promote tumor formation. The long-term chronic inflammatory stimulation of periodontal tissue can form an inflammatory microenvironment that is conducive to the development of OSCC [6]. Inflammation activates NF-κB through intracellular signaling pathways, induces the cytokines tumor necrosis factor-α (TNF-α) expression of cytokines and can form an inflammatory microenvironment that promotes tumor growth and progression [7]. Long-term inflammatory damage induces cell renewal and the repair of defective tissues. During the repair process, carcinogens or macrophages can cause cell DNA damage, cell proliferation and differentiation. Disruption occurs, creating conditions for the formation and metastasis of OSCC [8]. NF-κB can also upregulate vascular endothelial growth factor (VEGF) and cyclooxygenase-2 (COX-2) expression [9,10], inducing tumor angiogenesis involved in the invasion and metastasis of OSCC.

In this study, based on big data mining, the system identification method and genome-wide microarray data of patients of OSCC, a systems biology method is proposed to help us analyze macroscopically systematic relationships among proteins, genes and microenvironments in cancer. As shown in Figure 1, the systems biology method including system identification, system order detection and principle network projection (PNP) [11,12] has been widely used to construct core signal pathways to investigate the pathogenesis of diseases such as cancer [13] and virus infection [14] in recent years. Therefore, based on genome-wide microarray data of OSCC and non-OSCC, a systems biology method is used to find the carcinogenic mechanism of OSCC by comparing core signal pathways and their downstream abnormal cellular functions of OSCC and non-OSCC in this study. First, a systems biology method was employed to identify real GWGENs of OSCC and non-OSCC by prune false positives from the candidate GWGEN by their microarray data. Then, with the PNP approach, the core GWGENs of OSCC and non-OSCC were extracted from their real GWGENs, respectively. By the annotation of KEGG pathways, we could obtain core signal pathways of OSCC and non-OSCC from their corresponding core GWGENs. Then, we can investigate the carcinogenic mechanism of OSCC by comparing the discrepancy between core signal pathways and their downstream cellular disfunctions in the non-OSCC and OSCC. According to the investigated carcinogenic mechanism and cellular disfunctions in the core signaling pathways of OSCC, HES1, *TCF*, NF-κB and SP1 were identified as the significant carcinogenic biomarkers contributing to unnormal cellular functions such as inflammatory-dependent cell apoptosis, angiogenesis, tumor metastasis and tumor invasion, which were considered as drug targets for the systematic drug discovery design of OSCC.

The development process of a new drug is an arduous task because of the highly expensive cost and time-consuming investment. Moreover, it is estimated that it takes about 12–15 years and more than USD one billion to develop a new drug [15]. Pharmaceutical companies need to spend a large amount of time and effort on executing experiments to understand the properties and the possible bindings of the drug and selected targets. In addition, the efficacy and potency of the drug as well as the adverse influences on body should be considered. Therefore, researchers conduct a number of animal and clinical trials to check the safety and stability [16]. These complicated procedures vastly increase the risk of failure in designing drugs. Most failures are due to the poor clinical outcomes, and the results are usually lower than expected [17]. On the contrary, drug repurposing (also known as drug repositioning) has been employed to identify new therapeutic uses of approved or investigational drugs as a feasible and advantageous strategy [18]. Recently, deep learning schemes have been widely applied for some phenomenon predictions of molecular biological systems [19–21]. For these reasons, we developed systematic drug discovery and design strategies based on a DNN-based DTI model to predict candidate molecular drugs for biomarkers (drug targets) of OSCC. Then, these candidate molecular drugs of each drug target were sieved by drug design specifications of drug–target interaction (docking), adequate drug regulation ability, low toxicity and high sensitivity to select

adequate molecular drugs for biomarkers, which are combined as a multiple-molecular drug of OSCC, from the perspective of system engineering. As seen in the flowchart of the systematic drug design procedure in a DNN-based DTI model was trained by DTI databases in advance. Then, the well-trained DNN-based DTI model could predict a set of candidate molecular drugs for each biomarker (drug target). Eventually, with the help of the above drug design specifications, we chose and combined metformin, gefitinib and gallic-acid among the set of candidate molecular drugs as the multiple-molecule drug to target the biomarkers HES1, *TCF*, NF-κB and SP1 for OSCC. Taken together, we expect that the proposed systematic medicine discovery and design procedure can provide an efficient way to design a multiple-molecule drug as a new therapy for OSCC treatment before clinical trials.
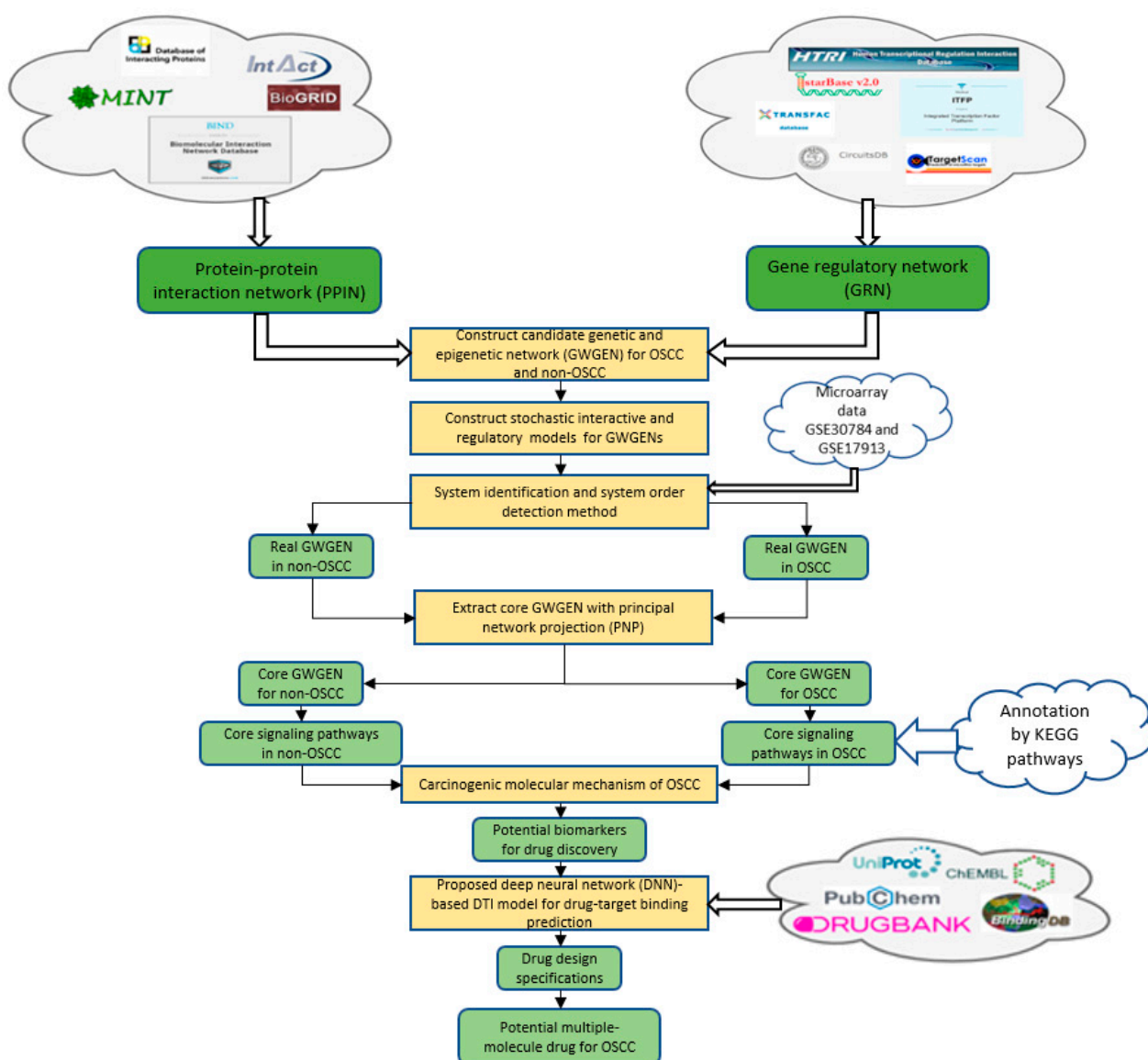

Figure 1. Flowchart of the systems biology method and the outline of the systematic drug discovery design. The candidate GWGEN consists of a gene regulation network (GRN) and protein–protein interaction network (PPIN), where the candidate GRN was constructed through integrating gene regulation databases, and candidate PPI was constructed via protein–protein interaction databases. The candidate GWGEN was identified to obtain real GWGEN by OSCC microarray data from GSE30784 and GSE17913 through system identification and system order detection, and core GWGEN was extracted from real GWGEN by the PNP method. The core signaling pathways of non-OSCC and OSCC are obtained by core GWGENs of non-OSCC and OSCC via the denotation of KEGG pathways,

respectively. The carcinogenic biomarkers were identified by comparing the core signaling pathways and their down streaming abnormal cellular functions of non-OSCC and OSCC. The DNN-based DTI model can be employed to predict candidate molecular drugs for these carcinogenic biomarkers, and drug design specifications are used to select a multiple-molecule drug for OSCC.

2. Results

2.1. Overview of the Systems Biology Method and the Systematic Drug Discovery and Design of OSCC

In order to obtain insight into the carcinogenic mechanism to identify significant carcinogenic biomarkers as drug targets of OSCC, we search for potential molecular drugs to target these significant biomarkers by a deep neural network (DNN)-based DTI model and drug design specifications from the viewpoint of system engineering. The first step is to construct a candidate GWGEN of non-OSCC and OSCC by big data mining from the databases DIP [22], IntAct [23], BioGRID [24], MINT [25] HTRIdb [26], ITFP [27], TRANSFAC [28], CircuitDB [29], TargetScanHuman [30] and starBase 2.0 [31]. Then, the system identification method in Equations (1)–(24) by the microarray data of non-OSCC and OSCC and the system order detection method in Equations (25)–(32) are employed to construct real GWGENs of non-OSCC and OSCC in Figure 2 by pruning off the false positives from candidate GWGEN, respectively. Since, at most, 6000 molecules in real GWGENs can be annotated by the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, the principal network projection method (PNP) in Equations (33)–(40) is used for extracting the core GWGENs in Figure 3, i.e., the core GWGEN network of non-OSCC and the core GWGEN network of OSCC are at the maximum with 6000 significant nodes. The core GWGENs extracted by the PNP method from the real GWGENs can also simplify the investigation of the carcinogenic mechanism of OSCC. The numbers of proteins, TFs, Receptors, LncRNAs and miRNAs of core GWGENs are also indicated. The core signaling pathways of non-OSCC and OSCC are constructed by projecting the corresponding core GWGENs to KEGG significant pathways in Figure 4. By comparing the core signaling pathways and the downstream abnormal cellular functions between non-OSCC and OSCC in Figure 4, we could investigate the carcinogenic mechanism of OSCC, from which the significant biomarkers were identified as drug targets for the therapeutic treatment of OSCC. Furthermore, the deep neural network of the drug–target interaction (DTI) model is trained in Figure 5 to predict candidate molecular drugs, which are selected by drug design specifications such as adequate drug regulatory ability, low toxicity and high sensitivity as potential molecular drugs that can be combined as a multiple-molecule drug of OSCC.
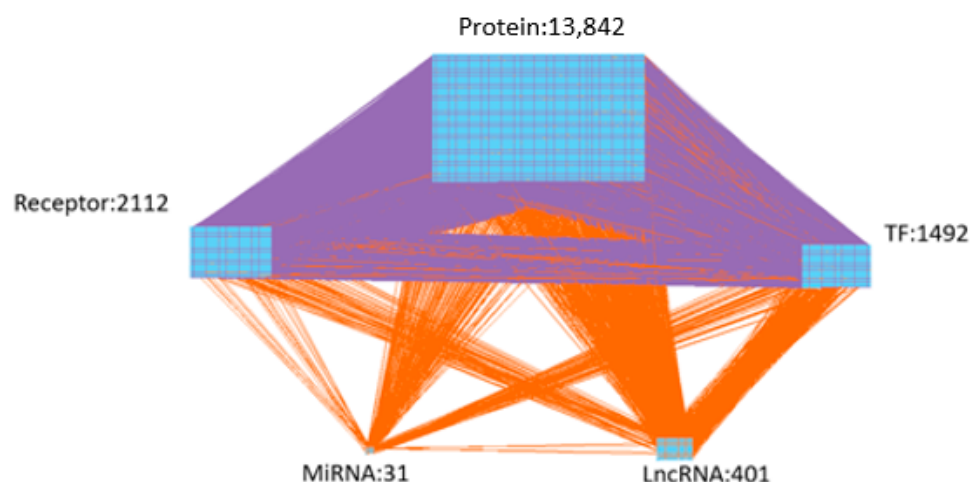
(**A**) Real GWGEN of non-OSCC

Figure 2. *Cont.*

Protein:13,855

Receptor:2112

TF:1492

MiRNA:31

LncRNA:419
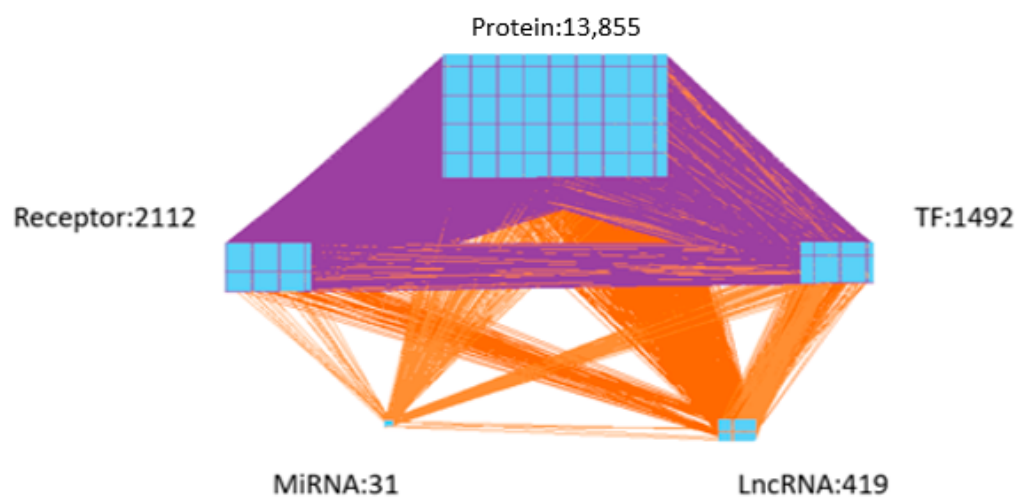
(**B**) Real GWGEN of OSCC

Figure 2. (**A**) The real GWGEN of non-OSCC. (**B**) The real GWGEN of OSCC. The numbers indicate the node numbers of proteins, TFs, Receptors, LncRNAs and miRNAs. The purple lines indicate the protein–protein interactions, and the orange lines denote the gene regulations.

Protein:4452

Receptor:680

TF:703

MiRNA:7

LncRNA:151

(**A**) Core GWGEN of non-OSCC

Protein:4621

Receptor:672

TF:620

MiRNA:2

LncRNA:85
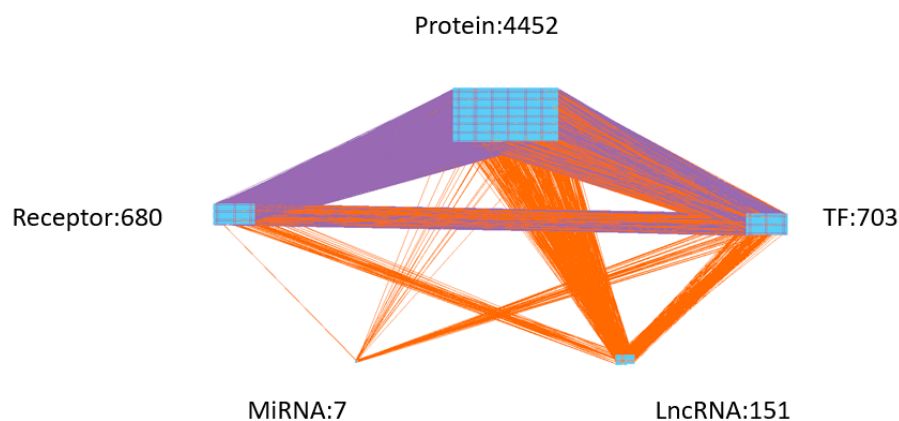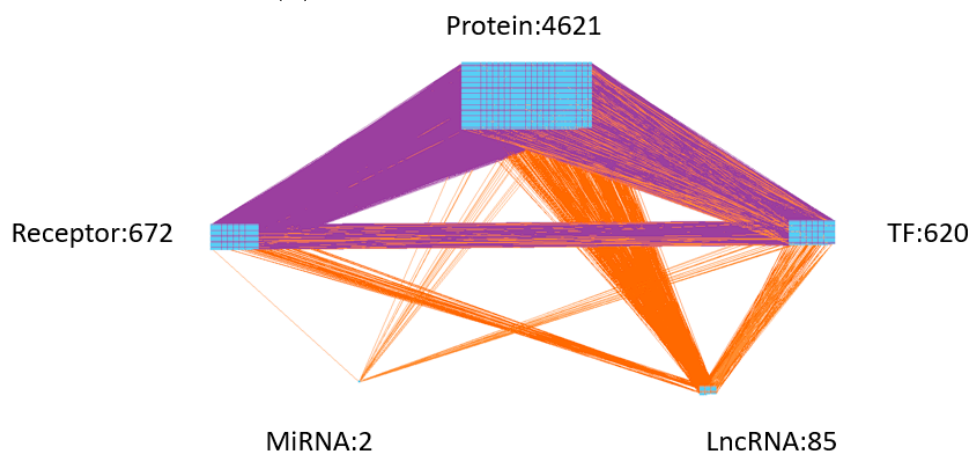
(**B**) Core GWGEN of OSCC

Figure 3. (**A**) The core GWGEN of non-OSCC. (**B**) The core GWGEN of OSCC. The core GWGENs were extracted by the PNP method from the real GWGENs to simplify the annotation of KEGG pathways

for the carcinogenic mechanism analysis of OSCC. The numbers denote the node numbers of proteins, TFs, Receptors, LncRNAs and miRNAs, respectively. The purple lines indicate the protein–protein interactions, and the orange lines denote the gene regulations.
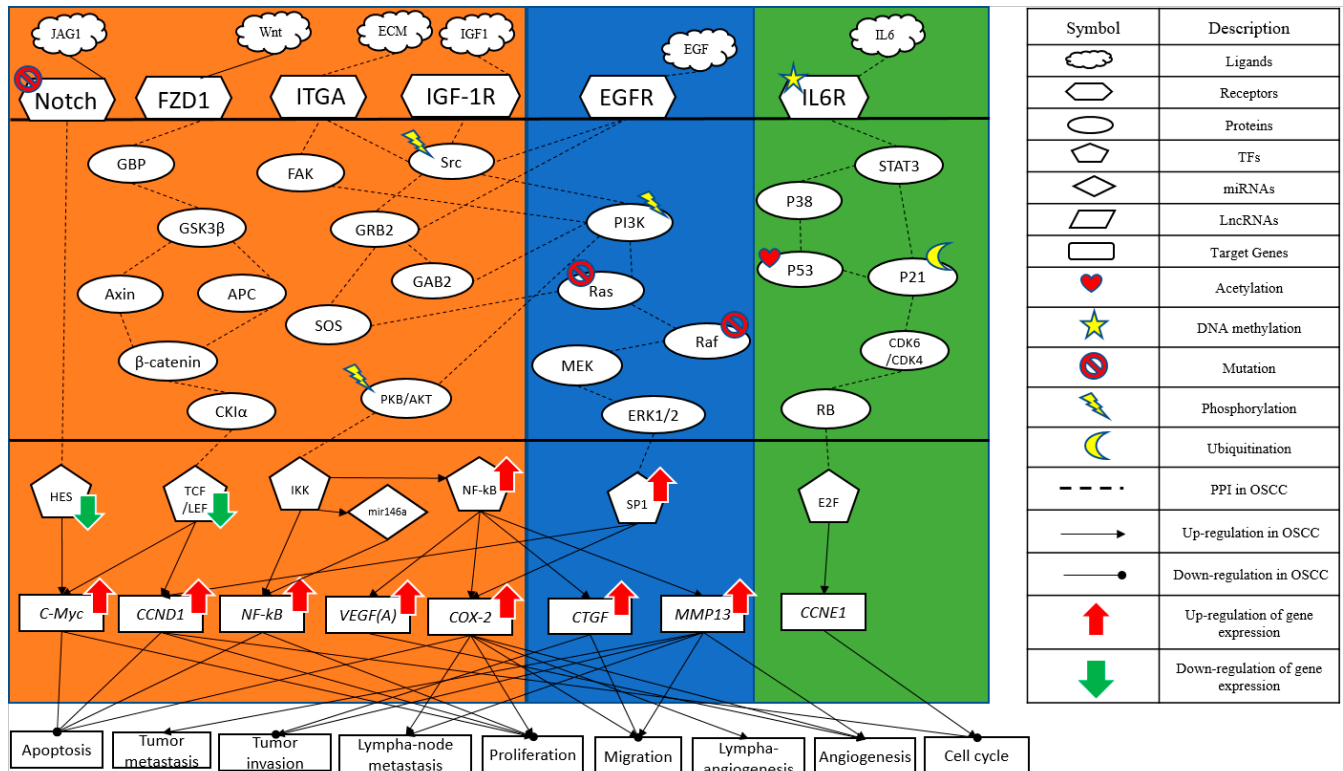

Figure 4. The core signaling pathways in three blocks represent specific OSCC, common non-OSCC and specific non-OSCC core signaling pathways from left to right, respectively. The core signaling pathways of non-OSCC and OSCC are based on the annotation of core GWGENs of non-OSCC and OSCC in Figure 3, respectively. For investigating the genetic and epigenetic carcinogenic mechanism of OSCC, the core signaling pathways and the downstream abnormal cellular functions of non-OSCC and OSCC are compared. The genes and proteins in the core signaling pathways were chosen from core GWGENs of the non-OSCC and OSCC by the annotation of KEGG pathways. The gene regulations and protein interactions were constructed based on the edges in core GWGENs of non-OSCC and OSCC. The low and high expression arrow-head marks are relative to non-OSCC.

The collected microarray data were classified into non-OSCC and OSCC groups, as shown in Table 1.

Table 1. Samples of microarray data from GSE30784 and GSE17913.

Microarray Data	Non-OSCC	OSCC
GSE30784 and GSE17913	102	167

Based on protein interaction and gene regulation databases, since the candidate GWGEN was constructed due to different biological conditions and computational predictions in these databases, there are many false positives in the candidate GWGEN. Therefore, the system identification method in Systems Biology [32,33] is employed to identify the protein interactions and gene regulations by the microarray data of OSCC and non-OSCC. Systems order detection is employed to prune off the false positive protein interactions out of the interaction order of each protein and the gene regulations out of the regulation order of each gene in candidate GWGEN to obtain the real GWGENs of OSCC and non-OSCC. The real GWGENs were too complex and large for analyzing the carcinogenesis of OSCC

by the annotation of KEGG pathways. To solve this problem, the core GWGENs were extracted from the real GWGENs of OSCC and non-OSCC by the PNP method to reduce the network size in order to simplify the following carcinogenic analysis of OSCC by the annotation through KEGG pathways. In Table 2, the sizes of the core GWGENs of OSCC are smaller than real OSCC GWGENs. Both the real OSCC GWGEN and the real non-OSCC were plotted by Cytoscape software in Figure 2. The corresponding core GWGENs for non-OSCC and OSCC were plotted by Cytoscape software in Figure 3. Moreover, the core signaling pathways were obtained by the annotation of core GWGENs by KEGG pathways to investigate the carcinogenic mechanism by comparing core signaling pathways. The KEGG pathway enrichment analyses of the core signaling pathways of OSCC and non-OSCC are given in Tables 3 and 4, respectively. The core signaling pathways for OSCC and non-OSCC are given in Figure 4. Then, based on the core signal pathways and their downstream abnormal cellular functions of OSCC and non-OSCC in Figure 4, we will investigate the genetic and epigenetic carcinogenic mechanism of OSCC in the following.

Table 2. The statistics of the nodes in real GWGEN and core GWGEN of OSCC.

	Real GWGEN of OSCC	Core GWGEN of OSCC
Protein	13,855	4621
Receptor	2112	672
TF	1492	620
miRNA	31	2
LncRNA	419	85
Total nodes	17,909	6000

Table 3. KEGG pathway enrichment analysis of core OSCC signaling pathways.

KEGG Pathway Enrichment Analysis of OSCC Core Signaling Pathways		
Pathway	Gene number	*p*-value
Cell cycle	91	8.3×10^{-20}
Pathways in cancer	267	3.1×10^{-19}
MAPK signaling pathway	149	3.6×10^{-11}
Apoptosis	80	3.2×10^{-10}
Proteoglycans in cancer	120	1.0×10^{-14}

Table 4. KEGG pathway enrichment analysis of core non-OSCC signaling pathways.

KEGG Pathway Enrichment Analysis of Non-OSCC Core Signaling Pathways		
Pathway	Gene number	*p*-value
Cellular senescence	95	2.3×10^{-13}
MAPK signaling pathway	166	2.4×10^{-18}
Human T-cell leukemia virus 1 infection	134	3.2×10^{-18}
Cell cycle	84	4.2×10^{-15}
P53 signaling pathway	46	1.6×10^{-7}

The content in the table shows the number of nodes. The rows of the table contain different types of nodes.

2.2. Investigating the Genetic and Epigenetic Carcinogenic Mechanism of OSCC

Cancer is a disease mainly caused by abnormal signaling pathways. Smoking and drinking can cause changes in the microenvironment of the human body [34,35]. In our

study, adverse factors such as JAG1, Wnt, ECM, EGF and IGF1 produced by the human microenvironment ultimately lead to cellular dysfunctions. From comparing the core signaling pathways of OSCC and non-OSCC in Figure 4, the core signaling pathways of OSCC and their downstream abnormal cellular functions need to be investigated into the genetic and epigenetic carcinogenic mechanism of OSCC. The core signal pathways and their downstream cellular functions of OSCC are investigated as follows:

(i)   Abnormal MAPK signaling pathway in OSCC

In Figure 4, the ligand epithelial growth factor (EGF) in the micro-environment of OSCC targets the EGF receptor (EGFR), activates the phosphorylation of its downstream signaling proteins Src and PI3K and then leads to the well-known estrogen-mediated Ras/Raf/MEK/ERK pathway [36]. Several mutations in the MAPK/ERK pathway have been identified in human cancers. The mitogen-activated protein kinase (MAPK) cascade is a critical pathway for human cancer cell survival, dissemination and resistance to drug therapy. The extracellular signal-regulated kinase (ERK) pathway is a convergent signaling node that receives input from numerous stimuli, including internal metabolic stress and DNA damage pathways and altered protein concentrations, as well as from external growth factors, cell–matrix interactions and communications from other cells. Mutated genes responsible for regulating cell fate, genome integrity and survival can overactivate this pathway by causing increased protein amplification and altering the tumor microenvironment. These mutations can occur upstream of membrane receptor genes [37]. Current and future drug development efforts will require altering and modulating tumor signaling in this complex network of codependent pathways. ERK ultimately redirects to the transcription factor SP1, which is abnormally upregulated in patients with OSCC, and SP1 promotes the expression of CCND1 and COX-2 [38,39]. COX-2 has been reported to be involved in cancer cell migration, cancer cell proliferation, lymph-angiogenesis and metastasis. There was a significant positive correlation between angiogenesis and apoptosis [40]. CCND1 was negatively correlated with apoptosis and an accelerated cell cycle [41,42].

(ii)   The impact of the Wnt signaling pathway on OSCC

The ligand Wnt is common in embryonic development and cancer [43]. Wnts are secreted glycoproteins that bind to frizzled class 1 (FZD1) receptors, which may be coupled to heterotrimeric G proteins. Intracellularly, signal transduction passes through GBP, glycogen synthase kinase 3β (GSK3β) and tumor suppressor gene product (APC) and then activates a key protein (β-catenin). Consequentially, stable β-catenin enters the nucleus and interacts with TF *TCF/LEF*, leading to the transcription of the Wnt target genes *C-Myc* and CCND1 [44]. According to reports and studies [45], there is a regulation of underexpressed TF *TCF/LEF* on *C-Myc* and CCND1. CCND1 was positively correlated with apoptosis and an accelerated cell cycle. *C-Myc* was positively correlated with proliferation and negatively correlated with apoptosis [46].

(iii)   Notch signaling pathway in OSCC

In Figure 4, the cell surface receptor Notch is activated and causes mutation by the microenvironmental factor JAG1 of OSCC [47,48]. The constitutive activation of the Notch pathway has been demonstrated in various types of malignancies [49]. In this study, the Notch pathway was also found to be extremely important in OSCC. The Notch signaling cascade affects several key aspects of normal development through proliferation and apoptosis [50]. All the signaling pathways transduce down to the TF HES1 [51]. Finally, the signal is transmitted to the proto-oncogene *C-Myc* [52]. According to reports and our study, HES1 is underexpressed and upregulates *C-Myc* and then promotes cell apoptosis and proliferation in OSCC patients [50].

(iv)   Key transcription factor NF-κB on OSCC

Integrin subunit alpha (ITGA) also plays a crucial role in OSCC. In Figure 4, it can be seen that, after the receptor ITGA is affected by the extrinsic factor extra cellular matrix

(ECM) in the micro-environment of OSCC [53], it will then successively affect the downstream signaling transduction proteins Src/PI3K/PKB/AKT. Insulin-like growth factor 1 (IGF-1R) signaling is partially mediated by the Src pathway. The activation of Src and IGF-1R also activates Akt. It is a key effector of the PI3K/Akt pathway. It is abnormally activated in most malignant tumors, promoting cell growth, proliferation and survival [54]. In many reports, AKT is underexpressed and phosphorylated in OSCC patients [55]. It then affects the TFs IKK and NF-κB. NF-κB affects numerous target genes. Many reports indicate that the genes COX-2 and VEGF are particularly important in OSCC patients [55,56]. The genes NF-κB, COX-2 and VEGF are abnormally up-regulated to promote angiogenesis, lympha-angiogenesis, migration, proliferation and apoptosis [55,56].

According to the description of OSCC above, poor diet and living habits can lead to changes in the human microenvironment. Then, a series of pathway signaling leads to downstream cellular dysfunction, among which abnormal apoptosis, proliferation, migration and angiogenesis play crucial roles on the carcinogenesis of OSCC.

2.3. Significant Biomarkers as Drug Targets for the Therapeutic Treatment of OSCC Utilizing the Systematic Drug Discovery Approach

After investigating the carcinogenic mechanism of OSCC from the core signaling pathways and their downstream abnormal cellular functions, the significant biomarkers of the carcinogenic mechanism of OSCC will be identified as drug targets for therapeutic treatment.

According to the investigation of carcinogenic mechanisms, OSCC suffers from proliferation and apoptosis. Based on the above core signaling pathways and their downstreaming abnormal cellular functions, we properly selected significant biomarkers that were related to carcinogenically abnormal inflammation, apoptosis, proliferation, angiogenesis and cell cycles. Consequently, we selected HES, *TCF*, NF-κB and SP1 as significant biomarkers and aimed to reverse their expression levels, i.e., to restore them to normal inflammation, apoptosis, proliferation, angiogenesis and cell cycles. HES plays an important role in the NOTCH pathway. *TCF* is also a main character in the Wnt pathway. NF-κB is involved in cellular responses to many stimuli. Sp1 has been shown to be involved in apoptosis.

After identifying the significant biomarkers as drug targets, we considered the chemical properties of drugs and targets to select candidate molecular drugs for these drug targets (biomarkers) based on their drug–target bindings (dockings) predicted by the DNN-based DTI model and to design a multiple-molecule drug for OSCC treatment before clinical trials based on some suitable drug design specifications, i.e., adequate drug regulation ability, low toxicity and high sensitivity. The flowchart of the systematic drug discovery and design method is described in Figure 5.
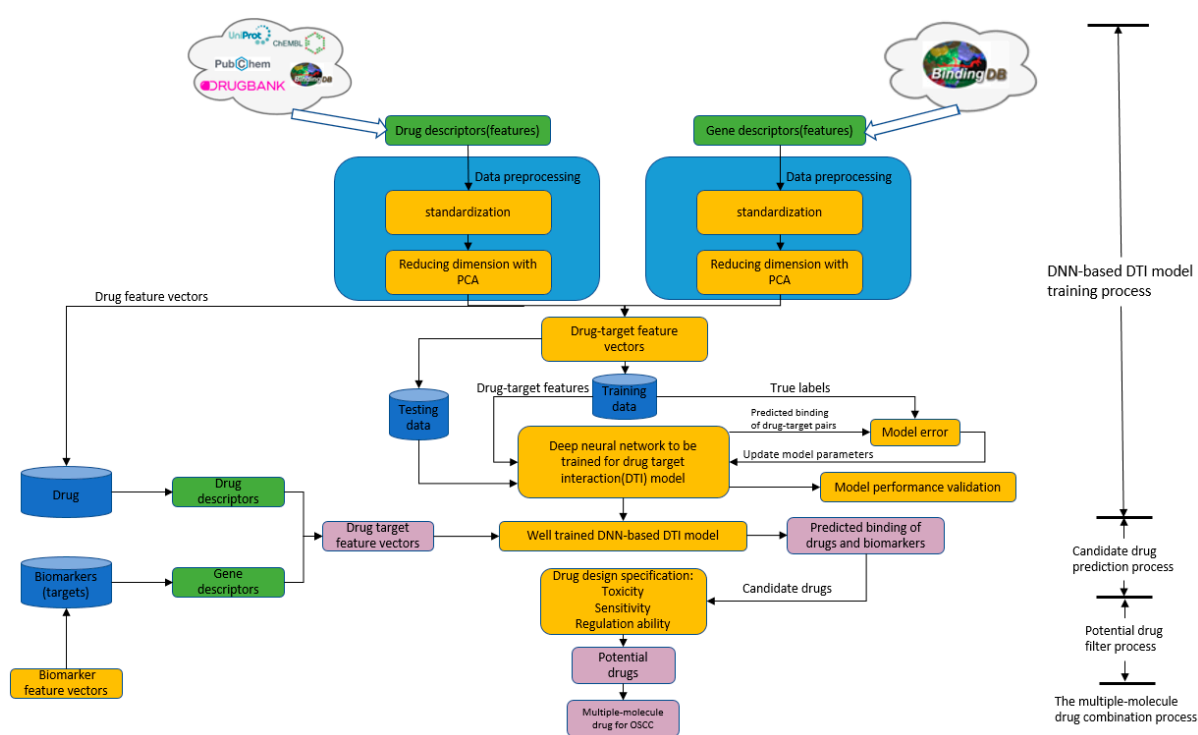
Figure 5. The flowchart of the systematic drug discovery and design procedure for OSCC. The drug–target binding datasets were obtained from the binding database BindingDB, which integrates the information of drugs and targets from several databases. Then, the drug and target features were preprocessed respectively, including descriptor transformation, standardization and PCA dimension reduction. Afterwards, the processed data were split into the training data for DNN-based DTI model training and the testing data for DTI model performance validation in Figures 5 and 6. We updated the model parameters through the model error between the true binding label and the predicted binding label of drug–target pairs. The well-trained DNN-based DTI model was used to predict the binding probability between drugs and targets (biomarkers). Therefore, candidate drugs were predicted for each biomarker in Table 5 by the well-trained DNN-based DTI model from the drug databases and further filtered as potential drugs by the drug design specifications of suitable regulation ability, low toxicity and high sensitivity, which are combined as a multiple-molecule drug for OSCC in Table 6.
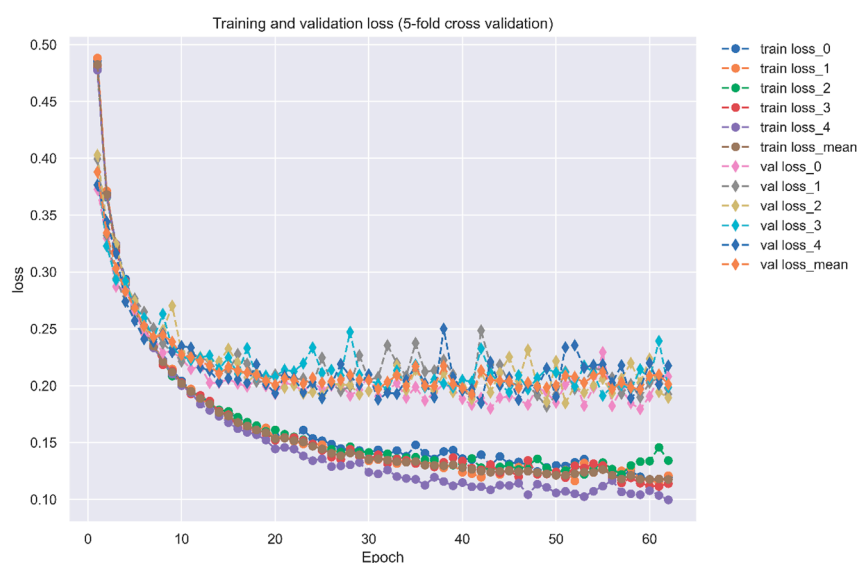

Figure 6. Training and validation loss of the DNN-based DTI model (five-fold cross validation).

Table 5. Some candidate drugs for biomarkers of OSCC and their regulability, toxicity and sensitivity information.
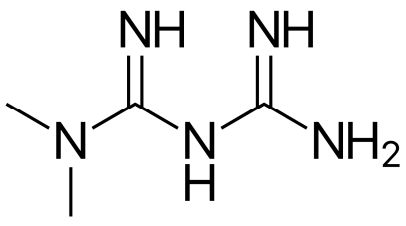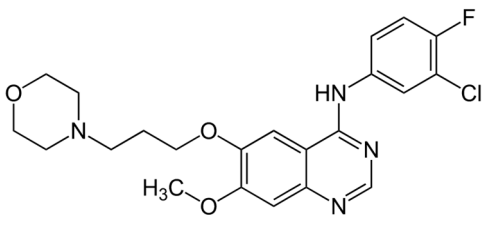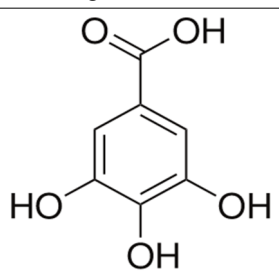
HES1 (-)			
Drug	Regulation Ability (L1000)	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
capsaicin	0.1690	−0.1217	4.202
gabazine	0.3716	−0.6103	3.079
phenolphthalein	−0.6116	−0.4833	5.297
tetramisole	0.1036	0.1136	4.111
gefitinib	0.2750	−0.5144	5.068
TCF (-)			
Drug	Regulation Ability (L1000)	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
carvedilol	−0.0787	−0.0906	5.014
fipronil	−0.1207	−0.0939	5.534
metformin	0.0770	−0.0789	2.039
diethylcarbamazine	0.0501	−0.0848	2.008
dyphylline	0.1372	0.0356	2.022
NF-κB (+)			
Drug	Regulation Ability (L1000)	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
sirolimus	−0.0866	−0.2058	3.486
terfenadine	−0.7665	−0.7406	5.437
metformin	−0.2607	−0.0789	2.039
gallic-acid	−1.0620	0.6208	3.262
gefitinib	−0.3428	−0.5144	5.068
SP1 (+)			
Drug	Regulation Ability (L1000)	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
niridazole	−0.6456	−0.1400	2.746
chlorambucil	−0.0559	−0.1424	3.249
bepridil	0.7249	0.2789	5.083
gallic-acid	−0.5239	0.6208	3.262
disopyramide	−0.3694	−0.1440	3.316

(+), abnormal overexpression; (-), abnormal low expression.

For our DNN-based DTI model in Figure 5, DNN was set with four hidden layers, and each of them is connected with a ReLU activation function layer after it. ReLU activation function could avoid vanishing gradient problems and converge much faster than the other activation functions. Although ReLU is not the best activation function, it is useful for the classification issue. Additionally, the dropout layer is incorporated after each hidden layer to prevent overfitting. The input layer has 996 nodes, and 512, 256, 128 and 64 neurons are embedded, respectively, in four hidden layers. Before the output layer, a sigmoid activation function is used to restrict the output value in the range of 0 to 1 as a probability value. In general, the sigmoid function is usually used for binary classification. The outcome of DTI is a probability value, where a higher value corresponds to a more reliable interaction (docking) between the drug and the target. The loss and accuracy during the training process are separately recorded in Figures 6 and 7, respectively.

Table 6. A potential multiple-molecule drug for OSCC from the candidate molecular drugs in Table 5 by the drug design specifications of suitable regulation ability, low toxicity and high sensitivity.

Drugs \ Targets	HES1	TCF	NF-κB	SP1	Toxicity (LC50, mol/kg)	Sensitivity (PRISM)
metformin		✓(0.0770)	✓(−0.2607)		2.039	−0.0789
gefitinib	✓(0.2750)		✓(−0.3428)		5.068	−0.5144
gallic-acid			✓(−1.0620)	✓(−0.5239)	3.262	0.6208

metformin	gefitinib

gallic-acid

✓ denotes the drug that can bind to the biomarker with a desired regulation ability in (·).

Figure 7. Training and validation accuracy of the DNN-based DTI model (five-fold cross validation).

Furthermore, we also plot the receiver operating characteristic (ROC) curve measure of the probability of the prediction accuracy of the DNN-based DTI model. The visualization of the ROC curve comparison is denoted in Figure 8. From Figure 8, the prediction perfor-

mance of our proposed DTI model indicates that the deep learning concept is promising to adapt to the large and complicated drug–target interaction data.
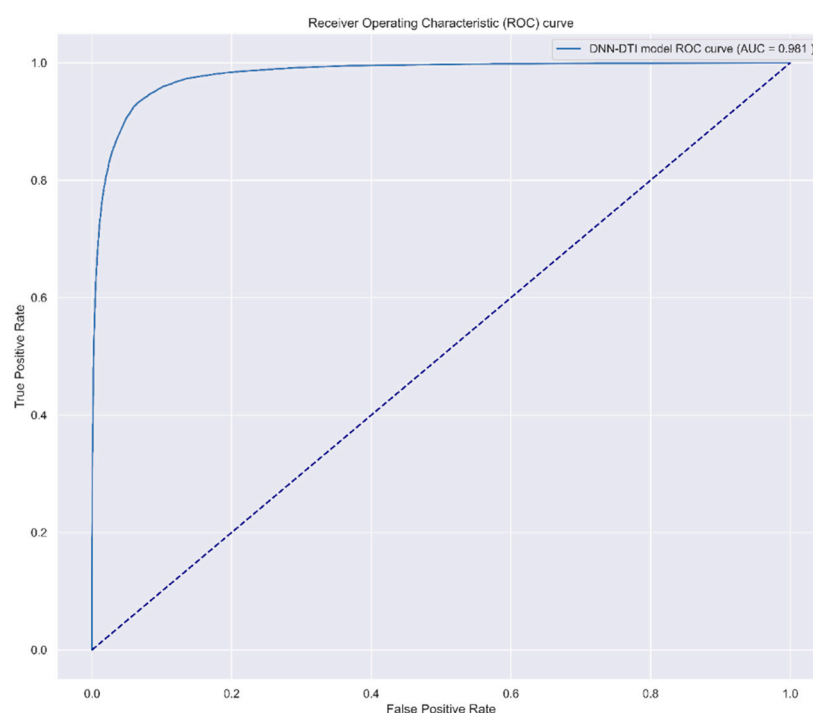

Figure 8. The receiver operating characteristic (ROC) curve measure of the probability of the prediction accuracy of the DNN-based DTI model, with the area under the curve of ROC (AUC-ROC) score of 0.981.

To find suitable drug candidates, predictions are made based on the high probability of the candidate drug binding (docking) to the selected biomarker. At the same time, since powerful drugs are often associated with a high risk of damage, attention should also be paid to the balance between drug efficacy and adverse effects. Correspondingly, we guarantee the stability and safety of the drug in clinical trials, taking into account the drug design specifications, such as regulatory ability, toxicity and sensitivity.

To measure the regulatory ability of drug candidates, available regulatory capacity data were downloaded from the L1000 level5 dataset [57,58], which contains 978 genes treated with 19,811 small molecule compounds in 78 different cell lines. In the accommodation ability data, positive values indicate up-regulations and negative values indicate down-regulations. Based on this criterion, we searched for molecular drugs in suitable cell lines that could reverse the expression of carcinogenic biomarkers in OSCC to restore their normal expressions to remedy their down streaming cellular dysfunctions of OSCC. In addition, a lower drug toxicity has the effect of reducing side effect by referring to the median lethal dose (LD50) value in DrugBank [59]. LD50 is often considered for disease and cancer drug design. The lower the LD50 value, the greater the toxicity. In addition, a higher drug sensitivity (lower EC50 value) can reduce the drug dosage. Drug susceptibility data were obtained from the PRISM dataset [60], which includes 4518 drugs tested in 578 human cell lines based on the half-maximal effective concentration (EC50). EC50 is a measure of the potency of a drug, where a lower EC50 indicates that the drug works best at lower doses. As mentioned above, aberrant expression in a disease can be well reversed by choosing the right drug. Some candidate drugs predicted by the DNN-based DTI model for identified biomarkers and their regulability, toxicity, and sensitivity information are shown in Table 5. The potential drugs metformin, gallic-acid and gefitinib are selected by the three above-mentioned drug design specifications (i.e., suitable regulation ability, high sensitivity and low toxicity) and are combined as a multiple-molecule drug for therapeutic treatment of OSCC in Table 6.

3. Discussion

Potential Multiple-Molecule Drug for the Identified Biomarkers of OSCC

Recently, Cisplatin has been the most common drug for the treatment of OSCC [61,62], but its powerful side effects are daunting; for example, renal toxicity, nausea, vomiting and neurotoxicity [63]. In this study, we tried to find novel multiple-molecular drugs to treat OSCC. With the consideration of sensitivity, toxicity and regulation ability as drug specifications, we combined deep learning and systems biology methods to find the right molecular drugs for the identified biomarkers of OSCC. Ultimately, the molecular drugs metformin, gefitinib and gallic-acid are detected and combined as a multi-molecule drug in Table 6 for the therapeutic treatment of OSCC.

Gefitinib is a drug that acts on the tyrosine kinase domain of the epidermal growth factor receptor [64]. EGFR is overexpressed in certain human cancers, such as breast and lung cancer [65,66]. The excessive activation of EGFR by the ligand EGF in the microenvironment of OSCC leads to the abnormal activation of anti-apoptotic Ras cell signaling, resulting in uncontrolled cell division [67,68]. Gefitinib binds covalently to the enzyme adenosine triphosphate (ATP) and inhibits the epidermal growth factor receptor tyrosine kinase [69].

Metformin, a low-cost antidiabetic drug [70], has been widely used to treat diabetes by inhibiting hepatic gluconeogenesis and enhancing glucose uptake by skeletal muscle [71]. Several studies have shown that metformin is being repurposed as an anticancer therapeutic for different types of cancer [72]. The insulin receptor transmits signals through growth factor receptor binding protein 2 (GRB2) to the Ras/Raf/ERK pathway that drives cell growth. There is evidence that these pathways play an important role in the changes in cellular metabolism that are typical of tumor cells. Elevated circulating insulin/IGF1 levels and the upregulation of insulin/IGF receptor signaling have been implicated in the development of various cancers. Metformin was found to reduce insulin levels, inhibit the insulin/IGF signaling pathway and alter cellular metabolism in normal and cancer cells [73]. Interestingly, metformin can increase the sensitivity of oral cancer cells to chemotherapeutic drugs such as gefitinib [74,75], improving therapeutic efficacy and reducing dose and toxicity [76]. Overall, metformin in combination with gefitinib may be a potential drug for the development of new therapeutic strategies for human OSCC.

Gallic-acid is a phenolic compound widely found in the plant kingdom [77], such as green vegetables, fruits and other plants [78]. The toxicity of gallic-acid to normal cells of humans is very low [79]. It is highly toxic to bad cells such as fibrotic cells and cancer cells [80,81] and can kill these harmful cells; but when it encounters normal cells, the toxicity will become weaker [82].

In summary, the molecular drugs metformin, gefitinib and gallic-acid are selected and combined as a multi-molecule drug of OSCC in Table 6 from the proposed systematic drug discovery and design perspectives.

4. Materials and Methods

4.1. A General Review of Constructing Core Genome-Wide Genetic and Epigenetic Networks (GWGENs) of OSCC and Non-OSCC

At first, we divided the data into a disease group and a control group. The disease group data came from GSE30784, and the control group data came from GSE30784 and GSE17913. Then, we used the following steps to construct core GWGENs of OSCC and non-OSCC.

(1) Constructing the candidate GWGEN: We use big database mining to construct a candidate PPIN and a candidate GRN, including genes, miRNAs and lncRNAs. Note that the candidate GWGEN includes a candidate PPIN and a candidate GRN.

(2) Identifying real GWGENs: We identify the parameters of PPIN and GRN through the system identification method by solving the corresponding constrained linear least squares estimation problems with the help of the microarray data of OSCC and non-OSCC. After performing system modeling and parameter identification for proteins, genes, miRNAs and lncRNAs in the candidate GWGEN, we used the system order

detection method AIC to prune the false positives in the regulation and interactions in the candidate GWGEN to obtain the real GWGENs of OSCC and non-OSCC.

(3) Extracting the core GWGENs: To extract the core GWGENs of OSCC and non-OSCC, we applied the PNP approach. By doing this, we could compute a projection value for each node in the real GWGENs to 85% significant network structures of real GWGENs. The top 6000 nodes of real GWGENs with the highest projection values have remained as core GWGENs.

(4) Building and comparing the core signaling pathways: The core signaling pathways for cells of OSCC and non-OSCC can be constructed by the annotation of the KEGG pathways of core GWGENs of OSCC and non-OSCC, respectively. We investigated the molecular mechanisms of carcinogenesis of OSCC by comparing the upstream microenvironmental factors, core signaling pathways and their corresponding downstream abnormal cellular functions of OSCC and non-OSCC.

4.2. Data Preprocessing for Constructing the Candidate GWGEN

In this research, we downloaded the dataset with accession numbers GSE30784 and GSE17913 from the National Center for Biotechnology Information (NCBI). We divided the data into a disease group and a control group. The disease group data came from GSE30784, and the control group data came from GSE30784 and GSE17913. Note that the candidate GWGEN included a candidate PPIN and a candidate GRN. The candidate GWGEN matrix is a binary matrix. If two nodes have an interaction or regulation, we assigned value one for it; otherwise, we assigned a value zero for it. For building the candidate PPIN, we referred to the following databases: DIP [22], IntAct [23], BioGRID [24] and MINT [25]. Moreover, to construct the candidate GRN, we considered the following databases: HTRIdb [26], ITFP [27], TRANSFAC [28], CircuitDB [29], TargetScanHuman [30] and StarBase 2.0 [31].

4.3. System Modeling of the Candidate GWGEN

To model the candidate GWGEN, we build system modeling for proteins, genes, miRNAs and lncRNAs. First, in the candidate GWGEN, the following interactive equation describes the q-th protein interaction with its G_q neighboring proteins in the candidate PPIN:

$$p_q[n] = \sum_{r=1}^{G_q} \kappa_{qr} p_q[n] p_r[n] + \lambda_{q,PPIM} + \mu_{q,PPIM}[n], \text{ for } q = 1,\dots,Q, n = 1,\dots,N$$

where $p_q[n]$ means the expression level of the q-th protein in the n-th sample and $p_r[n]$ is the expression level of the r-th protein in the n-th sample; κ_{qr} indicates the interaction ability between the q-th protein and the r-th protein; G_q represents the total number of proteins that interact with the q-th protein in the candidate PPIN; Q expresses the total number of proteins in the candidate PPIN; N denotes the total number of samples; $\lambda_{q,PPIM}$ is the basal level in the model of the q-th protein for unknown protein interactions of histone modifications such as phosphorylation, acetylation and ubiquitination; $\mu_{q,PPIM}[n]$ denotes the environment and measurement noise of the q-th protein.

Second, the transcriptional regulation of the x-th gene in GRN is described by the following equation:

$$g_x[n] = \sum_{u=1}^{U_x} \alpha_{xu} t_u[n] + \sum_{v=1}^{V_x} \beta_{xv} l_v[n] - \sum_{w=1}^{W_x} \gamma_{xw} m_w[n] g_x[n] + \lambda_{x,GRM} + \mu_{x,GRM}[n]$$
, for $x = 1,\dots,X, n = 1,\dots,N$

where $g_x[n]$ means the gene expression level of the x-th gene in the n-th sample;

$t_u[n]$, $l_v[n]$ and $m_w[n]$ individually indicate the gene expression level of the u-th TF, the v-th lncRNA and the w-th miRNA in the n-th sample; U_x, V_x and W_x individually denote the total binding number of TFs, lncRNAs and miRNAs; α_{xu} is the transcriptional regulatory ability of the u-th TF on the x-th gene; β_{xv} expresses the transcriptional regulatory ability

of the v-th lncRNA on the x-th gene; $\gamma_{xw} \geq 0$ denotes the post-transcriptional regulatory ability of the w-th miRNA on the x-th gene; X means the total number of genes in the candidate GWGEN; N denotes the total number of samples; $\lambda_{x,GRM}$ means the basal level of the x-th gene because of the unknown gene regulations such as methylation and genetic mutation; $\mu_{x,GRM}[n]$ represents the environment or measurment noise.

Third, TFs, lncRNAs and miRNAs also have a potential impact on the i-th lncRNA, and we can depict this behavior by the lncRNA model (LRM) in the candidate GWGEN. The regulatory equation is described as follows:

$$l_i[n] = \sum_{u=1}^{U_i} \sigma_{iu} t_u[n] + \sum_{v=1}^{V_i} \varsigma_{iv} l_v[n] - \sum_{w=1}^{W_i} \tau_{iw} m_w[n] l_i[n] + \lambda_{i,LRM} + \mu_{i,LRM}[n]$$
$$\text{, for } i = 1, \ldots, I, n = 1, \ldots, N$$

where $l_i[n]$ means the expression level of the i-th lncRNA; $t_u[n]$, $l_v[n]$ and $m_w[n]$ indicate the expression level of the u-th TF, the v-th lncRNA and the the w-th miRNA of the n-th sample, respectively; U_i, V_i and W_i individually represent the total binding number of TFs, lncRNAs and miRNAs on the i-th lncRNAs, respectively; σ_{iu} is the transcriptional regulatory ability from the u-th TF on the i-th lncRNA; ς_{iv} expresses the transcriptional regulatory ability from the v-th lncRNA on the i-th lncRNA; $\tau_{iw} \geq 0$ denotes the post-transcriptional regulatory ability from the w-th miRNA on the i-th lncRNA; I denotes the total number of lncRNAs and N means the total number of samples; $\lambda_{i,LRM}$ indicates the basal level of the i-th lncRNA; $\mu_{i,LRM}[n]$ is the data noise.

Fourth, the expression of the j-th miRNA is also affected by the TFs, lncRNAs and miRNAs. Furthermore, we can illustrate the regulation of miRNA in the miRNA model (MRM) of the candidate GWGEN through the following equation:

$$m_j[n] = \sum_{u=1}^{U_j} \omega_{ju} t_u[n] + \sum_{v=1}^{V_j} \xi_{jv} l_v[n] - \sum_{w=1}^{W_j} \psi_{jw} m_w[n] m_j[n] + \lambda_{j,MRM} + \mu_{j,MRM}[n]$$
$$\text{, for } j = 1, \ldots, J, n = 1, \ldots, N$$

where $m_j[n]$ means the expression level of the j-th miRNA; $t_u[n]$, $l_v[n]$ and $m_w[n]$ separately represent the expression level of the u-th TF, the v-th lncRNA and the w-th miRNA, respectively; U_j, V_j and W_j are the binding total numbers of TFs, lncRNAs and miRNAs, respectively; ω_{ju} expresses the transcriptional regulatory ability from the u-th TF on the j-th miRNA; ξ_{jv} denotes the transcriptional regulatory ability from the v-th lncRNA on the j-th miRNA; ψ_{jw} denotes the post-transcriptional regulatory ability from the w-th miRNA on the j-th miRNA; J indicates the total number of miRNAs and N is the total number of samples; $\lambda_{j,MRM}$ represents the basal level of the j-th miRNA; $\mu_{j,MRM}[n]$ means the data noise.

4.4. The System Identification and System Order Detection Methods for Real GWGENs of OSCC and Non-OSCC from the Candidate GWGEN

According to the interaction and regulation models we described above, the candidate PPIN is generated from PPI models in (1); the candidate GRN is described and combined by the corresponding regulation models in (2)–(4). We obtain the real GWGENs of OSCC and non-OSCC through pruning the false positive interactions and regulations by system identification and system order detection methods based on the microarray data of OSCC and non-OSCC, respectively. To identify the parameters of the above interactive and regulatory models, Equations (1)–(4) could be respectively expressed by the following linear regression forms.

$$p_q[n] = \begin{bmatrix} p_q[n]p_1[n] & p_q[n]p_2[n] & \cdots & p_q[n]p_{G_q}[n] & 1 \end{bmatrix} \times \begin{bmatrix} \kappa_{q1} \\ \kappa_{q2} \\ \vdots \\ \kappa_{qG_q} \\ \lambda_{q,PPIM} \end{bmatrix} + \mu_{q,PPIM}[n] \tag{1}$$

$$g_x[n] = \begin{bmatrix} t_1[n] & \cdots & t_{U_x} & l_1[n] & \cdots & l_{V_x} & m_1[n]g_x[n] & \cdots & m_{W_x}[n]g_x[n] & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_{x1} \\ \vdots \\ \alpha_{xU_x} \\ \beta_1 \\ \vdots \\ \beta_{xV_x} \\ -\gamma_1 \\ \vdots \\ -\gamma_{xW_x} \\ \lambda_{x,GRM} \end{bmatrix} + \mu_{x,GRM}[n] \tag{2}$$

$$l_i[n] = \begin{bmatrix} t_1[n] & \cdots & t_{U_i} & l_1[n] & \cdots & l_{V_i} & m_1[n]l_i[n] & \cdots & m_{W_i}[n]l_i[n] & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{iU_i} \\ \beta_1 \\ \vdots \\ \beta_{iV_i} \\ -\gamma_1 \\ \vdots \\ -\gamma_{iW_i} \\ \lambda_{i,LRM} \end{bmatrix} + \mu_{i,LRM}[n] \tag{3}$$

$$m_j[n] = \begin{bmatrix} t_1[n] & \cdots & t_{U_j} & l_1[n] & \cdots & l_{V_j} & m_1[n]m_j[n] & \cdots & m_{W_j}[n]m_j[n] & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_{j1} \\ \vdots \\ \alpha_{jU_j} \\ \beta_1 \\ \vdots \\ \beta_{jV_j} \\ -\gamma_1 \\ \vdots \\ -\gamma_{jW_j} \\ \lambda_{j,MRM} \end{bmatrix} + \mu_{j,MRM}[n] \tag{4}$$

for $q = 1, \ldots, Q$, $x = 1, \ldots, X$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $n = 1, \ldots, N$, where (5)–(8) are individually the regression forms for the protein interactions in PPIN and the regulations in GRN in the candidate GWGEN. Q, X, I and J are, respectively, the total number of proteins, genes, lncRNAs and miRNAs in the candidate GWGWN, and N means the total number of samples.

The linear regression equations in (5)–(8) could be simply expressed as the following equations:

$$p_q[n] = \phi_{q,PPIM}[n] \cdot \theta_{q,PPIM} + \varepsilon_{q,PPIM}, \text{ for } q = 1, \ldots, Q \tag{5}$$

$$g_x[n] = \phi_{x,GRM}[n] \cdot \theta_{x,GRM} + \varepsilon_{x,GRM}, \text{ for } x = 1, \ldots, X \tag{6}$$

$$l_i[n] = \phi_{i,LRM}[n] \cdot \theta_{i,LRM} + \varepsilon_{i,LRM}, \text{ for } i = 1, \ldots, I \tag{7}$$

$$m_j[n] = \phi_{j,MRM}[n] \cdot \theta_{j,MRM} + \varepsilon_{j,MRM}, \text{ for } j = 1, \ldots, J \tag{8}$$

where $\phi_{q,PPIM}[n]$, $\phi_{x,GRM}[n]$, $\phi_{i,LRM}[n]$ and $\phi_{j,MRM}[n]$ individually mean the regression vectors of proteins, genes, lncRNAs and miRNAs in the candidate GWGEN in the *n*-th sample; $\theta_{q,PPIM}$ is the parameter vector of the protein–protein interaction abilities and the protein basal levels; $\theta_{x,GRM}$, $\theta_{i,LRM}$ and $\theta_{j,MRM}$ respectively express the parameter vectors of the transcriptional regulatory abilities and basal levels of the genes, lncRNAs and miRNAs; $\varepsilon_{q,PPIM}$, $\varepsilon_{x,GRM}$, $\varepsilon_{i,LRM}$ and $\varepsilon_{j,MRM}$ are separately the noises for protein interactions and regulations in the candidate GWGEN.

The above linear regression forms for *N* samples are denoted, respectively, as the following:

$$\begin{bmatrix} p_q[1] \\ p_q[2] \\ \vdots \\ p_q[N] \end{bmatrix} = \begin{bmatrix} \phi_{q,PPIM}[1] \\ \phi_{q,PPIM}[2] \\ \vdots \\ \phi_{q,PPIM}[N] \end{bmatrix} \cdot \theta_{q,PPIM} + \begin{bmatrix} \varepsilon_{q,PPIM}[1] \\ \varepsilon_{q,PPIM}[2] \\ \vdots \\ \varepsilon_{q,PPIM}[N] \end{bmatrix}, \text{ for } q = 1, \ldots, Q \tag{9}$$

$$\begin{bmatrix} g_x[1] \\ g_x[2] \\ \vdots \\ g_x[N] \end{bmatrix} = \begin{bmatrix} \phi_{x,GRM}[1] \\ \phi_{x,GRM}[2] \\ \vdots \\ \phi_{x,GRM}[N] \end{bmatrix} \cdot \theta_{x,GRM} + \begin{bmatrix} \varepsilon_{x,GRM}[1] \\ \varepsilon_{x,GRM}[2] \\ \vdots \\ \varepsilon_{x,GRM}[N] \end{bmatrix}, \text{ for } x = 1, \ldots, X \tag{10}$$

$$\begin{bmatrix} l_i[1] \\ l_i[2] \\ \vdots \\ l_i[N] \end{bmatrix} = \begin{bmatrix} \phi_{i,LRM}[1] \\ \phi_{i,LRM}[2] \\ \vdots \\ \phi_{i,LRM}[N] \end{bmatrix} \cdot \theta_{i,LRM} + \begin{bmatrix} \varepsilon_{i,LRM}[1] \\ \varepsilon_{i,LRM}[2] \\ \vdots \\ \varepsilon_{i,LRM}[N] \end{bmatrix}, \text{ for } i = 1, \ldots, I \tag{11}$$

$$\begin{bmatrix} m_j[1] \\ m_j[2] \\ \vdots \\ m_j[N] \end{bmatrix} = \begin{bmatrix} \phi_{j,MRM}[1] \\ \phi_{j,MRM}[2] \\ \vdots \\ \phi_{j,MRM}[N] \end{bmatrix} \cdot \theta_{j,MRM} + \begin{bmatrix} \varepsilon_{j,MRM}[1] \\ \varepsilon_{j,MRM}[2] \\ \vdots \\ \varepsilon_{j,MRM}[N] \end{bmatrix}, \text{ for } j = 1, \ldots, J \tag{12}$$

The above equations could be individually expressed as the following algebraic equations:

$$P_q = \Phi_{q,PPIM} \cdot \Theta_{q,PPIM} + E_{q,PPIM}, \text{ for } q = 1, \ldots, Q \tag{13}$$

$$G_x = \Phi_{x,GRM} \cdot \Theta_{x,GRM} + E_{x,GRM}, \text{ for } x = 1, \ldots, X \tag{14}$$

$$L_i = \Phi_{i,LRM} \cdot \Theta_{i,LRM} + E_{i,LRM}, \text{ for } i = 1, \ldots, I \tag{15}$$

$$M_j = \Phi_{j,MRM} \cdot \Theta_{j,MRM} + E_{j,MRM}, \text{ for } j = 1, \ldots, J \tag{16}$$

where $\Phi_{q,PPIM}$, $\Phi_{x,GRM}$, $\Phi_{i,LRM}$ and $\Phi_{j,MRM}$ separately express the regression matrix of proteins, genes, lncRNAs and miRNAs of N samples. $\Theta_{q,PPIM}$, $\Theta_{x,GRM}$, $\Theta_{i,LRM}$ and $\Theta_{j,MRM}$ individually mean the corresponding interactive and regulatory parameter vectors. $E_{q,PPIM}$, $E_{x,GRM}$, $E_{i,LRM}$ and $E_{j,MRM}$ respectively denote the corresponding data noise vectors.

To identify the interactive and regulatory parameter of the candidate GWGEN, we estimate the parameter vectors $\theta_{q,PPIM}$, $\theta_{q,GRM}$, $\theta_{q,LRM}$ and $\theta_{q,MRM}$ by the least square method with the negative regulation constraint on miRNA, as follows:

$$\hat{\Theta}_{q,PPIM} = \underset{\Theta_{q,PPIM}}{\operatorname{argmin}} \frac{1}{2} \| \Phi_{q,PPIM} \cdot \Theta_{q,PPIM} - P_q \|_2^2 \tag{17}$$

$$\hat{\Theta}_{x,GRM} = \underset{\Theta_{x,GRM}}{\arg\min} \frac{1}{2} \|\Phi_{x,GRM} \cdot \Theta_{x,GRM} - G_x\|_2^2$$

subject to

$$\begin{bmatrix} \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{U_i} & \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{V_i} & \underbrace{\begin{matrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{matrix}}_{W_i} & \begin{matrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{matrix} \end{bmatrix} \Theta_{i,GRM} \leq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{18}$$

$$\hat{\Theta}_{i,LRM} = \underset{\Theta_{i,LRM}}{\arg\min} \frac{1}{2} \|\Phi_{i,LRM} \cdot \Theta_{i,LRM} - L_i\|_2^2$$

subject to

$$\begin{bmatrix} \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{U_i} & \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{V_i} & \underbrace{\begin{matrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{matrix}}_{W_i} & \begin{matrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{matrix} \end{bmatrix} \Theta_{i,LRM} \leq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{19}$$

$$\hat{\Theta}_{j,MRM} = \underset{\Theta_{j,MRM}}{\arg\min} \frac{1}{2} \|\Phi_{j,MRM} \cdot \Theta_{j,MRM} - M_j\|_2^2$$

subject to

$$\begin{bmatrix} \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{U_j} & \underbrace{\begin{matrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{matrix}}_{V_j} & \underbrace{\begin{matrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{matrix}}_{W_j} & \begin{matrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{matrix} \end{bmatrix} \Theta_{j,MRM} \leq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{20}$$

For the constrained optimization problems for the parameter estimation problems (21)–(24), we look for the optimal parameter estimates of the candidate GWGEN through the microarray data of OSCC and non-OSCC as follows: interactive parameters between proteins $\hat{\Theta}_{q,PPIM}$, the regulatory parameters of genes $\hat{\Theta}_{x,GRM}$, lncRNAs $\hat{\Theta}_{i,LRM}$ and miR-NAs $\hat{\Theta}_{j,MRM}$. These constrained optimization problems for the parameter estimation of the candidate GWGEN were solved by the MATLAB Optimization Toolbox. It is worth noting that the negative inequality constraints in (22)–(24) represent that the regulatory parameters of miRNAs should be less than or equal to zero to ensure the negative regulation of miRNAs on genes, lncRNAs and miRNAs.

After the parameter estimation of the candidate GWGEN of non-OSCC and OSCC by the respective microarray data, we used the system order detection method, AIC, to detect the system order (the number of interactions of each protein or the number of regulations of each gene, lncRNA and miRNA). The equations of AIC for each protein, gene, lncRNA and miRNA are given as follows.

$$AIC(Q_q) = \log(\Omega_{q,PPIM}) + \frac{2(G_q+1)}{N}, \text{ for } q = 1, \dots, Q$$
$$\text{where} \Omega_{q,PPIM} = \frac{(P_q - \Phi_{q,PPIM} \cdot \hat{\Theta}_{q,PPIM})^T (P_q - \Phi_{q,PPIM} \cdot \hat{\Theta}_{q,PPIM})}{N} \tag{21}$$

where $\Omega_{q,PPIM}$ is the estimated residual error of the q-th protein from the least square parameter estimation $\hat{\Theta}_{q,PPIM}$ in (21), and Q_q means the number of protein interactions with the q-th protein.

$$AIC(U_x, V_x, W_x) = \log(\Omega_{x,GRM}) + \frac{2(O_{x,GRM}+1)}{N}, \text{ for } x = 1, \ldots, X$$
$$\text{where } \Omega_{x,GRM} = \frac{(G_x - \Phi_{x,GRM} \cdot \hat{\Theta}_{x,GRM})^T (G_x - \Phi_{x,GRM} \cdot \hat{\Theta}_{x,GRM})}{N}, O_{x,GRM} = U_x + V_x + W_x \tag{22}$$

where $\Omega_{x,GRM}$ means the estimated residual error of the x-th gene in (22), and $O_{x,GRM}$ is the number of regulations of the genes, lncRNAs and miRNAs on the x-th gene; $\hat{\Theta}_{x,GRM}$ denotes the estimated parameters in (22).

$$AIC(U_i, V_i, W_i) = \log(\Omega_{i,LRM}) + \frac{2(O_{i,LRM}+1)}{N}, \text{ for } i = 1, \ldots, I$$
$$\text{where } \Omega_{i,LRM} = \frac{(L_i - \Phi_{i,LRM} \cdot \hat{\Theta}_{i,LRM})^T (L_i - \Phi_{i,LRM} \cdot \hat{\Theta}_{i,LRM})}{N}, O_{i,LRM} = U_i + V_i + W_i \tag{23}$$

where $\Omega_{i,LRM}$ denotes the estimated residual error of the i-th lncRNA in (23), and $O_{i,LRM}$ is the number of regulations of the genes, lncRNAs and miRNAs on the i-th lncRNA; $\hat{\Theta}_{i,LRM}$ indicates the estimated parameters in (23).

$$AIC(U_j, V_j, W_j) = \log(\Omega_{j,MRM}) + \frac{2(O_{j,MRM}+1)}{N}, \text{ for } j = 1, \ldots, J$$
$$\text{where } \Omega_{j,MRM} = \frac{(M_j - \Phi_{j,MRM} \cdot \hat{\Theta}_{j,MRM})^T (M_j - \Phi_{j,MRM} \cdot \hat{\Theta}_{j,MRM})}{N}, O_{j,MRM} = U_j + V_j + W_j \tag{24}$$

where $\Omega_{j,MRM}$ denotes the estimated residual error of the j-th miRNA in (24), and $O_{j,MRM}$ indicates the number of regulations of the genes, lncRNAs and miRNAs on the j-th miRNA; $\hat{\Theta}_{j,MRM}$ denotes the estimated parameters in (24).

The AIC system order detection method in (25) means that if the system order (number interactions) Q_q increases, the second term of AIC in (25) will increase and the first term of AIC will decrease, and vice versa. The real number of interactions (system order) Q_q* will balance two terms and achieve the minimum AIC (Q_q). Therefore, employ the AIC technique to determine the real number of interactions or regulations for each protein, gene, miRNA and lncRNA by their AICs in (25)–(28) in the candidate GWGEN to prune the false positive interactions and regulations and obtain the real GWGENs of OSCC and non-OSCC in the following.

Considering the order detection method of AIC in system identification, the real order of the system (i.e., the total number of interactions of the q-th protein in (1) or the number of regulations on the x-th in (2)) is to minimize the AIC problems of system identification in (21)–(28). Therefore, the real number of interactions or regulations for every protein, gene, lncRNA and miRNA in the candidate GWGEN can be obtained by solving the AIC minimization problems below:

$$Q_q^* = \underset{Q_q}{\text{argmin}} AIC(G_q), \text{ for } q = 1, \ldots, Q \tag{25}$$

$$(U_x^*, V_x^*, W_x^*) = \underset{U_x, V_x, W_x}{\text{argmin}} AIC(U_x, V_x, W_x), \text{ for } x = 1, \ldots, X \tag{26}$$

$$(U_i^*, V_i^*, W_i^*) = \underset{U_i, V_i, W_i}{\text{argmin}} AIC(U_i, V_i, W_i), \text{ for } i = 1, \ldots, I \tag{27}$$

$$(U_j^*, V_j^*, W_j^*) = \underset{U_j, V_j, W_j}{\text{argmin}} AIC(U_j, V_j, W_j), \text{ for } j = 1, \ldots, J \tag{28}$$

where Q_q^* means the real number of protein interactions for the q-th protein; U_x^*, V_x^* and W_x^* respectively express the real number of regulations of genes, lncRNAs and miRNAs on the x-th gene; U_i^*, V_i^* and W_i^* represent the real number of regulations of genes, lncRNAs and miRNAs on the i-th lncRNA, respectively; U_j^*, V_j^* and W_j^* are individually the real number of regulations of genes, lncRNAs and miRNAs on the j-th miRNA. Therefore, the

interactions of proteins and the regulations of the gene, miRNA and lncRNA, which are out of real order by solving the AIC minimization problems in (29)–(32), are thought of as false positives in the candidate GWGEN of non-OSCC and OSCC and should be pruned out to obtain the real GWGENs of non-OSCC and OSCC in Figure 2, respectively.

4.5. The Principal Network Projection (PNP) Method for the Core GWGENs by Extracting from Real GWGENs

We want to compare the signaling pathways of non-OSCC and OSCC to investigate their genetic and epigenetic carcinogenic molecular mechanism. Therefore, we need to transform real GWGENs of non-OSCC and OSCC to signaling pathways of non-OSCC and OSCC, respectively, by the annotation of KEGG pathways. However, at present, only a GWGEN with 6000 nodes at most can be annotated by KEGG signal pathways. Therefore, the principal network projection (PNP) method on the basis of the singular value decomposition (SVD) is employed to extract the core GWGENs with 6000 nodes from the real GWGENs for the annotation of core signaling pathways by KEGG pathways. Before we extract the core GWGENs, a network matrix H of real GWGENs should be introduced. The network matrix H consists of interactions among proteins and regulations of the TF-gene, TF-lncRNA, TF-miRNA, lncRNA-gene, lncRNA-lncRNA, lncRNA-miRNA, miRNA-gene, miRNA-lncRNA and miRNA-miRNA in real GWGENs, as follows:

$$H = \begin{bmatrix} h_{protein \Leftrightarrow protein} & 0 & 0 \\ h_{TF \Rightarrow \text{gene}} & h_{\ln cRNA \Rightarrow gene} & h_{miRNA \Rightarrow gene} \\ h_{TF \Rightarrow \ln cRNA} & h_{\ln cRNA \Rightarrow \ln cRNA} & h_{miRNA \Rightarrow \ln cRNA} \\ h_{TF \Rightarrow miRNA} & h_{\ln cRNA \Rightarrow miRNA} & h_{miRNA \Rightarrow miRNA} \end{bmatrix} \tag{29}$$

where $h_{protein \Leftrightarrow protein}$ is the sub-matrix of PPIs, of which the bidirectional arrow at the subscript of the sub-matrix means that the protein interaction is bidirectional; $h_{TF \Rightarrow \text{gene}}$, $h_{TF \Rightarrow \ln cRNA}$, $h_{TF \Rightarrow miRNA}$, $h_{\ln cRNA \Rightarrow gene}$, $h_{\ln cRNA \Rightarrow \ln cRNA}$, $h_{\ln cRNA \Rightarrow miRNA}$, $h_{miRNA \Rightarrow gene}$, $h_{miRNA \Rightarrow \ln cRNA}$ and $h_{miRNA \Rightarrow miRNA}$ represent the transcriptional regulatory sub-matrixes of TFs on genes, lncRNAs and miRNAs; lncRNAs on genes, lncRNAs and miRNAs; and miRNAs on genes, lncRNAs and miRNAs, respectively. The network matrix H of real GWGENs is given in detail, as follows:

$$H = \begin{bmatrix}
\hat{\kappa}_{11} & \hat{\kappa}_{12} & \cdots & \hat{\kappa}_{1r} & \cdots & \hat{\kappa}_{1Q_q} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\hat{\kappa}_{21} & \hat{\kappa}_{22} & \cdots & \hat{\kappa}_{2r} & \cdots & \hat{\kappa}_{2Q_q} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\kappa}_{q1} & \hat{\kappa}_{q2} & \cdots & \hat{\kappa}_{qr} & \cdots & \hat{\kappa}_{qQ_q} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\kappa}_{Q1} & \hat{\kappa}_{Q1} & \cdots & \hat{\kappa}_{Q1} & \cdots & \hat{\kappa}_{QQ_q} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\hat{\alpha}_{11} & \hat{\alpha}_{12} & \cdots & \hat{\alpha}_{1u} & \cdots & \hat{\alpha}_{1U_x} & \hat{\beta}_{11} & \hat{\beta}_{12} & \cdots & \hat{\beta}_{1v} & \cdots & \hat{\beta}_{1V_x} & \hat{\gamma}_{11} & \hat{\gamma}_{12} & \cdots & \hat{\gamma}_{1w} & \cdots & \hat{\gamma}_{1W_x} \\
\hat{\alpha}_{21} & \hat{\alpha}_{22} & \cdots & \hat{\alpha}_{2u} & \cdots & \hat{\alpha}_{2U_x} & \hat{\beta}_{21} & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2v} & \cdots & \hat{\beta}_{2V_x} & \hat{\omega}_{21} & \hat{\omega}_{22} & \cdots & \hat{\gamma}_{2w} & \cdots & \hat{\gamma}_{2W_x} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\alpha}_{x1} & \hat{\alpha}_{x2} & \cdots & \hat{\alpha}_{xu} & \cdots & \hat{\alpha}_{xU_x} & \hat{\beta}_{x1} & \hat{\beta}_{x2} & \cdots & \hat{\beta}_{xv} & \cdots & \hat{\beta}_{xV_x} & \hat{\gamma}_{x1} & \hat{\gamma}_{x2} & \cdots & \hat{\gamma}_{xw} & \cdots & \hat{\gamma}_{xW_x} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\alpha}_{X1} & \hat{\alpha}_{X2} & \cdots & \hat{\alpha}_{Xu} & \cdots & \hat{\alpha}_{XU_x} & \hat{\beta}_{X1} & \hat{\beta}_{X2} & \cdots & \hat{\beta}_{Xv} & \cdots & \hat{\beta}_{XV_x} & \hat{\gamma}_{X1} & \hat{\gamma}_{X2} & \cdots & \hat{\gamma}_{Xw} & \cdots & \hat{\gamma}_{XW_x} \\
\hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1u} & \cdots & \hat{\sigma}_{1U_i} & \hat{\varsigma}_{11} & \hat{\varsigma}_{12} & \cdots & \hat{\varsigma}_{1v} & \cdots & \hat{\varsigma}_{1V_i} & \hat{\tau}_{11} & \hat{\tau}_{12} & \cdots & \hat{\tau}_{1w} & \cdots & \hat{\tau}_{1W_i} \\
\hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2u} & \cdots & \hat{\sigma}_{2U_i} & \hat{\varsigma}_{21} & \hat{\varsigma}_{22} & \cdots & \hat{\varsigma}_{2v} & \cdots & \hat{\varsigma}_{2V_i} & \hat{\tau}_{21} & \hat{\tau}_{22} & \cdots & \hat{\tau}_{2w} & \cdots & \hat{\tau}_{2W_i} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{i1} & \hat{\sigma}_{i2} & \cdots & \hat{\sigma}_{iu} & \cdots & \hat{\sigma}_{iU_i} & \hat{\varsigma}_{i1} & \hat{\varsigma}_{i2} & \cdots & \hat{\varsigma}_{iv} & \cdots & \hat{\varsigma}_{iV_i} & \hat{\tau}_{i1} & \hat{\tau}_{i2} & \cdots & \hat{\tau}_{iw} & \cdots & \hat{\tau}_{iW_i} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{I1} & \hat{\sigma}_{I2} & \cdots & \hat{\sigma}_{Iu} & \cdots & \hat{\sigma}_{IU_i} & \hat{\varsigma}_{I1} & \hat{\varsigma}_{I2} & \cdots & \hat{\varsigma}_{Iv} & \cdots & \hat{\varsigma}_{IV_i} & \hat{\tau}_{I1} & \hat{\tau}_{I2} & \cdots & \hat{\tau}_{Iw} & \cdots & \hat{\tau}_{IW_i} \\
\hat{\omega}_{11} & \hat{\omega}_{12} & \cdots & \hat{\omega}_{1u} & \cdots & \hat{\omega}_{1U_j} & \hat{\hat{\varsigma}}_{11} & \hat{\hat{\varsigma}}_{12} & \cdots & \hat{\hat{\varsigma}}_{1v} & \cdots & \hat{\hat{\varsigma}}_{1U_j} & \hat{\psi}_{11} & \hat{\psi}_{12} & \cdots & \hat{\psi}_{1w} & \cdots & \hat{\psi}_{1W_j} \\
\hat{\omega}_{21} & \hat{\omega}_{22} & \cdots & \hat{\omega}_{2u} & \cdots & \hat{\omega}_{2U_j} & \hat{\hat{\varsigma}}_{21} & \hat{\hat{\varsigma}}_{22} & \cdots & \hat{\hat{\varsigma}}_{2v} & \cdots & \hat{\hat{\varsigma}}_{2U_j} & \hat{\psi}_{21} & \hat{\psi}_{22} & \cdots & \hat{\psi}_{2w} & \cdots & \hat{\psi}_{2W_j} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\omega}_{j1} & \hat{\omega}_{j2} & \cdots & \hat{\omega}_{ju} & \cdots & \hat{\omega}_{jU_j} & \hat{\hat{\varsigma}}_{j1} & \hat{\hat{\varsigma}}_{j2} & \cdots & \hat{\hat{\varsigma}}_{jv} & \cdots & \hat{\hat{\varsigma}}_{jU_j} & \hat{\psi}_{j1} & \hat{\psi}_{j2} & \cdots & \hat{\psi}_{jw} & \cdots & \hat{\psi}_{jW_j} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\omega}_{J1} & \hat{\omega}_{J2} & \cdots & \hat{\omega}_{Ju} & \cdots & \hat{\omega}_{JU_j} & \hat{\hat{\varsigma}}_{J1} & \hat{\hat{\varsigma}}_{J2} & \cdots & \hat{\hat{\varsigma}}_{Jv} & \cdots & \hat{\hat{\varsigma}}_{JU_j} & \hat{\psi}_{J1} & \hat{\psi}_{J2} & \cdots & \hat{\psi}_{Jw} & \cdots & \hat{\psi}_{JW_j}
\end{bmatrix} \in \mathbb{R}^{(Q^*+X^*+I^*+J^*) \times (U^*+V^*+W^*)} \tag{30}$$

where $\hat{\kappa}_{qr}$ denotes the interaction ability between the q-th protein and the r-th protein; $\hat{\alpha}_{xu}$, $\hat{\beta}_{xv}$ and $\hat{\gamma}_{xw}$ are respectively the regulation abilities of the u-th TF on the x-th gene, the v-th lncRNA on the x-th gene and the w-th miRNA on the x-th gene; $\hat{\sigma}_{iu}$, $\hat{\varsigma}_{iv}$ and $\hat{\tau}_{iw}$ individually express the regulation abilities of the u-th TF on the i-th lncRNA, the v-th lncRNA on the i-th lncRNA and the w-th miRNA on the i-th lncRNA; $\hat{\omega}_{ju}$, $\hat{\xi}_{jv}$ and $\hat{\psi}_{jw}$ respectively mean the regulation abilities of the u-th TF on the j-th miRNA, the v-th lncRNA on the j-th miRNA and the w-th miRNA on the w-th miRNA. Moreover, if there is neither interaction nor regulation between the source and the target, some zeros should be padded in the network matrix in (34). Then, the core GWGENs were obtained by using PNP on the network matrix H with a significant energy threshold of 85%. First, the network matrix H is decomposed by singular value decomposition (SVD), as follows [11,83]:

$$H = SVD^T \tag{31}$$

where $S \in \mathbb{R}^{(Q^*+X^*+I^*+J^*) \times (Q^*+X^*+I^*+J^*)}$ and $D^T \in \mathbb{R}^{(U^*+V^*+W^*) \times (U^*+V^*+W^*)}$ are unitary singular matrices; $V = diag(v_1, \cdots, v_i, \cdots v_{U^*+V^*+W^*}) \in \mathbb{R}^{(Q^*+X^*+I^*+J^*) \times (U^*+V^*+W^*)}$ means the diagonal matrix of which the components at the diagonal are the singular values of H and are arranged in descending order, i.e., $v_1 \geq v_2 \geq \cdots \geq v_i \geq \cdots \geq v_{U^*+V^*+W^*} \geq 0$.

$$V = \begin{bmatrix} v_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & v_{U^*+V^*+W^*} \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{32}$$

Moreover, the normalization of singular values in (36) is defined as follows.

$$E_i = \frac{v_i^2}{\sum\limits_{i=1}^{U^*+V^*+W^*} v_i^2} \text{ and } \sum\limits_{i=1}^{U^*+V^*+W^*} E_i = 1 \tag{33}$$

$$\sum\limits_{i=1}^{I} E_i \geq 0.85 \tag{34}$$

From the above equation, we chose the top I significant singular vector structures to indicate the significant energy of real GWGENs, which is equal or more than the threshold of 0.85. Then, we separately projected every node of real GWGENs (i.e., each row of the network matrix H) to the top I significant singular vectors, as follows.

$$Z(a,b) = h_{a,:} \cdot d_{b,:}^T, \text{ for } a = 1, \ldots, Q^* + X^* + I^* + J^*, \ b = 1, \ldots, I \tag{35}$$

where $Z(a,b)$ means the projection value of the a-th node on the b-th significant singular vector; $h_{a,:}$ is the a-th row vector of the network matrix H and $d_{:,b}^T$ expresses the b-th column of D^T, i.e., the transpose of the b-th singular vector. Then, we define the two-norm projection value of each node such as protein, gene, lnRNA and miRNA in real GWGENs on the top I significant singular vectors as follows:

$$S(a) = \sqrt{\sum\limits_{i=1}^{I} Z^2(a,b)}, \text{ for } a = 1, \ldots, Q^* + X^* + I^* + J^* \tag{36}$$

According to the two-norm projection values in (40), the top 6000 significant proteins, genes, miRNAs and lncRNAs with high projection values were chosen to comprise the core GWGENs for OSCC and non-OSCC in Figure 3. Then, the core GWGENs were submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID) website to conduct the KEGG pathway enrichment analysis in Tables 3 and 4, and the core signaling pathways of non-OSCC and OSCC in Figure 4 were obtained by exploring the annotation of KEGG pathways. The enrichment analysis was used to check which significant pathways of our results were important to OSCC. Finally, the potential biomarkers were identified in Table 5 for the OSCC carcinogenic mechanism investigated by comparing the core signaling pathways and their downstream abnormal cellular functions of non-OSCC and OSCC in Figure 4.

4.6. Systematic Drug Discovery Based on the Drug/Target Interaction Prediction by the DNN-Based DTI Model and Drug Design Specifications for OSCC

Based on drug/target (biomarker) prediction and drug design specifications, we want to discover a potential multiple-molecule drug for the carcinogenic biomarkers on OSCC. First, we trained a DTI model based on a deep neural network (DNN) to predict the drug–target interactions between the drugs and targets (biomarkers). However, it is not enough to only consider the interactions between the molecular drugs and the targets in drug design. Some drug design specifications, i.e., adequate regulation ability, low toxicity and high sensitivity, are necessary to filter the candidate drugs predicted by the DNN-based DTI model. This systems drug discovery method is introduced in the following paragraphs for designing a multiple-molecule drug for OSCC treatment before clinical trials.

The flowchart of the systematic drug discovery method is shown in Figure 5. First, drug–target interaction databases by UniProt [84], DrugBank [59], ChEMBL [85] and Pubchem [86] are combined to train DNN as the DTI model for drug–target interaction prediction. In a few years, the feature-based method, i.e., molecular descriptor, will be widely used to describe the structural and chemical properties of molecules such as characteristics from the 2D, 3D spectrum of the structure, molecular weight, hydrophilic and hydrophobicity, etc. The chemical properties of the drug and genomic sequence of the target could be described with the molecular descriptor for the purpose of convenient analysis in drug design since the molecular descriptor can transform complicated chemical properties into a feature vector. We used the molecule descriptor function and protein descriptor function of python package pyBioMed to transform the drug and target into descriptors as drug and target features, respectively, under the python2.7 environment. Drug features of molecule descriptors include constitutional descriptors, connectivity indices, E-state indices, charge descriptors, molecular properties and kappa shape indices. For the target features, the protein descriptor calculates the structural and physicochemical features of proteins and peptides from the amino acid sequence such as amino acid composition, dipeptide composition, etc. For more detailed information about the descriptor transformation, readers can access the documents of pyBioMed. The drug descriptor and the target descriptor are combined into the feature vector corresponding to the drug–target pair, as follows:

$$v_{\text{drug–target}} = [D, T] = [d_1, d_2, \cdots, d_M, t_1, t_2, \cdots, t_N] \tag{37}$$

where the former symbol D in $v_{\text{drug–target}}$ is defined as the drug descriptor and the latter symbol T is the target descriptor. d_1 means the first drug feature; t_1 represents the first target feature; M indicates the total number of features of the drug and N expresses the total number of features of the target. We have 363 features for the drug and 996 features for the target.

Before training our DNN-based DTI model, we encountered several problems: features of different scales will affect our results—for example, large-scale features will have a great dominance in the training process. There are far more unverified data in the data than verified data, so there exists a between-class imbalance issue which could lead to a biased parameter updating tendency to the larger class during the training process.

To remedy the imbalance issue, we have to solve the problem of scale first. We use the min-max normalization to deal with this problem. Min-max normalization can handle this problem effectively, but compared to normalization, it is very sensitive to outliers. Additionally, feature normalization is performed before principal component analysis (PCA) to improve the DNN-based DTI model training performance. Therefore, the normalization to the features is given as follows:

$$d_i^* = \frac{d_i - \mu_i}{\sigma_i}, \forall i = 1, \dots, M \tag{38}$$

$$t_j^* = \frac{t_j - \mu_j}{\sigma_j}, \forall j = 1, \dots, N \tag{39}$$

where d_i denotes the i-th drug feature and d_i^* expresses the i-th drug feature after the standardization; μ_i and σ_i respectively denote the mean and standard deviations of the i-th drug feature; t_j means the j-th feature of the target and t_j^* represents the j-th feature of the target after standardization; μ_j and σ_j separately indicate the mean and standard deviation of the j-th target feature; M expresses the total number of drug features and N denotes the total number of target features.

In our data, the number of proven drug–target interaction samples, called the positive class, is 1455, and the number of unverified drug–target interaction samples, called the negative class, is much larger than the positive class. There exists a huge difference in the amount between the negative class and the positive class, which leads to the between-class imbalance issue. Before training the DNN-based DTI model, we also implemented data preprocessing by the principal component analysis (PCA) method—(35)–(40)—to reduce the dimension of the features of the drug and target to the dimension of the input of DNN. The PCA extraction was conducted after down-sampling and standardization to ensure that the PCA could accurately project the original data on the feature plane. It is worth noting that data preprocessing, e.g., PCA, is only conducted in training data because testing data should be deemed unknown to the model. The drug features were selected by the top 85% with higher singular values. The insignificant features of the drug and the target lower than the dimension of the input of DNN are deleted, and the remainders were employed to train DNN as the DTI model to predict candidate drugs for biomarkers (drug targets) of OSCC.

We referenced the basic concept and knowledge of DNN to train a DNN-based DTI model to predict drug–target interaction through the python Tensorflow and Keras package under the python3.7 environment. In the model structure of the DNN-based DTI model in Figure 4, the function in a feedforward step can be denoted as follows:

$$\mathbf{h} = \sigma(\mathbf{w}\mathbf{x} + \mathbf{b}) \tag{40}$$

where $\mathbf{x}$ and $\mathbf{h}$ respectively denote the input and output; $\mathbf{w}$ is the weighting matrix and $\mathbf{b}$ is the bias vector; $\sigma(.)$ indicates the activation function with ReLU in the hidden layer and the sigmoid function at the output layer. Since the binary classification issue is concerned, the binary-cross entropy is chosen as the cost function to calculate the model loss [11]:

$$C_n(\mathbf{w}, \mathbf{b}) = -[\hat{p}_n \log p_n + (1 - \hat{p}_n) log(1 - p_n)] \tag{41}$$

$$L(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{n=1}^{N} C_n(\mathbf{w}, \mathbf{b}) \tag{42}$$

where p_n means the truth label of positive interaction; $\hat{p}_n$ indicates the predictive probability of positive interaction, and $1 - p_n$ shows the truth label of negative interaction; $1 - \hat{p}_n$ represents the predicted probability of negative interaction; $L(\mathbf{w}, \mathbf{b})$ denotes the average of total loss $C_n(\mathbf{w}, \mathbf{b})$. According to the cost function, the backward propagation algorithm is applied to update the model parameter set θ containing the weighting matrix w and the

bias vector b through calculating the gradient of the cost function in (46) and eventually obtain the optimal solution θ^* in (48), as follows.

$$\theta = \begin{bmatrix} \mathbf{w} \\ \mathbf{b} \end{bmatrix} \tag{43}$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} L(\theta) \tag{44}$$

$$\theta^l = \theta^{l-1} - \eta \nabla L(\theta^{l-1}) \tag{45}$$

where l is the l-th epoch of the learning procedure; η is learning rate and $\nabla L\left(\theta^{l-1}\right)$ is the gradient of $L\left(\theta^{l-1}\right)$, as follows:

$$\nabla L(\theta^{l-1}) = \begin{bmatrix} \frac{\partial L(\theta^{l-1})}{\partial \mathbf{w}} \\ \frac{\partial L(\theta^{l-1})}{\partial \mathbf{b}} \end{bmatrix} \tag{46}$$

Based on the backward propagation method, the DNN-based DTI model could adjust the parameters to fit the drug–target interaction data at each iteration well.

In addition, the hyperparameters were tuned to not only lower the training time but also achieve the best model performance. We used an optimizer with default settings and set the learning rate to 0.003 to make model parameter θ converge faster and precisely. We set 100 for epochs and 100 for batch size. For the data, we split one-fourth of the data as testing data and three-fourths of it as training data. Moreover, we divided the training data into five equal folds to perform the five-fold cross-validation strategy. In the five-fold data, four-fifths of them were for the model training and one-fifth of them was used as the validation data, which play the role of supervisor to check whether the model was better than that of the former epoch. Additionally, five-fold cross-validation could verify the stability of the data and model. To avoid model overfitting, we applied an early stopping strategy to check if the test accuracy decreased when the training accuracy increased continuously. Moreover, we embedded the dropout layer after each hidden layer to further prevent model overfitting and set 0.4 for the dropout rate. After training the DNN-based DTI model, we adopted a performance measurement AUC (area under the curve) score and ROC (receiver operating characteristics) curve in Figure 8 to check the model performance. It is one of the most useful evaluation metrics to visualize the model performance when it comes to the classification problems. The higher the AUC score (which means the area under the line is larger), the better the accuracy is for the DNN-based DTI model in predicting the true positive and true negative drug–target interaction. The formulas for the AUC score and ROC curve are shown below [83].

$$\text{TPR(True Positive Rate)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{47}$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{48}$$

$$\text{FPR(False Positive Rate)} = 1 - \text{specificity} = \frac{\text{FP}}{\text{TN} + \text{FP}} \tag{49}$$

where TP (True Positive) means that the real value is true and is judged correctly; TN shows that the real value is true and is judged by mistake; FP indicates that the real value is false and is judged accurately; FN represents that the real value is false and is judged in error.

5. Conclusions

In this study, based on our proposed combination of systems biology and systems drug discovery design methods, we investigated the complex carcinogenic molecular mechanisms of OSCC from genome-wide data, genetic and epigenetic network perspectives,

and designed prospective multiple-molecular drugs as a multi-molecule drug to target multiple biomarkers of OSCC. We construct a genetic and epigenetic biological network by exploring methods of systematic identification and systematic sequential detection of big data. Afterwards, we extracted core signaling pathways by the PNP method and KEGG pathway annotation to select important biomarkers from OSCC carcinogenesis by comparing core signaling pathways and their downstream abnormal cellular functions. To discover drug candidates that interact with these biomarkers, we trained a DNN-based DTI model by DTI databases to predict drug–target interaction probability values. In addition, we took the drug regulation ability, low toxicity and high sensitivity as the drug design criteria to screen out suitable potential drugs from the predicted drug candidates. Therefore, a set of combined multi-molecular drugs is proposed as a multi-molecule drug for OSCC treatment. In the future, more and more different types of genomics data will be available for epigenetic and epigenetic regulations. A combination of multiple types of genomics data is needed to help us enhance our work and gain a deeper understanding of the significant biomarkers of the carcinogenic mechanism of OSCC. It is anticipated that the proposed systems biology and systems drug discovery design approach may provide new directions for OSCC treatment.

References

1. Da Silva, S.D.; Marchi, F.A.; Xu, B.; Bijian, K.; Alobaid, F.; Mlynarek, A.; Rogatto, S.R.; Hier, M.; Kowalski, L.P.; Alaoui-Jamali, M.A. Predominant Rab-GTPase amplicons contributing to oral squamous cell carcinoma progression to metastasis. *Oncotarget* **2015**, *6*, 21950. [CrossRef] [PubMed]
2. Petersen, P.E. Oral cancer prevention and control–the approach of the World Health Organization. *Oral Oncol.* **2009**, *45*, 454–460. [CrossRef] [PubMed]
3. Andisheh-Tadbir, A.; Mehrabani, D.; Heydari, S.T. Epidemiology of squamous cell carcinoma of the oral cavity in Iran. *J. Craniofacial Surg.* **2008**, *19*, 1699–1702. [CrossRef] [PubMed]
4. Dissanayaka, W.L.; Pitiyage, G.; Kumarasiri, P.V.R.; Liyanage, R.L.P.R.; Dias, K.D.; Tilakaratne, W.M. Clinical and histopathologic parameters in survival of oral squamous cell carcinoma. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2012**, *113*, 518–525. [CrossRef] [PubMed]
5. Chen, Y.J.; Chang, J.T.C.; Liao, C.T.; Wang, H.M.; Yen, T.C.; Chiu, C.C.; Lu, Y.C.; Li, H.F.; Cheng, A.J. Head and neck cancer in the betel quid chewing area: Recent advances in molecular carcinogenesis. *Cancer Sci.* **2008**, *99*, 1507–1514. [CrossRef] [PubMed]
6. Otero-Rey, E.M.; Suarez-Alen, F.; Peñamaria-Mallon, M.; Lopez-Lopez, J.; Blanco-Carrion, A. Malignant transformation of oral lichen planus by a chronic inflammatory process. Use of topical corticosteroids to prevent this progression? *Acta Odontol. Scand.* **2014**, *72*, 570–577. [CrossRef]
7. Tang, D.; Tao, D.; Fang, Y.; Deng, C.; Xu, Q.; Zhou, J. TNF-alpha promotes invasion and metastasis via NF-kappa B pathway in oral squamous cell carcinoma. *Med. Sci. Monit. Basic Res.* **2017**, *23*, 141. [CrossRef]
8. Scully, C.; Field, J.; Tanzawa, H. Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. *Oral Oncol.* **2000**, *36*, 256–263. [CrossRef]

9.   Feller, L.; Altini, M.; Lemmer, J. Inflammation in the context of oral cancer. *Oral Oncol.* **2013**, *49*, 887–892. [CrossRef]

10.   Harada, K.; Ferdous, T.; Itashiki, Y.; Takii, M.; Mano, T.; Mori, Y.; Ueyama, Y. Cepharanthine inhibits angiogenesis and tumorigenicity of human oral squamous cell carcinoma cells by suppressing expression of vascular endothelial growth factor and interleukin-8. *Int. J. Oncol.* **2009**, *35*, 1025–1035. [CrossRef]

11.   Chang, S.; Chen, J.-Y.; Chuang, Y.-J.; Chen, B.-S. Systems Approach to Pathogenic Mechanism of Type 2 Diabetes and Drug Discovery Design Based on Deep Learning and Drug Design Specifications. *Int. J. Mol. Sci.* **2020**, *22*, 166. [CrossRef] [PubMed]

12.   Ting, C.-T.; Chen, B.-S. Repurposing Multiple-Molecule Drugs for COVID-19-Associated Acute Respiratory Distress Syndrome and Non-Viral Acute Respiratory Distress Syndrome via a Systems Biology Approach and a DNN-DTI Model Based on Five Drug Design Specifications. *Int. J. Mol. Sci.* **2022**, *23*, 3649. [CrossRef] [PubMed]

13.   Yeh, S.-J.; Chang, C.-A.; Li, C.-W.; Wang, L.H.-C.; Chen, B.-S. Comparing progression molecular mechanisms between lung adenocarcinoma and lung squamous cell carcinoma based on genetic and epigenetic networks: Big data mining and genome-wide systems identification. *Oncotarget* **2019**, *10*, 3760. [CrossRef]

14.   Li, C.-W.; Jheng, B.-R.; Chen, B.-S. Investigating genetic-and-epigenetic networks, and the cellular mechanisms occurring in Epstein–Barr virus-infected human B lymphocytes via big data mining and genome-wide two-sided NGS data identification. *PLoS ONE* **2018**, *13*, e0202537.

15.   Hughes, J.P.; Rees, S.; Kalindjian, S.B.; Philpott, K.L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249. [CrossRef]

16.   Akhondzadeh, S. The importance of clinical trials in drug development. *Avicenna J. Med. Biotechnol.* **2016**, *8*, 151.

17.   Weaver, M.F.; Hopper, J.A.; Gunderson, E.W. Designer drugs 2015: Assessment and management. *Addict. Sci. Clin. Pract.* **2015**, *10*, 8. [CrossRef] [PubMed]

18.   Li, A.; Huang, H.-T.; Huang, H.-C.; Juan, H.-F. LncTx: A network-based method to repurpose drugs acting on the survival-related lncRNAs in lung cancer. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3990–4002. [CrossRef]

19.   Cheng, L.-H.; Hsu, T.-C.; Lin, C. Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction. *Sci. Rep.* **2021**, *11*, 14914.

20.   Lai, Y.-H.; Chen, W.-N.; Hsu, T.-C.; Lin, C.; Tsao, Y.; Wu, S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Sci. Rep.* **2020**, *10*, 4679.

21.   Lee, K.-H.; Chang, Y.-C.; Chen, T.-F.; Juan, H.-F.; Tsai, H.-K.; Chen, C.-Y. Connecting MHC-I-binding motifs with HLA alleles via deep learning. *Commun. Biol.* **2021**, *4*, 1194. [CrossRef] [PubMed]

22.   Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The database of interacting proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32* (Suppl. 1), D449–D451. [CrossRef] [PubMed]

23.   Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; Del-Toro, N. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [CrossRef] [PubMed]

24.   Stark, C.; Breitkreutz, B.-J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34* (Suppl. 1), D535–D539. [CrossRef]

25.   Zanzoni, A.; Montecchi-Palazzi, L.; Quondam, M.; Ausiello, G.; Helmer-Citterich, M.; Cesareni, G. MINT: A Molecular INTeraction database. *FEBS Lett.* **2002**, *513*, 135–140. [CrossRef]

26.   Bovolenta, L.A.; Acencio, M.L.; Lemke, N. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genom.* **2012**, *13*, 405. [CrossRef]

27.   Zheng, G.; Tu, K.; Yang, Q.; Xiong, Y.; Wei, C.; Xie, L.; Zhu, Y.; Li, Y. ITFP: An integrated platform of mammalian transcription factors. *Bioinformatics* **2008**, *24*, 2416–2417. [CrossRef]

28.   Wingender, E.; Chen, X.; Hehl, R.; Karas, H.; Liebich, I.; Matys, V.; Meinhardt, T.; Prüß, M.; Reuter, I.; Schacherer, F. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res.* **2000**, *28*, 316–319. [CrossRef]

29.   Friard, O.; Re, A.; Taverna, D.; De Bortoli, M.; Corá, D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinform.* **2010**, *11*, 435. [CrossRef]

30.   Agarwal, V.; Bell, G.W.; Nam, J.-W.; Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *elife* **2015**, *4*, e05005. [CrossRef]

31.   Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [CrossRef] [PubMed]

32.   Chen, B.-S.; Wu, C.-C. Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering. *Cells* **2013**, *2*, 635–688. [CrossRef]

33.   Li, C.-W.; Chen, B.-S. Investigating core genetic-and-epigenetic cell cycle networks for stemness and carcinogenic mechanisms, and cancer drug design using big database mining and genome-wide next-generation sequencing data. *Cell Cycle* **2016**, *15*, 2593–2607. [CrossRef] [PubMed]

34.   Wang, G.; Pan, C.; Cao, K.; Zhang, J.; Geng, H.; Wu, K.; Wen, J.; Liu, C. Impacts of cigarette smoking on the tumor immune microenvironment in esophageal squamous cell carcinoma. *J. Cancer* **2022**, *13*, 413. [CrossRef]

35. Foy, J.-P.; Bertolus, C.; Michallet, M.-C.; Deneuve, S.; Incitti, R.; Bendriss-Vermare, N.; Albaret, M.-A.; Ortiz-Cuaran, S.; Thomas, E.; Colombe, A. The immune microenvironment of HPV-negative oral squamous cell carcinoma from never-smokers and never-drinkers patients suggests higher clinical benefit of IDO1 and PD1/PD-L1 blockade. *Ann. Oncol.* **2017**, *28*, 1934–1941. [CrossRef]

36. Cheskis, B.J.; Greger, J.; Cooch, N.; McNally, C.; Mclarney, S.; Lam, H.-S.; Rutledge, S.; Mekonnen, B.; Hauze, D.; Nagpal, S. MNAR plays an important role in ERa activation of Src/MAPK and PI3K/Akt signaling pathways. *Steroids* **2008**, *73*, 901–905. [CrossRef]

37. Burotto, M.; Chiou, V.L.; Lee, J.M.; Kohn, E.C. The MAPK pathway across different malignancies: A new perspective. *Cancer* **2014**, *120*, 3446–3456. [CrossRef]

38. Mittal, M.; Kapoor, V.; Mohanti, B.K.; Das, S.N. Functional variants of COX-2 and risk of tobacco-related oral squamous cell carcinoma in high-risk Asian Indians. *Oral Oncol.* **2010**, *46*, 622–626. [CrossRef]

39. Ramos-Garcia, P.; Gil-Montoya, J.; Scully, C.; Ayén, A.; González-Ruiz, L.; Navarro-Triviño, F.; González-Moles, M. An update on the implications of cyclin D1 in oral carcinogenesis. *Oral Dis.* **2017**, *23*, 897–912. [CrossRef] [PubMed]

40. Lakshminarayana, S.; Augustine, D.; Rao, R.S.; Patil, S.; Awan, K.H.; Venkatesiah, S.S.; Haragannavar, V.C.; Nambiar, S.; Prasad, K. Molecular pathways of oral cancer that predict prognosis and survival: A systematic review. *J. Carcinog.* **2018**, *17*, 7. [CrossRef]

41. Freier, K.; Sticht, C.; Hofele, C.; Flechtenmacher, C.; Stange, D.; Puccio, L.; Toedt, G.; Radlwimmer, B.; Lichter, P.; Joos, S. Recurrent coamplification of cytoskeleton-associated genes EMS1 and SHANK2 with CCND1 in oral squamous cell carcinoma. *Genes Chromosomes Cancer* **2006**, *45*, 118–125. [CrossRef] [PubMed]

42. Wang, Q.; Han, J.; Xu, P.; Jian, X.; Huang, X.; Liu, D. Silencing of LncRNA SNHG16 downregulates cyclin D1 (CCND1) to abrogate malignant phenotypes in oral squamous cell carcinoma (OSCC) through upregulating miR-17–5p. *Cancer Manag. Res.* **2021**, *13*, 1831. [CrossRef] [PubMed]

43. Giles, R.H.; Van Es, J.H.; Clevers, H. Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim. Biophys. Acta (BBA)-Rev. Cancer* **2003**, *1653*, 1–24. [CrossRef]

44. Huelsken, J.; Behrens, J. The Wnt signalling pathway. *J. Cell Sci.* **2002**, *115*, 3977–3978. [CrossRef] [PubMed]

45. Lyou, Y.; Habowski, A.N.; Chen, G.T.; Waterman, M.L. Inhibition of nuclear Wnt signalling: Challenges of an elusive target for cancer therapy. *Br. J. Pharmacol.* **2017**, *174*, 4589–4599. [CrossRef] [PubMed]

46. Xiao, X.; Gu, Y.; Wang, G.; Chen, S. c-Myc, RMRP, and miR-34a-5p form a positive-feedback loop to regulate cell proliferation and apoptosis in multiple myeloma. *Int. J. Biol. Macromol.* **2019**, *122*, 526–537. [CrossRef]

47. Walker, L.; Lynch, M.; Silverman, S.; Fraser, J.; Boulter, J.; Weinmaster, G.; Gasson, J.C. The Notch/Jagged pathway inhibits proliferation of human hematopoietic progenitors in vitro. *Stem Cells* **1999**, *17*, 162–171. [CrossRef]

48. Chai, A.W.Y.; Lim, K.P.; Cheong, S.C. Translational genomics and recent advances in oral squamous cell carcinoma. In *Seminars in Cancer Biology*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 71–83.

49. Hijioka, H.; Setoguchi, T.; Miyawaki, A.; Gao, H.; Ishida, T.; Komiya, S.; Nakamura, N. Upregulation of Notch pathway molecules in oral squamous cell carcinoma. *Int. J. Oncol.* **2010**, *36*, 817–822.

50. Sjölund, J.; Manetopoulos, C.; Stockhausen, M.-T.; Axelson, H. The Notch pathway in cancer: Differentiation gone awry. *Eur. J. Cancer* **2005**, *41*, 2620–2629. [CrossRef]

51. Ingram, W.; McCue, K.; Tran, T.; Hallahan, A.; Wainwright, B. Sonic Hedgehog regulates Hes1 through a novel mechanism that is independent of canonical Notch pathway signalling. *Oncogene* **2008**, *27*, 1489–1500. [CrossRef]

52. Subramaniam, D.; Ponnurangam, S.; Ramamoorthy, P.; Standing, D.; Battafarano, R.J.; Anant, S.; Sharma, P. Curcumin induces cell death in esophageal cancer cells through modulating Notch signaling. *PLoS ONE* **2012**, *7*, e30590. [CrossRef] [PubMed]

53. Bi, H.; Ming, L.; Cheng, R.; Luo, H.; Zhang, Y.; Jin, Y. Liver extracellular matrix promotes BM-MSCs hepatic differentiation and reversal of liver fibrosis through activation of integrin pathway. *J. Tissue Eng. Regen. Med.* **2017**, *11*, 2685–2698. [CrossRef] [PubMed]

54. Dayyani, F.; Parikh, N.U.; Varkaris, A.S.; Song, J.H.; Moorthy, S.; Chatterji, T.; Maity, S.N.; Wolfe, A.R.; Carboni, J.M.; Gottardis, M.M. Combined Inhibition of IGF-1R/IR and Src family kinases enhances antitumor effects in prostate cancer by decreasing activated survival pathways. *PLoS ONE* **2012**, *7*, e51189. [CrossRef] [PubMed]

55. Yu, Z.; Weinberger, P.M.; Sasaki, C.; Egleston, B.L.; Speier IV, W.F.; Haffty, B.; Kowalski, D.; Camp, R.; Rimm, D.; Vairaktaris, E. Phosphorylation of Akt (Ser473) predicts poor clinical outcome in oropharyngeal squamous cell cancer. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 553–558. [CrossRef]

56. Subarnbhesaj, A.; Miyauchi, M.; Chanbora, C.; Mikuriya, A.; Nguyen, P.T.; Furusho, H.; Ayuningtyas, N.F.; Fujita, M.; Toratani, S.; Takechi, M. Roles of VEGF-Flt-1 signaling in malignant behaviors of oral squamous cell carcinoma. *PLoS ONE* **2017**, *12*, e0187092. [CrossRef]

57. Seçilmiş, D.; Hillerton, T.; Morgan, D.; Tjärnberg, A.; Nelander, S.; Nordling, T.E.; Sonnhammer, E.L. Uncovering cancer gene regulation by accurate regulatory network inference from uninformative data. *NPJ Syst. Biol. Appl.* **2020**, *6*, 37. [CrossRef]

58. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **2017**, *171*, 1437–1452.e17. [CrossRef]

59. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V. DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2010**, *39* (Suppl. 1), D1035–D1041. [CrossRef]

60. Corsello, S.M.; Nagari, R.T.; Spangler, R.D.; Rossen, J.; Kocak, M.; Bryan, J.G.; Humeidi, R.; Peck, D.; Wu, X.; Tang, A.A. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **2020**, *1*, 235–248. [CrossRef]

61. Li, X.; Guo, S.; Xiong, X.-K.; Peng, B.-Y.; Huang, J.-M.; Chen, M.-F.; Wang, F.-Y.; Wang, J.-N. Combination of quercetin and cisplatin enhances apoptosis in OSCC cells by downregulating xIAP through the NF-κB pathway. *J. Cancer* **2019**, *10*, 4509. [CrossRef]

62. Li, L.; Liu, H.-C.; Wang, C.; Liu, X.; Hu, F.-C.; Xie, N.; Lü, L.; Chen, X.; Huang, H.-Z. Overexpression of β-catenin induces cisplatin resistance in oral squamous cell carcinoma. *BioMed Res. Int.* **2016**, *2016*, 5378567. [CrossRef] [PubMed]

63. Ding, L.; Ren, J.; Zhang, D.; Li, Y.; Huang, X.; Ji, J.; Hu, Q.; Wang, H.; Ni, Y.; Hou, Y. The TLR3 Agonist Inhibit Drug Efflux and Sequentially Consolidates Low-Dose Cisplatin-Based Chemoimmunotherapy while Reducing Side EffectsOrderly Combination poly (I: C)/Low-Dose DDP. *Mol. Cancer Ther.* **2017**, *16*, 1068–1079. [CrossRef] [PubMed]

64. Chou, T.-Y.; Chiu, C.-H.; Li, L.-H.; Hsiao, C.-Y.; Tzen, C.-Y.; Chang, K.-T.; Chen, Y.-M.; Perng, R.-P.; Tsai, S.-F.; Tsai, C.-M. Mutation in the tyrosine kinase domain of epidermal growth factor receptor is a predictive and prognostic factor for gefitinib treatment in patients with non–small cell lung cancer. *Clin. Cancer Res.* **2005**, *11*, 3750–3757. [CrossRef] [PubMed]

65. Taron, M.; Ichinose, Y.; Rosell, R.; Mok, T.; Massuti, B.; Zamora, L.; Mate, J.L.; Manegold, C.; Ono, M.; Queralt, C. Activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor are associated with improved survival in gefitinib-treated chemorefractory lung adenocarcinomas. *Clin. Cancer Res.* **2005**, *11*, 5878–5885. [CrossRef]

66. Moulder, S.L.; Yakes, F.M.; Muthuswamy, S.K.; Bianco, R.; Simpson, J.F.; Arteaga, C.L. Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erb B2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res.* **2001**, *61*, 8887–8895. [PubMed]

67. Khan, Z.; Bisen, P.S. Oncoapoptotic signaling and deregulated target genes in cancers: Special reference to oral cancer. *Biochim. Biophys. Acta Rev. Cancer* **2013**, *1836*, 123–145. [CrossRef]

68. Shortt, J.; Johnstone, R.W. Oncogenes in cell survival and cell death. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a009829. [CrossRef]

69. Peters, S.; Zimmermann, S.; Adjei, A.A. Oral epidermal growth factor receptor tyrosine kinase inhibitors for the treatment of non-small cell lung cancer: Comparative pharmacokinetics and drug–drug interactions. *Cancer Treat. Rev.* **2014**, *40*, 917–926. [CrossRef] [PubMed]

70. Sanchez-Rangel, E.; Inzucchi, S.E. Metformin: Clinical use in type 2 diabetes. *Diabetologia* **2017**, *60*, 1586–1593. [CrossRef]

71. Lv, Z.; Guo, Y. Metformin and its benefits for various diseases. *Front. Endocrinol.* **2020**, *11*, 191. [CrossRef]

72. Chae, Y.K.; Arya, A.; Malecek, M.-K.; Shin, D.S.; Carneiro, B.; Chandra, S.; Kaplan, J.; Kalyan, A.; Altman, J.K.; Platanias, L. Repurposing metformin for cancer treatment: Current clinical studies. *Oncotarget* **2016**, *7*, 40767. [CrossRef] [PubMed]

73. Kasznicki, J.; Sliwinska, A.; Drzewoski, J. Metformin in cancer prevention and therapy. *Ann. Transl. Med.* **2014**, *2*, 57. [PubMed]

74. Li, Y.; Wang, M.; Zhi, P.; You, J.; Gao, J.-Q. Metformin synergistically suppress tumor growth with doxorubicin and reverse drug resistance by inhibiting the expression and function of P-glycoprotein in MCF7/ADR cells and xenograft models. *Oncotarget* **2018**, *9*, 2158. [CrossRef] [PubMed]

75. Shah, R.R. Hyperglycaemia induced by novel anticancer agents: An undesirable complication or a potential therapeutic opportunity? *Drug Saf.* **2017**, *40*, 211–228. [CrossRef]

76. Peng, M.; Darko, K.O.; Tao, T.; Huang, Y.; Su, Q.; He, C.; Yin, T.; Liu, Z.; Yang, X. Combination of metformin with chemotherapeutic drugs via different molecular mechanisms. *Cancer Treat. Rev.* **2017**, *54*, 24–33. [CrossRef]

77. Haslam, E.; Cai, Y. Plant polyphenols (vegetable tannins): Gallic acid metabolism. *Nat. Prod. Rep.* **1994**, *11*, 41–66. [CrossRef]

78. Cai, Y.; Luo, Q.; Sun, M.; Corke, H. Antioxidant activity and phenolic compounds of 112 traditional Chinese medicinal plants associated with anticancer. *Life Sci.* **2004**, *74*, 2157–2184. [CrossRef]

79. Shabani, S.; Rabiei, Z.; Amini-Khoei, H. Exploring the multifaceted neuroprotective actions of gallic acid: A review. *Int. J. Food Prop.* **2020**, *23*, 736–752. [CrossRef]

80. Nikbakht, J.; Hemmati, A.A.; Arzi, A.; Mansouri, M.T.; Rezaie, A.; Ghafourian, M. Protective effect of gallic acid against bleomycin-induced pulmonary fibrosis in rats. *Pharmacol. Rep.* **2015**, *67*, 1061–1067. [CrossRef]

81. Chuang, C.-Y.; Liu, H.-C.; Wu, L.-C.; Chen, C.-Y.; Chang, J.T.; Hsu, S.-L. Gallic acid induces apoptosis of lung fibroblasts via a reactive oxygen species-dependent ataxia telangiectasia mutated-p53 activation pathway. *J. Agric. Food Chem.* **2010**, *58*, 2943–2951. [CrossRef]

82. Bai, J.; Zhang, Y.; Tang, C.; Hou, Y.; Ai, X.; Chen, X.; Zhang, Y.; Wang, X.; Meng, X. Gallic acid: Pharmacological activities and molecular mechanisms involved in inflammation-related diseases. *Biomed. Pharmacother.* **2021**, *133*, 110985. [CrossRef] [PubMed]

83. Yeh, S.-J.; Chen, S.-W.; Chen, B.-S. Investigation of the genome-wide genetic and epigenetic networks for drug discovery based on systems biology approaches in colorectal cancer. *Front. Genet.* **2020**, *11*, 117. [CrossRef]

84. Consortium, U. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

85. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **2007**, *35* (Suppl. 1), D198–D201. [CrossRef] [PubMed]

86. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef]