biology

MDPI

Article

Personalized Driver Gene Prediction Using Graph Convolutional Networks with Conditional Random Fields

Pi-Jing Wei [1], An-Dong Zhu [1], Ruifen Cao [2] and Chunhou Zheng [3,*]

[1] Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University, 111 Jiulong Road, Hefei 230601, China; weipj@ahu.edu.cn (P.-J.W.); q21301153@stu.ahu.edu.cn (A.-D.Z.)
[2] School of Computer Science and Technology, Anhui University, 111 Jiulong Road, Hefei 230601, China; rfcao@ahu.edu.cn
[3] School of Artificial Intelligence, Anhui University, 111 Jiulong Road, Hefei 230601, China
* Correspondence: zhengch99@126.com

Simple Summary: Identifying cancer driver genes plays a significant role in cancer diagnosis and treatment. With the advancement of next-generation sequencing technologies, a wealth of multi-omics cancer data, including genomic, epigenomic, and transcriptomic data, are now available for cancer research. Integrating these data to effectively identify cancer driver genes causally associated with cancer is a computational challenge. Methods for identifying cancer driver genes are mainly based on population levels. Considering the trend of precision medicine and the heterogeneity of patients, it is challenging but crucial to identify cancer driver genes at the individual level. We developed a method called PDGCN (Personalized Drivers of GCN), which constructs sample–gene interaction networks by integrating multiple types of data features and using network structural features extracted from Node2vec. Then, a graphical convolutional neural network model with a conditional random field layer is used to prioritize candidate driver genes in the network. The results show that PDGCN can identify driver genes at the individual level, providing a new perspective for predicting driver genes in individual samples.

Abstract: Cancer is a complex and evolutionary disease mainly driven by the accumulation of genetic variations in genes. Identifying cancer driver genes is important. However, most related studies have focused on the population level. Cancer is a disease with high heterogeneity. Thus, the discovery of driver genes at the individual level is becoming more valuable but is a great challenge. Although there have been some computational methods proposed to tackle this challenge, few can cover all patient samples well, and there is still room for performance improvement. In this study, to identify individual-level driver genes more efficiently, we propose the PDGCN method. PDGCN integrates multiple types of data features, including mutation, expression, methylation, copy number data, and system-level gene features, along with network structural features extracted using Node2vec in order to construct a sample–gene interaction network. Prediction is performed using a graphical convolutional neural network model with a conditional random field layer, which is able to better combine the network structural features with biological attribute features. Experiments on the ACC (Adrenocortical Cancer) and KICH (Kidney Chromophobe) datasets from TCGA (The Cancer Genome Atlas) demonstrated that the method performs better compared to other similar methods. It can identify not only frequently mutated driver genes, but also rare candidate driver genes and novel biomarker genes. The results of the survival and enrichment analyses of these detected genes demonstrate that the method can identify important driver genes at the individual level.

Keywords: cancer; driver genes; multi-omics features; graph convolutional neural network; conditional random field layer

Citation: Wei, P.-J.; Zhu, A.-D.; Cao, R.; Zheng, C. Personalized Driver Gene Prediction Using Graph Convolutional Networks with Conditional Random Fields. *Biology* **2024**, *13*, 184. https://doi.org/10.3390/biology13030184

Academic Editor: Angelo Fortunato

Received: 15 February 2024
Revised: 3 March 2024
Accepted: 10 March 2024
Published: 14 March 2024

1. Introduction

Cancer is an evolving and complicated illness causing high morbidity and mortality rates worldwide. GLOBOCAN 2020 projects that 28.4 million new cancer cases will be diagnosed worldwide in 2040, a 47% increase from 2020 [1]. The early identification and the subsequent diagnostic treatment of cancer lesions are two of the most successful approaches to minimizing cancer mortality, but they require a deeper knowledge of the molecular underpinnings of tumor development and progression [2]. Complex genetic changes, such as Single Nucleotide Variations (SNVs), copy number variations (CNVs), insertions and deletions, and structural abnormalities, are the main causes of the complicated genesis of cancer [3]. Driver mutations are often referred to as mutations that provide tumor cells with a selective growth advantage and hasten the development of cancer [4]. The term "passenger mutations" refers to mutations that happen randomly in tumor samples but are not necessarily connected to the development of cancer [4,5]. Cancer driver genes (CDGs) are those with cancer-causing mutations [5]. Finding cancer driver genes in various tumor types is one of the main goals of cancer genomics. For the clinical diagnosis, prevention, and therapy of cancer, the identification of driver genes is essential [6].

The advancement of computer technology and sequencing techniques in recent years has resulted in a rise in the number of researchers working on cancer driver gene identification. Most computational techniques focus on discovering driver genes for a cancer type at the population level. Typical methods include mutation-based and network-based methods. Mutation-based approaches identify cancer driver genes based on the gene mutation frequency, with more frequently mutated genes being more likely to be driver genes [7]. For example, the methods of Mutsig [8] and MuSic [9] estimate the background mutation rates of each gene and identify driver mutations that significantly deviate from this rate. OncodriveCLUST [10] constructs a background model using silent mutations to identify genes with a tendency to cluster significant mutations in protein sequences. However, it is difficult to accurately estimate the background mutation rate and identify infrequently or rarely mutated genes using mutation-based methods [11]. Considering that the biological network can describe the relationships between genes and gene features, some network-based approaches have been proposed to identify cancer driver genes by assessing their roles in biological networks. For instance, DriverNet [12] constructs a bipartite graph based on mutated genes and differentially expressed genes (DEGs) in tumor samples and prioritizes mutated genes according to their degree. HotNet2 [13] incorporates gene interaction networks to identify significantly altered gene modules in a cohort. Furthermore, because the graph convolutional network (GCN) algorithm can directly deal with graph-structured data, showing outstanding performances, some cancer driver gene identification methods based on GCNs have been proposed. The EMOGI [14] method successfully identifies known driver genes by combining the features of different genes using a GCN model. The MTGCN [15] method constructs a multichannel GCN network that can combine the driver gene identification task with link prediction. It was also shown that combining biometric and network features could improve prediction accuracy. However, population-based approaches have limitations in that they cannot identify rare driver genes occurring in small cohorts or in individual patients.

Cancer is a complex disease with high heterogeneity, as different patients may be driven by different genes and have different outcomes even if they receive the same treatment. Therefore, it is necessary to investigate personalized cancer driver genes that are specific to individual patients. In recent years, many researchers have proposed some driver gene identification methods for individual patients. For instance, OncoIMPACT [16] and DawnRank [17] prioritize patient-specific driver genes by exploiting the perturbations of the transcriptional programs through molecular networks. However, these approaches apply the same aggregated gene network to all patients, which may reduce the personalized information. Additionally, some approaches, such as SSN [18], SCS [19], and paired-SSN [20], have been proposed based on individual networks. The SSN algorithm constructs individual perturbation networks based on the expression data of diseased individual

samples against a group of given control samples [18]. The importance of each edge was quantified using Pearson's correlation coefficient to identify personalized driver genes. Similarly, SCS uses a random walk with a restart algorithm to construct personalized networks based on mutation data, expression data, and protein interaction network data [19]. The network-controlled strategy assesses the effect of mutations on expression patterns to identify personalized driver genes [19]. Paired-SSN uses paired sample expression data, i.e., normal and diseased data from the same sample, to construct individual networks [20]. It then identifies individual driver genes using a cybernetic approach, which provides a more personalized assessment of cancer driver genes [20]. Pham et al. proposed the pDriver [21] method, which constructs gene regulatory networks for each sample and uses network control strategies to identify personalized driver genes, including coding and non-coding genes. This method takes into account the use of known driver genes to localize to the patient's personalized driver genes, which is not considered by most other methods [21]. PRODIGY [22] identifies driver genes by optimizing the cost of subtrees in the gene interaction network as a score for mutated genes. However, PRODIGY [22] did not use known driver genes to localize to the patient's personalized driver genes. Overall, challenges remain in identifying the driver genes for each patient. In terms of methodology, some network-based methods will miss predictions for some patients. Many GCN-based methods are used to handle various bioinformatic tasks, but they still lack individual driver gene prediction. In terms of feature fusion, how to make better use of multi-omics features is also an issue that needs to be addressed.

In this work, we present a novel approach for predicting personalized cancer driver genes for individual patients, called PDGCN. It uses a GCN model with a conditional random field (CRF) layer to more adaptively fuse multi-omics features over a network of individual attributes. Unlike some previous methods, PDGCN constructs the networks for each patient, and then a GCN model with a CRF layer is used to learn the feature representation of the nodes by combining the structural features of the network with the biological property features of the genes. PDGCN obtains data from each sample for training, thus avoiding the possibility that individual samples may be missing to predict outcomes. Finally, the driver genes are obtained from the results predicted by the model. To evaluate the performance of the proposed model, we applied it to two TCGA datasets and compared it with other similar methods. The experimental results demonstrate that our model outperformed the other methods in detecting personalized cancer drivers in individuals.

2. Materials and Methods

2.1. Datasets

For this study, we downloaded two cancer datasets, those of adrenocortical carcinoma (ACC) and kidney chromophobe (KICH), through The Cancer Genome Atlas (TCGA) [22] Xena platform data portal [23] (http://xena.ucsc.edu/, accessed on 1 July 2022). Both datasets contained somatic mutation, expression, methylation, and copy number variation data. We selected only patients that had all four types of data, resulting in 77 samples for ACC and 66 samples for KICH. The STRING dataset (v11.0) [24] was used to construct individual networks for each patient. To obtain a positive set, we selected known cancer driver genes from the Network of Cancer Genes (NCG 6.0) database [25], which is a manually managed repository that collects well-studied cancer genes from various sources, and the Cancer Gene Census (CGC) from the COSMIC database (v90) [26], which is a popular cancer gene dataset containing 719 well-established driver genes. In addition, we collected the list of cancer type-specific genes published by Bailey et al. [27]. The above three components above made up the positive set. The negative set was formed by recursively filtering out the positive set from the processed data. We conducted a series of experiments on the two different datasets (i.e., ACC and KICH). For ACC, we generated 5467 positive and 15,448 negative datasets; for KICH, we generated 5982 positive and 13,815 negative datasets.

2.2. PDGCN

In this work, we propose a new framework called PDGCN that is based on a GCN and can predict driver genes at the individual level. PDGCN consists of two main steps, as illustrated in Figure 1. The first step (Figure 1a) is the construction of a patient–gene interaction network based on the known PPI network for each individual sample. Here, the genes we used were DEG specific to the cancer and mutated genes for each individual. The second step involves using the GCN to learn the representation of each node in the network. To force the aggregated representation of the neighbor nodes, a CRF layer was added to the GCN model. Nodes are scored based on the representation learned by the GCN, and the driver genes are identified. Next, we describe the above two steps in detail.

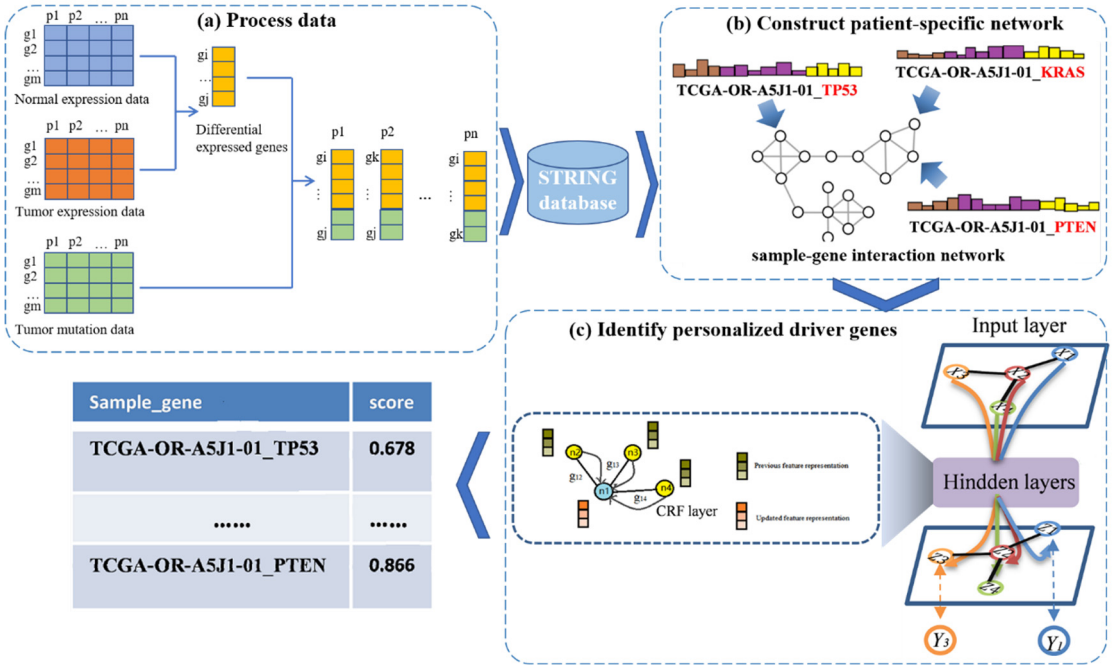


Figure 1. A flowchart of the PDGCN method. The method consists of three main steps. (**a**) Process data: A subset of genes for each sample are obtained using genes DEG specific to the cancer and mutated genes for each individual. (**b**) Construct a patient-specific network: A sample–gene interaction network for individual samples is constructed on the basis of the STRING database. (**c**) Identify personalized driver genes: Based on the representation of each node in the GCN, a CRF layer is inserted to force the aggregated representation of similar neighbors, and each node is scored based on the representation form learned by the GCN to identify driver genes.

2.2.1. Construction of Personalized Networks

In this study, we focused on coding genes; hence, the raw data sets from the TCGA were preprocessed by filtering out abnormal data, which mainly contained non-coding genes. To construct the individual-driven network, we used the processed data as follows. Firstly, the mutated genes of each patient were extracted from the mutation data. Next, the DEGs for each cancer type were selected by intersecting the results of the Anovar and Limma tools with an adjusted $p < 0.05$ and Log2 (FC) > 2 or Log2 (FC) < −2, respectively. This strategy can decrease the bias of a method. Then, the mutated genes and those DEG-specific to the individual were combined to create a subset of genes for each patient. These genes were then combined with the STRING dataset to obtain a sample-specific network, which we refer to as the sample–gene interaction network. Each node in this network was accompanied by a sample ID and a gene symbol, and a neighbor matrix A was constructed using this network.

In the sample–gene interaction network, the node's attribute features can be classified into three main types: molecular features, system-level features (gene properties), and

network structure features obtained through individual networks. Molecular features are extracted from somatic mutation, methylation, copy number, and expression data. For somatic mutation data, we used non-silent mutation data at the Multi-Center Mutation Calling in Multiple Cancers (MC3) gene level, where "1" indicates a mutation, and "0" indicates no mutation. Methylation data are represented by the beta values of the DNA methylation profiles measured experimentally, which are continuous variables between 0 and 1. The CNV data were processed using the GISTIC2 method, and the expression data were processed using log2-transformation. System-level features were obtained from sysSVM (from sysSVM method proposed by Nulsen et.al.) [28], which are the features of genes in global attributes. Here, the PPI network features were removed because the network we used was different from the sysSVM method. The features retained included 18-dimensional features, such as the length of the gene, the number of protein structural domains, the age of the gene, and the necessity of the gene. We matched these features to the genes of the individual samples to form complete system-level features. To obtain network structure features, the Node2vec [29], a model inherited from the random walk model in the DeepWalk (v1.0.2) [30] algorithm, was used to obtain features of different dimensions. To select the proper dimensions of the features, we used 10-, 20-, and 30-dimensional features for the subsequent experiments. Finally, all three types of features were stitched together to form the feature matrix X, which was then normalized to have values between 0 and 1.

2.2.2. Graph Convolutional Network for Node Embedding

A GCN is a multilayer graph convolutional neural network that aims to learn the node embedding by implementing convolutional operations on a graph. The basic concept of a GCN is to utilize the properties of neighboring nodes to improve the classification results. The topology of the graph is important, as the nodes can have a varying number of neighbors [31]. A GCN is a first-order local approximation of spectral graph convolutions that can handle first-order neighborhood information in each convolutional layer. Additionally, the multi-layer convolution permits the transfer of multi-order neighborhood information. From the adjacency matrix A and the feature matrix X, the simple propagation rules for each layer in a GCN can be defined as follows:

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \quad (1)$$

where $\widetilde{A} = A + I$, in which I is the unit matrix; $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$ represents the degree matrix; W denotes the learnable weight matrix; and σ denotes a nonlinear function, such as the ReLU activation function. H is the feature of each layer, and for the input layer, H is X. The first layer receives X as input, so $H^{(0)} = X$. The added self-connection matrix $\widetilde{A}$ helps to preserve the original node signals and incorporates them into the Laplace smoothing process.

2.2.3. CRF Layer for Embedding Update

The existing approach of the GCN considers all neighbors equally, so it cannot retain the similarity information of similar nodes when learning the node embedding. We enhanced the representational learning of the nodes by adding a CRF layer to encourage similar nodes to keep similar hidden features. The CRF is a probabilistic graphical model proposed by Lafferty et al. [32]. Combining the features of the maximum entropy model and the hidden Markov model, it is an undirected graph model that is commonly used to label or analyze sequence information, such as natural language text or biological sequences. The loss function of the CRF layer is as shown in Equation (2).

$$\ell_{CRF} = \sum_{i=1} \ell(H_i) \quad (2)$$

$$\ell(H_i) = \alpha\|H_i - Q_i\|_2^2 + \beta\sum_{j\in M_i} g_{ij}\|H_i - H_j\|_2^2 \quad (3)$$

where $\ell(H_i)$ is the loss of the H_i layer, and it can be calculated using Equation (3). Q_i denotes the initial embedding of node i obtained from the GCN convolution layer, and H_i denotes the updated embedding of node i in the CRF layer. In addition, g_{ij} denotes the importance of neighboring node pairs. M_i is the neighborhood of node i, and α and β are the balance factors of the two parts of Equation (2).

Motivated by Long et al. [33], we used self-attention [34] to distinguish the contributions of neighboring node pairs. The use of self-attention makes it easier to update the losses. Formally, the attention g_{ij} between nodes i and node j in Equation (3) is defined as follows.

$$\alpha_{ij} = att(W_t H_i, W_t H_j) \tag{4}$$

$$g_{ij} = softmax(\alpha_{ij}) = \frac{exp(a_{ij})}{\sum_{x \in N_i} exp(a_{ix})} \tag{5}$$

where $att()$ denotes a single-layer feedforward network to perform attention, and W_t denotes the potential trainable matrix.

2.2.4. Overall Loss and Optimization

Using the original binary cross-entropy loss function would result in a bias in data prediction toward the side with more samples, given the unbalanced nature of our positive and negative dataset. To address this issue, we attached weights of different multiples to the positive set. The modified loss function is as shown in Equation (6).

$$\ell_\theta = -(pylog(h) + (1-y)log(1-h)) \tag{6}$$

where p is the value of the different weights we added to the positive set, h is the output of the network after the sigmoid activation function, and y is the original node label (0 or 1). Finally, the overall loss ℓ_{Total} is defined as follows:

$$\ell_{Total} = \ell_{CRF} + \ell_\theta \tag{7}$$

The ADAM optimizer was used to train the GCN model.

3. Results

This section is divided into three parts. Firstly, we provide a brief description of our experimental setup. Then, we discuss the performance of the model and analyze it at both the population and individual levels. Lastly, we analyze the predicted rare mutant genes with novel biomarkers.

3.1. Experimental Setup

We implemented our model using Python 3.7 as the compiler. To achieve optimal model performance, we split all labeled genes, using 25% as the test set and 75% as the training set. Then, we used a grid search within a reasonable parameter range to optimize the hyperparameters, including learning rate, weight decay, dropout rate, and epoch, and selected the best performance hyperparameter combination as the final parameters by conducting five-fold cross-validation on the training set. For the Adam optimizer in the constructed individual gene network, we selected a learning rate of 0.001, weight decay of 0.005, and a dropout rate of 0.1 for 3000 epochs. We implemented the experimental code based on the open source machine learning framework TensorFlow. The experiments were conducted using the Windows 10 operating system, with an Intel® Core™ (Intel, Santa Clara, CA, USA) i5-8265U, 1.60 GHz CPU, GTX1060 graphics card, and 16 GB RAM.

3.2. Evaluation Metrics

To verify the effectiveness of the proposed method, the metrics of accuracy (ACC), area under the curve (AUC), and area under the precision–recall curve (AUPR) were considered.

ACC $= \frac{(TP+TN)}{(TP+FN+FP+TN)}$, and the AUC is defined as the area under the ROC curve. The closer the AUC is to 1, the higher the correct rate. Additionally, the AUPR is defined as the area under the P-R curve, and the P-R curve is a graph consisting of precision (P) and recall (R), where P $= \frac{TP}{(TP+FP)}$ and R $= \frac{TP}{(TP+FN)}$.

3.3. Effect of Node2vec Dimensions

In this study, the network features extracted using Node2vec, which is a graph embedding method that can automatically extract the spatial features of graphs, were combined with biological attribute features. To explore the effects of different network features' dimensions, we spliced three different dimensional features generated by Node2vec with the original features and performed the normalization operation on both sets of features. The model using the added 10-dimensional spatial features is denoted as GCN-CRF-Node10; the different results are shown in Table 1. The experiments demonstrated that combining the spatial features of graphs with biological attribute features can achieve a better performance. Although the best performance was not achieved in all metrics, we selected the relatively good features, with 20 dimensions added for subsequent experiments.

Table 1. Results of different feature dimensions from node2vec. The bold font in the table indicates the best performances of the model under these conditions.

	ACC Data			**KICH Data**		
	Acc	**Aupr**	**Auc**	**Acc**	**Aupr**	**Auc**
GCN-CRF-Node10	0.849	**0.813**	0.890	0.865	0.810	0.912
GCN-CRF-Node20	**0.855**	0.802	**0.898**	0.858	**0.814**	**0.919**
GCN-CRF-Node30	0.833	0.758	0.884	**0.866**	0.803	0.900

3.4. Ablation Experiments

The proposed method is based on a GCN with a CRF layer and network structure features. To evaluate their effects, ablation experiments were conducted. Three model variations are shown in Table 2: GCN is the traditional method with basic features, GCN-CRF is the GCN method with an added CRF layer, and GCN-CRF-Node20 represents the GCN-CRF model with an added network structure extracted using the Node2vec method. The results demonstrate that the GCN model with the CRF layer could better aggregate the information of similar nodes compared to using GCN alone. Additionally, the network structure features were useful and important in improving the performance.

Table 2. Results of ablation experiments. The bold font in the table indicates the best performances of the model under these conditions.

	ACC Data			**KICH Data**		
	Acc	**Aupr**	**Auc**	**Acc**	**Aupr**	**Auc**
GCN	0.821	0.772	0.879	0.804	0.807	0.904
GCN-CRF	0.850	0.796	0.892	0.821	0.813	0.902
GCN-CRF-Node20	**0.855**	**0.802**	**0.898**	**0.858**	**0.814**	**0.919**

3.5. Effects of Different Weights in Loss Function

The original GCN loss function is a binary cross-entropy loss function, and here, the original loss function could not make accurate judgments due to the imbalanced number of positive and negative sets. Based on the MODIG method proposed by Zhao et al. [35], different multiples of weights for the positive set in the best model we obtained in the previous step were applied, and the results are shown in Table 3. From the table, it can be seen that appropriate weights can further improve the performance of the model. Finally,

for the ACC data, we used three times the positive set weights, while for the KICH data, we used two times the positive set weights.

Table 3. Results of different weights for the positive set. The bold font in the table indicates the best performances of the model under these conditions.

	ACC Data			KICH Data		
	Acc	**Aupr**	**Auc**	**Acc**	**Aupr**	**Auc**
Weight of 1 time	0.855	0.802	0.898	**0.858**	0.814	0.919
Weight of 2 times	**0.861**	0.812	0.912	0.848	**0.816**	**0.929**
Weight of 3 times	0.851	**0.813**	**0.928**	0.830	0.803	0.916
Weight of 4 times	0.857	0.811	0.920	0.784	0.778	0.920
Weight of 5 times	0.846	0.790	0.920	0.767	0.795	0.897

3.6. Analysis at the Population Level

In this section, we compare the performance of our method with those of some existing methods at the population level since there is no factual benchmark for personalized driver gene identification. We selected some representative approaches in the field, including a frequency-based method, DriverMAPS [36], a network-based method, HotNet2 [13], and a machine learning-based method, sysSVM, which is also a method for identifying personalized cancer driver genes. The reason we selected the DriverMAPS and HotNet2 methods is that they have the best overall performances among 12 methods according to ref. [37].

To evaluate the performance of our method for the identification of driver genes, CGC was used as a basic benchmark. We selected the top 100 genes predicted by different methods and adopted three metrics, precision (P), recall (R), and the F1-score, to measure the performances of these methods. P represents the fraction of correctly predicted driver genes among predicted driver genes, while R represents the fraction of correctly predicted partial driver genes among the CGC driver genes. The F1-score, a combined metric of precision and recall, can effectively assess the ability of predicting cancer driver genes on the CGC database. It is calculated as $F1-score = 2 * \frac{P*R}{P+R}$. The results are shown in Figure 2, and it can be seen that our method outperformed the other methods in all three metrics for both datasets. This indicates the effectiveness of the method in identifying cancer driver genes at the population level.

3.7. Analysis at the Individual Level

The proposed method targets individuals. To validate the performance of the method at the individual level, we compared the proposed method with the individual-level sysSVM method [28]. Specifically, we analyzed each sample predicted by both methods, including 77 samples from ACC and 66 samples from KICH. The results of the ACC are shown in Figure 3 (for the results of KICH, see the Supplementary Materials, Section S4, see Figure S3), which indicates the number of driver genes in each sample according to the CGC database for the top 50 and top 100 driver genes predicted by the two methods. It is worth noting that our method performed better than sysSVM on most samples from both datasets, with sysSVM performing better on six samples from ACC and three samples from KICH.

3.8. Analysis of Identifying Rare Drivers

In this study, PDGCN could identify not only frequently mutated genes, but also rare driver genes, which are defined as those with a mutation frequency of <2% in the total patient cohort. We selected the identified rare genes in different samples and analyzed them according to previous studies. The results (see Supplementary Materials, Section S1) show that the identified rare genes play important roles in cancer.

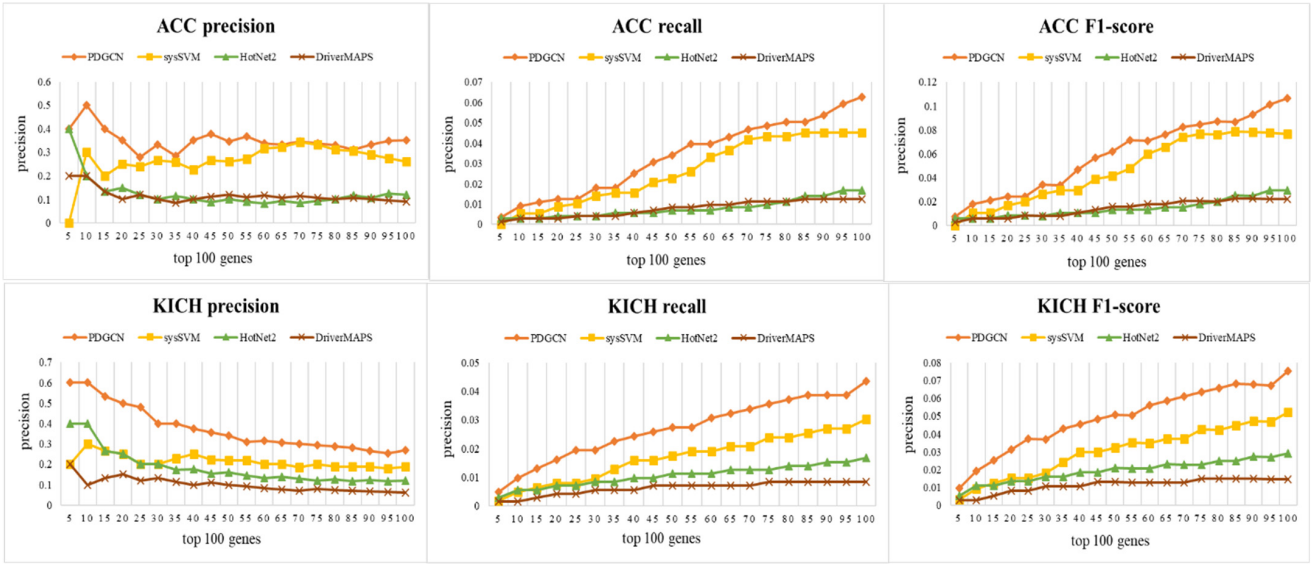
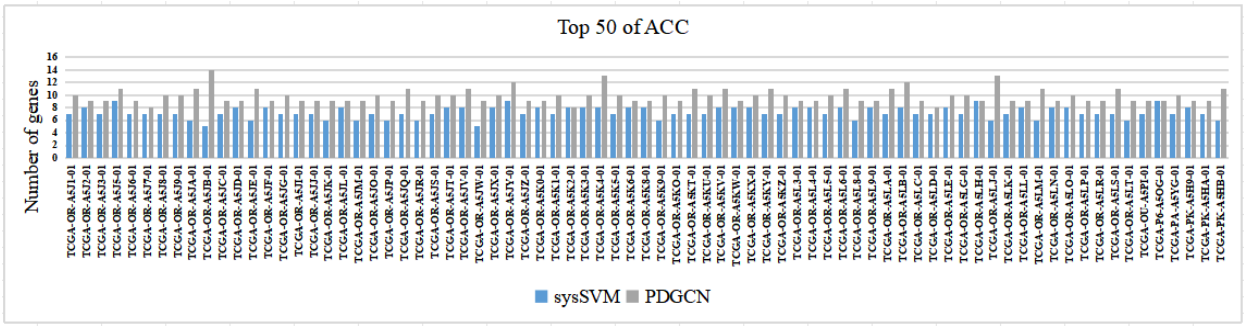
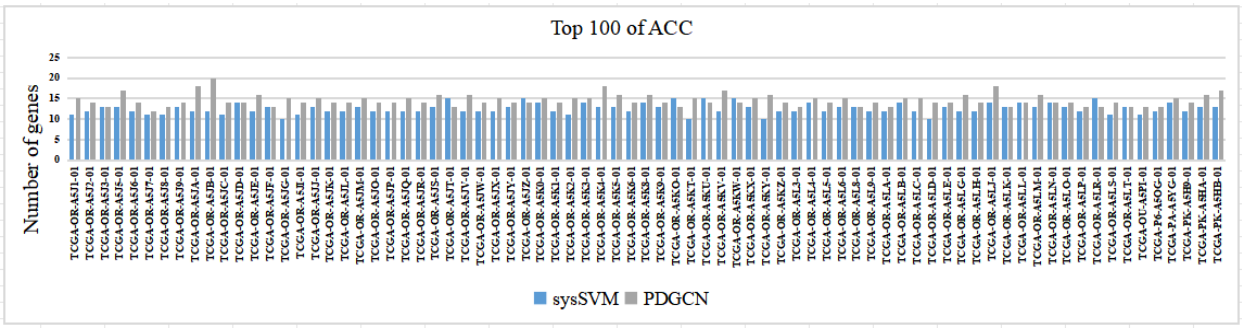


Figure 2. Performance comparisons (precision, recall, and F1-score) with PDGCN, sysSVM, HotNet2, and DriverMAPS methods on ACC and KICH datasets.



(**a**)



(**b**)

Figure 3. Comparisons of PDGCN and sysSVM at the individual level. The horizontal axis represents each sample, and the vertical axis represents the number of genes in the CGC: (**a**) Top 50 of ACC and (**b**) Top 100 of ACC.

3.9. Survival Analysis

To verify whether the genes identified using this method can differentiate the prognostic risk of cancer patients, we validated the identified genes by conducting a survival analysis. The top 50 candidate driver genes for cancer predicted using the PDGCN method were subjected to survival analysis using the online tool GEPIA2 (Gene Expression Profiling Interactive Analysis, http://gepia2.cancer-pku.cn) [38]. Genes with logrank $p < 0.05$ were considered significant biomarker genes [39]. Moreover, among these significant biomarker

genes, those not in the CGC dataset were considered novel candidate biomarker genes. Our method identified nine novel biomarker genes in the ACC data and seven novel biomarker genes in the KICH data. All novel biomarker genes identified were also mapped for survival analysis. The results of the ACC data are shown in Figure 4 (for the results of KICH, see Supplementary Materials, Section S2, see Figure S1). Furthermore, to validate the role of these novel biomarkers, a literature analysis was also conducted, and the results are shown in the Supplementary Materials, Section S2, see Table S1.

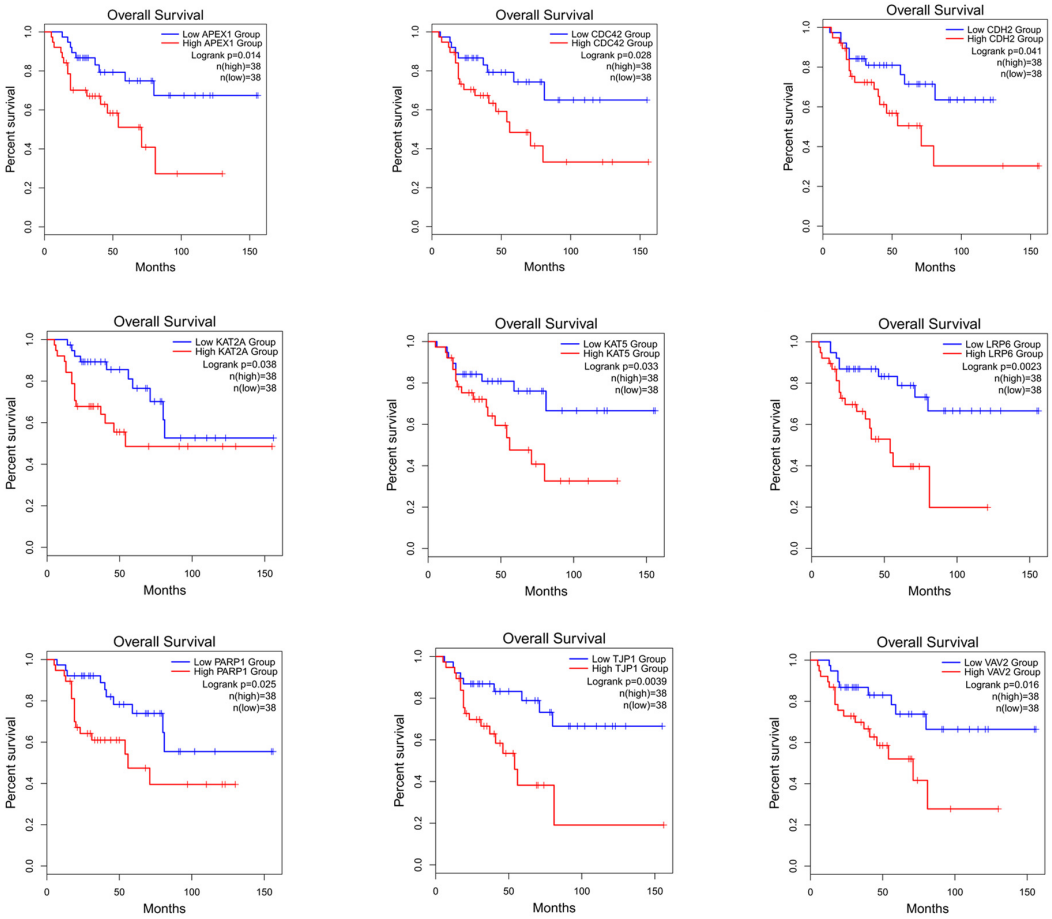


Figure 4. Survival analysis map of nine biomarker genes for ACC.

From the results, it can be seen that these novel biomarkers can efficiently distinguish the longer survival group from the shorter survival group.

3.10. Enrichment Analysis

To analyze the associations among the identified driver genes at the population level, we used the DAVID (the Database for Annotation, Visualization and Integrated Discovery) [40] online tool to perform KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis and GO (Gene Ontology) function enrichment analysis. KEGG is a database that specifically stores information on gene pathways in different species, while GO function enrichment analysis mainly annotates gene products in terms of the biological processes involved (GO-BP), cellular components (GO-CC), and molecular functions (GO-MF). The results are shown in Figure 5. For the ACC data, the KEGG pathway analysis mainly revealed that these genes are significantly enriched in some important cancer-related pathways. Regarding the GO functional enrichment, the identified driver genes are significantly enriched in cell migration regulation and promoter regulation in terms of GO-BP. These processes can interact with *Ago1* and positively affect gene expression in cancer cells [41]. Regarding GO-CC, these genes are significantly enriched in the

cytoplasm, cytosol, and cytoplasmic membrane. And specific hydrolase activity in ACC is positively correlated with cytoplasmic activity [42]. In terms of GO-MF, the analysis showed significant enrichment in the same protein binding sites and binding to kinase proteins. Corresponding inhibitors have been used as therapeutic agents for ACC [43]. For the KICH data, the KEGG results focused on some cancer-related pathways. In terms of the GO functional enrichment, driver genes were significantly enriched in some GO-BP terms; for example, the transformation of cellular proto-oncogenes into oncogenes leads to the over-activation of these signaling pathways, which in turn interact with the PI3K-Akt and Ras-ERK pathways to dysregulate cancer signaling and generate tumor cells [44]. Regarding GO-CC, the analysis showed significant enrichment in the cell membrane, nucleus, and, to a lesser extent, chromosomes. Related studies suggest that the deletion of DNA from chromosomes may be a unique feature of KICH [45]. In terms of GO-MF, a large number of proteins were enriched in protein binding, among which parvalbumin may be the KICH marker that distinguishes primary from metastatic tumors [46] (see Supplementary Materials, Section S3, see Figure S2).

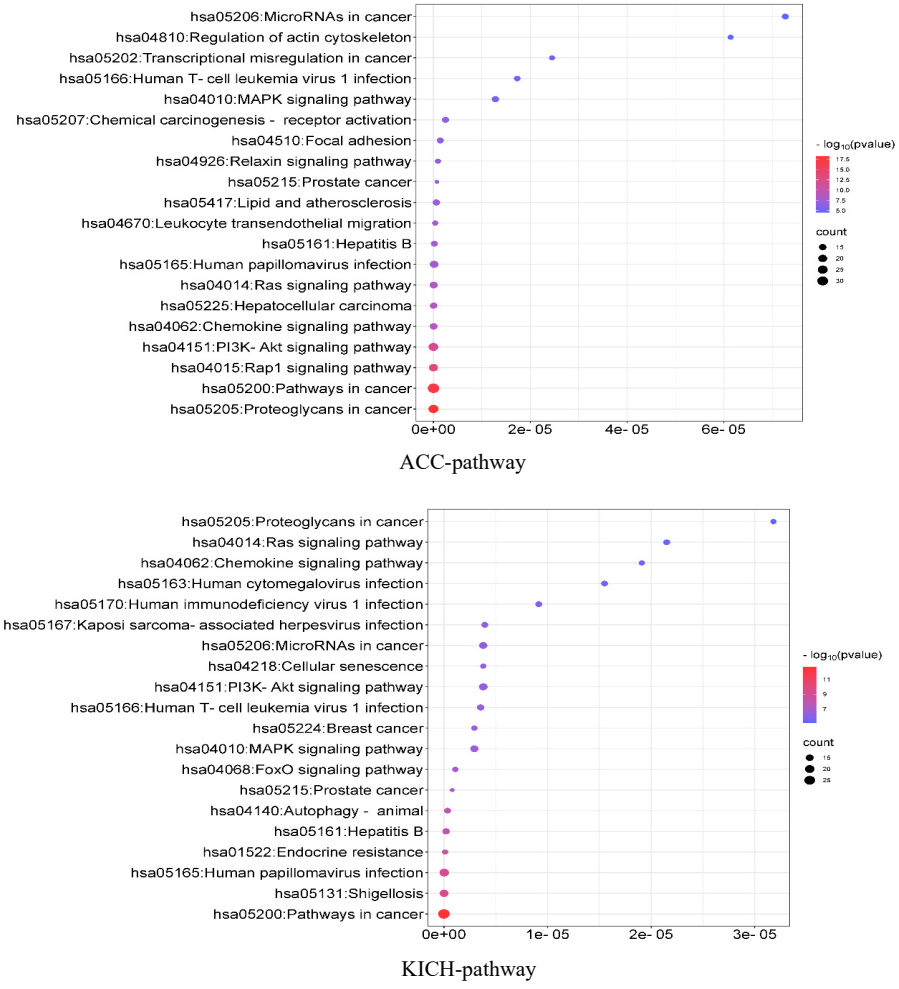


Figure 5. KEGG pathway enrichment analysis of ACC and KICH.

4. Conclusions and Discussion

Although research on cancer driver gene identification algorithms has made some progress, there is still a need for further improvements in overall performance and to address certain problems. Cancer samples are highly heterogeneous from one another, and with the increasing emphasis on precision and individualized medicine, it is necessary to develop individual driver gene prediction methods on a population basis. Some previous methods are more dependent on the gene interaction network constructed and

tend to ignore the similarity in characteristics between genes and typical known driver genes. Furthermore, some methods fail to achieve better fusion when integrating genomic, transcriptomic, and other multi-omics data and network data simultaneously. With the continuous development and application of machine learning, machine learning approaches have become increasingly successful in addressing various important biomedical problems. Machine learning plays an increasingly important role in developing models for predicting cancer driver genes.

Considering the limitations of previous driver gene identification methods and the advantages of machine learning, we propose a new machine learning approach, PDGCN, which applies a GCN to drive gene identification at the individual level. We constructed individual sample–gene networks in which each node was accompanied by information about each specific sample and gene. Additionally, the biological attribute features and spatial features of the genes were combined to constitute the enhanced features of the genes. We used a GCN with a CRF model to enhance the feature representations. We evaluated the performance of the method with different experiments and found that the method was more effective than other existing methods in identifying cancer driver genes at the population level. Furthermore, it was also able to identify novel biomarker genes, most of which have been confirmed in the literature to be associated with cancer. Personalized rare driver genes have also been detected and confirmed in the literature. Moreover, the enrichment analysis demonstrated that the predicted driver genes are significantly enriched in the GO terms and KEGG pathways. Overall, these findings suggest that our proposed method, PDGCN, can provide new insights into the molecular regulatory mechanisms during cancer development.

Although we constructed different networks based on different patients, the number of nodes in most patient gene networks did not significantly differ. This is because our approach was to construct personalized gene networks based on DEGs for specific cancers and mutated genes for individual samples. Moreover, we used only the STRING database to determine the interrelationship of patient genes, which may have resulted in less cancer type-specific information in the network. As a future research direction, we can consider specifying DEGs in different samples to make the specificity of the samples more apparent. Additionally, we can apply multiple gene interaction databases to enrich the relationships among genes and generate more possibilities. Good features are the key to improving model performance, and we will also consider feature selection techniques to optimize the way features are combined to achieve better results.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/biology13030184/s1: Figure S1: Survival analysis graph of 7 biomarker genes on KICH. Figure S2: KEGG pathway and GO function enrichment analysis. (a) ACC-GO-BP. (b) ACC-GOCC. (c) ACC-GO-MF. (d) KICH-GO-BP. (e) KICH-GO-CC. (f) KICH-GO-MF. Figure S3: The comparison of PDGCN and sysSVM at individual level. The horizontal axis represents each sample and the vertical axis represents the number of genes in CGC: (a) Top 50 of KICH, and (b) Top 100 of KICH. Table S1: The table shows the new biomarker genes identified in the two datasets and the corresponding descriptions. References [47–95] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, P.-J.W. and A.-D.Z.; methodology, P.-J.W. and A.-D.Z.; writing—original draft preparation, P.-J.W. and A.-D.Z.; writing—review and editing, P.-J.W., A.-D.Z., R.C. and C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Natural Science Foundation of China (No. 62202004) and Natural Science Foundation of Anhui Province (No. 2108085QF267).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The somatic mutation, expression, methylation, and copy number variation data of ACC and KICH are publicly available on the Xena platform: http://xena.ucsc.edu/ (accessed on 1 July 2022). The string dataset is publicly available at https://cn.string-db.org/ (accessed on 1 July 2022).

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

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef]
2. Loomans-Kropp, H.A.; Umar, A. Cancer prevention and screening: The next step in the era of precision medicine. *NPJ Precis. Oncol.* **2019**, *3*, 3. [CrossRef] [PubMed]
3. Anandakrishnan, R.; Varghese, R.T.; Kinney, N.A.; Garner, H.R. Estimating the number of genetic mutations (hits) required for carcinogenesis based on the distribution of somatic mutations. *PLoS Comput. Biol.* **2019**, *15*, e1006881. [CrossRef] [PubMed]
4. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719–724. [CrossRef] [PubMed]
5. Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A., Jr.; Kinzler, K.W. Cancer genome landscapes. *Science* **2013**, *339*, 1546–1558. [CrossRef]
6. Martínez-Jiménez, F.; Muiños, F.; Sentís, I.; Deu-Pons, J.; Reyes-Salazar, I.; Arnedo-Pac, C.; Mularoni, L.; Pich, O.; Bonet, J.; Kranas, H. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **2020**, *20*, 555–572. [CrossRef]
7. Ding, L.; Getz, G.; Wheeler, D.A.; Mardis, E.R.; McLellan, M.D.; Cibulskis, K.; Sougnez, C.; Greulich, H.; Muzny, D.M.; Morgan, M.B. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **2008**, *455*, 1069–1075. [CrossRef]
8. Banerji, S.; Cibulskis, K.; Rangel-Escareno, C.; Brown, K.K.; Carter, S.L.; Frederick, A.M.; Lawrence, M.S.; Sivachenko, A.Y.; Sougnez, C.; Zou, L.; et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **2012**, *486*, 405–409. [CrossRef]
9. Dees, N.D.; Zhang, Q.; Kandoth, C.; Wendl, M.C.; Schierding, W.; Koboldt, D.C.; Mooney, T.B.; Callaway, M.B.; Dooling, D.; Mardis, E.R.; et al. MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **2012**, *22*, 1589–1598. [CrossRef]
10. Tamborero, D.; Gonzalez-Perez, A.; Lopez-Bigas, N. OncodriveCLUST: Exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* **2013**, *29*, 2238–2244. [CrossRef]
11. Wood, L.D.; Parsons, D.W.; Jones, S.; Lin, J.; Sjoblom, T.; Leary, R.J.; Shen, D.; Boca, S.M.; Barber, T.; Ptak, J.; et al. The genomic landscapes of human breast and colorectal cancers. *Science* **2007**, *318*, 1108–1113. [CrossRef]
12. Bashashati, A.; Haffari, G.; Ding, J.; Ha, G.; Lui, K.; Rosner, J.; Huntsman, D.G.; Caldas, C.; Aparicio, S.A.; Shah, S.P. DriverNet: Uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol.* **2012**, *13*, 1–14. [CrossRef]
13. Leiserson, M.D.; Vandin, F.; Wu, H.-T.; Dobson, J.R.; Eldridge, J.V.; Thomas, J.L.; Papoutsaki, A.; Kim, Y.; Niu, B.; McLellan, M. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **2015**, *47*, 106–114. [CrossRef]
14. Schulte-Sasse, R.; Budach, S.; Hnisz, D.; Marsico, A. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nat. Mach. Intell.* **2021**, *3*, 513–526. [CrossRef]
15. Peng, W.; Tang, Q.; Dai, W.; Chen, T. Improving cancer driver gene identification using multi-task learning on graph convolutional network. *Brief. Bioinform.* **2022**, *23*, bbab432. [CrossRef] [PubMed]
16. Bertrand, D.; Chng, K.R.; Sherbaf, F.G.; Kiesel, A.; Chia, B.K.; Sia, Y.Y.; Huang, S.K.; Hoon, D.S.; Liu, E.T.; Hillmer, A. Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. *Nucleic Acids Res.* **2015**, *43*, e44. [CrossRef]
17. Hou, J.P.; Ma, J. DawnRank: Discovering personalized driver genes in cancer. *Genome Med.* **2014**, *6*, 56. [CrossRef] [PubMed]
18. Liu, X.; Wang, Y.; Ji, H.; Aihara, K.; Chen, L. Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res.* **2016**, *44*, e164. [CrossRef]
19. Guo, W.-F.; Zhang, S.-W.; Liu, L.-L.; Liu, F.; Shi, Q.-Q.; Zhang, L.; Tang, Y.; Zeng, T.; Chen, L. Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. *Bioinformatics* **2018**, *34*, 1893–1903. [CrossRef] [PubMed]
20. Guo, W.; Zhang, S.-W.; Zeng, T.; Li, Y.; Gao, J.; Chen, L. A novel network control model for identifying personalized driver genes in cancer. *PLoS Comput. Biol.* **2019**, *15*, e1007520. [CrossRef]
21. Pham, V.V.H.; Liu, L.; Bracken, C.P.; Nguyen, T.; Goodall, G.J.; Li, J.; Le, T.D. pDriver: A novel method for unravelling personalized coding and miRNA cancer drivers. *Bioinformatics* **2021**, *37*, 3285–3292. [CrossRef]
22. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M.; Canc Genome Atlas Res, N. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef]
23. Goldman, M.J.; Craft, B.; Hastie, M.; Repecka, K.; McDade, F.; Kamath, A.; Banerjee, A.; Luo, Y.; Rogers, D.; Brooks, A.N.; et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **2020**, *38*, 675–678. [CrossRef]
24. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. Available online: https://cn.string-db.org/ (accessed on 1 July 2022). [CrossRef] [PubMed]

25. Repana, D.; Nulsen, J.; Dressler, L.; Bortolomeazzi, M.; Venkata, S.K.; Tourna, A.; Yakovleva, A.; Palmieri, T.; Ciccarelli, F.D. The Network of Cancer Genes (NCG): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol.* **2019**, *20*, 1. [CrossRef] [PubMed]
26. Futreal, P.A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177–183. [CrossRef] [PubMed]
27. Bailey, M.H.; Tokheim, C.; Porta-Pardo, E.; Sengupta, S.; Bertrand, D.; Weerasinghe, A.; Colaprico, A.; Wendl, M.C.; Kim, J.; Reardon, B. Comprehensive characterization of cancer driver genes and mutations. *Cell* **2018**, *173*, 371–385. [CrossRef] [PubMed]
28. Nulsen, J.; Misetic, H.; Yau, C.; Ciccarelli, F.D. Pan-cancer detection of driver genes at the single-patient resolution. *Genome Med.* **2021**, *13*, 12. [CrossRef] [PubMed]
29. Grover, A.; Leskovec, J.; Assoc Comp, M. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
30. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online Learning of Social Representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, NY, USA, 24–27 August 2014; ACM: New York, NY, USA, 2014; pp. 701–710.
31. Niepert, M.; Ahmed, M.; Kutzkov, K. Learning Convolutional Neural Networks for Graphs. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 2014–2023.
32. Lafferty, J.; McCallum, A.; Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*; Williams College: Williamstown, MA, USA, 2001; pp. 282–289.
33. Long, Y.; Wu, M.; Kwoh, C.K.; Luo, J.; Li, X. Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* **2020**, *36*, 4918–4927. [CrossRef] [PubMed]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
35. Zhao, W.; Gu, X.; Chen, S.; Wu, J.; Zhou, Z. MODIG: Integrating multi-omics and multi-dimensional gene network for cancer driver gene identification based on graph attention network model. *Bioinformatics* **2022**, *38*, 4901–4907. [CrossRef]
36. Zhao, S.; Liu, J.; Nanga, P.; Liu, Y.; Cicek, A.E.; Knoblauch, N.; He, C.; Stephens, M.; He, X. Detailed modeling of positive selection improves detection of cancer driver genes. *Nat. Commun.* **2019**, *10*, 3399. [CrossRef]
37. Shi, X.; Teng, H.; Shi, L.; Bi, W.; Wei, W.; Mao, F.; Sun, Z. Comprehensive evaluation of computational methods for predicting cancer driver genes. *Brief. Bioinform.* **2022**, *23*, bbab548. [CrossRef]
38. Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **2019**, *47*, W556–W560. [CrossRef]
39. Zhang, S.-W.; Wang, Z.-N.; Li, Y.; Guo, W.-F. Prioritization of cancer driver gene with prize-collecting steiner tree by introducing an edge weighted strategy in the personalized gene interaction network. *Bmc Bioinform.* **2022**, *23*, 341. [CrossRef]
40. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4*, 44–57. [CrossRef]
41. Huang, V.; Zheng, J.; Qi, Z.; Wang, J.; Place, R.F.; Yu, J.; Li, H.; Li, L.-C. Ago1 Interacts with RNA Polymerase II and Binds to the Promoters of Actively Transcribed Genes in Human Cancer Cells. *PLoS Genet.* **2013**, *9*, e1003821. [CrossRef] [PubMed]
42. Papadopoulos, D.; Grondal, S.; Rydstrom, J.; DePierre, J.W. Levels of cytochrome P-450, steroidogenesis and microsomal and cytosolic epoxide hydrolases in normal human adrenal tissue and corresponding tumors. *Cancer Biochem. Biophys.* **1992**, *12*, 283–291. [PubMed]
43. Patalano, A.; Brancato, V.; Mantero, F. Adrenocortical Cancer Treatment. *Horm. Res.* **2009**, *71*, 99–104. [CrossRef] [PubMed]
44. Gomperts, B.D.; Tatham, P.E.; Kramer, I.M. *Signal Transduction*; Gulf Professional Publishing: Oxford, UK, 2002.
45. Akhtar, M.; Kardar, H.; Linjawi, T.; McClintock, J.; Ali, M.A. Chromophobe cell carcinoma of the kidney. A clinicopathologic study of 21 cases. *Am. J. Surg. Pathol.* **1995**, *19*, 1245–1256. [CrossRef] [PubMed]
46. Martignoni, G.; Pea, M.; Chilosi, M.; Brunelli, M.; Scarpa, A.; Colato, C.; Tardanico, R.; Zamboni, G.; Bonetti, F. Parvalbumin is constantly expressed in chromophobe renal carcinoma. *Mod. Pathol.* **2001**, *14*, 760–767. [CrossRef] [PubMed]
47. Liang, H.; Lin, Z.; Ye, Y.; Luo, R.; Zeng, L. ARRB2 promotes colorectal cancer growth through triggering WTAP. *Acta Biochim. Et Biophys. Sin.* **2021**, *53*, 85–93. [CrossRef]
48. Jiang, T.; Yu, J.T.; Wang, Y.L.; Wang, H.F.; Zhang, W.; Hu, N.; Tan, L.; Sun, L.; Tan, M.S.; Zhu, X.C. The genetic variation of ARRB2 is associated with late-onset Alzheimer's disease in Han Chinese. *Curr. Alzheimer Res.* **2014**, *11*, 408–412. [CrossRef]
49. Zhou, B.; Song, H.; Xu, W.; Zhang, Y.; Liu, Y.; Qi, W. The Comprehensive Analysis of Hub Gene ARRB2 in Prostate Cancer. *Dis. Markers* **2022**, *2022*, 8518378. [CrossRef]
50. Ma, T.; Zhao, Y.; Wei, K.; Yao, G.; Pan, C.; Liu, B.; Xia, Y.; He, Z.; Qi, X.; Li, Z. MicroRNA-124 functions as a tumor suppressor by regulating CDH2 and epithelial-mesenchymal transition in non-small cell lung cancer. *Cell. Physiol. Biochem.* **2016**, *38*, 1563–1574. [CrossRef]
51. Miao, J.; Wang, W.; Wu, S.; Zang, X.; Li, Y.; Wang, J.; Zhan, R.; Gao, M.; Hu, M.; Li, J. miR-194 suppresses proliferation and migration and promotes apoptosis of osteosarcoma cells by targeting CDH2. *Cell. Physiol. Biochem.* **2018**, *45*, 1966–1974. [CrossRef]

52. Chen, Q.; Cai, J.; Jiang, C. CDH2 expression is of prognostic significance in glioma and predicts the efficacy of temozolomide therapy in patients with glioblastoma. *Oncol. Lett.* **2018**, *15*, 7415–7422. [PubMed]
53. Cao, L.; Cheng, H.; Jiang, Q.; Li, H.; Wu, Z. APEX1 is a novel diagnostic and prognostic biomarker for hepatocellular carcinoma. *Aging* **2020**, *12*, 4573. [CrossRef] [PubMed]
54. Tummanatsakun, D.; Proungvitaya, T.; Roytrakul, S.; Limpaiboon, T.; Wongkham, S.; Wongkham, C.; Silsirivanit, A.; Somintara, O.; Sangkhamanon, S.; Proungvitaya, S. Serum apurinic/apyrimidinic endodeoxyribonuclease 1 (APEX1) level as a potential biomarker of cholangiocarcinoma. *Biomolecules* **2019**, *9*, 413. [CrossRef] [PubMed]
55. Yang, J.; Yang, D.; Cogdell, D.; Du, X.; Li, H.; Pang, Y.; Sun, Y.; Hu, L.; Sun, B.; Trent, J. APEX1 gene amplification and its protein overexpression in osteosarcoma: Correlation with recurrence, metastasis, and survival. *Technol. Cancer Res. Treat.* **2010**, *9*, 161–169. [CrossRef] [PubMed]
56. Li, J.; Feng, Q.; Qi, Y.; Cui, G.; Zhao, S. PPARGC1A is upregulated and facilitates lung cancer metastasis. *Exp. Cell Res.* **2017**, *359*, 356–360. [CrossRef] [PubMed]
57. Xiao, X.; Wang, W.; Li, Y.; Yang, D.; Li, X.; Shen, C.; Liu, Y.; Ke, X.; Guo, S.; Guo, Z. HSP90AA1-mediated autophagy promotes drug resistance in osteosarcoma. *J. Exp. Clin. Cancer Res.* **2018**, *37*, 201. [CrossRef] [PubMed]
58. Okino, K.; Nagai, H.; Hatta, M.; Nagahata, T.; Yoneyama, K.; Ohta, Y.; Jin, E.; Kawanami, O.; Araki, T.; Emi, M. Up-regulation and overproduction of DVL-1, the human counterpart of the Drosophila dishevelled gene, in cervical squamous cell carcinoma. *Oncol. Rep.* **2003**, *10*, 1219–1223. [CrossRef] [PubMed]
59. Nagahata, T.; Shimada, T.; Harada, A.; Nagai, H.; Onda, M.; Yokoyama, S.; Shiba, T.; Jin, E.; Kawanami, O.; Emi, M. Amplification, up-regulation and over-expression of DVL-1, the human counterpart of the Drosophila disheveled gene, in primary breast cancers. *Cancer Sci.* **2003**, *94*, 515–518. [CrossRef]
60. Smith, M.J.; O'Sullivan, J.; Bhaskar, S.S.; Hadfield, K.D.; Poke, G.; Caird, J.; Sharif, S.; Eccles, D.; Fitzpatrick, D.; Rawluk, D. Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nat. Genet.* **2013**, *45*, 295–298. [CrossRef]
61. Wang, P.; Xie, M.; Yang, D.; Wang, F.; Chen, E. Integrative multi-omics analysis reveals the landscape of Cyclin-Dependent Kinase (CDK) family genes in pan-cancer. *Res. Sq.* **2022**. [CrossRef]
62. Caliskan, A.; Andac, A.C.; Arga, K.Y. Novel molecular signatures and potential therapeutics in renal cell carcinomas: Insights from a comparative analysis of subtypes. *Genomics* **2020**, *112*, 3166–3178. [CrossRef]
63. Zhou, L.; Yin, B.; Liu, Y.; Hong, Y.; Zhang, C.; Fan, J. Mechanism and function of decreased FOXO1 in renal cell carcinoma. *J. Surg. Oncol.* **2012**, *105*, 841–847. [CrossRef]
64. Kojima, T.; Shimazui, T.; Horie, R.; Hinotsu, S.; Oikawa, T.; Kawai, K.; Suzuki, H.; Meno, K.; Akaza, H.; Uchida, K. FOXO1 and TCF7L2 genes involved in metastasis and poor prognosis in clear cell renal cell carcinoma. *Genes Chromosomes Cancer* **2010**, *49*, 379–389. [CrossRef]
65. Xu, J.; Perecman, A.; Wiggins, A.; Kalantzakos, T.; Das, S.; Sullivan, T.; Rieger-Christ, K. MetastamiRs in Renal Cell Carcinoma: An Overview of MicroRNA Implicated in Metastatic Kidney Cancer. *Exon Publ.* **2022**, 71–93.
66. Erdem, M.; Erdem, S.; Sanli, O.; Sak, H.; Kilicaslan, I.; Sahin, F.; Telci, D. Up-regulation of TGM2 with ITGB1 and SDC4 is important in the development and metastasis of renal cell carcinoma. In *Urologic Oncology: Seminars and Original Investigations*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 32, pp. 25.e13–25.e20.
67. Bruder, E.; Moch, H.; Ehrlich, D.; Leuschner, I.; Harms, D.; Argani, P.; Briner, J.; Graf, N.; Selle, B.; Rufle, A. Wnt signaling pathway analysis in renal cell carcinoma in young patients. *Mod. Pathol.* **2007**, *20*, 1217–1229. [CrossRef] [PubMed]
68. Cui, J.; Yuan, Y.; Shanmugam, M.K.; Anbalagan, D.; Tan, T.Z.; Sethi, G.; Kumar, A.P.; Lim, L.H.K. MicroRNA-196a promotes renal cancer cell migration and invasion by targeting BRAM1 to regulate SMAD and MAPK signaling pathways. *Int. J. Biol. Sci.* **2021**, *17*, 4254. [CrossRef] [PubMed]
69. Dirim, A.; Haberal, A.N.; Goren, M.R.; Tekin, M.I.; Peskircioglu, L.; Demirhan, B.; Ozkardes, H. VEGF, COX-2, and PCNA expression in renal cell carcinoma subtypes and their prognostic value. *Int. Urol. Nephrol.* **2008**, *40*, 861–868. [CrossRef] [PubMed]
70. Altintas, E.; Kaynar, M.; Celik, Z.E.; Celik, M.; Kilic, O.; Akand, M.; Goktas, S. Expression of Ring Box-1 protein and its relationship with Fuhrman grade and other clinical-pathological parameters in renal cell cancer. In *Urologic Oncology: Seminars and Original Investigations*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 38, pp. 6.e17–6.e22.
71. Chen, C.; Chi, H.; Min, L.; Junhua, Z. Downregulation of guanine nucleotide-binding protein beta 1 (GNB1) is associated with worsened prognosis of clearcell renal cell carcinoma and is related to VEGF signaling pathway. *JBUON* **2017**, *22*, 1441–1446.
72. Gara, S.K.; Wang, Y.; Patel, D.; Liu-Chittenden, Y.; Jain, M.; Boufraqech, M.; Zhang, L.; Meltzer, P.S.; Kebebew, E. Integrated genome-wide analysis of genomic changes and gene regulation in human adrenocortical tissue samples. *Nucleic Acids Res.* **2015**, *43*, 9327–9339. [CrossRef]
73. Kaidi, A.; Jackson, S.P. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. *Nature* **2013**, *498*, 70–74. [CrossRef]
74. Mouat, I.C.; Omata, K.; McDaniel, A.S.; Hattangady, N.G.; Talapatra, D.; Cani, A.K.; Hovelson, D.H.; Tomlins, S.A.; Rainey, W.E.; Hammer, G.D. Somatic mutations in adrenocortical carcinoma with primary aldosteronism or hyperreninemic hyperaldosteronism. *Endocr.-Relat. Cancer* **2019**, *26*, 217–225. [CrossRef]
75. Lin, S.; Qiu, L.; Liang, K.; Zhang, H.; Xian, M.; Chen, Z.; Wei, J.; Fu, S.; Gong, X.; Ding, K. KAT2A/E2F1 Promotes Cell Proliferation and Migration via Upregulating the Expression of UBE2C in Pan-Cancer. *Genes* **2022**, *13*, 1817. [CrossRef]

76. Altieri, B.; Ronchi, C.L.; Kroiss, M.; Fassnacht, M. Next-generation therapies for adrenocortical carcinoma. *Best Pract. Res. Clin. Endocrinol. Metab.* **2020**, *34*, 101434. [CrossRef]
77. Shaikh, L.H.; Zhou, J.; Teo, A.E.D.; Garg, S.; Neogi, S.G.; Figg, N.; Yeo, G.S.; Yu, H.; Maguire, J.J.; Zhao, W. LGR5 activates noncanonical Wnt signaling and inhibits aldosterone production in the human adrenal. *J. Clin. Endocrinol. Metab.* **2015**, *100*, E836–E844. [CrossRef]
78. Ruggiero, C.; Lalli, E. VAV2: A novel prognostic marker and a druggable target for adrenocortical carcinoma. *Oncotarget* **2017**, *8*, 88257. [CrossRef]
79. Ruggiero, C.; Doghman-Bouguerra, M.; Sbiera, S.; Sbiera, I.; Parsons, M.; Ragazzon, B.; Morin, A.; Robidel, E.; Favier, J.; Bertherat, J.; et al. Dosage-dependent regulation of VAV2 expression by steroidogenic factor-1 drives adrenocortical carcinoma cell invasion. *Sci. Signal.* **2017**, *10*, eaal2464. [CrossRef]
80. Parviainen, H.; Schrade, A.; Kiiveri, S.; Prunskaite-Hyyrylainen, R.; Haglund, C.; Vainio, S.; Wilson, D.B.; Arola, J.; Heikinheimo, M. Expression of Wnt and TGF-β pathway components and key adrenal transcription factors in adrenocortical tumors: Association to carcinoma aggressiveness. *Pathol.-Res. Pract.* **2013**, *209*, 503–509. [CrossRef]
81. Zhu, Y.; Wang, M.; Zhao, X.; Zhang, L.; Wu, Y.; Wang, B.; Hu, W. Rottlerin as a novel chemotherapy agent for adrenocortical carcinoma. *Oncotarget* **2017**, *8*, 22825. [CrossRef]
82. Gayarre, J.; Kamieniak, M.M.; Cazorla-Jiménez, A.; Muñoz-Repeto, I.; Borrego, S.; García-Donas, J.; Hernando, S.; Robles-Díaz, L.; García-Bueno, J.M.; y Cajal, T.R. The NER-related gene GTF2H5 predicts survival in high-grade serous ovarian cancer patients. *J. Gynecol. Oncol.* **2016**, *27*. [CrossRef] [PubMed]
83. Mukherjee, M.; Fogarty, E.; Janga, M.; Surendran, K. Notch signaling in kidney development, maintenance, and disease. *Biomolecules* **2019**, *9*, 692. [CrossRef]
84. Peri, S.; Devarajan, K.; Yang, D.H.; Knudson, A.G.; Balachandran, S. Meta-analysis identifies NF-κB as a therapeutic target in renal cancer. *PLoS ONE* **2013**, *8*, e76746. [CrossRef]
85. Lind, G.E.; Kleivi, K.; Meling, G.I.; Teixeira, M.R.; Thiis-Evensen, E.; Rognum, T.O.; Lothe, R.A. ADAMTS1, CRABP1, and NR3C1 identified as epigenetically deregulated genes in colorectal tumorigenesis. *Anal. Cell. Pathol.* **2006**, *28*, 259–272. [CrossRef] [PubMed]
86. Zhang, L.; Song, L.; Xu, Y.; Xu, Y.; Zheng, M.; Zhang, P.; Wang, Q. Midkine promotes breast cancer cell proliferation and migration by upregulating NR3C1 expression and activating the NF-κB pathway. *Mol. Biol. Rep.* **2022**, *49*, 2953–2961. [CrossRef] [PubMed]
87. Jakob, J.A.; Bassett, R.L., Jr.; Ng, C.S.; Curry, J.L.; Joseph, R.W.; Alvarado, G.C.; Rohlfs, M.L.; Richard, J.; Gershenwald, J.E.; Kim, K.B. NRAS mutation status is an independent prognostic factor in metastatic melanoma. *Cancer* **2012**, *118*, 4014–4023. [CrossRef] [PubMed]
88. Therkildsen, C.; Bergmann, T.K.; Henrichsen-Schnack, T.; Ladelund, S.; Nilbert, M. The predictive value of KRAS, NRAS, BRAF, PIK3CA and PTEN for anti-EGFR treatment in metastatic colorectal cancer: A systematic review and meta-analysis. *Acta Oncol.* **2014**, *53*, 852–864. [CrossRef] [PubMed]
89. Chang, S.; Cao, Y. Differentially expressed genes SNRPC and PRPF38A are potential biomarkers candidates for osteosarcoma. *Res. Sq.* **2020**. [CrossRef]
90. Liu, Y.; Ni, R.; Zhang, H.; Miao, L.; Wang, J.; Jia, W.; Wang, Y. Identification of feature genes for smoking-related lung adenocarcinoma based on gene expression profile data. *OncoTargets Ther.* **2016**, *9*, 7397. [CrossRef]
91. Sathe, A.; Nawroth, R. Targeting the PI3K/AKT/mTOR Pathway in Bladder Cancer. *Urothelial Carcinoma Methods Protoc.* **2018**, *1665*, 335–350.
92. Schiffman, M.; Doorbar, J.; Wentzensen, N.; De Sanjosé, S.; Fakhry, C.; Monk, B.J.; Stanley, M.A.; Franceschi, S. Carcinogenic human papillomavirus infection. *Nat. Rev. Dis. Primers* **2016**, *2*, 16086. [CrossRef] [PubMed]
93. Kim, E.K.; Choi, E.J. Pathological roles of MAPK signaling pathways in human diseases. *Biochim. Et Biophys. Acta (BBA)-Mol. Basis Dis.* **2010**, *1802*, 396–405. [CrossRef] [PubMed]
94. Di Leva, G.; Garofalo, M.; Croce, C.M. MicroRNAs in cancer. *Annu. Rev. Pathol.* **2014**, *9*, 287. [CrossRef]
95. Mazal, P.R.; Exner, M.; Haitel, A.; Krieger, S.; Thomson, R.B.; Aronson, P.S.; Susani, M. Expression of kidney-specific cadherin distinguishes chromophobe renal cell carcinoma from renal oncocytoma. *Hum. Pathol.* **2005**, *36*, 22–28. [CrossRef]