TARGET DISCOVERY FOR ANTICANCER THERAPY FACILITATED BY ARTIFICIAL INTELLIGENCE

EDITED BY: Feng Zhu, Yu Zong Chen and Weiwei Xue
PUBLISHED IN: Frontiers in Pharmacology and Froniters in Oncology

frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88971-200-7
DOI 10.3389/978-2-88971-200-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

TARGET DISCOVERY FOR ANTICANCER THERAPY FACILITATED BY ARTIFICIAL INTELLIGENCE

Topic Editors:
Feng Zhu, Zhejiang University, China
Yu Zong Chen, National University of Singapore, Singapore
Weiwei Xue, Chongqing University, China

Citation: Zhu, F., Chen, Y. Z., Xue, W., eds. (2021). Target Discovery for Anticancer Therapy Facilitated by Artificial Intelligence. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-200-7

Table of Contents

05 ***Probing the Druggablility on the Interface of the Protein–Protein Interaction and Its Allosteric Regulation Mechanism on the Drug Screening for the CXCR4 Homodimer***
Liting Shen, Yuan Yuan, Yanzhi Guo, Menglong Li, Chuan Li and Xuemei Pu

20 ***Improved Prediction of Aqueous Solubility of Novel Compounds by Going Deeper With Deep Learning***
Qiuji Cui, Shuai Lu, Bingwei Ni, Xian Zeng, Ying Tan, Ya Dong Chen and Hongping Zhao

29 ***Identifying Drug Targets in Pancreatic Ductal Adenocarcinoma Through Machine Learning, Analyzing Biomolecular Networks, and Structural Modeling***
Wenying Yan, Xingyi Liu, Yibo Wang, Shuqing Han, Fan Wang, Xin Liu, Fei Xiao and Guang Hu

43 ***Extracellular Vesicles in Renal Cell Carcinoma: Multifaceted Roles and Potential Applications Identified by Experimental and Computational Methods***
Zhiyuan Qin, Qingwen Xu, Haihong Hu, Lushan Yu and Su Zeng

58 ***Identification of Prognostic miRNA Signature and Lymph Node Metastasis-Related Key Genes in Cervical Cancer***
Shuoling Chen, Chang Gao, Yangyuan Wu and Zunnan Huang

72 ***Discovering Anti-Cancer Drugs* via *Computational Methods***
Wenqiang Cui, Adnane Aouidate, Shouguo Wang, Qiuliyang Yu, Yanhua Li and Shuguang Yuan

86 ***Preparation of RGD Peptide/Folate Acid Double-Targeted Mesoporous Silica Nanoparticles and Its Application in Human Breast Cancer MCF-7 Cells***
Huijie Yan, Yun You, Xinjian Li, Lei Liu, Fengqian Guo, Qiongling Zhang, Dewen Liu, Yan Tong, Shilan Ding and Jinyu Wang

96 ***Molecular Screening for Nigericin Treatment in Pancreatic Cancer by High-Throughput RNA Sequencing***
Zhihua Xu, Guanzhuang Gao, Fei Liu, Ye Han, Chen Dai, Sentai Wang, Guobang Wei, Yuting Kuang, Daiwei Wan, Qiaoming Zhi and Ye Xu

113 ***Integrated PPI- and WGCNA-Retrieval of Hub Gene Signatures Shared Between Barrett's Esophagus and Esophageal Adenocarcinoma***
Asma Sindhoo Nangraj, Gurudeeban Selvaraj, Satyavani Kaliamurthi, Aman Chandra Kaushik, William C. Cho and Dong Qing Wei

127 ***Artificial Intelligence to Decode Cancer Mechanism: Beyond Patient Stratification for Precision Oncology***
Sandip Kumar Patel, Bhawana George and Vineeta Rai

144 ***Combined Metabolomic Analysis of Plasma and Tissue Reveals a Prognostic Risk Score System and Metabolic Dysregulation in Esophageal Squamous Cell Carcinoma***
Zhongjian Chen, Yalan Dai, Xiancong Huang, Keke Chen, Yun Gao, Na Li, Ding Wang, Aiping Chen, Qingxia Yang, Yanjun Hong, Su Zeng and Weimin Mao

157 ***Evaluation of Artificial Intelligence in Participating Structure-Based Virtual Screening for Identifying Novel Interleukin-1 Receptor Associated Kinase-1 Inhibitors***
Jinxin Che, Ruiwei Feng, Jian Gao, Hongyun Yu, Qinjie Weng, Qiaojun He, Xiaowu Dong, Jian Wu and Bo Yang

169 ***Integrating Machine Learning-Based Virtual Screening With Multiple Protein Structures and Bio-Assay Evaluation for Discovery of Novel GSK3β Inhibitors***
Jingyu Zhu, Yuanqing Wu, Man Wang, Kan Li, Lei Xu, Yun Chen, Yanfei Cai and Jian Jin

180 ***Discovery of Icotinib-1,2,3-Triazole Derivatives as IDO1 Inhibitors***
Long-fei Mao, Yu-wei Wang, Jie Zhao, Gui-qing Xu, Xiao-jun Yao and Yue-Ming Li

190 ***Discovery of Novel IDH1 Inhibitor Through Comparative Structure-Based Virtual Screening***
Yuwei Wang, Shuai Tang, Huanling Lai, Ruyi Jin, Xu Long, Na Li, Yuping Tang, Hui Guo, Xiaojun Yao and Elaine Lai-Han Leung

201 ***Targeting Nrf2-Mediated Oxidative Stress Response Signaling Pathways as New Therapeutic Strategy for Pituitary Adenomas***
Xianquan Zhan, Jiajia Li and Tian Zhou

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 07 November 2019
doi: 10.3389/fphar.2019.01310

Probing the Druggablility on the Interface of the Protein–Protein Interaction and Its Allosteric Regulation Mechanism on the Drug Screening for the CXCR4 Homodimer

*Liting Shen[1], Yuan Yuan[2], Yanzhi Guo[1], Menglong Li[1], Chuan Li[3] and Xuemei Pu[1]**

[1] College of Chemistry, Sichuan University, Chengdu, China, [2] College of Management, Southwest University for Nationalities, Chengdu, China, [3] College of Computer Science, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Hai-Bin Luo,
Sun Yat-sen University, China
Tingjun Hou,
Zhejiang University, China
Jin Yu,
University of California, Irvine, United States

***Correspondence:**
Xuemei Pu
xmpuscu@scu.edu.cn

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *23 August 2019*
Accepted: *15 October 2019*
Published: *07 November 2019*

Citation:
Shen L, Yuan Y, Guo Y, Li M, Li C and Pu X (2019) Probing the Druggablility on the Interface of the Protein–Protein Interaction and Its Allosteric Regulation Mechanism on the Drug Screening for the CXCR4 Homodimer. Front. Pharmacol. 10:1310. doi: 10.3389/fphar.2019.01310

Modulating protein–protein interactions (PPIs) with small drug-like molecules targeting it exhibits great promise in modern drug discovery. G protein-coupled receptors (GPCRs) are the largest family of targeted proteins and could form dimers in living biological cells through PPIs. However, compared to drug development of the orthosteric site, there has been lack of investigations on the druggability of the PPI interface for GPCRs and its functional implication on experiments. Thus, in order to address these issues, we constructed a novel computational strategy, which involved in molecular dynamics simulation, virtual screening and protein structure network (PSN), to study one representative GPCR homodimer (CXCR4). One druggable pocket was identified in the PPI interface and one small molecule targeting it was screened, which could strengthen PPI mainly through hydrophobic interaction between the benzene rings of the PPI molecule and TM4 of the receptor. The PSN results further reveals that the PPI molecule could increase the number of the allosteric regulation pathways between the druggable pocket of the dimer interface to the orthostatic site for the subunit A but only play minor role for the other subunit B, leading to the asymmetric change in the volume of the binding pockets for the two subunits (increase for the subunit A and minor change for the subunit B). Consequently, the screening performance of the subunit A to the antagonists is enhanced while the subunit B is unchanged nearly, implying that the PPI molecule may be beneficial to enhance the drug efficacies of the antagonists. In addition, one main regulation pathway with the highest frequency was identified for the subunit A, which consists of Trp195$^{5.34}$–Tyr190^{ECL2}–Val196$^{5.35}$–Gln200$^{5.39}$–Asp262$^{6.58}$–Cys28$^{N\text{-}term}$, revealing their importance in the allosteric regulation from the PPI molecule. The observations from the work could provide valuable information for the development of the PPI drug-like molecule for GPCRs.

Keywords: target, GPCR dimer interface, druggability, regulation mechanism, computation

INTRODUCTION

In living cells, only a few proteins perform their biological functions independently, and the vast majority (more than 80%) of proteins function through interacting with other molecules (Keskin et al., 2016; Wang et al., 2018b). It is estimated that there are approximately 130,000 to 650,000 protein–protein interactions (PPIs) in the human interactome (Venkatesan et al., 2009; Sheng et al., 2015; Tortorella et al., 2016), and targeting protein–protein interactions (PPIs) with small drug-like molecules (Sheng et al., 2015; Shin et al., 2017; Han et al., 2018) become one of the most promising methods in modern drug discovery (Li et al., 2017; Tang et al., 2019a; Tang et al., 2019b). If drugs could strengthen the PPI interaction or damage it, the function of PPI will be inevitably influenced. With increasing understanding of PPIs, significant progress has been made for investigations on PPI small drug-like molecules (Wells and McClendon, 2007; Jubb et al., 2012; Song et al., 2016; Shin et al., 2017). It was observed that the PPI molecules commonly have larger molecular weight, higher hydrophobicity, and lower solubility than most of traditional drug molecules (Villoutreix et al., 2012; Wang et al., 2018b). In previous studies, most of the drugs bind a targeted protein and inhibit it to form functional complexes with its binding partners, in turn influencing the downstream signals. For example, small molecule LEDGINs could block the interaction between HIV-1 integrase and human LEDGF/p75 so that it could inhibit HIV replication(Reddy et al., 2014). The peptidemimics MAIT was found to inhibit the migration of colorectal cells by disrupting APC–Asef interaction (Jiang et al., 2017). Although researches on strengthening PPI interaction are very limited with respect to inhibiting one, it is also highly valuable for some specific proteins. For example, ISD could strengthen the interaction between Neph1 and ZO-1 so that it could prevent podocyte injury and preserve glomerular filtration function (Sagar et al., 2017).

G protein-coupled receptors (GPCRs) are the largest membrane protein families with more than 800 members, which play key roles in various signal transductions. Approximate 50 percent of drugs target them (Rosenbaum et al., 2009; Venkatakrishnan et al., 2013; Lao et al., 2017). Monomers have long been recognized as functional units of GPCR signaling (Whorton et al., 2007; Maurice et al., 2011). However, recently increasing biochemical and biophysical evidences have indicated that the GPCR dimers and oligomers also exist in living biological cells (Ferré et al., 2014; Navarro et al., 2018; Pediani et al., 2018), which could significantly affect the signal transduction process of GPCRs like receptor activation, internalization, ligand binding and coupling with G protein (Huang et al., 2013; Xue et al., 2015; Damian et al., 2018). Some experimental works already found that positive or negative cooperativity exists between the two subunits of the GPCR dimer (Maurice et al., 2011). For example, when the ligand binds to one of the subunits, it will increase or decrease the binding affinity of another subunit to the ligand (Gherbi et al., 2015; Liu et al., 2017). Therefore, the GPCR dimers possess unique pharmacological profiles, being potential drug targets for the discovery of novel drugs.

Chemokine receptors are members of family A GPCRs, which regulate cell migration in development, immune system function and inflammatory diseases, thus being important therapeutic targets (Kufareva et al., 2014; Van Hout et al., 2018). CXCR4 is one of 23 known human chemokine receptors, which plays a key role in leukocyte trafficking, hematopoiesis, organ development and cancer metastases (Zweemer et al., 2014). It was revealed that CXCR4 is associated with more than 23 types of cancers (Wu et al., 2010; Nguyen et al., 2018). CXCR4 and related CC chemokine receptor 5 (CCR5) are not only the key regulators of signal transduction, but also involve in the entry of HIV-1 virus as coreceptors of HIV-1 into leukocytes (Shaik et al., 2019). Several observations suggested that the dimer may be the minimal functional units of the chemokine receptors and CXCR4 was demonstrated to form homo- or hetero-dimers (Percherancier et al., 2005; Muñoz et al., 2012). In 2010, the crystal structure of the CXCR4 homo-dimer (PDBID: 3ODU) was resolved (Wu et al., 2010). Chemotaxis assay shows that the migration index of T-REx-293 cells stably transfected with CXCR4 gene changes with the oligomeric status of CXCR4, indicating a correlation between the functions and the oligomeric status of CXCR4 (Lao et al., 2017). These findings clearly indicate that the polymerization of GPCRs could affect the structure and the function of the receptors. Therefore, it is also valuable to design small drug-like molecules targeting PPIs of GPCRs, which are beneficial to their therapeutic effects, to enhance their polymerization. However, the investigations on drug-like small molecules targeting the interface to enhance PPIs have been lacked so far. Thus, many questions have been remained to be unclear. For example, is the PPI interface druggable for GPCRs like CXCR4? What drugs could target the interface? How does the drug regulate the dimerization and the structure of the receptor, in turn influence its drug efficacy? In fact, these questions mainly involve in microscopic structure changes of the receptor upon the ligand bound the PPI interface. Therefore, it is highly desired to introduce computational techniques to assist the experiments to probe these issues.

Molecular dynamics (MD) simulation could acquire the structural evolution of proteins at the atomic level. Therefore, it become a powerful tool to study the structural and functional mechanisms for biological systems (Xue et al., 2018), including GPCRs. However, previous MD researches on GPCRs were mainly focused on the GPCR monomers, including their structures (Liang et al., 2017; Zhang et al., 2018a), interactions with ligands (Bai et al., 2014; Sader et al., 2018), activation mechanisms (Miao et al., 2015; Stanley et al., 2016), water channels (Yuan et al., 2013; Yuan et al., 2015) and so on. In contrast, the studies on GPCR oligomers by MD are very limited, mainly concerning the self-assembly behavior (Provasi et al., 2015), activation mechanisms (Kim et al., 2017), interaction of dimers (Petersen et al., 2017). Recently, our group probed the effect of the dimerization on the activation and ligand-binding for some GPCRs (Wang et al., 2018a; Zhang et al., 2018b; Zhang et al., 2019). Based on our previous studies on the mechanism of GPCR dimers, we hope to further probe the druggability of their PPI interfaces and its regulation mechanism on the drug function of the receptor, using molecular dynamics simulation,

virtual screening and protein structure network. Herein, we selected the CXCR4 dimer (PDBID:3ODU) as a representative of the GPCR dimers, which is sole crystal-structure of the GPCR dimers resolved for the chemokine receptors so far. Ultimately, we screened one ligand, which could significantly enhance the dimer interaction, and revealed its regulation mechanism on the drug binding for the orthosteric site. The observations could provide valuable information for the development of the GPCR PPI drugs.

MATERIALS AND METHODS

Workflow

Figure 1 shows the entire workflow. Considering the protein flexibility, 300 ns MD is first carried out for the crystal structure of CXCR4 homo-dimer. Then, according to the root-mean-square-deviation (RMSD) of residues of the dimer interface, eight representative conformations are obtained through clustering. The druggable pocket of the dimer interface is identified by FTmap. Based on the pocket, ligands targeting the PPI interface are screened, and then 1us MD simulations are performed for the four representative complexes of the dimer bound by the PPI ligand. Finally, we discussed the effect of the ligand on the dimerization and the screening performance of the orthostatic site to antagonists, and revealed its regulation mechanism.

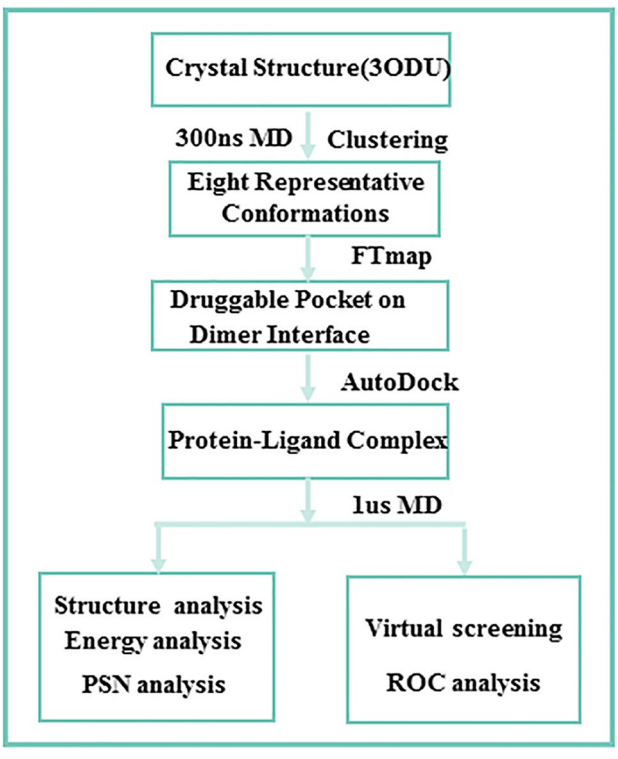


FIGURE 1 | Computational workflow.

System Preparation

X-ray crystal structure of CXCR4 dimer with a resolution of 2.5 Å was obtained from PDB bank (PDBID: 3 ODU) (Wu et al., 2010). We removed ligands and other non-essential components used for crystallizing and purification, including T4 lysozyme (T4L) inserted between transmembrane (TM) helices V and VI at the cytoplasmic side of the receptor, small isothiourea derivative (IT1t). In addition, crystal water molecules outside the receptor were also deleted. But the crystal water molecules inside the receptor were retained. All protein residues were set to the standard CHARMM protonation state under physiological pH. The receptor was inserted into a palmitoyl-oleoyl-phosphatidyl-choline (POPC) (Filizola et al., 2006) bilayer. Then, water molecules were added to the system, which was described by the TIP3P model. The whole system was neutralized with 0.15M NaCl by CHARMM-GUI (Lee et al., 2015). According to the tertiary structure information of the protein system, the two subunits were manually added with two disulfide bonds between $Cys28^{N\text{-}term}$ and $Cys274^{7.25}$, $Cys109^{3.25}$, and $Cys186^{5.50}$.

MD Simulation

All molecular dynamics simulations were performed by the sander module of AMBER 16 (Case et al., 2016). The MD trajectories were analyzed using the correlation analysis module of AMBER 16 and VMD, as well as some other specific trajectory analysis softwares. Ff14SB force field (Maier et al., 2015) was used for the receptor and the lipid14 force field (Dickson et al., 2014) was utilized for the POPC lipids. Twenty thousand step energy minimization was performed to eliminate bad contacts in the initial structures. After the minimization, the entire system was heated from 0 K to 310 K within 250 ps, then 5 ns NVT pre-equalization was performed at 310 K temperature. Finally, 300 ns and 1 us simulations were carried out using the NPT ensemble at 300 K and 1 bar for the apo dimer system and the dimer bound the PPI ligand, respectively. The cutoff distance of 10 Å was set for nonbonded interactions and the electrostatic interaction was computed by the particle mesh Ewald (PME) algorithm (Essmann et al., 1995). The SHAKE algorithm (Berendsen et al., 1984) was used to constrain all hydrogen-containing bonds. The time step was 2-fs and trajectories were saved at interval of 10 ps for further analysis.

Clustering Analysis

For the last 200 ns trajectory of the apo dimer system, clustering was carried out using the k-means algorithm (Han and Zhang, 2009; Li et al., 2014) embedded in the ptraj module of the AmberTools package in terms of RMSD of the backbone atoms of 136 residues of TM5-TM6/TM5-TM6 interface (Wu et al., 2010). Consequently, eight classes were obtained (vide in **Supplement 1**) and the center of each class was selected as a representative conformation for subsequent analysis.

FTMap Analysis

FTmap analysis (Kozakov et al., 2015) was performed in order to identify the druggable pocket in the dimer interface, using

FTMap computational map server. The server probes small molecule binding sites using CSM method (Dennis et al., 2002), which places molecular probes on a protein surface to identify the most favorable binding positions. The eight representative structures from the clustering above were individually computed using this server (www.ftmap.bu.edu). Pymol (Janson et al., 2016) was utilized to inspect visually the results.

Virtual Screening

A ligand set was constructed by a focused chemical compound collection (iPPI-lib) with a total of 51,232 ligand molecules, which was tuned to target PPIs. The PPI-specific database was provided by MTiOpenScreen (Labbé et al., 2015). First, the initial drug-like compounds containing 384,372 PubChem molecules was selected and collected. Then PPI-HitProfiler (Reynès et al., 2010) was used to select PPI-friendly compounds. Finally, these molecules were aggregated by Cluster Molecule Protocol (Accelrys Pipeline Pilot v8.5), resulting in 51,232 drug-like molecules in the final iPPI-lib. Approximately 4,000 molecules (including isomers) were obtained through preliminary screening of MTiOpenScreen, and further docking evaluations were performed using Autodock 4.2 (Morris et al., 2009). All docking input files were prepared by AutoDockTools 1.5.6 (Sanner, 1999) package, and Lattice files for active sites were generated by the AutoGrid 4.2. In order to cover the ligand-binding site, the box site was set to 75 Å × 75 Å × 75 Å with 0.375 Å spacing. The dockings with the flexible ligand and the rigid receptor were performed by AutoDock 4.2. To ensure the accuracy of the result, each ligand was done by 100 docking calculations separately, and 1,000,000 energy evaluations were carried out using Lamarck genetic algorithm for each docking calculation. We selected the docking pose with the lowest binding energy as the best binding mode for further analysis. The ROC (Metz, 1978) plot was used to assess virtual screening performance, which is a curve of true-positive rates versus false-positive rates. They could be calculated in terms of the following equations.

$$TPR = \frac{TP}{\left(TP + FN\right)} \tag{1}$$

$$FPR = \frac{FP}{\left(FP + TN\right)} \tag{2}$$

Where *TP* (true positive) and *FN* (false negative) refer to the number of active substances in positive and negative classes, respectively. *FP* (false positive) and *TN* (true negative) refer to the number of decoys in positive and negative classes, respectively. The AUC is the area under the receiver operating characteristic curve (Hanley and McNeil, 1982). The larger AUC value, the better the performance of the receptor in screening the active molecules from the decoys. For example, when the AUC value is 0.5, it represents random screening. When the AUC value is 1, the receptor has the strongest ability to screen the active molecules. The AUC value could reflect the affinity of the receptor to a class of active molecules in the ligand set. Therefore, it has been widely used to characterize the performance of virtual screening.

MMPBSA

As accepted, molecular Mechanics Poisson-Boltzmann surface area (MM/PBSA) (Sun et al., 2014; Sun et al., 2018; Wang et al., 2019; Weng et al., 2019) is a versatile method to calculate the binding free energy $\Delta G_{\text{binding}}$ between two molecules in terms of equation (3).

$$G_{\text{binding}} = G_{\text{complex}} - \left(G_{\text{receptor}} + G_{\text{ligand}}\right) \tag{3}$$

Herein, G_{complex}, G_{receptor} and G_{ligand} denote the free energies of the complex, receptor, and ligand, respectively, which could be calculated by MMPBSA.py.MPI algorithm (Miller III et al., 2012) of the SANDER module [vide equations (4)–(6)].

$$G = E_{\text{gas}} + G_{\text{sol}} - TS \tag{4}$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdw}} \tag{5}$$

$$G_{sol} = G_{\text{psolv}} + G_{\text{npsolv}} \tag{6}$$

The gas phase energy (E_{gas}) is calculated by the internal energy (E_{int}), the electrostatic interaction energy (E_{ele}) and van der Waals interaction energy (E_{vdw}) in equation (5). G_{sol} denotes the solvation energy, which consists of polar solvation energy (G_{psolv}) and the nonpolar solvation (G_{npsolv}) [vide equation (6)]. G_{psolv} could be obtained by solving the Poisson-Boltzmann equation while G_{npsolv} could be estimated by γ×SASA. Herein, γ uses 0.0072 kcal Å^{-2} value and SASA denotes the solvent-accessible area of the molecular. The dielectric constants are set to be 1 for the receptor interior and 80 for the external water. *T* represents absolute temperature and *S* is the total conformational entropy. Similar to many computational studies (Niu et al., 2017; Tu et al., 2018), the contribution of entropy is not considered in the calculation of free energy since we mainly concern with the relative change of the binding energy, rather than its absolute value.

Protein Structure Network

Protein structure network (PSN) (Kannan and Vishveshwara, 1999) could exhibit the structure of proteins as an interaction network. In PSN, residues are served as nodes. If the percentage of interaction [vide equation (7)] between the two nodes is greater than or equal to a given cutoff, the two nodes are connected to one edge.

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100 \tag{7}$$

In equation (7), I_{ij} represents the percentage of interaction between nodes *i* and *j*, and n_{ij} represents the pair number of side chain atoms within a given distance cut-off range (the default

cutoff is 4.5 Å). N_i and N_j are the normalization factors of residue *i* and *j*, respectively. Based on the network, we could gain insight into inter-residue communication, which play a vital role for proteins to execute their biological functions. Consequently, PSN has been successfully applied to study unfolding, stability and allosteric interaction (Brinda and Vishveshwara, 2005; Vishveshwara et al., 2009; Gao et al., 2016).

In addition, the shortest paths between pairs of nodes could be obtained through searching PSN by Dijkstra's algorithm (Dijkstra, 1959), which considers the PSN node inter-connectivities and residue correlated motions. The dynamic cross-correlation (DCC) (McCammon and Harvey, 1988) could be evaluated along an MD trajectory, in which DCC values (C_{ij}) are computed in terms of equation (8):

$$C_{ij} = \frac{\overline{\left(r_i(t) - \overline{r_i}\right)\left(r_j(t) - \overline{r_j}\right)}}{\sqrt{\left(r_i(t)^2 - \overline{r_i^2}\right)\left(r_j(t)^2 - \overline{r_j^2}\right)}} \quad (8)$$

i and *j* denotes atoms or residues, and r_i(t) and r_j(t) are the corresponding position vectors at time *t*. $\overline{r}$ means the ensemble average over a period time. DCC could characterize the extent of atom or residue movement correlations within a range from 1.0 to -1.0, where 1.0 indicates completely correlated displacements and -1.0 denotes completely anti-correlated displacements. Cross correlation analysis and PSN were performed using Wordom software (Seeber et al., 2011).

RESULTS AND DISCUSSION

Prediction of the Druggable Pocket in the Interface of the CXCR4 Dimer Based on Representative Conformations

The crystal structure is not completely equal to the functional conformation due to the flexibility of protein, which play a crucial role in the protein function. Thus, we first performed 300 ns MD simulation to obtain representative conformations for the apo dimer. **Figure 2** shows the root-mean-square-deviation (RMSD) of the backbone atoms with respect to its crystal structure for the dimer. It can be seen that the RMSD values present minor fluctuations after 100 ns. Thus, we used the k-means algorithm to cluster the last 200 ns trajectories, based on RMSD of the backbone atoms of 136 residues of the dimer interface (Wu et al., 2010). Consequently, eight classes of the conformations were obtained, as shown in Supplement 1. **Figure 3** shows the proportion of conformations for each class and populations of the *a* and *b* classes are significantly higher

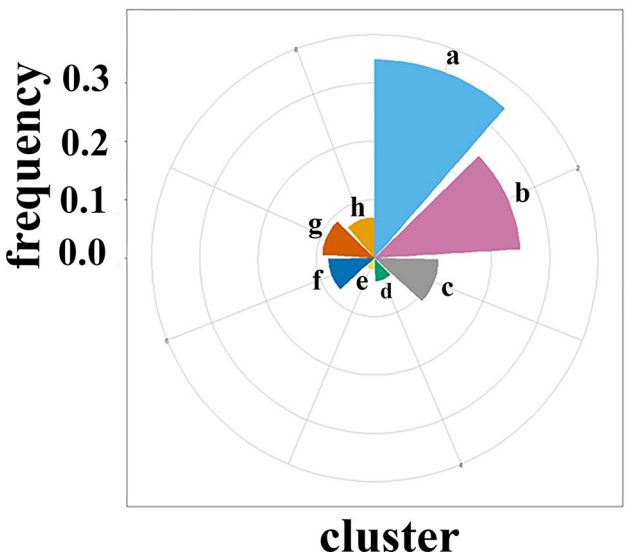


FIGURE 3 | The proportion of the eight representative conformations obtained by clustering in the last 200-ns trajectory for the apo CXCR4 dimer. One druggable pocket was identified by FTmap only for four types of conformations *a*, *b*, *d*, and *g*.

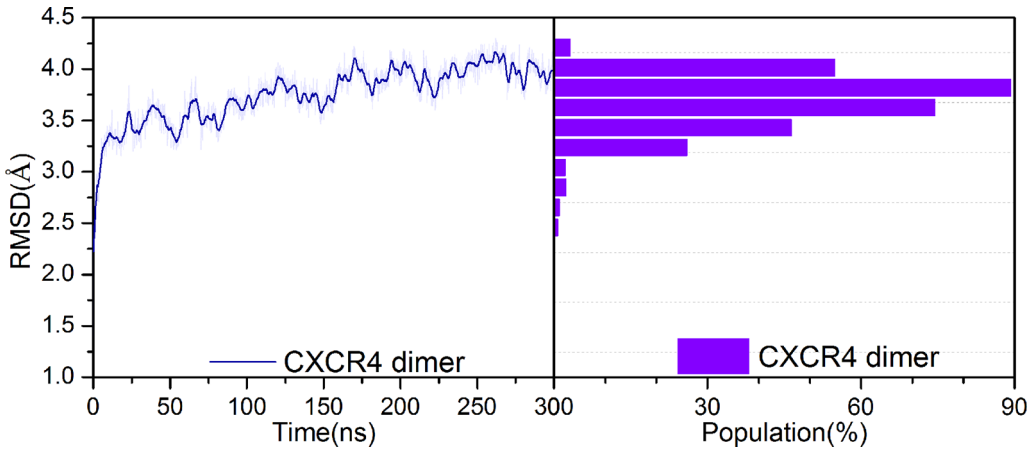


FIGURE 2 | Changes in RMSD values of backbone atoms for the apo CXCR4 dimer along with simulation time (left) and its distribution (right).

than the other classes. The center of each class was selected as representative conformation to probe its druggability with the aid of FTmap method. Only one druggable pocket was identified in the dimer interface for the classes *a*, *b*, *d*, and *g*, which account for 72% conformations of the last 200 ns trajectories, thus being highly representative. Furthermore, the druggable pockets in the dimer interfaces are highly similar for the four classes, which are mainly involved in Trp195$^{5.34}$, Val198$^{5.37}$, Phe199$^{5.38}$ of the subunit A, Val197$^{5.36}$, Gln200$^{5.39}$, Phe201$^{5.40}$, Ile259$^{6.55}$, Ser260$^{6.56}$, Ser263$^{6.59}$, and Leu267$^{6.63}$ of subunit B. It was revealed from the CXCR4 crystal structure that the residue Trp195$^{5.34}$, Val197$^{5.36}$, Val198$^{5.37}$, Phe201$^{5.40}$, and Leu267$^{6.63}$ play an important role in the dimerization of dimer (Wu et al., 2010). Thus, it can be assumed that a ligand targeting the pocket could significantly influence the dimerization of CXCR4.

Screening Potential Ligands to the Druggable Pocket in the Dimer Interface

The ligand set was constructed by a focused chemical compound collection (iPPI-lib) with a total of 51,232 ligand molecules, which were docked to the four representative conformations of the *a*, *b*, *d*, and *g* classes. The complex with the lowest binding energy was selected as the best binding mode for each of the four classes. Consequently, four small molecules (vide in **Figure 4**) were screened for the four representative conformations, which present best binding. The ligands 1, 2, 3, and 4 correspond to the conformations *a*, *b*, *d*, and *g*, respectively. **Table 1** lists some important physicochemical properties calculated by the SwissADME (Daina et al., 2017) for the four small molecules. It can be seen that their molecular weights are between 340 and 500. LogP values are between 3 and 5. LogS values are between -5 and -6. These properties are in line with those of the PPI drugs reported. Furthermore, the four molecules satisfy "Rule-of-Five" proposed by Lipinski (Lipinski et al., 1997; Lipinski, 2004), which indicates MW≤500, Log P≤5, N or O ≤10, NH or OH≤5, maybe potential drugs.

TABLE 1 | Properties of the four ligands targeting the PPI interface.

Ligands	MW (g/mol)[a]	LogP[b]	LogS[c]	TPSA(Å^2)[d]	Lipinski[e]
CBMicro_026776(1)	341.45	4.82	–5.64	20.31	Yes
CHEMBL2133598(2)	400.51	4.64	–5.82	49.41	
CHEMBL2136779(3)	434.92	3.90	–5.05	71.94	
CHEMBL1895118(4)	439.56	4.43	–5.66	77.21	

[a]*Molecular weight.*
[b]*Lipid-water partition coefficient.*
[c]*Aqueous solubility.*
[d]*Topological polar surface area.*
[e]*Rule-of-five.*

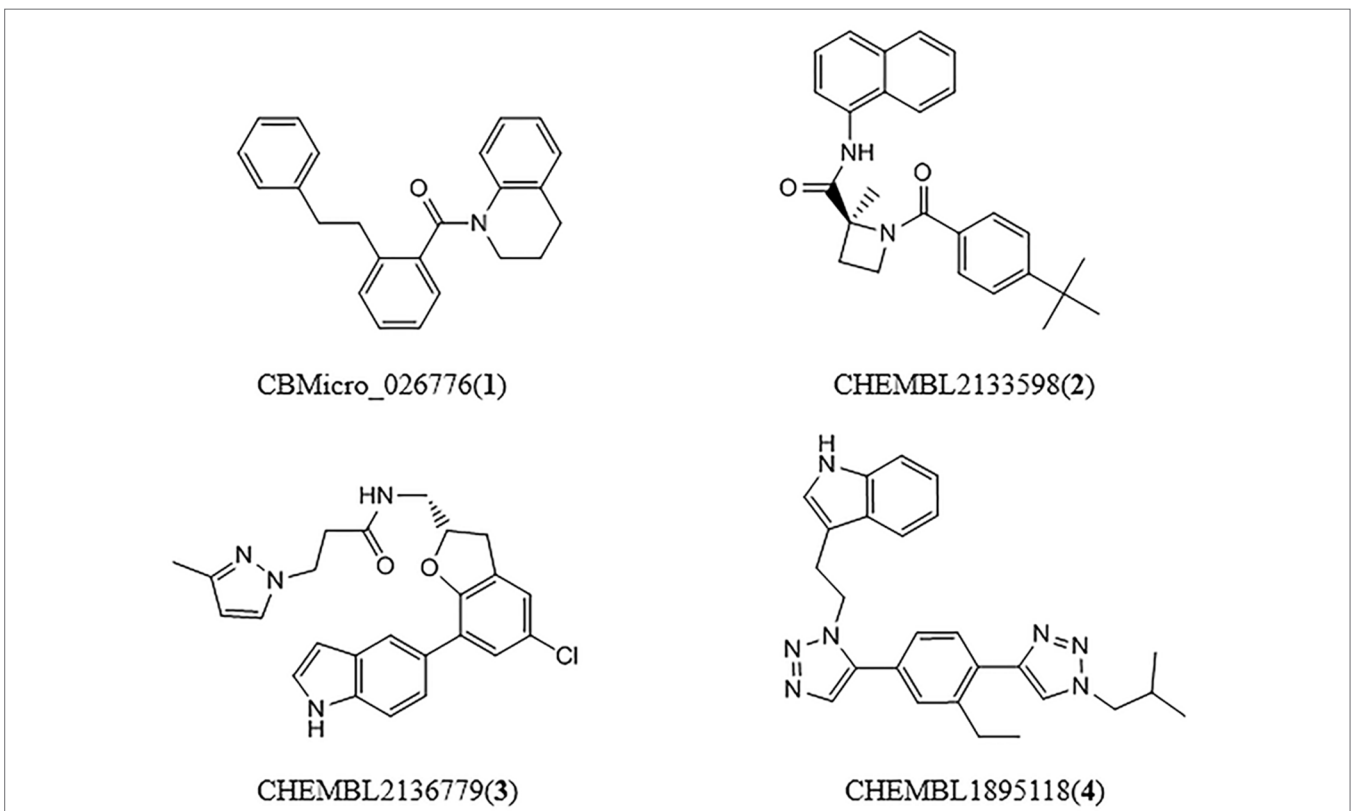


FIGURE 4 | Chemical structural formulas of four ligands (ligand 1 binds to conformational *a*, ligand 2 binds to conformational *b*, ligand 3 binds to conformational *d*, ligand 4 binds to conformational *g*).

Effect of the Four Ligands on the CXCR4 Dimerization

In order to probe the impact of the four ligands on the dimerization of CXCR4, we further performed 1us MD simulation for the four dimer conformations, the interfaces of which were docked by the individual ligand. The centroid distance and the contact area between the two subunits of the CXCR4 dimer were calculated based on the 1 us trajectory, as shown in **Figure 5**. It can be seen from **Figure 5** that only the ligand 1 targeting the conformation *a* reduces the centroid distance between the two subunits and increases their contact area, suggesting enhanced dimerization. However, an opposite trend is presented for the conformations *b* and *d*. For the conformation *g*, the two parameters change little. The observations indicate that the ligand 1 could enhance the dimerization of CXCR4 while the ligands 2 and 3 disfavor the dimerization. The ligand 4 only plays a negligible role in the dimerization. Since our objective is to search the PPI ligand enhancing the dimerization, we only focused on the ligand 1 in the following analysis.

Interaction Energy Between the Ligand 1 and the CXCR4 Dimer

In order to estimate the interaction strength between the dimer and the ligand, the binding free energy between them was calculated using the MM-PBSA method, based on the last 100 ns trajectory, as shown in **Table 2**. The $\Delta G_{binding}$ value is -46.77 kcal/mol and van der Waals interaction is main driving force, as judged from -46.18 kcal/mol of ΔE_{vdw}. In the other words, van der Waals interaction devotes main contributions for the interaction between the CXCR4 dimer and the ligand 1.

To identify important residues contributed to the ligand binding, we decomposed the binding free energy into the

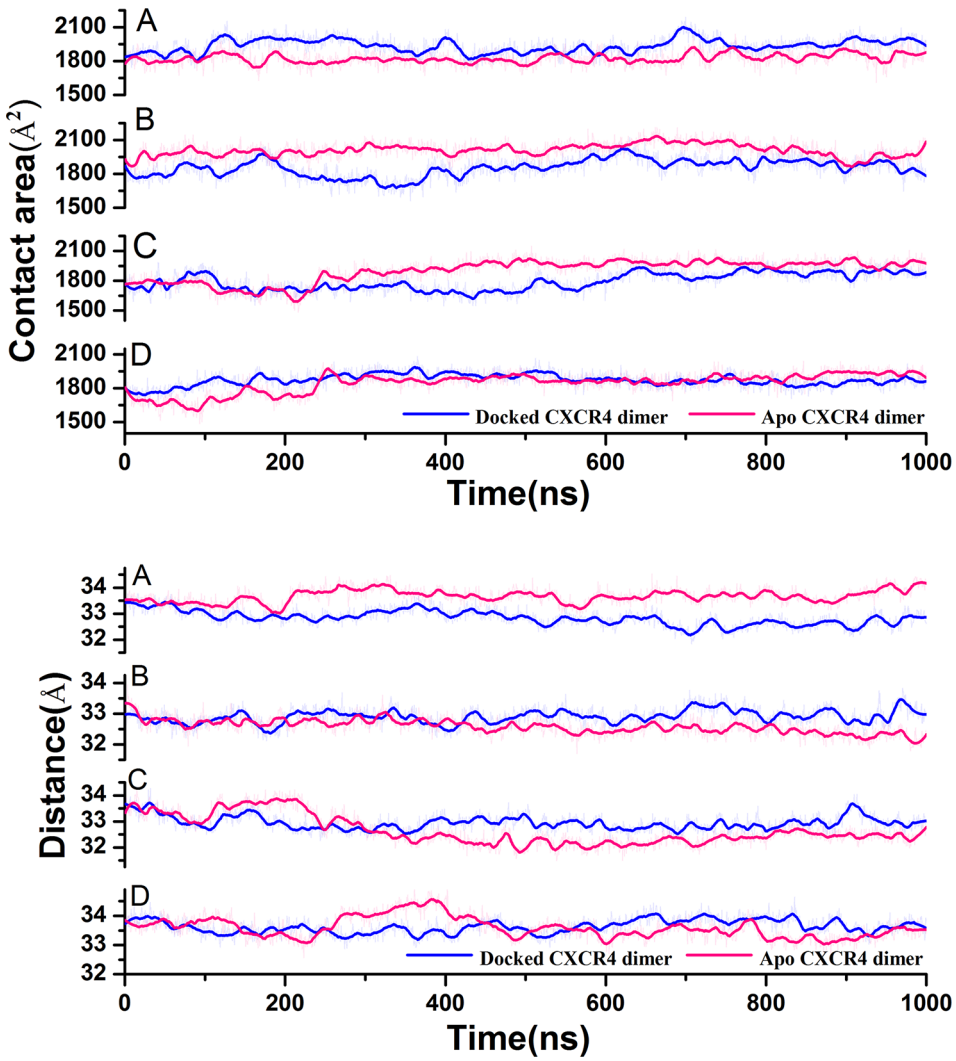


FIGURE 5 | Variations of the contact area and the centroid distance between the two protomers along with simulation time for the CXCR4 dimer with and without the ligand targeting the PPI interface **(A)** Corresponds to conformation *a*, **(B)** Corresponds to conformation *b*, **(C)** Corresponds to conformation *d*, **(D)** corresponds to conformation *g*).

TABLE 2 | The binding free energy (in kcal mol^{-1}) between the CXCR4 dimer and the ligand 1.

Components	Energy (kcal/mol)
ΔE_{vdw}[a]	–46.18 ± 2.86
ΔE_{ele}[b]	–3.28 ± 1.20
ΔE_{int}[c]	0.00 ± 0.00
ΔE_{gas}[d]	–49.46 ± 3.15
ΔG_{npsolv}[e]	–4.38 ± 0.20
ΔG_{psolv}[f]	7.07 ± 1.11
ΔG_{solv}[g]	2.69 ± 1.08
$\Delta G_{binding}$[h]	–46.77 ± 2.82

[a]*Non-bonded van der walls contribution from MM force field.*
[b]*Non-bonded electrostatic energy as calculated by the MM force field.*
[c]*Internal energy arising from bond, angle, and dihedral terms in the MM force field.*
[d]*Total gas phase energy.*
[e]*Nonpolar contribution to the solvation free energy.*
[f]*Polar contribution to the solvation free energy calculated.*
[g]*Solvation free energy.*
[h]*Binding energy.*
$\Delta E_{gas} = \Delta E_{ele} + \Delta E_{vdw} + \Delta E_{int}$, $\Delta G_{solv} = \Delta G_{npsolv} + \Delta G_{psolv}$, $\Delta G_{binding} = \Delta E_{gas} + \Delta G_{solv}$

corresponding residue. **Figure 6** shows residues with binding energy less than -1 kcal mol^{-1}, including residues Phe201$^{5.40}$, Ile204$^{5.43}$, and Phe264$^{6.60}$of the subunit A, residues Ile169$^{4.58}$, Pro170$^{4.59}$, Ile173$^{4.62}$, Val198$^{5.37}$, and Phe199$^{5.38}$of the subunit B. To identify important groups of the ligand contributed to the binding, we also calculated the interaction between the CXCR4 dimer and the ligand using protein–ligand interaction analysis software (PLIP) (Salentin et al., 2015). **Figure 7** shows the interaction mode between the CXCR4 dimer and the ligand 1 before the simulation and after that. Herein, the snapshot of the lowest energy in the last 100 ns MD trajectory was selected as representative conformation for calculating the binding mode after the simulation. It can be seen from **Figure 7** that the benzene ring of the small molecule devotes main contribution to the hydrophobic interaction between the ligand and the dimer, indicating the importance of the benzene group of the ligand in enhancing PPI. A comparison of the interaction modes in **Figure 7** indicates that TM5 mainly contributes to the binding before the simulation while TM4 also devotes to the binding besides TM5 after 1us simulation. Thus, it should be the interaction between TM4 and the ligand that drives the two subunits closer.

Effect of Ligand 1 on Drug Screening of Orthosteric Site

Since the existing drugs targeting CXCR4 are mainly antagonists, we, herein, focused the impact of the PPI ligand on the selectivity of the orthosteric site to the antagonists. One ligand set was constructed. The active molecules are extracted from the ZINC database (Irwin et al., 2012), GPCR-ligand database (Okuno et al., 2007) and PubChem database (Kim et al., 2015). The decoys stem from the DUD-E database (Mysinger et al., 2012). The ratio of decoys to the active molecules (N activity/N decoy) is 1:36. Consequently, the ligand set contains 1,480 small molecules (40 antagonists and 1,440 decoys). The ligand set was docked to the orthosteric site of the receptor (Wu et al., 2010; Venkatakrishnan et al., 2013; Qin et al., 2015), which consists of Lys25$^{1.19}$, Cys28$^{1.22}$, Arg30$^{1.24}$, Asp97$^{2.64}$, His113$^{3.29}$, Asp171$^{4.60}$, Cys186$^{5.25}$, Asp187$^{5.26}$, Asp262$^{6.58}$, Glu277$^{7.28}$, His281$^{7.32}$, and Glu288$^{7.32}$. **Figure 8** shows the receiver operating characteristic (ROC) curve and the area (AUC) under the ROC.

It can be seen that the screening performance of the subunit A increases with the simulation time (AUC=0.4661 at 100 ns, AUC=0.6711 at 500 ns, and AUC=0.7329 at 1,000 ns). However, there is little change for the subunit B (AUC=0.6698 at 100 ns, AUC=0.6612 at 500 ns, and AUC=0.6503 at 1 us). In addition, we

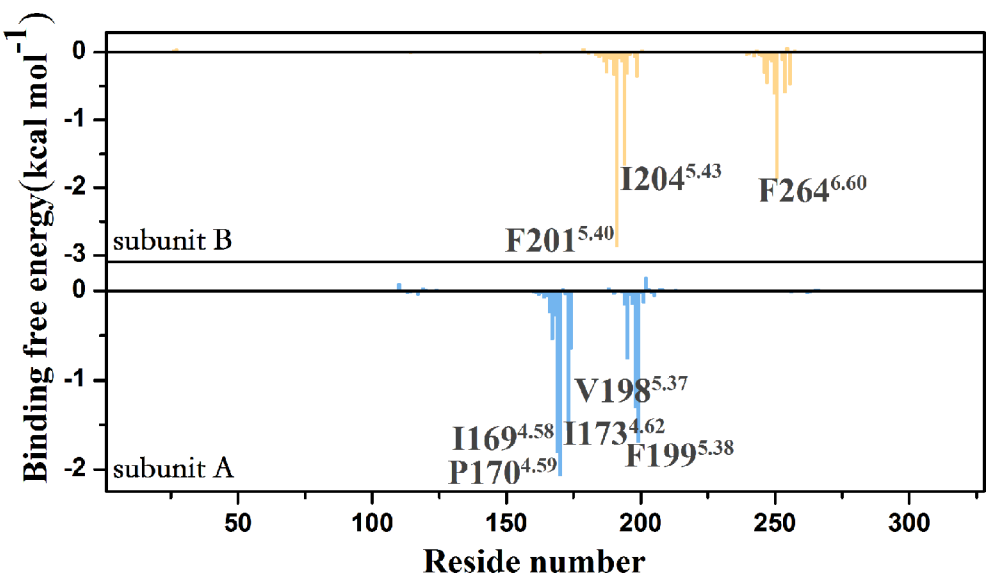


FIGURE 6 | Per-residue decomposition of the binding free energy for the CXCR4 dimer.

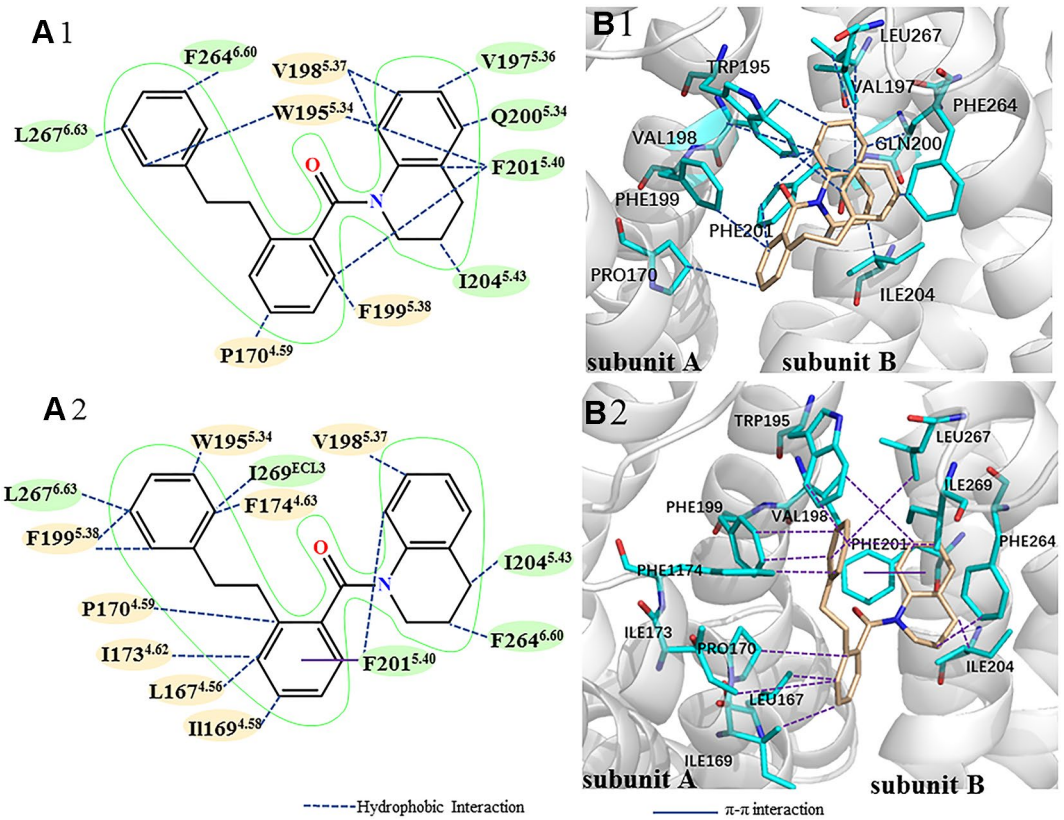


FIGURE 7 | The comparison of binding modes between the CXCR4 dimer and the PPI ligand before (Top) and after (Bottom) 1us simulation. **(A)** 2D diagram of binding mode between the CXCR4 dimer and the ligand (yellow represents residues from the subunit A, green represents residues from the subunit B). **(B)** 3D diagram of binding mode between the CXCR4 dimer and the ligand, blue for residues, orange for the ligand. Hydrophobic and π-π interactions are represented by different lines.

also compared the drug screening performance of the orthosteric site between the CXCR4 dimer bound the PPI ligand and one without the ligand, as shown in **Figure 9**. Similarly, the PPI ligand improves the screening performance of the subunit A but nearly has no effect on the subunit B, exhibiting asymmetric regulation. The asymmetric effect was also observed for the activation and the ligand binding for some GPCR dimers (Han et al., 2009; Liu et al., 2017).

In order to probe the origin of the asymmetric impact of the PPI molecule on the ligand binding of the orthosteric site for the two subunits, we calculated the pocket volumes of the orthosteric sites of CXCR4, as shown in **Figure 10**. It is clear that the PPI ligand significantly increases the volume of the orthosteric pocket for the subunit A but plays a minor role in the subunit B, which should contribute to the asymmetric screening.

The Allosteric Pathway for the Regulation Impact of the PPI Molecule on the Ligand Binding of the Orthosteric Site

In order to probe how the PPI molecule regulates the ligand binding of the orthosteric site of the receptor, we used the protein structure network to identify the allosteric pathway between the druggable pocket in the dimer interface and the orthosteric site of the two subunits. The residues consisted of the two types of the binding pockets are served as the starting and ending nodes in the PSN calculation, respectively, based on the last 100 ns equilibrium trajectory. **Table 3** lists the number of main pathways with frequency higher than 30%. Compared to the dimer without the PPI small molecule, the binding of the PPI molecule significantly increases the number of the pathway for the subunit A while there is little change for the subunit B. The observation suggests that the PPI ligand enhances the role of the interface in regulating the orthosteric site of the subunit A but only plays minor role for the subunit B. As a result, the volume of the orthosteric pocket is increased for the subunit A while the slight change is observed for that of the subunit B. In order to identify important residues in the allosteric regulation pathway, we searched the shortest pathway with the highest frequency between the PPI pocket and the orthosteric pocket for the subunit A. It can be seen from **Figure 11** that the pathway is composed of Trp195$^{5.34}$–Tyr190^{ECL2}–Val196$^{5.35}$–Gln200$^{5.39}$–Asp262$^{6.58}$–Cys28$^{N\text{-term}}$. As revealed above, Trp195$^{5.34}$ is an important residue contributed to the binding of the PPI ligand. Residue Tyr190^{ECL2} locates in ECL2, which was revealed to be switch for the ligand binding in the orthosteric site (Scarselli et al., 2007; Arkin et al., 2014).

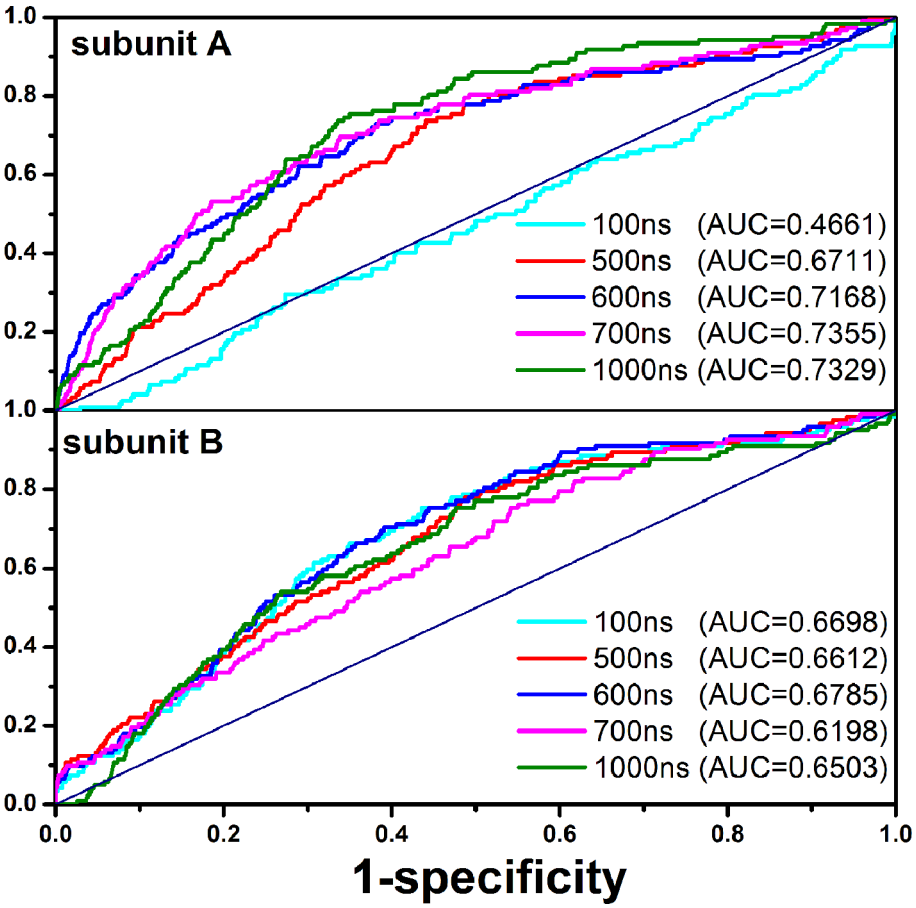


FIGURE 8 | Effects of the PPI ligand on the screening performance of the orthosteric site to the antagonists for the subunit A (top) and the subunit B (bottom).

Residue Gln200$^{5.39}$ plays a specific role in the dimerization of CXCR4 dimer (Altwaijry et al., 2017). The residues Cys274$^{7.25}$ and Cys28$^{N\text{-term}}$ bind closely through disulfide bonds, which play an important role in the formation of entrance to the ligand binding pocket at orthosteric site (Wu et al., 2010; Pawig et al., 2015). Residue Asp262$^{6.58}$ is an important residue for the binding of orthosteric site ligands (Wu et al., 2010; Qin et al., 2015). It can be seen that most of the residues composed of the allosteric pathway are associated with the ligand binding, which should be the reason why the PPI ligand significantly affect the ligand binding pockets, in turn influence its screening performance to the ligands. Although there is no report on the importance of the residue Val196$^{5.35}$ of this pathway, our observations suggest that the residue Val196$^{5.35}$ is also important for the ligand binding of the dimer and should be concerned by experiments. In addition, we also searched the shortest pathway with the highest frequency between the PPI pocket and the orthosteric pocket of the subunit B, as shown in **Supplement 2**. The pathway is composed of Ile269^{ECL3}–Phe264$^{6.60}$–Ile270^{ECL3}–Ile265$^{6.61}$–Glu277$^{7.28}$, only Glu277$^{7.28}$ of which was reported to be the pocket residue of the orthosteric site (Wu et al., 2010; Venkatakrishnan et al., 2013; Qin et al., 2015). Compared to the pathway of the subunit A, there are fewer residues involved in the ligand binding for that of the subunit B, which should contribute to the observation above that the PPI ligand plays a minor role in influencing the screening ability of the subunit B to the antagonists.

CONCLUSIONS

PPIs offer a rich source of novel drug targets. As the largest family of drug-targeted proteins, it was evidenced that GPCRs could form the dimers through the protein–protein interaction. Unfortunately, the drugs targeting the PPI interface of the GPCR dimers have not been explored so far. In the work, we utilized molecular dynamics simulation coupled with the virtual screening and the protein structure network to probe the druggability in the PPI interface of CXCR4 homodimer and its regulation mechanism on the receptor structure and the drug screening ability of the orthosteric site.

One druggable pocket is identified in the PPI interface. One small molecule is screened from the PPI drug-like small molecule dataset which could enhance the dimerization mainly through hydrophobic interactions between the benzene rings of the PPI molecule and TM4 of the receptor. The enhancement of PPI by the small molecule changes the screening performance of the

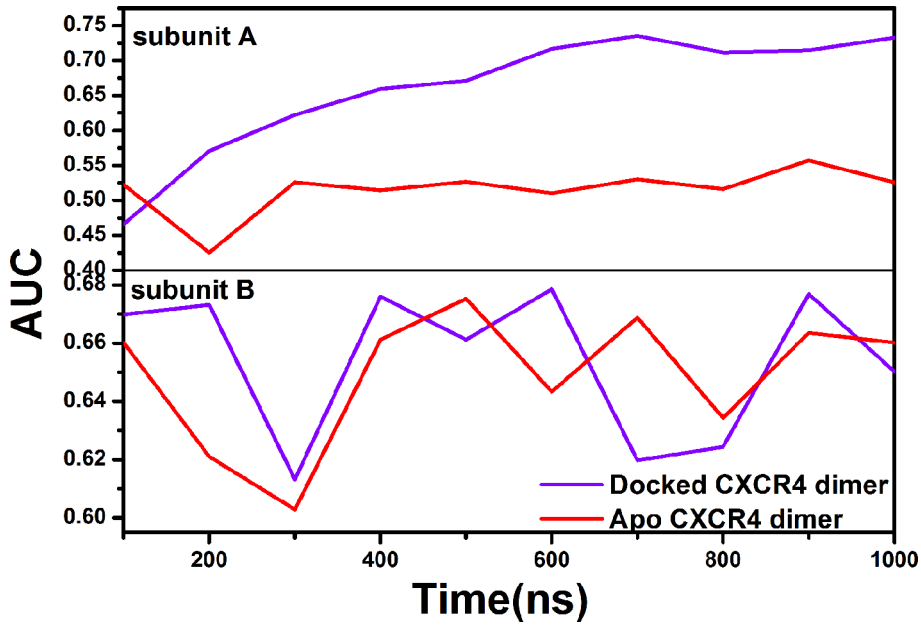


FIGURE 9 | Comparison of the screening performance for the orthosteric sites of the subunit A and the subunit B between the apo CXCR4 dimer (red lines) and the CXCR4 dimer bound by the PPI ligand (purple lines).

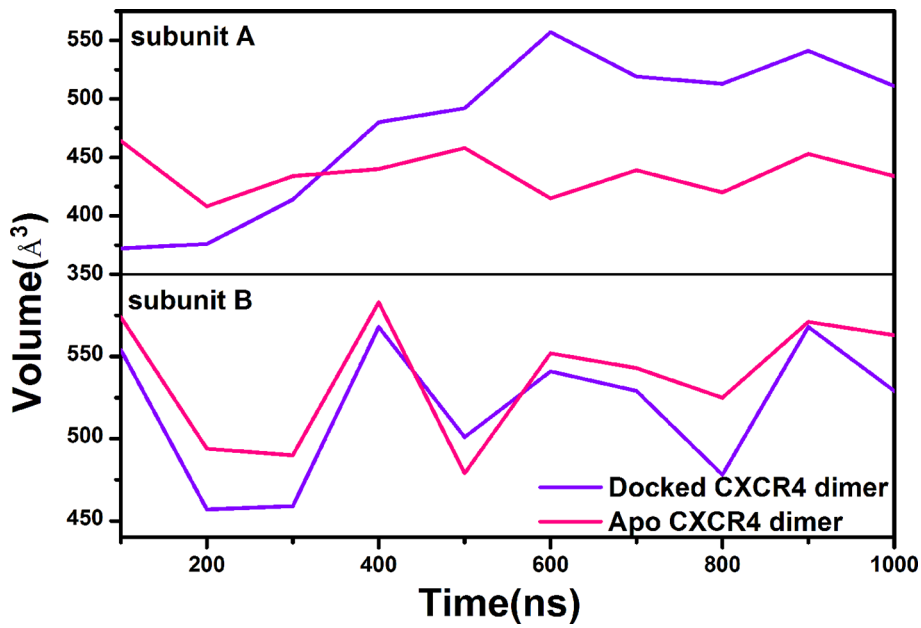


FIGURE 10 | Comparison of volumes for the orthosteric pockets of the subunit A and the subunit B between the apo CXCR4 dimer (red lines) and the CXCR4 dimer bound by the PPI ligand (purple lines).

two subunits to the antagonists targeting the orthosteric pocket. One subunit exhibits an enhanced screening performance to the antagonists while the minor change is observed for the other subunit, exhibiting an asymmetric cooperativity. The structural analysis indicates that the negative cooperativity should be attributed to the asymmetric change in the orthosteric pocket volumes induced by the binding of the PPI molecule, which leads to the significant increase in the pocket volume of the subunit A but only plays a minor role for the subunit B.

The results of PSN reveal that the number of the regulatory pathways from the PPI pocket to the orthosteric pocket is significantly increased for the subunit A while a minor change

TABLE 3 | The number of communication pathways between the binding pocket of the dimer interface and the orthosteric binding pocket (frequency above 30%), derived from the last 100 ns trajectory of the 1 us simulation.

	[a]Docked-dimer	[b]Free-dimer
Subunit A	78	56
Subunit B	79	77

[a]*The dimer docked by the PPI ligand.*
[b]*The apo dimer.*

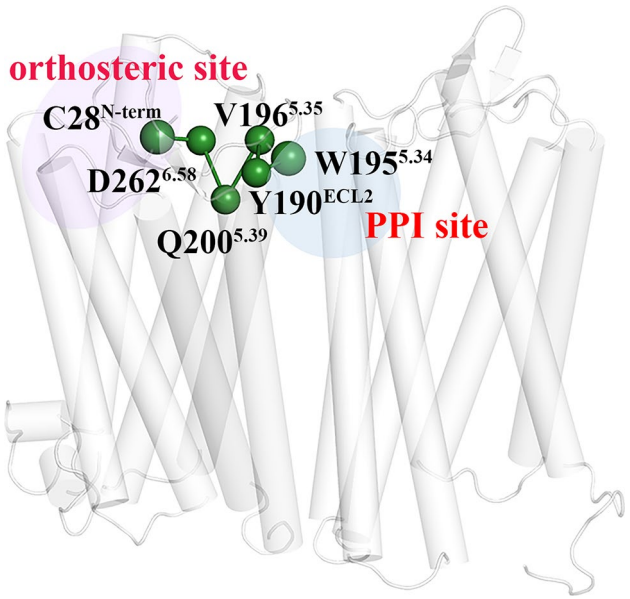


FIGURE 11 | The shortest pathway with the highest frequency between the druggable pocket on the dimer interface and the orthosteric site of subunit A. The blue region represents the PPI pocket. The purple region represents the orthosteric site of the subunit A. The green balls represent the residues composed of the pathway.

is observed for the subunit B, which should contribute to the asymmetric change of the binding pockets between the two subunits. In addition, one main regulatory pathway from the PPI binding site to the pocket of the subunit A is identified, revealing that the PPI ligand molecule allosterically regulates the structural change of the orthosteric pocket of the subunit A mainly through the pathway consisted of Trp195$^{5.34}$–Tyr190^{ECL2}–Val196$^{5.35}$–Gln200$^{5.39}$–Asp262$^{6.58}$–Cys28$^{N\text{-term}}$. These residues were revealed to significantly contribute to the dimerization and the ligand binding to the PPI interface and the orthosteric site. Consequently, the PPI small molecule could significantly regulate the dimerization and the screening ability of the orthosteric site to the ligands.

It is first time revealed the druggability of the GPCR dimer interface and its role in influencing the drug recognition ability of the orthosteric site. Since the antagonists of CXCR4 are used to treat CXCR4-related diseases like AIDS and some cancers, it is reasonable to assume that the PPI molecule identified from the work should enhance their drug efficacies. In addition, the strategy proposed by the work could be applied to probe the other GPCR PPI drugs.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

XP designed the experiments. YY and YG finished all the molecular dynamic simulations and analyzed the trajectories. LS organized the main manuscript text and plotted all the figures. XP, CL, and ML revised the manuscript. All authors contributed to the work.

FUNDING

This project is supported by the National Science Foundation of China (Grant No. 21573151), NSAF (Grand No. U1730127), and National Natural Science Youth Foundation of China (Grand No. 2170511).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2019.01310/full#supplementary-material

REFERENCES

Altwaijry, N. A., Baron, M., Wright, D. W., Coveney, P. V., and Townsend-Nicholson, A. (2017). An ensemble-based protocol for the computational prediction of helix–helix interactions in G Protein-coupled receptors using coarse-grained molecular dynamics. *J. Chem. Theroy. Comput.* 13 (5), 2254–2270. doi: 10.1021/acs.jctc.6b01246

Arkin, M. R., Tang, Y., and Wells, J. A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* 21 (9), 1102–1114. doi: 10.1016/j.chembiol.2014.09.001

Bai, Q., Zhang, Y., Li, X., Chen, W., Liu, H., and Yao, X. (2014). Computational study on the interaction between CCR5 and HIV-1 entry inhibitor maraviroc: insight from accelerated molecular dynamics simulation and free energy calculation. *Phys. Chem. Chem. Phys.* 16 (44), 24332–24338. doi: 10.1039/C4CP03331K

Berendsen, H. J., Postma, J.v., van Gunsteren, W. F., DiNola, A., and Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81 (8), 3684–3690. doi: 10.1063/1.448118

Brinda, K., and Vishveshwara, S. (2005). A network representation of protein structures: implications for protein stability. *Biophys. J.* 89 (6), 4159–4170. doi: 10.1529/biophysj.105.064485

Case, D. A., Betz, R., Cerutti, D., Cheatham, T. III, Darden, T., Duke, R., et al. (2016). *AMBER 2016 Reference Manual*. CA, USA: University of California: San Francisco, 1–923.

Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717. doi: 10.1038/srep42717

Damian, M., Pons, V., Renault, P., M'Kadmi, C., Delort, B., Hartmann, L., et al. (2018). GHSR-D2R heteromerization modulates dopamine signaling through

an effect on G protein conformation. *Proc. Natl. Acad. Sci. U. S. A.* 115 (17), 4501–4506. doi: 10.1073/pnas.1712725115

Dennis, S., Kortvelyesi, T., and Vajda, S. (2002). Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl. Acad. Sci. U. S. A.* 99 (7), 4290–4295. doi: 10.1073/pnas.062398499

Dickson, C. J., Madej, B. D., Skjevik, Å.A., Betz, R. M., Teigen, K., Gould, I. R., et al. (2014). Lipid14: the amber lipid force field. *J. Chem. Theroy. Comput.* 10 (2), 865–879. doi: 10.1021/ct4010307

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numer. Math.* 1 (1), 269–271. doi: 10.1007/BF01386390

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103 (19), 8577–8593. doi: 10.1063/1.470117

Ferré, S., Casadó, V., Devi, L. A., Filizola, M., Jockers, R., Lohse, M. J., et al. (2014). G protein–coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacol. Rev.* 66 (2), 413–434. doi: 10.1124/pr.113.008052

Filizola, M., Wang, S. X., and Weinstein, H. (2006). Dynamic models of G-protein coupled receptor dimers: indications of asymmetry in the rhodopsin dimer from molecular dynamics simulations in a POPC bilayer. *J. Comput. Aid. Mol. Des.* 20 (7-8), 405–416. doi: 10.1007/s10822-006-9053-3

Gao, N., Liang, T., Yuan, Y., Xiao, X., Zhao, Y., Guo, Y., et al. (2016). Exploring the mechanism of F282L mutation-caused constitutive activity of GPCR by a computational study. *Phys. Chem. Chem. Phys.* 18 (42), 29412–29422. doi: 10.1039/c6cp03710k

Gherbi, K., May, L. T., Baker, J. G., Briddon, S. J., and Hill, S. J. (2015). Negative cooperativity across β1-adrenoceptor homodimers provides insights into the nature of the secondary low-affinity CGP 12177 β1-adrenoceptor binding conformation. *FASEB J.* 29 (7), 2859–2871. doi: 10.1096/fj.14-265199

Han, L., Yang, Q., Liu, Z., Li, Y., and Wang, R. (2018). Development of a new benchmark for assessing the scoring functions applicable to protein–protein interactions. *Future Med. Chem.* 10 (13), 1555–1574. doi: 10.4155/fmc-2017-0261

Han, M., and Zhang, J. Z. (2009). Class I phospho-inositide-3-kinases (PI3Ks) isoform-specific inhibition study by the combination of docking and molecular dynamics simulation. *J. Chem. Inf. Model.* 50 (1), 136–145. doi: 10.1021/ci900175n

Han, Y., Moreira, I. S., Urizar, E., Weinstein, H., and Javitch, J. A. (2009). Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nat. Chem. Biol.* 5 (9), 688–695. doi: 10.1038/nchembio.199

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143 (1), 29–36. doi: 10.1148/radiology.143.1.7063747

Huang, J., Chen, S., Zhang, J. J., and Huang, X.-Y. (2013). Crystal structure of oligomeric β 1-adrenergic G protein–coupled receptors in ligand-free basal state. *Nat. Struct. Mol. Biol.* 20 (4), 419–425. doi: 10.1038/nsmb.2504

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 52 (7), 1757–1768. doi: 10.1021/ci3001277

Janson, G., Zhang, C., Prado, M. G., and Paiardini, A. (2016). PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* 33 (3), 444–446. doi: 10.1093/bioinformatics/btw638

Jiang, H., Deng, R., Yang, X., Shang, J., Lu, S., Zhao, Y., et al. (2017). Peptidomimetic inhibitors of APC–Asef interaction block colorectal cancer migration. *Nat. Chem. Biol.* 13 (9), 994–1001. doi: 10.1038/nchembio.2442

Jubb, H., Higueruelo, A. P., Winter, A., and Blundell, T. L. (2012). Structural biology and drug discovery for protein–protein interactions. *Trends Pharmacol. Sci.* 33 (5), 241–248. doi: 10.1016/j.tips.2012.03.006

Kannan, N., and Vishveshwara, S. (1999). Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* 292 (2), 441–464. doi: 10.1006/jmbi.1999.3058

Keskin, O., Tuncbag, N., and Gursoy, A. (2016). Predicting protein–protein interactions from the molecular to the proteome level. *Chem. Rev.* 116 (8), 4884–4909. doi: 10.1021/acs.chemrev.5b00683

Kim, S.-K., Chen, Y., Abrol, R., Goddard, W. A., and Guthrie, B. (2017). Activation mechanism of the G protein-coupled sweet receptor heterodimer with sweeteners and allosteric agonists. *Proc. Natl. Acad. Sci. U. S. A.* 114 (10), 2568–2573. doi: 10.1073/pnas.1700001114

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2015). PubChem substance and compound databases. *Nucleic Acids Res.* 44 (D1), D1202–D1213. doi: 10.1093/nar/gkv951

Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* 10 (5), 733–755. doi: 10.1038/nprot.2015.043

Kufareva, I., Stephens, B. S., Holden, L. G., Qin, L., Zhao, C., Kawamura, T., et al. (2014). Stoichiometry and geometry of the CXC chemokine receptor 4 complex with CXC ligand 12: Molecular modeling and experimental validation. *Proc. Natl. Acad. Sci. U. S. A.* 111 (50), E5363–E5372. doi: 10.1073/pnas.1417037111

Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Becot, J., Sperandio, O., et al. (2015). MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res.* 43 (W1), W448–W454. doi: 10.1093/nar/gkv306

Lao, J., He, H., Wang, X., Wang, Z., Song, Y., Yang, B., et al. (2017). Single-Molecule imaging demonstrates ligand regulation of the oligomeric status of CXCR4 in living cells. *J. Phys. Chem. B.* 121 (7), 1466–1474. doi: 10.1021/acs.jpcb.6b10969

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., et al. (2015). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theroy. Comput.* 12 (1), 405–413. doi: 10.1021/acs.jctc.5b00935

Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., et al. (2017). NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* 45 (W1), W162–W170. doi: 10.1093/nar/gkx449

Li, Y., Li, X., Ma, W., and Dong, Z. (2014). Conformational transition pathways of epidermal growth factor receptor kinase domain from multiple molecular dynamics simulations and Bayesian clustering. *J. Chem. Theroy. Comput.* 10 (8), 3503–3511. doi: 10.1021/ct500162b

Liang, T., Yuan, Y., Wang, R., Guo, Y., Li, M., Pu, X., et al. (2017). Structural features and ligand selectivity for 10 intermediates in the activation process of β2-Adrenergic receptor. *ACS Omega.* 2 (12), 8557–8567. doi: 10.1021/acsomega.7b01031

Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* 1 (4), 337–341. doi: 10.1016/j.ddtec.2004.11.007

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliver. Rev.* 23 (1-3), 3–25. doi: 10.1016/S0169-409X(96)00423-1

Liu, J., Zhang, Z., Moreno-Delgado, D., Dalton, J. A., Rovira, X., Trapero, A., et al. (2017). Allosteric control of an asymmetric transduction in a G protein-coupled receptor heterodimer. *Elife* 6, e26985. doi: 10.7554/eLife.26985

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theroy. Comput.* 11 (8), 3696–3713. doi: 10.1021/acs.jctc.5b00255

Maurice, P., Kamal, M., and Jockers, R. (2011). Asymmetry of GPCR oligomers supports their functional relevance. *Trends Pharmacol. Sci.* 32 (9), 514–520. doi: 10.1016/j.tips.2011.05.006

McCammon, J. A., and Harvey, S. C. (1988). *Dynamics of proteins and nucleic acids*. New York: Cambridge University Press. doi: 10.1063/1.2811564.

Metz, C. E. (1978). Basic principles of ROC analysis. *Semin. Nucl. Med.* 8 (4), 283–298. doi: 10.1016/S0001-2998(78)80014-2

Miao, Y., Caliman, A. D., and McCammon, J. A. (2015). Allosteric effects of sodium ion binding on activation of the m3 muscarinic g-protein-coupled receptor. *Biophys. J.* 108 (7), 1796–1806. doi: 10.1016/j.bpj.2015.03.003

Miller, B. R. III, McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA. py: an efficient program for end-state free energy calculations. *J. Chem. Theroy. Comput.* 8 (9), 3314–3321. doi: 10.1021/ct300418h

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30 (16), 2785–2791. doi: 10.1002/jcc.21256

Muñoz, L. M., Holgado, B. L., Martínez-A, C., Rodríguez-Frade, J. M., and Mellado, M. (2012). Chemokine receptor oligomerization: a further step

toward chemokine function. *Immunol. Lett.* 145 (1-2), 23–29. doi: 10.1016/j.imlet.2012.04.012

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (14), 6582–6594. doi: 10.1021/jm300687e

Navarro, G., Cordomí, A., Casadó-Anguera, V., Moreno, E., Cai, N.-S., Cortés, A., et al. (2018). Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase. *Nat. Commun.* 9 (1), 1242. doi: 10.1038/s41467-018-03522-3

Nguyen, H. H., Kim, M. B., Wilson, R. J., Butch, C. J., Kuo, K. M., Miller, E. J., et al. (2018). Design, synthesis, and pharmacological evaluation of second-generation tetrahydroisoquinoline-based CXCR4 antagonists with favorable ADME properties. *J. Med. Chem.* 61 (16), 7168–7188. doi: 10.1021/acs.jmedchem.8b00450

Niu, Y., Shi, D., Li, L., Guo, J., Liu, H., and Yao, X. (2017). Revealing inhibition difference between PFI-2 enantiomers against SETD7 by molecular dynamics simulations, binding free energy calculations and unbinding pathway analysis. *Sci. Rep.* 7, 46547. doi: 10.1038/srep46547

Villoutreix, O., B., M., Labbe, C., Lagorce, D., Laconde, G., and Sperandio, O. (2012). A leap into the chemical space of protein-protein interaction inhibitors. *Curr. Pharm. Design.* 18 (30), 4648–4667. doi: 10.2174/138161212802651571

Okuno, Y., Tamon, A., Yabuuchi, H., Niijima, S., Minowa, Y., Tonomura, K., et al. (2007). GLIDA: GPCR—ligand database for chemical genomics drug discovery—database and tools update. *Nucleic Acids Res.* 36 (suppl_1), D907–D912. doi: 10.1093/nar/gkm948

Pawig, L., Klasen, C., Weber, C., Bernhagen, J., and Noels, H. (2015). Diversity and inter-connections in the CXCR4 chemokine receptor/ligand family: molecular perspectives. *Front. Immunol.* 6, 429. doi: 10.3389/fimmu.2015.00429

Pediani, J. D., Ward, R. J., Marsango, S., and Milligan, G. (2018). Spatial intensity distribution analysis: studies of G protein-coupled receptor oligomerisation. *Trends Pharmacol. Sci.* 39 (2), 175–186. doi: 10.1016/j.tips.2017.09.001

Percherancier, Y., Berchiche, Y. A., Slight, I., Volkmer-Engert, R., Tamamura, H., Fujii, N., et al. (2005). Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo-and heterodimers. *J. Biol. Chem.* 280 (11), 9895–9903. doi: 10.1074/jbc.M411151200

Petersen, J., Wright, S. C., Rodríguez, D., Matricon, P., Lahav, N., Vromen, A., et al. (2017). Agonist-induced dimer dissociation as a macromolecular step in G protein-coupled receptor signaling. *Nat. Commun.* 8 (1), 226. doi: 10.1038/s41467-017-00253-9

Provasi, D., Boz, M. B., Johnston, J. M., and Filizola, M. (2015). Preferred supramolecular organization and dimer interfaces of opioid receptors from simulated self-association. *PloS Comput. Biol.* 11 (3), e1004148. doi: 10.1371/journal.pcbi.1004148

Qin, L., Kufareva, I., Holden, L. G., Wang, C., Zheng, Y., Zhao, C., et al. (2015). Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* 347 (6226), 1117–1122. doi: 10.1126/science.1261064

Reddy, K. K., Singh, P., and Singh, S. K. (2014). Blocking the interaction between HIV-1 integrase and human LEDGF/p75: mutational studies, virtual screening and molecular dynamics simulations. *Mol. Biosyst.* 10 (3), 526–536. doi: 10.1039/C3MB70418A

Reynès, C., Host, H., Camproux, A.-C., Laconde, G., Leroux, F., Mazars, A., et al. (2010). Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PloS Comput. Biol.* 6 (3), e1000695. doi: 10.1371/journal.pcbi.1000695

Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature* 459 (7245), 356–363. doi: 10.1038/nature08144

Sader, S., Anant, K., and Wu, C. (2018). To probe interaction of morphine and IBNtxA with 7TM and 6TM variants of the human μ-opioid receptor using all-atom molecular dynamics simulations with an explicit membrane. *Phys. Chem. Chem. Phys.* 20 (3), 1724–1741. doi: 10.1039/c7cp06745c

Sagar, A., Arif, E., Solanki, A. K., Srivastava, P., Janech, M. G., Kim, S.-H., et al. (2017). Targeting Neph1 and ZO-1 protein-protein interaction in podocytes prevents podocyte injury and preserves glomerular filtration function. *Sci. Rep.* 7 (1), 12047. doi: 10.1038/s41598-017-12134-8

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., and Schroeder, M. (2015). PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* 43 (W1), W443–W447. doi: 10.1093/nar/gkv315

Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 17 (1), 57–61. doi: 10.1016/S1093-3263(99)00019-4

Scarselli, M., Li, B., Kim, S.-K., and Wess, J. (2007). Multiple residues in the second extracellular loop are critical for M3 muscarinic acetylcholine receptor activation. *J. Biol. Chem.* 282 (10), 7385–7396. doi: 10.1074/jbc.M610394200

Seeber, M., Felline, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., et al. (2011). Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* 32 (6), 1183–1194. doi: 10.1002/jcc.21688

Shaik, M. M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., et al. (2019). Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* 565 (7739), 318–323. doi: 10.1038/s41586-018-0804-9

Sheng, C., Dong, G., Miao, Z., Zhang, W., and Wang, W. (2015). State-of-the-art strategies for targeting protein–protein interactions by small-molecule inhibitors. *Chem. Soc Rev.* 44 (22), 8238–8259. doi: 10.1039/C5CS00252D

Shin, W.-H., Christoffer, C. W., and Kihara, D. (2017). In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods* 131, 22–32. doi: 10.1016/j.ymeth.2017.08.006

Song, X., Chen, J., Zhao, M., Zhang, C., Yu, Y., Lonard, D. M., et al. (2016). Development of potent small-molecule inhibitors to drug the undruggable steroid receptor coactivator-3. *Proc. Natl. Acad. Sci. U. S. A.* 113 (18), 4970–4975. doi: 10.1073/pnas.1604274113

Stanley, N., Pardo, L., and De Fabritiis, G. (2016). The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* 6, 22639. doi: 10.1038/srep22639

Sun, H., Duan, L., Chen, F., Liu, H., Wang, Z., Pan, P., et al. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* 20 (21), 14450–14460. doi: 10.1039/c7cp07623a

Sun, H., Li, Y., Shen, M., Tian, S., Xu, L., Pan, P., et al. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 5. improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* 16 (40), 22035–22045. doi: 10.1039/c4cp03179b

Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., et al. (2019a). ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief. Bioinform.* 00 (00), 1–16. doi: 10.1093/bib/bby127

Tang, J., Fu, J., Wang, Y., Luo, Y., Yang, Q., Li, B., et al. (2019b). Simultaneous improvement in the precision, accuracy and robustness of label-free proteome quantification by optimizing Data manipulation chains. *Mol. Cell. Proteomics* 18 (8), 1683–1699. doi: 10.1074/mcp.RA118.001169

Tortorella, P., Laghezza, A., Durante, M., Gomez-Monterrey, I., Bertamino, A., Campiglia, P., et al. (2016). An effective virtual screening protocol to identify promising p53–MDM2 inhibitors. *J. Chem. Inf. Model.* 56 (6), 1216–1227. doi: 10.1021/acs.jcim.5b00747

Tu, J., Li, J. J., Song, L. T., Zhai, H. L., Wang, J., and Zhang, X. Y. (2018). Molecular modeling study on resistance of WT/D473H SMO to antagonists LDE-225 and LEQ-506. *Pharmacol. Res.* 129, 491–499. doi: 10.1016/j.phrs.2017.11.025

Van Hout, A., Klarenbeek, A., Bobkov, V., Doijen, J., Arimont, M., Zhao, C., et al. (2018). CXCR4-targeting nanobodies differentially inhibit CXCR4 function and HIV entry. *Biochem. Pharmacol.* 158, 402–412. doi: 10.1016/j.bcp.2018.10.015

Venkatakrishnan, A., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., and Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature* 494 (7436), 185–194. doi: 10.1038/nature11896

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6 (1), 83–90. doi: 10.1038/nmeth.1280

Vishveshwara, S., Ghosh, A., and Hansia, P. (2009). Intra and inter-molecular communications through protein structure network. *Curr. Protein Pept. Sci.* 10 (2), 146–160. doi: 10.2174/138920309787847590

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z., et al. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* 119 (16), 9478–9508. doi: 10.1021/acs.chemrev.9b00055

Wang, L., Yuan, Y., Chen, X., Chen, J., Guo, Y., Li, M., et al. (2018a). Probing the cooperative mechanism of the μ–δ opioid receptor heterodimer by multiscale simulation. *Phys. Chem. Chem. Phys.* 20 (47), 29969–29982. doi: 10.1039/C8CP06652C

Wang, Z., Kang, Y., Li, D., Sun, H., Dong, X., Yao, X., et al. (2018b). Benchmark study based on 2P2IDB to gain insights into the discovery of small-molecule PPI inhibitors. *J. Phys. Chem. B.* 122 (9), 2544–2555. doi: 10.1021/acs.jpcb.7b12658

Wells, J. A., and McClendon, C. L. (2007). Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* 450 (7172), 1001–1009. doi: 10.1038/nature06526

Weng, G., Wang, E., Chen, F., Sun, H., Wang, Z., and Hou, T. (2019). Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein–peptide complexes. *Phys. Chem. Chem. Phys.* 21 (19), 10135–10145. doi: 10.1039/c9cp01674k

Whorton, M. R., Bokoch, M. P., Rasmussen, S. G., Huang, B., Zare, R. N., Kobilka, B., et al. (2007). A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl. Acad. Sci. U. S. A.* 104 (18), 7682–7687. doi: 10.1073/pnas.0611448104

Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., et al. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330 (6007), 1066–1071. doi: 10.1126/science.1194396

Xue, L., Rovira, X., Scholler, P., Zhao, H., Liu, J., Pin, J.-P., et al. (2015). Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nat. Chem. Biol.* 11 (2), 134–140. doi: 10.1038/nchembio.1711

Xue, W., Yang, F., Wang, P., Zheng, G., Chen, Y., Yao, X., et al. (2018). What contributes to serotonin–norepinephrine reuptake inhibitors' dual-targeting mechanism? the key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.* 9 (5), 1128–1140. doi: 10.1021/acschemneuro.7b00490

Yuan, S., Hu, Z., Filipek, S., and Vogel, H. (2015). W2466. 48 opens a gate for a continuous intrinsic water pathway during activation of the adenosine A2A receptor. *Angew. Chem. Int. Edit.* 54 (2), 556–559. doi: 10.1002/anie.201409679

Yuan, S., Vogel, H., and Filipek, S. (2013). The role of water and sodium ions in the activation of the μ-Opioid receptor. *Angew. Chem. Int. Edit.* 52 (38), 10112–10115. doi: 10.1002/anie.201302244

Zhang, F., Yuan, Y., Xiang, M., Guo, Y.-Z., Li, M.-L., Liu, Y., et al. (2019). Molecular mechanism regarding allosteric modulation of ligand binding and the impact of mutations on dimerization for CCR5 homodimer. *J. Chem. Inf. Model.* 59 (5), 1965–1976. doi: 10.1021/acs.jcim.8b00850

Zhang, L., Yuan, Y., Ren, T., Guo, Y., Li, C., and Pu, X. (2018a). Shining light on molecular mechanism for odor-selectivity of CNT-immobilized olfactory receptor. *Sci. Rep.* 8 (1), 7824. doi: 10.1038/s41598-018-26105-0

Zhang, X., Yuan, Y., Wang, L., Guo, Y., Li, M., Li, C., et al. (2018b). Use multiscale simulation to explore the effects of the homodimerizations between different conformation states on the activation and allosteric pathway for the μ-opioid receptor. *Phys. Chem. Chem. Phys.* 20 (19), 13485–13496. doi: 10.1039/C8CP02016G

Zweemer, A. J., Toraskar, J., Heitman, L. H., and IJzerman, A. P. (2014). Bias in chemokine receptor signalling. *Trends Immunol.* 35 (6), 243–252. doi: 10.1016/j.it.2014.02.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shen, Yuan, Guo, Li, Li and Pu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fonc.2020.00121

Improved Prediction of Aqueous Solubility of Novel Compounds by Going Deeper With Deep Learning

Qiuji Cui[1†], *Shuai Lu*[1†], *Bingwei Ni*[1], *Xian Zeng*[2], *Ying Tan*[3], *Ya Dong Chen*[1*] *and Hongping Zhao*[1*]

[1] School of Science, China Pharmaceutical University, Nanjing, China, [2] Department of Biological Medicine, School of Pharmacy, Fudan University, Shanghai, China, [3] The State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, Shenzhen Technology and Engineering Laboratory for Personalized Cancer Diagnostics and Therapeutics, The Graduate School at Shenzhen, Shenzhen Kivita Innovative Drug Discovery Institute, Tsinghua University, Shenzhen, China

Aqueous solubility is an important physicochemical property of compounds in anti-cancer drug discovery. Artificial intelligence solubility prediction tools have scored impressive performances by employing regression, machine learning, and deep learning methods. The reported performances vary significantly partly because of the different datasets used. Solubility prediction on novel compounds needs to be improved, which may be achieved by going deeper with deep learning. We constructed deeper-net models of ∼20-layer modified ResNet convolutional neural network architecture, which were trained and tested with 9,943 compounds encoded by molecular fingerprints. Retrospectively tested by 62 recently-published novel compounds, one deeper-net model outperformed four established tools, shallow-net models, and four human experts. Deeper-net models also outperformed others in predicting the solubility values of a series of novel compounds newly-synthesized for anti-cancer drug discovery. Solubility prediction may be improved by going deeper with deep learning. Our deeper-net models are accessible at http://www.npbdb.net/solubility/index.jsp.

Keywords: aqueous solubility, deep learning, artificial intelligence, compounds, chemical, anti-cancer drug discovery

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Jiangning Song,
Monash University, Australia
Jing Tang,
Chongqing Medical University, China

***Correspondence:**
Ya Dong Chen
ydchen@cpu.edu.cn
Hongping Zhao
zhaohongping@cpu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Oncology

Received: *12 December 2019*
Accepted: *23 January 2020*
Published: *11 February 2020*

Citation:
Cui Q, Lu S, Ni B, Zeng X, Tan Y, Chen YD and Zhao H (2020) Improved Prediction of Aqueous Solubility of Novel Compounds by Going Deeper With Deep Learning. Front. Oncol. 10:121. doi: 10.3389/fonc.2020.00121

INTRODUCTION

Aqueous solubility is an important physicochemical property of compounds in anti-cancer drug discovery and development, impacting pharmacokinetic properties and formulations (1, 2). To facilitate solubility assessment, a number of artificial intelligence (AI) solubility prediction tools have been developed by employing regression and modeling (3, 4), machine learning (5–9), and deep learning (10–12) methods. These tools have scored impressive performances with high R^2 (e.g., 0.62–0.97) and low RMSE (e.g., 0.29–0.89) values (5, 13). However, the reported performances vary significantly, even among the same tools, partly because of the different datasets used. For instance, the reported R^2 and RMSE values of MOE software V2010.10 are 0.62 and 0.51 (8) and those in a 2014 publication are 0.27 and 1.05 (14). The reported R^2 and RMSE values of QikProp software V1.6, V2.1, and V3.2 are 0.9 and 0.8 (6), 0.95 and 0.63 (15), and 0.45 and 0.86 (8), respectively.

AI solubility prediction tools may be critically tested by newly-published novel compounds. Tested by 62 novel compounds published since November 2017 (Methods section), four established

tools MOE V2016.0802, QikProp QP18 and CIQP18, and AlogGPS V2.1 scored significantly lower R^2 (<0.2) and higher RMSE (0.814–1.162) values (Results section) than the typically-reported values (5, 6, 8, 14, 15). Our own-developed deep learning model of typically-employed shallow-net architecture (Methods section), trained and tested with 9,943 compounds, also scored lower R^2 (0.307) and higher RMSE (0.739) values (Results section). Hence, there is a need for improved solubility prediction particularly on novel compounds to promote oral anti-cancer drug development. In AI field, deep learning methods with distinguished learning capabilities (16) [which has been proved by prediction of CRISPR-Cpf1 guide RNA activity (17) and prediction of protein-ligand binding affinity (18)] are useful for this task, but their potential has yet to be fully realized.

The published deep learning solubility prediction models are primarily shallow-nets (3–7 layers) (10–12). Deep learning performances have been routinely enhanced by going deeper (adding more layers to shallow-nets) (19–21). Although performances can also be enhanced by going wider (22), it may be practically easier to develop deeper-nets by tapping into the well-established architectures that require fewer parameters (19–21). The depth of deeper-nets or the width of wider-nets is constrained by the limited number of compounds with experimental solubility data. The architecture with fewer parameters, convolutional neural networks (CNN), is therefore preferred. A question is whether the superior local-feature learning capability of CNN can adequately learn molecular features of compounds. To fit with the local-feature learning capability of CNN, compounds are better represented by substructure-encoded molecular fingerprints (23) instead of molecular descriptors used for solubility prediction by previously-developed deep learning models (10–12). Molecular fingerprints are vectors with individual components encoding specific sub-structures of molecules. Hence, the superior local-feature learning capability of CNN is expected to be useful for capturing the key sub-structural elements and their combinations contributing the solubility of molecules.

We constructed N-layer CNN models (N = 14, 20, and 26) using 9,943 compounds and based on a residual network (ResNet) architecture (20), which are significantly deeper than the previously-developed 3–7 layers shallow-net models (10–12). The solubility prediction capability of our deeper-net models was tested by retrospective prediction of the experimental solubility of 62 recently-published novel compounds beyond the training and testing compounds. These performances were compared with those of four established tools, shallow-net models and four human experts. Our deeper-net models and others were further tested by a real anti-cancer drug discovery project with a series of novel compounds newly-synthesized for discovering FLT3 inhibitors. These compounds were considered difficult for solubility estimation by medicinal chemistry experts, which are ideal for rigorous test of solubility prediction models. Our models are accessible at http://www.npbdb.net/solubility/index.jsp for supporting broader tests.

MATERIALS AND METHODS

Data Collection and Processing

A total of 10,166 compounds with experimental aqueous solubility value were collected from ChemIDplus database (24) and Pubmed (9, 25, 26) literature search up to November 2017. Another 62 recently-published novel compounds with experimental aqueous solubility value (**Supplementary Figure 1**, 6 representative compounds in **Figure 1**) were collected from PMC database (27–31) search using keyword combination of "novel", "new," and "solubility" and under the following criteria: published between November 2017 and May 2018, and solubility measured at room-temperature and around pH 7.0. For the 10,166 compounds, their SMILES strings (which encode sub-structures), InChIKeys (chemical structure identifiers) and aqueous solubility values were collected from the searched sources. For the 62 novel compounds, their structures were drawn from literature-reported structures by using ChemDraw 18.0 and then converted to the SMILES strings by using RDKit[1]. Solubility S values in different units (e.g., μg/mL, mg/mL, and mg/L) were converted to mol/L and transformed into logS (in logarithmic units) values. The SMILES strings were converted to canonical SMILES strings for consistency by using Open Babel (32). Duplicates were removed by InChIKeys comparisons. The canonical SMILES of the remaining non-redundant 9,943 compounds (**Supplementary Table 1**, the basic physical properties detailed in **Supplementary Table 2**) and the 62 novel compounds were converted into the Pubchem molecular fingerprints (which encode sub-structures by 881 bits) using PaDEL (33).

Established Tools and a Deep Learning Model of Typically-Employed Shallow-Net Architecture for Solubility Prediction

Solubility prediction performances were comparatively evaluated with respect to four established software tools [MOE V2016.0802[2], QikProp 2018-4 QP18 and CIQP18[3], and AlogGPS V2.1 based on an artificial neural network method (5)]. The deep learning model was developed based on a typically-employed shallow-net deep neural network (DNN) architecture for solubility prediction (11), which is a 4 hidden-layers DNN (**Supplementary Figure 2**) with the network architecture and parameter sets re-constructed based on the literature descriptions (11) with the following minor variations: the activation function was changed from SReLU to ReLU and the compounds were represented by pubchem molecular fingerprints instead of fp6 molecular fingerprints. The numbers of nodes of the hidden layers are 512, 1,024, 2,048, and 4,096. The parameters of L2 regularization and dropout regularization are 0.001 and 0.5. The 9,943 compounds were randomly divided into 90% training and 10% testing datasets for training the DNN model.

[1] http://www.rdkit.org/
[2] http://www.chemcomp.com/index.htm
[3] https://www.schrodinger.com/QikProp

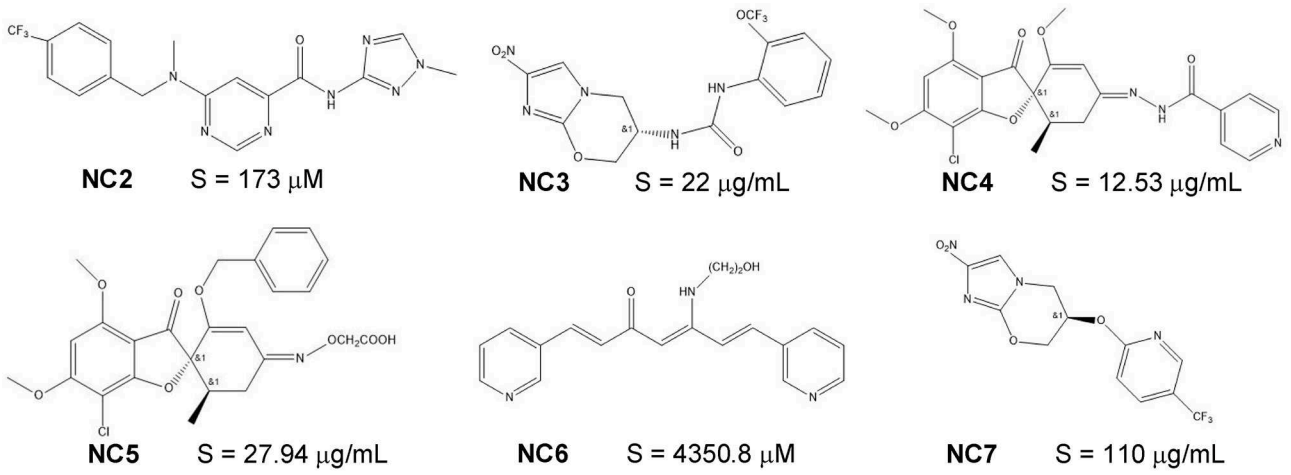


FIGURE 1 | The molecular structures and experimental solubility S values of six recently-published novel compounds.

Development of Deep learning Models of Deeper-Net Architecture for Solubility Prediction

The deeper-net models were based on the ResNet architecture (20) with the usual matrix forms of the ResNet layers, filters and feature maps replaced by vector forms. The numbers of layers N are 14, 20 (**Figure 2**), and 26 (**Supplementary Figure 3**) (N-1 CNN layers and 1 fully-connected layer). The vector forms were used because the inputs are 881-dimensional vectors (Pubchem fingerprints) instead of matrices of image pixel values. These CNN models were trained by the 10-fold cross validation method used for the development of two shallow-net deep learning solubility prediction models (10, 12). In the 10-fold cross validation method, the 9,943 compounds were randomly divided into 10 sets of approximately equal sizes, with each set used once as a testing dataset, and the remaining 9 sets as training dataset for training the CNN models. The CNN hyperparameters were optimized based on the overall performance of the 10 training/testing datasets. These hyperparameters include loss function, kernel sizes, number of filters, stride lengths, number of fully-connected hidden layers, number of neurons of the fully-connected layer, activation function, optimizer, learning rate, weight initialization, regularization, batch size, and epochs. Multiple activation functions (Sigmoid, ReLU, Softmax) were evaluated in both activation layers and the activation arguments of all forward layers. The weight initialization was uniform. L2 regularization was added by small amounts of L2 weight decay. A solubility value regression model was trained by least squares fit ($-R^2 = -(1 - (\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 / \sum_{i=0}^{n-1} (y_i - \bar{y})^2))$) between the predicted ($\hat{y}_i$) and experimental (y_i) solubility values of the n training compounds as the loss function of the output of our deeper-net models.

Performance Evaluation Metrics

The solubility prediction performances of the developed deep learning models were assessed by two metrics used in the evaluation of previously-developed shallow-net deep learning models (10, 12). One is the R^2 value, where R is the Pearson correlation coefficient defined by:

$$R^2 = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2}$$

The second is the root mean squared error RMSE defined by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{n}}$$

where $\hat{y}_i$ is the predicted and y_i is experimental solubility values of the training compounds.

In statistics, R^2, the coefficient of determination, is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It is a statistical measure used in a regression model to indicate that how well the model fits the data. Theoretically, it denotes a goodness-of-fit indicator that can vary from $-\infty$ to 1. The closer the R^2 value is to 1, the better the model fits the data, and vice versa. The other metric, RMSE, is the square root of the average of squared errors. It is a statistical measure of the differences between the values predicted by a model and the true values. RMSE is always non-negative, and the value closer to 0 indicates the better fit to the data.

Chemical Synthesis and Experimental Aqueous Solubility Determination

In one of our drugs discovering projects toward antitumor therapeutics, a series of novel FLT3 inhibitors were designed and synthesized using the structure-based drug design methods. The aqueous solubilities (pH = 7) of these compounds were measured using the modified shake flask method and RP-HPLC (34, 35). Each compound was added into a 1.5 mL Eppendorf tube containing Milli-Q water (1 mL) to form the precipitates at

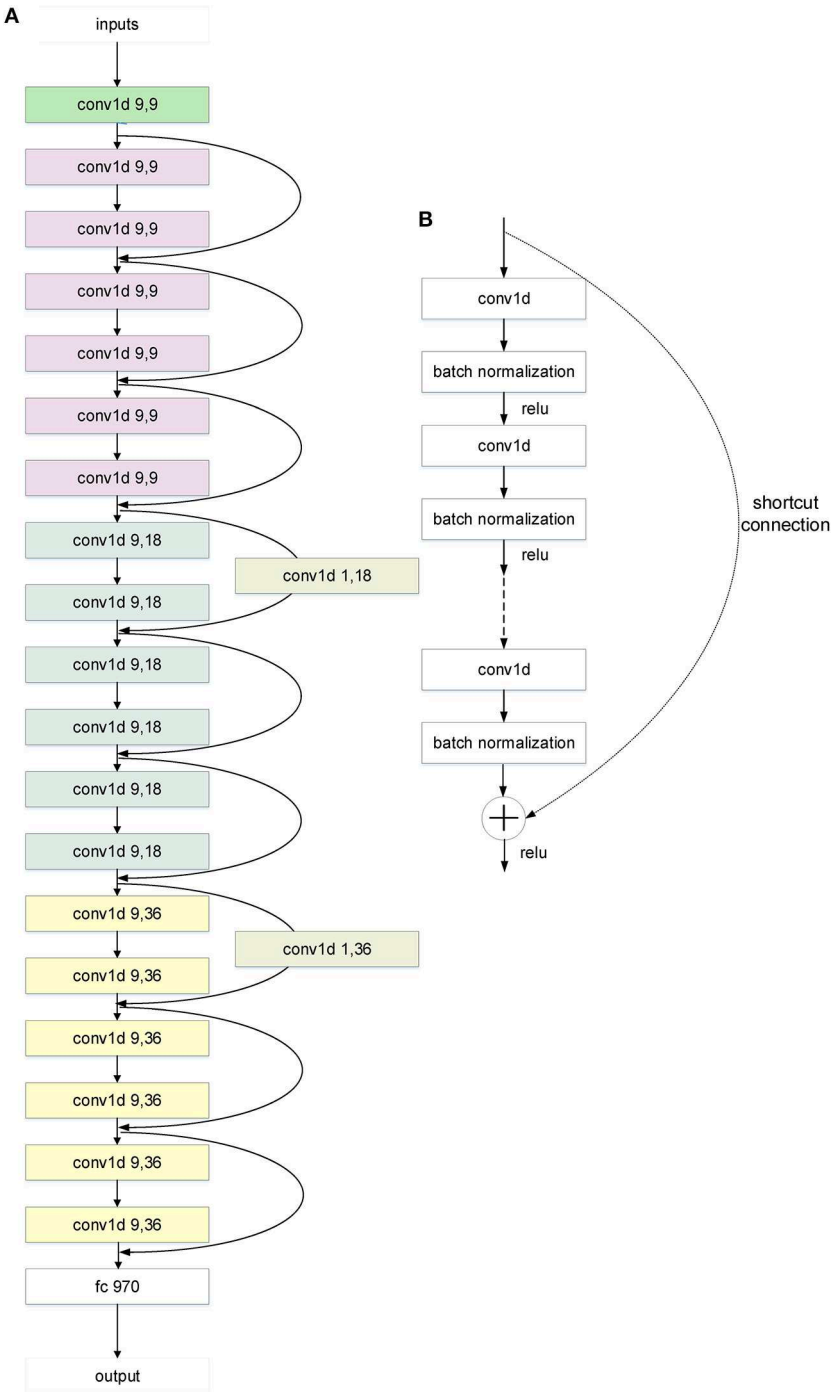


FIGURE 2 | The architecture of the 20-layer CNN ResNet-like deep learning model. **(A)** A CNN ResNet-like deep learning model with 20 parameter layers. The "conv1d x,y" is a 1D convolution layer with x kernel sizes and y filters. And the curvy arrows are the shortcut connections. The shortcut connection with a parameter layer increases dimensions. The different color means different layer class in the architecture. "Green" means the first layer, "white" means the last layer, "gray" means the parameter layer of the shortcut connection, and the others mean the residual layers. The color change of the residual layers from purple to blue to yellow indicates the tensor dimension change from 9 to 18 to 36. **(B)** The shortcut connection in the architecture of CNN ResNet-like deep learning model. Shortcut connections simply perform identity mapping by skipping one or more layers (20). Their outputs are added to the outputs of the stacked layers without extra parameter and computational complexity.

25°C. Then the mixture was subjected to a solubility-equilibrium stage. The tube was shook at 300 rpm at 25°C for 24 h. The precipitate was separated by centrifugation at 23,000 g for 20 min. Subsequently, 0.25 mL of supernatant was transferred into a 1 mL Eppendorf tube, and it was centrifuged again with the same settings used above. The supernatant was then used for HPLC analysis. An Agilent 1260 Infinity LC system (Agilent Technologies, Inc., Santa Clara, California) was used. For HPLC conditions, a ZORBAX SB-C18 column (5 μM, 4.6 × 150 mm; Agilent), a flow rate of 0.8 mL/min for mobile phase, a UV wavelength of 250 nM and a column temperature of 30°C were used. The sample was injected automatically by a mechanical arm and separated by a constant mixture of methanol/PBS (pH 5.6), 90:10. For each compound, a standard curve consisting of four concentrations was established. The synthetic methods of all but compound SC5 and SC6 have been published in literatures (36–39). The synthetic methods of SC5 and SC6 are described in **Supplementary Method 1**.

RESULTS

The Training of the Deeper-Net Models and Solubility Prediction Performance Evaluation

Using 9,943 compounds and 10-fold cross validation method, three deeper-net models of 14-, 20-, and 26-layer were developed. The ranges and the optimal hyperparameter values for the 20-layer model (which is the top performing model based on the loss function R^2 values) are given in **Supplementary Table 3**. The 10-fold cross validation performances of the 14-, 20-, and 26-layer models are R^2 = 0.72–0.78, 0.74–0.79, and 0.72–0.79, and RMSE = 0.988–1.144, 1.006–1.112, 1.015–1.151, respectively (detailed in **Supplementary Table 4**). In spite of different depths, these models performed similarly well, possibly because the superior predictive capability of these deeper-net models cannot be fully tested by 1-fold (1/10) testing datasets. The test by novel compounds may be better for probing the predictive capabilities. The reported 10-fold cross validation performances of the two previously-developed shallow-net models are R^2 = 0.86–0.92 and 0.90–0.92, and RMSE = 0.58–0.79 and 0.45–0.50, respectively (10, 12), which are substantially better than those of our deeper-net models. It is noted that our datasets (testing 994 compounds, training 8,949 compounds) are significantly larger than those of the two previously-developed shallow-net models (testing 102–287 and 129–154 compounds, training 923–2,586 and 1,161–1,537 compounds, respectively) (10, 12). Caution is needed in a direct comparison of the performance statistics of these models. The significantly more diverse testing datasets may partly contribute to the lower performance statistics. But the more diverse training datasets likely lead to more robust prediction capability than the less diverse training datasets. Because of the inaccessibility of the previously-published shallow-net models, it is impossible to test these models on a common set of diverse compounds. Therefore, these models were tested on the 62 newly-published novel compounds and a series of novel compounds from our anti-cancer drug discovery project with solubility measured for the first time in this work.

Prediction of the Solubility Values of Literature-Reported Novel Compounds by the Deeper-Net Models in Comparison With the Established Tools and Shallow-Net Models

The solubility prediction capability of our deeper-net models was tested by the 62 newly-published novel compounds. We also trained 1-layer DNN model, 6-layer DNN model, and 8-layer ResNet-like model as our shallow-net models. The testing results of these models are included in **Table 1**, and the predicted logS values of these models with respect to experimental logS values are in **Supplementary Table 5**. Based on the R^2 and RMSE values, the 20-layer deeper-net model (R^2 = 0.412, RMSE = 0.681) performed substantially better than all the other models including the four established tools and the shallow-net models (R^2 in the range of <0.2 to 0.307, RMSE = 0.739–0.982). The R^2 and RMSE values of four established tools, shallow-net and deeper-net deep learning models were evaluated by the bootstrap sampling method. The mean, standard deviation and 95% confidence interval of R^2 and RMSE values for 10,000 bootstrap samples of 62 recently-published novel compounds were detailed in **Supplementary Table 6**. Judged by the percent of predicted logS

TABLE 1 | Performance on the logS prediction of 62 recently-published novel compounds[a].

Model	R^2	RMSE	PCT-10-fold[b] (%)
Established tools			
MOE V2016.0802	<0.2	0.908	74.2
QikProp 2018-4 QP18	<0.2	0.926	69.4
QikProp 2018-4 ClQP18	<0.2	1.162	54.8
AlogGPS V2.1	0.160	0.814	77.4
Shallow-net deep learning model of a typically-employed architecture for solubility prediction			
4-layer DNN model	0.307	0.739	80.7
Shallow-net deep learning models developed in this work			
1-layer DNN model	0.086	0.849	72.6
6-layer DNN model	0.264	0.762	79.0
8-layer ResNet-like model	<0.2	0.982	66.1
Deeper-net deep learning models developed in this work			
14-layer ResNet-like model	0.133	0.827	74.2
20-layer ResNet-like model	**0.412**	**0.681**	**82.3**
26-layer ResNet-like model	0.075	0.854	77.4

[a] *The performance of the established tools, and the shallow-net and deeper-net deep learning models in the prediction of experimental logS values of 62 recently-published novel compounds. The best performance values are in bold font.*
[b] *Percent of predicted logS value within 10-fold of experimental value.*

TABLE 2 | Performance on the solubility category prediction[a].

Human expert or established tool	Percent of 62 compounds with correct classification (%)	Deep learning model	Percent of 62 compounds with correct classification (%)
Expert 1	6.5	4-layer DNN model	79.0
Expert 2	8.1	1-layer DNN model	79.0
Expert 3	11.3	6-layer DNN model	82.3
Expert 4	74.2	8-layer ResNet-like model	80.7
MOE V2016.0802	91.9	14-layer ResNet-like model	87.1
QikProp 2018-4 QP18	85.5	20-layer ResNet-like model	85.5
QikProp 2018-4 CIQP18	87.1	26-layer ResNet-like model	83.9
AlogGPS V2.1	82.3		

[a] *The performance of human experts, the established tools, and the shallow-net and deeper-net deep learning models in the prediction of solubility category of 62 recently-published novel compounds. The solubility categories are practically insoluble or insoluble (<0.1 g/L), slightly soluble (0.1–10 g/L), soluble (10–100 g/L), and freely soluble (>100 g/L).*

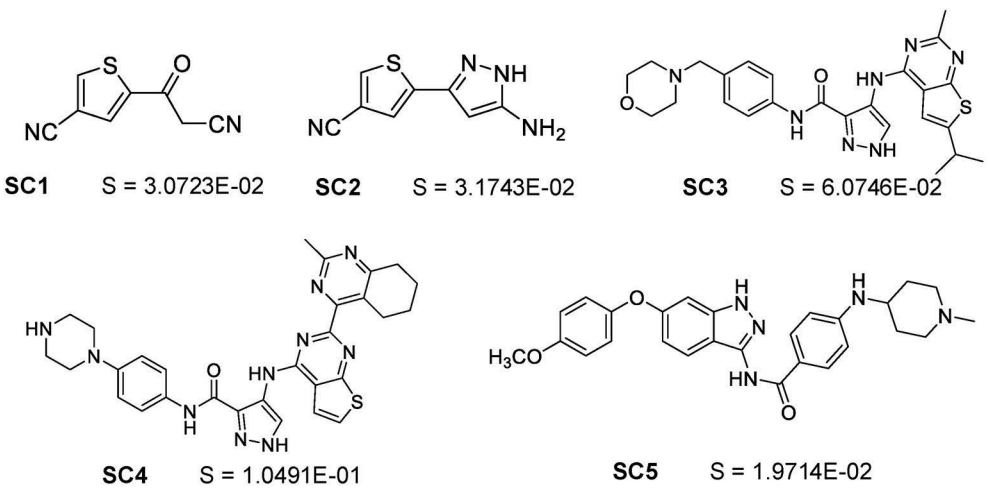


FIGURE 3 | The molecular structures and experimental solubility S values (in mg/mL) of the five synthetic novel compounds for a drug discovery project with solubility values measured for the first time by this work.

values within 10-fold of experimental value, all but one model achieved high performances (66.1%), suggesting the usefulness of both established tools and deep learning models for accessing solubility categories. Nonetheless, the 20-layer deeper-net model substantially outperforms all other models. These suggested that going deeper with deep learning at appropriate depth may give rise to significantly improved solubility prediction on novel compounds. The lower R^2 and RMSE values of the 26-layer model (R^2 = 0.075, RMSE = 0.854) over the 20-layer model indicated signs of overfitting in going further deeper beyond ∼20-layer.

Comparison With Human Experts in Coarse-Grained Classification of the Solubility Categories of the Literature-Reported Novel Compounds

Four human experts in medicinal chemistry were selected from the China Pharmaceutical University using the criterion of a recent machine vs. human comparative solubility prediction study (9), i.e., a human expert is someone with medicinal chemistry expertise working or studying in a university. These four experts include one assistant professor and three PhD students. They were tasked to conduct coarse-grained classification of the aqueous solubility of the 62 novel compounds at room temperature into one of the following categories: practically insoluble or insoluble (<0.1 g/L), slightly soluble (0.1∼10 g/L), soluble (10∼100 g/L), and freely soluble (>100 g/L). The classification performance of these four experts together with those of the established tools, and shallow and deeper-net models are in **Table 2**. All tools and models achieved high classification accuracies of 79.0–91.9%, which significantly outperformed the human experts (6.5–74.2%). These indicated the more superior capability of both established tools and deep learning models over human experts in coarse-grained classification of the solubility categories on novel compounds. However, no definite conclusion could be deduced on which was better between the established tools and the deep learning models. No improving trend was found with the increasing of the deep learning models' depth.

It seemed that the coarse-grained classification method was not discriminative enough to differentiate the capabilities of the established tools and deep learning models as revealed by the more quantitatively-precise evaluations of R^2 and RMSE values.

Solubility Prediction of a Series of Novel Compounds From a Real Anti-cancer Drug Discovery Project

A series of 17 novel compounds were synthesized by using the method described in **Supplementary Method 1** and the published literatures (36–39) for discovering FLT3 inhibitors. These compounds are structurally novel based on SciFinder search. They are difficult for solubility estimation based on our surveys with medicinal chemistry experts. The solubility values of these 17 compounds (**Supplementary Figure 4**) were experimentally measured using the method described in the Methods section. We were unable to determine the exact solubility values for 12 compounds because they are insoluble below 1.0000E-2 mg/mL in neutral water. Hence, only the remaining five compounds (**Figure 3**) with exact experimental solubility values were used for testing our deeper-net models and other models. Partly because of the novelty and low number of compounds, the R^2 values of all models are well below statistically meaningful values. Hence only the RMSE values and the percent of predicted logS values within 10-fold of experimental value were used for performance evaluation (**Table 3**). Judged by the RMSE values, the deeper-net models substantially outperformed all other models, with the 26-layer model as the best one in spite of minor level of overfitting. This further indicated the advantage of going deeper for improved solubility prediction. Judged by the percent of predicted logS values within 10-fold of experimental value, the majority of the models (including 14- and 20-layer deeper-net models) achieved equally good performances (60%) with the 26-layer model as the best one (80%). This again showed that both the established tools and deep learning models are useful for rough estimation of the solubility values of novel compounds.

TABLE 3 | Performance on the logS prediction of 5 novel compounds[a].

Model	RMSE	PCT-10fold[b] (%)
Established tools		
MOE V2016.0802	2.293	<20
QikProp 2018-4 QP18	2.717	20
QikProp 2018-4 ClQP18	2.308	20
AlogGPS V2.1	1.073	60
Shallow-net deep learning model of a typically-employed architecture for solubility prediction		
4-layer DNN model	1.325	60
Shallow-net deep learning models developed in this work		
1-layer DNN model	1.502	60
6-layer DNN model	1.494	40
8-layer ResNet-like model	1.646	60
Deeper-net deep learning models developed in this work		
14-layer ResNet-like model	0.982	60
20-layer ResNet-like model	0.811	60
26-layer ResNet-like model	**0.689**	**80**

[a] *The performance of the established tools, and the shallow-net and deeper-net deep learning models in the prediction of experimental logS values of 5 novel compounds (quantitative values measured in this work). The best performance value is in bold font.*
[b] *Percent of predicted logS value within 10-fold of experimental value.*

DISCUSSIONS

Like successful applications of deep learning methods in other fields (19–21), the superior learning capability of deeper-net models may be exploited to improve solubility prediction of novel compounds, including those compounds considered by medicinal chemistry experts as difficult for solubility estimations. To better explore the learning capability of deeper-net architectures, the molecular representations of the compounds may be selected for conforming to these architectures. Specifically, the superior local-feature learning capability of the CNN architectures may be better exploited by using the substructure-encoded molecular fingerprints for representing compounds. Our studies consistently scored the substantially better solubility prediction performances of the deeper-net deep learning models on novel compounds than the established tools and shallow-net models. Nonetheless, the prediction performance of the deeper-net models on novel compounds is affected by the limited number of 9,943 compounds for training these models. Solubility prediction capability of the deeper-net methods may be further enhanced with the expanded experimental solubility data and by means of algorithm development. Our novel approach may find broader applications in the development of high-performance deep learning models for the prediction of various pharmacodynamic, pharmacokinetic, and toxicological properties.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YC and HZ conceived and designed this research, contributed to the preparation, and write-up of this manuscript. HZ designed and conducted computation. QC processed data, built models, predicted results, analyzed results data, and participated in this manuscript writing. SL synthesized compounds and experimented to test the aqueous solubility values of these compounds, and participated in this manuscript writing. BN collected and processed information on 62 recently-published novel compounds and assisted in optimizing the models. XZ

predicted the aqueous solubility values of novel compounds using commercial software. YT assisted in the design of this study.

FUNDING

The research was funded by the Fundamental Research Funds for the Central Universities (No. 2632018YX03), Singapore Academic Research Fund (R-148-000-273-114), and Shenzhen Municipal Government Grants (JCYJ2016032416, 3734374, JCYJ20170413113448742, and 20151030A0610001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00121/full#supplementary-material

REFERENCES

1. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* (2001) 46:3–26. doi: 10.1016/s0169-409x(00)00129-0
2. Di L, Fish PV, Mano T. Bridging solubility between drug discovery and development. *Drug Discov Today.* (2012) 17:486–95. doi: 10.1016/j.drudis.2011.11.007.
3. Ran Y, Yalkowsky SH. Prediction of drug solubility by the general solubility equation (GSE). *J Chem Inf Comput Sci.* (2001) 41:354–7. doi: 10.1021/ci000338c.
4. Konczol A, Dargo G. Brief overview of solubility methods: recent trends in equilibrium solubility measurement and predictive models. *Drug Discov Today Technol.* (2018) 27:3–10. doi: 10.1016/j.ddtec.2018.06.001.
5. Tetko IV, Tanchuk VY, Kasheva TN, Villa AE. Estimation of aqueous solubility of chemical compounds using E-state indices. *J Chem Inf Comput Sci.* (2001) 41:1488–93. doi: 10.1021/ci000392t
6. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. *Adv Drug Deliv Rev.* (2002) 54:355–66. doi: 10.1016/s0169-409x(02)00008-x
7. Gozalbes R, Pineda-Lucena A. QSAR-based solubility model for drug-like compounds. *Bioorg Med Chem.* (2010) 18:7078–84. doi: 10.1016/j.bmc.2010.08.003
8. Chevillard F, Lagorce D, Reynes C, Villoutreix BO, Vayer P, Miteva MA. *In silico* prediction of aqueous solubility: a multimodel protocol based on chemical similarity. *Mol Pharm.* (2012) 9:3127–35. doi: 10.1021/mp300234q
9. Boobier S, Osbourn A, Mitchell JBO. Can human experts predict solubility better than computers? *J Cheminform.* (2017) 9:63. doi: 10.1186/s13321-017-0250-y
10. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *J Chem Inf Model.* (2013) 53:1563–75. doi: 10.1021/ci400187y
11. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Mol Pharm.* (2017) 14:4462–75. doi: 10.1021/acs.molpharmaceut.7b00578
12. Wu K, Zhao Z, Wang R, Wei GW. TopP-S: persistent homology-based multi-task deep neural networks for simultaneous predictions of partition coefficient and aqueous solubility. *J Comput Chem.* (2018) 39:1444–54. doi: 10.1002/jcc.25213
13. Wang J, Hou T. Recent advances on aqueous solubility prediction. *Comb Chem High T Scr.* (2011) 14:328–38. doi: 10.2174/138620711795508331
14. Palmer DS, Mitchell JB. Is experimental data quality the limiting factor in predicting the aqueous solubility of druglike molecules? *Mol Pharm.* (2014) 11:2962–72. doi: 10.1021/mp500103r
15. Ioakimidis L, Thoukydidis L, Mirza A, Naeem S, Reynisson J. Benchmarking the reliability of qikprop. Correlation between experimental and predicted values. *QSAR Comb Sci.* (2008) 27:445–56. doi: 10.1002/qsar.200730051
16. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539.
17. Kim HK, Min S, Song M, Jung S, Choi JW, Kim Y, et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat Biotechnol.* (2018) 36:239–41. doi: 10.1038/nbt.4061
18. Joseph G, Bharath R, Evan NF, Vijay SP. Atomic convolutional networks for predicting protein-ligand binding affinity. *arXiv:1703.10603.* (2017). Available online at: https://arxiv.org/abs/1703.10603 (accessed April 4, 2019).
19. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *arXiv:1409.4842.* (2014). Available online at: https://arxiv.org/abs/1409.4842 (accessed April 4, 2019).
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv:1512.03385.* (2015). Available online at: https://arxiv.org/abs/1512.03385 (accessed April 4, 2019).
21. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *arXiv:1608.06993v3.* (2016). Available online at: https://arxiv.org/abs/1608.06993v3 (accessed April 4, 2019).
22. Ba LJ, Caruana R. Do deep nets really need to be deep? *arXiv:1312.6184.* (2013). Available online at: https://arxiv.org/abs/1312.6184 (accessed April 4, 2019).
23. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today.* (2006) 11:1046–53. doi: 10.1016/j.drudis.2006.10.005
24. Tomasulo P. ChemIDplus-super source for chemical and drug information. *Med Ref Serv Q.* (2002) 21:53–9. doi: 10.1300/J115v21n01_04
25. Wang J, Krudy G, Hou T, Zhang W, Holland G, Xu X. Development of reliable aqueous solubility models and their application in druglike analysis. *Chem Inf Model.* (2007) 47:1395–404. doi: 10.1021/ci700096r
26. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* (2019) 47:D23–8. doi: 10.1093/nar/gky1069
27. Hamdy AK, Sheha MM, Abdel-Hafez AA, Shouman SA. Design, synthesis, and cytotoxicity evaluation of novel griseofulvin analogues with improved water solubility. *Int J Med Chem.* (2017) 2017:7386125. doi: 10.1155/2017/7386125
28. Ortiz D, Guiguemde WA, Hammill JT, Carrillo AK, Chen Y, Connelly M, et al. Discovery of novel, orally bioavailable, antileishmanial compounds using phenotypic screening. *PLoS Negl Trop Dis.* (2017) 11:e0006157. doi: 10.1371/journal.pntd.0006157
29. Theppawong A, Van de Walle T, Grootaert C, Bultinck M, Desmet T, Van Camp J, et al. Synthesis of novel aza-aromatic curcuminoids with improved biological activities towards various cancer cell lines. *ChemistryOpen.* (2018) 7:381–92. doi: 10.1002/open.201800029
30. Thompson AM, O'Connor PD, Marshall AJ, Blaser A, Yardley V, Maes L, et al. Development of (6 R)-2-Nitro-6-[4-(trifluoromethoxy)phenoxy]-6,7-dihydro-5 H-imidazo[2,1- b][1,3]oxazine (DNDI-8219): a new lead for visceral leishmaniasis. *J Med Chem.* (2018) 61:2329–52. doi: 10.1021/acs.jmedchem.7b01581
31. Wilson CR, Gessner RK, Moosa A, Seldon R, Warner DF, Mizrahi V, et al. Novel antitubercular 6-dialkylaminopyrimidine carboxamides from phenotypic whole-cell high throughput screening of a softfocus library: structure-activity relationship and target identification studies. *J Med Chem.* (2017) 60:10118–34. doi: 10.1021/acs.jmedchem.7b01347
32. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform.* (2011) 3:33. doi: 10.1186/1758-2946-3-33
33. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem.* (2011) 32:1466–74. doi: 10.1002/jcc.21707

34. Paschke A, Neitzel PL, Walther W, Schüürmann G. Octanol/water partition coefficient of selected herbicides: determination using shake-flask method and reversed-phase high-performance liquid chromatography. *J Chem Eng Data.* (2004) 49:1639–42. doi: 10.1021/je049947x
35. Bergstrom CA, Wassvik CM, Johansson K, Hubatsch I. Poorly soluble marketed drugs display solvation limited solubility. *J Med Chem*. (2007) 50:5858–62. doi: 10.1021/jm0706416
36. Heng H, Zhi Y, Yuan H, Wang Z, Li H, Wang S, et al. Discovery of a highly selective FLT3 inhibitor with specific proliferation inhibition against AML cells harboring FLT3-ITD mutation. *Eur J Med Chem.* (2019) 163:195–206. doi: 10.1016/j.ejmech.2018.11.063
37. Zhi Y, Li B, Yao C, Li H, Chen P, Bao J, et al. Discovery of the selective and efficacious inhibitors of FLT3 mutations. *Eur J Med Chem.* (2018) 155:303–15. doi: 10.1016/j.ejmech.2018.06.010
38. Wang Y, Zhi Y, Jin Q, Lu S, Lin G, Yuan H, et al. Discovery of 4-((7H-Pyrrolo[2,3-d]pyrimidin-4-yl)amino)-N-(4-((4-methylpiperazin-1-yl)methyl)p henyl)-1H-pyrazole-3-carboxamide (FN-1501), an FLT3- and CDK-kinase inhibitor with potentially high efficiency against acute myelocytic leukemia. *J Med Chem.* (2018) 61:1499–518. doi: 10.1021/acs.jmedchem.7b01261
39. Lu S, Jin Q, Wang Y, Chen Y, Lu T. *4-(Fused Heterocycle Substituted-Amino)-1H-Pyrazole-3-Formamide Useful in Treatment of Various Diseases and Its Preparation*. CHN Patent No 2015100407173. Beijing: National Intellectual Property Administration, PRC (2015).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cui, Lu, Ni, Zeng, Tan, Chen and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 30 April 2020
doi: 10.3389/fphar.2020.00534

Identifying Drug Targets in Pancreatic Ductal Adenocarcinoma Through Machine Learning, Analyzing Biomolecular Networks, and Structural Modeling

Wenying Yan[1†], *Xingyi Liu*[1†], *Yibo Wang*[1†], *Shuqing Han*[1†], *Fan Wang*[1], *Xin Liu*[1], *Fei Xiao*[1*] and *Guang Hu*[1,2*]

[1] Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China, [2] State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Xiaojun Tian,
Arizona State University, United States
Junfeng Xia,
Anhui University, China

***Correspondence:**
Fei Xiao
xiaofei@suda.edu.cn
Guang Hu
huguang@suda.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: 04 March 2020
Accepted: 06 April 2020
Published: 30 April 2020

Citation:
Yan W, Liu X, Wang Y, Han S, Wang F, Liu X, Xiao F and Hu G (2020) Identifying Drug Targets in Pancreatic Ductal Adenocarcinoma Through Machine Learning, Analyzing Biomolecular Networks, and Structural Modeling. Front. Pharmacol. 11:534. doi: 10.3389/fphar.2020.00534

Pancreatic ductal adenocarcinoma (PDAC) is one of the leading causes of cancer-related death and has an extremely poor prognosis. Thus, identifying new disease-associated genes and targets for PDAC diagnosis and therapy is urgently needed. This requires investigations into the underlying molecular mechanisms of PDAC at both the systems and molecular levels. Herein, we developed a computational method of predicting cancer genes and anticancer drug targets that combined three independent expression microarray datasets of PDAC patients and protein-protein interaction data. First, Support Vector Machine–Recursive Feature Elimination was applied to the gene expression data to rank the differentially expressed genes (DEGs) between PDAC patients and controls. Then, protein-protein interaction networks were constructed based on the DEGs, and a new score comprising gene expression and network topological information was proposed to identify cancer genes. Finally, these genes were validated by "druggability" prediction, survival and common network analysis, and functional enrichment analysis. Furthermore, two integrins were screened to investigate their structures and dynamics as potential drug targets for PDAC. Collectively, 17 disease genes and some stroma-related pathways including extracellular matrix-receptor interactions were predicted to be potential drug targets and important pathways for treating PDAC. The protein-drug interactions and hinge sites predication of ITGAV and ITGA2 suggest potential drug binding residues in the Thigh domain. These findings provide new possibilities for targeted therapeutic interventions in PDAC, which may have further applications in other cancer types.

Keywords: pancreatic ductal adenocarcinoma, drug targets, support vector machine–recursive feature elimination, protein-protein interactions, structural dynamics, integrins

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most malignant solid tumors (Bailey et al., 2016). PDAC is difficult to treat due to the stage of diagnosis, severe cachexia and poor metabolic status, the resistance of cancer stem cells (CSCs) to current drugs, and the marked desmoplastic response that facilitates growth and invasion, provides a physical barrier to therapeutic drugs, and prevents immunosurveillance (Al Haddad and Adrian, 2014). PDAC is also a drug-resistant disease, and the response of pancreatic cancer to most chemotherapy drugs is poor. Until now, most of research effort in PDAC has been directed at identifying the important disease-driving genes and pathways (Waddell et al., 2015). These studies have shown that *KRAS*, *CDKN2A*, *TP53*, and *SMAD4* are the four most common driver genes in PDAC (Carr and Fernandez-Zapico, 2019). With the development of multi-omics data, a series of new regulators that are strongly correlated with survival have been proposed to be PDAC biomarkers (Rajamani and Bhasin, 2016; Mishra et al., 2019), including genes (e.g., *IRS1*, *DLL1*, *HMGA2*, *ACTN1*, *SKI*, *B3GNT3*, *DMBT1*, and *DEPDC1B*) and lncRNAs (e.g., PVT1 and GATA6-AS). The integrated transcriptomic analysis of five PDAC datasets identified four-hub gene modules, which were used to build a diagnostic risk model for the diagnosis and prognosis of PDAC (Zhou et al., 2019). Integrated genomic analysis of 456 PDAC cases identified 32 recurrently mutated genes that aggregate into 10 pathways: KRAS, TGF-β, WNT, NOTCH, ROBO/SLIT signaling, G1/S transition, SWI-SNF, chromatin modification, DNA repair, and RNA processing (Bailey et al., 2016). Previous treatments for pancreatic cancer have focused on targeting some of these PDAC-associated pathways, including TGFβ (Craven et al., 2016), PI3K (Conway et al., 2019), Src (Parkin et al., 2019), and RAF→MEK→ERK (Kinsey et al., 2019) and NFAT1-MDM2-MDMX (Qin et al., 2017) signaling, as well as cell-cell communication within the tumor microenvironment (Shi et al., 2019). The discovery of novel drug targets provides extremely valuable resource towards the discovery of drugs. Although the human genome comprises approximately 30,000 genes, proteins encoded by fewer than 400 are used as drug targets in disease treatments. A range of therapeutic targets in PDAC have been proposed, including suppressing the abovementioned genes and pathways (Tang and Chen, 2014). However, the current drug targets for PDAC will not be 100% effective due to the heterogeneous nature of the disease. To tackle this challenge, a complete understanding of the molecular mechanism of PDAC is urgently needed.

Improving PDAC therapy will require a greater knowledge of the disease at both the systems and molecular levels. At the systems level, protein-protein interaction (PPI) networks provide a global picture of cellular function and biological processes (BPs); thus, the network approach is used to understand the molecular mechanisms of disease, particularly in cancer (Conte et al., 2019; Sonawane et al., 2019). Some proteins act as hub proteins that are highly connected to others, thus cancer drug targets can be predicted by hubs in PPI networks (Li et al., 2018; Lu et al., 2018; Zhu et al., 2019). However, there are some conflicting results that suggest disease genes or drug targets have no significant degree of prominence (Mitsopoulos et al., 2015), but higher betweenness, centrality, smaller average shortest path length, and smaller clustering coefficient (Zhao and Liu, 2019). Recent advances in systems biology have led to a plethora of new network-based methods and parameters for predicting essential genes (Li et al., 2019), disease genes, and drug targets (Csermely et al., 2013; Vinayagam et al., 2016; Zhang et al., 2017; Fotis et al., 2018; Liu et al., 2018). Additionally, the structural annotation of PPI networks that has highlighted key residues has enriched the fields of both systems biology and rational drug design (Kar et al., 2009; Winter et al., 2012). The prediction of binding sites, allosteric sites, and genetic variations based on systems-level data is critical for suggesting therapeutic approaches to complex diseases and personalized medicine (Duran-Frigola et al., 2013; Yan et al., 2018). Combined with PPI network analysis, molecular docking studies of target genes can further help to find drug molecules and protein-drug interactions for lung adenocarcinoma (Selvaraj et al., 2018).

Together with advances in "-omics" data, including gene expression and PPI data, machine learning (ML), and artificial intelligence (AI) techniques are powerful tools that can assess gene and protein "druggability" from such massive and noisy datasets (Kandoi et al., 2015; Zhavoronkov, 2018). As the most used ML method, support vector machine (SVM) has been used for cancer genomic classification or subtyping, which may be useful for obtaining a better understanding of cancer driver genes and discovering new biomarkers and drug targets (Huang et al., 2018). ML-based methods have been applied to study PDAC for different purposes. By applying ML algorithms to proteomics and other molecular data from The Cancer Genome Atlas (TCGA), two subtypes of pancreatic cancer can be classified (Sinkala et al., 2020). A meta-analysis of PDAC microarray data could help predict biomarkers that can be used to build AI-based computational predictors for classifying PDAC and normal samples (Bhasin et al., 2016), as well as predicting sample status (Almeida et al., 2020). To predict and validate novel drug targets for cancer, including PDAC, a ML-based classifier that integrates a variety of genomic and systems datasets was built to prioritize drug targets (Jeon et al., 2014).

In this study, we developed a computational framework that integrates various types of high-throughput data, including transcriptomics, interactomics, and structural data, for the genome-wide identification of therapeutic targets in PDAC. A novel centrality metric, referred to as SVM-REF and Network topological score (RNs), was proposed for the identification of disease genes and drug targets. This method incorporates gene expression and network topology information from ML and PPI analyses. Moreover, the predicted genes were validated by "druggability" prediction, survival, and comparative network analyses, as well as functional enrichment analysis. Finally, the structural and dynamic properties of two integrins (ITGAV and ITGA2) as drug targets were investigated. The workflow of these methods is shown in **Figure 1**.

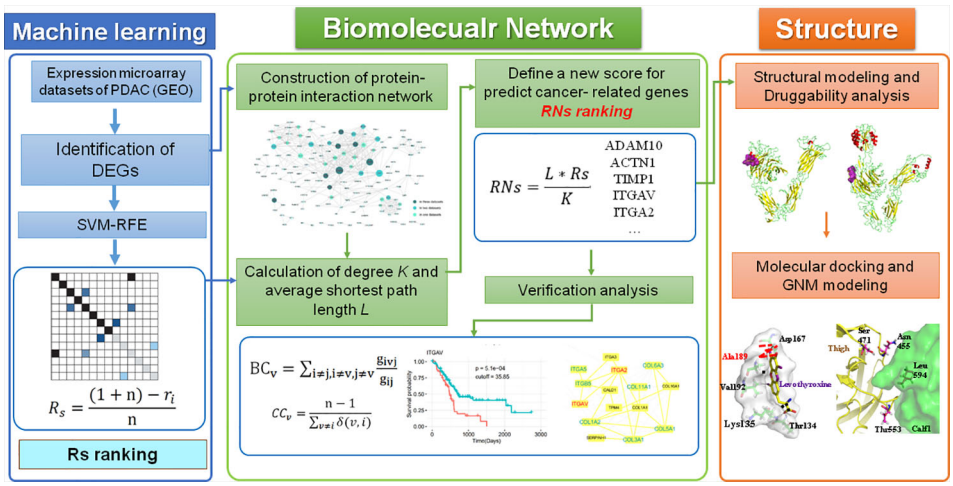


FIGURE 1 | The computational pipeline proposed in this work included three steps. Overall, a machine learning method was used to identify DEGs in PDAC, which were then combined with two parameters of the PPI network to define a new score that predicted disease genes and drug targets in PDAC. All potential targets were then further verified by other bioinformatics analyses and investigated by a "druggability" analysis of structural and dynamic properties.

MATERIALS AND METHODS

Identification of DEGs

In this study, three independent PDAC expression microarray datasets with 184 pancreas samples (95 cancer and 89 nonmalignant samples) were used. The datasets were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). Details of each dataset are listed in **Table 1**. The GSE15471 dataset included 36 PDAC samples and matching normal pancreas samples from pancreatic cancer patients in Romania (Badea et al., 2008). There were also matched samples in the GSE28735 dataset, which contains gene expression profiles of 45 matched pairs of pancreatic tumor and adjacent non-tumor tissues from PDCA patients in Germany (Zhang et al., 2012; Zhang et al., 2013). The GSE71989 dataset contained expression profiles of eight normal pancreas and 14 PDAC tissues (Jiang et al., 2016). The normalized data were downloaded from GEO and then analyzed to identify DEGs using t-tests, with p-values adjusted by the Benjamini-Hochberg method. Only genes with adjusted p-values < 0.01 and |FC| > 1.5 were chosen as DEGs.

TABLE 1 | Information on the included GEO datasets.

Accessions	Platforms	Samples (tumor vs. non-tumor tissues)	References
GSE15471	Affymetrix Human Genome U133 Plus 2.0 Array	36 vs. 36	(Badea et al., 2008)
GSE28735	Affymetrix Human Gene 1.0 ST Array	45 vs. 45	(Zhang et al., 2012; Zhang et al., 2013)
GSE71989	Affymetrix Human Genome U133 Plus 2.0 Array	14 vs. 8	(Jiang et al., 2016)

Gene Prioritization Pipeline

Disease genes and drug targets usually have large degree in PPI networks, but there is no single network parameter that can accurately predict them (Li et al., 2016). Protein targets do not exert their function in isolation; rather they are affected by interactions within their PPI network, which are governed by protein localization and environment. In the same way, topological information from PPI networks alone is not enough to identify disease genes and drug targets without biological information. To overcome these limitations, we developed a new three-step pipeline to identify cancer-related genes that may be candidate drug targets in PDAC. The pipeline integrated information from gene expression data and local and global topological characteristics of genes in PPI networks.

Step 1: For each gene expression dataset, we employed SVM methods based on a Recursive Feature Elimination (SVM-RFE) algorithm (Guyon et al., 2002), which is an embedded method to specifically deal with gene selection for cancer classification (Bolón-Canedo et al., 2014), rank DEGs, and select the most relevant features (Jeon et al., 2014). SVM-RFE can remove redundant features (genes) to generalize performance, implement backward feature elimination, search an optimal subset of genes, and provide a ranking for each gene. We ranked genes by SVM-RFE score (R_s), according the following formula:

$$R_S = \frac{(1+\mathrm{n}) - r_i}{\mathrm{n}},$$

where n is the number of DEGs and r_i is the rank of gene i.

Step 2: A PPI network of DEGs was constructed with the STRING database (von Mering et al., 2003; Szklarczyk et al., 2017) using scores > 0.9. The topological parameters degree and shortest path length for each gene in the PPI network were

calculated. The degree (K) of a node in the PPI network is the number of links attached to that node, which is one of the measures of centrality of a node in the network. The average path length (L) of node v in the network is the average length of the shortest paths between v and all other nodes and was defined as:

$$L_v = \frac{\Sigma_{v \neq i}^{n} \delta(v,\ i)}{n-1},$$

where $\delta(v,i)$ is the length of the shortest path between nodes v and I, and n is the node number in the network.

Step 3: Finally, we incorporated Network topological properties into R_s and defined a new score (RNs) for each gene as:

$$RNs = \frac{L \star R_s}{K}.$$

Accordingly, this new RNs score (SVM-RFE and Network topological score) considers the cancer status of each gene by including information about gene expression and two levels of topological features in PPI networks, namely, degree K indicates the importance of the node, while the shortest path length L shows the effects from other nodes. The code for gene prioritization is freely available on GitHub for download at: https://github.com/CSB-SUDA/RNs.

PPI Network Analysis

Once the PPI network was constructed, two other analyses were performed. The first analysis was the calculation of two commonly used centrality parameters: betweenness and closeness centrality. The betweenness centrality (BC) (Freeman, 1977) of node v was defined as:

$$\mathrm{BC_v} = \Sigma_{\mathrm{i} \neq \mathrm{j}, \mathrm{i} \neq \mathrm{v}, \mathrm{j} \neq \mathrm{v}} \frac{g_{ivj}}{g_{ij}},$$

where g_{ivj} is the number of the shortest paths from i to j that pass through node v, and g_{ij} is the number of shortest paths from i to j.

The closeness (CC) of node v is the reciprocal of the average shortest path length, which was calculated as:

$$CC_v = \frac{\mathrm{n}-1}{\Sigma_{v \neq i} \delta(v,i)}.$$

Proteins are often incorporated into modules that can be shared between several different cellular activities. The second analysis was module detection of PPIs by integrating a Gaussian network (GN) algorithm (Newman and Girvan, 2004) and functional semantic similarity (Wang et al., 2007). In general, this involved using the GN algorithm to detect the module of PPI networks, and then applying functional semantic similarity to filter links. Thus, the genes in the detected modules not only had topological similarity, but also functional similarity.

Survival Analysis

To evaluate the prognostic value of candidate genes, a survival analysis was performed using data from the human protein atlas (Uhlen et al., 2017), which contains gene expression data and clinical information of 176 pancreatic cancer patients. P-values < 0.01 were considered significantly correlated with overall survival.

Functional Enrichment Analysis

Functional enrichment analysis, including cellular component (CC), molecular function (MF), and BP, from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of genes was performed using the R package cluster Profiler (Yu et al., 2012). Terms with adjusted p-value < 0.05 were considered significant.

Structural Modeling and "Druggability" Analysis

The protein structures of potential drug targets were retrieved from the Protein Data Bank (PDB) if they were available. The Swiss model (Waterhouse et al., 2018) and I-TASSER (Roy et al., 2010) were used for the structural modeling of genes if protein structures were unavailable. We choose the Swiss model when the sequence similarity between searched models was >30%; otherwise, we used I-TASSER, which predicts protein structure using modeling by iterative threading assembly. Based on model structures, Fpocket (Le Guilloux et al., 2009) was used to detect druggable pockets and calculate "druggability" scores, which were based on several physicochemical descriptors on a genomic scale. The pocket with the highest score in the entire PDB was defined as the reference druggable score. The score of each pocket was classified as: 0.0–0.5: non-druggable; 0.5–0.7: druggable; and 0.7–1.0: highly druggable.

Molecular Docking and GNM Modeling

To study the interactions and binding mode of small molecules with the potential drug targets, molecular docking was performed using AutoDock 4.2 (Khodade et al., 2007). The target, drug, and related disease information were collected from the Drug Bank database (Version 5.0) (Wishart et al., 2018) and the Therapeutic Target Database 2020 (Wang et al., 2020). A normal mode analysis of the GN model (GNM) was performed to investigate collective dynamics *via* the DynOmics online tool (Danne et al., 2017). The default cutoff distance of 7.3 Å between GNM model nodes was used.

RESULTS AND DISCUSSION

Identification of Disease Genes and Drug Targets in PDAC

From the three datasets GSE28735, GSE71989, and GSE15471, we identified 3,079, 1,225, and 2,257 DEGs between PDAC and adjacent tissues, respectively. The top 10 genes with the smallest p-values are marked in **Figure 2**. In GSE28735, 1,724 genes showed increased expression in PDAC tissues, while 1,355 genes showed decreased expression (**Figure 2A**). In GSE71989, 766 genes were upregulated and 459 genes were downregulated in PDAC tissues compared with normal tissues (**Figure 2B**). In GSE15471, 1713 genes were overexpressed, while 544 genes showed decreased expression in tumor tissues (**Figure 2C**). Together, there were 313 common DEGs between PDAC and adjacent tissues in all three datasets (**Figure 2D**).

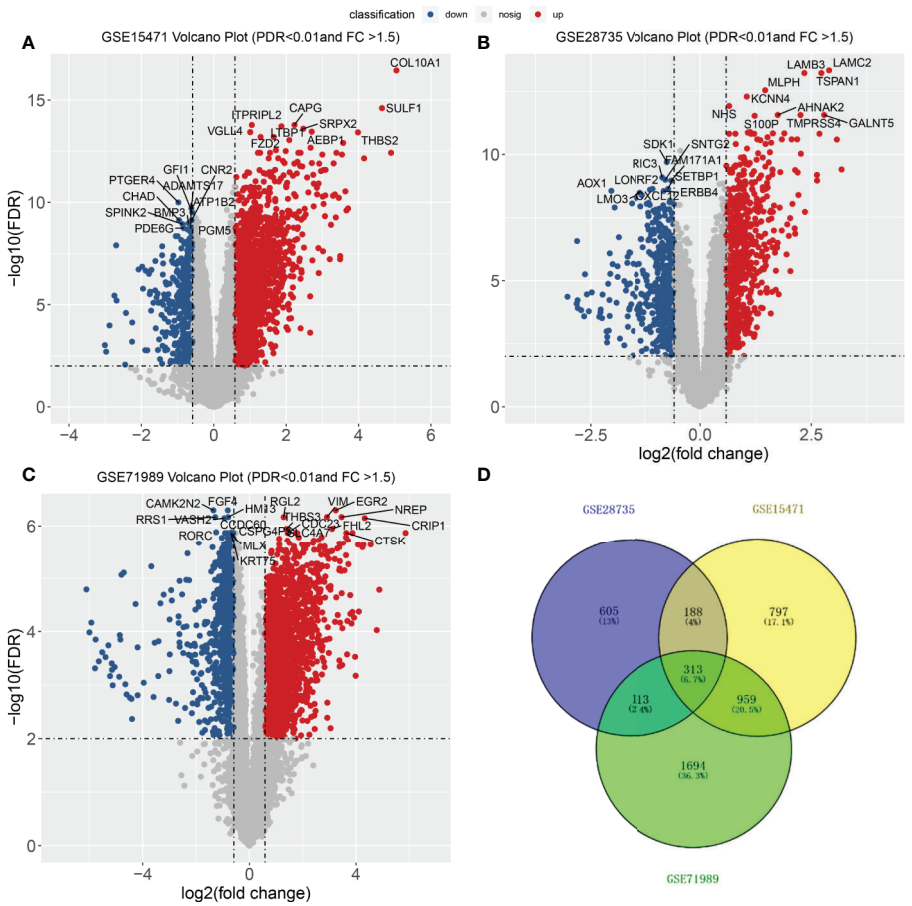


FIGURE 2 | Differentially expressed genes (DEGs) between PDACs and normal tissues. **(A–C)** Volcano plot of –log10 (FDR) vs. log2 (fold change) of DEGs in the three datasets. **(D)** Venn diagram with the number of overlapping DEGs from the different datasets.

Additionally, we evaluated gene expression as an input feature for ML and selected the most relevant genes for PDAC using SVM-RFE (Almeida et al., 2020), which provided a ranking for the genes. Then, each DEG was assigned an R_s value (see *Materials and Methods*), which was used to further rank all genes. As an illustration, the top 100 R_s values of the DEGs in each dataset are listed in **Table S1**. This shows that there is little overlap of results between the different datasets. This means that calculating R_s based on SVM-RFE can provide information for classification, but not enough for ranking.

The DEGs were next mapped to the STRING database, which yielded a PPI network with 144 genes and 440 links (**Figure 3**). Then, degree and shortest path length of each gene in the network were calculated. Finally, we ranked the genes according to our designed score RNs, which integrated these two topological parameters and was based on gene expression profile. The top 20 genes predicted based on at least two datasets were considered potential drug targets. As shown in **Table 2** and **Table S2**, eight genes (*ADAM10, TIMP1, MATN3, PKM, APLP2, ACTN1, CALU*, and *VCAN*) were identified in all three datasets, and nine genes (*LGALS1*, *ITGA2*, *BST2*, *MFGE8*, *ITGAV*, *EGF*, *APOL1*, *ALB*, and *MSLN*) were identified in two of three datasets. We propose that genes predicted by at least two datasets could serve as disease genes and/or drug targets. Taken together, 17 genes predicted by RNs score are listed in **Table 3**, and most have been previously reported to be PDAC-associated genes. There are only four that have not been previously associated with PDAC. This suggests that our metric RNs is useful for identifying novel disease genes and drug targets.

It is also useful to compare our results predicted by RNs with other common network parameters. The genes predicted by calculating betweenness and closeness centrality are also listed in **Table S2**. Among our 20 predicted potential drug targets, six and nine were also found by betweenness and closeness centrality, respectively. Notably, *ADAM10*, *ACTN1*, and *TIMP1* were in all three lists, which suggested they had important roles in PDAC. Moreover, two other genes (*ITGAV* and *ITGA2*) were in the top 20 of two datasets, which suggested they should be investigated. Overall, compared with the top 20 genes predicted by these two common network parameters, our RNs parameter identified more extracellular matrix (ECM) proteins, including integrins and collagens. The other

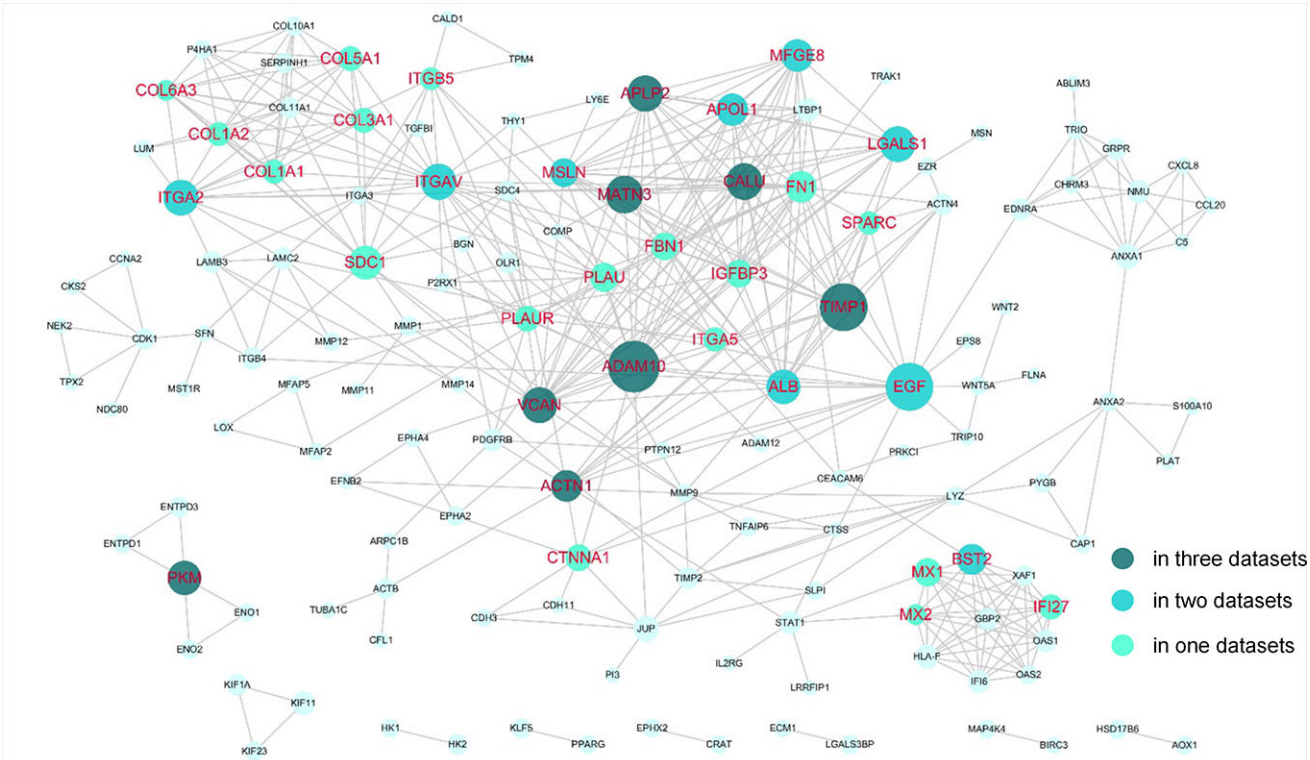


FIGURE 3 | Potential drug targets in the PPI network. The genes that were predicted by our pipeline are marked with red labels. The node size denotes the average RNs of the gene in two or three datasets.

TABLE 2 | Identified potential drug targets for PDAC.

In three datasets	*ADAM10*, *TIMP1*, *MATN3*, *PKM*, *APLP2*, *ACTN1*, *CALU*, *VCAN*
In two of three datasets	*LGALS1*, *ITGA2*, *BST2*, *MFGE8*, *ITGAV*, *EGF*, *APOL1*, *ALB*, *MSLN*
In only one dataset	*COL5A1*, *CTNNA1*, *MX1*, *COL1A2*, *COL6A3*, *SPARC*, *IFI27*, *SDC1*, *FN1*, *PLAU*, *PLAUR*, *IGFBP3*, *FBN1*, *COL1A1*, *COL3A1*, *ITGB5*, *ITGA5*, *MX2*

interesting finding was that four common genes (*ALB*, *EGF*, *ITGA2*, and *VCAN*) were identified by isolating the nodes with large degrees (hubs) in PPI network construction based on other PDAC GSE datasets (Lu et al., 2018).

Survival analysis was also performed to evaluate whether the expression of our 17 identified candidates was related to the prognosis of PDAC. Using Kaplan-Meier analysis with the log-rank test for 176 pancreatic cancer patients from the human protein atlas (Uhlen et al., 2017), we found that higher expression levels of 11 genes were significantly correlated with decreased overall survival ($p < 0.01$, **Figure 4**). For the eight genes identified in all three datasets, five (*ADAM10*, *PKM*, *APLP2*, *CALU*, and *VCAN*) were associated with poor prognosis when highly expressed. The other six highly expressed genes (*LGALS1*, *ITGA2*, *BST2*, *ITGAV*, *APOL1*, and *MSLN*) associated with poor prognosis that were identified in two of three datasets are shown in **Table 2**. Accordingly, the survival analysis showed significant prognostic values for most of the predicted genes.

Characterization of Predicted Drug Targets for PDAC

Table 3 shows the genes predicted above shortlisted based on our RNs criteria. After searching the drug bank, these 17 predicted genes were classified into two types: 11 genes were drug targets, while six were non-drug targets. We also annotated drug targets in the drug bank by their related drugs and diseases. It should be noted that *MSLN* was the only proven drug target for PDAC, and there are many drugs that inhibit *ALB*. Thus, we concluded that these two genes had been studied widely and would not give us more insight regarding discovering new targets. Considering the potential of other predicted genes as drug targets for PDAC, we performed functional and "druggability" annotations for all. Among the 15 genes, 11 (*ADAM10*, *TIMP1*, *EGF*, *APLP2*, *ITGAV*, *VCAN*, *ITGA2*, *PKM*, *APOL1*, *ACTN1*, and *BST2*) have been reported to be contributing factors in PDAC invasion, growth, or metastasis, which indicated that our pipeline had good performance for finding potential drug targets for PDAC.

The protease *ADAM10* was predicted as the highest ranked gene, and it has been reported that *ADAM10* influences the

TABLE 3 | List of prioritized protein targets with their drug target information and "druggability" features.

Gene	RNs	Drug targets*	Drug(s)#	Disease(s)#	PDB	DS
ADAM10	5.34	Yes	XL784	Solid tumor/cancer, Breast cancer	6BE6	0.694
TIMP1	4.79	No	NA	NA	1LQN	0.839
EGF	4.77	Yes	Sucralfate, Tesevatinib, Alpha-Aminobutyric Acid, Cholecystokinin	Oral mucositis, Vulnerary	template: 5GJE	0.968
MATN3	3.31	No	NA	NA	template: 6BXJ	0.545
CALU	3.17	No	NA	NA	template: 2Q4U	0.677
APLP2	3.12	Yes	Zinc, Zinc acetate, Zinc chloride	NA	5TPT	0.912
ITGAV	3.08	Yes	Abituzumab, Levothyroxine	Colorectal cancer, Solid tumour/ cancer	3IJE	0.663
VCAN	3.05	Yes	Hyaluronic acid	NA	template: 4CSY	NA
LGALS1	3.03	Yes	Thiodigalactoside, 1,4-Dithiothreitol, Mercaptoethanol, Artenimol	NA	3W59	NA
ITGA2	3.027	No	NA	NA	Templates: 3K71, 4NEH, 3K6S	0.672
ALB	2.85	Yes	Gadobenate Dimeglumine, Glycyrrhizic acid, Patent Blue, (365 drugs)	Hemophilia, Schizophrenia	4BKE	1.000
PKM	2.82	Yes	Pyruvic acid, L-Phospholactate, 2-Phosphoglycolic Acid, et al.	Pain, Renal cell carcinoma;	6GG5	0.996
MFGE8	2.54815	No	NA	NA	template: 4DEQ	NA
APOL1	2.52	Yes	Zinc, Zinc acetate, Zinc chloride	NA	template: 5J2L	0.503
ACTN1	2.45	Yes	Copper, Human calcitonin	NA	template: 4D1E	0.673
BST2	2.31	No	NA	NA	3MQB	0.821
MSLN	2.0	Yes	Amatuximab	Ovarian/Pancreatic cancer	4F3F	0.727

**"YES" means drug target, and "NO" means non-drug target; #"NA" means no drug and disease information, or no druggable pockets.*

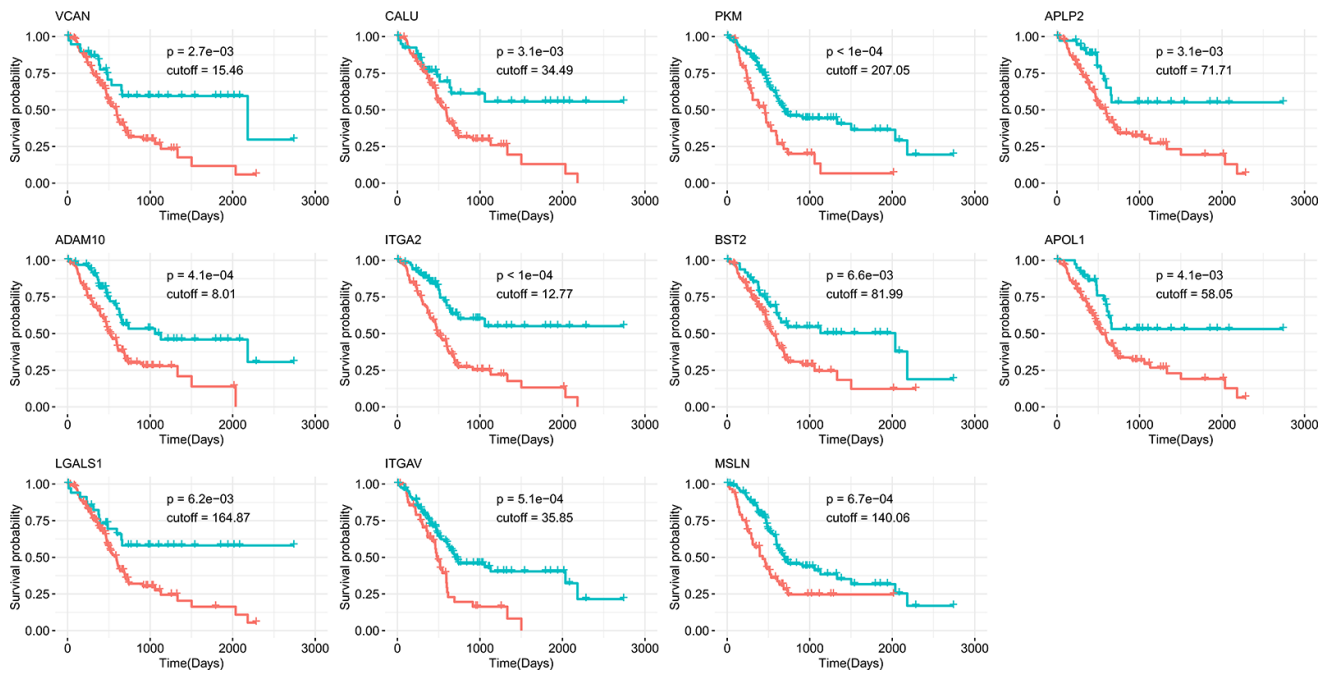


FIGURE 4 | Kaplan-Meier survival curves of overall survival from the human protein atlas datasets for potential drug targets divided by high (red) or low (green) expression level.

progression and metastasis of cancer cells, as it promotes PDAC cell migration and invasion (Gaida et al., 2010). Inhibiting *ADAM10* could be a novel approach for natural killer (NK) cell-based immunotherapy (Pham et al., 2017). Tissue inhibitor of metalloproteinases-1 (*TIMP-1*) correlated with tumor progression, and elevated levels of *TIMP-1* in tumor tissue and peripheral blood were associated with poor clinical outcomes in numerous malignancies, including PDAC (Prokopchuk et al., 2018). The third gene was epidermal growth factor (*EGF*), which was a common disease gene for many cancers, and EGF mutations were associated with PDAC (Grapa et al., 2019). Amyloid precursor-like protein 2 (*APLP2*) affects the actin cytoskeleton and also increases PDAC growth and metastasis (Pandey et al., 2015). *ITGAV* (Villani et al., 2019), *VCAN* (Skandalis et al., 2006), and *ITGA2* (Nones et al., 2014) are matrix proteins that have been shown to contribute to pancreatic cancer cell migration, invasion, and metastasis. *PKM2* is one of the isoforms of pyruvate kinase muscle isozyme (*PKM*) and promotes the invasion and metastasis of PDAC through the phosphorylation and stabilization of PAK2 (Cheng et al., 2018). The final three genes, *APOL1* (Liu et al., 2017), *ACTN1* (Rajamani and Bhasin, 2016), and *BST2* (Grutzmann et al., 2005) have previously been reported to be effective biomarkers for PDAC.

Although 11 genes were already known drug targets, "druggability" annotations based on protein structures can improve our knowledge and understanding of the mechanisms of proteins as drug targets. The "druggability" of proteins is a measure of their ability to bind drug-like molecules based on molecular shapes. For the "druggability" of all 17 genes, we first obtained their structural modes by retrieved data from the PDB database or homology modeling. The PDB codes of proteins or their templates are listed in **Table 3**. Then, Fpocket was used to compute all possible pockets and their corresponding "druggability score" (DS). The "druggability" of the protein was defined as the DS of the highest scoring pocket. As expected, most of the predicted proteins were druggable (DS ≥ 0.5), except VCAN, IGALS1, and MFGE8. ALB had the largest DS (1.00), which can partially explain why so many ALB inhibitors exist. Among the six non-drug targets, TIMP1, ITGA2, and BST2 were predicted as highly druggable (DS ≥ 0.5), which meant that these three genes had the structural abilities to be drug targets. In particular, the non-drug target ITGA2 had a larger DS than ITGAV, suggesting that a more detailed structural comparison between these two integrin proteins is needed.

Identification of Functional Modules and Pathways

Within PPI networks, cancer targets interact with different modules to perform biological functions. A module within a network is defined a set of nodes that are densely connected within subsets of the network but may not all directly interact with each other. To get further insight into the topological and biological functions of potential targets, we performed module detection in the PPI network using a GN algorithm and functional semantic similarity. As shown in **Figure 5**, we

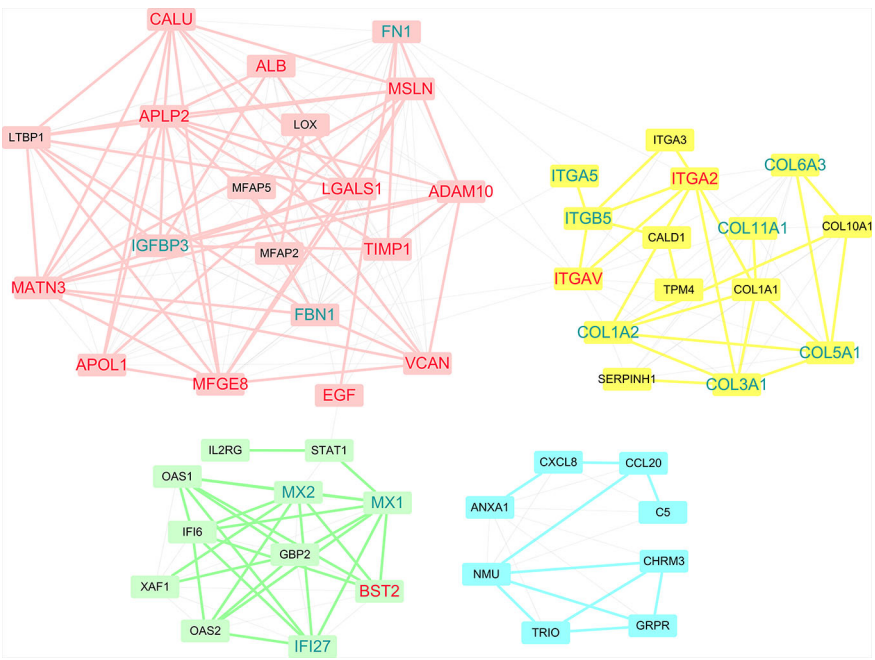


FIGURE 5 | Four modules were discovered within PPI networks. Genes that were predicted in at least two datasets are marked red, while genes that were predicted in only one dataset are marked blue.

identified four modules (the pink, yellow, green, and blue nodes) and labeled the genes that were predicted in at least two datasets (red) or in only one dataset (blue). Except *PKM* and *ACTN1*, 15 of the 17 predicted genes were detected by the modular analysis and are included in these four modules. The top module (pink) was formed of 19 genes, including the most of our predicted genes (12/17, *ADAM10*, *CALU*, *ALB*, *APLP2*, *MSLN*, *LGALS1*, *TIMP1*, *MATN3*, *VCAN*, *EGF*, *MFGE8*, and *APOL1*). Most of these genes have been previously reported as disease genes in PDAC or drug targets in other cancers. Another three predicted genes were included in two other modules, while *ITGAV* and *ITGA2* were detected in the second largest module (yellow). Although there were only two predicted genes, this module deserves more attention, as it primarily contains two types of gene targets: integrins (*ITGA5*, *ITGA3*, *ITGB5*, *ITGA2*, and *ITGAV*) and collagens (*COL6A3*, *COL11A1*, *COL1A1*, *COL10A1*, *COL5A1*, *COL1A2*, and *COL3A1*). Research into integrins and collagens and their interactions may provide more insights into the molecular mechanisms of PDAC.

We next performed an enrichment analysis on genes in the PPI network (**Figure 6** and **Table 4**). The genes were enriched for the GO terms related to extracellular structure and matrix, such as extracellular structure and matrix organization in BP, ECM in CC, and ECM structural constituent and binding in MF. **Table 4** shows the top 10 most significantly enriched KEGG pathways. Most of the pathways are associated with cancer, such as ECM-receptor interaction, focal adhesion, and proteoglycans in cancer. Moreover, integrins were enriched in most of the carcinogenesis-associated pathways, such as focal adhesion, which play essential roles in important BPs, including cell motility, proliferation, and differentiation. Interestingly, several altered molecular pathways were identified, which suggests that genes in the secondary module were involved in these pathways. These modules and pathways not only contained integrins, but also another group of collagens. In particular, two predicted integrins (*ITGAV* and *ITGA2*) were involved in nine out of the top 10 pathways, while the top four pathways (ECM-receptor interaction, focal adhesion, proteoglycans in cancer, and human papillomavirus infection) also contained collagens, especially *COL1A1* and *COL1A2*. Except for these pathways, the list of integrins and collagens was used to define the traditional cancer-related PI3K/AKT pathway. It was previously known that collagen is a major component of the tumor microenvironment that participates in cancer fibrosis, which can influence tumor cell behavior through integrins (Xu et al., 2019). Our results indicated that *ITGAV*, *ITGA2*, and their interactions with *COL1A1* and *COL1A2* may play important roles in PDAC, suggesting they could serve as potential drug targets. For example, the predicted genes and their interactions were highlighted in the ECM-receptor interaction pathway (**Figure S1**). This systems biology evidence of gene cluster- and pathway-based distributions suggested that targeting several key genes together could be a more promising approach.

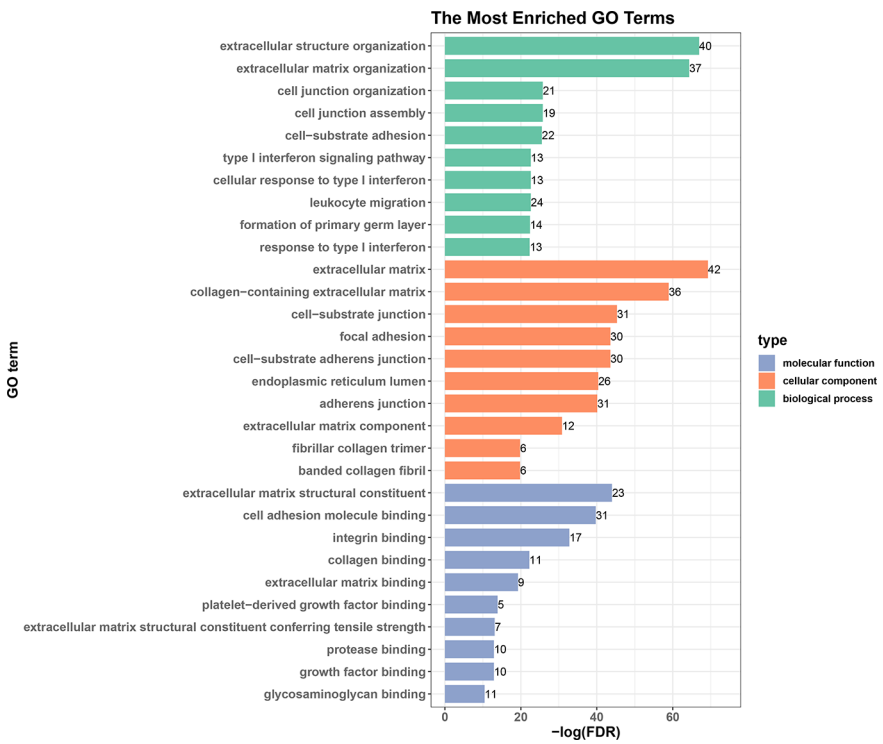


FIGURE 6 | Top 10 enriched GO terms in biological processes, cellular components, and molecular functions.

TABLE 4 | Top 10 enriched KEGG pathways (integrins and collagens are marked in bold).

KEGG term	Gene(s)	Count	Adjust p-value
ECM-receptor interaction	***COL1A1***, ***COL1A2***, ***COL6A3***, *COMP*, *FN1*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *LAMB3*, *LAMC2*, *SDC1*, *SDC4*	15	2.62E-11
Focal adhesion	*ACTB*, *ACTN4*, *ACTN1*, *BIRC3*, ***COL1A1***, ***COL1A2***, ***COL6A3***, *COMP*, *EGF*, *FLNA*, *FN1*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *LAMB3*, *LAMC2*, *PDGFRB*	20	2.67E-11
Proteoglycans in cancer	*ACTB*, ***COL1A1***, ***COL1A2***, *FLNA*, *FN1*, ***ITGA2***, ***ITGA5***, ***ITGAV***, ***ITGB5***, *LUM*, *MMP9*, *MSN*, *PLAU*, *PLAUR*, *SDC1*, *SDC4*, *EZR*, *WNT2*, *WNT5A*	19	2.97E-10
Human papillomavirus infection	***CCNA2***, ***COL1A1***, ***COL1A2***, ***COL6A3***, *COMP*, *EGF*, *FN1*, *HLA-F*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *LAMB3*, *LAMC2*, *MX1*, *PDGFRB*, *PKM*, *PRKCI*, *STAT1*, *WNT2*, *WNT5A*	23	4.12E-10
Regulation of actin cytoskeleton	*ACTB*, *ACTN4*, *ACTN1*, *CFL1*, *CHRM3*, *EGF*, *FN1*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *MSN*, *PDGFRB*, *EZR*, *ARPC1B*	17	8.45E-10
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	*ACTB*, *CTNNA1*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *JUP*	9	1.25E-05
PI3K-Akt signaling pathway	***COL1A1***, ***COL1A2***, ***COL6A3***, *COMP*, *EGF*, *EPHA2*, *FN1*, *IL2RG*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *LAMB3*, *LAMC2*, *PDGFRB*	17	3.87E-05
Amoebiasis	*ACTN4*, *ACTN4*, *ACTN1*, ***COL1A1***, ***COL1A2***, ***COL3A1***, *FN1*, *CXCL8*, *LAMB3*, *LAMC2*	9	1.03E-04
Hypertrophic cardiomyopathy (HCM)	*ACTB*, ***ITGA2***, ***ITGA3***, ***ITGA5***, ***ITGAV***, ***ITGB4***, ***ITGB5***, *TPM4*	8	3.01E-04
Small cell lung cancer	*BIRC3*, *CKS2*, *FN1*, ***ITGA2***, ***ITGA3***, ***ITGAV***, *LAMB3*, *LAMC2*	8	3.01E-04

ITGAV and ITGA2 as Potential Drug Targets for PDAC

By combining SVM-RFE, PPI network, and survival analysis, 11 out of 17 candidate genes have been predicted as biomarkers in pancreatic cancer patients. Among them, two integrins of ITGAV and ITGA2 were further screened as two potential drug targets according to the following evidences: 1) Both ITGAV and ITGA2 are involved in all PDAC-related pathways include ECM-receptor interaction and focal adhesion pathways, suggesting that ITGAV and ITGA2 may play an important role in PDAC progression; 2) Based on the druggability criteria, ITGAV and ITGA2 have relatively high DS. In addition, ITGAV is already a drug target for other cancer. Due to the structural similarity, ITGA2 can also be considered as a potential drug target; 3) Current experimental data suggest that several other integrins are overexpressed in various cancer types, being involved in tumor progression through tumor cell invasion and metastases. For example, the therapeutic potential of ITGA5 in the PDAC stroma has been proved efficacy (Kuninty et al., 2019). Collectively, our data together with some know results point towards ITGAV and ITGA2 as two potential drug targets for PDAC. Thus, the emerging understanding of their structural properties will guide the development of new strategies for anticancer therapy.

Integrins are transmembrane receptors that are central to the biology of many human pathologies. Classically, integrins are known for mediating cell-ECM and cell-cell interaction, and they have been shown to have an emerging role as local activators of TGF-β, influencing cancer, fibrosis, thrombosis, and inflammation (Raab-Westphal et al., 2017). Integrins are composed of α and β subunits to form a complete signaling molecule. Their ligand binding and some regulatory sites are extracellular and sensitive to pharmacological intervention, as proven by the clinical success of seven drugs that target integrins (Hamidi et al., 2016). Although peptides and small molecules are generally designed to target integrin $\alpha\beta$ dimers, the individual integrin α subunits may also be therapeutic targets. ITGAV always bind with five β subunits that form receptors for vitronectin, cytotactin, fibronectin, fibrinogen, and laminin. ITGAV has mostly been investigated for its role in malignant tumor cells and tumor vasculature (Xiong et al., 2001; Xiong et al., 2009). ITGAV recognizes the Arg-Gly-Asp (RGD) sequence in a wide array of ligands at the interface between the α and β subunits (Xiong et al., 2002). ITGA2 forms with β_1 and belongs to the collagen receptor subfamily of integrins (Emsley et al., 2000).

The structure of ITGAV was taken from chain A of the *x*-ray structure of complete integrin $\alpha V\beta_3$ (PDB code: 3IJE). It contains a β-propeller domain of seven 60-amino-acid repeats, and three other domains including the Thigh, Calf-1, and Calf-2 domains (**Figure 7A**). The PDB repository contains no crystal structure for full-length ITGA2. The highest sequence similarity between ITGA2 and searched models (PDB code: 5ES4) was 28%, so we employed I-TASSER to generate a composite model of ITGA2 based on several templates. A subsequent analysis of the structure of ITGA2 revealed similar domain structures with ITGAV but with the addition of an *I* domain (Emsley et al., 1997) and a *WK*π *GfFkR* helix tail, which may suggest more drug-targeting possibilities for ITGA2. Based on the structures of ITGAV and ITGA2, Fpocket was used to detect their druggable pockets. For ITGAV, there were two highly druggable pockets, both located within the β-propeller domain. The largest druggable pocket was located on the outer side of the β-barrel, consisted of Val192, Lys104, Ala189, Asp132, Val188, Ala189, Asp167, Leu130, Gln187, Glu190, Lys135, Val137, and Gln131, and had a DS of 0.663 (**Figure 7A**). The second largest druggable pocket was located at the hole of the β-barrel, consisted of Trp93, Leu111, Gln156, Phe159, Pro110, Ala96, Phe21, Tyr406, Tyr224, and Phe278, and had a DS of 0.599 (**Figure S2A**). For ITGA2, only one highly druggable pocket was found at the β-propeller domain and had a DS of 0.92. This pocket consisted of His416, Phe162, His414, Ser159, Phe156, Leu417, Ser161, Val409,

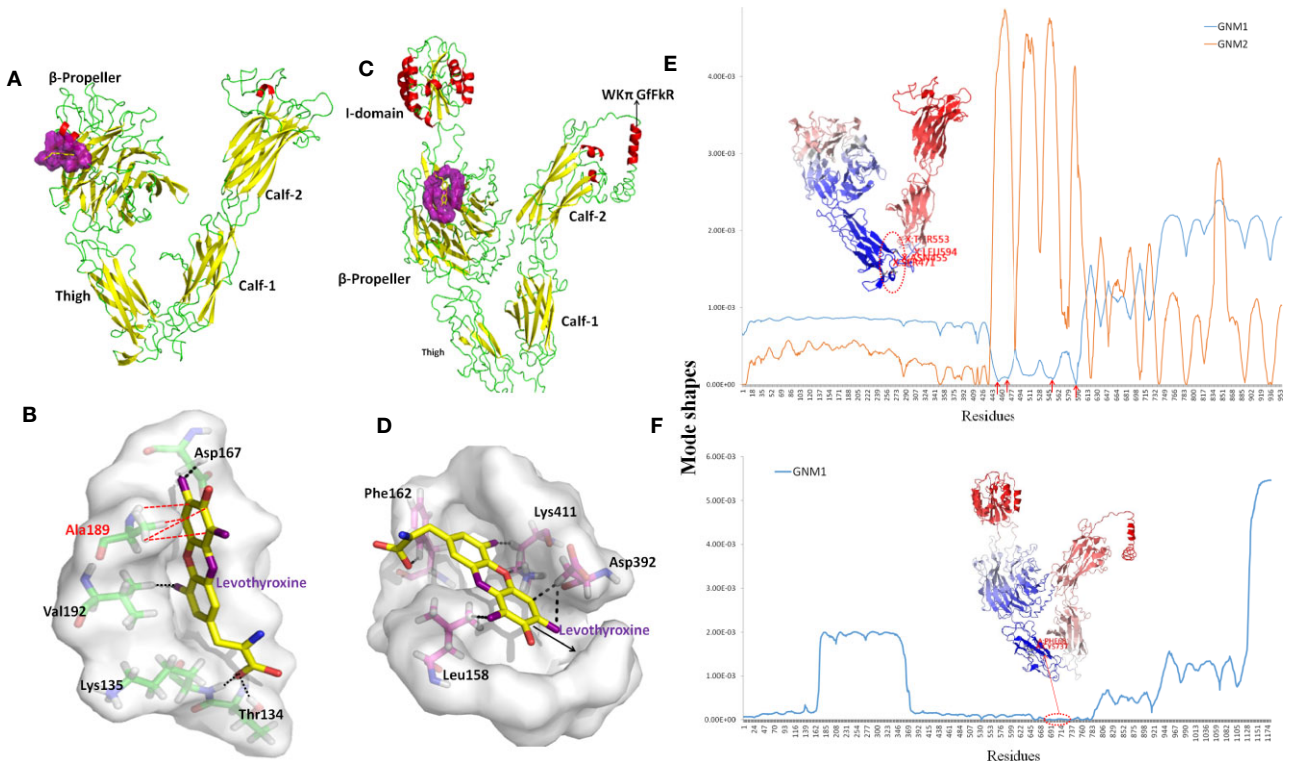


FIGURE 7 | Structures and dynamics of ITGAV and ITGA2. **(A)** The structure of ITGAV including the *β*-propeller, Thigh, Calf-1, and Calf-2 domains, and the most druggable pocket (purple), which is located along the outer side of the *β*-barrel. **(B)** The binding poses by docking Levothyroxine into the most druggable pocket of ITGAV. Levothyroxine and interacting residues are represented as colored sticks. **(C)** The structure of ITGA2 including the *I*-, *β*-propeller, Thigh, Calf-1, and Calf-2 domains, and the most druggable pocket (purple), which is located at the hole of the *β*-barrel; the binding pose with Levothyroxine and this pocket is shown in **(D)**. **(E)** The shapes of first and second GNM modes of ITGAV. The minimum of the shapes indicate the hinge region, which corresponds to the structure in dark blue. Mode 1 predicts Asn455, Ser471, Arg553, and Gly594 within the Thigh domain are hinge sites (red arrows). **(F)** The shape of the first GNM mode of ITGA2, where the region of Phe681 to Ser737 within the Thigh domain was predicted to contain hinge sites (red circle).

Leu396, Lys411, Leu158, Gln157, Leu394, Ala160, Leu417, Asp155, Asp392, Val381, Gly415, and Ser413 (**Figure 7B**).

Despite progress in the development of drugs that target different integrins, there are only two clinical approved drugs in the drug bank for ITGAV (Levothyroxine and Antithymocyte immunoglobulin) (**Table 3**). Thymoglobulin is a polyclonal antibody, while Levothyroxine is currently the only approved small molecule that targets ITGAV. The small ligand Levothyroxine was docked to the two druggable pockets in ITGAV to study the stability of the complex and protein-drug interactions. When docked to the largest druggable pocket, Levothyroxine formed hydrogen bonds with Asp167, Thr134, Lys135, and Val192, and a hydrophobic interaction with Ala189, and the binding free energy was −8.3 kcal/mol (**Figure 7C**). For the other pocket, hydrogen bonds were formed between Levothyroxine and Phe21, Trp93, Ala96, and Pro110 with the binding free energy of −10.08 kcal/mol (**Figure S2B**). We further docked Levothyroxine to ITGA2 at its druggable pocket. The binding free energy of −9.09 kcal/mol suggested a good interaction between ITGA2 and Levothyroxine, with the potential binding sites at Phe162, Lys411, Asp392, and Leu158 (**Figure 7D**).

To determine residues that play a key role in the global dynamics of ITGAV and ITGA2, we performed a GNM analysis. GNM analysis provides information on the mechanisms of collective movements intrinsically accessible to the structure, which usually enable structural changes relevant to function (Bahar et al., 2010). The most discriminative feature in dynamic analysis is hinge prediction, which are expected to be sites for drug development (Sumbul et al., 2015). We predicted hinges sites by the minima of corresponding GNM slow modes. By applying GNM to ITGAV (**Figure 7E**), GNM mode 1 highlights the hinge region located in the Thigh domain, especially at Asn455, Ser471, Arg553, and Gly594, which are located at the interface between the Thigh and Calf-1 domains. We also note that the *β*-propeller domain became the major hinge region in GNM mode 2, while Ile286, Asn287, Asp352, Phe377, Ser389, Thr413, Asp414, Pro421, and Tyr436 have minimal fluctuations. Hinge sites located at the β-propeller domain in GNM mode 2 may correspond to pocket sites, as the first and second largest druggable pockets were within the β-propeller domain. For ITGA2 (**Figure 7F**), both GNM modes 1 and 2 highlighted the same hinge regions within the β-propeller domain and the Thigh domain, with critical positioning of

Phe681 to Ser737. Accordingly, our GNM modeling suggested that both the β-propeller domain and the Thigh domain play important roles in modulating the collective movements of ITGAV and ITGA2. The β-propeller domain has been indicated to be a druggable domain by pocket detection. Here, some hinge sites located within the Thigh domain offer other reasonable starting points for inhibitor design.

CONCLUSIONS

In this study, we developed a computational framework that integrated ML (SVM-RFE), biomolecular networks (PPI network analysis), and structural modeling analysis (homology modeling, molecular docking, and GNM modeling) to help future drug targets for PDAC. The core of the new method was that we defined a new score, termed RNs, based on cancer-related information from gene expression data and topological information obtained from PPI network analysis. Research using three GEO datasets (GSE28735, GSE71989, and GSE15471) yielded 17 genes (*ADAM10*, *TIMP1*, *MATN3*, *PKM*, *APLP2*, *ACTN1*, *CALU*, *VCAN*, *LGALS1*, *ITGA2*, *BST2*, *MFGE8*, *ITGAV*, *EGF*, *APOL1*, *ALB*, and *MSLN*) that were predicted to be potential drug targets. The survival and "druggability" analysis of these genes showed that most of the identified genes had poor survival associations and good DS values, further providing evidence that they can be used as therapeutic targets in PDAC. The important roles of integrins as well as their interactions with collagens were highlighted by combining network modules and KEGG pathway analysis, in term of four pathways, ECM-receptor interaction, focal adhesion, proteoglycans in cancer, and human papillomavirus infection pathways. By focusing on ITGAV and ITGA2, we identified druggable pockets, drug binding sites, and hinge sites that are potential sites for designing small molecules. In summary, this new methodology will provide new avenues for discovering drug targets in PDAC and other cancers.

Of course, our method in this work has some limitations. Firstly, our method only used SVM-REF to the gene expression data to rank the DEGs. With the growth of other omics data, we need to apply our method by including more kinds of data, such as RNA-Seq data for PDAC (Raphael et al., 2017), which will make our method more practical. Secondly, our method just combined the systems level analysis of PPI construction and analysis and the molecular level analysis of "druggability" prediction, and thus, the drug target prediction needs some structural research experience to some extent. To address this, the real integration of structure knowledge into PPI networks is still needed.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

WY, XYL, and GH analyzed the data and wrote the manuscript. XYL and WY conducted the SVM calculation and network analysis. FW assisted in network analysis. FX and SH conducted the structural modeling and docking. XL assisted in molecular docking. WY, FX, and GH conceived and designed all experiments, and interpreted all results. GH revised the manuscript. All authors contributed to the work.

FUNDING

This study was supported by the National Natural Science Foundation of China (31872723), the Project of State Key Laboratory of Radiation Medicine and Protection, Soochow University (No. GZK1201902), and a project funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.00534/full#supplementary-material

REFERENCES

Al Haddad, A. H., and Adrian, T. E. (2014). Challenges and future directions in therapeutics for pancreatic ductal adenocarcinoma. *Expert Opin. Invest. Drugs* 23 (11), 1499–1515. doi: 10.1517/13543784.2014.933206

Almeida, P. P., Cardoso, C. P., and de Freitas, L. M. (2020). PDAC-ANN: an artificial neural network to predict pancreatic ductal adenocarcinoma based on gene expression. *BMC Cancer* 20 (1), 82. doi: 10.1186/s12885-020-6533-0

Badea, L., Herlea, V., Dima, S. O., Dumitrascu, T., and Popescu, I. (2008). Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology* 55 (88), 2016–2027.

Bahar, I., Lezon, T. R., Yang, L. W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39, 23–42. doi: 10.1146/annurev.biophys.093008.131258

Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531 (7592), 47–52. doi: 10.1038/nature16965

Bhasin, M. K., Ndebele, K., Bucur, O., Yee, E. U., Otu, H. H., Plati, J., et al. (2016). Meta-analysis of transcriptome data identifies a novel 5-gene pancreatic adenocarcinoma classifier. *Oncotarget* 7 (17), 23263–23281. doi: 10.18632/oncotarget.8139

Bólon-Canedo, V., Sánchez-Maroñoa, N., Alonso-Betanzosa, A., Benítezb, J. M., and Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Inf. Sci.* 282, 111–135. doi: 10.1016/j.ins.2014.05.042

Carr, R. M., and Fernandez-Zapico, M. E. (2019). Toward personalized TGFbeta inhibition for pancreatic cancer. *EMBO Mol. Med.* 11 (11), e11414. doi: 10.15252/emmm.201911414

Cheng, T. Y., Yang, Y. C., Wang, H. P., Tien, Y. W., Shun, C. T., and Huang, H. Y. (2018). Pyruvate kinase M2 promotes pancreatic ductal adenocarcinoma

invasion and metastasis through phosphorylation and stabilization of PAK2 protein. *Oncogene* 37 (13), 1730–1742. doi: 10.1038/s41388-017-0086-y

Conte, F., Fiscon, G., Licursi, V., Bizzarri, D., D'Antò, T., Farina, L., et al. (2019). A paradigm shift in medicine: A comprehensive review of network-based approaches. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863 (6), 194416. doi: 10.1016/j.bbagrm.2019.194416

Conway, J. R., Herrmann, D., Jeffry Evans, T. R., Morton, J. P., and Timpson, P. (2019). Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine. *Gut* 68 (4), 742–758. doi: 10.1136/gutjnl-2018-316822

Craven, K. E., Gore, J., Wilson, J. L., and Korc, M. (2016). Angiogenic gene signature in human pancreatic cancer correlates with TGF-beta and inflammatory transcriptomes. *Oncotarget* 7 (1), 323–341. doi: 10.18632/oncotarget.6345

Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138 (3), 333–408. doi: 10.1016/j.pharmthera.2013.01.016

Danne, R., Poojari, C., Martinez-Seara, H., Rissanen, S., Lolicato, F., and Róg, T. (2017). doGlycans-Tools for Preparing Carbohydrate Structures for Atomistic Simulations of Glycoproteins, Glycolipids, and Carbohydrate Polymers for GROMACS. *J. Chem. Inf. Model.* 57 (10), 2401–2406. doi: 10.1021/acs.jcim.7b00237

Duran-Frigola, M., Mosca, R., and Aloy, P. (2013). Structural systems pharmacology: the role of 3D structures in next-generation drug development. *Chem. Biol.* 20 (5), 674–684. doi: 10.1016/j.chembiol.2013.03.004

Emsley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997). Crystal structure of the I domain from integrin alpha2beta1. *J. Biol. Chem.* 272 (45), 28512–28517. doi: 10.1074/jbc.272.45.28512

Emsley, J., Knight, C. G., Farndale, R. W., and Barnes, M. J. (2000). Structural basis of collagen recognition by integrin alpha2beta1. *Cell* 101 (1), 47–56. doi: 10.1016/S0092-8674(00)80622-4

Fotis, C., Antoranz, A., Hatziavramidis, D., Sakellaropoulos, T., and Alexopoulos, L. G. (2018). Network-based technologies for early drug discovery. *Drug Discovery Today* 23 (3), 626–635. doi: 10.1016/j.drudis.2017.12.001

Freeman, L. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40 (1), 35–41. doi: 10.2307/3033543

Gaida, M. M., Haag, N., Günther, F., Tschaharganeh, D. F., Schirmacher, P., Friess, H., et al. (2010). Expression of A disintegrin and metalloprotease 10 in pancreatic carcinoma. *Int. J. Mol. Med.* 26 (2), 281–288. doi: 10.3892/ijmm_00000463

Grapa, C. M., Mocan, T., Gonciar, D., Zdrehus, C., Mosteanu, O., and Pop, T. (2019). Epidermal Growth Factor Receptor and Its Role in Pancreatic Cancer Treatment Mediated by Nanoparticles. *Int. J. Nanomed.* 14, 9693–9706. doi: 10.2147/IJN.S226628

Grutzmann, R., Boriss, H., Ammerpohl, O., Lüttges, J., Kalthoff, H., Schackert, H. K., et al. (2005). Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene* 24 (32), 5079–5088. doi: 10.1038/sj.onc.1208696

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46 (1-3), 389–422. doi: 10.1023/A:1012487302797

Hamidi, H., Pietila, M., and Ivaska, J. (2016). The complexity of integrins in cancer and new scopes for therapeutic targeting. *Br. J. Cancer* 115 (9), 1017–1023. doi: 10.1038/bjc.2016.312

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics* 15 (1), 41–51. doi: 10.21873/cgp.20063

Jeon, J., Nim, S., Teyra, J., Datti, A., Wrana, J. L., Sidhu, S. S., et al. (2014). A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med.* 6 (7), 57. doi: 10.1186/s13073-014-0057-7

Jiang, J., Azevedo-Pouly, A. C., Redis, R. S., Lee, E. J., Gusev, Y., and Allard, D. (2016). Globally increased ultraconserved noncoding RNA expression in pancreatic adenocarcinoma. *Oncotarget* 7 (33), 53165–53177. doi: 10.18632/oncotarget.10242

Kandoi, G., Acencio, M. L., and Lemke, N. (2015). Prediction of Druggable Proteins Using Machine Learning and Systems Biology: A Mini-Review. *Front. Physiol.* 6, 366. doi: 10.3389/fphys.2015.00366

Kar, G., Gursoy, A., and Keskin, O. (2009). Human cancer protein-protein interaction network: a structural perspective. *PloS Comput. Biol.* 5 (12), e1000601. doi: 10.1371/journal.pcbi.1000601

Khodade, P., Prabhu, R., Chandra, N., Raha, S., and Govindarajan, R. (2007). Parallel implementation of AutoDock. *J. Appl. Crystallogr.* 40, 598–599. doi: 10.1107/S0021889807011053

Kinsey, C. G., Camolotto, S. A., Boespflug, A. M., Guillen, K. P., Foth, M., Truong, A., et al. (2019). Protective autophagy elicited by RAF–>MEK–>ERK inhibition suggests a treatment strategy for RAS-driven cancers. *Nat. Med.* 25 (4), 620–627. doi: 10.1038/s41591-019-0367-9

Kuninty, P. R., Bansal, R., De Geus, S. W. L., Mardhian, D. F., Schnittert, J., and van Baarlen, J. (2019). ITGA5 inhibition in pancreatic stellate cells attenuates desmoplasia and potentiates efficacy of chemotherapy in pancreatic cancer. *Sci. Adv.* 5 (9)eaax2770. doi: 10.1126/sciadv.aax2770.

Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinf.* 10, 168. doi: 10.1186/1471-2105-10-168

Li, S., Yu, X., Zou, C., Gong, J., Liu, X., and Li, H. (2016). Are Topological Properties of Drug Targets Based on Protein-Protein Interaction Network Ready to Predict Potential Drug Targets? *Comb. Chem. High Throughput Screen* 19 (2), 109–120. doi: 10.2174/1386207319666151110122145

Li, S., Su, Z., Zhang, C., Xu, Z., Chang, X., Zhu, J., et al. (2018). Identification of drug target candidates of the swine pathogen Actinobacillus pleuropneumoniae by construction of protein-protein interaction network. *Genes Genomics* 40 (8), 847–856. doi: 10.1007/s13258-018-0691-3

Li, X., Li, W., Zeng, M., Zheng, R., and Li, M. (2019). Network-based methods for predicting essential genes or proteins: a survey. *Brief Bioinform*. doi: 10.1093/bib/bbz017

Liu, X., Zheng, W., Wang, W., Shen, H., Liu, L., and Lou, W. (2017). A new panel of pancreatic cancer biomarkers discovered using a mass spectrometry-based pipeline. *Br. J. Cancer* 117 (12), 1846–1854. doi: 10.1038/bjc.2017.365

Liu, L., Chen, X., Hu, C., Zhang, D., Shao, Z., and Jin, Q. (2018). Synthetic Lethality-based Identification of Targets for Anticancer Drugs in the Human Signaling Network. *Sci. Rep.* 8 (1), 8440. doi: 10.1038/s41598-018-26783-w

Lu, Y., Li, C., Chen, H., and Zhong, W. (2018). Identification of hub genes and analysis of prognostic values in pancreatic ductal adenocarcinoma by integrated bioinformatics methods. *Mol. Biol. Rep.* 45 (6), 1799–1807. doi: 10.1007/s11033-018-4325-2

Mishra, N. K., Southekal, S., and Guda, C. (2019). Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma. *Front. Genet.* 10, 624. doi: 10.3389/fgene.2019.00624

Mitsopoulos, C., Schierz, A. C., Workman, P., and Al-Lazikani, B. (2015). Distinctive Behaviors of Druggable Proteins in Cellular Networks. *PloS Comput. Biol.* 11 (12), e1004597. doi: 10.1371/journal.pcbi.1004597

Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69 (2), 026113. doi: 10.1103/PhysRevE.69.026113

Nones, K., Waddell, N., Song, S., Patch, A. M., Miller, D., Johns, A., et al. (2014). Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int. J. Cancer* 135 (5), 1110–1118. doi: 10.1002/ijc.28765

Pandey, P., Rachagani, S., Das, S., Seshacharyulu, P., Sheinin, Y., and Naslavsky, N. (2015). Amyloid precursor-like protein 2 (APLP2) affects the actin cytoskeleton and increases pancreatic cancer growth and metastasis. *Oncotarget* 6 (4), 2064–2075. doi: 10.18632/oncotarget.2990

Parkin, A., Man, J., Timpson, P., and Pajic, M. (2019). Targeting the complexity of Src signalling in the tumour microenvironment of pancreatic cancer: from mechanism to therapy. *FEBS J.* 286 (18), 3510–3539. doi: 10.1111/febs.15011

Pham, D. H., Kim, J. S., Kim, S. K., Shin, D. J., Uong, N. T., and Hyun, H. (2017). Effects of ADAM10 and ADAM17 Inhibitors on Natural Killer Cell Expansion and Antibody-dependent Cellular Cytotoxicity Against Breast Cancer Cells In Vitro. *Anticancer Res.* 37 (10), 5507–5513. doi: 10.21873/anticanres.11981

Prokopchuk, O., Grünwald, B., Nitsche, U., Jäger, C., Prokopchuk, O. L., and Schubert, E. C. (2018). Elevated systemic levels of the matrix metalloproteinase inhibitor TIMP-1 correlate with clinical markers of cachexia in patients with chronic pancreatitis and pancreatic cancer. *BMC Cancer* 18 (1), 128. doi: 10.1186/s12885-018-4055-9

Qin, J. J., Li, X., Wang, W., Zi, X., and Zhang, R. (2017). Targeting the NFAT1-MDM2-MDMX Network Inhibits the Proliferation and Invasion of Prostate Cancer Cells, Independent of p53 and Androgen. *Front. Pharmacol.* 8, 917. doi: 10.3389/fphar.2017.00917

Raab-Westphal, S., Marshall, J. F., and Goodman, S. L. (2017). Integrins as Therapeutic Targets: Successes and Cancers. *Cancers (Basel)* 9 (9), 110. doi: 10.3390/cancers9090110

Rajamani, D., and Bhasin, M. K. (2016). Identification of key regulators of pancreatic cancer progression through multidimensional systems-level analysis. *Genome Med.* 8 (1), 38. doi: 10.1186/s13073-016-0282-3

Raphael, B. J., Hruban, R. H., Aguirre,, A. J., Moffitt, R. A., Yeh, J. J., Stewart, C., et al. (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 32 (2), 185–18+. doi: 10.1016/j.ccell.2017.07.007

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5 (4), 725–738. doi: 10.1038/nprot.2010.5

Selvaraj, G., Kaliamurthi, S., Kaushik, A. C., Khan, A., Wei, Y. K., and Cho, W. C. (2018). Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. *J. BioMed. Inform* 86, 120–134. doi: 10.1016/j.jbi.2018.09.004

Shi, Y., Gao, W., Lytle, N. K., Huang, P., Yuan, X., and Dann, A. M. (2019). Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* 569 (7754), 131–135. doi: 10.1038/s41586-019-1130-6

Sinkala, M., Mulder, N., and Martin, D. (2020). Machine Learning and Network Analyses Reveal Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. *Sci. Rep.* 10 (1), 1212. doi: 10.1038/s41598-020-58290-2

Skandalis, S. S., Kletsas, D., Kyriakopoulou, D., Stavropoulos, M., and Theocharis, D. A. (2006). The greatly increased amounts of accumulated versican and decorin with specific post-translational modifications may be closely associated with the malignant phenotype of pancreatic cancer. *Biochim. Biophys. Acta* 1760 (8), 1217–1225. doi: 10.1016/j.bbagen.2006.03.021

Sonawane, A. R., Weiss, S. T., Glass, K., and Sharma, A. (2019). Network Medicine in the Age of Biomedical Big Data. *Front. Genet.* 10, 294. doi: 10.3389/fgene.2019.00294

Sumbul, F., Acuner-Ozbabacan, S. E., and Haliloglu, T. (2015). Allosteric Dynamic Control of Binding. *Biophys. J.* 109 (6), 1190–1201. doi: 10.1016/j.bpj.2015.08.011

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., and Simonovic, M. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Tang, S. C., and Chen, Y. C. (2014). Novel therapeutic targets for pancreatic cancer. *World J. Gastroenterol.* 20 (31), 10825–10844. doi: 10.3748/wjg.v20.i31.10825

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., and Bidkhori, G. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357 (6352) eaan2507. doi: 10.1126/science.aan2507

Villani, V., Thornton, M. E., Zook, H. N., Crook, C. J., Grubbs, B. H., and Orlando, G. (2019). SOX9+/PTF1A+ Cells Define the Tip Progenitor Cells of the Human Fetal Pancreas of the Second Trimester. *Stem Cells Transl. Med.* 8 (12), 1249–1264. doi: 10.1002/sctm.19-0231

Vinayagam, A., Gibson, T. E., Lee, H. J., Yilmazel, B., Roesel, C., Hu, Y., et al. (2016). Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proc. Natl. Acad. Sci. U. S. A* 113 (18), 4976–4981. doi: 10.1073/pnas.1603992113

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31 (1), 258–261. doi: 10.1093/nar/gkg034

Waddell, N., Pajic, M., Patch, A. M., Chang, D. K., and Kassahn, K. S. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 518 (7540), 495–501. doi: 10.1038/nature14169

Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., and Chen, C. F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23 (10), 1274–1281. doi: 10.1093/bioinformatics/btm087

Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z, et al.. (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* 48 (D1), D1031–D1041. doi: 10.1093/nar/gkz981

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., and Gumienny, R. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46 (W1), W296–W303. doi: 10.1093/nar/gky427

Winter, C., Henschel, A., Tuukkanen, A., and Schroedera, M. (2012). Protein interactions in 3D: from interface evolution to drug discovery. *J. Struct. Biol.* 179 (3), 347–358. doi: 10.1016/j.jsb.2012.04.009

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., and Grant, J. R. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46 (D1), D1074–D1082. doi: 10.1093/nar/gkx1037

Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., and Scott, D. L. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294 (5541), 339–345. doi: 10.1126/science.1064535

Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., and Goodman, S. L. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 296 (5565), 151–155. doi: 10.1126/science.1069040

Xiong, J. P., Mahalingham, B., Alonso, J. L., Borrelli, L. A., Rui, X., and Anand, S. (2009). Crystal structure of the complete integrin alphaVbeta3 ectodomain plus an alpha/beta transmembrane fragment. *J. Cell Biol.* 186 (4), 589–600. doi: 10.1083/jcb.200905085

Xu, S., Xu, H., Wang, W., Li, S., Li, H., and Li, T. (2019). The role of collagen in cancer: from bench to bedside. *J. Transl. Med.* 17 (1), 309. doi: 10.1186/s12967-019-2058-1

Yan, W., Zhang, D., Shen, C., Liang, Z., and Hu, G. (2018). Recent Advances on the Network Models in Target-based Drug Discovery. *Curr. Top. Med. Chem.* 18 (13), 1031–1043. doi: 10.2174/1568026618666180719152258

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Zhang, G., Schetter, A., He, P., Funamizu, N., Gaedcke, J., and Ghadimi, B. M. (2012). DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PloS One* 7 (2), e31507. doi: 10.1371/journal.pone.0031507

Zhang, G., He, P., Tan, H., Budhu, A., Gaedcke, J., and Ghadimi, B. M. (2013). Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin. Cancer Res.* 19 (18), 4983–4993. doi: 10.1158/1078-0432.CCR-13-0209

Zhang, P., Tao, L., Zeng, X., Qin, C., Chen, S., and Zhu, F. (2017). A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. *Brief Bioinform.* 18 (6), 1057–1070. doi: 10.1093/bib/bbw071

Zhao, X., and Liu, Z. P. (2019). Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network. *Genes (Basel)* 10 (2), 143. doi: 10.3390/genes10020143

Zhavoronkov, A. (2018). Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry. *Mol. Pharm.* 15 (10), 4311–4313. doi: 10.1021/acs.molpharmaceut.8b00930

Zhou, Y. Y., Chen, L. P., Zhang, Y., Hu, S. K., Dong, Z. J., and Wu, M. (2019). Integrated transcriptomic analysis reveals hub genes involved in diagnosis and prognosis of pancreatic cancer. *Mol. Med.* 25 (1), 47. doi: 10.1186/s10020-019-0113-2

Zhu, W., Chen, X., Ning, L., and Jin, K. (2019). Network Analysis Reveals TNF as a Major Hub of Reactive Inflammation Following Spinal Cord Injury. *Sci. Rep.* 9 (1), 928. doi: 10.1038/s41598-018-37357-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yan, Liu, Wang, Han, Wang, Liu, Xiao and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

REVIEW
published: 07 May 2020
doi: 10.3389/fonc.2020.00724

Extracellular Vesicles in Renal Cell Carcinoma: Multifaceted Roles and Potential Applications Identified by Experimental and Computational Methods

*Zhiyuan Qin, Qingwen Xu, Haihong Hu, Lushan Yu and Su Zeng**

College of Pharmaceutical Sciences, Institute of Drug Metabolism and Pharmaceutical Analysis, Zhejiang University, Hangzhou, China

Renal cell carcinoma (RCC) is the most common type of kidney cancer. Increasingly evidences indicate that extracellular vesicles (EVs) orchestrate multiple processes in tumorigenesis, metastasis, immune evasion, and drug response of RCC. EVs are lipid membrane-bound vesicles in nanometer size and secreted by almost all cell types into the extracellular milieu. A myriad of bioactive molecules such as RNA, DNA, protein, and lipid are able to be delivered via EVs for the intercellular communication. Hence, the abundant content of EVs is appealing reservoir for biomarker identification through computational analysis and experimental validation. EVs with excellent biocompatibility and biodistribution are natural platforms that can be engineered to offer achievable drug delivery strategies for RCC therapies. Moreover, the multifaceted roles of EVs in RCC progression also provide substantial targets and facilitate EVs-based drug discovery, which will be accelerated by using artificial intelligence approaches. In this review, we summarized the vital roles of EVs in occurrence, metastasis, immune evasion, and drug resistance of RCC. Furthermore, we also recapitulated and prospected the EVs-based potential applications in RCC, including biomarker identification, drug vehicle development as well as drug target discovery.

Keywords: renal cell carcinoma, extracellular vesicles, exosomes, biomarkers, drug targets, drug vehicles, artificial intelligence, machine learning

OPEN ACCESS

Edited by:
Yu Zong Chen, National University of Singapore, Singapore

Reviewed by:
Ying Hong Li, Chongqing University of Posts and Telecommunications, China
Xiaoqiang Xiang, Fudan University, China
Mariangela Garofalo, University of Padova, Italy

****Correspondence:***
Su Zeng
zengsu@zju.edu.cn

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Oncology

Received: *18 February 2020*
Accepted: *16 April 2020*
Published: *07 May 2020*

Citation:
Qin Z, Xu Q, Hu H, Yu L and Zeng S (2020) Extracellular Vesicles in Renal Cell Carcinoma: Multifaceted Roles and Potential Applications Identified by Experimental and Computational Methods. Front. Oncol. 10:724. doi: 10.3389/fonc.2020.00724

INTRODUCTION

Renal cell carcinoma, or RCC for short, is one of the most common type of urological cancers that represents ∼90% of all kidney malignancies (1). According to updated data provided by the World Health Organization, over 400,000 people were diagnosed with kidney cancer worldwide in 2018, accounting for nearly 3% of all cancers (2). It has been estimated that there will be about 74,000 new cases and 15,000 deaths associated with kidney cancer in the United States in 2020 (3). The 5-year survival rate among RCC patients increased for decades due to the improvement of early-detection techniques and targeted-therapies. The current overall 5-year survival rate of RCC is 75%, decreasing to 70% among patients with regional metastases and 12% among patients with distant metastases (4). Still around one-third of patients diagnosed with RCC had metastases

(5). The most common metastatic sites of RCC are lungs, bone, brain, lymph node, and liver might also be involved (6). Surgery is the mainstay curative treatment for localized RCC (7). However, around 40% RCC patients will suffer tumor recurrence after curative surgical resection (8). For patients who present with metastatic RCC or relapses after local therapy, typically require systemic treatment. The current landscape of systemic therapies are consist of small molecule kinase inhibitors, cytokines, and monoclonal antibodies, including checkpoint inhibitors, which have been tested as first-line or second-line therapies (9).

Extracellular vesicles (EVs) are nanometer sized vesicles composed of a lipid bilayer membrane packaging a wealth of bioactive molecules such as RNA, DNA, protein, and lipid. Currently, EVs can be broadly divided into two main types based on the mechanism of biogenesis: one is exosomes which originate from the endosomal system and another one is microvesicles that directly shed from the plasma membrane (10). As Thery et al. mentioned in a review, both exosomes and microvesicles may be co-isolated due to the overlapping characteristics between these two forms of EVs and the limitations of current isolation methods. Therefore, the term exosomes is generally used in literatures to designate a mixed population of EVs without adequate characterization of the intracellular origin (11). Hereafter, we chose to use the generic term "EVs" in this review independent of the term used in the original articles.

With the nanoscale size and double-layered lipid membrane appropriately protecting the cargoes from degradation, EVs stably exist in blood, urine, saliva, and many other kinds of biological fluids. Accumulating evidences indicate that EVs traffic between donor and recipient cells are fundamental phenomenon of the intercellular information exchange, especially in tumor microenvironment (TME). EVs within TME are emerging as crucial contributor to carcinogenesis, angiogenesis, premetastatic niche (PMN) formation, dysfunction of immune system and the dissemination of anti-cancer drugs resistance, adding novel dimension to the complexity of TME (12). Thus, the contents of tumor-derived EVs may be applied as abundant sources to biomarker discovery identified by experimental and computational methods. In addition, EVs with naturally excellent biocompatibility and biodistribution are ideal materials to be exploited or engineered which may offer us achievable drug delivery strategies for cancer therapies (13). Furthermore, it is increasingly clear that mechanisms of EVs biogenesis, secretion and uptake could also provide promising targets for cancer therapy (14).

The past decades have witnessed unprecedented research progresses of EVs, especially for the roles of EVs in different malignant tumors. Nevertheless, to the best of our knowledge, few researchers paid close attention to the roles of EVs in urological malignancies, especially for RCC (15–22). There is still no comprehensive summary highlighting the EVs-based potential applications in RCC either. Hence, this review serves to introduce the latest research progresses in the burgeoning field of EVs, recapitulate the multifaceted functions of EVs in RCC progression. Accordingly, we will also give a perspective of the potential applications of EVs in RCC identified by both experimental and computational methods.

BIOLOGICAL FEATURES OF EVS AND RESEARCH TECHNIQUES

Biogenesis, Secretion and Uptake of EVs

The biogenesis of two EVs subtypes are different as shown in **Figure 1**. Diameter of microvesicles range from 50 to 1,000 nm but can up to 10 μm in the case of oncosomes, which refers to cancer cells-derived microvesicles that contain oncogenic molecules (10, 23). Microvesicles are generated through the direct budding and fission of the cytoplasmic membrane then released into the extracellular space (24). Exosomes originate from multivesicular bodies (MVBs) within endosomal system, ranging from 30 to 150 nm. The endosomal membrane invaginate intraluminal vesicles (ILVs) in the lumen during the mature process of early endosomes into late endosomes or MVBs. The endosomal sorting complex required for transport (ESCRT) machinery plays critical role in this process (10, 25). Moreover, members of the Rab GTPases family, including Rab27a/b, Rab11, and Rab35, are essential coordinators for MVBs trafficking and exosomes secretion (26, 27). The last step of secretion requires the fusion of MVBs with plasma membrane. This process primarily is mediated by soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) and synaptotagmin family members to release ILVs as exosomes (28). Several studies have also found that Ca^{2+} may be involved in the activation of SNAREs (29, 30).

Once secreted into the extracellular milieu and absorbed by recipient cells, EVs cargoes can be transmitted to recipient cells to induce functional responses and confer new properties then result in phenotypic changes (10). This EVs-mediated interaction requires docking at the plasma membrane of recipient cells via several mediators such as clathrin, tetraspanins, and integrins to activate surface receptors and signaling pathways, being followed by vesicle endocytosis or membrane fusion of recipient cells (10, 31–33). The secretion processes of EVs are evolutionarily conserved among eukaryotes, bacteria, and archaea, which lay the foundation for interspecies transfer of genetic molecules via EVs (34). However, the whole process of exosomes biogenesis and secretion may be influenced by the heterogeneity of donor and recipient cells, different physiological or pathological conditions, making the detailed mechanisms remains elusive (35, 36).

EVs Composition

Diverse bioactive molecules such as RNA, DNA, proteins, and lipids can be packaged into EVs and secreted out of cell membrane at both local regional and systemic levels (37). A "routine passenger" of EVs is RNA. Both mRNA and microRNA (miRNA) could be loaded and transported through EVs then functioned in recipient cells (38–40). Besides, numerous long non-coding RNA (lncRNA) could also be transferred via EVs, inducing signals and phenotypes changes in a variety of cells in TME (41, 42). Furthermore, more than 1,000 circular RNA (circRNA) were identified in EVs derived from human serum. Interestingly, several circRNAs were highly enriched in EVs compared to the donor cells, which may provide more achievable applications in biomarker discovery (43, 44). Other RNA species were also detected in EVs by RNA deep sequencing analysis,

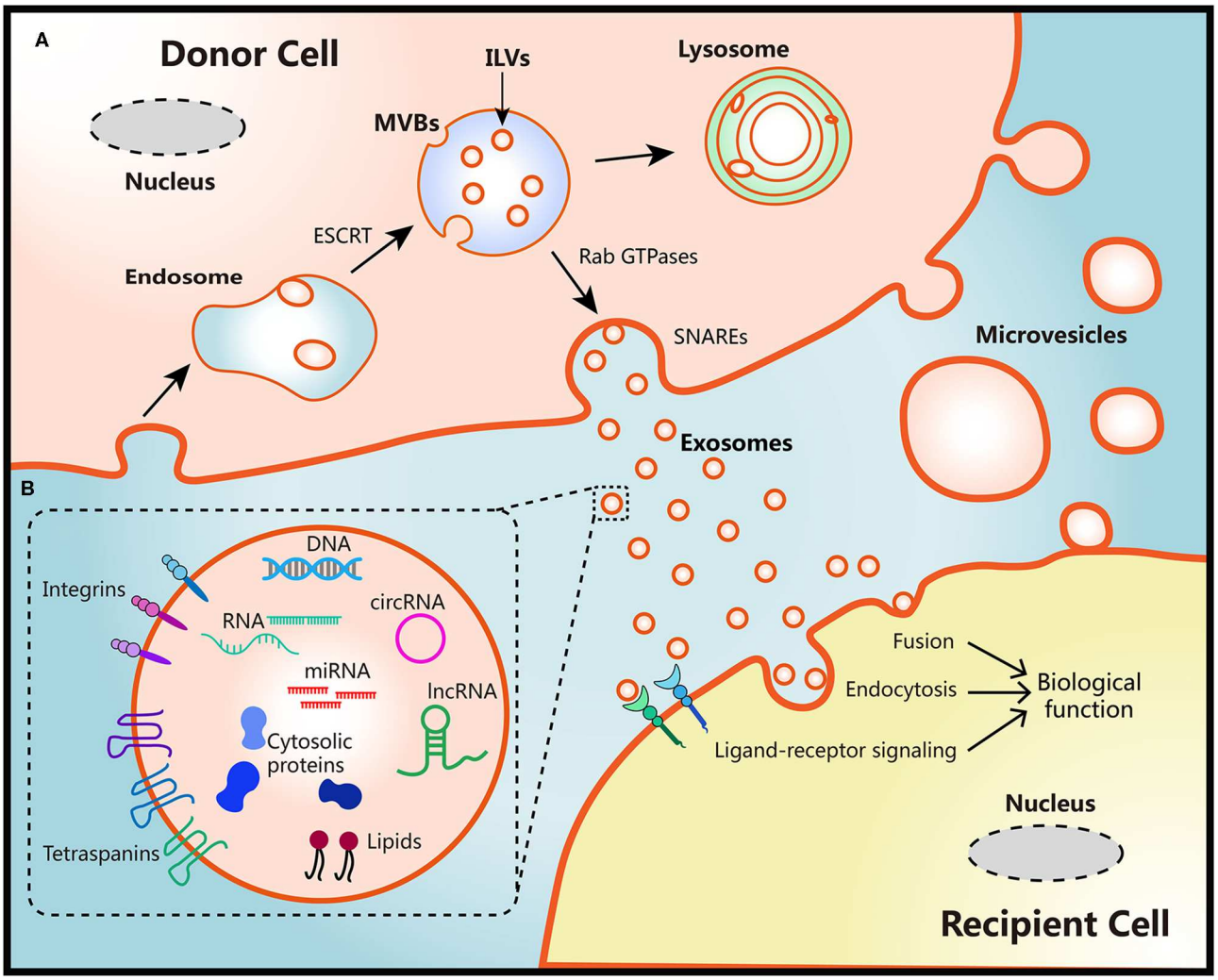


FIGURE 1 | Schematic diagram of the biological features of EVs. **(A)** Biogenesis, secretion and uptake of EVs. During the process of early endosome mature into MVBs, the endosomal membrane invaginate ILVs in the lumen of donor cells, which mediated by the ESCRT machinery. MVB fuse with cell surface and release ILVs as exosomes or degrade in lysosomes. Protein members of Rab GTPases, SNAREs, and synaptotagmin family play vital roles in MVBs trafficking and exosomes secretion. Microvesicles originate from the plasma membrane of donor cells directly. There are three ways to uptake EVs and induce biological functions in recipient cells: fusion with membrane of recipient cells directly, internalization by endocytosis, or activation of ligand-receptor signaling. **(B)** Representative structure and composition of EVs. EVs are nanometer sized vesicles composed of a lipid bilayer membrane. Size of exosomes range from 30 to 150 nm, Diameter of microvesicles range from 50 to 1,000 nm but can up to 10 μm in the case of oncosomes. EVs package various bioactive molecules such as RNA, DNA, proteins, and lipids. Transmembrane including integrins and tetraspanins are also contained in EVs.

including transfer RNA, ribosomal RNA and piwi-interacting RNA (44, 45).

The presence of DNA within EVs also provide novel insights into the cellular homeostasis and open another intriguing mode of intracellular communication (46). It has been reported that EVs secretion removed various length of chromosomal DNA fragments which were harmful to normal human cells (47). Moreover, studies demonstrated that retrotransposon elements, oncogene amplifications, and other functional DNA fragments that reflected the genetic status of the parent tumor cells were found in EVs (48, 49). Notably, these transposable elements could be encapsulated and transferred from tumor cells to normal cells (50). Thereby it can be inferred that tumor-derived EVs may function as novel mediators of horizontal gene transfer and make contribution to tumor evolution in local or systematical level (51).

As a consequence of the biogenesis, EVs derived from different cell types contain substantial cytosolic proteins, such as Rab GTPase, SNAREs, and Annexins (52). Tetraspanins is a highly conserved family of transmembrane proteins which have been found in EVs from diverse cell types. It is believed that tetraspanins interact or coordinate with other proteins and involve in membrane compartmentalization (53). Members of this family, including CD9, CD63, and CD81, consist part of

the most abundant proteins in EVs, thus commonly be used as protein markers for EVs characterization (54). In addition, increasing evidences have demonstrated the presence of several transporters and enzymes in EVs with full activity (55–57). Thus, it can be inferred that the change of EVs components can be connected with the *in vivo* fate of drugs.

EVs Isolation and Characterization

Since research field of EVs has achieved high-speed development in the past few decades, many techniques have been used to isolate and characterize EVs. At present, the frequently used techniques for EVs isolation can be summarized into five broad categories: differential ultracentrifugation (UC), polymer-based precipitation, particle size-based techniques, immunological capture, and microfluidic techniques (58). As one of the most traditionally and widely used method, differential UC is suitable for most sources of EVs, even though it is laborious, time-consuming, and inaccessible. Several commercial isolation kits are developed based on above theories and techniques to isolate EVs more efficiently and precisely. However, according to results of a recent benchmark study, a large quantity of non-vesicular contaminants may be co-isolated by these kits. While the purity of EVs isolated by differential UC was much higher than commercial kits (59). More recently, microfluidic-based platforms have generated heightened interest. Based on specific capture of the surface marker or the specific size and density of EVs subsets, microfluidic-based platform can provide advantages such as low consumption, ready portability, with high throughput, and high precision (60). Since there is still no consensus on a "gold standard" method for EVs isolation and purification, comparison study is still needed to analyze the parameters of EVs isolated by different methods. According to a global survey in 2015 conducted by the International Society for Extracellular Vesicles (ISEV), around 81% of respondents chose differential UC as their primary isolation method, around 59% of respondents used a combination methods of differential UC with other techniques (61). In terms of EVs characterization, multiple techniques based on biophysics and molecular biology have been developed and applied. Three of the most common methods are western blotting for identification of specific protein marker, electron microscopy for detection of structural information and nanoparticle tracking analysis for quantification of EVs size and concentration, respectively. Generally speaking, two or more complementary methods are necessary to assess the results of separation methods as ISEV recommended (62).

ROLES OF EVS IN RCC

EVs is employed by tumor cells to deliver bioactive molecules directing to not only tumor cells but also tumor-associated cells including fibroblasts, endothelial cells, immune cells, and cancer stem cells (CSCs) (63, 64). Reciprocally, EVs derived from non-tumor cells also have influence on tumor progression in TME. Therefore, these multidirectional communications via EVs make TME becoming a more complex network, which draw accumulating attention of researchers in recent years. Herein we reviewed the latest studies about roles of EVs in carcinogenesis, cancer metastasis, immune evasion, and drug resistance of RCC (**Figure 2**).

Tumorigenesis

EVs secreted by different cells in TME may make contributions to RCC progression and development. Jiang et al. revealed that EVs secreted by RCC cell line OS-RC-2 could inhibit hepaCAM expression, a tumor suppressor frequently lost in various types of human cancers, and promote cell proliferation in a p-AKT-dependent pathway (65). By use of cell culture and nude mice xenograft model, Du et al. claimed that EVs released by human Wharton's jelly mesenchymal stem cells induced HGF expression, activated AKT and ERK1/2 signaling pathways, then promoted the proliferation and aggressiveness of RCC cells both *in vitro* and *in vivo* (66). By using next-generation sequencing, Song et al. found the levels of EVs-contained miR-30c-5p in RCC cell lines 786-O and ACHN were significant lower than that in human renal proximal tubular cell line HK-2. Consistently, the expression pattern of miR-30c-5p was significant different in urinary EVs from healthy controls and patients of clear cell RCC (ccRCC), which is the predominant RCC type. Heat-shock protein 5 was identified as a direct target of miR-30c-5p. Gain-of-function study showed that overexpression of miR-30c-5p inhibited ccRCC progression both *in vitro* and *in vivo* (67). Considered together, these data suggest that EVs may transfer various cargoes between heterogeneous cells within TME, initiate the critical regulation of the tumorigenesis to support the growth of RCC cells.

Hypoxia is one of the distinguishing features of TME in many solid tumors including RCC. Carbonic anhydrase IX (CAIX), a cellular hypoxia biomarker that overexpress in RCC with *von Hippel-Lindau* (*VHL*) gene mutation, is involved in proliferation and transformation of RCC cells (68). It has been revealed that abundant CAIX proteins were detected in EVs released from RCC cell lines. Result of *in vitro* angiogenesis assays demonstrated that hypoxic RCC cells could release EVs containing CAIX and promote the migration and tube formation abilities of human umbilical vein endothelial cells (69). Several researchers have also provided direct evidences that hypoxia not only regulated the tumorigenic potential of epithelial cells, but also contributed to EVs production of tumor cells in response to low pH and oxidative stress (70, 71). Wang et al. reported that acute hypoxia condition induced by $CoCl_2$ treatment upregulated miR-210 expression in EVs which derived from both normal renal cells and RCC cells, especially for metastatic RCC cell line (72). Interestingly, EVs secreted by hypoxic cells are more easily absorbed by hypoxic cells (73). Hitherto, there is limited knowledge about the mechanism of how hypoxia orchestrate the biogenesis and secretion of EVs. Nevertheless, it can be concluded that hypoxia-induced EVs derived from stromal and tumor cells are crucial mediator in the process of tumorigenesis and TME rebuilding.

Tumor Metastasis

Recent years, numerous investigations have revealed the significant influence of EVs on both regional and distant metastatic processes, including coagulation, vascular leakiness,

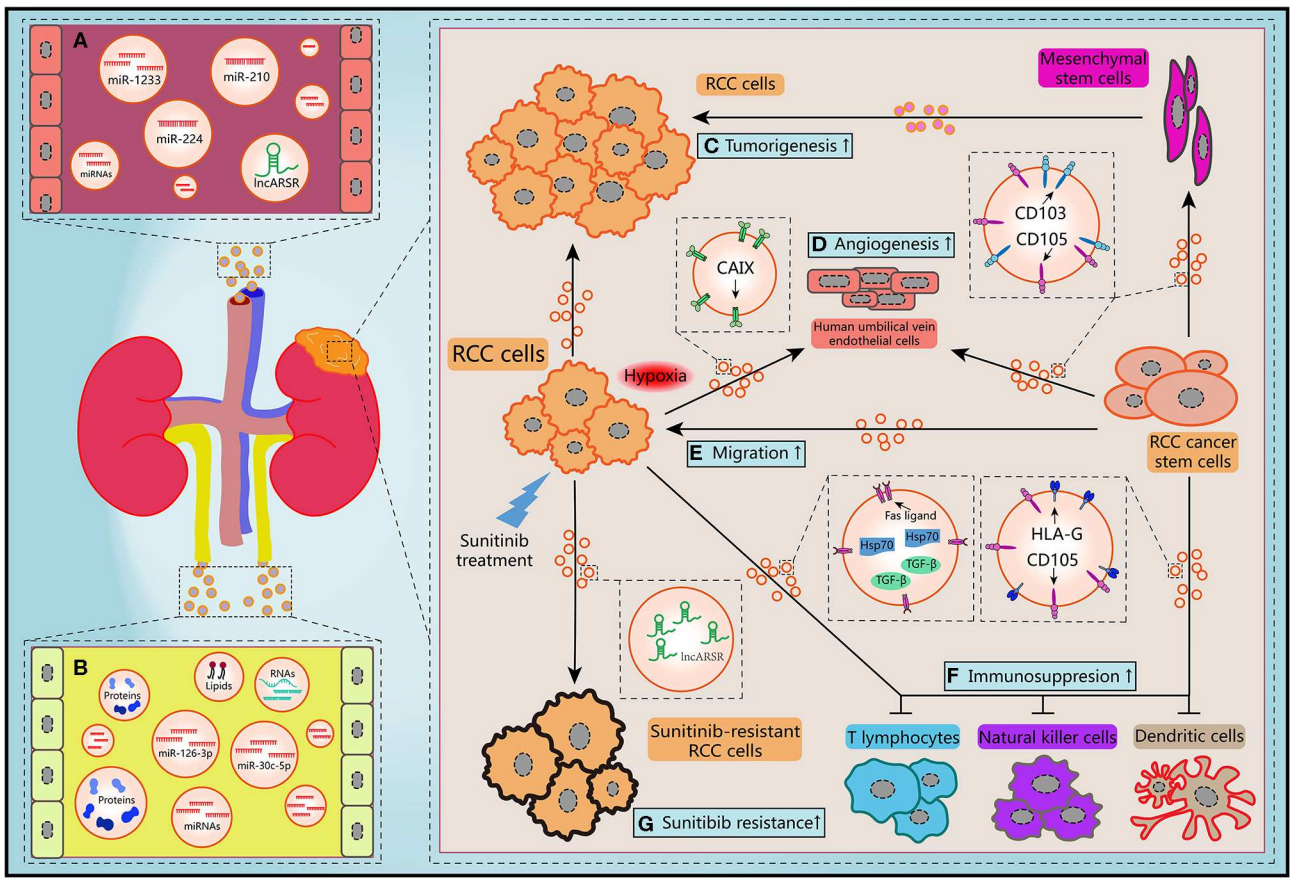


FIGURE 2 | Schematic diagram of the biological features of EVs. **(A)** Circulating EVs in blood contain potential biomarkers of RCC. **(B)** Circulating EVs in urine contain potential biomarkers of RCC. **(C)** RCC-derived EVs and mesenchymal stem cells-derived EVs promoted the tumorigenesis of RCC cells. **(D,E)** Migration ability of RCC cells and angiogenesis of human umbilical vein endothelial cells and could be improved by hypoxic RCC cells released EVs containing CAIX, CD103-positive or CD105-positive RCC CSCs-derived EVs. **(F)** RCC cells-derived EVs and RCC CSCs-derived EVs facilitated the immunosuppression of immune cells. **(G)** Sunitinib treatment induced RCC cells secreted EVs delivering lncARSR to increase the drug resistance of RCC cells.

reprogram of stromal recipient cells, and formation of PMN (74). However, the roles of EVs in RCC metastasis are still need to be unraveled. It has been shown that MMP-9 and CXCR-4 are closely associate with tumor metastasis and highly express in different cancer types. Chen et al. revealed that expression levels of these two proteins were upregulated after co-cultured RCC cell line 786-O with EVs shed from itself, which resulted in the improvements of the migration and invasion abilities and suppression of the adhesion ability (75). Camussi's team identified a subset of tumor-initiating cells expressing mesenchymal stem cell marker CD105 from human RCC specimens in a previous work. They found that EVs released by renal CD105$^+$ CSCs could trigger angiogenesis both *in vitro* and *in vivo*, and enhanced the lung metastases induced by injection of renal tumor cells intravenously. Furthermore, mRNAs and miRNAs implicating in tumor progression and metastasis were identified through molecular characterization of EVs (76). Subsequently, Camussi et al. reported that renal CSCs-derived EVs could stimulate persistent phenotypical changes in mesenchymal stem cells *in vitro* and support the tumor growth and vascularization when co-injected with RCC cells *in vivo* (77). Their conclusions unveiled that EVs shed from a subtype of renal CSCs may play critical roles in the TME modification, PMN formation, and metastasis of RCC in lung, which is one of the most common site of RCC metastasis.

Recently, Wang et al. demonstrated that CD103$^+$ CSCs, another subtype of renal CSCs, could release EVs enwrapping miR-19b-3p and deliver to RCC cells to initiate epithelial-mesenchymal transition (EMT) via suppressing the expression of PTEN. Quantitative detection of expression changes of EMT markers such as N-cadherin, Vimentin and Twist showed that CD103$^+$ CSCs EVs derived from RCC patients with lung metastasis presented significant effects on EMT. Notably, results of flow cytometry quantification also showed that the ratio of CD103$^+$ EVs over total EVs was higher in blood samples of RCC patients with lung metastasis than non-metastasis patients (78). Therefore, it can be inferred that EVs-contained CD103 may be involved in the organotropism of RCC. Additionally, previous work suggested that tetraspanins and integrins were also associated with metastasis organotropism (79, 80). Typically,

integrins α_6 and α_v were closely relevant to lung and liver metastases, respectively (33). Since lung and liver are common sites for RCC metastasis, we can believe that integrins α_6 and α_v may present in RCC-derived EVs and address these EVs to specific organs. Hence these endogenous surface molecules of EVs provide us crucial clues to understand the complex mechanism of tumor metastasis. It will be promising to develop indicators of metastatic prognosis and selective target-binding therapeutics for RCC treatment through unraveling the functions these transmembrane proteins.

Immune Evasion

In the past decade, the deep comprehension of communication between the immune cells and malignant tumor cells in TME has become a popular research field. Emerging investigations advocated that EVs are active players in this scenario (81). However, this interaction can be hijacked by tumor cells to facilitate immune evasion and stick many anti-cancer therapeutic strategies. Studies showed that the activation of T cells and the differentiation processes of monocytes to dendritic cells (DC cells) were both impaired by EVs derived from renal $CD105^+$ CSCs (82). This immune inhibitory effect was mediated by HLA-G, an antigen highly overexpressing in RCC and facilitating to immunosuppression (83). HLA-G blockade markedly relieved the inhibitory effect of EVs on DC cells differentiation. It has also been verified that EVs purified from RCC cell line ACHN contained Fas ligand and contributed to apoptosis of Jurkat T lymphocyte and immune evasion of RCC cells. These effects could be rescued by soluble Fas treatment (84). Natural killer (NK) cells are crucial player in the innate immune system, possessing strong abilities to control and kill tumor cells. Xia et al. found that EVs derived from primary RCC cells contained TGF-β, a major immunosuppressive cytokine. Co-culturing these EVs with NK cells exacerbated the dysfunctions of NK cells in a TGF-β/SMAD-dependent manner (85). Furthermore, Diao et al. elucidated that Hsp70 protein was more enriched in EVs than that in whole-cell lysates of Renca cells which is a cancer cell line of murine kidney. EVs-contained Hsp70 triggered the phosphorylation of Stat3 through regulating TLR2-MyD88 pathway and impeding the activity of the myeloid-derived suppressor cells (86). Considered together, these conclusions suggest that RCC cells may secrete EVs to interfere the immune system and support evasion of innate immune surveillance. Potential drug targets or biomarkers of the immunotherapy can be developed by clarifying the detailed mechanism of intercellular communication between cancer cells and immune cells.

Immunotherapy is one of the most promising therapeutic approach in multiple cancer types including RCC. Immune checkpoint protein inhibitors, especially antibodies against programmed cell death-1 (PD-1) and its ligand programmed death-ligand 1 (PD-L1), have elicited anti-cancer effects and long-lasting alleviation in melanoma, lymphoma, bladder cancer, non-small-cell lung cancer, RCC, and many other malignancies (87). However, only limited subset of patients exhibited durable response to immunotherapies. The total respond rate of anti-PD-1/PD-L1 therapy is merely around 10–30% (88). Previous studies have identified EVs-contained PD-L1 in diverse sources, including plasma of head and neck cancer glioblastoma, and melanoma patients as well as culture medium of breast cancer cell lines (89–94). A recent work demonstrated that EVs could support tumor growth by carrying PD-L1 and suppressing T cell activation in draining lymph nodes. Genetic blockade of EVs-contained PD-L1 induced long-term and systemic anti-tumor effects (95). Most recently, several novel methods were developed to quantitate the PD-L1 level in EVs. These newly approaches were higher in sensitivity, time-saving, and easily operated compared with ELISA-based canonical methods (96, 97). However, to the best of our knowledge, yet still no research focus on the PD-L1 in RCC-derived EVs. Above findings enlighten us that inhibition of EVs-contained PD-L1 may be an alternative therapy for RCC treatment, especially for RCC patients that are resistant to anti-PD-L1 antibodies. Meanwhile, EVs-carried molecules represented by PD-L1 may serve as reliable biomarkers for immunotherapies.

Drug Resistance

Accumulating evidences corroborate that EVs make non-negligible contributions to the resistance of anti-cancer drugs. The horizontally intercellular transmit of drug resistance are mediated by EVs cargoes including drug-efflux transporters, miRNAs, lncRNAs (98). Corcoran et al. established and characterized docetaxel-resistant variants of two prostate cancer cell lines by a serial assays including cross-resistance, morphology, multi-category phenotypes, and EVs secretion. They revealed that EVs released from docetaxel-resistant prostate cancer cells subverted sensitive cells to docetaxel-resistant phenotype through the involvement of EVs delivering multidrug resistance protein 1. Consistent results were presented when co-cultured docetaxel-sensitive prostate cancer cells with serum-derived EVs from prostate cancer patients before and after commencing docetaxel treatment (99). As a vital organ for the elimination and reabsorption of therapeutic drugs, kidney contain various drug transporters in proximal tubules. Thereby the variability of renal drug transporters will impact the processes of drug disposition (100). However, there is still no study focus on the drug resistance in RCC mediated by EVs-contained drug transporters.

Since several receptor tyrosine kinases relevant to angiogenesis and homeostasis of TME are overexpressed predominantly due to inactivation of *VHL* gene in ccRCC, inhibitors targeted receptor tyrosine kinases such as sunitinib have become the one of first-line therapies for RCC treatment (101). However, the clinical benefit of sunitinib treatment in ccRCC patients is limited due to inherent or acquired resistance. As such, the biological basis for resistance to sunitinib therapy and the clinical approach in this setting is of heightened interest of investigators (102). Qu et al. obtained sunitinib-resistant RCC cells through cycles of sunitinib treatment to nude mice with serial xenografts. Then lncRNA required for sunitinib resistance in RCC was identified by three rounds of screening sequentially. Firstly, lncRNA expression profiles between parental and sunitinib-resistant RCC cells was compared by lncRNA microarray. Then they established patient-derived xenograft models of RCC and mimic sunitinib therapy. Eight

lncRNA candidates were consequently selected to loss-of-function analysis by RNAi in sunitinib-resistant RCC cells. LncARSR was eventually identified as a highly abundant lncRNA in sunitinib-resistant ccRCC cells, which could favor sunitinib resistance via competitively binding both miR-34 and miR-449 to improve AXL and c-MET expression. More interestingly, lncARSR could be secreted and delivered via EVs to transform the phenotype of recipient cells from sunitinib-sensitive to sunitinib-resistant and lead to the dissemination of sunitinib resistance (103). Overall, it is valuable to clarify the various mechanisms of anti-cancer drugs resistance mediated by EVs, which may further help us to identify desirable biomarkers that can be used in drug response and identify novel targets to restore therapeutic approaches.

CLINICAL IMPLICATIONS OF EVS IN RCC

Recent years, many reviews have summarized the clinical implications of EVs in a variety of cancer types. Since the composition of the original cells can be reflected in the cargoes of EVs in a real-time mode, the initial interest of clinical implications is to find vital biomarkers from this favorable reservoir. EVs are natural nanoscale vesicles as ideal engineering platform owing to their unique advantages such as low toxicity and long-term stability in biofluids (104). EVs-based drug target discovery is also draw considerable attention of researchers due to recent findings. Moreover, RCC is still a malignant tumor with unpredictable progression, limited effective therapies and poor clinical prognosis. The progresses of clinical application of EVs in RCC is also relatively lag behind than that in other cancer types. Accordingly, the demonstrated and conceivable clinical implications of EVs in RCC will be discussed here from following aspects.

EVs-Derived Biomarkers for RCC

Owing to the encapsulation by vesicle membrane, the bioactive molecules within EVs are free from degradation by exogenous nucleases or proteases and stable in biological fluids (15). These abundant content which may be reliable biomarkers for prediction of RCC progression have been extensively investigated. Previously, Zhao et al. reported that the expression level of miR-210 was differentially higher in primary RCC tissues of 32 patients than non-tumor renal parenchymas. Results of receiver operating characteristic (ROC) analysis also showed that ccRCC patients and healthy individuals could be discriminated by the average level of cell-free miR-210 in serum (105). They assessed expression levels of three miRNAs (miR-210, miR-1233, and miR-15a) in serum-derived EVs in a follow-up work. Results of ROC analysis showed that it was feasible to use miR-210 and miR-1233 but not miR-15a as diagnostic biomarkers (106). Consistently, a recent study confirmed the expression level of miR-210 in serum-derived EVs was significantly higher in RCC patients than healthy controls (72). Similarly, expression level of miR-224 was also overexpressed in cancer tissues of ccRCC patients (107, 108). The level of serum EVs-contained miR-224 was significantly correlated with progression-free survival (PFS) or overall survival (OS) of ccRCC patients (109). Moreover, a study evaluated the possibility of miRNAs from plasma-derived EVs for RCC prognosis by RNA sequencing. Results of Kaplan-Meier analysis confirmed the correlations of three miRNAs with OS of RCC patients, including miR-let-7i-5p, miR-26a-1-3p, and miR-615-3p (110).

Urine as a dynamic biofluid is also a promising source for RCC biomarker development rather than a waste product of body. Urinary EVs can be released from every renal epithelial cell type facing the urinary tract. Therefore, the cargoes of urinary EVs may be accessibly real-time signals for renal dysfunction. However, only few researchers attempted to find bioactive molecules from urinary EVs and these snapshots need to be further characterized (18). Study reported that combinations of urinary EVs-derived miRNAs (miR-449a, miR-34b-5p, or miR-486-5p with miR-126-3p) had the power to distinguish healthy controls, patients with benign renal tumors, and patients with early-stage or advanced ccRCC (111). It has also claimed that the level of miR-30c-5p within the urinary EVs was significantly decreased in ccRCC patients but not in other urological malignancies samples (67). In addition, differential levels of miR-150 and miR-205 were found in EVs isolated from 786-O and HK-2 cell lines (112). Our previous work showed that the lost expression of organic cation transporter 2 were partly due to the downregulation by miR-489-3p and miR-630. Interestingly, miR-489-3p and miR-630 were more abundant in EVs than donor cells (113, 114). Therefore, these findings of fundamental work may also have translational value to provide clues for RCC biomarker discovery in a certain extent.

In addition to miRNAs, other content of EVs also have potential to be developed as biomarkers for RCC. As mentioned above, lncARSR was elucidated as a mediator of the transmission of sunitinib resistance, which could be enwrapped and delivered through EVs. Qu et al. further revealed that circulating lncARSR could be utilized as indicator to predict sunitinib response in RCC patients (103). Moreover, Palma et al. reported that the mRNA levels of *GSTA1*, *CEBPA*, and *PCBD1* genes in urinary EVs were lower in RCC patients than that in control subjects and this pattern backed to normal level after 1 month of nephrectomy (115). In 2012, Boccio et al. established a hyphenated micro LC-Q-TOF-MS platform to profile the lipid repertoire of human urinary EVs. A comparative analysis for lipid content in urinary EVs purified from RCC patients and healthy subjects was performed for the first time (116). Similarly, a proteomics study in 2013 reported that the protein composition of urinary EVs was substantially different in RCC patients and control subjects. Results presented for the first time that considerable number of proteins were significantly enriched in RCC patients, including Ceruloplasmin, Podocalyxin, Dickkopf related protein 4, MMP9 and CAIX (117). A recent work reported that Azurocidin was highly enriched in EVs isolated from tumor tissues of ccRCC patients than adjacent normal tissues. Importantly, Azurocidin content was also significantly higher in serum EVs from ccRCC patients compared to healthy controls (118). These tentative work provided valuable indications for exploiting potential mRNA, lipid, and protein biomarker for RCC from urinary EVs. Taken together, it can be concluded that multiple EVs cargoes derived from different kinds of biofluids are promising

non-invasive biomarkers for early diagnosis and treatment of RCC. The potential biomarkers derived from EVs which have been validated in clinical samples of RCC are listed in **Table 1**.

EVs-Based Drug Vehicles and Targets in RCC

The biological characteristics make EVs can be harnessed as vehicles for therapeutic agents to improve curative effect. Numerous clinical and preclinical trials have suggested that these EVs-based drug vehicles and therapies are promising, feasible and well-tolerated (119–121). There are two basic approaches to load cargoes into EVs: exogenous loading and endogenous loading. Exogenous modification can be achieved after collection of EVs, with encapsulation of small molecules, proteins, and RNAs into or onto EVs via diverse methods including co-incubation, electroporation, and sonication (121). Tian et al., developed a tumor-targeting EVs from mouse immature DC cells expressing a well-characterized EVs membrane protein (Lamp2b) fused to integrin α_v-specific iRGD, which is a new tumor-homing and penetrating peptide. After loaded with doxorubicin via electroporation, this delivery platform showed high efficiency in tumor-targeting and doxorubicin delivery to integrin α_v-positive breast cancer cells both *in vitro* and *in vivo* (122). Wan et al. developed a nucleolin-targeting aptamer AS1411 which covalently conjugated to cholesterol-PEG and anchored onto membrane of mouse DC cells. Subsequently, EVs were obtained from this modified DC cell model and loaded with paclitaxel by sonication. Results of cancer treatment in xenograft nude mice showed that engineered EVs enhanced therapeutic efficacy with low systemic toxicity (123). We can believe that along with the detailed mechanism of EVs-mediated metastasis organotropism are being clarified, EVs are promising material to achieve drug-targeting delivery for cancer treatment. However, the immune responses are need to be considered seriously. Additionally, the production yield is also a challenge for applying engineered EVs in tumor-targeting delivery.

Alternatively, cargo of EVs can be endogenously loaded through genetically manipulating the donor cells to overexpress bioactive molecules and employed as EVs-based vaccines or imaging tools. With significant higher level in surface of RCC cells than normal renal cells, RCC-associated antigen G250 could be served as one of the therapeutic targets (124). EVs containing G250 or other RCC-specific antigens may be novel approaches to develop EVs-based cancer vaccines for RCC treatment. It has been shown that modified RCC cells released EVs expressing both glycolipid-anchored-IL-12 and G250, which efficiently promoted the proliferation of antigen-specific cytotoxic T lymphocytes and enhanced cytotoxic effects (125). Notably, there is a risk of mixing pathogens such as viruses with EVs since these nanometric vesicles have similar biophysical properties (126). Hence a standard operating procedure is very necessary when isolate EVs as cancer vaccine. By combining a Cre recombinase-based system with high-resolution fluorescence imaging techniques, Zomer et al. realized the visualization of intracellular EVs exchange within local and distant tumor sites *in vivo*. Results showed that less malignant tumor cells presented heightened migratory ability after taken up the EVs released by highly malignant tumor cells (127). Moreover, several other molecular imaging strategies have also been utilized to monitor and determine the biodistribution of EVs *in vivo*, including bioluminescence, nuclear, and magnetic resonance imaging techniques (128). These interesting findings and advanced techniques make it clear that EVs-based modification can be used to achieve the phenocopying of tumor cells and visualize cancer development process *in vivo* in the future.

Drugs targeting vital steps in formation, release or uptake of EVs may also be served as effective adjuvants for cancer treatment. Datta et al. utilized quantitative high throughput screen assay to find active compounds targeting the formation and release of EVs in prostate cancer cells. Totally five and six lead compounds were validated as potent inhibitors and activators, respectively (129). In another review, two groups of candidate drugs were broadly classified according to the mechanisms of modulating EVs biogenesis or secretion. One is compounds that specifically inhibit EVs trafficking, including calpeptin, manumycin A, and Y27632. Another group is compounds that specifically disrupt lipid metabolism, including pantethine, imipramine, and GW4869 (130). Interestingly, Ortiz et al. identified that reserpine, a commonly used anti-hypertensive drug since 1955, could alter the fusion process of lipid membrane and then inhibit PMN formation that was induced by melanoma-derived EVs. Their findings indicated that tumor-derived EVs could "educate" healthy cells to facilitate tumor metastasis. Meanwhile agents like reserpine can interfere this education process and play a defensive role on EVs uptake. Thus, it is valuable to repurpose these drugs as adjuvant treatment for metastatic cancer therapy (131). More recently, sulfisoxazole, an oral antibacterial drug approved by US FDA, was screened out as inhibitor of EVs secretion in breast cancer cells. Through targeting endothelin receptor A, sulfisoxazole promoted the degradation of ESCRT-dependent MVB, suppressed biogenesis and secretion of EVs, as well as significantly inhibited the growth and metastasis of breast cancer cells without notable toxicity (132). These important findings enlighten us drug repurposing can be harnessed as approaches to block EVs functions in tumor progression.

Potential Application of Artificial Intelligence in EVs Research

Artificial intelligence (AI) refers to the simulation of human intelligence in machines. AI approaches have the potential to enhance the qualitative interpretation of cancer imaging by expert clinicians in three main tasks: computer-aided detection of tumor sites, characterization of intra-tumor heterogeneity and variation, as well as temporal monitoring of tumor changes (133). As a specific subset of AI approaches, machine learning (ML) are able to interpret complex data and leverage the detailed information to make accurate prediction or decision. Studies have demonstrated that deep learning frameworks can be applied to distinguish major subtypes of RCC using histological or computed tomography images (134, 135). Similarly, ML algorithms also have the power to analyze a substantial

TABLE 1 | EVs derived potential biomarkers with clinical significance for RCC.

Type	EVs source	EVs cargoes	Analysis method	Cohorts	Clinical significance	Year	References
Lipid	Urine	LysoPE etc. 196 differential signals	microLC-Q-TOF-MS	8 ccRCC patients, 8 HS	48 differential lipidomes (22 upregulated and 26 downregulated in RCC)	2012	(116)
lncRNA	Plasma	Circulating lncARSR	qRT-PCR	71 advanced ccRCC patients, 32 HS	Differentiated ccRCC patients from healthy controls; High lncARSR levels in pre-therapy correlated with PFS independent of clinical characteristics	2016	(103)
mRNA	Urine	GSTA1, CEBPA, PCBD1	Microarray, qRT-PCR	46 RCC patients (33 with ccRCC), 22 HS	Significant lower in ccRCC patients than HS and increased to normal level 1 month after nephrectomy	2016	(115)
miRNA	Plasma	miR-let-7i-5p, miR-26a-1-3p, miR-615-3p	RNA-sequencing, qRT-PCR	44 and 65 metastatic RCC patients for screening and validate cohort, respectively	Low levels correlated with poor OS of mRCC patients, independent of age, gender, tumor grade, stage at diagnosis, coagulative necrosis, or sarcomatoid differentiation	2017	(110)
	Serum	miR-1233, miR-210	qRT-PCR	82 ccRCC patients, 80 HS	Both significant higher in ccRCC patients than HS independent of gender, age, or ccRCC grade	2018	(106)
	Serum	miR-210	Microarray, qRT-PCR	45 pre-operative and 35 post-operative ccRCC patients, 30 HS	Significant higher in ccRCC patients than HS, and in pre-operative than post-operative samples	2019	(72)
	Serum	miR-224	qRT-PCR	108 ccRCC patients	High level correlated with shorter PFS, CSS and OS of ccRCC patients	2017	(109)
	Urine	miR-126-3p	Microarray, qRT-PCR	81 ccRCC patients, 33 HS	Differentiated ccRCC patients from HS	2016	(111)
	Urine	miR-126-3p combined miR-449a	Microarray, qRT-PCR	81 ccRCC patients, 33 HS	Differentiated ccRCC patients from HS		
	Urine	miR-126-3p combined miR-34b-5p	Microarray, qRT-PCR	81 ccRCC patients, 33 HS	Differentiated ccRCC and small renal masses (pT1a, $\leq$4 cm) patients from HS, respectively		
	Urine	miR-126-3p combined miR-486-5p	Microarray, qRT-PCR	24 benign renal tumor patients, 33 HS	Differentiated benign patients from HS		
	Urine	miR-30c-5p	RNA-sequencing, qRT-PCR	70 early-stage ccRCC patients, 30 HS	Significant lower in early-stage ccRCC patients than HS	2019	(67)
Protein	Urine	Matrix metalloproteinase 9, Ceruloplasmin, Podocalyxin, Dickkopf related protein 4, Carbonic anhydrase IX	LC-MS/MS, western blotting	9 ccRCC patients, 9 HS	Significant higher in ccRCC patients than HS	2013	(117)
	Urine	Aquaporin-1, Extracellular matrix metalloproteinase inducer, Neprilysin, Dipeptidase 1, Syntenin-1	LC-MS/MS, western blotting	9 ccRCC patients, 9 HS	Significant lower in ccRCC patients than HS		
	Serum	CD103	Flow cytometry	76 and 133 metastatic or non-metastatic ccRCC patients, respectively	Higher ratio of $CD103^+$ EVs over total EVs in samples of metastatic patients than non-metastatic patients	2019	(78)
	Serum	Azurocidin	LC-MS/MS	19 ccRCC patients, 10 HS	Significant higher in ccRCC patients than HS	2018	(118)
	Tissue	Azurocidin	LC-MS/MS	20 paired tumor and adjacent normal tissues of ccRCC patients	Significant higher in ccRCC patients than HS		

amount of images that are produced by EVs purification and characterization processes. Studies showed that these biophysical parameters of EVs could be assessed by ML algorithms to identify the subpopulation of EVs or even further predict the original donor cells (136, 137). Due to the incredible amount of EVs and the need for downstream analysis during

each study, multiparameter results of EVs characterization are particularly amenable to ML algorithms. A preliminary work of Borgovan et al. reported that ML algorithms could distinguish the heterogeneous EVs derived from blood samples with healthy or leukemic phenotypes based on data sets collected from a nanoparticle tracking analysis, thus improved the accurate of EVs classification (138).

At present, one of the most challenges in the field of biomarker discovery is how to decipher the huge amount of garbled information within EVs. AI approaches are becoming trustworthy solutions to this given problem as they are able to modelize complicated network and leverage valuable information within observed data to accurately estimate and predict new samples. Early in 2003, Won et al. had identified five protein biomarkers of serum by using a mass spectrometry-based protein profiling and AI analysis and then successfully differentiated RCC from healthy subjects and other urological diseases (139). Moreover, Zheng et al. developed a novel diagnosis tool to predict early-stage RCC patients which depended on a biomarker cluster that was identified by serum metabolomics method and ML algorithms (140). Meanwhile, unprecedentedly massive data of EVs are also being generated by various "omics" technologies including genomics, transcriptomics, proteomics, metabolomics, glycomics, and lipidomics (141). Several online databases have been established to categorize the RNAs, lipids, proteins, and metabolites within EVs, which have been summarized in **Table 2** (142–151). These integrative resources will favor researchers to outline the landscape of EVs in cancer progression and identify relevant biomarkers more quickly and more accurately.

Integrating EVs-derived biomarkers with ML algorithms to analyze patterns in massive data sources such as gene expression, protein expression, or digital pathology data may obtain a higher diagnostic efficacy of the diagnosis. Chen et al. profiled four surface biomarkers including HER2, GPC-1, EpCAM, and EGFR from serum-derived EVs through DNA points accumulation for imaging in nanoscale topography. They implemented an integrated platform combining EVs identification with quantitative analysis and accurately differentiated pancreatic cancer and breast cancer from unknown samples (152). Additionally, advanced techniques such as microfluidic make it possible to separate EVs on a single chip. In a previous study, Ko et al. developed a multichannel microfluidic platform combining with ML algorithms that specifically isolated EVs from clinical plasma samples, quantitatively detected the RNA profile inside of EVs, and distinguished pancreatic cancer patients with healthy controls (153). They subsequently exploited another workflow that integrated a magnetic capture system with RNA sequencing and ML algorithms. This system purified a subpopulation of EVs and identified a panel of 11 miRNAs from EVs which could classify distinct cancer states in a transgenic mouse model (154). Thus, it is also a feasible strategy to combine upstream isolation methods with downstream ML algorithms to realize the development of "on-a-chip" platform for systemically purification and determination of EVs-derived biomarkers.

Moreover, AI approaches promises to make great strides in almost all stage of drug discovery, including target validation, biomarker identification, and analysis of clinical trial information (155). Since the drug data sets are becoming dynamic, heterogeneous and large scale, state-of-the-art AI approaches such as deep learning and innovative modeling methods provide new answers to efficacy and safety evaluations of drug candidates based on big data modeling and analysis (156). Donner et al. reported a novel method for computational drug repositioning by taking advantage of neural network. They revealed previously unnoticed functional relationships between different compounds based on denoise gene expression data rather than structural similarity (157). Hence AI approaches can build bridges between abundant data sources from high-throughput experiments with gene expression profiles and massive drug candidates. The information of EVs content is also increasingly rich in data. Meanwhile the downstream effects of EVs in cancer progression are non-linear. It is reasonable to assume that the ability of AI

TABLE 2 | EVs related online databases.

Database	Publish date	Overview	Update date	References
EVmiRNA	2019	Comprehensive miRNA expression profiles in 462 EVs small RNA-sequencing datasets from 17 tissues/diseases	2019	(142)
EVpedia	2013	High-throughput datasets of EVs components (proteins, RNAs, and lipids) from prokaryotic and eukaryotic EVs	2013	(143)
EV-TRACK	2017	Experimental parameters of EV-related studies	2019	(144)
ExoCarta	2009	Identified contents (protein, mRNA, miRNA, and lipids) of exosomes in multiple organisms from 286 studies	2016	(145)
exoRBase	2018	Exosomal RNA (circRNA, lncRNA, and mRNA) derived from RNA-sequencing data analyses of human blood	2019	(146)
Exosome Gene Ontology Annotation Initiative	2015	GO annotations of human exosomal proteins	2015	(147)
Plasma Proteome Database	2014	Annotation of 318 identified proteins of EVs from plasma	2014	(148)
Urinary Exosome Protein Database	2004	Mass spectrometry data of 1,160 proteins derived from urinary exosomes isolated from healthy human volunteers	2009	(149, 150)
Vesiclepedia	2012	Compendium of molecular data (lipid, RNA, and protein) identified in different classes of EVs from 1,254 studies	2019	(151)

to mining valuable information presents new opportunities for novel target identification and validation for EVs-based anti-cancer therapies. Therapeutic Target Database (TTD, http://db.idrblab.org/ttd/) has been established to integrate information of early drug candidates and therapeutic targets that contain expanded knowledge of target regulators such as miRNAs, transcription factors and other interacting proteins (158). Database with molecular information about drugs such as DrugBank (https://www.drugbank.ca/) include comprehensive data of the influence of hundreds drugs on metabolite levels, gene expression levels and protein expression levels, enabling us to find more connections of EVs content changes with drugs (159). Altogether, these important approaches may provide novel research tools to fundamental studies of EVs biology and translational studies of EVs-based therapies. Clearly, more work is need to be deployed in this scenario to figure out the completed mechanisms of EVs biogenesis, secretion and uptake, which may reward us valuable drug targets by using advanced AI approaches.

PROSPECTS

EVs are attracting increasing attention in cancer research due to its various roles in intracellular communication during cancer progression. However, RCC is relative unnoticed in this research hotspot compared with other cancer types. In this review, we recapitulated the roles and clinical implications of EVs in RCC. Diverse bioactive molecules carried by EVs regulate almost all processes of RCC, such as tumorigenesis, metastasis, immunosuppression, and drug resistance. Due to the unique function of kidney in urinary system, both blood and urine are valuable biofluids with abundant EVs, which are readily accessible sources for biomarkers discovery. Moreover, multiple potential applications can be developed to provide novel strategies for diagnosis and treatment of RCC, including but not limited to EVs-based cancer vaccine, *in vivo* imaging technique, targeted drug delivery system, and drug discovery. But it is noteworthy the detailed mechanisms and effects of EVs on RCC progression are still to be further clarified. The gaps between digital analysis and experimental validation are still need to be solved. Meantime there are still a variety of challenges for the clinical use of EVs in RCC. Standard operating procedure for EVs isolation, quantification, and analysis are still deficiency, especially for biofluids sample. The stability and the unknown side effects of EVs-based therapy must to be considered and assessed. Moreover, High-quality data sets are required in terms of the AI-aided drug target discovery based on EVs. Taken together, extensive work need to be launched to make a better understanding of roles of EVs in RCC progression and make the potential clinical utilities for EVs in RCC therapies come true.

AUTHOR CONTRIBUTIONS

ZQ was a major contributor in searching the literature and writing the manuscript. QX and HH reviewed the manuscript and provided significant revisions. LY and SZ gave critical advice and guidance throughout the whole process of this study. All the authors read and approved the final manuscript.

FUNDING

This project was supported by the National Key R&D Program of China (2017YFE0102200), NSFC (81773817) and the Leading Talent of Ten Thousand Plan-National High-Level Talents Special Support Plan.

REFERENCES

1. Shuch B, Amin A, Armstrong AJ, Eble JN, Ficarra V, Lopez-Beltran A, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. *Eur Urol*. (2015) 67:85–97. doi: 10.1016/j.eururo.2014.04.029
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. (2020) 70:7–30. doi: 10.3322/caac.21590
4. Surveillance, Epidemiology, End Results Program. *SEER Stat Fact Sheets: Kidney and Renal Pelvis Cancer*. Bethesda, MD: National Cancer Institute. Available online at: https://seer.cancer.gov/statfacts/html/kidrp.html (accessed December 20, 2019).
5. Choueiri TK, Motzer RJ. Systemic therapy for metastatic renal-cell carcinoma. *N Engl J Med*. (2017) 376:354–66. doi: 10.1056/NEJMra1601333
6. Capitanio U, Montorsi F. Renal cancer. *Lancet*. (2016) 387:894–906. doi: 10.1016/S0140-6736(15)00046-X
7. Bhatt JR, Finelli A. Landmarks in the diagnosis and treatment of renal cell carcinoma. *Nat Rev Urol*. (2014) 11:517–25. doi: 10.1038/nrurol.2014.194
8. Koul H, Huh JS, Rove KO, Crompton L, Koul S, Meacham RB, et al. Molecular aspects of renal cell carcinoma: a review. *Am J Cancer Res*. (2011) 1:240–54.
9. Kumbla RA, Figlin RA, Posadas EM. Recent advances in the medical treatment of recurrent or metastatic renal cell cancer. *Drugs*. (2017) 77:17–28. doi: 10.1007/s40265-016-0665-1
10. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol*. (2018) 19:213–28. doi: 10.1038/nrm.2017.125
11. Tkach M, Thery C. Communication by extracellular vesicles: where we are and where we need to go. *Cell*. (2016) 164:1226–32. doi: 10.1016/j.cell.2016.01.043
12. Kalluri R. The biology and function of exosomes in cancer. *J Clin Invest*. (2016) 126:1208–15. doi: 10.1172/JCI81135
13. He CJ, Zheng S, Luo Y, Wang B. Exosome theranostics: biology and translational medicine. *Theranostics*. (2018) 8:237–55. doi: 10.7150/thno.21945
14. Vader P, Breakefield XO, Wood MJA. Extracellular vesicles: emerging targets for cancer therapy. *Trends Mol Med*. (2014) 20:385–93. doi: 10.1016/j.molmed.2014.03.002
15. Nawaz M, Camussi G, Valadi H, Nazarenko I, Ekstrom K, Wang XQ, et al. The emerging role of extracellular vesicles as biomarkers for urogenital cancers. *Nat Rev Urol*. (2014) 11:688–701. doi: 10.1038/nrurol.2014.301
16. Franzen CA, Blackwell RH, Foreman KE, Kuo PC, Flanigan RC, Gupta GN. Urinary exosomes: the potential for biomarker utility, intercellular signaling and therapeutics in urological malignancy. *J Urol*. (2016) 195:1331–9. doi: 10.1016/j.juro.2015.08.115

17. Junker K, Heinzelmann J, Beckham C, Ochiya T, Jenster G. Extracellular vesicles and their role in urologic malignancies. *Euro Urol*. (2016) 70:323–31. doi: 10.1016/j.eururo.2016.02.046
18. Merchant ML, Rood IM, Deegens JKJ, Klein JB. Isolation and characterization of urinary extracellular vesicles: implications for biomarker discovery. *Nat Rev Nephrol*. (2017) 13:731–49. doi: 10.1038/nrneph.2017.148
19. Panfoli I. Cancer exosomes in urine: a promising biomarker source. *Transl Cancer Res*. (2017) 6:S1389–93. doi: 10.21037/tcr.2017.10.17
20. Urabe F, Kosaka N, Kimura T, Egawa S, Ochiya T. Extracellular vesicles: toward a clinical application in urological cancer treatment. *Int J Urol*. (2018) 25:533–43. doi: 10.1111/iju.13594
21. Gai C, Pomatto MAC, Grange C, Deregibus MC, Camussi G. Extracellular vesicles in onco-nephrology. *Exp Mol Med*. (2019) 51:1–8. doi: 10.1038/s12276-019-0213-7
22. Grange C, Brossa A, Bussolati B. Extracellular vesicles and carried miRNAs in the progression of renal cell carcinoma. *Int J Mol Sci*. (2019) 20:1832. doi: 10.3390/ijms20081832
23. Di Vizio D, Kim J, Hager MH, Morello M, Yang W, Lafargue CJ, et al. Oncosome formation in prostate cancer: association with a region of frequent chromosomal deletion in metastatic disease. *Cancer Res*. (2009) 69:5601–9. doi: 10.1158/0008-5472.CAN-08-3860
24. Tricarico C, Clancy J, D'Souza-Schorey C. Biology and biogenesis of shed microvesicles. *Small GTPases*. (2017) 8:220–32. doi: 10.1080/21541248.2016.1215283
25. Colombo M, Raposo G, Thery C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev Biol*. (2014) 30:255–89. doi: 10.1146/annurev-cellbio-101512-122326
26. Stenmark H. Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol*. (2009) 10:513–25. doi: 10.1038/nrm2728
27. Kowal J, Tkach M, Thery C. Biogenesis and secretion of exosomes. *Curr Opin Cell Biol*. (2014) 29:116–25. doi: 10.1016/j.ceb.2014.05.004
28. Jahn R, Scheller RH. SNAREs: engines for membrane fusion. *Nat Rev Mol Cell Biol*. (2006) 7:631–43. doi: 10.1038/nrm2002
29. Savina A, Fader CM, Damiani MT, Colombo MI. Rab11 promotes docking and fusion of multivesicular bodies in a calcium-dependent manner. *Traffic*. (2005) 6:131–43. doi: 10.1111/j.1600-0854.2004.00257.x
30. Faure J, Lachenal G, Court M, Hirrlinger J, Chatellard-Causse C, Blot B, et al. Exosomes are released by cultured cortical neurones. *Mol Cell Neurosci*. (2006) 31:642–8. doi: 10.1016/j.mcn.2005.12.003
31. Horibe S, Tanahashi T, Kawauchi S, Murakami Y, Rikitake Y. Mechanism of recipient cell-dependent differences in exosome uptake. *BMC Cancer*. (2018) 18:47. doi: 10.1186/s12885-017-3958-1
32. Nazarenko I, Rana S, Baumann A, McAlear J, Hellwig A, Trendelenburg M, et al. Cell surface tetraspanin Tspan8 contributes to molecular pathways of exosome-induced endothelial cell activation. *Cancer Res*. (2010) 70:1668–78. doi: 10.1158/0008-5472.CAN-09-2470
33. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. *Nature*. (2015) 527:329–35. doi: 10.1038/nature15756
34. Deatherage BL, Cookson BT. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect Immun*. (2012) 80:1948–57. doi: 10.1128/IAI.06014-11
35. Wang Z, Chen JQ, Liu JL, Tian L. Exosomes in tumor microenvironment: novel transporters and biomarkers. *J Transl Med*. (2016) 14:297. doi: 10.1186/s12967-016-1056-9
36. Mathieu M, Martin-Jaular L, Lavieu G, Thery C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nat Cell Biol*. (2019) 21:9–17. doi: 10.1038/s41556-018-0250-9
37. Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW, Simpson RJ. Extracellular vesicles in cancer: implications for future improvements in cancer care. *Nat Rev Clin Oncol*. (2018) 15:617–38. doi: 10.1038/s41571-018-0036-9
38. Ratajczak J, Miekus K, Kucia M, Zhang J, Reca R, Dvorak P, et al. Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery. *Leukemia*. (2006) 20:847–56. doi: 10.1038/sj.leu.2404132
39. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. (2007) 9:654–72. doi: 10.1038/ncb1596
40. Skog J, Wurdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol*. (2008) 10:1470–9. doi: 10.1038/ncb1800
41. Hinger SA, Cha DJ, Franklin JL, Higginbotham JN, Dou YC, Ping J, et al. Diverse long RNAs are differentially sorted into extracellular vesicles secreted by colorectal cancer cells. *Cell Rep*. (2018) 25:715–25.e4. doi: 10.1016/j.celrep.2018.09.054
42. Chen F, Chen JN, Yang LB, Liu J, Zhang XQ, Zhang Y, et al. Extracellular vesicle-packaged HIF-1α-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat Cell Biol*. (2019) 21:498–510. doi: 10.1038/s41556-019-0299-0
43. Li Y, Zheng QP, Bao CY, Li SY, Guo WJ, Zhao J, et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res*. (2015) 25:981–4. doi: 10.1038/cr.2015.82
44. Bai H, Lei K, Huang F, Jiang Z, Zhou X. Exo-circRNAs: a new paradigm for anticancer therapy. *Mol Cancer*. (2019) 18:56. doi: 10.1186/s12943-019-0986-2
45. Huang XY, Yuan TZ, Tschannen M, Sun ZF, Jacob H, Du MJ, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics*. (2013) 14:319. doi: 10.1186/1471-2164-14-319
46. Kalluri R, LeBleu VS. Discovery of double-stranded genomic DNA in circulating exosomes. *Cold Spring Harb Symp Quant Biol*. (2016) 81:275–80. doi: 10.1101/sqb.2016.81.030932
47. Takahashi A, Okada R, Nagao K, Kawamata Y, Hanyu A, Yoshimoto S, et al. Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun*. (2017) 8:15287. doi: 10.1038/ncomms15287
48. Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, et al. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat Commun*. (2011) 2:180. doi: 10.1038/ncomms1180
49. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res*. (2014) 24:766–9. doi: 10.1038/cr.2014.44
50. Kawamura Y, Yamamoto Y, Sato T-A, Ochiya T. Extracellular vesicles as trans-genomic agents: emerging roles in disease and evolution. *Cancer Sci*. (2017) 108:824–30. doi: 10.1111/cas.13222
51. Carvalho J, Oliveira C. Extracellular vesicles - powerful markers of cancer evolution. *Front Immunol*. (2014) 5:685. doi: 10.3389/fimmu.2014.00685
52. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol*. (2013) 200:373–83. doi: 10.1083/jcb.201211138
53. Charrin S, Jouannet S, Boucheix C, Rubinstein E. Tetraspanins at a glance. *J Cell Sci*. (2014) 127:3641–8. doi: 10.1242/jcs.154906
54. Thery C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and function. *Nat Rev Immunol*. (2002) 2:569–79. doi: 10.1038/nri855
55. Console L, Scalise M, Tonazzi A, Giangregorio N, Indiveri C. Characterization of exosomal SLC22A5 (OCTN2) carnitine transporter. *Sci Rep*. (2018) 8:3758. doi: 10.1038/s41598-018-22170-7
56. Taylor J, Bebawy M. Proteins regulating microvesicle biogenesis and multidrug resistance in cancer. *Proteomics*. (2019) 19:e1800165. doi: 10.1002/pmic.201800165
57. Gerth K, Kodidela S, Mahon M, Haque S, Verma N, Kumar S. Circulating extracellular vesicles containing xenobiotic metabolizing CYP enzymes and their potential roles in extrahepatic cells via cell-cell interactions. *Int J Mol Sci*. (2019) 20:6178. doi: 10.3390/ijms20246178
58. Li P, Kaslan M, Lee SH, Yao J, Gao Z. Progress in exosome isolation techniques. *Theranostics*. (2017) 7:789–804. doi: 10.7150/thno.18133
59. Tian Y, Gong M, Hu Y, Liu H, Zhang W, Zhang M, et al. Quality and efficiency assessment of six extracellular vesicle isolation methods by nano-flow cytometry. *J Extracell Vesicles*. (2020) 9:1697028. doi: 10.1080/20013078.2019.1697028
60. Guo SC, Tao SC, Dawn H. Microfluidics-based on-a-chip systems for isolating and analysing extracellular vesicles. *J Extracell Vesicles*. (2018) 7:1508271. doi: 10.1080/20013078.2018.1508271

61. Gardiner C, Di Vizio D, Sahoo S, Thery C, Witwer KW, Wauben M, et al. Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. *J Extracell Vesicles*. (2016) 5:32945. doi: 10.3402/jev.v5.32945
62. Thery C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the international society for extracellular vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. (2018) 7:1535750. doi: 10.1080/20013078.2018.1461450
63. Maacha S, Bhat AA, Jimenez L, Raza A, Haris M, Uddin S, et al. Extracellular vesicles-mediated intercellular communication: roles in the tumor microenvironment and anti-cancer drug resistance. *Mol Cancer*. (2019) 18:55. doi: 10.1186/s12943-019-0965-7
64. Maia J, Caja S, Moraes MCS, Couto N, Costa-Silva B. Exosome-based cell-cell communication in the tumor microenvironment. *Front Cell Dev Biol*. (2018) 6:18. doi: 10.3389/fcell.2018.00018
65. Jiang XL, Zhang Y, Tan B, Luo CL, Wu XH. Renal tumor-derived exosomes inhibit hepaCAM expression of renal carcinoma cells in a p-AKT-dependent manner. *Neoplasma*. (2014) 61:416–23. doi: 10.4149/neo_2014_051
66. Du T, Ju G, Wu S, Cheng Z, Cheng J, Zou X, et al. Microvesicles derived from human Wharton's Jelly mesenchymal stem cells promote human renal cancer cell growth and aggressiveness through induction of hepatocyte growth factor. *PLoS ONE*. (2014) 9:e96836. doi: 10.1371/journal.pone.0096836
67. Song S, Long M, Yu G, Cheng Y, Yang Q, Liu J, et al. Urinary exosome miR-30c-5p as a biomarker of clear cell renal cell carcinoma that inhibits progression by targeting HSPA5. *J Cell Mol Med*. (2019) 23:6755–65. doi: 10.1111/jcmm.14553
68. Chamie K, Klöpfer P, Bevan P, Störkel S, Said J, Fall B, et al. Carbonic anhydrase-IX score is a novel biomarker that predicts recurrence and survival for high-risk, nonmetastatic renal cell carcinoma: data from the phase III ARISER clinical trial. *Urol Oncol*. (2015) 33:204.e25–33. doi: 10.1016/j.urolonc.2015.02.013
69. Horie K, Kawakami K, Fujita Y, Sugaya M, Kameyama K, Mizutani K, et al. Exosomes expressing carbonic anhydrase 9 promote angiogenesis. *Biochem Biophys Res Commun*. (2017) 492:356–61. doi: 10.1016/j.bbrc.2017.08.107
70. Parolini I, Federici C, Raggi C, Lugini L, Palleschi S, de Milito A, et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. *J Biol Chem*. (2009) 284:34211–22. doi: 10.1074/jbc.M109.041152
71. Guo J, Jayaprakash P, Dan J, Wise P, Jang GB, Liang C, et al. PRAS40 connects microenvironmental stress signaling to exosome-mediated secretion. *Mol Cell Biol*. (2017) 37:e00171-17. doi: 10.1128/MCB.00171-17
72. Wang X, Wang T, Chen C, Wu Z, Bai P, Li S, et al. Serum exosomal miR-210 as a potential biomarker for clear cell renal cell carcinoma. *J Cell Biochem*. (2019) 120:1492–502. doi: 10.1002/jcb.27347
73. Jung KO, Jo H, Yu JH, Gambhir SS, Pratx G. Development and MPI tracking of novel hypoxia-targeted theranostic exosomes. *Biomaterials*. (2018) 177:139–48. doi: 10.1016/j.biomaterials.2018.05.048
74. Becker A, Thakur BK, Weiss JM, Kim HS, Peinado H, Lyden D. Extracellular vesicles in cancer: cell-to-cell mediators of metastasis. *Cancer Cell*. (2016) 30:836–48. doi: 10.1016/j.ccell.2016.10.009
75. Chen G, Zhang Y, Wu XH. 786-O renal cancer cell line-derived exosomes promote 786-O cell migration and invasion *in vitro*. *Oncol Lett*. (2014) 7:1576–80. doi: 10.3892/ol.2014.1962
76. Grange C, Tapparo M, Collino F, Vitillo L, Damasco C, Deregibus MC, et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Res*. (2011) 71:5346–56. doi: 10.1158/0008-5472.CAN-11-0241
77. Lindoso RS, Collino F, Camussi G. Extracellular vesicles derived from renal cancer stem cells induce a pro-tumorigenic phenotype in mesenchymal stromal cells. *Oncotarget*. (2015) 6:7959–69. doi: 10.18632/oncotarget.3503
78. Wang L, Yang G, Zhao D, Wang J, Bai Y, Peng Q, et al. CD103-positive CSC exosome promotes EMT of clear cell renal cell carcinoma: role of remote miR-19b-3p. *Mol Cancer*. (2019) 18:86. doi: 10.1186/s12943-019-0997-z
79. Rana S, Yue S, Stadel D, Zoller M. Toward tailored exosomes: the exosomal tetraspanin web contributes to target cell selection. *Int J Biochem Cell Biol*. (2012) 44:1574–84. doi: 10.1016/j.biocel.2012.06.018
80. Singh A, Fedele C, Lu H, Nevalainen MT, Keen JH, Languino LR. Exosome-mediated transfer of $\alpha_v\beta_3$ integrin from tumorigenic to nontumorigenic cells promotes a migratory phenotype. *Mol Cancer Res*. (2016) 14:1136–46. doi: 10.1158/1541-7786.MCR-16-0058
81. Thery C, Ostrowski M, Segura E. Membrane vesicles as conveyors of immune responses. *Nat Rev Immunol*. (2009) 9:581–93. doi: 10.1038/nri2567
82. Grange C, Tapparo M, Tritta S, Deregibus MC, Battaglia A, Gontero P, et al. Role of HLA-G and extracellular vesicles in renal cancer stem cell-induced inhibition of dendritic cell differentiation. *BMC Cancer*. (2015) 15:1009. doi: 10.1186/s12885-015-2025-z
83. Kochan G, Escors D, Breckpot K, Guerrero-Setas D. Role of non-classical MHC class I molecules in cancer immunosuppression. *Oncoimmunol*. (2013) 2:e26491. doi: 10.4161/onci.26491
84. Yang L, Wu X, Wang D, Luo C, Chen L. Renal carcinoma cell-derived exosomes induce human immortalized line of Jurkat T lymphocyte apoptosis *in vitro*. *Urol Int*. (2013) 91:363–9. doi: 10.1159/000348747
85. Xia Y, Zhang Q, Zhen Q, Zhao Y, Liu N, Li T, et al. Negative regulation of tumor-infiltrating NK cell in clear cell renal cell carcinoma patients through the exosomal pathway. *Oncotarget*. (2017) 8:37783–95. doi: 10.18632/oncotarget.16354
86. Diao JJ, Yang X, Song XD, Chen SY, He YF, Wang QS, et al. Exosomal Hsp70 mediates immunosuppressive activity of the myeloid-derived suppressor cells via phosphorylation of stat3. *Med Oncol*. (2015) 32:453. doi: 10.1007/s12032-014-0453-2
87. Zou WP, Wolchok JD, Chen LP. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: mechanisms, response biomarkers, and combinations. *Sci Transl Med*. (2016) 8:328rv4. doi: 10.1126/scitranslmed.aad7118
88. Page DB, Postow MA, Callahan MK, Allison JP, Wolchok JD. Immune modulation in cancer with antibodies. *Annu Rev Med*. (2014) 65:185–202. doi: 10.1146/annurev-med-092012-112807
89. Theodoraki MN, Yerneni SS, Hoffmann TK, Gooding WE, Whiteside TL. Clinical significance of PD-L1(+) exosomes in plasma of head and neck cancer patients. *Clin Cancer Res*. (2018) 24:896–905. doi: 10.1158/1078-0432.CCR-17-2664
90. Ricklefs FL, Alayo Q, Krenzlin H, Mahmoud AB, Speranza MC, Nakashima H, et al. Immune evasion mediated by PD-L1 on glioblastoma-derived extracellular vesicles. *Sci Adv*. (2018) 4:aar2766. doi: 10.1126/sciadv.aar2766
91. Yang Y, Li CW, Chan LC, Wei YK, Hsu JM, Xia WY, et al. Exosomal PD-L1 harbors active defense function to suppress T cell killing of breast cancer cells and promote tumor growth. *Cell Res*. (2018) 28:862–4. doi: 10.1038/s41422-018-0060-4
92. Chen G, Huang AC, Zhang W, Zhang G, Wu M, Xu W, et al. Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. *Nature*. (2018) 560:382–6. doi: 10.1038/s41586-018-0392-8
93. Song Y, Wu L, Yang C. Exosomal PD-L1: an effective liquid biopsy target to predict immunotherapy response. *Natl Sci Rev*. (2019) 6:1103–4. doi: 10.1093/nsr/nwy154
94. Xie F, Xu M, Lu J, Mao L, Wang S. The role of exosomal PD-L1 in tumor progression and immunotherapy. *Mol Cancer*. (2019) 18:146. doi: 10.1186/s12943-019-1074-3
95. Poggio M, Hu TY, Pai CC, Chu B, Belair CD, Chang A, et al. Suppression of exosomal PD-L1 induces systemic anti-tumor immunity and memory. *Cell*. (2019) 177:414–27. doi: 10.1016/j.cell.2019.02.016
96. Huang M, Yang J, Wang T, Song J, Xia J, Wu L, et al. Homogeneous, low-volume, efficient and sensitive quantitation of circulating exosomal PD-L1 for cancer diagnosis and immunotherapy response prediction. *Angew Chem Int Ed Engl*. (2020) 59:4800–5. doi: 10.1002/ange.201916039
97. Pang Y, Shi J, Yang X, Wang C, Sun Z, Xiao R. Personalized detection of circling exosomal PD-L1 based on Fe_3O_4@TiO_2 isolation and SERS immunoassay. *Biosens Bioelectron*. (2020) 148:111800. doi: 10.1016/j.bios.2019.111800
98. Sousa D, Lima RT, Vasconcelos MH. Intercellular transfer of cancer drug resistance traits by extracellular vesicles. *Trends Mol Med*. (2015) 21:595–608. doi: 10.1016/j.molmed.2015.08.002
99. Corcoran C, Rani S, O'Brien K, O'Neill A, Prencipe M, Sheikh R, et al. Docetaxel-resistance in prostate cancer: evaluating associated phenotypic changes and potential for resistance transfer via exosomes. *PLoS ONE*. (2012) 7:e50999. doi: 10.1371/journal.pone.0050999

100. Yin J, Sun W, Li F, Hong J, Li X, Zhou Y, et al. VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res*. (2020) 48:D1042–50. doi: 10.1093/nar/gkz878
101. Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Eng J Med*. (2007) 356:115–24. doi: 10.1056/NEJMoa065044
102. Rini BI, Atkins MB. Resistance to targeted therapy in renal-cell carcinoma. *Lancet Oncol*. (2009) 10:992–1000. doi: 10.1016/S1470-2045(09)70240-2
103. Qu L, Ding J, Chen C, Wu ZJ, Liu B, Gao Y, et al. Exosome-transmitted lncARSR promotes sunitinib resistance in renal cancer by acting as a competing endogenous RNA. *Cancer Cell*. (2016) 29:653–68. doi: 10.1016/j.ccell.2016.03.004
104. Tang Z, Li D, Hou S, Zhu X. The cancer exosomes: clinical implications, applications and challenges. *Int J Cancer*. (2020) 11:2946–59. doi: 10.1002/ijc.32762
105. Zhao A, Li GR, Peoc'h M, Genin C, Gigante M. Serum miR-210 as a novel biomarker for molecular diagnosis of clear cell renal cell carcinoma. *Exp Mol Pathol*. (2013) 94:115–20. doi: 10.1016/j.yexmp.2012.10.005
106. Zhang W, Ni M, Su Y, Wang H, Zhu S, Zhao A, et al. MicroRNAs in serum exosomes as potential biomarkers in clear-cell renal cell carcinoma. *Eur Urol Focus*. (2018) 4:412–9. doi: 10.1016/j.euf.2016.09.007
107. Cheng TT, Wang LN, Li YY, Huang C, Zeng LX, Yang J. Differential microRNA expression in renal cell carcinoma. *Oncol Lett*. (2013) 6:769–76. doi: 10.3892/ol.2013.1460
108. Chen X, Lou N, Ruan A, Qiu B, Yan Y, Wang X, et al. miR-224/miR-141 ratio as a novel diagnostic biomarker in renal cell carcinoma. *Oncol Lett*. (2018) 16:1666–74. doi: 10.3892/ol.2018.8874
109. Fujii N, Hirata H, Ueno K, Mori J, Oka S, Shimizu K, et al. Extracellular miR-224 as a prognostic marker for clear cell renal cell carcinoma. *Oncotarget*. (2017) 8:109877–88. doi: 10.18632/oncotarget.22436
110. Du MJ, Giridhar KV, Tian YJ, Tschannen MR, Zhu J, Huang CC, et al. Plasma exosomal miRNAs-based prognosis in metastatic kidney cancer. *Oncotarget*. (2017) 8:63703–14. doi: 10.18632/oncotarget.19476
111. Butz H, Nofech-Mozes R, Ding Q, Khella HWZ, Szabó PM, Jewett M, et al. Exosomal microRNAs are diagnostic biomarkers and can mediate cell-cell communication in renal cell carcinoma. *Eur Urol Focus*. (2016) 2:210–8. doi: 10.1016/j.euf.2015.11.006
112. Crentsil VC, Liu H, Sellitti DF. Comparison of exosomal microRNAs secreted by 786-O clear cell renal carcinoma cells and HK-2 proximal tubule-derived cells in culture identifies microRNA-205 as a potential biomarker of clear cell renal carcinoma. *Oncol Lett*. (2018) 16:1285–90. doi: 10.3892/ol.2018.8751
113. Liu Y, Zheng X, Yu Q, Wang H, Tan F, Zhu Q, et al. Epigenetic activation of the drug transporter OCT2 sensitizes renal cell carcinoma to oxaliplatin. *Sci Transl Med*. (2016) 8:348ra97. doi: 10.1126/scitranslmed.aaf3124
114. Chen L, Chen L, Qin Z, Lei J, Ye S, Zeng K, et al. Upregulation of miR-489-3p and miR-630 inhibits oxaliplatin uptake in renal cell carcinoma by targeting OCT2. *Acta Pharm Sin B*. (2019) 9:1008–20. doi: 10.1016/j.apsb.2019.01.002
115. de Palma G, Sallustio F, Curci C, Galleggiante V, Rutigliano M, Serino G, et al. The three-gene signature in urinary extracellular vesicles from patients with clear cell renal cell carcinoma. *J Cancer*. (2016) 7:1960–7. doi: 10.7150/jca.16123
116. Del Boccio P, Raimondo F, Pieragostino D, Morosi L, Cozzi G, Sacchetta P, et al. A hyphenated microLC-Q-TOF-MS platform for exosomal lipidomics investigations: application to RCC urinary exosomes. *Electrophoresis*. (2012) 33:689–96. doi: 10.1002/elps.201100375
117. Raimondo F, Morosi L, Corbetta S, Chinello C, Brambilla P, Della Mina P, et al. Differential protein profiling of renal cell carcinoma urinary exosomes. *Mol Biosyst*. (2013) 9:1220–33. doi: 10.1039/c3mb25582d
118. Jingushi K, Uemura M, Ohnishi N, Nakata W, Fujita K, Naito T, et al. Extracellular vesicles isolated from human renal cell carcinoma tissues disrupt vascular endothelial cell morphology via azurocidin. *Int J Cancer*. (2018) 142:607–17. doi: 10.1002/ijc.31080
119. Kamerkar S, LeBleu VS, Sugimoto H, Yang S, Ruivo CF, Melo SA, et al. Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature*. (2017) 546:498–503. doi: 10.1038/nature22341
120. Mendt M, Kamerkar S, Sugimoto H, McAndrews KM, Wu CC, Gagea M, et al. Generation and testing of clinical-grade exosomes for pancreatic cancer. *JCI Insight*. (2018) 3:e99263. doi: 10.1172/jci.insight.99263
121. Wiklander OPB, Brennan MÁ, Lötvall J, Breakefield XO, El Andaloussi S. Advances in therapeutic applications of extracellular vesicles. *Sci Transl Med*. (2019) 11:eaav8521. doi: 10.1126/scitranslmed.aav8521
122. Tian Y, Li S, Song J, Ji T, Zhu M, Anderson GJ, et al. A doxorubicin delivery platform using engineered natural membrane vesicle exosomes for targeted tumor therapy. *Biomaterials*. (2014) 35:2383–90. doi: 10.1016/j.biomaterials.2013.11.083
123. Wan Y, Wang L, Zhu C, Zheng Q, Wang G, Tong J, et al. Aptamer-conjugated extracellular nanovesicles for targeted drug delivery. *Cancer Res*. (2018) 78:798–808. doi: 10.1158/0008-5472.CAN-17-2880
124. Oosterwijk E, Ruiter DJ, Hoedemaeker PJ, Pauwels EK, Jonas U, Zwartendijk J, et al. Monoclonal antibody G250 recognizes a determinant present in renal-cell carcinoma and absent from normal kidney. *Int J Cancer*. (1986) 38:489–94. doi: 10.1002/ijc.2910380406
125. Zhang Y, Luo CL, He BC, Zhang JM, Cheng G, Wu XH. Exosomes derived from IL-12-anchored renal cancer cells increase induction of specific antitumor response *in vitro*: a novel vaccine for renal cell carcinoma. *Int J Oncol*. (2010) 36:133–40. doi: 10.3892/ijo_00000484
126. van der Grein SG, Defourny KAY, Slot EFJ, Nolte-'t Hoen ENM. Intricate relationships between naked viruses and extracellular vesicles in the crosstalk between pathogen and host. *Semin Immunopathol*. (2018) 40:491–504. doi: 10.1007/s00281-018-0678-9
127. Zomer A, Maynard C, Verweij FJ, Kamermans A, Schafer R, Beerling E, et al. *In vivo* imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. *Cell*. (2015) 161:1046–57. doi: 10.1016/j.cell.2015.04.042
128. Gangadaran P, Hong CM, Ahn BC. An update on *in vivo* imaging of extracellular vesicles as drug delivery vehicles. *Front Pharmacol*. (2018) 9:169. doi: 10.3389/fphar.2018.00169
129. Datta A, Kim H, McGee L, Johnson AE, Talwar S, Marugan J, et al. High-throughput screening identified selective inhibitors of exosome biogenesis and secretion: a drug repurposing strategy for advanced cancer. *Sci Rep*. (2018) 8:8161. doi: 10.1038/s41598-018-26411-7
130. Catalano M, O'Driscoll L. Inhibiting extracellular vesicles formation and release: a review of EV inhibitors. *J Extracell Vesicles*. (2020) 9:1703244. doi: 10.1080/20013078.2019.1703244
131. Ortiz A, Gui J, Zahedi F, Yu P, Cho C, Bhattacharya S, et al. An interferon-driven oxysterol-based defense against tumor-derived extracellular vesicles. *Cancer Cell*. (2019) 35:33–45.e6. doi: 10.1016/j.ccell.2018.12.001
132. Im E-J, Lee C-H, Moon P-G, Rangaswamy GG, Lee B, Lee JM, et al. Sulfisoxazole inhibits the secretion of small extracellular vesicles by targeting the endothelin receptor A. *Nat Commun*. (2019) 10:1387. doi: 10.1038/s41467-019-09387-4
133. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552
134. Tabibu S, Vinod PK, Jawahar CV. Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. *Sci Rep*. (2019) 9:10509. doi: 10.1038/s41598-019-46718-3
135. Han S, Hwang SI, Lee HJ. The classification of renal cancer in 3-phase CT images using a deep learning method. *J Digit Imaging*. (2019) 32:638–43. doi: 10.1007/s10278-019-00230-2
136. Ito K, Ogawa Y, Yokota K, Matsumura S, Minamisawa T, Suga K, et al. Host cell prediction of exosomes using morphological features on solid surfaces analyzed by machine learning. *J Phys Chem B*. (2018) 122:6224–35. doi: 10.1021/acs.jpcb.8b01646
137. Gomez-de-Mariscal E, Maska M, Kotrbova A, Pospichalova V, Matula P, Munoz-Barrutia A. Deep-learning-based segmentation of small extracellular vesicles in transmission electron microscopy images. *Sci Rep*. (2019) 9:13211. doi: 10.1038/s41598-019-49431-3
138. Borgovan T, Crawford L, Nwizu C, Quesenberry P. Stem cells and extracellular vesicles: biological regulators of physiology and disease. *Am J Physiol Cell Physiol*. (2019) 317:C155–66. doi: 10.1152/ajpcell.00017.2019
139. Won Y, Song HJ, Kang TW, Kim JJ, Han BD, Lee SW. Pattern analysis of serum proteome distinguishes renal cell carcinoma from

other urologic diseases and healthy persons. *Proteomics*. (2003) 3:2310–6. doi: 10.1002/pmic.200300590
140. Zheng H, Ji J, Zhao L, Chen M, Shi A, Pan L, et al. Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps. *Oncotarget*. (2016) 7:59189–98. doi: 10.18632/oncotarget.10830
141. Gezsi A, Kovacs A, Visnovitz T, Buzas EI. Systems biology approaches to investigating the roles of extracellular vesicles in human diseases. *Exp Mol Med*. (2019) 51:1–11. doi: 10.1038/s12276-019-0226-2
142. Liu T, Zhang Q, Zhang J, Li C, Miao YR, Lei Q, et al. EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res*. (2019) 47:D89–93. doi: 10.1093/nar/gky985
143. Kim DK, Kang B, Kim OY, Choi DS, Lee J, Kim SR, et al. EVpedia: an integrated database of high-throughput data for systemic analyses of extracellular vesicles. *J Extracell Vesicles*. (2013) 2:20384. doi: 10.3402/jev.v2i0.20384
144. Consortium E-T, van Deun J, Mestdagh P, Agostinis P, Akay O, Anand S, et al. EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat Methods*. (2017) 14:228–32. doi: 10.1038/nmeth.4185
145. Mathivanan S, Simpson RJ. ExoCarta: a compendium of exosomal proteins and RNA. *Proteomics*. (2009) 9:4997–5000. doi: 10.1002/pmic.200900351
146. Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, et al. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res*. (2018) 46:D106–12. doi: 10.1093/nar/gkx891
147. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, et al. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res*. (2015) 43:D1057–63. doi: 10.1093/nar/gku1113
148. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma proteome database as a resource for proteomics research: 2014 update. *Nucleic Acids Res*. (2014) 42:D959–D65. doi: 10.1093/nar/gkt1251
149. Pisitkun T, Shen RF, Knepper MA. Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci USA*. (2004) 101:13368–73. doi: 10.1073/pnas.0403453101
150. Gonzales PA, Pisitkun T, Hoffert JD, Tchapyjnikov D, Star RA, Kleta R, et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol*. (2009) 20:363–79. doi: 10.1681/ASN.2008040406
151. Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, et al. Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol*. (2012) 10:e1001450. doi: 10.1371/journal.pbio.1001450
152. Chen C, Zong S, Liu Y, Wang Z, Zhang Y, Chen B, et al. Profiling of exosomal biomarkers for accurate cancer identification: combining DNA-PAINT with machine-learning-based classification. *Small*. (2019) 15:e1901014. doi: 10.1002/smll.201901014
153. Ko J, Bhagwat N, Yee SS, Ortiz N, Sahmoud A, Black T, et al. Combining machine learning and nanofluidic technology to diagnose pancreatic cancer using exosomes. *ACS Nano*. (2017) 11:11182–93. doi: 10.1021/acsnano.7b05503
154. Ko J, Bhagwat N, Black T, Yee SS, Na YJ, Fisher S, et al. miRNA profiling of magnetic nanopore-isolated extracellular vesicles for the diagnosis of pancreatic cancer. *Cancer Res*. (2018) 78:3688–97. doi: 10.1158/0008-5472.CAN-17-3703
155. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. (2019) 18:463–77. doi: 10.1038/s41573-019-0024-5
156. Zhu H. Big data and artificial intelligence modeling for drug discovery. *Annu Rev Pharmacol Toxicol*. (2020) 60:573–89. doi: 10.1146/annurev-pharmtox-010919-023324
157. Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep embeddings of gene expression profiles. *Mol Pharma*. (2018) 15:4314–25. doi: 10.1021/acs.molpharmaceut.8b00284
158. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, et al. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res*. (2020) 48:D1031–41. doi: 10.1093/nar/gkz981
159. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the drugBank database for 2018. *Nucleic Acids Res*. (2018) 46:D1074–82. doi: 10.1093/nar/gkx1037

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Qin, Xu, Hu, Yu and Zeng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/fphar.2020.00544

Identification of Prognostic miRNA Signature and Lymph Node Metastasis-Related Key Genes in Cervical Cancer

Shuoling Chen[1,2†], *Chang Gao*[1,3†], *Yangyuan Wu*[1,2] *and Zunnan Huang*[1,4*]

[1] *Key Laboratory of Big Data Mining and Precision Drug Design of Guangdong Medical University, Research Platform Service Management Center, Guangdong Medical University, Dongguan, China,* [2] *The Second School of Clinical Medicine, Guangdong Medical University, Dongguan, China,* [3] *Key Laboratory for Research and Development of Natural Drugs of Guangdong Province, School of Pharmacy, Guangdong Medical University, Dongguan, China,* [4] *Institute of Marine Biomedical Research, Guangdong Medical University, Zhanjiang, China*

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Bifang He,
Guizhou University, China
Kangyu Lin,
University of Texas MD Anderson Cancer Center, United States

****Correspondence:***
Zunnan Huang
zn_huang@yahoo.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *27 February 2020*
Accepted: *09 April 2020*
Published: *08 May 2020*

Citation:
Chen S, Gao C, Wu Y and Huang Z (2020) Identification of Prognostic miRNA Signature and Lymph Node Metastasis-Related Key Genes in Cervical Cancer. Front. Pharmacol. 11:544. doi: 10.3389/fphar.2020.00544

Background: miRNAs and genes can serve as biomarkers for the prognosis and therapy of cervical tumors whose metastasis into lymph nodes is closely associated with disease progression and poor prognosis.

Methods: R software and Bioconductor packages were employed to identify differentially expressed miRNAs (DEMs) from The Cancer Genome Atlas (TCGA) database. GEO2R detected differentially expressed genes (DEGs) in the GSE7410 dataset originating from the Gene Expression Omnibus (GEO). A Cox proportional hazard regression model was established to select prognostic miRNA biomarkers. Online tools such as TargetScan and miRDB predicted target genes, and overlapping DEGs and target genes were defined as consensus genes. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment and Gene Ontology (GO) function annotations were performed to discern the potential functions of consensus genes. STRING and Cytoscape screened key genes and constructed a regulatory network.

Results: A combination of four miRNAs (down-regulated miR-502 and miR-145, up-regulated miR-142 and miR-33b) was identified as an independent prognostic signature of cervical cancer. A total of 94 consensus genes were significantly enriched in 7 KEGG pathways and 19 GO function annotations including the cAMP signaling pathway, the plasma membrane, integral components of the plasma membrane, cell adhesion, etc. The module analysis suggested that *CXCL12*, *IGF1*, *PTPRC CDH5*, *RAD51B*, *REV3L*, and *WDHD1* are key genes that significantly correlate with cervical cancer lymph node metastasis.

Conclusions: This study demonstrates that a four-miRNA signature can be a prognostic biomarker, and seven key genes are significantly associated with lymph node metastasis in cervical cancer patients. These miRNAs and key genes have the potential to be therapeutic targets for cervical cancer. Among them, two miRNAs (miR-502 and miR-33b) and two key genes (*PTPRC* and *CDH5*) were first reported to be potential novel

biomarkers for cervical cancer. The current study further characterizes the progression of lymph node metastasis and mechanism of cervical tumors; therefore, it provides a novel diagnostic indicator and therapeutic targets for future clinical treatments.

Keywords: cervical cancer, miRNA, key gene, prognostic signature, lymph node metastasis

INTRODUCTION

Worldwide, cervical cancer is the fourth most common female malignancy with an extremely high morbidity and mortality rate (Cohen et al., 2019). According to the World Health Organization (WHO), there were 569,874 new cases of cervical tumors in 2018, accounting for 3.2% of all new cancer cases in that same year (Bray et al., 2018). Most cervical tumors are attributed to recurrent HPV infections. Aberrant expression of oncoproteins encoded by HPV genetic material, such as E6 and E7, partially leads to epigenetic instability, which affects the carcinogenesis and metastasis of cervical cancer (Doorbar, 2006).

MicroRNAs (miRNAs) are associated with the development of a wide range of cancers, including cervical cancer (Park et al., 2017). miRNAs are a group of non-coding, single-stranded RNA molecules that are approximately 22 nucleotides (nt) in length and encoded by endogenous genes (Bartel, 2004; Chen and Kang, 2015). Abnormally expressed miRNAs regulate various biological processes such as apoptosis, proliferation, and metabolism (Zhang et al., 2007; Lee and Dutta, 2009). Moreover, miRNA dysregulation plays a critical role in cancer progression and metastasis in multiple cancers, including cervical cancer (Gómez-Gómez et al., 2013; Huang et al., 2017). For example, Yao et al. (Yao et al., 2018) demonstrated that decreased expression of HPGD by miR-146b-3p induced proliferation, migration, and anchorage-independent growth of cervical cancer cells by activating the STAT3 and AKT signaling pathways. This indicated that miRNAs may be clinically applicable as potential biomarkers and therapeutic targets.

Tumor metastasis is the leading cause of death for most cervical cancer patients. It generally involves complex processes such as extracellular matrix degradation, lymphangiogenesis, angiogenesis, clonal growth at secondary sites, etc. (Zhang et al., 2015). The metastasis and invasion of tumor cells through blood and lymph nodes are crucial processes in the progression of cervical cancer (Dai et al., 2017). Newly formed lymphatic vessels are comprised of endothelial cells that are not tightly connected. Therefore, cervical cancer cells can easily invade this endothelial layer and metastasize to lymph nodes (Zhang et al., 2015). In addition, the poor prognosis of cervical cancer patients is correlated with high invasiveness and diffuse lymph node metastasis (Wang et al., 2016). Lymph node metastasis is the main metastatic pathway and the most critical factor in the prognosis and recurrence of cervical cancer cases (Huang and Fang, 2018). However, a detailed understanding of the specific signatures and molecular mechanisms of lymph node metastasis is lacking. Therefore, the search for a reliable lymphatic signature is important in determining malignant metastasis, and it also provides information that can guide the clinical treatment of cervical tumor patients.

Many have proposed conducting a large-scale systematic analysis of miRNAs, genes, and clinical data using bioinformatics to further characterize the functions of miRNAs and genes in certain disease states, clarify their potential as disease-related signatures, and discover new disease biomarkers and drug targets (Liang et al., 2017). Researchers can obtain tumor data from public databases and conduct differential expression, survival, and prognosis analyses as well as target gene prediction and functional characterization using R language (Sepulveda, 2020), TargetScan (Agarwal et al., 2015), miRDB (Wong and Wang, 2015), DAVID (Huang et al., 2009), etc. With these tools, tumor biomarkers can be screened, and their mechanism of action can be further elucidated. For example, Liang et al. (Liang et al., 2017) identified a three-miRNA signature (miR-145, miR-200c, and miR-218-1) that is a prognostic factor of cervical tumors by conducting Cox univariate and multivariate analyses on differentially expressed miRNAs screened from clinical samples in TCGA database.

In this study, we conducted a multi-step analysis using various R language packages on clinical samples downloaded from the TCGA (Tomczak et al., 2015) and GEO (Barrett et al., 2007) databases to identify DEMs and DEGs. A Cox proportional hazard regression model was then established to determine potential prognostic biomarkers from the available DEMs. Subsequently, the target genes of the miRNAs biomarkers were predicted by the TargetScan and miRDB online tools and the consensus genes were further determined based on overlap between DEGs and these target genes. Finally, MCODE (Bader

Abbreviations: AUC, area under the curve; BP, biological process; CC, cellular component; CDH5, cadherin 5; CEA, carcinoembryonic antigen; CI, confidence interval; Coef, partial regression coefficient; CXCL12, C-X-C motif chemokine ligand 12; CXCR4, C-X-C motif chemokine receptor 4; CXCR7, C-X-C motif chemokine receptor 7; DAVID, Database for Annotation, Visualization, and Integrated Discovery; DEGs, differentially expressed genes; DEMs, differentially expressed miRNAs; EMT, epithelial-mesenchymal transition; Exp (Coef), relative risk; FDR, false discovery rate; GDF15, growth differentiation factor 15; GEO, Gene Expression Omnibus; GO, gene ontology; HMGA2, high-mobility group AT-hook 2; HR, hazard ratio; IGF1, insulin like growth factor 1; IGF1R, insulin like growth factor 1 receptor; KEGG, Kyoto Encyclopedia of Genes and Genome; LNM, lymph node metastasis; MALAT1, metastasis-associated lung adenocarcinoma transcript-1; MCODE, Molecular Complex Detection; MF, molecular function; Pol ζ, DNA translation synthesis polymerase ζ; PPI, protein-protein interaction; PTPRC, protein tyrosine phosphatase receptor type C; qRT-PCR, quantitative real-time polymerase chain reaction (reverse transcription-quantitative polymerase chain reaction (RT-qPCR)); RAD51B, RAD51 paralog B; REV3L, recovery protein 3 like; Se (Coef), standard error of regression coefficient; siRNA, small interfering RNA; STRING, Search Tool for the Retrieval of Interacting Genes; TCGA, The Cancer Genome Atlas; TNM, tumor node metastasis; WDHD1, WD repeat and HMG-box DNA binding protein 1.

and Hogue, 2003) software in Cytoscape (Shannon et al., 2003) was used to identify key genes related to lymph node metastasis caused by early-stage cervical tumors. Together, the prognostic miRNAs and key genes mined in this study will provide new insights in elucidating the molecular mechanisms of cervical cancer and contribute to finding new therapeutic targets and prognostic biomarkers for cervical cancer patients.

MATERIALS AND METHODS

Data Preparation and Differential Expression Analysis

The expression profiles and clinical data for 310 samples were obtained from TCGA (https://portal.gdc.cancer.gov/) database on July 27, 2018. These samples included 307 primary solid cervical cancer tissue samples and three normal tissue samples. Even though this study used many more cancer samples than normal samples, previous studies by Yang et al. (2019), Zeng et al. (2018), etc. have demonstrated that appropriate prognostic models with diagnostic indicators and clinically significant therapeutic targets can be obtained with similar unbalanced sample sets. DEMs were analyzed using the edgeR (Robinson et al., 2010), gplots (Li S. et al., 2020), and limma (Ritchie et al., 2015) R language packages (Version 3.5.1) according to the screening criteria of $|\log2FC| > 1$ and $P_{adj} < 0.05$. The GSE7410 expression profile data from 24 samples were obtained from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database. The samples included 19 early-stage cervical tumor tissues with lymph node metastasis and five healthy cervical tissues. DEGs were analyzed using GEO2R based on the screening criteria of $|\log2FC| > 1$ and $P_{adj} < 0.05$.

Establishment of a Cox Proportional Hazard Regression Model

The Cox proportional hazard regression model (George et al., 2014) was established to analyze the association between DEMs and overall survival. The survival package was used in the univariate and multivariate Cox analyses of DEMs (Liang et al., 2017). miRNAs with $P < 0.05$ as calculated by the univariate Cox analysis were considered strongly correlated with overall patient survival. The multivariate Cox analysis used stepwise regression to screen a prognostic model based on the Akaike information criterion (AIC) value (Vrieze, 2012), where the model with the smallest AIC value contains the smallest number of miRNAs that best predict cervical cancer patient prognosis. miRNAs with $P < 0.05$ as calculated by the multivariate Cox analysis were considered as independent prognostic factors.

Prognostic Model Construction

Using the results of the multivariate Cox analysis, the risk score was calculated as follows: Risk Score = Exp ($miRNA_1$) × β_1 + Exp ($miRNA_2$) × β_2 +… + Exp ($miRNA_n$) × β_n. The patients were divided into low- and high-risk groups based on their median risk score. A risk score curve was plotted to demonstrate the risk score differences according to the model. A survival status map was plotted to demonstrate the survival status of each sample. A heatmap was plotted to demonstrate the expression level differences of the four prognostic miRNAs in the low- and high-risk groups. A survival curve was plotted to demonstrate the 5-year survival in high- and low-risk groups. And an ROC curve of the model was constructed to determine its predictive ability. A model with an AUC value larger than 0.7 possesses a strong prediction function.

Target Gene Prediction

TargetScan and miRDB predicted the target genes of miRNAs obtained from the multivariate Cox analysis. Target genes that were predicted by both tools were considered target genes of those miRNAs. And the consensus genes were further obtained from the overlap between DEGs related to lymph node metastasis and candidate target genes from the prognosis miRNAs.

KEGG Pathways and GO Function Annotation Enrichment Analysis

To further understand the underlying biological significance of DEMs and DEGs, DAVID (https://david.ncifcrf.gov/) produced KEGG pathway (Kanehisa and Goto, 2000) enrichment and GO functional annotations (Ashburner et al., 2000) of the consensus genes using $P < 0.05$ as a demarcation criterion. The pathways or annotations with the smallest *P*-value or the largest count value were considered as key KEGG pathways and GO function annotations. GO function annotations include three parts: biological process (BP), cellular component (CC), and molecular function (MF).

PPI Network Construction and Module Analysis

STRING (http://string-db.org) (Szklarczyk et al., 2017) analyzed the consensus genes and obtained protein interaction data. Proteins with a minimum required interaction score greater than or equal to 0.400 were selected to construct the PPI (protein-protein interaction) network, and nodes with network interruption were hidden. The combined score was imported into Cytoscape software (Version 3.7.1, https://cytoscape.org/). The Molecular Complex Detection (MCODE) plug-in in Cytoscape calculated the MCODE score and selected the significant modules of key genes using the screening criteria of Degree Cut-off = 2, Haircut on, Node Score Cut-off = 0.2, k-core = 2, and Max. Depth = 100. Moreover, the logFC value of genes in the interaction network was also imported into Cytoscape to show the up/down regulation status.

Network Visualization

Regulatory relationship data were imported into Cytoscape along with miRNAs, key genes, consensus genes, key KEGG pathways, and GO functional annotations. The visualization network was performed with Cytoscape to explore the potential regulatory relationship among them.

Analysis Procedure

Figure 1 shows the analysis procedure of the data mining processes that was used to screen tumor biomarkers and key genes in this study. It was based on the extensive use of R

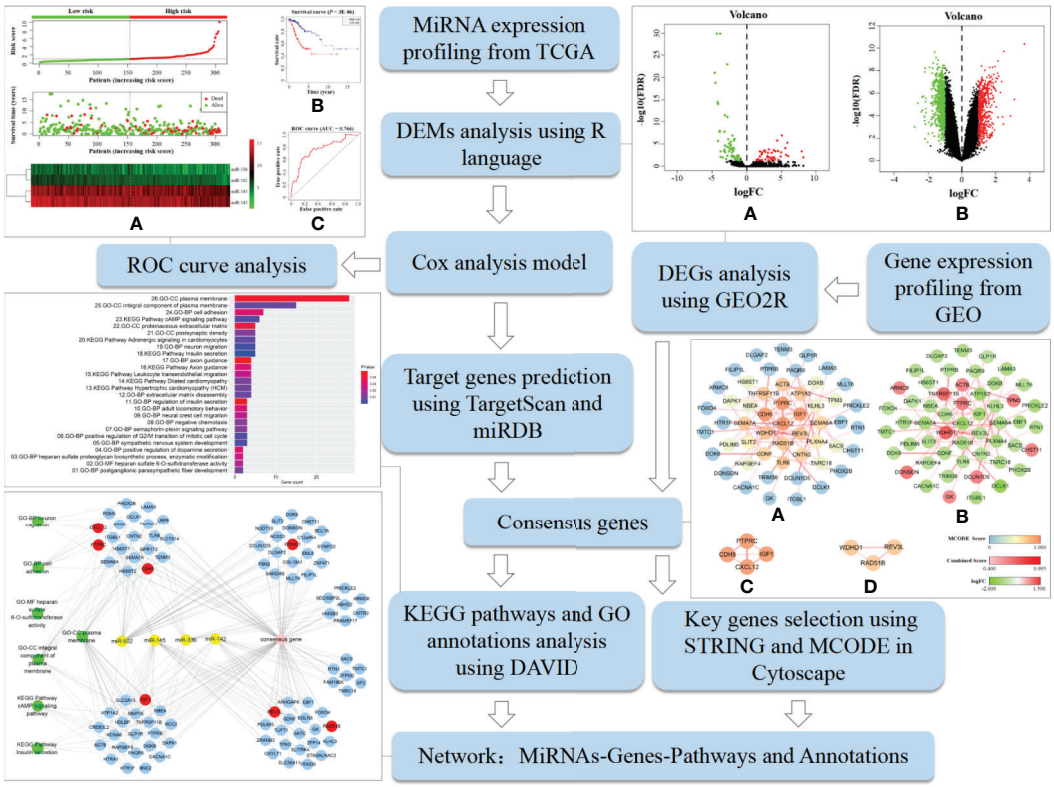


FIGURE 1 | Analysis procedure of the data mining process used to screen tumor biomarkers and key genes in this study. It includes specific bioinformatics methods, data processing tools, and partial research results.

language and various online analytical tools. We first obtained DEMs and DEGs using R language and GEO2R, respectively, by analyzing cervical cancer-related miRNAs and gene expression profiles downloaded from TCGA and GEO databases. A prognostic model was constructed *via* Cox analysis to detect key miRNAs to form a best-available prognostic signature whose target genes were predicted by TargetScan and miRDB. Subsequently, we performed KEGG pathway and GO annotation analyses using DAVID to clarify the function of the consensus genes from the overlap between DEGs and the predicted target genes of key miRNAs. Furthermore, STRING and MCODE selected key genes among consensus genes. Finally, we constructed a miRNAs-Genes-Pathways and Annotations network to elucidate potential regulatory mechanisms. The approach described here is a feasible protocol to identify potential tumor biomarkers and key genes.

RESULTS

miRNA and Gene Differential Expression Analyses

A total of 110 DEMs were obtained after analyzing miRNA expression profiles from TCGA with R language using $P_{adj} < 0.05$ and $|\log2FC| > 1$ as screening criteria. Among them, 64 miRNAs were significantly down-regulated, and 46 miRNAs were significantly up-regulated (**Table S1**). At the same time, 1840 DEGs related to early cervical cancer lymph node metastasis were identified by analyzing the GSE7410 expression profile with GEO2R using $P_{adj} < 0.05$ and $|\log2FC| > 1$ as screening criteria. Among them, 1,298 genes were significantly down-regulated, and 542 genes were significantly up-regulated (**Table S2**). The volcano map illustrates the significant differences and distribution of the fold change in DEMs and DEGs (**Figure 2**).

miRNA-Based Signature Identification as a Prognostic Biomarker

A total of 15 miRNAs related to patient survival were obtained from 110 DEMs using univariate Cox analysis ($P < 0.05$) (**Table 1**). Four miRNAs related to patient prognosis (miR-502, miR-142, miR-145, and miR-33b) were further screened from the above 15 miRNAs by multivariate Cox analysis (**Table 2**). Among them, miR-502 and miR-145 were down-regulated, and miR-142 and miR-33b were up-regulated in cervical cancer tissues. The multivariate Cox analysis demonstrated that these four miRNAs could be used as independent prognostic factors in cervical cancer ($P < 0.05$).

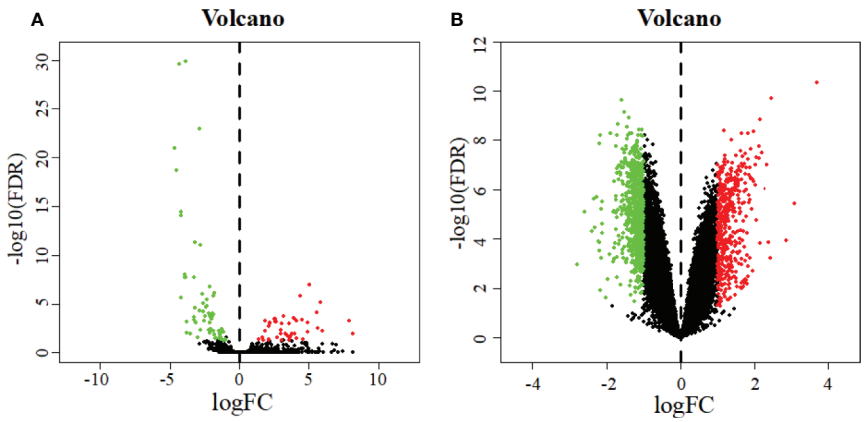


FIGURE 2 | Volcano plot of DEMs **(A)** and DEGs **(B)**. The abscissa represents the log2 transformation value of the differential expression fold change between the cervical cancer samples and the normal samples. The larger the | logFC | value is, the greater the fold change is. The ordinate represents the -log10 transformation value of the FDR value. The larger the -log10 transformation value is, the more significant the difference is. Green dots represent significantly down-regulated miRNAs or genes. Red dots represent significantly up-regulated miRNAs or genes.

TABLE 1 | Univariate analysis of cervical cancer patients.

miRNA	HR	z	*P*
miR-502	0.47295993	−4.710714692	2.47E-06
miR-142	0.726460896	−3.311319458	0.000928571
miR-362	0.634317782	−3.089446858	0.002005296
miR-101-2	0.70026668	−2.832903633	0.004612729
miR-101-1	0.699830285	−2.831809241	0.004628545
miR-145	0.753658772	−2.628062686	0.008587269
miR-1468	0.71581814	−2.570507809	0.010154954
miR-204	0.852831645	−2.377156425	0.017446688
miR-140	0.633673302	−2.368455333	0.017862537
miR-33b	0.803090114	−2.212258218	0.026948828
miR-126	0.700379333	−2.186430538	0.028784121
miR-218-1	0.815625887	−2.121987729	0.033838769
miR-504	0.757023197	−2.003615093	0.045111308
miR-99a	0.879776376	−1.986775064	0.046947329
miR-331	0.691475709	−1.965107538	0.049401792

Bolded miRNAs are the prognostic miRNAs.

TABLE 2 | Multivariate analysis of cervical cancer patients.

	Coef	Exp (Coef)	Se (Coef)	z	*P*
miR-502	−0.676	0.509	0.156	−4.33	1.50E-05
miR-142	−0.297	0.743	0.106	−2.82	0.0049
miR-145	−0.29	0.748	0.101	−2.87	0.0042
miR-33b	−0.209	0.811	0.101	−2.07	0.0382

Application of a Cox Proportional Hazard Regression Model in Disease Prognosis

The prognostic model consisting of four miRNAs was constructed by multivariate cox analysis. The formula of the model was defined as follows: Risk score = (Exp (miR-502) × (−0.676) + Exp (miR-145) × (−0.290) + Exp (miR-142) × (−0.297) + Exp (miR-33b) × (−0.209)). The survival risk score based on the four miRNAs was calculated by the model formula and divided patients into low- (154 samples) and high-risk (153 samples) groups according to their median risk score. The risk score curve demonstrates that the risk score of individuals in the low-risk group is small and consistent while the risk score of individuals in the high-risk group is larger and rises significantly (**Figure 3A** top). The survival status map demonstrates that

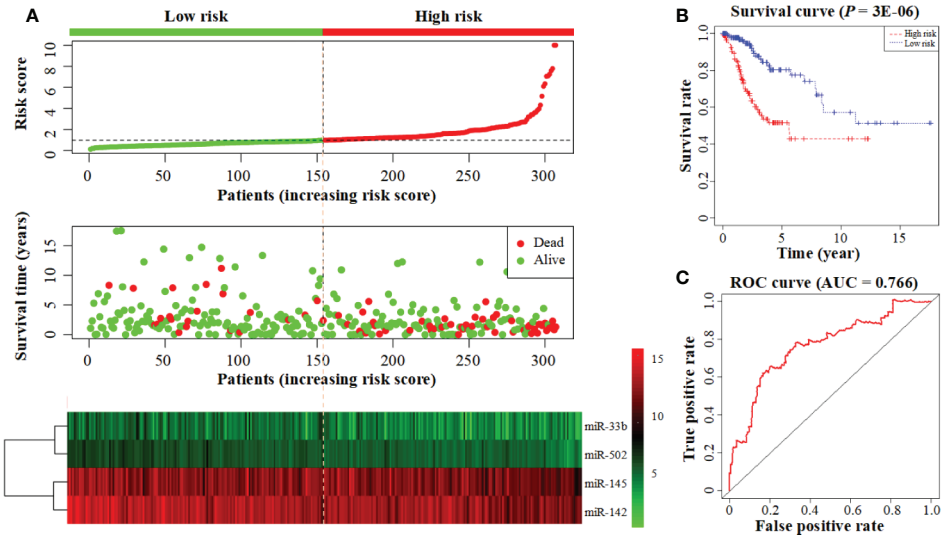


FIGURE 3 | Construction of a prognostic model based on a 4-miRNA signature. **(A)** From top to bottom: the risk score curve, survival status map, and expression heatmap between the low- and high-risk groups. The color bar shows the relative miRNA expression value with red indicating high expression and green indicating low expression. **(B)** Survival curve for the low- and high-risk groups. **(C)** The ROC curve for survival predictions.

individuals in the low-risk group have low mortality rates, and individuals in the high-risk group have a short survival time (most survived for less than 5 years) (**Figure 3A** middle). The heatmap demonstrates that low miRNA expression levels are associated with high-risk scores, and high miRNA expression levels are associated with low-risk scores (**Figure 3A** bottom). These all demonstrate the reliability of the prognostic model in classifying high- and low-risk groups. And the survival curve demonstrates that the survival rate in the low-risk group is notably higher than in the high-risk group (P = 3E-06; low-risk group: 80.3%, 95% CI = 71.3–90.4%; high-risk group: 51.6%; 95% CI = 41.6–64.0%) (**Figure 3B**). The time-dependent ROC curve analysis indicates that this prognostic model has a high level of credibility, sensitivity, and specificity with an AUC value of 0.766 (**Figure 3C**).

Target Gene Prediction for the Four Prognostic miRNAs

The prediction results combined with TargetScan and miRDB identified 267 target genes of miR-502, 423 target genes of miR-145, 370 target genes of miR-142, and 279 target genes of miR-33b (**Figure S1**). After excluding 111 repetitive target genes co-regulated by multiple miRNAs, 1,228 target genes of the four prognostic miRNAs were obtained. Moreover, 94 consensus genes (**Table 3**) were identified from the overlap between 1840 DEGs (**Table S3**) and 1228 target genes (**Table S4**).

KEGG Pathway and GO Function Annotation Details of Consensus Genes

Enrichment analyses produced 7 KEGG pathways and 19 GO function annotations from the consensus genes using the screening criteria of $P < 0.05$ (**Figure 4**). Consensus genes enriched in each pathway or function annotations were showed in **Table S5**. KEGG pathways involving insulin secretion, cAMP signaling, adrenergic signaling in cardiomyocytes, leukocyte transendothelial migration, etc. were mainly enriched. Among them, the insulin secretion pathway had the smallest P-value, and this pathway is associated with five consensus genes (P = 1.05E-03). The cAMP signaling pathway had the largest number (count = 6) of involved consensus genes, and its P-value was 3.87E-03. BP GO annotations involved in neuron migration, positive regulation of the G2/M mitotic cell cycle transition, extracellular matrix disassembly, cell adhesion, etc. were mainly enriched. The annotation with the smallest P-value in the BP involved neuron migration (P = 2.07E-03), and it correlated with five consensus genes. The largest number of related consensus genes involved cell adhesion (count = 7), and its P-value was 3.03E-02. And GO annotations in CC involving the plasma membrane, actin cytoskeleton, postsynaptic density, etc. were mainly enriched. The annotation with the smallest P-value in the CC was an integral component of the plasma membrane (P = 6.92E-03), and it involved 15 consensus genes. The largest number of linked consensus genes involved the plasma membrane (count = 28), and its P-value was 4.33E-02. Furthermore, GO annotations in MF involving heparan sulfate 6-O-sulfotransferase activity were mainly enriched and had a P-value of 1.98E-02 and two associated consensus genes.

Identification of Key Modules Using a Cluster Analysis of Protein-Protein Interactions

The interaction network presents protein-protein interactions, the strength of the interactions within the protein modules (**Figure 5A**), and up/down regulation conditions of the

TABLE 3 | Consensus genes of four prognostic miRNAs discovered in the overlap between predicted target genes and DEGs related to cervical cancer lymph node metastasis.

miRNA	Consensus Genes				
miR-502	ATP1A2	DLGAP2	FILIP1L	C12orf54	TLR6
	DAPK1	DCLK1	DONSON	SLIT3	WDHD1
	FBN2	DCUN1D5	CHST11	SLC7A14	LAMA3
	BCL7A	CDH5	ZNF471	CNTN2	PLXNA4
	COL10A1	DOK6	GLP1R	PGM5	SYNPO2
	EML6	MLLT6	ACSS3	HS6ST2	NUDT10
	RCC2	PHOX2B	SAMD4B		
miR-142	GDNF	ZFPM2	TRIM36	NBEA	IGF1
	FOXO4	FAM199X	HDLBP	TMTC1	GK
	BNC2	REV3L	SLC2A13	PDLIM5	SP2
	TNRC18	SACS	ST6GALNAC3	EGLN3	RTN1
miR-145	MMP16	ITGBL1	SEMA6A	ACTB	HS6ST1
	REV3L	ZFP14	SLC38A11	TNFRSF11B	PAQR9
	ST6GALNAC3	TUFT1	ARHGAP6	CREB3L2	GATC
	HTRA1	KLHL3	TPM3	KCNA6	GXYLT1
	EBF1	HTR1F	RAD51B	ZRANB3	SLITRK4
	RAPGEF4	DGKB			
miR-33b	LRP8	PRAMEF17	GPR173	SEMA7A	MMP16
	ATP1A2	TENM3	CACNA1C	DSC3	SECISBP2L
	HMGB3	PTPRB	ABHD2	GDNF	PTPRC
	CMTR2	CXCL12	PRICKLE2	ARMC8	

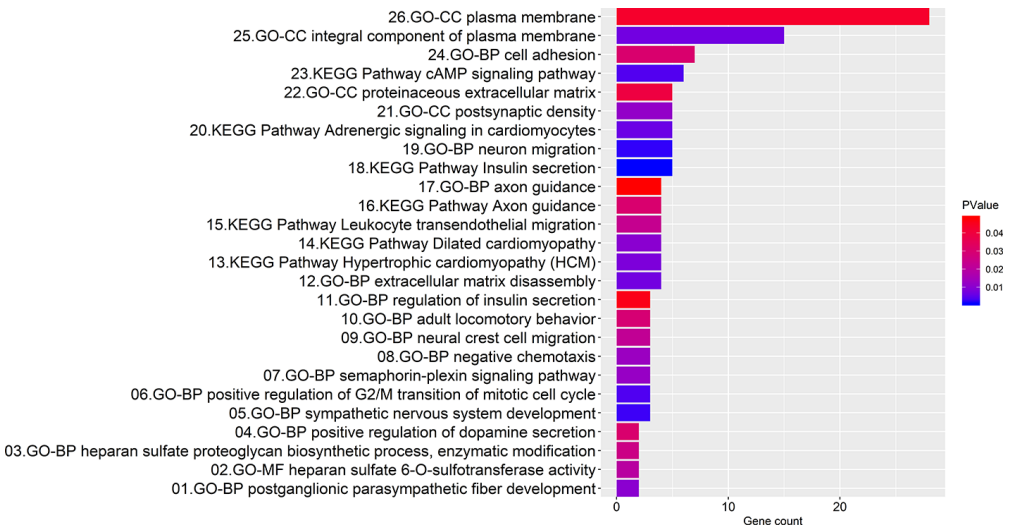


FIGURE 4 | Bar graph illustrating the enrichment analysis. The abscissa represents the number of consensus genes involved in KEGG pathways or GO function annotations. The ordinate represents items of the primary KEGG pathways or GO function annotations.

involved genes (**Figure 5B**). It consists of 52 nodes and 59 edges. Two important modules with MCODE scores greater than 2.0 were screened by MCODE and included seven key genes: *CXCL12*, *IGF1*, *PTPRC*, and *CDH5* (**Figure 5C**); *RAD51B*, *REV3L*, and *WDHD1* (**Figure 5D**). Among them, *CXCL12*, *IGF1*, *CDH5*, *RAD51B*, and *REV3L* were significantly down-regulated while *WDHD1* and *PTPRC* were notably up-regulated. It is worth to note that we also performed topology parameters analysis of PPI and obtained the same seven key genes above (**Figure S2**). The readers who are interested in this process can read the supplementary TopologyParametersAnalysis.doc.

Visualization Network of MiRNAs-Genes-Pathways and Annotations

The visualization network demonstrated that seven key genes are regulated by four miRNAs and are involved in seven key KEGG pathways and GO functional annotations (**Figure 6**). *WDHD1* and *CDH5* are regulated by miR-502; *RAD51B* and *REV3L* are

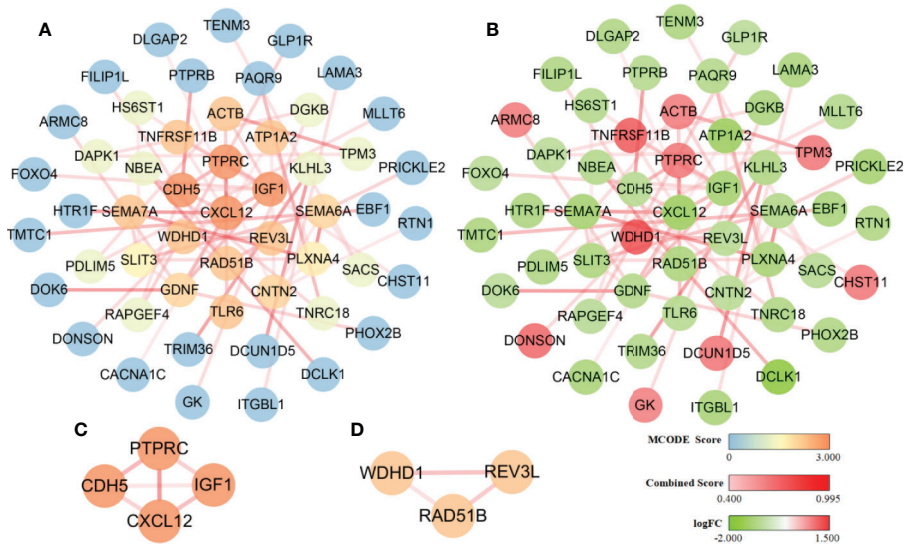


FIGURE 5 | Diagram of the protein-protein interaction network. The dark and light shading of the lines between solid circles indicates high and low interaction relationships, respectively, which is represented by the Combined Score. **(A)** The color of solid circles representing protein targets is distinguished by the MCODE score value. The blue circles are protein targets with an MCODE score value of 0, and these proteins did not participate in the construction of the key module. The dark and light orange circles are protein targets with a high MCODE score value of 3.000 and 2.000, respectively, and these proteins participated in the construction of the key module. **(B)** The color of solid circles representing protein targets is distinguished by the logFC value. The green circles are protein targets with a low logFC value and represent significantly down-regulated genes. The red circles are protein targets with a high logFC value and represent significantly up-regulated genes. **(C)** A module was constructed using the four protein targets with MCODE score values of 3.000. **(D)** A module was constructed using the three protein targets with MCODE score values of 2.000.

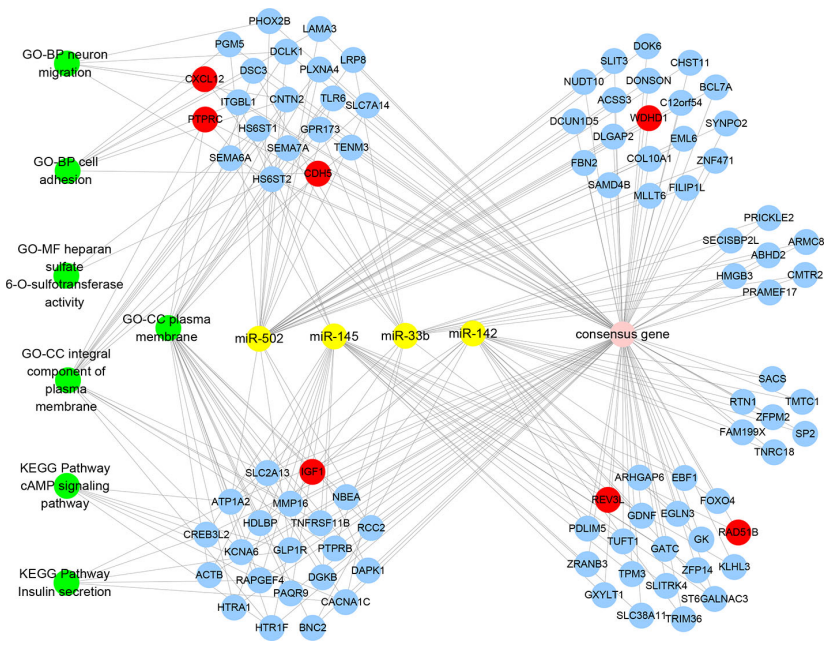


FIGURE 6 | MiRNAs-Genes-Pathways and Annotations visualization network denoting the relationships between miRNAs, consensus genes, key genes, key KEGG pathways, and GO functional annotations. Yellow solid circles represent miRNAs, blue solid circles represent consensus genes, red solid circles represent key genes, and green solid circles represent KEGG pathways and GO functional annotations.

regulated by miR-145; *REV3L* and *IGF1* are regulated by miR-142; and *CXCL12* and *PTPRC* are regulated by miR-33b. The enrichment analysis of the consensus genes showed that there was no key gene enrichment in the key KEGG pathways, while *CXCL12*, *IGF1*, *CDH5*, and *PTPRC* were involved in the key GO functional annotations. Specifically, *CXCL12* was enriched in neuron migration and cell adhesion of BP. *CDH5* was enriched in cell adhesion of BP. *IGF1*, together with *CDH5* and *PTPRC*, were enriched in the plasma membrane of CC. *PTPRC* was enriched in the integral component of the plasma membrane of CC.

DISCUSSION

Cervical carcinoma is one of the most common malignancies in females. Approximately 90% of cervical cancers occur in low- and middle-income countries. Lymph node metastasis and recurrence are the main manifestations in cervical cancer patients with a poor prognosis (Cohen et al., 2019). miRNAs collectively regulate thousands of human cancer-related, protein-coding genes and regulate many important biological processes that facilitate cancer development (i.e. cell growth, invasion, and apoptosis) (Lu et al., 2005). Therefore, miRNAs have become a hotspot of tumor research in the past decade. To identify novel and reliable prognostic biomarkers and important regulatory genes related to cervical cancer and lymph node metastasis, this study first identified 110 DEMs and 1840 DEGs, separately, from expression profiling and clinical data uploaded into TCGA and GEO databases, respectively. Next, a cancer prognosis model based on four prognostic miRNAs (miR-502, miR-145, miR-142, and miR-33b) was established using univariate Cox, multivariate Cox, and survival analyses. Subsequently, 1,228 target genes were predicted using TargetScan and miRDB, and 94 consensus genes were obtained from the overlap of DEGs and predicted target genes. Finally, seven key genes related to lymph node metastasis (i.e. *CXCL12*, *IGF1*, *PTPRC*, *CDH5*, *RAD51B*, *REV3L*, and *WDHD1*) were identified using STRING and Cytoscape.

According to the prognostic model, the expression levels of miR-502 and miR-145 were down-regulated, and the expression levels of miR-142 and miR-33b were up-regulated in cervical cancer. Among these four miRNAs, miR-145 and miR-142 have been reported that they were related to cervical cancer in previous experimental studies.

Previous studies have shown that the down-regulation of miR-145 is closely related to cervical cancer and its lymph node metastasis. Azizmohammadi et al. (2017) used qRT-PCR and a multivariate Cox analysis to demonstrate that the expression of miR-145 is reduced in cervical cancer tissues. They also showed that reduced expression of miR-145 is related to lymph node metastasis (P = 0.02), advanced Federation International of Gynecology and Obstetrics (FIGO) stage (P = 0.007), and vascular invasion (P = 0.026), which confirms miR-145's potential as a prognostic biomarker for the early detection of cervical cancer (Azizmohammadi et al., 2017). Ma et al. (2019) demonstrated that miR-145 is also down-regulated in cervical tumor cells, and up-regulation of miR-145 reduces cell proliferation by directly suppressing its target gene, FSCN1. Shi et al. (2012) also showed miR-145 is decreased in cervical cancer cells, and increasing miR-145 expression enhances chemo-sensitivity and inhibits invasion and migration by enhancing p53. The above experimental results are consistent with our prediction that the expression of miR-145 is suppressed in cervical cancer cells.

While our prognostic model predicted that miR-142 expression is increased, existing experimental studies illustrated that miR-142 expression is decreased in cervical cancer cells. Li et al. (2017) revealed that when compared with normal tissue, miR-142 expression is lower in cervical cancer cells and correlates with a poor prognosis. Deng et al. (2015) demonstrated that miR-142-3p is down-regulated in cervical tumors. The overregulation of miR-142-3p inhibits the expression of its target gene, *FZD7*, and further halts the proliferation and invasion of cervical cancer (Deng et al., 2015). Xia et al. (2018) reported that Metformin, an anti-cancer drug, up-regulates miR-142-3p expression in cervical cancer cells. They also showed that Metformin inhibits the invasion and migration of tumor cells by decreasing the sponge effect of MALAT1, up-regulating the expression of miR-142, and down-regulating the expression of the target gene, *HMGA2* (Xia et al., 2018). These three studies showed that miR-142 is a tumor suppressor gene. Thus, the decreased expression of miR-142 in cervical cancer cells is contrary to our results from the data mining calculation that shows an increase in its expression. We do not know the reason(s) for this contradiction, but it should be clarified in the future.

Although no experiment has demonstrated a correlation between miR-502 and miR-33b in cervical cancer, studies have shown that the down-regulation of miR-502 and up-regulation of miR-33b are involved in other types of cancer. Sun et al. (2014) demonstrated that miR-502, which is down-regulated in breast cancer cells, suppresses early apoptosis by targeting TRAF2 and restraining the NF-κB signaling pathway. Furthermore, Li et al. (2009) demonstrated that the activation of NF-κB results in a lower tumor grade, larger tumor volume, higher invasiveness, and increased metastasis in cervical cancer tissues. Therefore, miR-502 may affect cervical cancer lymph node metastasis by participating in the NF-κB signaling pathway. Zhang et al. (2019) showed that the up-regulation of miR-33b inhibits the Wnt/β-catenin signaling pathway by decreasing ZEB1 expression and promoting endometriosis. Although endometriosis is a common and benign disease, it has similar characteristics to malignancies including cell proliferation, invasion, metastasis, and recurrence. Also, Ramachandran et al. (2012) showed that abnormal activation of the Wnt/β-catenin pathway is common in cervical tumors, which may enhance proliferation and prevent apoptosis of cervical cancer cells. Therefore, up-regulation of miR-33b may affect the proliferation and apoptosis of cervical cancer cells by inhibiting the Wnt/β-catenin signaling pathway. To summarize, miR-502 and miR-33b might be involved in cervical cancer formation through distinctive ways as explained above, but such speculation needs to be experimentally validated.

To date, one computational paper exploring the relationship between cervical cancer and the expression of miRNA through data mining has been published. Liang et al. (2017) constructed a three-miRNA signature containing miR-145, miR-200c, and miR-218-1 by processing miRNA data from TCGA database. Their study demonstrated that the expression of miR-145 is significantly decreased in cervical tumor tissues. Furthermore, they found that the expression levels of miR-142 and miR-33b are up-regulated (Liang et al., 2017). This study provided results consistent with ours regarding the expression levels of miR-145, miR-142, and miR-33b. It is worth noting that although the expression of miR-142 was experimentally shown to be down-regulated in cervical cancer cells, as we mentioned previously, this earlier computational study circumstantially confirms the results of our research.

According to our predictions, among the seven key genes involved in early-stage cervical cancer lymph node metastasis, *CXCL12*, *IGF1*, *RAD51B*, *REV3L*, and *CDH5* are down-regulated, and *WDHD1* and *PTPRC* are up-regulated. **Table 4** compares the expression of these seven key genes in cervical cancer, other cancers, and lymph node metastasis as reported in previous experimental studies and our calculated results. The decreased expression of *CXCL12* in cervical cancer cells and its role in lymph node metastasis was confirmed in previous experiments (Yadav et al., 2016), and the decreased expression of *IGF1* and the increased expression of *WDHD1* were also experimentally validated in cervical cancer cells but only in lymph node metastasis of other cancers (Serrano et al., 2006; Kümmel et al., 2007; Huang et al., 2008; Zhou et al., 2016; Liu et al., 2019). Interestingly, the expression of *RAD51B* is also decreased in cervical cancer cells, but experiments showed that its expression is increased in other cancer lymph node metastases (Cheng et al., 2016; Hang et al., 2016). In addition, *REV3L* expression is increased in cervical cancer cells and lymph node metastases of other cancers in the laboratory, which is inconsistent with our prediction results (Yang et al., 2015; Zhu et al., 2016). Finally, *PTPRC* and *CDH5* have not been previously reported to be associated with cervical cancer or its lymph node metastasis. While experiments demonstrate that *PTPRC* expression is increased in other cancers and their lymph node metastases (Collette et al., 2007; Camacho et al., 2018), *CDH5* is up-regulated in other cancers but down-regulated in lymph node metastasis of colorectal cancer (Tacconi et al., 2015; Hung et al., 2016; Higuchi et al., 2017).

CXCL12 is the ligand for the G-protein coupled receptor-like chemokine (C-X-C motif) receptors 4 and 7. It affects many cellular processes such as immune monitoring, inflammatory response, tumor growth, and metastasis (Colamussi et al., 2001; Yadav et al., 2016). Yadav et al. (2016) demonstrated that *CXCL12* expression is absent in cervical cancer. They also illustrated that *CXCL12* silencing enables cells to evade apoptosis and leads to the progression and metastasis of cervical cancer (Yadav et al., 2016). Meanwhile, Müller et al. (2001) reported that *CXCR4* is usually highly expressed in tumor tissues, and the *CXCR4*/*CXCL12* axis could prevent breast tumor lymph-node metastasis and lung metastasis when *CXCR4* is neutralized. Thus, *CXCL12* can act as a tumor suppressor in lymph node metastasis of cervical and other cancers, and it might inhibit tumor progression through the *CXCR4*/*CXCL12* axis.

IGF1 is a cytokine that mediates cell growth and development (Macaulay, 1992). Serrano et al. (2006) reported low serum *IGF1* expression in cervical cancer cells, and reduced expression of *IGF1* is associated with an increased risk of cervical cancer. Huang et al. (2008) showed that lower levels of *IGF1* can effectively predict survival in patients with cervical cancer. Furthermore, they found that a combination of increased carcinoembryonic antigen (CEA) levels and decreased *IGF1* levels is significantly associated with an increased risk of death and could accurately predict patients with a poor prognosis. Though no experiment is available to confirm that *IGF1* is involved in cervical cancer lymph node metastasis, Kümmel et al. (2007) reported that in breast tumor patients, plasma *IGF1* expression is increased after dose-intensified chemotherapy, and they showed a more prominent increase in *IGF1* expression in patients with positive lymph nodes than other patients. *WDHD1*, also known as *AND1*, is involved in signal transduction, pre-mRNA processing, replication, transcription, cytoskeleton assembly, chromosome assembly, etc. (Park et al., 2012). Zhou et al. (2016) showed that *WDHD1* is up-regulated in primary human keratinocytes and spontaneously immortalized human foreskin keratinocytes cells expressing oncogene E7 in HPV-induced carcinogenesis. They also showed that *WDHD1* can increase E7-induced G1 checkpoint abolition and duplication. They further demonstrated that the polyploidy ratio of cells expressing E7 is significantly reduced after knocking down *WDHD1* with siRNA (Zhou et al., 2016). Although the role of *WDHD1* in cervical cancer lymph node metastasis has not been explained, Liu et al.

TABLE 4 | The expression of key genes reported from previous experimental studies.

	CXCL12	IGF1	WDHD1	RAD51B	REV3L	PTPRC	CDH5
Cancer	▼	▼	▲	▼	△	⧋	⧊ △
LNM	▼ ⧊	⧊	⧋	△	△	⧋	⧊

▲ *up-regulated in cervical cancer, agreed with our calculated result.*
▼ *down-regulated in cervical cancer, agreed with our calculated result.*
△ *up-regulated in cervical cancer, disagreed with our calculated result.*
⧋ *up-regulated in other cancers, agreed with our calculated result in cervical cancer.*
⧊ *down-regulated in other cancers, agreed with our calculated result in cervical cancer.*
△ *up-regulated in other cancers, disagreed with our calculated result in cervical cancer.*

(Liu et al., 2019) found that miR-494 can inhibit EMT and lymph node metastasis of cholangiocarcinoma (CCA) cells by targeting overexpressed *WDHD1*. In summary, *IGF1* and *WDHD1* are likely connected with cervical cancer lymph node metastasis.

RAD51B is one of the *RAD51* gene families involved in homologous recombination-mediated DNA repair (Thacker, 2005; Nagathihalli and Nagaraju, 2011). Hang et al. (2016) demonstrated that *RAD51B* is a tumor suppressor gene in cervical cancer cells, which is consistent with our prediction. They further showed that the miRNA binding sites of *RAD51B* genetic variants in cervical cancer cells may increase tumor susceptibility, and *RAD51B* is vital in gauging the cervical cancer risk of individuals and improving the effectiveness of preventive intervention (Hang et al., 2016). Currently, there are no reports that demonstrate *RAD51B* affects cervical cancer lymph node metastasis, but Cheng et al. (2016) showed that the overexpression of *RAD51B* in gastric cancer cells is significantly associated with lymph node metastasis (P = 0.001), advanced stage (P = 0.009), and invasive differentiation (P = 0.022), and it may act as a potential signature for early detection and poor prognosis. *RAD51B* expression in gastric cancer with lymph node metastasis is contrary to our prediction: it tended to represent the particularity of *RAD51B* expression in cervical cancer lymph node metastasis. The specific reasons for this contradiction need to be further clarified.

REV3L encodes the protein representing the catalytic sub-unit of Polζ, and inhibiting *REV3L* expression enables cancer cells to tolerate DNA damage and stunted growth (Knobel et al., 2011). Yang et al. (2015) showed that overexpression of *REV3L* promotes proliferation and colony formation and inhibits cervical cancer cell sensitivity to cisplatin. Thus, *REV3L* could be a potential therapeutic target for cervical cancer treatment. Zhu et al. (2016) also showed that *REV3L* is significantly up-regulated in esophageal squamous cell carcinoma tissues, and it positively correlates with lymph node metastasis (P <0.05) and clinical stage (P < 0.05). Additionally, overexpression of *REV3L* increases the expression levels of cyclin D1 and survivin, which work together to promote the growth and invasion of esophageal cancer cells (Zhu et al., 2016). These results demonstrate that *REV3L* is closely linked to cervical cancer lymph node metastasis. However, these *REV3L* experimental results are inconsistent with the inferences obtained from our data mining analysis. The specific reasons behind this contradiction need to be clarified.

PTPRC, also known as *CD45*, is a key regulator of cell growth, differentiation, mitosis, and carcinogenic transformation (Rheinländer et al., 2018). Camacho et al. (2018) showed that *PTPRC* is significantly overexpressed in head and neck squamous cell cancer cells, and tumor samples overexpressing *PTPRC* have significantly higher tumor-infiltrating lymphocytes (TIL) scores than tumor samples expressing low levels of *PTPRC*, leading to a poorer prognosis. Collette et al. (2007) noted that the expression of *PTPRC* plays a critical role in determining the signal transduction and proliferation response of human myeloma cells to growth factors such as IL-6 and IGF1. IL-6 and IGF1 separately induced CD45+ and CD45- myeloma cell colony formation through the MAPK/ERK signaling pathway in which CD45 is critical for myeloma proliferation (Collette et al., 2007). In addition, many experimental studies reported that the MAPK/ERK signaling pathway is crucial in cervical cancer formation. For example, Li et al. (2018) demonstrated that MAPK/ERK signaling pathway activation promotes cervical cancer cell proliferation. Tao et al. (2018) illustrated that miR-497 acts as a tumor suppressor by blocking the MAPK/ERK signaling pathway in cervical cancer cells. Furthermore, Wang et al. (2020) showed that the anti-cancer drug sclareol targets the MAPK/ERK signaling pathway and induces cervical cancer cell apoptosis and cell cycle arrest. Therefore, considering these observations, we can speculate that *PTPRC* and *IGF1* may negatively regulate the occurrence and development of cervical cancer and its lymph node metastasis by affecting the MAPK/ERK signaling pathway.

CDH5, also known as VE-cadherin, represses endothelial cell apoptosis and participates in endothelial cell growth contact inhibition (Cavallaro et al., 2006). Abnormal *CDH5* was found in cancer cells. Hung et al. (2016) and Higuchi et al. (2017) demonstrated that *CDH5* is over-expressed in lung cancer and gastric cancer, respectively. However, Tacconi et al. (2015) demonstrated that *CDH5* is down-regulated in colorectal cancer, and is negatively associated with its lymph node metastasis. They found that up-regulated *VEGFC*/*VEGFR3* reduces *CDH5* expression, enhances permeability, and increases trans-endothelial migration. Thus, it promotes lymphatic vessel density and colorectal cancer lymph node metastasis (Tacconi et al., 2015). Furthermore, Chaudary et al. (2011) demonstrated that suppressing the over-expression of *VEGFC*/*VEGFR3* in cervical cancer cells inhibits proliferation and cervical cancer lymph node metastasis. We suspect that the over-expression of *VEGFC*/*VEGFR3* restrains *CDH5* in cervical cancer cells and further promotes lymph node metastasis. But this assumption needs to be further studied.

Table 3 shows that eight targeting relationships are available between four prognosis miRNAs and seven key genes. Specifically, miR-502 targets *WDHD1* and *CDH5*, miR-145 targets *REV3L* and *RAD51B*, miR-142 targets *REV3L* and *IGF1*, and miR-33b targets *CXCL12* and *PTPRC*. Among these relationships, miR-145-REV3L and miR-142-IGF1 have been reported previously. Chen et al. (2019) demonstrated that miR-145 can directly regulate the expression of *REV3L* in esophageal squamous cell carcinoma (ESCC) cells using a dual luciferase reporter assay and Western blot analysis. Also, low levels of miR-145 augmented *REV3L* mRNA and protein expression in ESCC. Xiong et al. (2018) showed that eight out of 12 target gene prediction programs including TargetScan and miRDB predicted the miR-142-IGF1 relationship. Subsequently, they suggested that decitabine exerts its therapeutic effect on hepatocellular carcinoma by inhibiting miR-142 DNA methylation, which enhances miRNA expression and further down-regulates the target genes of miR-142-5p such as *IGF1* (Xiong et al., 2018). To summarize, the above studies demonstrated target/regulatory relationships between miR-145-REV3L and miR-142-IGF1. The remaining six regulatory relationships have yet to be reported, which might lay the foundation for future research on cervical and other cancers with lymph node metastasis.

In the key KEGG pathways and GO functional annotations of key gene enrichment, it had been reported that cell adhesion and integral component of the plasma membrane were closely related to tumor proliferation and metastasis. For example, Mierke (2008) showed that the integrity of the endothelial barrier is preserved by a complex balance of cell adhesion factors. Once the integrity is compromised, tumor cells can metastasize through blood vessels or lymphatic channels (Mierke, 2008). More importantly, Zheng et al. (2017) suggested that the *CXCR4*/*CXCL12* axis plays a role in reducing the adhesion ability of colon cancer cells by regulating the Akt and IGF1R signaling pathways. Therefore, the key gene *CXCL12*, enriched in cell adhesion, may lead to cervical cancer lymph node metastasis by affecting cell-cell adhesion. In addition, protein components of the cell membrane are involved in cell signal transduction, cell interaction, and other important steps in the process of cancer cell metastasis (Lund et al., 2009). Most of the key molecules involved in the signaling pathway are distributed in the cell membrane. *PTPRC*, as a transmembrane tyrosine phosphatase, is expressed in all leukocytes, and it is involved in lymphocyte immunity against tumor cells (Tchilian et al., 2001). Therefore, we speculate that *PTPRC*, which is enriched in an integral component of the plasma membrane, may be an important signal molecule in cervical cancer metastasis. Moreover, it may be involved in a new approach to studying the mechanism of cervical cancer lymph node metastasis.

CONCLUSION

By data-mining differentially expressed miRNAs and genes along with other clinical information in a multi-step analysis, we obtained a prognostic model of cervical cancer with lymph node metastasis containing four miRNAs and seven genes. We showed that four miRNAs (miR-502, miR-145, miR-142, and miR-33b) are independent and common prognostic biomarkers for patients with cervical cancer, and seven proteins (*CXCL12*, *IGF1*, *PTPRC CDH5*, *RAD51B*, *REV3L*, and *WDHD1*) are key genes significantly related to lymph node metastasis. Among them, miR-145, miR-142, *CXCL12*, *IGF1*, and *WDHD1* have been confirmed, while miR-502, miR-33b, *PTPRC*, and *CDH5* are reported, here, for the first time. Also, the expression levels and/or roles of miR-142, miR-33b, *RAD51B*, *REV3L*, and *CDH5* in cervical cancer lymph node metastasis need further clarification. In summary, our study may improve our understanding of the progression and lymph node metastasis mechanism of cervical cancer, and, therefore, provide a novel diagnostic indicator and therapeutic targets for future clinical treatments.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://portal.gdc.cancer.gov/, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7410.

AUTHOR CONTRIBUTIONS

Conceptualization: SC, CG, and ZH. Methodology: SC, CG, and YW. Software: SC, CG, and YW. Validation: SC, YW, and ZH. Formal analysis: SC, CG, and ZH. Investigation: SC and YW. Data curation: SC and CG. Writing—original draft preparation: SC and CG. Writing—review and editing: ZH. Visualization: SC and CG. Supervision: ZH. Project administration: ZH. Funding acquisition: ZH. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31770774), the Provincial Major Project of Basic or Applied Research in Natural Science, Guangdong Provincial, Education Department (2016KZDXM038) and the higher education reform project of Guangdong Province (2019268).

ACKNOWLEDGMENTS

We thank Wordvice for their help in revising the English grammar.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.00544/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005. doi: 10.7554/eLife.05005

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Azizmohammadi, S., Safari, A., Azizmohammadi, S., Kaghazian, M., Sadrkhanlo, M., Yahaghi, E., et al. (2017). Molecular identification of miR-145 and miR-9 expression level as prognostic biomarkers for early-stage cervical cancer detection. *J. Assoc. Physicians* 110 (1), 11–15. doi: 10.1093/qjmed/hcw101

Bader, G. D., and Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* 4, 2. doi: 10.1186/1471-2105-4-2

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2007). NCBI GEO: mining tens of millions of expression profiles–database and tools update. *Nucleic Acids Res.* 35 (Database issue), D760–D765. doi: 10.1093/nar/gkl887

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116 (2), 281–297. doi: 10.1016/s0092-8674(04)00045-5

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* 68 (6), 394–424. doi: 10.3322/caac.21492

Camacho, M., Agüero, A., Sumarroca, A., López, L., Pavón, M.-Á., Avilés-Jurado, F.-X., et al. (2018). Prognostic value of CD45 transcriptional expression in head and neck cancer. *Eur. Arch. Oto-Rhino-Laryngol* 275 (1), 225–232. doi: 10.1007/s00405-017-4806-2

Cavallaro, U., Liebner, S., and Dejana, E. (2006). Endothelial cadherins and tumor angiogenesis. *Exp. Cell Res.* 312 (5), 659–667. doi: 10.1016/j.yexcr.2005.09.019

Chaudary, N., Milosevic, M., and Hill, R. P. (2011). Suppression of vascular endothelial growth factor receptor 3 (VEGFR3) and vascular endothelial growth factor C (VEGFC) inhibits hypoxia-induced lymph node metastases in cervix cancer. *Gynecol. Oncol.* 123 (2), 393–400. doi: 10.1016/j.ygyno.2011.07.006

Chen, L., and Kang, C. (2015). miRNA interventions serve as 'magic bullets' in the reversal of glioblastoma hallmarks. *Oncotarget* 6 (36), 38628–38642. doi: 10.18632/oncotarget.5926

Chen, Q., Hou, J., Wu, Z., Zhao, J., and Ma, D. (2019). miR-145 Regulates the sensitivity of esophageal squamous cell carcinoma cells to 5-FU via targeting REV3L. *Pathol. Res. Pract.* 215 (7), 152427. doi: 10.1016/j.prp.2019.04.019

Cheng, Y., Yang, B., Xi, Y., and Chen, X. (2016). RAD51B as a potential biomarker for early detection and poor prognostic evaluation contributes to tumorigenesis of gastric cancer. *J. Int. Soc. Oncodevelopmental Biol. Med.* 37 (11), 14969–14978. doi: 10.1007/s13277-016-5340-3

Cohen, P. A., Jhingran, A., Oaknin, A., and Denny, L. (2019). Cervical cancer. *Lancet (London England)* 393 (10167), 169–182. doi: 10.1016/S0140-6736(18)32470-X

Colamussi, M. L., Secchiero, P., Gonelli, A., Marchisio, M., Zauli, G., and Capitani, S. (2001). Stromal derived factor-1 alpha (SDF-1 alpha) induces CD4+ T cell apoptosis via the functional up-regulation of the Fas (CD95)/Fas ligand (CD95L) pathway. *J. Leukocyte Biol.* 69 (2), 263–270.

Collette, M., Descamps, G., Pellat-Deceunynck, C., Bataille, R., and Amiot, M. (2007). Crucial role of phosphatase CD45 in determining signaling and proliferation of human myeloma cells. *Eur. Cytokine Network* 18 (3), 120–126. doi: 10.1684/ecn.2007.0095

Dai, Y., Tong, R., Guo, H., Yu, T., and Wang, C. (2017). Association of CXCR4, CCR7, VEGF-C and VEGF-D expression with lymph node metastasis in patients with cervical cancer. *Eur. J. Obstetrics Gynecol. Reprod. Biol.* 214, 178–183. doi: 10.1016/j.ejogrb.2017.04.043

Deng, B., Zhang, Y., Zhang, S., Wen, F., Miao, Y., and Guo, K. (2015). MicroRNA-142-3p inhibits cell proliferation and invasion of cervical cancer cells by targeting FZD7. *J. Int. Soc. Oncodevelopmental Biol. Med.* 36 (10), 8065–8073. doi: 10.1007/s13277-015-3483-2

Doorbar, J. (2006). Molecular biology of human papillomavirus infection and cervical cancer. *Clin. Sci. (London Engl: 1979)* 110 (5), 525–541. doi: 10.1042/CS20050369

Gómez-Gómez, Y., Organista-Nava, J., and Gariglio, P. (2013). Deregulation of the miRNAs expression in cervical cancer: human papillomavirus implications. *BioMed. Res. Int.* 2013, 407052. doi: 10.1155/2013/407052

George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *J. Nuclear Cardiol.* 21 (4), 686–694. doi: 10.1007/s12350-014-9908-2

Hang, D., Zhou, W., Jia, M., Wang, L., Zhou, J., Yin, Y., et al. (2016). Genetic variants within microRNA-binding site of RAD51B are associated with risk of cervical cancer in Chinese women. *Cancer Med.* 5 (9), 2596–2601. doi: 10.1002/cam4.797

Higuchi, K., Inokuchi, M., Takagi, Y., Ishikawa, T., Otsuki, S., Uetake, H., et al. (2017). Cadherin 5 expression correlates with poor survival in human gastric cancer. *J. Clin. Pathol.* 70 (3), 217–221. doi: 10.1136/jclinpath-2016-203640

Huang, B.-X., and Fang, F. (2018). Progress in the Study of Lymph Node Metastasis in Early-stage Cervical Cancer. *Curr. Med. Sci.* 38 (4), 567–574. doi: 10.1007/s11596-018-1915-0

Huang, Y. F., Shen, M. R., Hsu, K. F., Cheng, Y. M., and Chou, C. Y. (2008). Clinical implications of insulin-like growth factor 1 system in early-stage cervical cancer. *Br. J. Cancer* 99 (7), 1096–1102. doi: 10.1038/sj.bjc.6604661

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi: 10.1038/nprot.2008.211

Huang, Z., Zhu, D., Wu, L., He, M., Zhou, X., Zhang, L., et al. (2017). Six Serum-Based miRNAs as Potential Diagnostic Biomarkers for Gastric Cancer. *Cancer Epidemiol. Biomarkers Prev.* 26 (2), 188–196. doi: 10.1158/1055-9965.EPI-16-0607

Hung, M.-S., Chen, I. C., Lung, J.-H., Lin, P.-Y., Li, Y.-C., and Tsai, Y.-H. (2016). Epidermal Growth Factor Receptor Mutation Enhances Expression of Cadherin-5 in Lung Cancer Cells. *PloS One* 11 (6), e0158395. doi: 10.1371/journal.pone.0158395

Kümmel, S., Eggemann, H., Lüftner, D., Gebauer, N., Bühler, H., Schaller, G., et al. (2007). Significant changes in circulating plasma levels of IGF1 and IGFBP3 after conventional or dose-intensified adjuvant treatment of breast cancer patients with one to three positive lymph nodes. *Int. J. Biol. Markers* 22 (3), 186–193.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Knobel, P. A., Kotov, I. N., Felley-Bosco, E., Stahel, R. A., and Marti, T. M. (2011). Inhibition of REV3 expression induces persistent DNA damage and growth arrest in cancer cells. *Neoplasia (N.Y.)* 13 (10), 961–970. doi: 10.1593/neo.11828

Lee, Y. S., and Dutta, A. (2009). MicroRNAs in cancer. *Annu. Rev. Pathol.* 4, 199–227. doi: 10.1146/annurev.pathol.4.110807.092222

Li, J., Jia, H., Xie, L., Wang, X., Wang, X., He, H., et al. (2009). Association of constitutive nuclear factor-kappaB activation with aggressive aspects and poor prognosis in cervical cancer. *Int. J. Gynecol. Cancer* 19 (8), 1421–1426. doi: 10.1111/IGC.0b013e3181b70445

Li, M., Li, B. Y., Xia, H., and Jiang, L. L. (2017). Expression of microRNA-142-3p in cervical cancer and its correlation with prognosis. *Eur. Rev. Med. Pharmacol. Sci.* 21 (10), 2346–2350.

Li, S., Ma, Y.-M., Zheng, P.-S., and Zhang, P. (2018). GDF15 promotes the proliferation of cervical cancer cells by phosphorylating AKT1 and Erk1/2 through the receptor ErbB2. *J. Exp. Clin. Cancer Res.* 37 (1), 80. doi: 10.1186/s13046-018-0744-0

Li, S., Cui, Z., Zhao, Y., Ma, S., Sun, Y., Li, H., et al. (2020). Candidate lncRNA-microRNA-mRNA networks in predicting non-small cell lung cancer and related prognosis analysis. *J. Cancer Res. Clin. Oncol.* 146 (4), 883–896. doi: 10.1007/s00432-020-03161-6

Liang, B., Li, Y., and Wang, T. (2017). A three miRNAs signature predicts survival in cervical cancer using bioinformatics analysis. *Sci. Rep.* 7 (1), 5624. doi: 10.1038/s41598-017-06032-2

Liu, B., Hu, Y., Qin, L., Peng, X.-B., and Huang, Y.-X. (2019). MicroRNA-494-dependent WDHDI inhibition suppresses epithelial-mesenchymal transition, tumor growth and metastasis in cholangiocarcinoma. *Digestive Liver Dis.* 51 (3), 397–411. doi: 10.1016/j.dld.2018.08.021

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435 (7043), 834–838. doi: 10.1038/nature03702

Lund, R., Leth-Larsen, R., Jensen, O. N., and Ditzel, H. J. (2009). Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. *J. Proteome Res.* 8 (6), 3078–3090. doi: 10.1021/pr801091k

Müller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410 (6824), 50–56. doi: 10.1038/35065016

Ma, L., and Li, L.-L. (2019). miR-145 Contributes to the Progression of Cervical Carcinoma by Directly Regulating FSCN1. *Cell Transplant.* 28 (9-10), 1299–1305. doi: 10.1177/0963689719861063

Macaulay, V. M. (1992). Insulin-like growth factors and cancer. *Br. J. Cancer* 65 (3), 311–320. doi: 10.1038/bjc.1992.65

Mierke, C. T. (2008). Role of the endothelium during tumor cell metastasis: is the endothelium a barrier or a promoter for cell invasion and metastasis? *J. Biophysics (Hindawi Publishing Corporation Online)* 2008, 183516. doi: 10.1155/2008/183516

Nagathihalli, N. S., and Nagaraju, G. (2011). RAD51 as a potential biomarker and therapeutic target for pancreatic cancer. *Biochim. Biophys. Acta* 1816 (2), 209–218. doi: 10.1016/j.bbcan.2011.07.004

Park, S.-Y., Im, J.-S., Park, S.-R., Kim, S.-E., Wang, H.-J., and Lee, J.-K. (2012). Mimosine arrests the cell cycle prior to the onset of DNA replication by preventing the binding of human Ctf4/And-1 to chromatin via Hif-1α

activation in HeLa cells. *Cell Cycle (Georgetown Tex.)* 11 (4), 761–766. doi: 10.4161/cc.11.4.19209

Park, S., Eom, K., Kim, J., Bang, H., Wang, H.-Y., Ahn, S., et al. (2017). MiR-9, miR-21, and miR-155 as potential biomarkers for HPV positive and negative cervical cancer. *BMC Cancer* 17 (1), 658. doi: 10.1186/s12885-017-3642-5

Ramachandran, I., Thavathiru, E., Ramalingam, S., Natarajan, G., Mills, W. K., Benbrook, D. M., et al. (2012). Wnt inhibitory factor 1 induces apoptosis and inhibits cervical cancer growth, invasion and angiogenesis in vivo. *Oncogene* 31 (22), 2725–2737. doi: 10.1038/onc.2011.455

Rheinländer, A., Schraven, B., and Bommhardt, U. (2018). CD45 in human physiology and clinical medicine. *Immunol. Lett.* 196, 22–32. doi: 10.1016/j.imlet.2018.01.009

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinf. (Oxford England)* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Sepulveda, J. L. (2020). Using R and Bioconductor in Clinical Genomics and Transcriptomics. *J. Mol. Diagn.* 22 (1), 3–20. doi: 10.1016/j.jmoldx.2019.08.006

Serrano, M. L., Romero, A., Cendales, R., Sánchez-Gómez, M., and Bravo, M. M. (2006). Serum levels of insulin-like growth factor-I and -II and insulin-like growth factor binding protein 3 in women with squamous intraepithelial lesions and cervical cancer. *Rev. Del Instituto Nacional Salud* 26 (2), 258–268.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shi, M., Du, L., Liu, D., Qian, L., Hu, M., Yu, M., et al. (2012). Glucocorticoid regulation of a novel HPV-E6-p53-miR-145 pathway modulates invasion and therapy resistance of cervical cancer cells. *J. Pathol.* 228 (2), 148–157. doi: 10.1002/path.3997

Sun, L.-L., Wang, J., Zhao, Z.-J., Liu, N., Wang, A.-L., Ren, H.-Y., et al. (2014). Suppressive role of miR-502-5p in breast cancer via downregulation of TRAF2. *Oncol. Rep.* 31 (5), 2085–2092. doi: 10.3892/or.2014.3105

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Tacconi, C., Correale, C., Gandelli, A., Spinelli, A., Dejana, E., D'Alessio, S., et al. (2015). Vascular endothelial growth factor C disrupts the endothelial lymphatic barrier to promote colorectal cancer invasion. *Gastroenterology* 148 (7), 1438–51.e8. doi: 10.1053/j.gastro.2015.03.005

Tao, L., Zhang, C.-Y., Guo, L., Li, X., Han, N.-N., Zhou, Q., et al. (2018). MicroRNA-497 accelerates apoptosis while inhibiting proliferation, migration, and invasion through negative regulation of the MAPK/ERK signaling pathway via RAF-1. *J. Cell. Physiol.* 233 (10), 6578–6588. doi: 10.1002/jcp.26272

Tchilian, E. Z., Wallace, D. L., Wells, R. S., Flower, D. R., Morgan, G., and Beverley, P. C. (2001). A deletion in the gene encoding the CD45 antigen in a patient with SCID. *J. Immunol. (Baltimore Md.: 1950)* 166 (2), 1308–1313. doi: 10.4049/jimmunol.166.2.1308

Thacker, J. (2005). The RAD51 gene family, genetic instability and cancer. *Cancer Lett.* 219 (2), 125–135. doi: 10.1016/j.canlet.2004.08.018

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology (Poznan, Poland)* 19, 1A, A68–A77. doi: 10.5114/wo.2014.47136

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *psychol. Methods* 17 (2), 228–243. doi: 10.1037/a0027127

Wang, L., Zhou, P., Gao, X.-M., Luo, M.-L., Meng, M.-J., He, Y., et al. (2020). Naturally Occurring Sclareol Diterpene Augments the Chemosensitivity of Human Hela Cervical Cancer Cells by Inducing Mitochondrial Mediated Programmed Cell Death, S-Phase Cell Cycle Arrest and Targeting Mitogen-Activated Protein Kinase (MAPK)/Extracellular-Signal-Regulated Kinase (ERK) Signaling Pathway. *Int. Med. J. Exp. Clin. Res.* 26, e920248. doi: 10.12659/MSM.920248

Wang, W., Chu, H. J., Shang, C. L., Gong, X., Liu, T. Y., Zhao, Y. H., et al. (2016). Long-Term Oncological Outcomes After Laparoscopic Versus Abdominal Radical Hysterectomy in Stage IA2 to IIA2 Cervical Cancer: A Matched Cohort Study. *Int. J. Gynecol. Cancer* 26 (7), 1264–1273. doi: 10.1097/IGC.0000000000000749

Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 43 (Database issue), D146–D152. doi: 10.1093/nar/gku1104

Xia, C., Liang, S., He, Z., Zhu, X., Chen, R., and Chen, J. (2018). Metformin, a first-line drug for type 2 diabetes mellitus, disrupts the MALAT1/miR-142-3p sponge to decrease invasion and migration in cervical cancer cells. *Eur. J. Pharmacol.* 830, 59–67. doi: 10.1016/j.ejphar.2018.04.027

Xiong, D.-D., Dang, Y.-W., Lin, P., Wen, D.-Y., He, R.-Q., Luo, D.-Z., et al. (2018). A circRNA-miRNA-mRNA network identification for exploring underlying pathogenesis and therapy strategy of hepatocellular carcinoma. *J. Trans. Med.* 16 (1), 220. doi: 10.1186/s12967-018-1593-5

Yadav, S. S., Prasad, S. B., Prasad, C. B., Pandey, L. K., Pradhan, S., Singh, S., et al. (2016). CXCL12 is a key regulator in tumor microenvironment of cervical cancer: an in vitro study. *Clin. Exp. Metastasis* 33 (5), 431–439. doi: 10.1007/s10585-016-9787-9

Yang, L., Shi, T., Liu, F., Ren, C., Wang, Z., Li, Y., et al. (2015). REV3L, a promising target in regulating the chemosensitivity of cervical cancer cells. *PloS One* 10 (3), e0120334. doi: 10.1371/journal.pone.0120334

Yang, G., Zhang, Y., and Yang, J. (2019). A Five-microRNA Signature as Prognostic Biomarker in Colorectal Cancer by Bioinformatics Analysis. *Front. Oncol.* 9, 1207. doi: 10.3389/fonc.2019.01207

Yao, S., Xu, J., Zhao, K., Song, P., Yan, Q., Fan, W., et al. (2018). Down-regulation of HPGD by miR-146b-3p promotes cervical cancer cell proliferation, migration and anchorage-independent growth through activation of STAT3 and AKT pathways. *Cell Death Dis.* 9 (11), 1055. doi: 10.1038/s41419-018-1059-y

Zeng, Y., Wang, K.-X., Xu, H., and Hong, Y. (2018). Integrative miRNA analysis identifies hsa-miR-3154, hsa-miR-7-3, and hsa-miR-600 as potential prognostic biomarker for cervical cancer. *J. Cell. Biochem.* 119 (2), 1558–1566. doi: 10.1002/jcb.26315

Zhang, B., Pan, X., Cobb, G. P., and Anderson, T. A. (2007). microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* 302 (1), 1–12. doi: 10.1016/j.ydbio.2006.08.028.

Zhang, Z.-Q., Han, Y.-Z., Nian, Q., Chen, G., Cui, S.-Q., and Wang, X.-Y. (2015). Tumor Invasiveness, Not Lymphangiogenesis, Is Correlated with Lymph Node Metastasis and Unfavorable Prognosis in Young Breast Cancer Patients (≤35 Years). *PloS One* 10 (12), e0144376. doi: 10.1371/journal.pone.0144376

Zhang, H., Li, G., Sheng, X., and Zhang, S. (2019). Upregulation of miR−33b promotes endometriosis via inhibition of Wnt/β−catenin signaling and ZEB1 expression. *Mol. Med. Rep.* 19 (3), 2144–2152. doi: 10.3892/mmr.2019.9870

Zheng, F., Zhang, Z., Flamini, V., Jiang, W. G., and Cui, Y. (2017). The Axis of CXCR4/SDF-1 Plays a Role in Colon Cancer Cell Adhesion Through Regulation of the AKT and IGF1R Signalling Pathways. *Anticancer Res.* 37 (8), 4361–4369. doi: 10.21873/anticanres.11830

Zhou, Y., Zhang, Q., Gao, G., Zhang, X., Liu, Y., Yuan, S., et al. (2016). Role of WDHD1 in Human Papillomavirus-Mediated Oncogenesis Identified by Transcriptional Profiling of E7-Expressing Cells. *J. Virol.* 90 (13), 6071–6084. doi: 10.1128/JVI.00513-16

Zhu, X., Zou, S., Zhou, J., Zhu, H., Zhang, S., Shang, Z., et al. (2016). REV3L, the catalytic subunit of DNA polymerase ζ, is involved in the progression and chemoresistance of esophageal squamous cell carcinoma. *Oncol. Rep.* 35 (3), 1664–1670. doi: 10.3892/or.2016.4549

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Chen, Gao, Wu and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice.

frontiers in Pharmacology

REVIEW
published: 20 May 2020
doi: 10.3389/fphar.2020.00733

Discovering Anti-Cancer Drugs *via* Computational Methods

Wenqiang Cui[1,2], *Adnane Aouidate*[1], *Shouguo Wang*[1], *Qiuliyang Yu*[1], *Yanhua Li*[2] *and Shuguang Yuan*[1]*

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2] College of Veterinary Medicine, Northeast Agricultural University, Harbin, China

New drug discovery has been acknowledged as a complicated, expensive, time-consuming, and challenging project. It has been estimated that around 12 years and 2.7 billion USD, on average, are demanded for a new drug discovery *via* traditional drug development pipeline. How to reduce the research cost and speed up the development process of new drug discovery has become a challenging, urgent question for the pharmaceutical industry. Computer-aided drug discovery (CADD) has emerged as a powerful, and promising technology for faster, cheaper, and more effective drug design. Recently, the rapid growth of computational tools for drug discovery, including anticancer therapies, has exhibited a significant and outstanding impact on anticancer drug design, and has also provided fruitful insights into the area of cancer therapy. In this work, we discussed the different subareas of the computer-aided drug discovery process with a focus on anticancer drugs.

Keywords: anti-cancer, CADD, drug discovery, AI, computational methods

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Zhiwei Feng,
University of Pittsburgh, United States
Shiliang Li,
East China University of Science and Technology, China

***Correspondence:**
Shuguang Yuan
shuguang.yuan@siat.ac.cn

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *12 March 2020*
Accepted: *01 May 2020*
Published: *20 May 2020*

Citation:
Cui W, Aouidate A, Wang S, Yu Q, Li Y and Yuan S (2020) Discovering Anti-Cancer Drugs via Computational Methods. Front. Pharmacol. 11:733. doi: 10.3389/fphar.2020.00733

INTRODUCTION

Up to now, cancer remains a global and serious public health challenge. It is estimated that there are more than 200 different types of cancer, generally named according to the tissue where the cancer was recognized for the first time. Cancer is considered to be one of the significant causes for death in the 21st century and the most critical obstacle for the increase of global life expectancy. According to an analysis by the world health organization (WHO) in 2015, cancer is the second leading cause of death for patients younger than 70 years old in 91 countries and the third or fourth leading cause of death among 22 other countries (Yan et al., 2019). Moreover, a global increase of 18.1 million new cancer cases and 9.6 million cancer-related deaths have been reported in a previous study (Bray et al., 2018), especially 70% of the death caused by cancer occur in low-income and middle-income countries. The fast growth of the cancer incidence and mortality has turned out to be global health challenges. How to reduce the cancer-related death rate has attracted significant attention from the government, society, medical industry, as well as scientific communities, expecting the rapid development of effective and safe drugs for cancer treatment.

Despite of the impressive progress in biotechnologies and further understandings of the disease biology, the development of new, practical and innovative small molecule drugs remains an arduous, time-consuming, and expensive project, which requires collaborations from many expertise in multidisciplinary fields, including medicinal chemistry, computational chemistry, biology, drug metabolism, clinical research, etc. Furthermore, it has been illustrated that the

successful discovery and development of a new drug costs 12 years, and expensive investment (Kapetanovic, 2008). Thus, novel drug development strategies with a reduced cost of time and money, as well as an enhanced efficiency are in high demand, which would contribute to a significant improvement in global health and life expectancy. Since the successful development of HIV protease inhibitor Viracept in the USA in 1997, which was the first drug design fully driven by its target structure (Kaldor et al., 1997), computational methods have served as an essential tool in drug discovery projects and have been a cornerstone for new drug development approaches. This makes the drug developmental process faster and cheaper. Recently, the fast growth in computational power, including massively parallel computing on graphical processing units (GPUs), the continuous advances in artificial intelligence (AI) tools (Chan et al., 2019; Yang et al., 2019), have translated fundamental research into practical applications (Zhavoronkov et al., 2019) in the drug discovery field. This attracted considerable attention for their outstanding performance on providing new promising perspectives and solutions to overcome life-threatening diseases.

In this review, we aim at providing an overview of different subjects of the computational-method-aided new drug discovery processes in general, and anti-cancer therapy discovery in particular. We reviewed some of the most representative examples and clarified fundamental principles by exploring studies on anticancer drug designs with the help of computational methods. A workflow of computational drug discovery is explained in **Figure 1**.

ANTI-CANCER DRUG TARGET PREDICTION

Human contains approximately 30,000 genes, among which around 6,000 to 8,000 sites are estimated as potential pharmacological targets. However, less than 400 encoded proteins have been proved to be effective for drug development until now (Drews, 2000; Chen et al., 2016). Cancer, compared to many other human diseases, now has a plethora of potential molecular targets for therapeutic development (Lazo and Sharlow, 2016). Traditional drug discovery mainly follows the paradigm of "one molecule - one target - one disease", without considering the interactions between drugs and proteins. However, an important fact that many complex diseases are relevant to a variety of target proteins (Hopkins, 2008; Yamanishi et al., 2008; Chen et al., 2012) has been overlooked. Furthermore, unexpected drug functions derived from off-targets are an accidental and uncontrollable activities because of the "poly-pharmacological" properties of certain drugs, which might result in undesirable side effects. Those are particularly pronounced for cancer drugs. On the other hand, there are some positive examples that benefit from the different pathways targeted by one given molecule. For example, sildenafil (viagra)

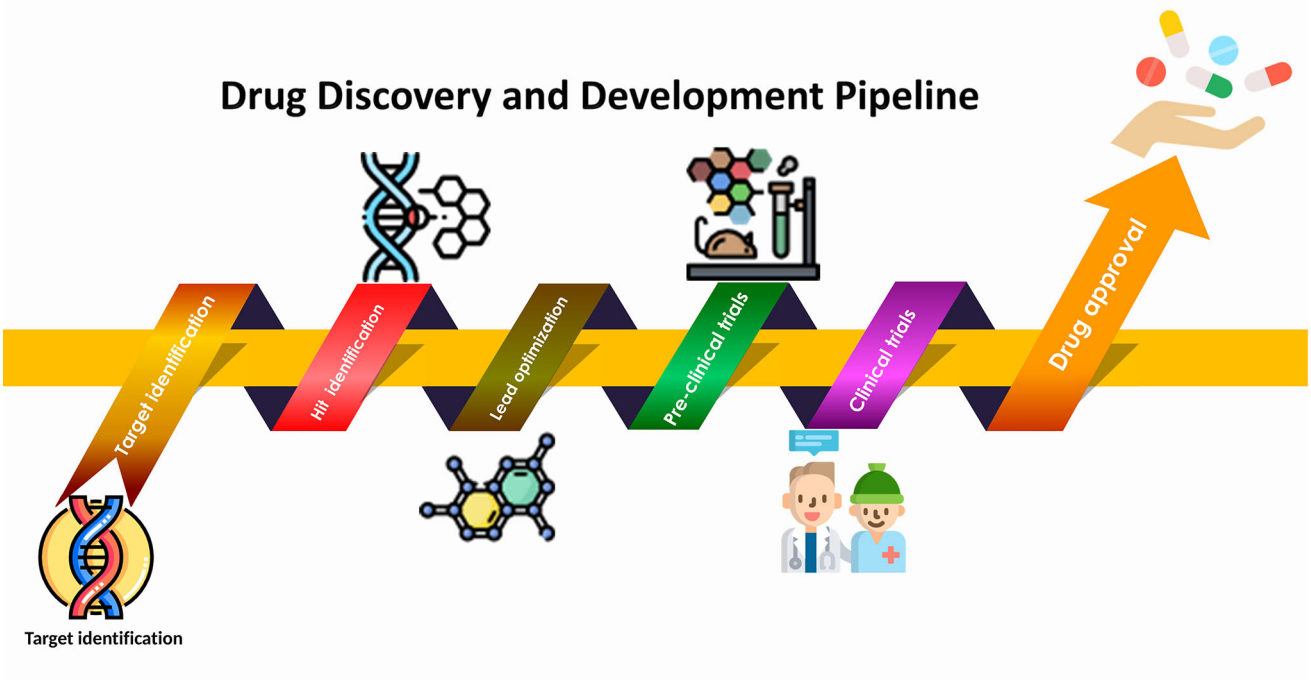


FIGURE 1 | A workflow for drug discovery: from target identification to drug approval.

was developed to treat angina, but now it is used for erectile dysfunction therapy (Ghofrani et al., 2006). There are several drugs, including anticancer drugs, whose corresponding target proteins (both primary and non-target) remain yet unidentified or unknown (Takarabe et al., 2012). Furthermore, some attractive and potentially effective cancer targets remain outside of the scope of pharmacological regulation. Some of these targets such as phosphatases, transcription factors, and RAS family members have been described as undruggable, as they lack effective enzymatic active sites (Lazo and Sharlow, 2016). To make the full use of known drugs to treat new indications, the characterization of all potential new ligand binding sites has been illustrated as a key point in drug repositioning and repurposing. Therefore, new and highly qualitative bioinformatic target prediction methods are required for the accurate prediction of drug targets.

Up to now, a wide range of drug target interactive web servers has been established, providing a series of drug-target databases and prediction tools (see **Tables 1** and **2**). Moreover, various computational approaches have been used to study potential interactions between proteins and drugs. In particular, network-based models and ML-based models have emerged as important tools. A review by Chen et al. summarizes several available computational models for this application (Chen et al., 2016). Interestingly, a method proposed by Campillos et al. that uses the similarity of drug side effects to determine whether multiple drugs could interact with the same target proteins attracted our attention (Campillos et al., 2008). Based on this research, Takarabe et al. took advantage of the US FDA's adverse event reporting system (AERS) to define the pharmacological similarity of all potential medicines and developed a novel system to predict large-scale interactions between unknown drug-targets (Takarabe et al., 2012). Notably, AERS was employed to predict interactions between drugs and targets for the first time. In 2010, Klipp et al. summarized several available computational models for network-based drug-target prediction (Klipp et al., 2010). Moreover, various biological data settings, including structures of bioactive compounds, sequences of target proteins, and information of ligand-target interactions, have been combined. A series of machine learning-based approaches have been demonstrated as efficient tools in detecting relationships among drug structures and corresponding target proteins from a large amount of data, such as supervised learning method (Srivastava et al., 2014), bipartite graph learning method (Li and Chen, 2013), bipartite local model (Yildirim et al., 2007), and so on. A recent review by Mayr et al. compared the predictive performance of deep learning with other prediction approaches for multiple drug targets in the comparative studies of composite target prediction methods. As a result, feed-forward neural networks were identified with better performance in drug target prediction than other methods (Mayr et al., 2018).

As above, since a large number of compounds and vigorous efforts are abandoned and wasted due to the off-target effects during the classical drug development procedure, a greatly enhanced development of target prediction in new drug exploration exhibited attractive advantages and further expansion in this area are still highly desirable

TABLE 1 | Drug-target database.

Databases	Websites
DrugBank	https://www.drugbank.ca/
TTD	http://bidd.nus.edu.sg/group/ttd/ttd.asp
MATADOR	http://matador.embl.de/
SuperTarget	http://insilico.charite.de/supertarget/
TDR targets	http://tdrtargets.org/
PDTD	http://www.dddc.ac.cn/pdtd/
ChEMBL	https://www.ebi.ac.uk/chembldb
STITCH	http://stitch.embl.de/
BindingDB	http://www.bindingdb.org/
CancerDR	http://crdd.osdd.net/raghava/cancerdr/
DCDB	http://www.cls.zju.edu.cn/dcdb/

TABLE 2 | Computational tools for target prediction.

Computational tools	Websites
SEA	https://omictools.com/sea-2-tool
Pharmmapper	http://www.lilab-ecust.cn/pharmmapper/
Chemmapper	https://omictools.com/chemmapper-tool
Tide	http://sysbio.molgen.mpg.de/tide
DINIES	http://www.genome.jp/tools/dinies/
SuperPred	http://prediction.charite.de/
SwissTarget Prediction	http://www.swisstargetprediction.ch/

STRUCTURE-BASED DRUG DISCOVERY

Structure-based strategy relies on the known structural information to define the interaction effect between bioactive compounds and the corresponding receptors. (Wang et al., 2000). With the development of biomolecular spectroscopic technologies such as X-ray crystallography and nuclear magnetic resonance (NMR), remarkable progress has been made in this field, leading to considerable improvements in our structural understanding of the drug target. Taking advantages of the three-dimensional structure of the proteins, new ligands could be rationally designed to trigger therapeutic effects. Hence, structure-based design (SBD) could provide critical insights into new drug design and development *via* discovering and optimizing the initial lead compounds (Prada-Gracia et al., 2016; Lu et al., 2018a). The high affinity ligand regulates validated drug targets selectively to influence specific cellular activities, ultimately achieving the desired pharmacological, and therapeutic effects (Urwyler, 2011). Capoten (captopril), the first ACE (angiotensin-converting enzyme) inhibitor, was one of the first successful examples of using structural information to optimize drug designs in the 1980s (Anthony et al., 2012). Since this study, structure-based

drug development started to serve as a novel and powerful algorithm and technique to promote faster, cheaper, and more effective drug development. In the past decade, extensive efforts have been made to promote the strategy of SBD, more and more successful applications played important roles in new medical research (Debnath et al., 2019; Hong et al., 2019; Mendoza et al., 2019; Itoh, 2020; Tondo et al., 2020).

Molecular Docking

Molecular docking is a typical structure-based protocol in rational drug design by studying and predicting the binding patterns and interaction affinities among the ligand and receptor biomolecules (Ferreira et al., 2015). It could be categorized as rigid docking and flexible docking according to the flexibility of the ligands involved in the computational process (Halperin et al., 2002; Dias and De Azevedo, 2008). The rigid docking method is a binding model which only considers the static geometrical, physical, and chemical complementarity between the ligand and the target proteins, while ignores the flexibility and the induced-fit theory (Salmaso and Moro, 2018). In general, the rigid docking, which is fast and highly effective, is applied to the high throughput virtual screening with a large number of small-molecule databases to be time-efficient. While the flexible docking method considers more detailed and accurate information. With the rapid improvement of computing resources and efficiency, flexible docking methods developed continuously and became more easily accessible. There are different types of software available for docking, such as Glide, FlexX, DOCK, AutoDock, Discovery Studio, Sybyl, etc.

The molecular docking process is mainly composed of three steps. First, the structures of small molecules and target proteins should be prepared in advance. In this step, abundant experimentally solved structures are available in the open access PDB database (http://www.rcsb.org), which can be used to understand many physiological processes based on the crystal structures, and also for homologous template models if docking structures are of interest. Second, it can act as an engine for predicting conformations, orientations, and positional spaces in the ligand binding site (Mathi et al., 2018). Conformational search algorithms carry out this task to predict the conformations of binary complexes by applying the methods of systematic and stochastic search. Systematic search techniques include: (i) Exhaustive search; (ii) Fragmentation; (iii) Conformational Ensemble. On the other hand, stochastic methods include: (i) Monte Carlo (MC) methods; (ii) Tabu search methods; (iii) Evolutionary Algorithms (EA); (IV) Swarm optimization (SO) methods (Ferreira et al., 2015). Finally, these programs evaluate the putative binding-free energy, which associates the scoring function to determine which compounds are more likely to bind to targets during the molecular docking (Huang et al., 2010). There are four essential types of scoring functions, including: (i) Consensus scoring functions (ii) Empirical scoring functions; (iii) Knowledge-based scoring functions; (iv) Force-field based scoring functions (Kortagere and Ekins, 2010). Furthermore, new scoring capabilities have been developed, for example (i) machine learning technologies; (ii) interactive fingerprints; (iii) quantum mechanical scores (Yuriev et al., 2015).

Structure-Based Pharmacophore Mapping

With the development in the past decades, the pharmacophore mapping method has been considered as one of the most useful technology during the process of drug discovery. All kinds of structure-based approaches have been conducted to improve pharmacophore modeling, which has been widely used for virtual screening, *de novo* design as well as lead optimization (Yang, 2010; Lu et al., 2018a). The structure-based pharmacophore (SBP) is another useful method. Based on the availability of ligand structures, SBP modeling methods can be cataloged into two types: target-ligand complex-based methods and target-binding site-based (without ligand) methods (Pirhadi et al., 2013). The approach based on the target-ligand complex can conveniently locate the ligand-binding pocket of the protein and assess the main ligand-protein interactions. This is exampled by LigandScout (Wolber et al., 2006), Pocket v.2 (Chen and Lai, 2006), and GBPM (Ortuso et al., 2006). It is worth noting that they cannot be used to the situations where ligands are unknown. The macromolecule (without ligand)-based method implemented in Discovery Studio (Lu et al., 2018b) is an obvious example which is not dependent on the ligands and the receptor-ligand interactions. The LUDI program (Bohm, 1992) defines the interactions within the binding site as pharmacological characteristics. Although this purely SBP method has the advantage of describing the entire interaction capability of a binding pocket, the main limitation of this method is that the derived interaction maps typically involve many unprioritized interaction features.

LIGAND-BASED DRUG DISCOVERY

Similarity Searching

The main principle and motivation behind the ligand-based approaches in drug discovery is a concept known as molecular similarity; based on this principle, molecules tend to perform similar biological effects due to the high structural similarity (Zhavoronkov et al., 2019). In other words, ligand-based drug discovery methods rely on the structural information of the active ligand that interacts with the target protein, and such a compound with interesting biological properties can be used as a query template in identifying and predicting new chemical entities with similar properties. Since only the structure of the known active small molecules are required, this methodology is considered as an indirective protocol for drug discovery. It offers an option when the 3D target protein structure is unknown or cannot be predicted. Hence, this approach is commonly applied to screen novel ligands with interesting biological activities *in silico* and to optimize the biological activities of ligands to improve drug pharmacokinetics including Adsorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) properties.

This simple and most widely used technique is based on molecular descriptors. Physicochemical properties (e.g., molecular weight, logP, Energy of high occupied molecular orbital (EHOMO), Energy of lowest unoccupied orbital (ELUMO), charges), as well as 2D fingerprint and 3D shape-similarity searches can be introduced as coordinates to represent

the reference compounds. The 2D fingerprint (Molprint2D and Unity 2D) and 3D shape similarity methods (MACCS), extended-connectivity fingerprints (ECFP), rapid overlay of chemical structures (ROCS), and Phase Shape, are more often used for molecular representation in virtual screening (Rush et al., 2005). For example, Bologa et al. (2006) applied 2D fingerprint and 3D shape-similarity methods to identify novel agonists of the estradiol receptor family receptor GPR30 (Bologa et al.). Furthermore, both methods have been successfully applied in virtual screenings, and both technology have exhibited better performance against a number of targets than docking methods in terms of the scalability and computational time. However, the main problem of the similar methods is their preference for input molecules and the difficulty in deciding which input structures to be used (Hu et al., 2012).

Ligand-Based Pharmacophore Mapping

Another more precise approach in comparison with the molecular descriptors is the pharmacophore-based approach, in which a pharmacophore model (PH4) is developed based on a group of active compounds. The IUPAC (International Union of Pure and Applied Chemistry) pointed out that a pharmacophore is "a collection of spatial and electronic characteristics necessary to ensure optimal supramolecular interactions with specific biological targets and to trigger (or block) their biological reactions" (Buckle et al., 2013). Thus, structural overlap of key molecular features derived from active compounds or a binding site in space are used as a pattern to represent the most probable chemical characteristics. The newly identified molecules that match and show a high complementation to the developed pharmacophore are likely to be active against the target protein of interests. Therefore, they can be selected as candidates for more further investigations. This approach has become a key computational strategy to promote and guide drug discoveries in the absence of macromolecular structures (Chao et al., 2007).

The process of pharmacophore modelling can be summarized as following: (i) Selection of a training set of ligands (active and inactive compounds). (ii) Molecular preparation (low energy conformations). (iii) Ligand alignment/superimposition and pharmacophore model generation. (iv) Validation of pharmacophore models (Chiang et al., 2009). Ligand-based pharmacophore modeling highly depends on the availability of a good training set of compounds manifesting the same binding mode.

QSAR Modeling

QSAR (Quantitative Structure Activity Relationship) is another ligand-based approach that relies on analyzing the biological activities of drugs using various molecular descriptors (MDs) or fingerprints (FPs). These models mathematically describe how the activities response to the targets according to the ligand's structural characteristics. QSAR was obtained by calculating the correlations between the properties of the ligand binding agent and the biological activity measured by experiments. Different ML and deep learning (DL) approaches have also been applied to develop QSAR models (Mendenhall and Meiler, 2016): including Support Vector Machine (SVM), Random Forest (RF), Polynomial Regression (PR), Multi Linear Regression (MLR), Artificial Neural Network (ANN). Unlike the pharmacophore models, QSAR models can measure biological activities quantitatively and can even find positive or negative effects according to certain characteristics of the molecule on its activity.

QSAR has been applied to many other molecular design purposes, such as predicting the new molecule analog activity, optimizing lead, and predicting new structural leads in drug discovery. In the classical 2D-QSAR approaches, the biological activity is related to physical and chemical features consisting of steric, electronic, and hydrophobic characters of drugs, and the relationships are represented as mathematical equations (Hansch and Fujita, 1964). More advanced 3D-QSAR approaches, such as comparative molecular field analysis (Cramer et al., 1988) and molecular similarity indexes in a comparative analysis (Klebe et al., 1994), are based on the force field calculations. The structural information of molecules is needed, and developed models are represented in 3D contour maps facilitating the visualization and interpretation.

USING MD SIMULATION TO FIND NEW DRUG BINDING SITES

Many important biological events rely on the information of protein-ligand complex interactions. The recognition and characterization of LBP is the key to understand the function of endogenous ligands and synthetic drug molecules. GPCRs perform an important role in a variety of physiological processes. GPCRs are a class of commonly used targets in drug discovery (Conn et al., 2009). Recent discovery indicated that beside binding to orthosteric sites, ligands could bind to different allosteric sites that are far away from the targeted binding pockets (Tautermann, 2014; Flock et al., 2015; Devree et al., 2016). Unfortunately, the position of such allosteric pocket is unclear without the information of experimental structures, and predicting the existence of such sites could facilitate the discovery of new drugs (Tautermann, 2014). A recent overview described the progresses in important computational tools for the prediction of functional sites, such as 3DLigandStie (http://www.sbg.bio.ic.ac.uk/~3dligandsite/), COACH-D (http://yanglab.nankai.edu.cn/COACH-D/), or SiteMap (https://www.schrodinger.com/sitemap), and many others. However, these reported tools often create multiple possible ligand binding sites, and sometimes it is not easy for the user to confirm which active pocket is real one for the compound binding. To overcome this limitation, methods based on molecular dynamics (MD) have been developed in recent years. For example, the supervised MD is an efficient approach for precise sampling and the identification of ligand-binding sites (Sabbadin and Moro, 2014; Deganutti et al., 2015; Cuzzolin et al., 2016). The conventional long-timescale MD has also been successfully applied for new drug binding sites (Chan et al., 2018). Similarly, a study by Chan et al. (2020) reported that an additional sodium ion, which located in the vicinity of the orthosteric binding site, by MD simulations (Chan et al., 2020). MD could also be applied for the recognition of the allosteric sites involved in protein kinases (Tong and Seeliger, 2015),

Ras proteins (Hancock, 2003), and *Staphylococcus aureus* Sortase A (Mazmanian et al., 1999). As above, information obtained from MD predictions provides new opportunities of drug discovery.

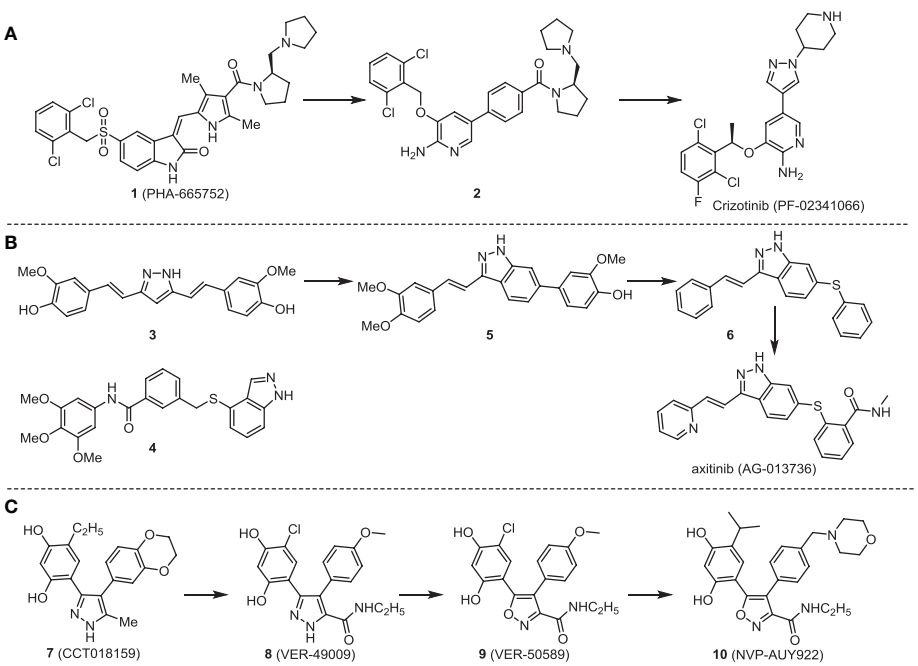
ARTIFICIAL INTELLIGENCE IN ANTI-CANCER DRUG DISCOVERY

Computational drug design has successfully promoted the discovery of several new anticancer drugs, which has become a milestone in this area. Gefitinib (Muhsin et al., 2003), Erlotinib (Grunwald and Hidalgo, 2003), Sorafenib (Wilhelm et al., 2006), Lapatinib (Wood et al., 2004), Abiraterone (Jarman et al., 1998), Crizotinib (Butrynski et al., 2010) are all approved drugs that have been discovered based on computational drug methods. Until now, the anticancer drug research is rapidly progressing: computational, and AI methods are generating new promising results. As an example, SR13668 is optimized from indole-3-carbinol (I3C) using PH4 design. SR13668 has shown a strong effect on different cancers in phase I (Chao et al., 2007). Recently, Rodrigues et al. have successfully identified a potent inhibitor for 5-lipoxygenase by using machine learning (ML)-based method which was developed from physicochemical and pharmacophore characteristics (Reker et al., 2014; Rodrigues et al., 2018). With the arrival of AI, the design of anticancer drugs *in silico* has undergone unprecedented changes, and state-of-the-art deep learning approaches have the potential to produce the excellent chemical properties needed for new molecules (Gomez-Bombarelli et al., 2018). Similarly, Jann et al. have developed the first ML-based anti-cancer compound generator using variational autoencoders (VAEs) and have demonstrated that the compound production may be selective toward molecules with high predicted inhibition to a specific cancer (Born et al., 2019). This implied that models could be developed to yield drug candidates with highly desired efficacy (IC_{50}) against a target of interest. This breakthrough could transform the design of anticancer drugs *in silico* by taking advantage of the bimolecular features of the disease to improve the success rate of lead compound discovery.

SUCCESSFUL STORIES OF COMPUTATIONAL DRUG DISCOVERY

Computational methods have proved to play an essential role in modern drug discovery. Since computational methods could cover almost all stages of the drug discovery pipeline, the applications of computational methods in anticancer drug discoveries have shown great advantages in terms of the required investment, resources, and time. More recently, computational methods have become a potent and powerful tool in several successful cases of anticancer drug development. Herein, we list several successful applications of computational methods for small molecule drugs, which have been applied to cancer treatment or are at later stages in the clinical trial.

The development of Crizotinib is a successful example of applying structure-based design techniques (Cui et al., 2011; Kung et al., 2015). Crizotinib has been considered as a selective and potent cMet/ALK dual inhibitor, which was approved by FDA in 2011 (Cui et al., 2013). c-Met, also known as HGFR (hepatocyte growth factor receptor), and its corresponding natural ligand HGF (hepatocyte growth factor) play a critical role in different cell activities (Christensen et al., 2005). The over-expression of c-Met protein has been often detected in human



SCHEME 1 | Successful applications of computational methods in anti-cancer drug discovery.

cancers (including SCLC and NSCLC) (Bottaro et al., 1991; Liu et al., 2008), and abnormal function of c-Met signaling was observed in various solid and blood tumor cancers. Thus, c-MET is an attractive and promising oncology target.

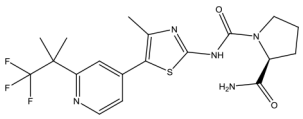
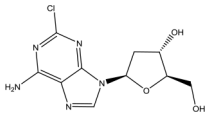
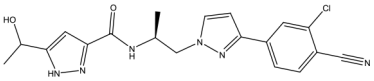
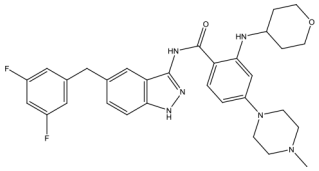
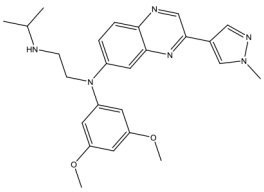
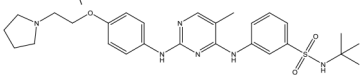
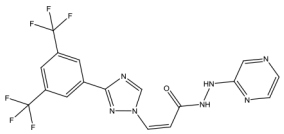
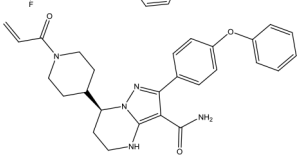
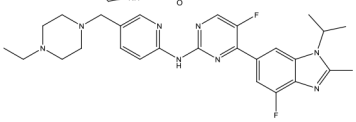
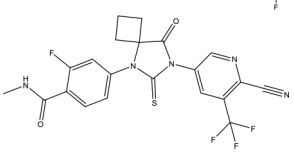
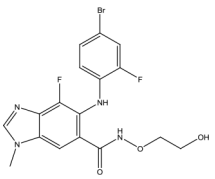
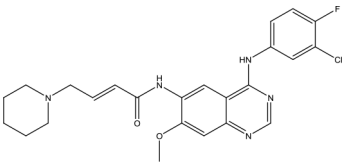
The investigation started with evaluating a series of 3-substituted indolin-2-ones, a potent class of kinase inhibitors, indolin-2-one derivatives for c-MET inhibition. Among the derivatives, compound 1 (PHA-665752, **Scheme 1**) showed strong activity against the c-MET autophosphorylation process and the corresponding biological activations both *in vitro* and *in vivo*. However, the bad drug-like characteristics of compound 1 (PHA-665752) limited its further study. The co-crystal structure analysis of compound 1 with the kinase domain of c-MET elucidated the key inhibitor binding site, presenting opportunities for more efficient drug designs. In combination with re-designing the central rings of compound 1 (PHA-665752), a new set of 5-substituted 3-benzyloxy-2-aminopyridine series has been developed. Among these newly designed derivatives, compound 2 displayed promising inhibition against c-MET. It is noted that lipophilic efficiency (LipE) was employed as the parameter for the binding effectiveness to monitor the progress of optimization. To further improve the c-Met inhibitory potency, a docked structure of compound 2 with the c-Met kinase domain was carried out to guide the application of structure-based design techniques. Followed by optimization of 3-benzyloxy group, the functional group at 5-position of the 2-aminopyridine, and examination of the chiral center, crizotinib (PF-02341066) with effective tumor growth inhibition and good drug performance has been achieved (see **Scheme 1A**). Moreover, Crizotinib has demonstrated remarkable clinical efficacy on c-MET gene amplification against lung cancer, lymphoma, and esophageal cancers (Cui et al., 2011; Lennerz et al., 2011; Schwab et al., 2014).

In 2012, Axitinib (AG-013736) was approved by the FDA as as a new therapy for advanced renal cell carcinoma (Meadows and Hurwitz, 2012) to treat VEHG. Axitinib was developed with a structure-based drug design strategy and served as an inhibitor by binding to the VEGF kinase domain in the DFG-out conformation (Kania, 2009; Kania et al., 2016). The VEGF (vascular endothelial growth factor) family functions as important regulators of many signaling networks which involves in angiogenesis. VEGF signaling was identified in tumor cells, and the VEGF signaling plays a crucial role in the development of malignant diseases. As the key receptors of VEGF, VEGFRs serve as ligands in the VEGF signaling network. The VEGF receptors are known as a class of the tyrosine kinases (RTKs), including VEGFR-1 (also called FLT1), VEGFR-2 (also called FLK1 and KDR) and VEGFR-3 (also called FLT4). Blocking the action of VEGFRs with a pan kinase inhibitor against VEGFR-1, VEGFR-2, and VEGFR-3 has been proved to be an efficient way of anti-angiogenic drug development.

During the developmental process, the crystal structure of phosphorylated construct (p-VEGFR2Δ50), the resolved structures of inhibitor–VEGFR2Δ50 (unphosphorylated kinase) complexes, and robust SAR provided important guidance to the rational drug design (Kania, 2009). Combining with the complex structure information, a collection of compounds has been evaluated, generating pyrazoles 3 and benzamide 4 as the starting point for the drug design. Further efforts have been made by the modeling of pyrazole 3 into the ligand-free p-VEGFR2Δ50 structure to modify the conformation of pyrazole 3 further, leading to the generation of indazole compounds as novel kinase inhibitors. Among these derivatives, compound 5 with a styryl functional group at the 3-position of the indazole ring was identified to exhibit potent inhibitory effect (Ki of 0.3 nM), with a high level of LipE and LE. The crystal structure of VEGFR2Δ50 with compound 5 revealed the detailed enzyme-ligand mode, showing the indazole core binding to the "open" DFG-in conformation of VEGFR2Δ50. Superimposing the other two VEGFR2Δ50–inhibitor co-crystal complex structures demonstrated a more precise 3D structure of the key binding sites for the induction of the DFG-out conformation. Inspired by the superposition result, a chimera design protocol was applied for the subsequent design to capture the above described inhibitor interactions, giving access to 6–sulfur linked indazole compound 6 and the corresponding amide analogs. Further studies on the overlay of VEGFR2Δ50 bound co-crystal structures of benzamide 4 and indazole 6 demonstrated that an additional amide group on the orthosteric site of S-phenyl group would help to make the two important hydrogen bonds with the hydrogen bonding groups from Glu885 and Asp1046 of VEGFR2Δ50 and provide highly potent inhibitors. Further applying the truncation strategy generated axitinib (AG-013736) (see **Scheme 1B**), which exhibited a remarkable improvement on cellular potency, desirable physiochemical, and PK properties. Very recently, axitinib (Inlyta®), in combination with pembrolizumab (KEYTRUDA®), was approved as the first-line anticancer drug against renal cell carcinoma (RCC)(Atkins et al., 2018).

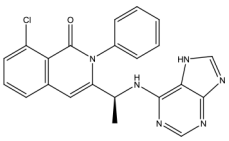
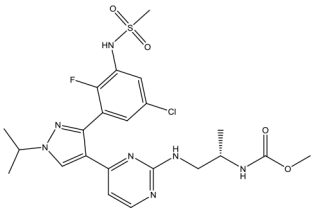
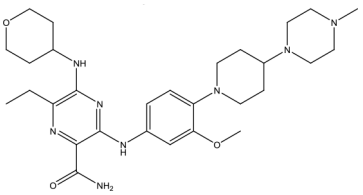
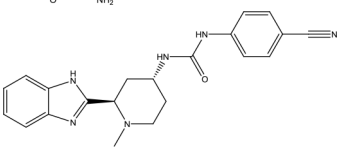
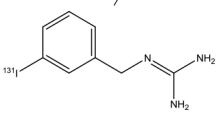
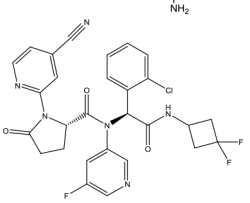
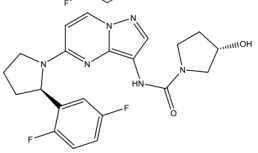
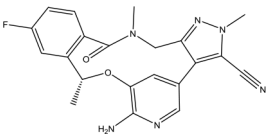
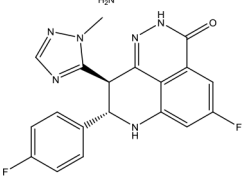
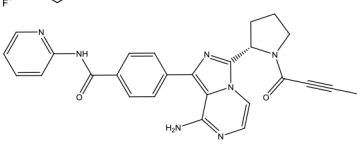
Heat shock protein 90 (HSP90) has direct and essential effects on the correct performance of different proteins with their activation, conformation, stabilization, and localization functions, whose alterations are associated with cancer development. Thus, HSP90 has become a promising target for cancer treatment (Whitesell and Lindquist, 2005; Pearl and Prodromou, 2006; Sharp and Workman, 2006; Workman et al., 2007). The biological functions of HSP90 have been identified. Its crystal structures indicated that HSP90 has four functional domains: a middle domain, an N-term domain, ATP/ADP-binding domain, and a C-term dimerization domain (Pearl and Prodromou, 2006). Based on the structural information of HSP90, a high-throughput screening was conducted which generated the active drug inhibitor: compound 7 (CCT018159) (Cheung et al., 2005; Smith et al., 2006; Sharp et al., 2007). The subsequently obtained co-crystal structure of HSP90-compound 7 (CCT018159) complex revealed that further modification of compound 7 (CCT018159) by replacing or adding certain functional groups could improve the pharmacokinetic properties. Moreover, replacing the methyl group to an amide group (VER-49009), changing pyrazolyl ring to isoxazole aromatic ring (VER-50589), and modifying some other chemical groups (see **Scheme 1C**) led to a potent effect in animal cancer models. Followed by toxicology and safety evaluation, Luminespib (NVP-AUY922) has been proved to be a strong HSP90 inhibitor which is now in clinical trials. More

TABLE 3 | The list of FDA-approved anticancer drugs in recent 3 years from the National Cancer Institute database.

Name	Chemical Structure	Therapeutic area	Target and functiuon	Year of Approval
Alpelisib		Breast cancer	PI3K inhibitor	2019 (Markham, 2019a)
Cladribine		Hairy cell leukemia	Adenosine deaminase inhibitor	2019 (Bryson and Sorkin, 1993)
Darolutamide		Prostate cancer	Androgen receptor inhibitor	2019 (Markham and Duggan, 2019)
Entrectinib		Non-small cell lung cancer and Solid tumors	Tyrosine kinase inhibitor	2019 (Al-Salama and Keam, 2019)
Erdafitinib		Urothelial carcinoma	FGFR tyrosine inhibitor	2019 (Markham, 2019b)
Fedratinib Hydrochloride		Myelofibrosis	Tyrosine kinase inhibitor	2019 (Zhang et al., 2014)
Selinexor		Multiple myeloma	Nuclear export inhibitor	2019 (Syed, 2019)
Zanubrutinib		Mantle cell lymphoma	Bruton's tyrosine kinase inhibitor	2019 (Syed, 2020)
Abemaciclib		Breast cancer	Cyclin-dependent kinase inhibitor	2018 (Kim, 2017b)
Apalutamide		Prostate cancer	Androgen receptor inhibitor	2018 (Al-Salama, 2019)
Binimetinib		Melanoma	MEk1 and MEK2 inhibitor	2018 (Shirley, 2018)
Dacomitinib		Non-small cell lung cancer	Oral kinase inhibitor	2018 (Sidaway, 2018)

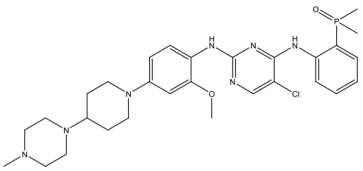
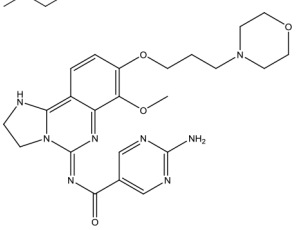
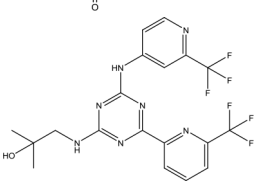
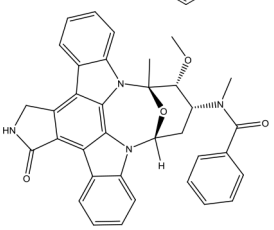
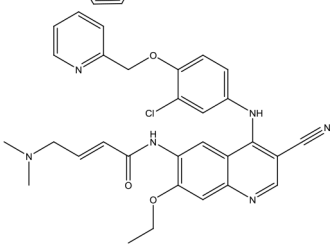
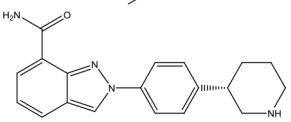
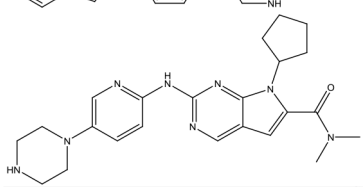
(Continued)

TABLE 3 | Continued

Name	Chemical Structure	Therapeutic area	Target and functiuon	Year of Approval
Duvelisib		Chronic lymphocytic leukemia (CLL) and follicular lymphoma (FL)	PI3K Kinase inhibitor	2018 (Blair, 2018)
Encorafenib		Colorectal cancer and Melanoma	BRAF Kinase inhibitor	2018 (Shirley, 2018)
Gilteritinib Fumarate		Acute myeloid leukemia	Tyrosine kinase inhibitor	2018 (Dhillon, 2019)
Glasdegib Maleate		Acute myeloid leukemia	Hedgehog pathway inhibitor	2018 (Shaik et al., 2019)
Iobenguane I 131		Pheochromocytoma	Radioactive therapeutic agent	2018 (Giammarile et al., 2008)
Ivosidenib		Acute myeloid leukemia	Isocitrate dehydrogenase-1 (IDH1) inhibitor	2018 (Dhillon, 2018)
Larotrectinib Sulfate		Solid tumors	Tropomyosin-related kinase (Trk) inhibitor	2018 (Gajdosik, 2017)
Lorlatinib		Non-small cell lung cancer	Tyrosine kinase inhibitor	2018 (Su et al., 2019)
Talazoparib Tosylate		Breast cancer	Poly (ADP-ribose) polymerase (PARP) inhibitor	2018 (Eskiler, 2019)
Acalabrutinib		Chronic lymphocytic leukemia, small lymphocytic lymphoma, and mantle cell lymphoma	Bruton's tyrosine kinase inhibitor	2017 (Markham and Dhillon, 2018)

(Continued)

TABLE 3 | Continued

Name	Chemical Structure	Therapeutic area	Target and functiuon	Year of Approval
Brigatinib		Non-small cell lung cancer	Anaplastic lymphoma kinase (ALK) and epidermal growth factor receptor (EGFR) kinase inhibitor	2017 (Markham, 2017a)
Copanlisib Hydrochloride		Follicular lymphoma	Phosphoinositide 3-kinase (PI3K) inhibitor	2017 (Markham, 2017b)
Enasidenib Mesylate		Acute myeloid leukemia	Isocitrate dehydrogenase-2 inhibitor	2017 (Gras, 2017)
Midostaurin		Acute myeloid leukemia	Synthetic indolocarbazole multikinase inhibitor	2017 (Kim, 2017a)
Neratinib Maleate		Breast cancer	Receptor tyrosine kinases (RTKs), Human epidermal growth factor receptor 2 (HER2; ERBB2), and Human epidermal growth factor receptor (EGFR) inhibitor	2017 (Kotecki et al., 2019)
Niraparib Tosylate Monohydrate		Recurrent epithelial ovarian, fallopian tube and primary peritoneal cancer	Poly (ADP-ribose) polymerase (PARP) inhibitor	2017 (Mittica et al., 2018)
Ribociclib		Breast cancer	Cyclin-dependent kinase (CDK) inhibitor	2017 (Syed, 2017)

We further manually screened the database to remove drugs that do not directly target cancer. Drugs for ameliorating conditions related to cancer or limiting side effects of cancer therapies are not listed in this short list. We then identified the FDA label of the drugs in the shortlist by searching in the U.S. National Library of Medicine database "DailyMed". The FDA approval date, drug function, and therapeutic area are retrieved from DailyMed database.

recently, Luminespib, a drug in phase one clinical trials, exhibited positive results for patients with ALK rearrangements (Felip et al., 2018). Luminespib (NVP-AUY922) also exhibited potent anti-tumor activity in lung adenocarcinomas targeting EGFR exon 20 insertion mutations and cellular models in a confirmatory clinical trial (Jorge et al., 2018; Piotrowska et al., 2018). Moreover, Luminespib (NVP-AUY922) serves as one of the components in anticancer combination therapies, which are now at different stages of clinical trials (Garcia-Carbonero et al., 2013; Rong and Yang, 2018). To depict how computational drug discovery facilitates to the development of anticancer drugs, we listed the FDA-approved anticancer drug in recent 3 years which was obtained from National Cancer Institute database (Heller, 1951) in **Table 3**.

CONCLUSION AND PERSPECTIVE

Cancer has become a tangible threat to human health. About 9.6 million people are estimated to die from the various forms of cancer each year, according to a statistic report (Collaborators, 2019). Cancer has become the second-largest disease that causes human death (Reimann et al., 2020). However, developing a new drug molecule costs 12 years and 2.7 billion USD on average (Hauser et al., 2017). The drug development for cancer even becomes more complicated, especially considering the molecular pharmacology is still not well understood. Hence, the discovery and development of new drugs is considered very expensive and time-consuming. In this respect, computational methods could be constructive for performing different tasks including protein-interaction network analysis, drug-target prediction, binding site prediction, virtual screening, and many others. All these innovative methods could considerably facilitate the anti-cancer drug discovery. In recent years, with the advance of AI, more sophisticated methods, such as retro-synthetic routine plan, drug scaffold generation, drug binding affinity predictions, were developed. The useful predictions generated by computational models combined with experimental validations could further speed up the anti-cancer drug development.

AUTHOR CONTRIBUTIONS

SY designed the whole review. WC directed the completion of the review. AA, SW, QY, and YL were supportive during the review.

FUNDING

This work was supported by the internal funding of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

REFERENCES

Al-Salama, Z. T., and Keam, S. J. (2019). Entrectinib: first global approval. *Drugs* 79, 1477–1483. doi: 10.1007/s40265-019-01177-y

Al-Salama, Z. T. (2019). Apalutamide: A Review in Non-Metastatic Castration-Resistant Prostate Cancer. *Drugs* 79, 1591–1598. doi: 10.1007/s40265-019-01194-x

Anthony, C. S., Masuyer, G., Sturrock, E. D., and Acharya, K. R. (2012). Structure Based Drug Design of Angiotensin-I Converting Enzyme Inhibitors. *Curr. Med. Chem.* 19, 845–855. doi: 10.2174/092986712799034950

Atkins, M. B., Plimack, E. R., Puzanov, I., Fishman, M. N., Mcdermott, D. F., Cho, D. C., et al. (2018). Axitinib in combination with pembrolizumab in patients with advanced renal cell cancer: a non-randomised, open-label, dose-finding, and dose-expansion phase 1b trial. *Lancet Oncol.* 19, 405–415. doi: 10.1016/S1470-2045(18)30081-0

Blair, H. A. (2018). Duvelisib: First global approval. *Drugs* 78, 1847–1853. doi: 10.1007/s40265-018-1013-4

Bohm, H. J. (1992). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comp. Aided Mol. Design* 6, 61–78. doi: 10.1007/BF00124387

Bologa, C. G., Revankar, C. M., Young, S. M., Edwards, B. S., Arterburn, J. B., Kiselyov, A. S., et al. (2006). Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* 2, 207–212. doi: 10.1038/nchembio775

Born, J., Manica, M., Oskooei, A., and Rodriguez Martínez, M. (2019). *PaccMannRL: Designing anticancer drugs from transcriptomic data via reinforcement learning*. New York: Cornell University Press.

Bottaro, D. P., Rubin, J. S., Faletto, D. L., Chan, A. M., Kmiecik, T. E., Vande Woude, G. F., et al. (1991). Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. *Sci. (New York N.Y.)* 251, 802–804. doi: 10.1126/science.1846706

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca-a Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Bryson, H. M., and Sorkin, E. M. (1993). Cladribine—A review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential in hematological malignancies. *Drugs* 46, 872–894. doi: 10.2165/00003495-199346050-00007

Buckle, D. R., Erhardt, P. W., Ganellin, C. R., Kobayashi, T., Perun, T. J., Proudfoot, J., et al. (2013). Glossary of terms used in medicinal chemistry. Part II (IUPAC Recommendations 2013). *Pure Appl. Chem.* 85, 1725–1758. doi: 10.1351/PAC-REC-12-11-23

Butrynski, J. E., D'adamo, D. R., Hornick, J. L., Dal Cin, P., Antonescu, C. R., Jhanwar, S. C., et al. (2010). Crizotinib in ALK-Rearranged Inflammatory Myofibroblastic Tumor. *New Engl. J. Med.* 363, 1727–1733. doi: 10.1056/NEJMoa1007056

Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science* 321, 263–266. doi: 10.1126/science.1158140

Chan, H. C. S., Wang, J., Palczewski, K., Filipek, S., Vogel, H., Liu, Z.-J., et al. (2018). Exploring a new ligand binding site of G proteincoupled receptors. *Chem. Sci.* 9, 11. doi: 10.1039/C8SC01680A

Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., and Yuan, S. (2019). Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* 40, 592–604. doi: 10.1016/j.tips.2019.06.004

Chan, H. C. S., Xu, Y., Tan, L., Vogel, H., Cheng, J., Wu, D., et al. (2020). Enhancing the Signaling of GPCRs via Orthosteric Ions. *ACS Cent. Sci.* 6, 274–282. doi: 10.1021/acscentsci.9b01247

Chao, W.-R., Yean, D., Amin, K., Green, C., and Jong, L. (2007). Computer-aided rational drug design: A novel agent (SR13668) designed to mimic the unique anticancer mechanisms of dietary indole-3-carbinol to block akt signaling. *J. Med. Chem.* 50, 3412–3415. doi: 10.1021/jm070040e

Chen, J., and Lai, L. (2006). Pocket v.2: Further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.* 46, 2684–2691. doi: 10.1021/ci600246s

Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). Drug-target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.* 8, 1970–1978. doi: 10.1039/c2mb00002d

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., et al. (2016). Drug-target interaction prediction: databases, web servers and computational models. *Briefings Bioinf.* 17, 696–712. doi: 10.1093/bib/bbv066

Cheung, K. M. J., Matthews, T. P., James, K., Rowlands, M. G., Boxall, K. J., Sharp, S. Y., et al. (2005). The identification, synthesis, protein crystal structure and in vitro biochemical evaluation of a new 3,4-diarylpyrazole class of Hsp90 inhibitors. *Bioorg. Med. Chem. Lett.* 15, 3338–3343. doi: 10.1016/j.bmcl.2005.05.046

Chiang, Y. K., Kuo, C. C., Wu, Y. S., Chen, C. T., Coumar, M. S., Wu, J. S., et al. (2009). Generation of Ligand-Based Pharmacophore Model and Virtual Screening for Identification of Novel Tubulin Inhibitors with Potent Anticancer Activity. *J. Med. Chem.* 52, 4221–4233. doi: 10.1021/jm801649y

Christensen, J. G., Burrows, J., and Salgia, R. (2005). c-Met as a target for human cancer and characterization of inhibitors for therapeutic intervention. *Cancer Lett.* 225, 1–26. doi: 10.1016/j.canlet.2004.09.044

Collaborators, G.B.D.S (2019). Global, regional, and national burden of stroke 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 18, 439–458. doi: 10.1016/s1474-4422(19)30034-1

Conn, P. J., Christopoulos, A., and Lindsley, C. W. (2009). Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat. Rev. Drug Discovery* 8, 41–54. doi: 10.1038/nrd2760

Cramer, R. D., Patterson, D. E., and Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110, 5959–5967. doi: 10.1021/ja00226a005

Cui, J. J., Tran-Dube, M., Shen, H., Nambu, M., Kung, P.-P., Pairish, M., et al. (2011). Structure Based Drug Design of Crizotinib (PF-02341066), a Potent and Selective Dual Inhibitor of Mesenchymal-Epithelial Transition Factor (c-MET) Kinase and Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* 54, 6342–6363. doi: 10.1021/jm2007613

Cui, J. J., Mctigue, M., Kania, R., and Edwards, M. (2013). Case History: XalkoriTM (Crizotinib), a Potent and Selective Dual Inhibitor of Mesenchymal Epithelial Transition (MET) and Anaplastic Lymphoma Kinase (ALK) for Cancer Treatment.*Annu. Rep. Med. Chem.* 48, 421–434. doi: 10.1016/b978-0-12-417150-3.00025-9

Cuzzolin, A., Sturlese, M., Deganutti, G., Salmaso, V., Sabbadin, D., Ciancetta, A., et al. (2016). Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J. Chem. Inf. Model.* 56, 687–705. doi: 10.1021/acs.jcim.5b00702

Debnath, S., Kanakaraju, M., Islam, M., Yeeravalli, R., Sen, D., and Das, A. (2019). In silico design, synthesis and activity of potential drug-like chrysin scaffoldd-erived selective EGFR inhibitors as anticancer agents. *Comput. Biol. Chem.* 83. doi: 10.1016/j.compbiolchem.2019.107156

Deganutti, G., Cuzzolin, A., Ciancetta, A., and Moro, S. (2015). Understanding allosteric interactions in G protein-coupled receptors using Supervised Molecular Dynamics: A prototype study analysing the human A(3) adenosine receptor positive allosteric modulator LUF6000. *Bioorg. Med. Chem.* 23, 4065–4071. doi: 10.1016/j.bmc.2015.03.039

Devree, B. T., Mahoney, J. P., Velez-Ruiz, G. A., Rasmussen, S. G. F., Kuszak, A. J., Edwald, E., et al. (2016). Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature* 535, 182–18+. doi: 10.1038/nature18324

Dhillon, S. (2018). Ivosidenib: First Global Approval. *Drugs* 78, 1509–1516. doi: 10.1007/s40265-018-0978-3

Dhillon, S. (2019). Gilteritinib: First Global Approval. *Drugs* 79, 331–339. doi: 10.1007/s40265-019-1062-3

Dias, R., and De Azevedo, W. F.Jr. (2008). Molecular Docking Algorithms. *Curr. Drug Targets* 9, 1040–1047. doi: 10.2174/138945008786949432

Drews, J. (2000). Drug discovery: a historical perspective. *Sci. (New York N.Y.)* 287, 1960–1964. doi: 10.1126/science.287.5460.1960

Eskiler, G. G. (2019). Talazoparib to treat BRCA-positive breast cancer. *Drugs Today* 55, 459–467. doi: 10.1358/dot.2019.55.7.3015642

Felip, E., Barlesi, F., Besse, B., Chu, Q., Gandhi, L., Kim, S.-W., et al. (2018). Phase 2 Study of the HSP-90 Inhibitor AUY922 in Previously Treated and Molecularly Defined Patients with Advanced Non-Small Cell Lung Cancer. *J. Thoracic Oncol.* 13, 576–584. doi: 10.1016/j.jtho.2017.11.131

Ferreira, L. G., Dos Santos, R. N., Oliva, G., and Andricopulo, A. D. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* 20, 13384–13421. doi: 10.3390/molecules200713384

Flock, T., Ravarani, C. N. J., Sun, D., Venkatakrishnan, A. J., Kayikci, M., Tate, C. G., et al. (2015). Universal allosteric mechanism for G alpha activation by GPCRs. *Nature* 524, 173–17+. doi: 10.1038/nature14663

Gajdosik, Z. (2017). Larotrectinib sulfate. *Drugs Future* 42, 275–280. doi: 10.1358/dof.2017.042.05.2623108

Garcia-Carbonero, R., Carnero, A., and Paz-Ares, L. (2013). Inhibition of HSP90 molecular chaperones: moving into the clinic. *Lancet Oncol.* 14, E358–E369. doi: 10.1016/S1470-2045(13)70169-4

Ghofrani, H. A., Osterloh, I. H., and Grimminger, F. (2006). Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat. Rev. Drug Discovery* 5, 689–702. doi: 10.1038/nrd2030

Giammarile, F., Chiti, A., Lassmann, M., Brans, B., and Flux, G. (2008). EANM procedure guidelines for I-131-meta-iodobenzylguanidine (I-131-mIBG) therapy. *Eur. J. Nuclear Med. Mol. Imaging* 35, 1039–1047. doi: 10.1007/s00259-008-0715-3

Gomez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernandez-Lobato, J. M., Sanchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* 4, 268–276. doi: 10.1021/acscentsci.7b00572

Gras, J. (2017). Enasidenib mesylate. *Drugs Future* 42, 15–20. doi: 10.1358/dof.2017.042.01.2579894

Grunwald, V., and Hidalgo, M. (2003). Development of the epidermal growth factor receptor inhibitor Tarceva (TM) (OSI-774). *New Trends in Cancer for the 21st Century*. 235–246. doi: 10.1007/978-1-4615-0081-0_19

Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409–443. doi: 10.1002/prot.10115

Hancock, J. F. (2003). Ras proteins: Different signals from different locations. *Nat. Rev. Mol. Cell Biol.* 4, 373–384. doi: 10.1038/nrm1105

Hansch, C., and Fujita, T. (1964). Additions and Corrections-ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* 86, 5710–5710. doi: 10.1021/ja01078a623

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B., and Gloriam, D. E. (2017). Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discovery* 16, 829–842. doi: 10.1038/nrd.2017.178

Heller, J.R. (1951). National Canaer Institute. *JAMA-J. Am. Med. Assoc.* 146, 1248–1248. doi: 10.1001/jama.1951.03670130070024

Hong, J. Y., Price, I. R., Bai, J. J., and Lie, H. (2019). A Glycoconjugated SIRT2 Inhibitor with Aqueous Solubility Allows Structure-Based Design of SIRT2 Inhibitors. *ACS Chem. Biol.* 14, 1802–1810. doi: 10.1021/acschembio.9b00384

Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690. doi: 10.1038/nchembio.118

Hu, G., Kuang, G., Xiao, W., Li, W., Liu, G., and Tang, Y. (2012). Performance Evaluation of 2D Fingerprint and 3D Shape Similarity Methods in Virtual Screening. *J. Chem. Inf. Model.* 52, 1103–1113. doi: 10.1021/ci300030u

Huang, S.-Y., Grinter, S. Z., and Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12, 12899–12908. doi: 10.1039/c0cp00151a

Itoh, Y. (2020). Drug Discovery Researches on Modulators of Lysine-Modifying Enzymes Based on Strategic Chemistry Approaches. *Chem. Pharmaceut. Bull.* 68, 34–45. doi: 10.1248/cpb.c19-00741

Jarman, M., Barrie, S. E., and Llera, J. M. (1998). The 16,17-double bond is needed for irreversible inhibition of human cytochrome P450(17 alpha) by abiraterone (17-(3-pyridyl)androsta-5,16-dien-3 beta-ol) and related steroidal inhibitors. *J. Med. Chem.* 41, 5375–5381. doi: 10.1021/jm981017j

Jorge, S. E., Lucena-Araujo, A. R., Yasuda, H., Piotrowska, Z., Oxnard, G. R., Rangachari, D., et al. (2018). EGFR Exon 20 Insertion Mutations Display Sensitivity to Hsp90 Inhibition in Preclinical Models and Lung Adenocarcinomas. *Clin. Cancer Res.* 24, 6548–6555. doi: 10.1158/1078-0432.CCR-18-1541

Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K., et al. (1997). Viracept (nelfinavir mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40, 3979–3985. doi: 10.1021/jm9704098

Kania, R. S., Bender, S. L., Borchardt, A. J., Cripps, S. J., Hua, Y., Johnson, M. D., et al. (2016) *Indazole compounds and pharmaceutical compositions for Inhibiting protein kinases, and methods for their use* [Accessed November 28, 2006].

Kania, R. (2009). *"Structure-Based Design and Characterization of Axitinib. Kinase Inhibitor Drugs, Drug Discovery and Development*. Eds. R. Li and J. A. Stafford (New York: Wiley).

Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chem. Biol. Interact.* 171, 165–176. doi: 10.1016/j.cbi.2006.12.006

Kim, E. S. (2017a). Midostaurin: First Global Approval. *Drugs* 77, 1251–1259. doi: 10.1007/s40265-017-0779-0

Kim, E. S. (2017b). Abemaciclib: first global approval. *Drugs* 77, 2063–2070. doi: 10.1007/s40265-017-0840-z

Klebe, G., Abraham, U., and Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 37, 4130–4146. doi: 10.1021/jm00050a010

Klipp, E., Wade, R. C., and Kummer, U. (2010). Biochemical network-based drug-target prediction. *Curr. Opin. Biotechnol.* 21, 511–516. doi: 10.1016/j.copbio.2010.05.004

Kortagere, S., and Ekins, S. (2010). Troubleshooting computational methods in drug discovery. *J. Pharmacol. Toxicol. Methods* 61, 67–75. doi: 10.1016/j.vascn.2010.02.005

Kotecki, N., Gombos, A., and Awada, A. (2019). Adjuvant therapeutic approaches of HER2-positive breast cancer with a focus on neratinib maleate. *Expert Rev. Anticancer Ther.* 19, 447–454. doi: 10.1080/14737140.2019.1613892

Kung, P. P., Jones, R. A., and Richardson, P. (2015). *Crizotinib (Xalkori): The First-in-Class ALK/ROS Inhibitor for Non-small Cell Lung Cancer* (Inc: John Wiley & Sons).

Lazo, J. S., and Sharlow, E. R. (2016). Drugging Undruggable Molecular Cancer Targets.. *Annu. Rev. Phar. Toxicol.* 56, 23–40. doi: 10.1146/annurev-pharmtox-010715-103440

Lennerz, J. K., Kwak, E. L., Ackerman, A., Michael, M., Fox, S. B., Bergethon, K., et al. (2011). MET Amplification Identifies a Small and Aggressive Subgroup of Esophagogastric Adenocarcinoma With Evidence of Responsiveness to Crizotinib. *J. Clin. Oncol.* 29, 4803–4810. doi: 10.1200/JCO.2011.35.4928

Li, X., and Chen, H. C. (2013). Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach. *Decision Support Syst.* 54, 880–890. doi: 10.1016/j.dss.2012.09.019

Liu, X., Yao, W., Newton, R. C., and Scherle, P. A. (2008). Targeting the c-MET signaling pathway for cancer therapy. *Expert Opin. Investigational Drugs* 17, 997–1011. doi: 10.1517/13543784.17.7.997

Lu, P., Bevan, D. R., Leber, A., Hontecillas, R., Tubau-Juni, N., and Bassaganya-Riera, J. (2018a). "Computer-aided drug discovery," in *Accelerated Path to Cures* (Springer), 7–24. doi: 10.1007/978-3-319-73238-1_2

Lu, X., Yang, H., Chen, Y., Li, Q., He, S.-Y., Jiang, X., et al. (2018b). The Development of Pharmacophore Modeling: Generation and Recent Applications in Drug Discovery. *Curr. Pharmaceut. Design* 24, 3424–3439. doi: 10.2174/1381612824666180810162944

Markham, A., and Dhillon, S. (2018). Acalabrutinib: First Global Approval. *Drugs* 78, 139–145. doi: 10.1007/s40265-017-0852-8

Markham, A., and Duggan, S. (2019). Darolutamide: First Approval. *Drugs* 79, 1813–1818. doi: 10.1007/s40265-019-01212-y

Markham, A. (2017a). Brigatinib: First Global Approval. *Drugs* 77, 1131–1135. doi: 10.1007/s40265-017-0776-3

Markham, A. (2017b). Copanlisib: First Global Approval. *Drugs* 77, 2057–2062. doi: 10.1007/s40265-017-0838-6

Markham, A. (2019a). Alpelisib: first global approval. *Drugs* 79, 1249–1253. doi: 10.1007/s40265-019-01161-6

Markham, A. (2019b). Erdafitinib: First Global Approval 79, 1017–1021. doi: 10.1007/s40265-019-01142-9

Mathi, P., Prasad, M. V., Botlagunta, M., Ravi, M., and Ramachandran, D. (2018). De novo design of selective Sortase-A inhibitors: Synthesis, structural and in vitro characterization. *Chem. Data Collect.* 15, 126–133. doi: 10.1016/j.cdc.2018.04.007

Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M., Wegner, J. K., Ceulemans, H., et al. (2018). Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* 9, 5441–5451. doi: 10.1039/C8SC00148K

Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999). Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Sci. (New York N.Y.)* 285, 760–763. doi: 10.1126/science.285.5428.760

Meadows, K. L., and Hurwitz, H. I. (2012). Anti-VEGF Therapies in the Clinic. *Cold Spring Harbor Perspect. Med.* 2, 27. doi: 10.1101/cshperspect.a006577

Mendenhall, J., and Meiler, J. (2016). Improving quantitative structure-activity relationship models using Artificial Neural Networks trained with dropout. *J. Comp. Aided Mol. Design* 30, 177–189. doi: 10.1007/s10822-016-9895-2

Mendoza, J. L., Escalante, N. K., Jude, K. M., Bellon, J. S., Su, L., Horton, T. M., et al. (2019). Structure of the IFN gamma receptor complex guides design of biased agonists. *Nature* 567, 56–5+. doi: 10.1038/s41586-019-0988-7

Mittica, G., Ghisoni, E., Giannone, G., Genta, S., Aglietta, M., Sapino, A., et al (2018). PARP Inhibitors in Ovarian Cancer. *Recent Pat. Anticancer Drug Discov.* 13, 392–410. doi: 10.2174/1574892813666180305165256

Muhsin, M., Graham, J., and Kirkpatrick, P. (2003). Fresh from the pipeline - Gefitinib. *Nat. Rev. Drug Discovery* 2, 515–516. doi: 10.1038/nrd1136

Ortuso, F., Langer, T., and Alcaro, S. (2006). GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics* 22, 1449–1455. doi: 10.1093/bioinformatics/btl115

Pearl, L. H., and Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* 75, 271–294. doi: 10.1146/annurev.biochem.75.103004.142738

Piotrowska, Z., Costa, D. B., Oxnard, G. R., Huberman, M., Gainor, J. F., Lennes, I. T., et al. (2018). Activity of the Hsp90 inhibitor luminespib among non-small-cell lung cancers harboring EGFR exon 20 insertions. *Ann. Oncol.* 29, 2092–2097. doi: 10.1093/annonc/mdy336

Pirhadi, S., Shiri, F., and Ghasemi, J. B. (2013). Methods and applications of structure based pharmacophores in drug discovery. *Curr. Top. Med. Chem.* 13, 1036–1047. doi: 10.2174/1568026611313090006

Prada-Gracia, D., Huerta-Yepez, S., and Moreno-Vargas, L. M. (2016). Application of computational methods for anticancer drug discovery, design, and optimization. *Boletin Med. Del Hosp. Infantil. Mexico* 73, 411–423. doi: 10.1016/j.bmhimx.2016.10.006

Reimann, Z., Miller, J. R., Dahle, K. M., Hooper, A. P., Young, A. M., Goates, M. C., et al. (2020). Executive functions and health behaviors associated with the leading causes of death in the United States: A systematic review. *J. Health Psychol.* 25, 186–196. doi: 10.1177/1359105318800829

Reker, D., Rodrigues, T., Schneider, P., and Schneider, G. (2014). Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. United States America* 111, 4067–4072. doi: 10.1073/pnas.1320001111

Rodrigues, T., Werner, M., Roth, J., Da Cruz, E. H. G., Marques, M. C., Akkapeddi, P., et al. (2018). Machine intelligence decrypts -lapachone as an allosteric 5-lipoxygenase inhibitor. *Chem. Sci.* 9, 6. doi: 10.1039/c8sc02634c.

Rong, B., and Yang, S. (2018). Molecular mechanism and targeted therapy of Hsp90 involved in lung cancer: New discoveries and developments (Review). *Int. J. Oncol.* 52, 321–336. doi: 10.3892/ijo.2017.4214

Rush, T. S., Grant, J. A., Mosyak, L., and Nicholls, A. (2005). A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein?Protein Interaction. *J. Med. Chem.* 48, 1489–1495. doi: 10.1021/jm040163o

Sabbadin, D., and Moro, S. (2014). Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR-Ligand Recognition Pathway in a Nanosecond Time Scale. *J. Chem. Inf. Model.* 54, 372–376. doi: 10.1021/ci400766b

Salmaso, V., and Moro, S. (2018). Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.* 9. doi: 10.3389/fphar.2018.00923

Schwab, R., Petak, I., Kollar, M., Pinter, F., Varkondi, E., Kohanka, A., et al. (2014). Major partial response to crizotinib, a dual MET/ALK inhibitor, in a squamous cell lung (SCC) carcinoma patient with de novo c-MET amplification in the absence of ALK rearrangement. *Lung Cancer* 83, 109–111. doi: 10.1016/j.lungcan.2013.10.006

Shaik, N., Hee, B., Wei, H., and Labadie, R. R. (2019). Evaluation of the effects of formulation, food, or a proton-pump inhibitor on the pharmacokinetics of glasdegib (PF-04449913) in healthy volunteers: a randomized phase I study. *Cancer Chemother. Pharmacol.* 83, 463–472. doi: 10.1007/s00280-018-3748-8

Sharp, S., and Workman, P. (2006). Inhibitors of the HSP90 molecular chaperone: Current status. *Adv. Cancer Res.* 95. doi: 10.1016/s0065-230x(06)95009-x

Sharp, S. Y., Boxall, K., Rowlands, M., Prodromou, C., Roe, S. M., Maloney, A., et al. (2007). In vitro biological characterization of a novel, synthetic diaryl pyrazole resorcinol class of heat shock protein 90 inhibitors. *Cancer Res.* 67, 2206–2216. doi: 10.1158/0008-5472.CAN-06-3473

Shirley, M. (2018). Encorafenib and Binimetinib: First Global Approvals. *Drugs* 78, 1277–1284. doi: 10.1007/s40265-018-0963-x

Sidaway, P. (2018). Cemiplimab effective in cutaneous SCC. *Nat. Rev. Clin. Oncol.* 15, 472–472. doi: 10.1038/s41571-018-0056-5

Smith, N. F., Hayes, A., James, K., Nutley, B. P., Mcdonald, E., Henley, A., et al. (2006). Preclinical pharmacokinetics and metabolism of a novel diaryl pyrazole resorcinol series of heat shock protein 90 inhibitors. *Mol. Cancer Ther.* 5, 1628–1637. doi: 10.1158/1535-7163.MCT-06-0041

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 15, 1929–1958. doi: 10.5555/2627435.2670313

Su, Y., Long, X., Song, Y., Chen, P., Li, S., Yang, H., et al. (2019). Distribution of ALK Fusion Variants and Correlation with Clinical Outcomes in Chinese Patients with Non-Small Cell Lung Cancer Treated with Crizotinib. *Targeted Oncol.* 14, 159–168. doi: 10.1007/s11523-019-00631-x

Syed, Y. Y. (2017). Ribociclib: First Global Approval. *Drugs* 77, 799–807. doi: 10.1007/s40265-017-0742-0

Syed, Y. Y. (2019). Selinexor: First Global Approval. *Drugs* 79, 1485–1494. doi: 10.1007/s40265-019-01188-9

Syed, Y. Y. (2020). Zanubrutinib: First Approval. *Drugs*. 80, 91–97. 1–7. doi: 10.1007/s40265-019-01252-4

Takarabe, M., Kotera, M., Nishimura, Y., Goto, S., and Yamanishi, Y. (2012). Drug target prediction using adverse event report systems: a pharmacogenomic approach. *Bioinformatics* 28, I611–I618. doi: 10.1093/bioinformatics/bts413

Tautermann, C. S. (2014). GPCR structures in drug design, emerging opportunities with new structures. *Bioorg. Med. Chem. Lett.* 24, 4073–4079. doi: 10.1016/j.bmcl.2014.07.009

Tondo, A. R., Caputo, L., Mangiatordi, G. F., Monaci, L., Lentini, G., Logrieco, A. F., et al. (2020). Structure-Based Identification and Design of Angiotensin Converting Enzyme-Inhibitory Peptides from Whey Proteins. *J. Agric. Food Chem.* 68, 541–548. doi: 10.1021/acs.jafc.9b06237

Tong, M., and Seeliger, M. A. (2015). Targeting Conformational Plasticity of Protein Kinases. *ACS Chem. Biol.* 10, 190–200. doi: 10.1021/cb500870a

Urwyler, S. (2011). Allosteric Modulation of Family C G-Protein-Coupled Receptors: from Molecular Insights to Therapeutic Perspectives. *Pharmacol. Rev.* 63, 59–126. doi: 10.1124/pr.109.002501

Wang, J. L., Liu, D., Zhang, Z. J., Shan, S., Han, X., Srinivasula, S. M., et al. (2000). Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc. Natl. Acad. Sci. United States America* 97, 7124–7129. doi: 10.1073/pnas.97.13.7124

Whitesell, L., and Lindquist, S. L. (2005). HSP90 and the chaperoning of cancer. *Nat. Rev. Cancer* 5, 761–772. doi: 10.1038/nrc1716

Wilhelm, S., Carter, C., Lynch, M., Lowinger, T., Dumas, J., Smith, R. A., et al. (2006). Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. *Nat. Rev. Drug Discovery* 5, 835–844. doi: 10.1038/nrd2130

Wolber, G., Dornhofer, A. A., and Langer, T. (2006). Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comp. Aided Mol. Design* 20, 773–788. doi: 10.1007/s10822-006-9078-7

Wood, E. R., Truesdale, A. T., Mcdonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., et al. (2004). A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): Relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* 64, 6652–6659. doi: 10.1158/0008-5472.CAN-04-1168

Workman, P., Burrows, F., Neckers, L., and Rosen, N. (2007). ""Drugging the cancer chaperone HSP90 combinatorial therapeutic exploitation of oncogene addiction and tumor stress," in *Stress Responses in Biology and Medicine: Stress of Life in Molecules, Cells, Organisms, and Psychosocial Communities*. Eds. P. Csermely, T. Korcsmaros and K. Sulyok. 1113, 202–216. doi: 10.1196/annals.1391.012

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, I232–I240. doi: 10.1093/bioinformatics/btn162

Yan, B., Yang, W. J., Han, X. Y., and Han, L. H. (2019). Crystal structures and antitumor activity evaluation against gastric carcinoma of two novel coordination polymers. *Main Group Chem.* 18, 239–246. doi: 10.3233/MGC-180748

Yang, X., Wang, Y., Byrne, R., Schneider, G., and Yang, S. (2019). Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem. Rev.* 119, 10520–10594. doi: 10.1021/acs.chemrev.8b00728

Yang, S.-Y. (2010). Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* 15, 444–450. doi: 10.1016/j.drudis.2010.03.013

Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., and Vidal, M. (2007). Drug-target network. *Nat. Biotechnol.* 25, 1119–1126. doi: 10.1038/nbt1338

Yuriev, E., Holien, J., and Ramsland, P. A. (2015). Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *J. Mol. Recogn.* 28, 581–604. doi: 10.1002/jmr.2471

Zhang, Q., Zhang, Y., Diamond, S., Boer, J., Harris, J. J., Li, Y., et al. (2014). The Janus kinase 2 inhibitor fedratinib inhibits thiamine uptake: a putative mechanism for the onset of Wernicke's encephalopathy. *Drug Metab. Dispos.* 42, 1656–1662. doi: 10.1124/dmd.114.058883

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–103+. doi: 10.1038/s41587-019-0224-x

Conflict of Interest: SY is the cofounder of AlphaMol Science Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cui, Aouidate, Wang, Yu, Li and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 16 June 2020
doi: 10.3389/fphar.2020.00898

Preparation of RGD Peptide/Folate Acid Double-Targeted Mesoporous Silica Nanoparticles and Its Application in Human Breast Cancer MCF-7 Cells

Huijie Yan, Yun You, Xinjian Li, Lei Liu, Fengqian Guo, Qiongling Zhang, Dewen Liu, Yan Tong, Shilan Ding and Jinyu Wang**

Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Anil Tiwari,
University of Pittsburgh, United States
Anju Gupta,
University of Toledo, United States

****Correspondence:***
Shilan Ding
dsl2009.ok@163.com
Jinyu Wang
jinyu024@163.com

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *26 March 2020*
Accepted: *02 June 2020*
Published: *16 June 2020*

Citation:
Yan H, You Y, Li X, Liu L, Guo F, Zhang Q, Liu D, Tong Y, Ding S and Wang J (2020) Preparation of RGD Peptide/Folate Acid Double-Targeted Mesoporous Silica Nanoparticles and Its Application in Human Breast Cancer MCF-7 Cells. Front. Pharmacol. 11:898. doi: 10.3389/fphar.2020.00898

Pharmacological Relevance: Paclitaxel (PTX) is currently the only botanical drug that can control the growth of cancer cells. Paclitaxel is widely used in the treatment of breast cancer, ovarian cancer, uterine cancer, non-small cell lung cancer and other cancers.

Aim: Folate receptor and integrin $\alpha_v\beta_3$ are highly expressed on the surface of human breast cancer cells MCF-7. Folic acid and arginine-glycine-aspartate (Arg-Gly-Asp, RGD) tripeptide sequence have a high affinity for folate receptor and integrin $\alpha_v\beta_3$, respectively. To enhance the effect on breast cancer, we constructed the folate acid and RGD peptide dual-targeted (MSNs-NH_2-FA-RGD) drug-carrier based on mesoporous silica nanoparticles.

Methods: The structure of mesoporous nanocarriers was characterized by Fourier transform infrared spectroscopy, nitrogen adsorption-desorption analysis, transmission electron microscopy, laser particle size analyzer, and thermogravimetric analysis. Paclitaxel was chosen as the model drug. The targeting-ability was verified by observing the uptake of mesoporous carriers loaded with rhodamine in MCF-7, MCF-10A, and HeLa cells using a fluorescence microscope. The cytotoxicity of the blank carrier MSNs-NH_2-FA-RGD and the efficacy of the drug carrier PTX@MSNs-NH_2-FA-RGD were assessed by cell experiments.

Results: The characterization showed successful construction of a dual-targeted mesoporous silica nanocarrier. Obvious differences were detected in the fluorescence intensity of the three cell lines. The results of the pharmacological tests indicated that the blank nanoparticles do not cause any apparent toxicity on these cells. The IC_{50} of free PTX and PTX@MSNs-NH_2-FA-RGD on MCF-7 cells line treated for 48 h were 35.25±2.57 ng·ml^{-1} and 22.21±3.4 ng·ml^{-1} respectively, which indicated that the inhibitory efficacy of PTX@MSNs-NH_2-FA-RGD on MCF-7 was 1.6 times than that of free PTX.

Conclusions: The dual-targeted nanocarrier MSNs-NH_2-FA-RGD could target breast cancer cells, and sever as a potential candidate in future of drug development.

Keywords: mesoporous silica, active targeted, folic acid, RGD peptide, MCF-7 cells, paclitaxel

INTRODUCTION

Breast cancer seriously harms women's health. Statistics show that 1.2 million women worldwide suffer from breast cancer each year, and 500,000 women die from breast cancer (Raviv, 2004). In recent years, the patient population has shown a trend of being younger. Traditional chemotherapy exposes many issues, such as poor specificity by chemotherapeutics, drug resistance caused by repeated and large doses of multi-drugs, and side effects to normal tissues. Fortunately, the treatment has received widespread attention in the medical community. Many researchers have put a lot of effort into targeted treatment field. With in-depth research, nanocarriers are found to play an increasingly important role in targeted therapy. Owing to passive or active targeted delivery of drugs, the nanocarriers have shown great potential in improving drug concentration and bioavailability in tumor sites (Darvishi and Farahmand, 2017).

Mesoporous silica nanoparticles (MSNs) are unique among numerous inorganic nanomaterials due to their good biocompatibility, high load capacity, and uniform adjustable pore size (Song et al., 2007; Kazuki et al., 2013). By modifying the surface of MSNs with different substances and groups, the MSNs carrier can be endowed with the ability of targeting and stimulate-responsive, avoiding the early leakage of drugs and increasing the concentration of drugs at the lesion sites. The surface of MSNs could also be easily functionalized with a variety of targeted groups, such as antibody (Zhang et al., 2015; Tao et al., 2016), protein (Pourjavadi and Tehrani, 2016), peptides, and small molecules (Alejandro et al., 2014). Kazuki et al (Kazuki et al., 2013) wrapped the peptide Ac-$(VKVS)_4$E-NH_2 on the surface of MSNs. The conformation of Ac-$(VKVS)_4$E-NH_2 changed in different pH environments, controlling the exposure and coverage of the mesoporous orifice, thereby the carrier system showed pH-dependent release behavior. Xue et al. (2011) utilized the supramolecular force between benzimidazole and β-cyclodextrin, choosing fluorescent dye Hoechst 33342 as model cargo, to construct a cyclodextrin-based silica pH controlled release system. This carrier exhibits an acid-responsive ability to release drugs and induce cancer cell apoptosis in human pancreatic cancer cell PANC-1. Martinez-Carmona et al., (2018) modified the targeted ligand plant lectin concanavalin (ConA) on MSNs to enable the nanocarrier to specifically recognize human osteosarcoma cells. Tian et al., (2016) utilized iron-binding glycoprotein (Tf) not only as an entrant into the target sites but also as a blocking agent to inhibit the release of the drug before entering the tumor cells. He et al., (2012) functionalized MSNs with nucleic acid sequence polyadenylic acid and loaded with coralyne and near-infrared photothermal dye indocyanine green (ICG), constructing a nano-therapy platform combining chemical and photothermal therapy.

Folic acid receptors (FR) and integrin $\alpha_v\beta_3$ have the characteristic of expression specificity, that is, they are highly expressed on the surface of tumor cells but not expressed on the surface of normal cells. RGD peptide is a type of short peptide containing arginine-glycine-aspartate (Arg-Gly-Asp, RGD) peptide sequence, which is the smallest unit that can be recognized by integrin $\alpha_v\beta_3$ (Hee Dong et al., 2010; Porta et al., 2013). Folic acid (FA) and RGD peptides have strong affinity for FR and integrin $\alpha_v\beta_3$, so FA and RGD peptides are often used as targeted modification groups. There have been studies on separately grafting FA or RGD peptide onto mesoporous silica carriers. In this experiment, for the first time, we grafted both FA and RGD peptide on the surface of MSNs as shown in **Figure 1**, constructing a dual-targeting nanocarrier MSNs-NH_2-FA-RGD. Modifying folic acid groups (NHS-PEG-FA) and RGD peptide groups (NHS-PEG-RGD) onto the surface of MSNs endows the carrier with the ability to actively target cancer sites, and the PEG long chains enhance the *in vivo* stability of the carrier.

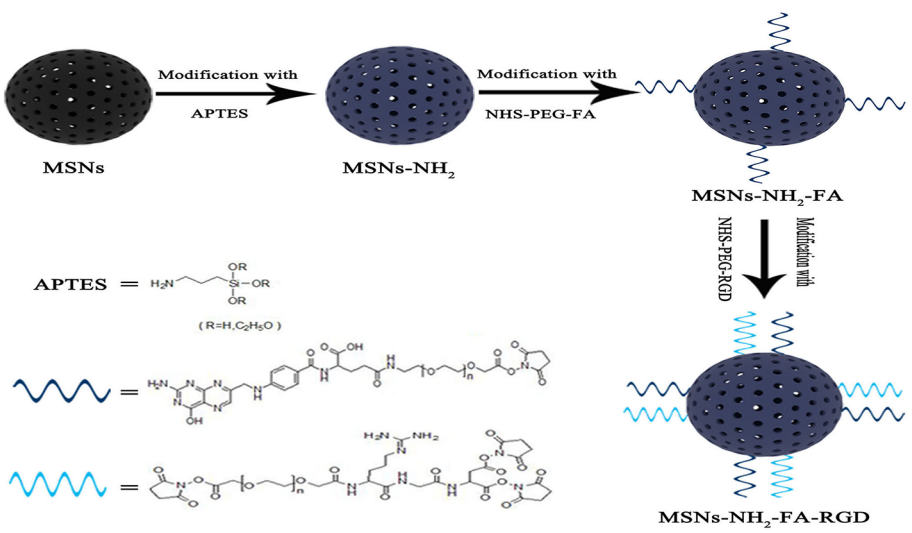


FIGURE 1 | Schematic illustration of preparation process of MSNs-NH_2-FA-RGD.

Paclitaxel (PTX) is a tetracyclic diterpenoid compound with potent effects on cancers, such as breast cancer, colon cancer, bronchial cancer, and endometrial cancer. PTX could specifically bind to different sites of the tubules and aggregate the microtubules, thereby inhibiting the normal reorganization of the microtubule network and interfering with the cell mitosis. It is one of the most energetic anti-tumor drugs since doxorubicin (Ren et al., 2005). However, PTX has low bioavailability and extremely poor solubility (water solubility of 0.006 $g{\cdot}L^{-1}$), which brings some difficulties to clinical application (Lv et al., 2014). We chose PTX as a model drug to evaluate the drug loading of MSNs-NH_2-FA-RGD@PTX. Loading PTX of nanometer-sized into MSNs-NH_2-FA-RGD would not only solve the problem of poor solubility but also significantly improve bioavailability. FR and integrin $\alpha_v\beta_3$ are highly expressed on the surface of MCF-7 cells but not on the surface of human normal breast epithelial cells MCF-10A, and human cervical cancer cells HeLa only express FR on the surface (Shen et al., 2011). Cell experiments were used to evaluate the biocompatibility, anti-cancer efficacy and cellular uptake of nanocarriers.

MATERIALS AND METHODS

Materials

Cetyltrimethylammonium bromide (CTAB), 3-aminopropyltrietho xysilane (APTES), tetraethylorthosilicate (TEOS), and rhodamine (RhB) were purchased from Aladdin Chemical Reagents (Shanghai, China,). Folic acid polyethylene glycol succinimidyl activated ester (FA-PEG-NHS, PEG= 2000) and RGD tripeptide polyethylene glycol succinimidyl activated ester (RGD-PEG-NHS, PEG= 2000) were purchased from Pengshuo Biotechnology Co., Ltd (Shanghai, China). Paclitaxel (purity ≥ 98 %) was purchased from Ruifensi Co., Ltd (Chengdu, China). Lysotracker green DND-22 was obtained from Invitrogen Life Technologies Corporation (Tianjing, China). Other reagents and solvents were provided by Dingguo reagent company (Beijing, China). All the chemical reagents used in this experiment were of analytical grade and used without further purification.

Characterization

The mesoporous structure and morphology of the nanoparticles were characterized by transmission electron microscopy (TEM) (Tecnai G2 F30, USA) at an accelerated voltage of 300 kV. The dispersing agent is water and the dispersions are stabile in 12 h. Nitrogen adsorption-desorption analysis at 77 K was carried out on an adsorption analyzer (ASAP 2460, Micromeritics, USA). Zeta potential and particle diameter experiments were performed at 25 °C using Malvern ZetaSizer Nano-S90. The Fourier transform infrared (FTIR) spectra were obtained on an FTIR spectrometer (Nexus, Thermo Nicolet, USA). Thermogravimetric analysis (TGA) was performed by a Thermo Gravimetric Analyzer (STA8000, Perkin Elmer, USA) under N_2 atmosphere at a heating rate of 10 $°C{\cdot}min^{-1}$. All fluorescence spectra were obtained on a fluorescence microscope (Hitachi F-7000 FL Spectrophotometer).

Preparation of MSNs-NH_2-FA-RGD

3 g CTAB was solubilized in 1440 ml of deionized water in three-necked flask heated to 80°C in an oil bath. Then the temperature of the CTAB solution was adjusted to 80°C before adding 10.5 ml sodium hydroxide aqueous solution (2.0 $mol{\cdot}L^{-1}$), followed by dropwise addition of 15 ml TEOS under vigorous stirring. After 2 h, the resultant product was collected by filtration using a suction pump and rinsed with ethanol. To remove the surfactant template CTAB, the product was calcined at 550°C for 4 h in a muffle furnace to obtain MSNs.

The introduction of aminopropyl groups through the post-grafting process was conducted by dispersing 0.9 g MSNs in 90 ml toluene, followed by the addition of 434 μl APTES. The mixture was refluxed and stirred at 90°C in an oil bath for 6 h, followed by centrifugation with 10000 rpm for 15 min and washing with ethanol and distilled water at room temperature. The resultant product was dried to a constant weight under vacuum to obtain MSNs-NH_2.

An equivalent of 40 mg of NHS-PEG-FA was dispersed by ultrasonication in 50 ml dimethyl sulfoxide (DMSO), and the pH of the system was adjusted to be alkaline by triethylamine. Then, 200 mg of MSNs-NH_2 was mixed in DMSO by then magnetic stirring for 4 h. The solids were collected by centrifugation and washing with ethanol. After drying under a vacuum atmosphere, MSNs-NH_2-FA was obtained. MSNs-NH_2-FA-RGD was synthesized similarly. 200 mg of MSNs-NH_2-FA was dispersed in DMSO-triethylamine with 40 mg of NHS-PEG-RGD. After 4 h, the reaction product was collected by centrifugation, washing, and vacuum drying.

Preparation of Drug-Loaded Nanocarriers

10 mg PTX and 20 mg MSNs-NH_2-FA-RGD were ultrasonically dispersed in 20 ml absolute ethanol and magnetically stirred for 24 h at room temperature to load the drug. Subsequently, the PTX-loaded MSNs-NH_2-FA-RGD (denoted as PTX@MSNs-NH_2-FA-RGD) was collected by suction filtration, with the surface adsorbed PTX washed away by phosphate-buffered saline (PBS) (pH=7.4). PTX@MSNs-NH_2-FA-RGD were collected after vacuum drying.10 mg drug-loaded particles was placed in a volumetric flask with methanol, followed by sonication for 1 h and analysis by high-performance liquid chromatography (HPLC). Drug loading rate (%) and entrapment rate (%) were calculated by HPLC at maxima wavelength of 229 nm using the following equation which was quoted from the Pharmacopoeia of the People's Republic of China:

$$\text{Drug loading rate} = (W_1 - W_2)/W_{nanocarriers} \times 100\,\%$$

$$\text{Encapsulation rate} = (W_1 - W_2)/W_1 \times 100\,\%$$

where W_1, W_2 and $W_{nanocarriers}$ represented the weight of PTX added, the weight of PTX in supernatant and the weight of nanocarriers.

RhB-Labeled Nanocarriers

RhB was used as a guest molecule to evaluate the ability of targeting tumor sites because of its fluorescence properties. 200 mg of MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD nanoparticles were mixed in the ethanol solution of RhB (0.4 $mg{\cdot}ml^{-1}$) for 4h. After centrifugation for 15 min at room temperature with 10000 rpm, the solid particles were dried in vacuum to constant weight. The RhB-labeled mesoporous silica nanoparticles were termed as RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, RhB@MSNs-NH_2-FA-RGD, respectively.

Cell Culture

The cell culture tests were performed using HeLa, MCF-7, and MCF-10A cells purchased from the American Type Culture Collection (Manassas, VA, USA). MCF-7 cells and HeLa cells were cultured in RPMI 1640 medium with 10% heat-inactivated fetal bovine serum (FBS). MCF-10A cells were cultured in DMEM/F12 medium with 5 % horse serum, 10 $\mu g{\cdot}ml^{-1}$ insulin, 20 $ng{\cdot}ml^{-1}$ EGF, 100 $ng{\cdot}ml^{-1}$ cholera toxin, and 0.5 $\mu g{\cdot}ml^{-1}$ hydrocortisone. All cells were cultivated in an incubator with 5 % CO_2 at 37°C.

Cell Uptake and Location

Collect HeLa, MCF-7, and MCF-10A cells in the logarithmic growth phase and seed them in 96 wells at a density of 6×10^4, 6×10^4, and 1.5×10^5 cells/ml. After the cells were incubated for 24 h, aspirate the medium. The cells were incubated with RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, and RhB@MSNs-NH_2-FA-RGD (20 $\mu g{\cdot}ml^{-1}$) for 4 h. Each well was washed three times with cold PBS to remove the nanoparticles not internalized into the cells and then the cell morphology was fixed with 4% paraformaldehyde for 5 min. Subsequently, the nucleus was stained with DAPI for 5 min, while lysosomes were identified using the dye named Lysotracker. Fluorescence microscopy of fluorescein-labeled cells was performed with an Imaging System equipped with three Led Lights Cubes (BioFlux 1000Z, USA, Fluxion Biosciences).

In Vitro Toxicity Test of Blank Carrier

The CCK- 8 method was used to determine the toxicity of the blank nanocarrier MSNs-NH_2-FA-RGD to MCF-7 cells. Collect MCF-7 cells in the logarithmic growth phase and seed them in 96 wells at a density of 6×10^4 cells/ml. After the cells were incubated for 24 h, we aspirated the medium and added 100 µl of complete medium containing different concentrations of MSNs-NH_2-FA-RGD (concentrations of 20, 40, 80, and 160 $\mu g{\cdot}ml^{-1}$) to each well. Cultivate MCF-7 cells in a constant temperature incubator for 24 h or 48 h. Measure the absorbance of each well at 450 nm by micro plate reader (Thermo scientific, USA) and calculate the inhibition rate.

In Vitro Antitumor Drug Efficacy

Collect MCF-7 cells in the logarithmic growth phase and seed them in 96 wells at 6×10^4 cells/ml. Configure the complete medium for PTX@MSNs-NH_2-FA-RGD and free PTX to different concentrations (based on the PTX concentration as a quantitative basis, and set the concentration gradient to 10, 30, 100, 300, 1000 $ng{\cdot}ml^{-1}$). After culturing MCF-7 cells for 24 h, aspirate the medium and add complete medium with different concentrations of PTX mentioned above. Cultivate MCF-7 cells in constant temperature incubator for 24 h and 48 h. Measure the absorbance of each well at 450 nm by a microplate reader (Thermo scientific, USA) and calculate the inhibition rate.

RESULTS AND DISCUSSION

Preparation and Characterization of MSNs-NH_2- FA-RGD Nanocarrier

TEM images showed that the MSNs and MSNs-NH_2-FA-RGD nanoparticles were spherical, with smooth surface and even distribution (**Figures 2A, B**). After modification, the ordered mesopores could still be directly observed from **Figure 2B**, which proved that modified process would not affect the mesoporous structure of them. The laser particle size analyzer showed that the average particle sizes of MSNs and MSNs-NH_2-FA-RGD were 188.6 nm (PDI= 0.267) and 204.1 nm (PDI= 0.269), respectively (**Figure 2C**). Zeta potential of MSNs in distilled water was -18.4 ± 4.30 mV. Because of the amino group on the surface, zeta potential of MSNs-NH_2 was reversed to 25.6 ± 3.8 mV after the process of amination. Due to the PEG long chains on the targeted group covering the positive charge of MSNs-NH_2, the positive potential of MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD have decreased to 24.4 ± 7.36 mV and 22.9 ± 3.9 mV, respectively (**Figure 2D**).

Figure 3A showed the FT-IR spectra of a) MSNs, b) MSNs-NH_2, c) MSNs-NH_2-FA, and d) MSNs-NH_2-FA-RGD. As shown in curve (a), the strongest absorption peak at 1083 cm^{-1} was the symmetric stretching vibration peak of Si-O-Si. 3400 cm^{-1} and 1640 cm^{-1} were the stretching vibration peak and bending vibration peak of Si-OH, respectively, indicating that there were hydroxyl groups with different bonds and states on the surface of SiO_2. In the curve (b), the stretching vibration peak of the methylene group at 2926 cm^{-1} and the bending vibration peak of the amino group at 1470 cm^{-1} both indicated that the amination process was successful. In the curve (c), the C=O vibration absorption peak and the O=C-N-H absorption peak at 1737 cm^{-1} and 1556 cm^{-1} indicated that FA was grafted on the surface. The new absorption peak at 1538 cm^{-1} in curve (d) was assigned to the amide bond which was affected by the RGD peptide and shifted to the direction of short wave number. The process of modification was quantified by TGA analysis. **Figure 3B** showed the thermograms of a) MSNs, b) MSNs-NH_2, c) MSNs-NH_2-FA, and d) MSNs-NH_2-FA-RGD. According to analysis of Pyris software, the weight loss from room temperature to 800°C of MSNs, MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD were 4.4%, 14.51%, 19.04%, and 24.37%, respectively. Weightlessness in different temperature ranges represents

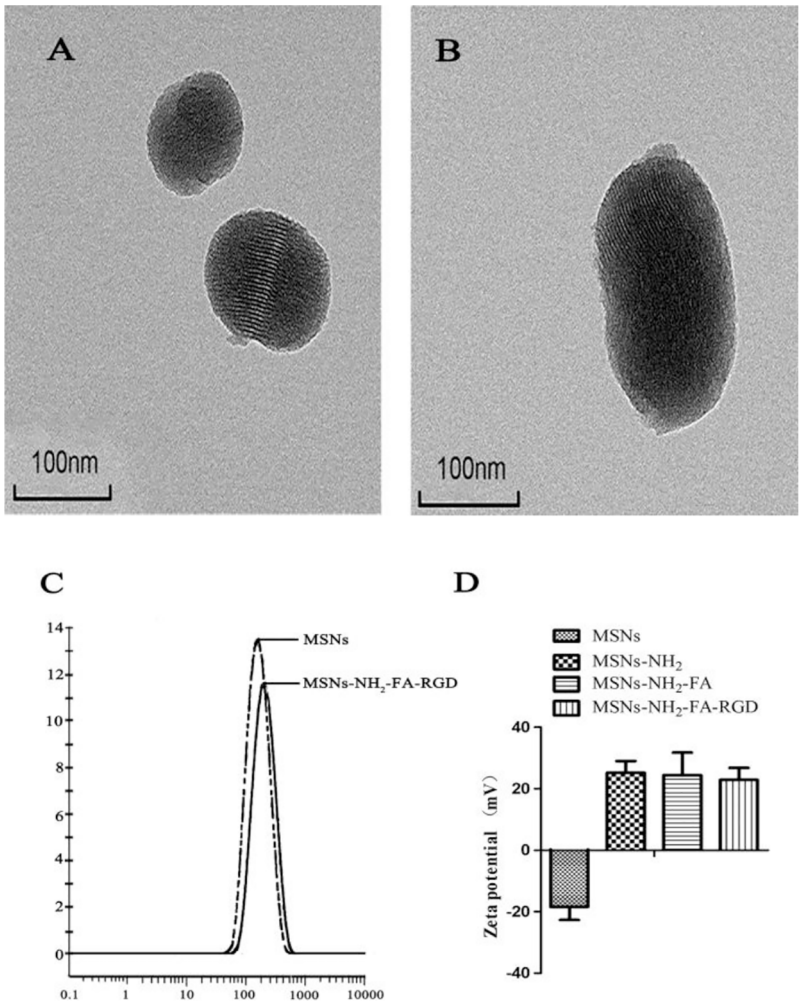


FIGURE 2 | TEM images of **(A)** MSNs and **(B)** MSNs-NH_2-FA-RGD. **(C)** Particle size distribution of MSNs and MSNs-NH_2-FA-RGD in deionized water. **(D)** Zeta potential of MSNs, MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD in deionized water.

different meaning (Pourjavadi and Tehrani, 2014). Take MSNs as an example, the weight loss below 200°C was 2.03%, causing by absorbed water in the air, and the proportion of this part of weightlessness can be directly read on the graph through the software. The weightlessness between 200°C and 800°C could be attributed to the removal of organic groups. The weightlessness of MSNs above 200°C was 2.37 %, which could be indicated as the incomplete removal of CTAB. After deducting the proportion of CTAB and absorbed water, the weightlessness between 200° C to 800°C of MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD were 10.30%, 14.17%, and 18.56%, respectively. And the graft ratios of the amino group, FA, and RGD peptide were about 7.93%, 3.87%, and 4.39%.

Figure 4 showed the nitrogen adsorption-desorption isotherms and the corresponding pore size distributions of the MSNs, MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD. Textural parameters were listed in **Table 1**. It could be observed that the adsorption amount of N_2 increased slowly when P/P_0 was less than 0.25 from the adsorption isotherms of MSNs, because the adsorption of N_2 on the surface of the sample channel occurred in single-molecule and multi-molecular layers. When P/P_0 was between 0.25-0.4, the adsorption amount of N_2 raised sharply. A steep platform peak was observed on the adsorption isotherm curve with the reason that N_2 could cause capillary condensation in the sample channel at low temperature, indicating that there was ordered mesoporous structure and uniform pore size distribution in the MSNs sample. When P/P_0 was between 0.4-0.95, the curve was relatively flat, due to the adsorption of N_2 on the outer surface. When P/P_0 was more than 0.95, the curve appeared a small jump, the reason was that the capillary condensation caused by the pores between the particles. With the introduction of FA and RGD

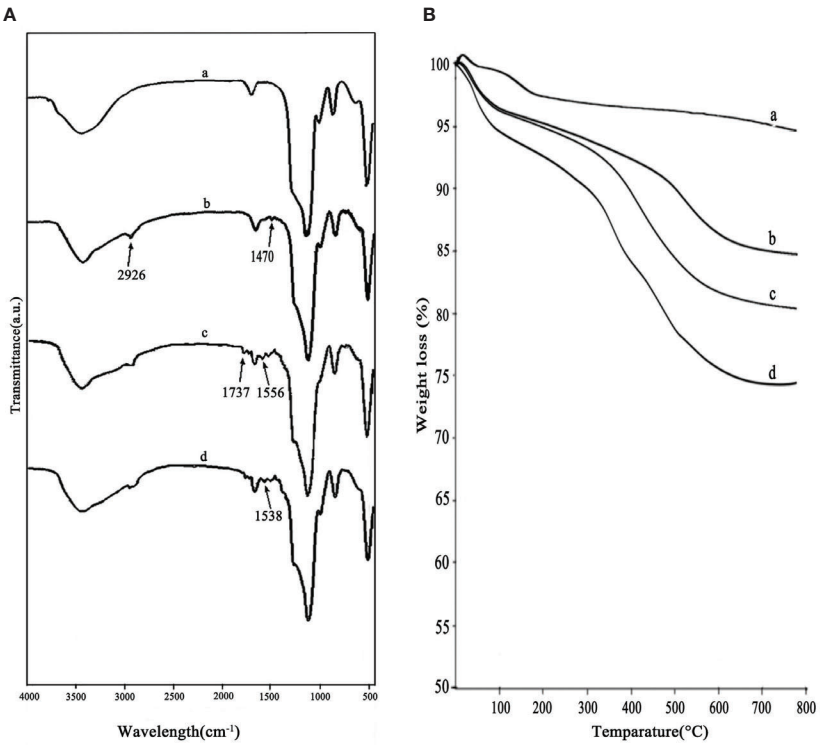


FIGURE 3 | FTIR spectra **(A)** and thermograms **(B)** of MSNs (a), MSNs-NH_2 (b), MSNs-NH_2-FA (c), and MSNs-NH_2-FA-RGD (d).

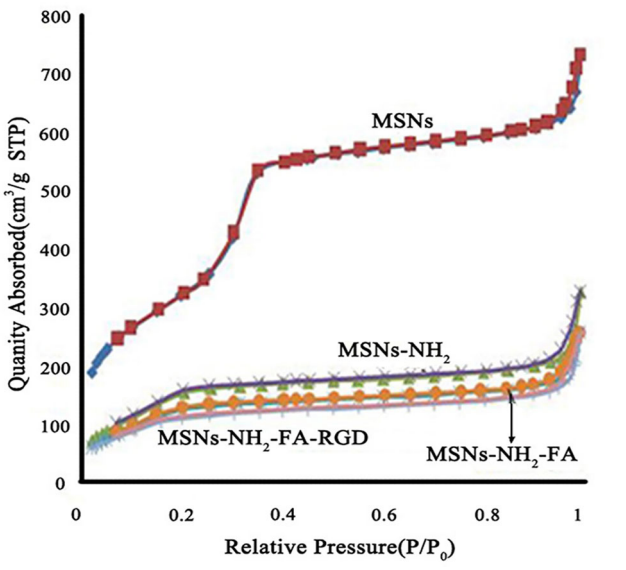


FIGURE 4 | Nitrogen adsorption-desorption isotherms of MSNs, MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD.

peptide, the relative pressure value P/P_0 of the steep peak decreased, and the specific surface area, pore volume and pore size of MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD also decreased, proving that the amino group, NHS-PEG-FA and NHS-PEG-RGD covered on the surface of MSNs.

Drug Loading Efficiency

In order to estimate the ability of drug-loading of MSNs-NH_2-FA-RGD, PTX was chosen as the model drug. The loading and entrapment efficiency were estimated at 18.7% and 85.2%, respectively.

In Vitro Cytotoxicity of Blank Nanoparticles

A safe and effective nanocarrier system is a prerequisite to *in vivo* therapy. Therefore, we explored the *in vitro* cytotoxicity of

TABLE 1 | The nitrogen adsorption-desorption parameters of different functionalized MSNs samples.

Samples	Surface area (m^2/g)	Pore size (nm)	Pore volume (cm^3/g)
MSNs	1203.41	3.77	1.13
MSNs-NH_2	659.34	3.09	0.51
MSNs-NH_2-FA	514.86	3.07	0.40
MSNs-NH_2-FA-RGD	426.68	3.03	0.40

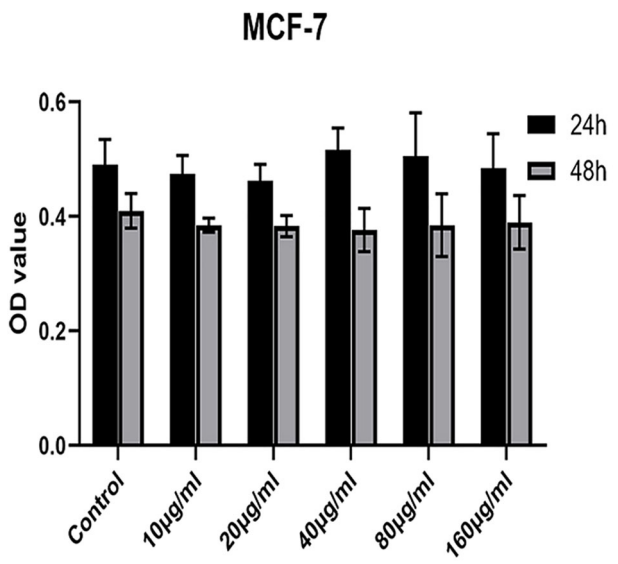


FIGURE 5 | Cell viability of MCF-7 cells incubation with MSNs-NH_2-FA-RGD at different concentrations (μg·ml^{-1}) after 24 h and 48 h.

MSNs-NH_2-FA-RGD nanoparticles. As shown in **Figure 5**, MSNs-NH_2-FA-RGD did not show any cytotoxicity compared with the blank control group at any of the concentrations used, even incubated after for 48 h, demonstrating that MSNs-NH_2-FA-RGD nanoparticles had excellent cytocompatibility.

Cell Uptake Assay

Firstly, we evaluated the cytotoxic potential of MSNs-NH_2, MSNs-NH_2-FA, and MSNs-NH_2-FA-RGD at 20 μg·ml^{-1} in HeLa, MCF-7, and MCF-10A cells for 24 h. The above three blank nanoparticles did not exert any apparent toxicity on the viability of these cells from **Figure 6**.

As shown in **Figure 7**, the three nanocarriers RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, RhB@MSNs-NH_2-FA-RGD showed no red RhB fluorescence in MCF-10A cells, indicating that MCF-10A cells hardly took up these three nanocarriers. Because of an excessive amount of FR on the surface, HeLa cells could specifically uptake the nanocarriers modified with FA groups. As shown in **Figure 8**, the red fluorescence representing nanocarriers coincided with the green lysosome region, proving that RhB@MSNs-NH_2-FA and RhB@MSNs-NH_2-FA-RGD were endocytosed and distributed in the cytoplasm of HeLa cells, but the fluorescence was not significantly different. There was no obvious fluorescence in the RhB@MSNs-NH_2 group. **Figure 9** showed that compared with RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, and RhB@MSNs-NH_2-FA-RGD showed obvious red fluorescence in MCF-7 cells, and RhB@MSNs-NH_2-FA-RGD had the strongest fluorescence intensity in them. This indicated that targeted modification with FA and RGD peptides enhanced the enrichment of nanocarriers in MCF-7 cells, and the double-targeted effect of modification was better, fully highlighted the advantages of receptor-mediated targeting.

In Vitro Antitumor Drug Efficacy

Figure 10 showed the inhibition rate of MCF-7 cells by PTX and PTX@MSNs-NH_2-FA-RGD at different concentrations after 24 h and 48 h. MSNs-NH_2-FA-RGD had no significant inhibitory effect on MCF-7 cells, therefore the toxic effects of PTX@MSNs-NH_2-FA-RGD on MCF-7 cells were all from PTX. The experimental results showed that both free PTX and PTX@MSNs-NH_2-FA-RGD showed inhibitory effects on MCF-7 cells, and the inhibitory effects were concentration-dependent and time-dependent. At 24 h of incubation, the

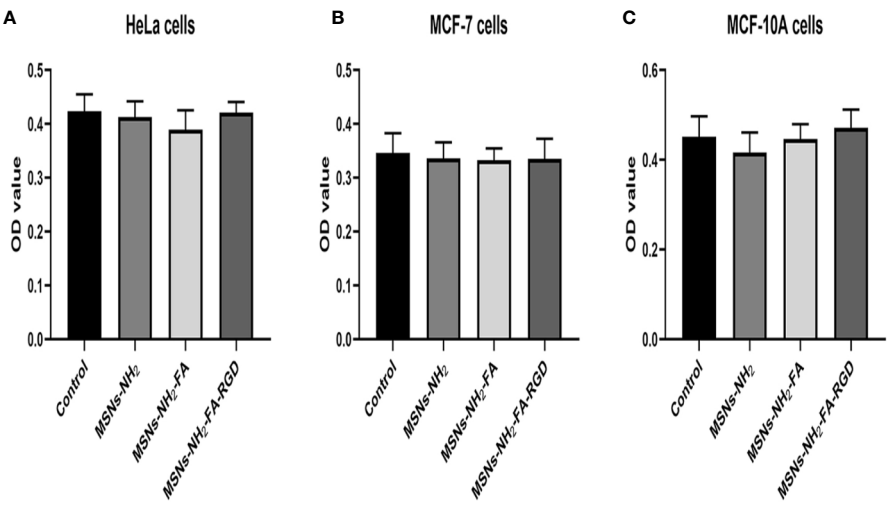


FIGURE 6 | Cell viability of HeLa **(A)**, MCF-7 **(B)** and MCF-10A **(C)** cells after 24 h incubation with MSNs-NH_2, MSNs-NH_2-FA and MSNs- NH_2-FA-RGD at 20 μg·ml^{-1}.

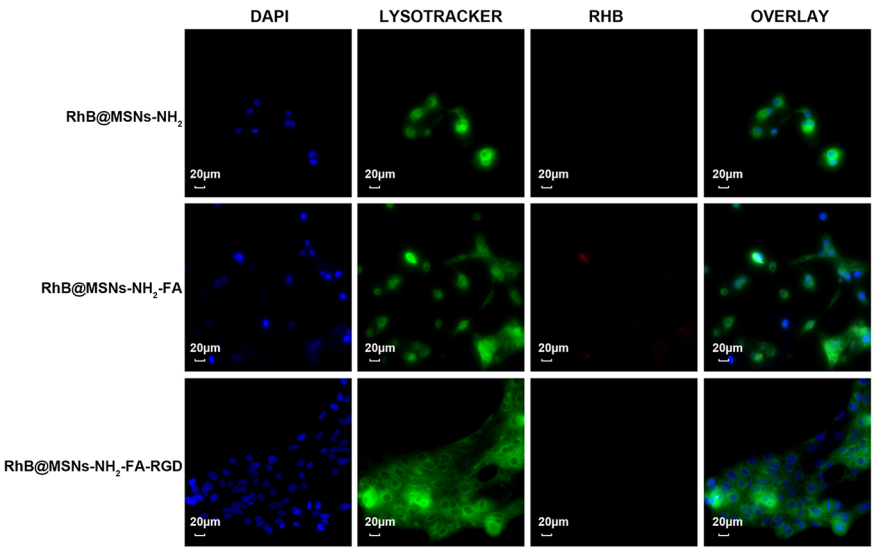


FIGURE 7 | Fluorescence microscopy images of MCF-10A cells incubation with RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, and RhB@MSNs-NH_2-FA-RGD for 4 h. Blue fluorescence field: nucleus; green fluorescence field: cytoplasm; red fluorescence field: a dye used to label nanocarriers. Scale bar: 20 μm.

inhibitory effects of PTX@MSNs-NH_2-FA-RGD and free PTX on MCF-7 cells did not show significant difference ($P > 0.05$). But after 48 h, it could be clearly seen that PTX@MSNs-NH_2-FA-RGD had stronger inhibitory effect on MCF-7 cells. The IC_{50} of free PTX and PTX@MSNs-NH_2-FA-RGD on MCF-7 cells line treated for 48 h were 35.25 ± 2.57 ng·ml^{-1} and 22.21 ± 3.4 ng·ml^{-1}, respectively, which indicated that the inhibitory efficacy of PTX@MSNs-NH_2-FA-RGD on MCF-7 was 1.6 times than that of free PTX.

CONCLUSION

Based on the verification of a series of chemical and cell experiments, we synthesized FA and RGD dual-targeted nanocarrier MSNs-NH_2-FA-RGD. RhB@MSNs-NH_2-FA-RGD exhibited excellent fluorescence property *in vitro*. The fluorescence microscopy experiment illustrated that RhB@MSNs-NH_2-FA-RGD had a higher cellular uptake by MCF-7 cells and HeLa cells than MCF-10A cells via receptor-mediated

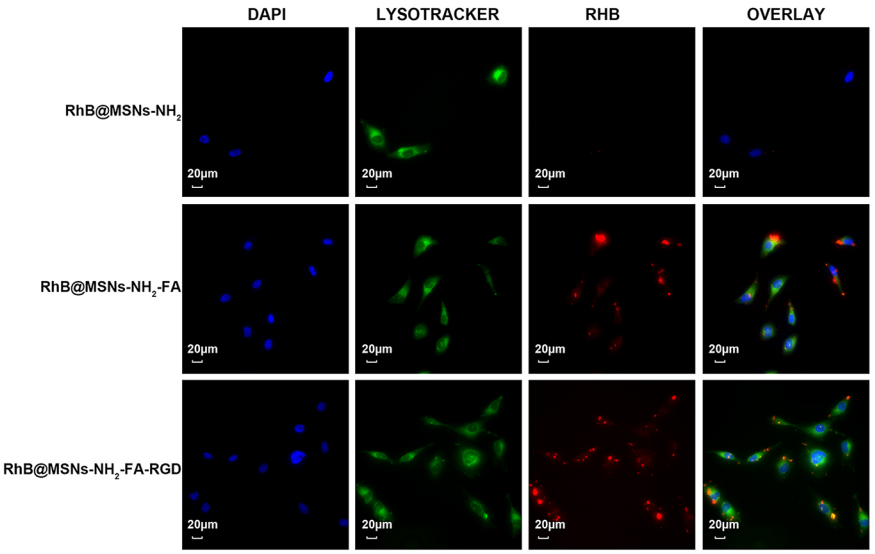


FIGURE 8 | Fluorescence microscopy images of HeLa cells incubation with RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, and RhB@MSNs-NH_2-FA-RGD for 4 h. Blue fluorescence field: nucleus; green fluorescence field: cytoplasm; red fluorescence field: a dye used to label nanocarriers. Scale bar: 20 μm.

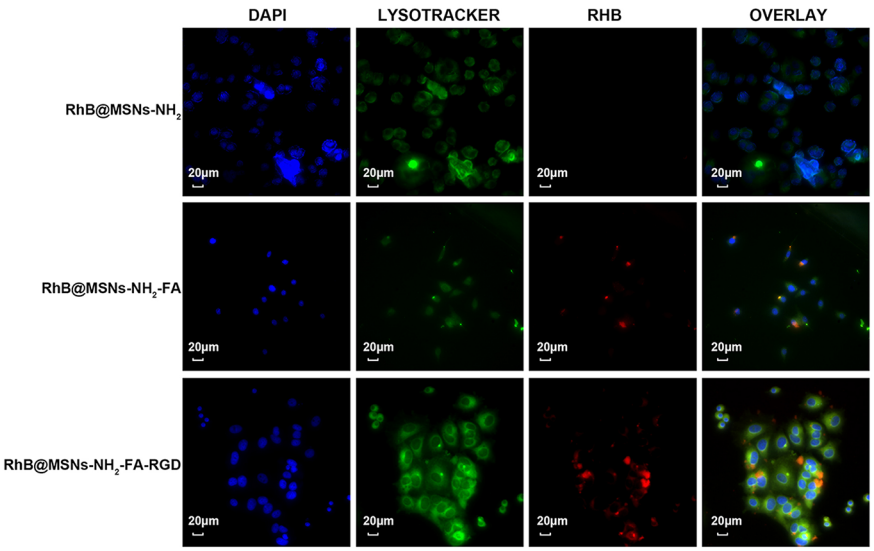


FIGURE 9 | Fluorescence microscopy images of MCF-7 cells incubation with RhB@MSNs-NH_2, RhB@MSNs-NH_2-FA, and RhB@MSNs-NH_2-FA-RGD for 4 h. Blue fluorescence field: nucleus; green fluorescence field: cytoplasm; red fluorescence field: a dye used to label nanocarriers. Scale bar: 20 μm.

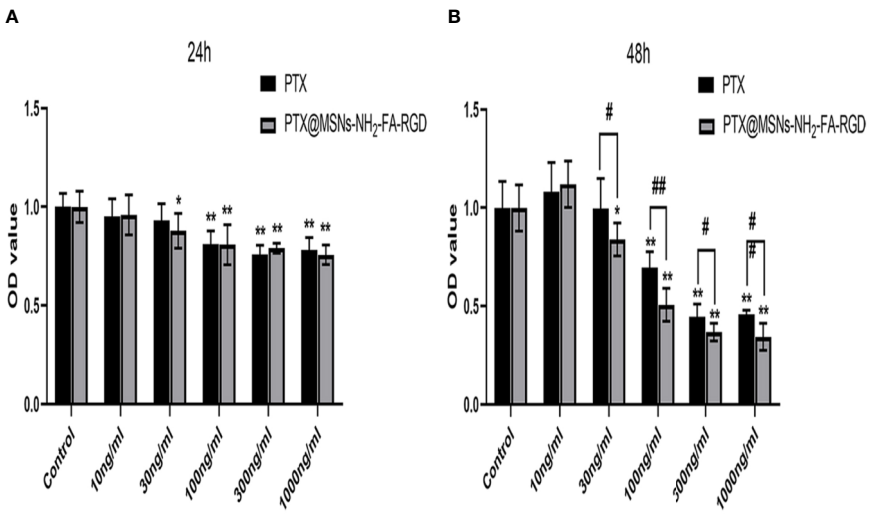


FIGURE 10 | Cytotoxicity for free PTX and PTX@MSNs-NH_2-FA-RGD at different concentrations ($ng{\cdot}ml^{-1}$) against MCF-7 cells evaluated at 24 h **(A)** and 48 h **(B)**. ($^{**}p < 0.01$, $^{*}p < 0.05$ as compared with the data of control group; $^{\#\#}p < 0.01$, $^{\#}p < 0.05$ when PTX group compared with PTX@MSNs-NH_2-FA-RGD at the same concentration).

endocytosis. In addition, The IC_{50} of free PTX and PTX@MSNs-NH_2-FA-RGD for 48h were 35.25 ± 2.57 $ng{\cdot}ml^{-1}$ and 22.21 ± 3.4 $ng{\cdot}ml^{-1}$, respectively. The killing capacity of PTX@MSNs-NH_2-FA-RGD to MCF-7 cells was 1.6 times than that of free PTX, indicating that PTX@MSNs-NH_2-FA-RGD had higher antitumor activity. The animal experiments are under going. These results indicated that MSNs-NH_2-FA-RGD could target breast cancer cells, and sever as a potential candidate in future of drug development.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HY participated in the literature search, study design, surgery operation, data collection, data analysis, data interpretation, and

writing of the manuscript. YY and XL carried out the data collection and analysis and provided the critical revision of the manuscript. LL, FG, QZ, and YT conceived the study and participated in its design and coordination. All authors read and approved the final manuscript. JW and SD participated in the study design and provided critical revision. All authors contributed to the article and approved the submitted version.

FUNDING

The work was supported by the Major National Science and Technology Program of China for Innovative Drug (2017ZX09101002-001-001) and the Fundamental Research Funds for the Central public welfare research institutes (ZXKT17025).

REFERENCES

Alejandro, B., Montserrat, C., and María, V. R. (2014). Advances in mesoporous silica nanoparticles for targeted stimuli-responsive drug delivery. *Expert Opin. Drug Del.* 12 (2), 319. doi: 10.1517/17425247.2014.953051

Darvishi, B., and Farahmand, L. (2017). Majidzadeh-A K. Stimuli-Responsive Mesoporous Silica NPs as Non-viral Dual siRNA/Chemotherapy Carriers for Triple Negative Breast Cancer. *Mol. Ther-Nucl. Acids* 7, 164–180. doi: 10.1016/j.omtn.2017.03.007

He, D., He, X., Wang, K., Chen, M., Jie, C., and Zhao, Y. (2012). Reversible stimuli-responsive controlled release using mesoporous silica nanoparticles functionalized with a smart DNA molecule-gated switch. *J. Mater. Chem.* 22 (29), 14715–14721. doi: 10.1039/c2jm32185h

Hee Dong, H., Mangala, L. S., Jeong Won, L., Mian, M. K. S., Hye Sun, K., Deyu, S., et al. (2010). Targeted gene silencing using RGD-labeled chitosan nanoparticles. *Clin. Cancer Res.* 16 (15), 3910–3922. doi: 10.1158/1078-0432.CCR-10-0005

Kazuki, M., Masahiro, H., Takatoshi, K., Kenji, N., and Katsuya, K. (2013). Design of a nanocarrier with regulated drug release ability utilizing a reversible conformational transition of a peptide, responsive to slight changes in pH. *Phys. Chem. Chem. Phys.* 15 (27), 11454–11460. doi: 10.1039/c3cp50916h

Lv, F., Cao, J., Zhang, J., Qian, J., Peng, W., Sun, S., et al. (2014). Phase I and pharmacokinetic study of polymeric micelle-formulated paclitaxel in adult Chinese patients with advanced solid tumors. *Cancer Chemother. Pharmacol.* 73 (6), 1173–1179. doi: 10.1007/s00280-014-2452-6

Martinez-Carmona, M., Lozano, D., Colilla, M., and Vallet-Regi, M. (2018). Lectin-conjugated pH-responsive mesoporous silica nanoparticles for targeted bone cancer treatment. *Acta Biomater.* 65, 393–404. doi: 10.1016/j.actbio.2017.11.007

Porta, F., Lamers, G. E., Morrhayim, J., Chatzopoulou, A., Schaaf, M., Dulk, H. d., et al. (2013). Folic acid-modified mesoporous silica nanoparticles for cellular and nuclear targeted drug delivery. *Adv. Healthc. Mater.* 2 (2), 281–286. doi: 10.1002/adhm.201200176

Pourjavadi, A., and Tehrani, Z. M. (2014). Mesoporous Silica Nanoparticles (MCM-41) Coated PEGylated Chitosan as a pH-Responsive Nanocarrier for Triggered Release of Erythromycin. *Int. J. Polym. Mater. Polym. Biomater.* 63 (13), 692–697. doi: 10.1080/00914037.2013.862534

Pourjavadi, A., and Tehrani, Z. M. (2016). Mesoporous silica nanoparticles with bilayer coating of poly(acrylic acid-co-itaconic acid) and human serum albumin (HSA): A pH-sensitive carrier for gemcitabine delivery. *Mat. Sci. Eng. C-Mater.* 61, 782–790. doi: 10.1016/j.msec.2015.12.096

Raviv, Z. (2004). MEK5 and ERK5 are localized in the nuclei of resting as well as stimulated cells, while MEKK2 translocates from the cytosol to the nucleus upon stimulation. *J. Cell Sci.* 117 (9), 1773–1784. doi: 10.1242/jcs.01040

Ren, Y., Wei, Y., and Zhao, F. (2005). Application of paclitaxel in glaucoma filtration surgery. *J. Diffcult Complicated cases*. 2005 (06), 328–330.

Shen, Z., Li, Y., Kohama, K., Oneill, B., and Bi, J. (2011). Improved drug targeting of cancer cells by utilizing actively targetable folic acid-conjugated albumin nanospheres. *Pharmacol. Res.* 63 (1), 51–58. doi: 10.1016/j.phrs.2010.10.012

Song, S. W., Hidajat, K., and Kawi, S. (2007). pH-controllable drug release using hydrogel encapsulated mesoporous silica. *Chem. Commun.* 2007 (42), 4396–4398. doi: 10.1039/b707626f

Tao, L., Song, C., Huo, C., Sun, Y., Zhang, C., Li, X., et al. (2016). Anti-CD155 and anti-CD112 monoclonal antibodies conjugated to a fluorescent mesoporous silica nanosensor encapsulating rhodamine 6G and fluorescein for sensitive detection of liver cancer cells. *Analyst* 141 (16), 4933–4940. doi: 10.1039/C5AN01908G

Tian, Y., Guo, R., Jiao, Y., Sun, Y., Shen, S., Wang, Y., et al. (2016). Redox stimuli-responsive hollow mesoporous silica nanocarriers for targeted drug delivery in cancer therapy. *Nanoscale Horiz.* 1 (6), 480–487. doi: 10.1039/C6NH00139D

Xue, M., Zhong, X., Shaposhnik, Z., Qu, Y., Tamanoi, F., Duan, X., et al. (2011). pH-Operated Mechanized Porous Silicon Nanoparticles. *J. Am. Chem. Soc.* 133 (23), 8798–8801. doi: 10.1021/ja201252e

Zhang, Y., Guo, J., Zhang, X. L., Li, D. P., Zhang, T. T., Gao, F. F., et al. (2015). Antibody fragment-armed mesoporous silica nanoparticles for the targeted delivery of bevacizumab in ovarian cancer cells. *Int. J. Pharm.* 496 (2), 1026–1033. doi: 10.1016/j.ijpharm.2015.10.080

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yan, You, Li, Liu, Guo, Zhang, Liu, Tong, Ding and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fonc.2020.01282

Molecular Screening for Nigericin Treatment in Pancreatic Cancer by High-Throughput RNA Sequencing

Zhihua Xu[1†], Guanzhuang Gao[1†], Fei Liu[2], Ye Han[1], Chen Dai[1], Sentai Wang[1], Guobang Wei[1], Yuting Kuang[1], Daiwei Wan[1], Qiaoming Zhi[1,3*] and Ye Xu[3*]*

[1] Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China, [2] Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou, China, [3] Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Shanghai, China

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Liang Cheng,
Harbin Medical University, China
Shikha Kumari,
University of Nebraska Medical Center, United States

***Correspondence:**
Daiwei Wan
dv19880114@163.com
Qiaoming Zhi
strexboy@163.com
Ye Xu
xu_shirley021@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Oncology

Received: 20 March 2020
Accepted: 19 June 2020
Published: 31 July 2020

Citation:
Xu Z, Gao G, Liu F, Han Y, Dai C, Wang S, Wei G, Kuang Y, Wan D, Zhi Q and Xu Y (2020) Molecular Screening for Nigericin Treatment in Pancreatic Cancer by High-Throughput RNA Sequencing. Front. Oncol. 10:1282. doi: 10.3389/fonc.2020.01282

Objectives: Nigericin, an antibiotic derived from *Streptomyces hygroscopicus*, has been proved to exhibit promising anti-cancer effects on a variety of cancers. Our previous study investigated the potential anti-cancer properties in pancreatic cancer (PC), and demonstrated that nigericin could inhibit the cell viabilities in concentration- and time-dependent manners via differentially expressed circular RNAs (circRNAs). However, the knowledge of nigericin associated with long non-coding RNA (lncRNA) and mRNA in pancreatic cancer (PC) has not been studied. This study is to elucidate the underlying mechanism from the perspective of lncRNA and mRNA.

Methods: The continuously varying molecules (lncRNAs and mRNAs) were comprehensively screened by high-throughput RNA sequencing.

Results: Our data showed that 76 lncRNAs and 172 mRNAs were common differentially expressed in the nigericin anti-cancer process. Subsequently, the bioinformatics analyses, including Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, coding and non-coding co-expression network, cis- and trans-regulation predictions and protein-protein interaction (PPI) network, were applied to annotate the potential regulatory mechanisms among these coding and non-coding RNAs during the nigericin anti-cancer process.

Conclusions: These findings provided new insight into the molecular mechanism of nigericin toward cancer cells, and suggested a possible clinical application in PC.

Keywords: nigericin, high-throughput sequencing, long non-coding RNA, bioinformatics, pancreatic cancer

INTRODUCTION

Ductal adenocarcinoma of the exocrine pancreas, commonly known as pancreatic cancer (PC), is a highly aggressive malignancy with few effective therapies. At the time of diagnosis, ~20% of PC patients are considered eligible for surgery and of these, about a half undergoes successful resection (1). But unfortunately, a majority of patients with PC are diagnosed at advanced stages, at which patients can hardly receive surgical R0 resection (2) with a 5-years survival rate of 3% and a median survival of <6 months (3–5). In spite of significant advances in surgical care, chemotherapy and radiotherapy, no effective systemic therapy for the aggressive pathology of this cancer is available.

One of the reasons for the treatment failures is due to resistance to chemotherapy or radiotherapy (6). Thus, novel therapeutic agents are needed to treat PC and improve the overall survival of patients with this disease.

Up to date, standard treatments for cancer involve chemotherapy with anti-tumor antibiotic. Adriamycin, an orally administered DNA alkylating agent, has been the most potent chemotherapy applied in clinic, in addition to surgical excision. Bleomycin had been emerged as another vital chemotherapeutic agent in many types of cancer, including Hodgkin lymphoma (7), testicular cancer (8), and squamous-cell carcinoma (9). In the previous study, we firstly identified salinomycin as a novel identified cancer stem cells (CSCs) killer in gastric cancer cells (10). Afterwards, we also found that salinomycin could specifically target on cisplatin-resistant colorectal cancer cells by accumulating reactive oxygen species (11). Recently, Moxifloxacin and ciprofloxacin induced cell apoptosis and S-phase arrest via ERK activation in PC (12). Similar anti-cancer influences of antibiotic on PC cells were found also in recent literatures. In 2012, Yadav et al. demonstrated that gatifloxacin possessed anti-proliferative activity against PC cell lines by causing S/G2 phase cell cycle arrest without induction of apoptosis through p21, p27, and p53 dependent pathway (13). They also investigated the effect of moxifloxacin and ciprofloxacin on survival and proliferation of PC cell lines, and found that both were able to suppress the proliferation of PC cells and induce apoptosis through the similar mechanism (12).

Nigericin is a monocarboxylic polyether antibiotic potassium ionophore that is widely used as a coccidiostatic agent in chickens (14). In 1972, the effects of nigericin on intracellular pH, glycolysis, and K^+ concentration of ehrlich ascites tumor cells were firstly reported (15). Since then, emerging evidence confirmed the promising anti-cancer activity of nigericin in a variety of cancers, including prostate cancer (16), nasopharyngeal carcinoma (17), bladder cancer (18), chronic lymphocytic leukemia cells (19), and lung cancer (20). In 2004, Vaupel et al. reported that nigericin could inhibit breast cancer stem cells at least 100 times more effectively than paclitaxel in mice (21). Zhou et al. demonstrated that nigericin could suppress the colorectal cancer metastasis through inhibition of epithelial-mesenchymal transition (EMT) (22). Recently, our study explored the circular RNA (circRNA) expression profiles after nigericin exposure on PC cells through bioinformatics method, and discussed the potential function of nigericin in PC (23). However, our knowledge of nigericin, which correlates with long non-coding RNA (lncRNA) and mRNA in PC, has not been studied yet.

In this study, we attempted to ascertain the specific activities of nigericin on human PC cell lines and investigate its possible molecular mechanism in PC. The continuously varying molecules (lncRNAs and mRNAs) were displayed by the high-throughput sequencing. Through analyzing the aberrant expressions of lncRNAs and mRNAs as well as their potential relationships, the molecular mechanisms of nigericin treatment on PC were discussed.

MATERIALS AND METHODS

Cell Culture and Reagents

Human PC cell lines (PANC-1) were purchased from Shanghai Institute of Biochemistry and Cell Biology at the Chinese Academy of Sciences (Shanghai, China). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco) at 37°C in a humidified incubator containing 5% CO_2. Cells were in the logarithmic phase of growth for all experiments. Nigericin was purchased from Sigma Aldrich (USA). The stock solutions (100 mmol/L) were prepared with dimethyl sulfoxide (DMSO) and stored at −20°C.

High-Throughput RNA Sequencing Analysis

PANC-1 cells were exposed to a proper concentration of nigericin (5 μmol/L) according to the results of 50% inhibitory concentration (IC50) for different time periods (0, 8, 16, or 32 h), and then total RNA was extracted from cells, respectively. The quantity and integrity of total RNAs were measured by the NanoDrop™ ND-2000 (Thermo Fisher Scientifc, Scotts Valley, CA, USA) and Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), respectively. lncRNAs and mRNAs were quantitatively analyzed by Shanghai OE Biotech (Shanghai, China). After removal of ribosomal RNA and then constructing a library, a high-throughput RNA sequencing was performed. The clean reads were aligned to the reference genome by Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml). For unmapped reads, the junctions were picked out using back-splice algorithm. Finally, lncRNAs and mRNAs were verified with software developed by Shanghai OE Biotech, which were considered as the reference sequence for further analysis.

Differentially Expressed lncRNA and mRNA Screen and Clustering Analysis

Differentially expressed lncRNAs and mRNAs were detected by the negative binomial distribution test based on the DESeq package. These lncRNAs and mRNAs with statistical significance were screened with $p < 0.05$, false discovery rate (FDR) <0.05 and fold change (FC) more than 2.0. Difference integration analysis (Venn analysis) was used to show the often characteristic elements among these 3 compared groups (0 vs. 8 h, 0 vs. 16 h, 0 vs. 32 h). The common differentially expressed lncRNAs and mRNAs were showed in pies with different colors. The non-supervised hierarchical clustering of the differentially expressed lncRNAs and mRNAs was used in the form of heat map to display the expression patterns of the differential lncRNAs and mRNAs between different groups.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) Validation

Total RNA from cell lines was extracted using Trizol solution (Invitrogen, USA) and converted into cDNA by using M-MLV reverse transcriptase (Invitrogen, USA). The quantities and qualities of isolated RNAs were evaluated using absorbance

measurements at 260 and 280 nm. Then reverse transcription (RT) was performed in a 20 μl reaction system using the ReverAid First Stand cDNA Synthesis (Thermo Scientific, Mountain View, CA, USA). RT-PCR with SybGreen I (Generay Bio Co., Shanghai, China) was performed using the 7500 real-time PCR system (Applied Biosystems, Hayward, CA, USA) with the follow program: initial denature at 95°C for 10 min, followed by 40 cycles of 95°C for 10 s and 60°C for 60 s. β-actin was used as control. Results were harvested in three independent wells. The sequences of primers were listed as follows: LINC00667:6 (F: 5′CCCGACTTTTTGATGCAGGC3′; R: 5′CCCGACTGTTTCCTACCCAC3′), Lnc-HMGN1-1:12 (F: 5′GATCATGGCTCTCTCTGCCA3′; R: 5′AGCTGTTACATA CGGCCCAC3′), Lnc-LRRC24-2:1 (F: 5′GATTCGCTGGAC GATCGCA3′; R: 5′CCTGTAAAGGGAACGCGTCA3′), Lnc-AC007952.1.1-3:1 (F: 5′GCGAGAAAGGTTTTCGCCTC3′; R: 5′ACAATAGGAGGTGCCACACA3′), Lnc-CCNB1IP1-1:2 (F: 5′TGTCCCTTGGGAAGGTCTGA3′; R: 5′CCCG TTCTCTGGGAACTCAC3′), GADD45A (F: 5′GAGAGCAGAA GACCGAAAGGA3′; R: 5′CACAACACCACGTTATCGGG3′), HBP1 (F: 5′TCATCACCATTGGAAGGAGGA3′; R: 5′TTGCAC CATCCCAAATCATCA3′), SESN2 (F: 5′AAGGACTACC TGCGGTTCG3′; R: 5′CGCCCAGAGGACATCAGTG3′), KIF20A (F: 5′TTGAGGGTTAGGCCCTTGTTA3′; R: 5′GTCCTTGGGTGCTTGTAGAAC3′), TOP2A (F: 5′ACCAT TGCAGCCTGTAAATGA3′; R: 5′GGGCGGAGCAAAATAT GTTCC3′), and β-actin (F: 5′CCTGTACGCCAACACAGTGC3′; R: 5′ATACTCCTGCTTGCTGATCC3′).

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Analysis

GO analysis was conducted to construct meaningful annotations of genes and gene products in a wide variety of organisms through DAVID database (http://david.abcc.ncifcrf.gov). Our GO analysis provided the ontology of defined terms which represented gene product properties, and covered three domains: cellular components, biological process and molecular function. The top 10 enriched GO terms, which were derived from the common differentially expressed mRNAs and ranked by enrichment score, were presented. KEGG pathway analysis was also adopted to map differentially expressed mRNAs in different biological pathways. The top 20 enriched pathways among the four groups ranked by enrichment score were calculated and shown.

lncRNA-mRNA Co-expression Network

To elucidate the potential functions of differentially expressed lncRNAs and explore the relationships between common differentially expressed lncRNAs and mRNAs, the lncRNA-mRNA co-expression network was constructed. For each differentially expressed lncRNA, we calculated the Pearson Correlation of its expression value with the expression value of each differentially expressed mRNA. It was considered to be correlated when the *P*-value of the correlation coefficient of lncRNAs and mRNAs' expression value was not higher than 0.05, and the absolute value of correlation was not <0.7. A total of 66 lncRNAs and mRNAs were selected to generate the network map.

Cis- and Trans-Regulation Predictions

As previous studies defined, a cis-regulator is the one that exerts its functions on the neighboring genes which were located at the same chromosome. lncRNAs are showed that they can regulate gene expressions in a cis-manner (24, 25). The cis-regulation regions in this study were identified by the following procedures. For each common differentially expressed lncRNA, we identified the mRNAs as "cis-regulated mRNAs" when: (1) the mRNAs loci were within 100 k windows up- and downstream of the given lncRNA. (2) the Pearson Correlation of lncRNA-mRNA expression was statistically significant (p-value of correlation ≤ 0.05).

For trans-regulation prediction, we focused on the manner that lncRNAs played their functions via transcription factors (TFs). The TF-lncRNA and TF-lncRNA-gene network were constructed, respectively. For each differentially expressed lncRNAs, the coding genes co-expressed with them were calculated, and the significance of the gene enrichment in each TF entry was calculated using the hypergeometric distribution test method. The result of the calculation returned a p-value that was enriched for significance. A small p-value indicated that gene has been enriched in the TF entry. We calculated the intersection of lncRNAs co-expressed gene sets with target gene sets of transcription factor/chromatin regulated complex, and calculated the degree of enrichment of the intersection through hypergeometric distribution method. Then we obtained the TFs which were significantly associated with lncRNAs, and identified possible transcription factor/chromatin regulated factors that might play a combined regulatory role with lncRNAs. Subsequently, we used the analysis results of hypergeometric distribution to visualize the network diagram. Through the hypegeometric distribution calculation, each lncRNA got multiple TF-lncRNA relationship pairs, and each TF-lncRNA pair was the results of enrichments of multiple genes. According to the p-value from small to large sort, the top 200 lines of regulatory relationships were used to construct the TF-lncRNA binary relationship network, and the top 10 lines of regulatory relationships were applied to construct the TF-lncRNA-gene ternary relationship network.

Protein-Protein Interaction (PPI) Network Construction

The Search Tool for the Retrieval of Interacting Genes (STRING, http://string.embl.de/) database was used to construct the PPI network of the common differentially expressed mRNAs. The PPI network was subsequently visualized using Cytoscape. Confidence score ≥ 0.7 was set as the cut-off criterion, and Molecular Complex Detection (MCODE) was conducted to screen modules of PPI network with degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. depth = 100. In addition, a sub-network was constructed by selecting several candidate mRNAs.

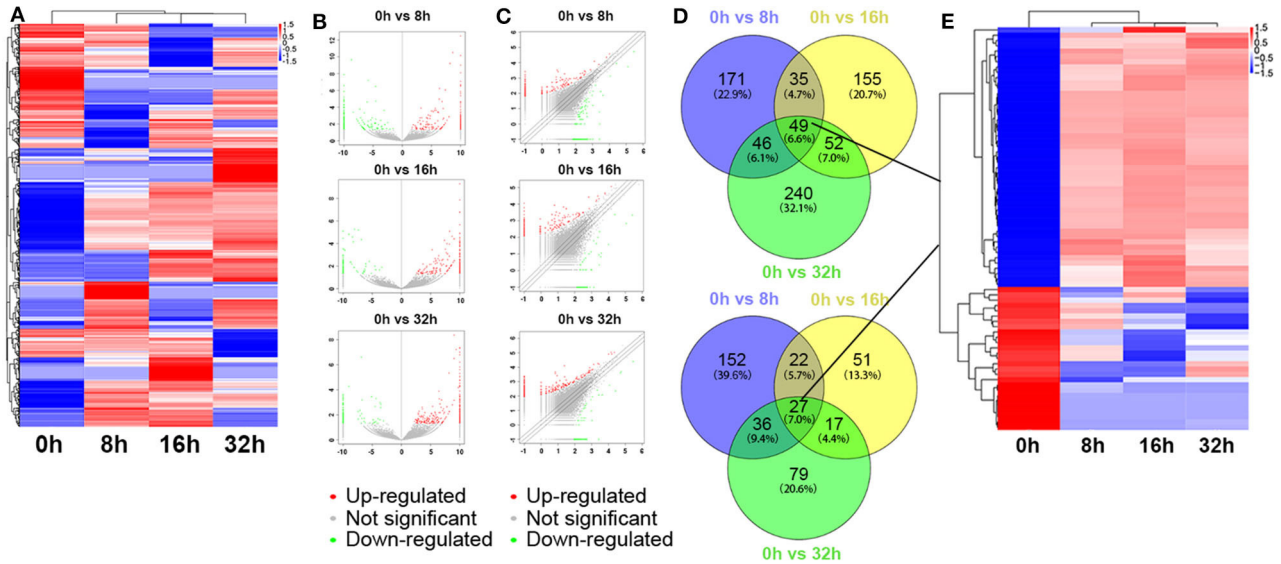


FIGURE 1 | Differentially expressed lncRNA profile by sequencing. **(A)** The cluster heatmap showed all differentially expressed lncRNAs at different time points (0, 8, 16, and 32 h). **(B,C)** The volcano and scatter plots presented differentially expressed lncRNAs between different compared groups (0 h vs. 8 h, 0 h vs. 16 h, 0 h vs. 32 h), respectively. **(D,E)** Venn analysis analyzed the common differentially expressed lncRNAs among the three compared groups, and described them as a cluster heatmap.

Statistical Analysis

Statistically significant differences among groups were estimated by the Student's *t*-test using SPSS 19.0 software (SPSS Inc.). $P < 0.05$ was considered to be statistically significant.

RESULTS

Differentially Expressed lncRNA and mRNA Profile by Sequencing

The global expression profile of lncRNAs at 4 different time points (0, 8, 16, and 32 h) was determined by a custom sequencing platform. In total, 118,314 lncRNAs were detected, and hundreds of lncRNAs showed differential expressions in each group of different time points (**Figure 1A**). Three compared groups were set according to the nigericin-treated time points (0 vs. 8 h, 0 vs. 16 h, 0 vs. 32 h). Compared to the 0 h group, 538 days-regulated lncRNAs (more than 2 folds) were found in 8 h, of which 301 lncRNAs were up-regulated and 237 ones were down-regulated. Similarly, 408 lncRNAs were differentially expressed in 16 h group with 291 up-regulated and 117 down-regulated ones, compared to the 0 h group. With the change of treatment time at 32 h, 387 differential lncRNAs were up-regulated, and 159 ones were down-regulated. All differentially expressed lncRNAs with statistical significance were selected with $p < 0.05$, FDR < 0.05, and FC > 2.0 (**Figures 1B,C**). Venn analysis was used to determine the common differentially expressed lncRNAs among the three compared groups. Our data confirmed that 76 common dys-regulated lncRNAs including 49 up-regulated and 27 down-regulated ones might participate in the process of nigericin damage (**Figures 1D,E**). To systematically predict the function of lncRNAs, lncRNA subgroup analyses were performed. These lncRNAs were widely distributed on all chromosomes except for sex chromosome X (**Figure 2A**). Moreover, we adapted specific probes for these lncRNAs to classify several kinds of lncRNAs. Among these dys-regulated lncRNAs, there were 71.1% sense-overlapping, 23.7% intergenic, 1.3% intronic, 1.3% bidirectional, 1.3% antisense, and 1.3% undefined (**Figure 2B**).

Besides, the global expression profile of mRNAs was also observed. Among the 98,121 coding transcripts examined, hundreds of coding transcripts were differentially expressed at different time points (**Figure 3A**). Three compared groups were also set as described above. A total of 2,468 mRNAs exhibited significant differential expression between the 0 and 8 h group, in which 1,282 mRNAs were up-regulated and 1,186 ones were down-regulated. For 0 vs. 16 h group, 616 mRNAs were up-regulated and 1,000 ones were down-regulated. Similarly, for 0 vs. 32 h group, 918 mRNAs were up-regulated and 1,082 ones were down-regulated (**Figures 3B,C**). All the differentially expressed mRNAs with statistical significance were screened with $p < 0.05$, FDR < 0.05, and FC > 2.0. Venn analysis was also used to determine the common differentially expressed mRNAs among the three compared groups (0 vs. 8 h, 0 vs. 16 h, and 0 vs. 32 h). The results showed that 172 mRNAs were common up-regulated and 85 ones were down-regulated (**Figure 3D**). As shown in **Figure 3E**, the common dys-regulated mRNAs were evaluated by the hierarchical clustering analysis.

Validation of Sequencing Data by qRT-PCR

To ensure that our results were reliable, we assessed the expressions of 5 lncRNAs and 5 mRNAs between the 0 and 32 h group by qRT-PCR. Our results showed that the LINC00667:6 and lnc-HMGN1-1:12 were both up-regulated at 32 h, compared

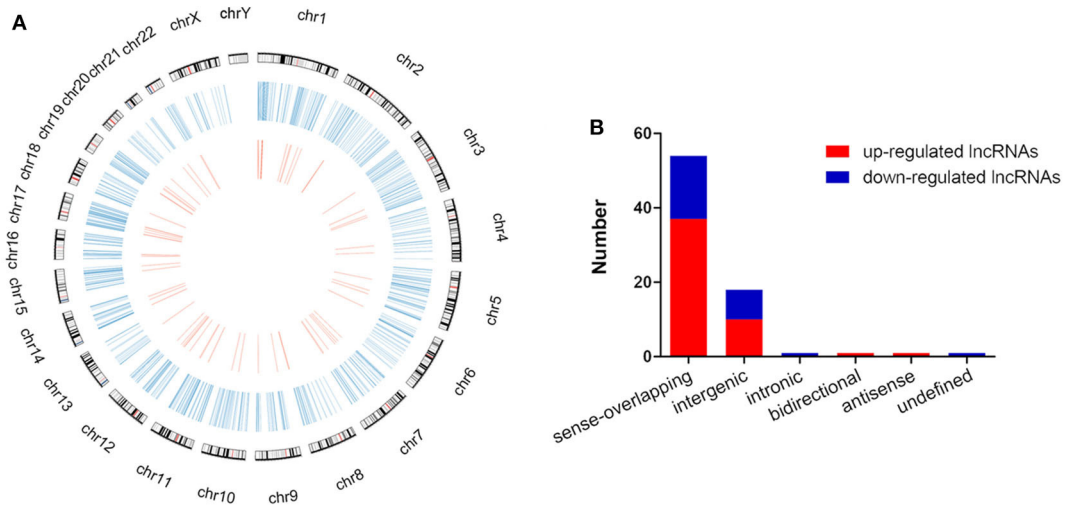


FIGURE 2 | Distribution and classification of the common differentially expressed lncRNAs. **(A)** The circos plot showed the distribution of lncRNAs on human chromosomes. The outermost layer of circos plot was chromosome map of the human genome (the black and white bars were chromosome cytobands, the red bars represented centromeres). The larger inner circle (blue) represented all differentially expressed lncRNAs at different time points, and the smaller inner circle (red) indicated the 76 common differentially expressed lncRNAs with fold change > 2.0, $p < 0.05$ and FDR < 0.05. **(B)** These 76 common differentially expressed lncRNAs were classified into 6 types, including sense-overlapping, intergenic, intronic, bidirectional, antisense and undefined.

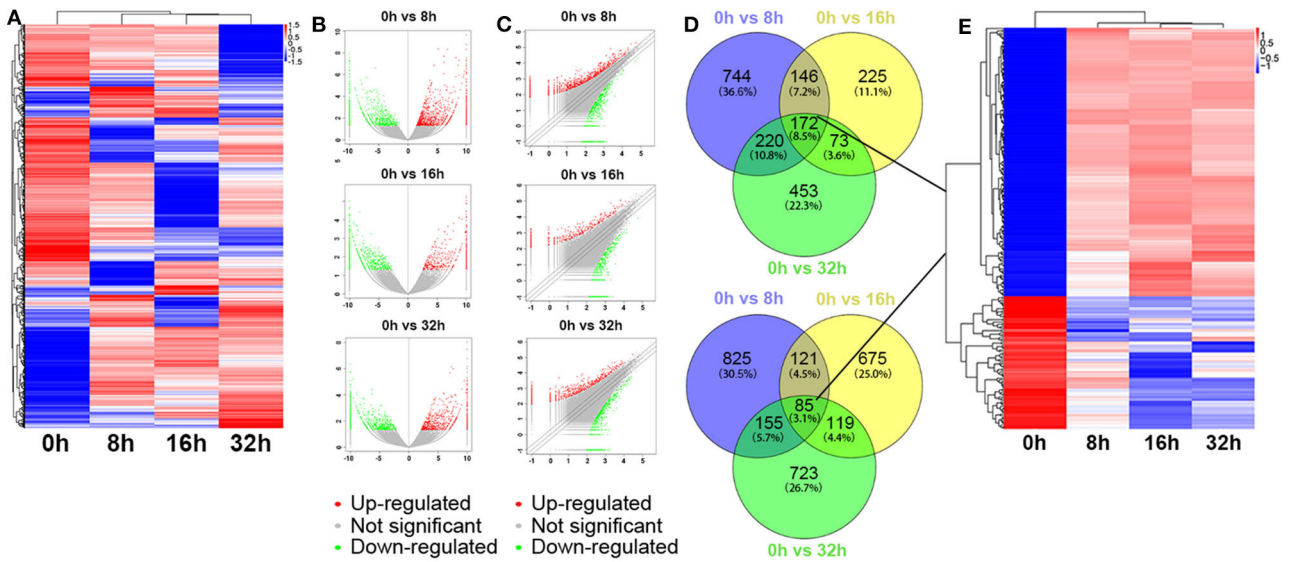


FIGURE 3 | Differentially expressed mRNA profile by sequencing. **(A)** The cluster heatmap showed all differentially expressed mRNAs at different time points (0, 8, 16, and 32 h). **(B,C)** The volcano and scatter plots presented differentially expressed mRNAs between different compared groups (0 vs. 8 h, 0 vs. 16 h, 0 vs. 32 h), respectively. **(D,E)** Venn analysis analyzed the common differentially expressed mRNAs among the three compared groups, and described them as a cluster heatmap.

to the 0 h group, whereas other 3 lncRNAs (lnc-LRRC24-2:1, lnc-AC007952.1.1-3:1, and lnc-CCNB1IP1-1:2) were down-regulated (**Figure 4A**, $^*P < 0.05$). Meanwhile, we also chose 5 random cancer-related genes for mRNA detection, and found that GADD45A, HBP1, and SESN2 were significantly up-regulated, whereas KIF20A and TOP2A were down-regulated, compared to the 0 h group (**Figure 4B**, $^*P < 0.05$). These data were consistent well with the sequencing data, which demonstrated the high reliability and validity of the sequencing expression results.

GO and KEGG Pathway Analysis

According to the common differentially expressed mRNAs among the three compared groups (0 vs. 8 h, 0 vs. 16 h, 0 vs. 32 h), the GO biological processes classification was

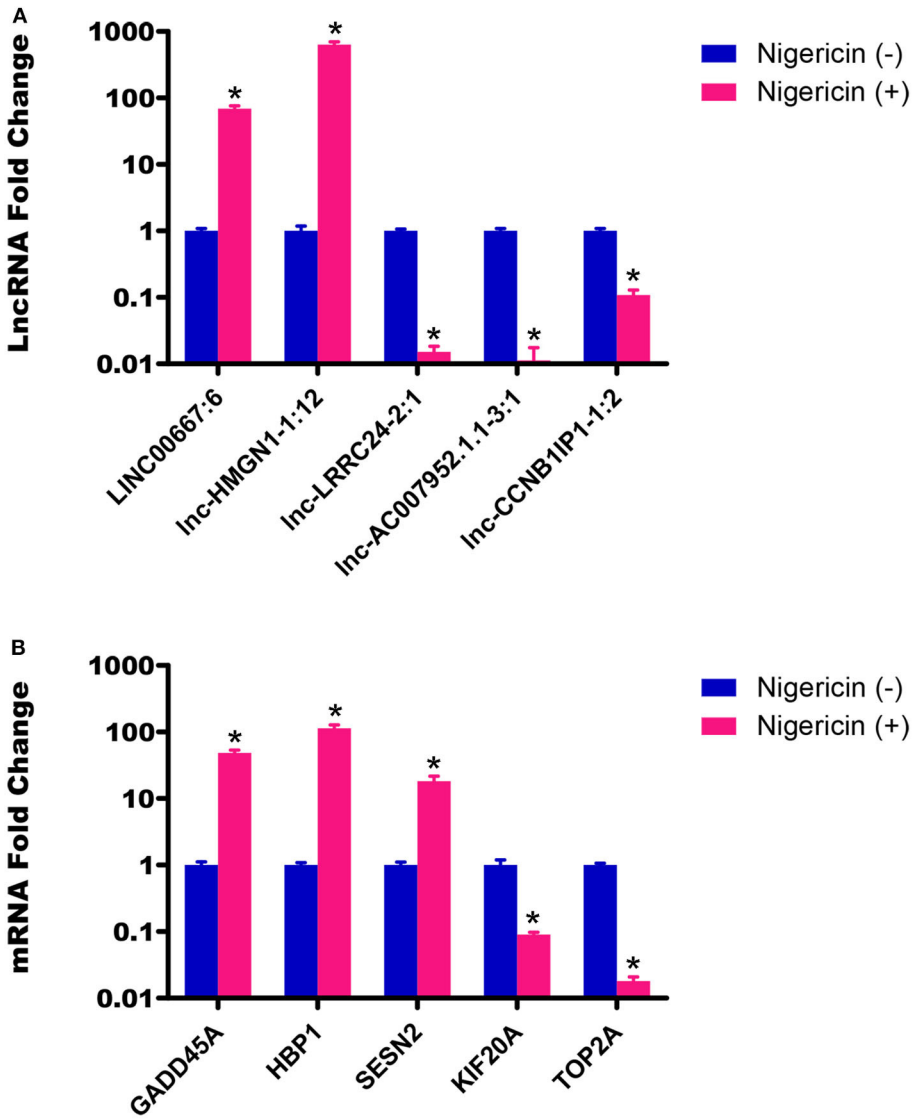
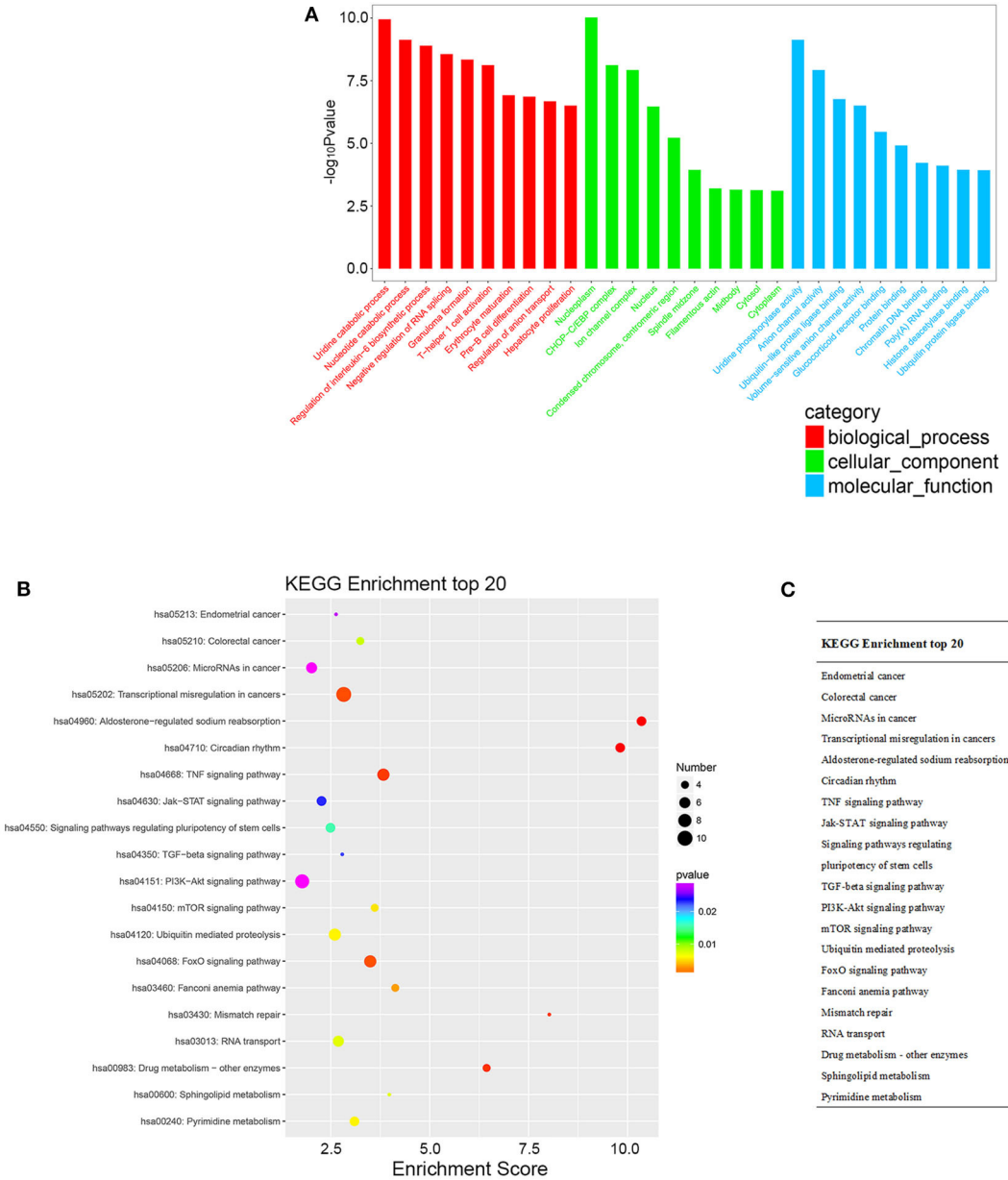


FIGURE 4 | Validation of sequencing data by qRT-PCR. **(A)** The expressions of 5 lncRNAs between the 0 and 32 h group were detected by qRT-PCR. **(B)** The expressions of 5 mRNAs between the 0 and 32 h group were also determined by qRT-PCR (*$P < 0.05$).

calculated. The top 10 GO biological processes such as uridine catabolic process, nucleotide catabolic process and regulation of interleukin-6 biosynthetic process were involved in the nigericin damage. Meanwhile, the top 10 cellular components and molecular functions were also analyzed and presented in **Figure 5A**. KEGG pathway analysis for the common differentially expressed mRNAs was used to elucidate the pathways related to these mRNAs. Our data showed that differentially expressed mRNAs were significantly enriched in top 20 KEGG signaling pathways, including Aldosterone-regulated sodium reabsorption, Circadian rhythm, Mismatch repair, Drug metabolism-other enzymes, TNF signaling pathway, Transcriptional misregulation in cancers, TGF-beta signaling pathway, PI3K-Akt signaling pathway and so on (**Figure 5B**). The corresponding *p*-value and enrichment score of the top 20 enrichment pathways were shown in **Figure 5C**. Using the results of KEGG enrichment analysis of genes, the network between all KEGG pathways and their corresponding genes was analyzed. One hundred sixty-nine pathways and 94 genes were included, and some genes involved in multiple KEGG pathways could be found to provide auxiliary reference for selection of candidate genes. For instance, PRKCA was found to participate in 60 KEGG pathways, including mTOR signaling pathway, PI3K-Akt signaling pathway, MicroRNAs in cancer, Choline metabolism in cancer, Wnt signaling pathway, MAPK signaling pathway, Pancreatic secretion, VEGF signaling pathway, Ras signaling pathway, Pathways in cancer and so on (**Figure 6**).



C

KEGG Enrichment top 20	p-value	Enrichment score
Endometrial cancer	0.0277237565644612	2.63243533655691
Colorectal cancer	0.00800187604839237	3.24960917144346
MicroRNAs in cancer	0.0299540180840073	2.01533625172606
Transcriptional misregulation in cancers	0.000823801914806438	2.82986798679868
Aldosterone-regulated sodium reabsorption	9.50E-06	10.3531755614586
Circadian rhythm	1.29E-05	9.81457105248097
TNF signaling pathway	0.000510112569834237	3.82784075509
Jak-STAT signaling pathway	0.0235232524522196	2.2669169453661
Signaling pathways regulating pluripotency of stem cells	0.0155266324300333	2.49327575929399
TGF-beta signaling pathway	0.0230017981336404	2.79110267191103
PI3K-Akt signaling pathway	0.0298405969477694	1.78103579588728
mTOR signaling pathway	0.00517425945244094	3.61259742995576
Ubiquitin mediated proteolysis	0.00565104011929243	2.60075394848019
FoxO signaling pathway	0.000910205529976006	3.50086761253445
Fanconi anemia pathway	0.00294500916721332	4.1286827770923
Mismatch repair	0.000554629642777106	8.02167303344508
RNA transport	0.00708723658868931	2.6951123683797
Drug metabolism - other enzymes	0.000412327177635301	6.43761437755515
Sphingolipid metabolism	0.00703265289044141	3.97950185643564
Pyrimidine metabolism	0.005666360633829	3.1040599489565

FIGURE 5 | GO and KEGG pathway analysis of the common differentially expressed mRNAs. **(A)** GO analysis was conducted and covered three domains: cellular components, biological process and molecular function. The top 10 enriched GO terms were presented. **(B)** KEGG pathway analysis was also adopted and the top 20 enriched pathways were calculated and shown. **(C)** *P*-value and enrichment score of the top 20 enriched pathways were included and shown.

Construction of Coding and Non-coding Co-expression Network

Common differentially expressed mRNAs (five up-regulated and five down-regulated ones, respectively), which were proved to implicate in multiple biological processes including cell cycle, apoptosis, angiogenesis and metastasis, were selected to build this network (**Figure 7**). The network implied a complex relationship that one gene could correlate with multiple lncRNAs and one lncRNA might also regulate numerous mRNAs in different ways. As shown in **Figure 7**, up-regulated lnc-AGRN-2_9 was positively correlated with HBP1, GADD45A, SIK1, and SESN2, and negatively associated with TOP2A, CKAP2, while these mRNAs were implicated in apoptosis. Meanwhile, down-regulated SSX2IP, which was involved in tumorigenesis and

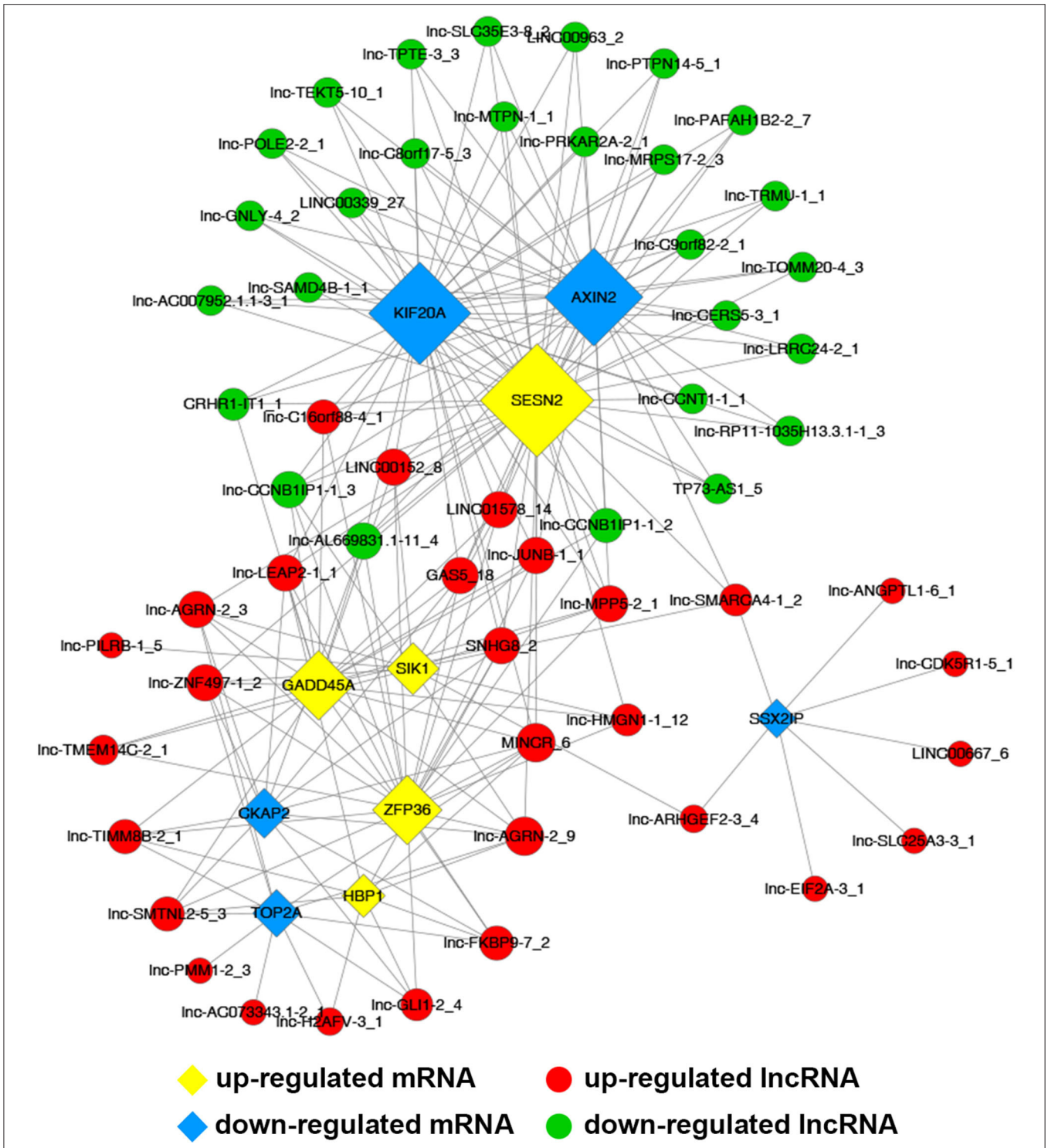


FIGURE 6 | The network between KEGG pathways and their corresponding genes was analyzed, in which 169 pathways and 94 genes were included.

metastasis, was negatively correlated with lnc-SMARCA4-1_2, lnc-EIF2A-3_1, lnc-SLC25A3-3_1, lnc-ANGPTL1-6_1, lnc-ARHGEF2-3_4, lnc-CDK5R1-5_1, and LINC00667_6. The co-expression network might imply the potential regulatory mechanisms between lncRNAs and mRNAs in the nigericin anti-cancer process.

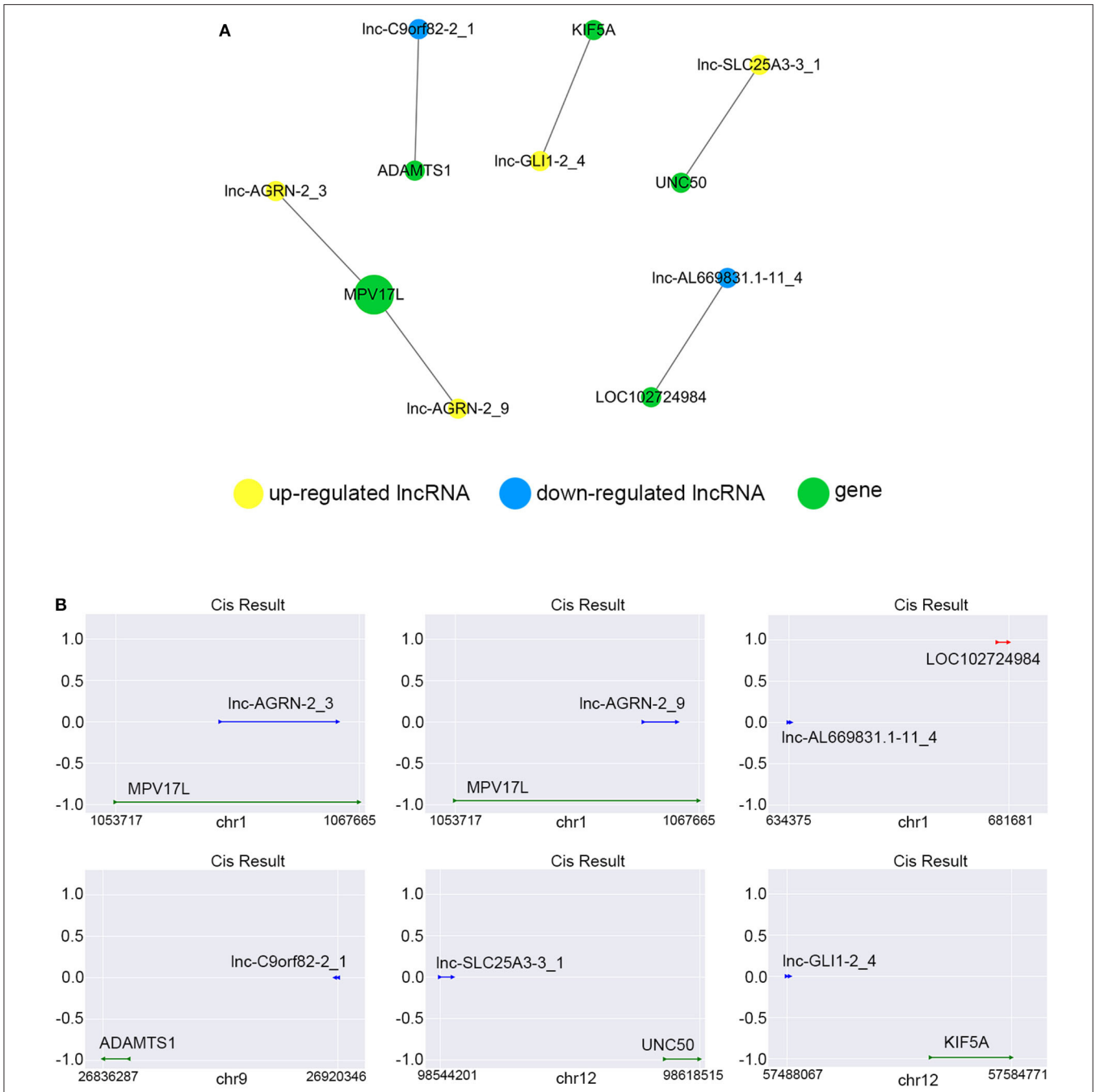


FIGURE 7 | Construction of coding and non-coding co-expression network. **(A)** Common differentially expressed mRNAs (five up-regulated and five down-regulated ones, respectively) were selected to construct the network with their co-expressed common differentially expressed lncRNAs. The network consisted of 66 nodes and 221 connections. **(B)** The cis result of the coding and non-coding co-expression network.

Cis-Regulating Function Prediction of lncRNAs

We constructed the correlated expression networks to elucidate the relationship between the common differentially expressed lncRNAs and their co-expressed adjacent coding genes. Among all the 76 common differentially expressed lncRNAs, only 6 lncRNAs were found to own neighboring protein-coding genes, and these 6 lncRNAs' potential cis-regulation networks were described in **Figure 8A**. However, each lncRNA had only one nearby coding gene. For example, lnc-AGRN-2_3 and

MPV17L, lnc-AGRN-2_9 and MPV17L, lnc-AL669831.1-11_4 and LOC102724984, lnc-C9orf82-2_1 and ADAMTS1, lnc-GLI1-2_4 and KIF5A, lnc-SLC25A3-3_1 and UNC50 were shown in **Figure 8B**. The networks might furnish valuable clue for these lncRNAs with nearby coding genes.

Trans-Regulation of lncRNAs (TF-lncRNA and TF-lncRNA-Gene Network)

Despite the prevalence of lncRNA-mediated cis-regulation, examples of trans-acting lncRNAs have also been reported (25, 26). For trans-regulation prediction, we constructed a co-expression network combined by these common differentially expressed lncRNAs with TFs. With a threshold of $P < 0.01$ and FDR < 0.01, the top 200 closest relationships were selected, while we constructed a TF-lncRNA binary network. The network showed that 44 up-regulated lncRNAs were found to correspond to 31 TFs, and 27 down-regulated lncRNAs corresponded to 12 TFs (**Figures 9A,B**). Moreover, we introduced target genes to build TF-lncRNA-gene ternary network. 10 up-regulated lncRNAs correspond to 3 TFs and 283 target genes, while six down-regulated lncRNAs were found to associate with 3 TFs and 125 target genes (**Figure 10**). Interestingly, up to 14 dys-regulated lncRNAs were regulated by 5 TFs, such as MYC, TAF1, E2F4, STAT1, and STAT2. The results implied that these TFs might also participate in the nigericin anti-cancer damage.

PPI Network Construction

As shown in **Figure 11**, a total of 152 genes of the 257 common differentially expressed genes were filtered into the PPI network containing 152 nodes and 644 edges. The nodes with high degrees were defined as hub proteins in the PPI networks and degree >10 was set as the cut-off criterion. In this network, a total of 12 nodes were selected as hub proteins, including TOP2A, MYC, ANAPC1, FBXW7, KIF20A, MTOR, CREB1, EXO1, MELK, NEDD4L, RACGAP1, and HERC2. The most significant hub proteins were TOP2A (degree = 40) and MYC (degree = 21). This network exhibited the interactions among these genes which might play a significant role in the nigericin treatment.

DISCUSSION

Recently, the anti-cancer effect of nigericin has drawn increasing attentions, and its molecular mechanisms toward cancer cells were gradually discovered. A newly study by Yakisich et al. demonstrated that nigericin might be used in a co-therapy model of lung cancer in combination with other chemotherapeutic agents (27). Coincidentally, our lab also implied that Wnt/β-catenin signaling might have an essential role in colorectal cancer progression, and nigericin exerted anti-cancer effects on colorectal cancer cells by directly targeting the β-catenin destruction complex (28). Furthermore, our recent study has proved the potential toxicity of nigericin on human PC, and revealed the molecular mechanism of nigericin toward PC cells from the perspective of circRNA (23). However, the knowledge of nigericin needs to be further elucidated from multiple perspectives.

Along with the deepening of research on PC, numerous lncRNAs have shown to be essential for the tumorigenesis and progression by serving as tumor oncogenes or suppressors. In 2016, Li et al. found that long non-coding RNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) could facilitate the advanced progression of PC by promoting autophagy *in vitro* (29). lncRNA myocardial infarction-associated transcript (MIAT) was found remarkably increased in PC tissues and cell lines, and PC patients with high MIAT levels had poor prognosis than those with low MITA levels (30). In contrast, Lnc-PCTST might exhibit as a potential tumor suppressor in PC, which inhibited cell proliferation, invasion, tumorigenesis and EMT by modulating TACC-3 (31). To further explore the anti-cancer mechanism of nigericin, we used high-throughput and bioinformatics methods to predict the changes of coding and non-coding RNAs when cells were exposed to the drug.

Firstly, the global expression profile of lncRNAs and mRNAs for four different nigericin-treated time points was determined by a custom sequencing platform. By venn analysis, our data confirmed that 76 common dys-regulated lncRNAs including 49 up-regulated and 27 down-regulated ones might participate in the process of nigericin damage. These lncRNAs were widely distributed on all chromosomes except for sex chromosome X. Meanwhile, the common differentially expressed mRNAs among the 3 compared groups were also found, in which 172 mRNAs were common up-regulated and 85 ones were down-regulated. Subsequently, we chose 5 random lncRNAs and 5 cancer-related genes for PCR detection between the 0 and 32 h group. The data were consistent well with our sequencing data, which demonstrated the high reliability and validity of the sequencing expression results. Of these common differentially expressed mRNAs, GADD45A was found to be variously expressed in cell lines derived from PC, and adenoviral-mediated expression of GADD45A (Ad-G45a) in these cells resulted in apoptosis via caspase activation and cell-cycle arrest in the G2/M phase (32). HMG-box transcription factor 1 (HBP1) had been described as a negative regulator of the Wnt/β-catenin signaling in many cancers, including breast cancer (33), osteosarcoma (34), glioma (35), and colorectal carcinoma (36). A recent study by Chan also indicated that HBP1 acted as a direct downstream target of FOXO1, and potently suppressed the phenotypes of oral cancer (37). Besides, other 3 validated genes (SESN2, SIK1, and KIF20A) were also proved to influence the proliferation, migration and invasion of PC cells (38–41). These results might provide clues to the potential mechanisms of nigericin in PC.

Next, we conducted GO and KEGG pathway analyses to uncover the roles of these common differentially expressed mRNAs after nigericin treatment. The top 10 GO biological processes such as uridine catabolic process, nucleotide catabolic process and regulation of interleukin-6 biosynthetic process were found in the nigericin damage. Meanwhile, the differentially expressed mRNAs were significantly enriched in top 20 KEGG signaling pathways, including Aldosterone-regulated sodium reabsorption, Circadian rhythm, Mismatch repair, Drug metabolism-other enzymes, TNF signaling pathway, Transcriptional misregulation in cancers, TGF-beta signaling

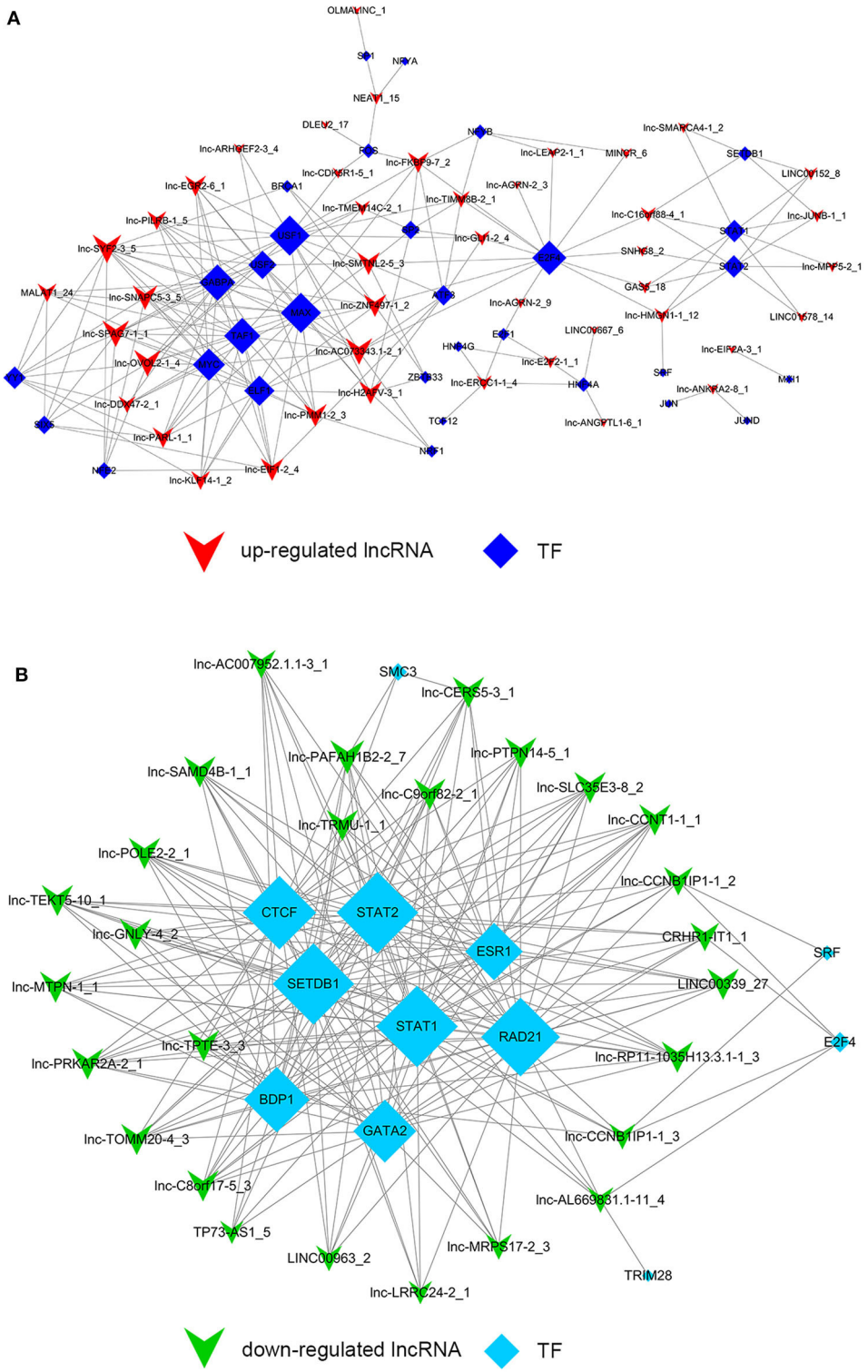


FIGURE 8 | Cis-regulation prediction of lncRNAs. **(A)** Among the 76 common differentially expressed lncRNAs, 6 lncRNAs were found to own neighboring protein-coding genes coding genes. The potential cis-regulation network was described. **(B)** The location and distances between lncRNAs and their nearby coding genes on the chromosome were presented.

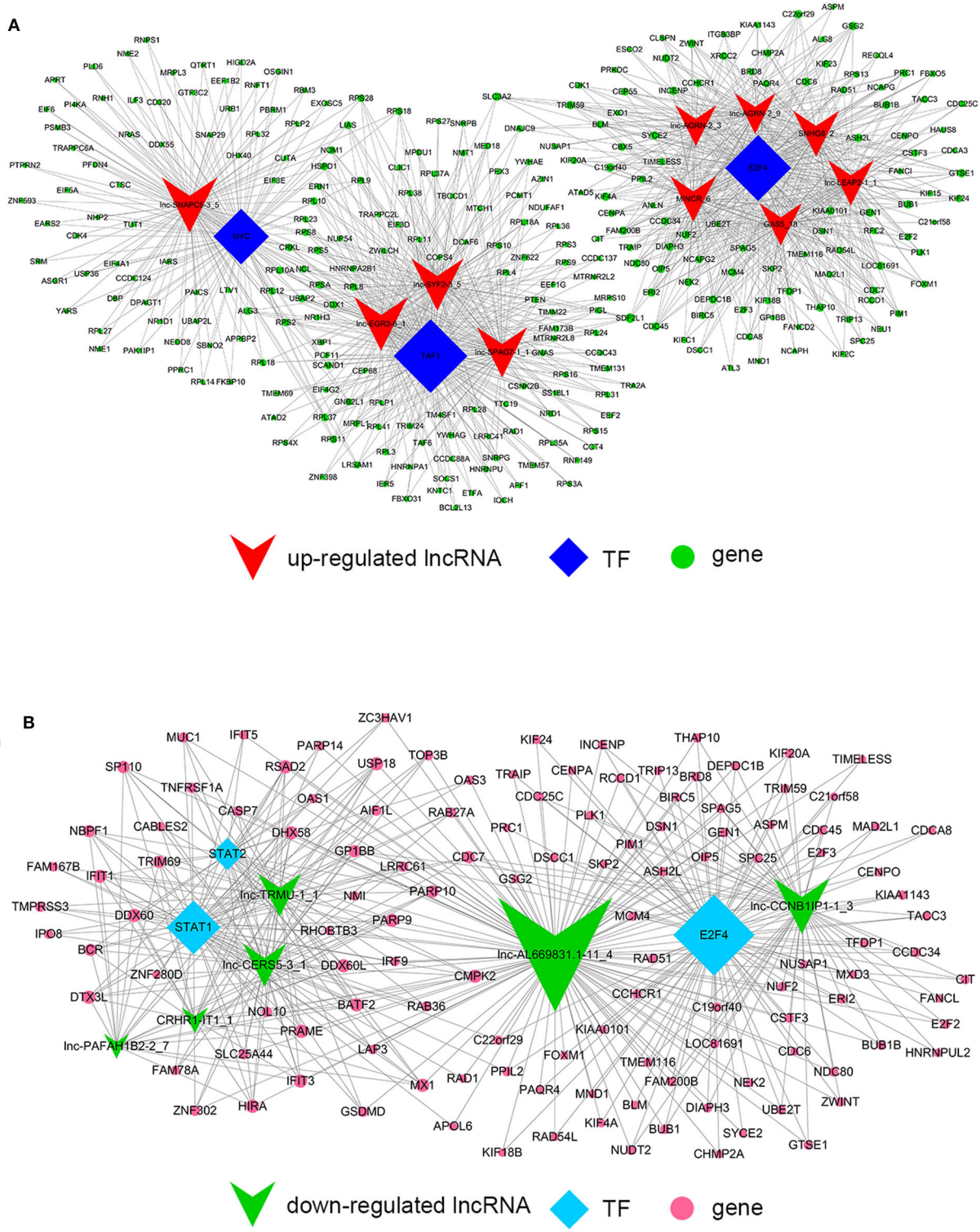


FIGURE 9 | Trans-regulation prediction of lncRNAs (TF-lncRNA binary network). **(A)** Among the 76 common differentially expressed lncRNAs, 44 up-regulated lncRNAs were found to correspond to 31 TFs. **(B)** 27 down-regulated lncRNAs corresponded to 12 TFs.

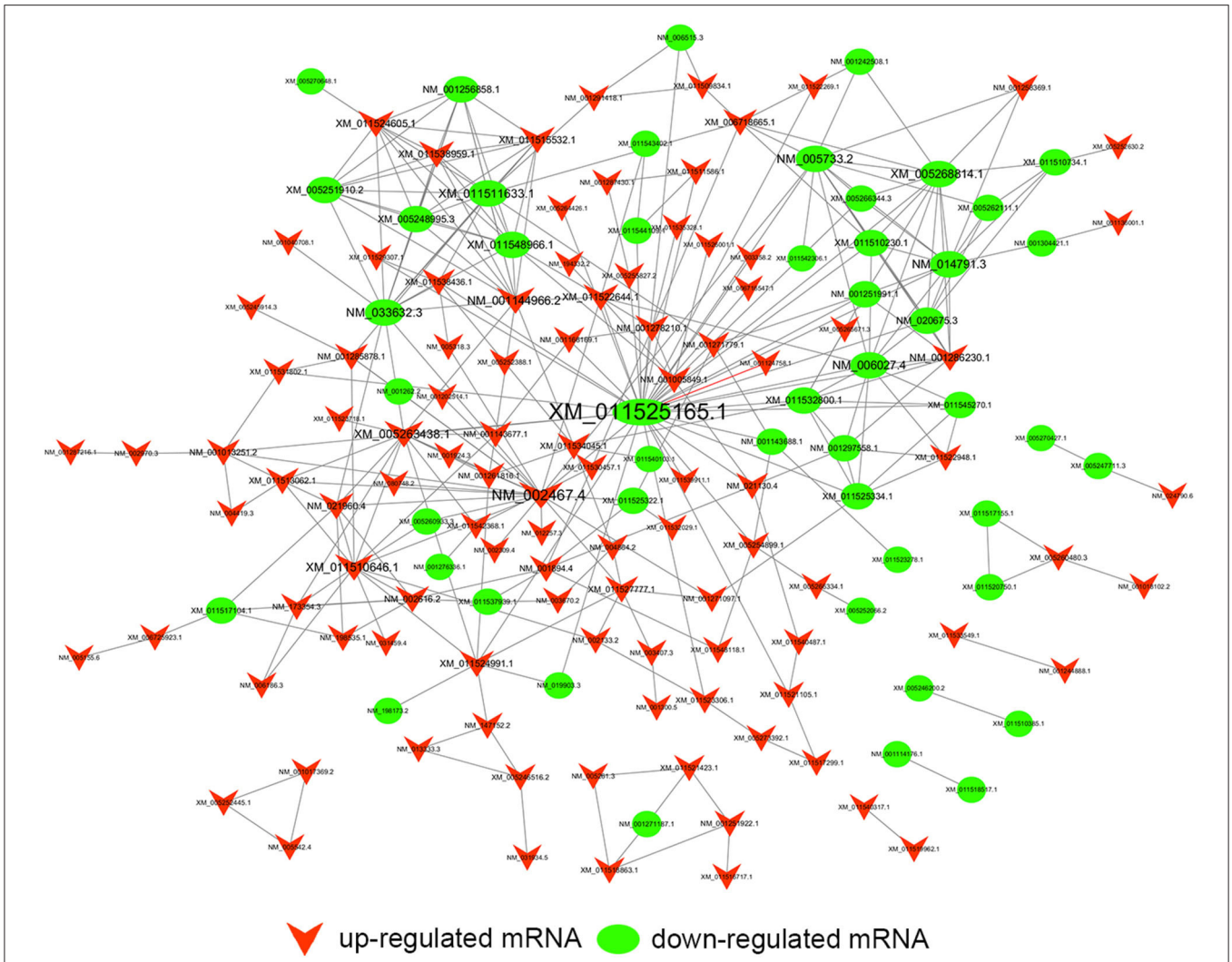


FIGURE 10 | Trans-regulation prediction of lncRNAs (TF-lncRNA-gene ternary network). Target genes were introduced to build TF-lncRNA-gene ternary network. 10 up-regulated lncRNAs correspond to 3 TFs and 283 target genes, while 6 down-regulated lncRNAs were found to associate with 3 TFs and 125 target genes.

pathway, PI3K-Akt signaling pathway and so on. Moreover, the network between all KEGG pathways and their corresponding genes was also analyzed. These nigericin-related pathways have been also reported in PC. For example, the PI3K/Akt signaling pathway is related with PC metastasis. Tanno et al. showed that increased insulin-like growth factor I receptor expression induced by active Akt markedly enhanced the invasiveness of human PC cells (42). A recent review from Murthy et al. also described the role of PI3K signaling in PC development and progression (43). In 2014, Zhu et al. provided valuable baseline information regarding the TGF-β pathway in PC, which could be utilized in targeted therapy clinical trials (44). These involved non-coding RNAs (lncRNAs and mRNAs) and GO/KEGG analyses might partly explain the phenomena that nigericin had the anti-cancer properties.

To better understand the mechanisms of nigericin in PC cells, we built the co-expression network between lncRNAs and mRNAs. The network implied a complex relationship that one gene could correlate with multiple lncRNAs and one lncRNA might also regulate numerous mRNAs in different ways. For instance, up-regulated lnc-AGRN-2_9 was positively correlated with HBP1, GADD45A, SIK1, and SESN2, and negatively associated with TOP2A, CKAP2, while these mRNAs were implicated in tumorigenesis (32, 37–41, 45). The co-expression network might imply the potential regulatory mechanisms between lncRNAs and mRNAs in the nigericin anti-cancer process.

It has been known that lncRNAs can cis-regulate the co-expressed and nearby coding genes (24). In this study, we constructed a cis-regulated network with the criterion that coding genes located at 100 k bp upstream and downstream of lncRNAs on the chromosome. Our results showed that 6 of 76 common differentially expressed lncRNAs possessed cis-regulated genes, and each of the 6 lncRNAs only had

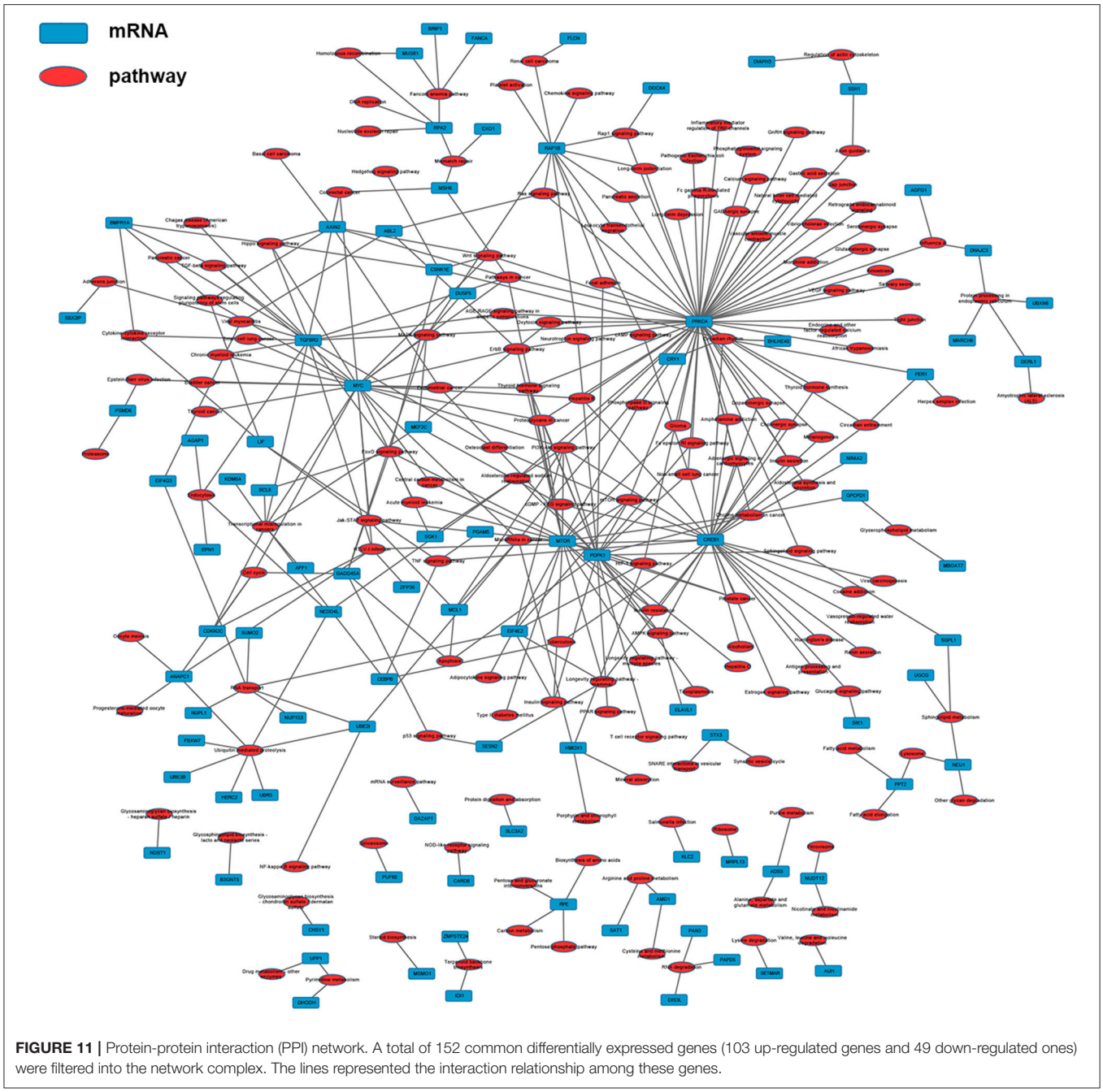


FIGURE 11 | Protein-protein interaction (PPI) network. A total of 152 common differentially expressed genes (103 up-regulated genes and 49 down-regulated ones) were filtered into the network complex. The lines represented the interaction relationship among these genes.

one neighboring protein-coding gene. For example, we found that lnc-AGRN-2_3 and lnc-AGRN-2_9 shared the same cis-regulated gene MPV17L, which indicated that these two lncRNAs might play a similar role. Lnc-C9orf82-2_1 cis-regulated ADAMTS1, and Masui et al. also suggested that ADAMTS1 was a potential biomarker to detect early-stage PCs (46). UNC50 has long been recognized as a Golgi apparatus protein in yeast, and is involved in nicotinic receptor trafficking in Caenorhabditis elegans. In 2015, Fang et al. found that UNC50 was correlated with G1/S transition and proliferation in hepatocellular carcinoma via the influencing epidermal growth factor receptor trafficking (47). Interestingly, our data showed that UNC50 was involved with the nigericin damage, which could be cis-regulated by lnc-SLC25A3-3_1. These results revealed the prevalence of lncRNA-mediated cis-regulations on nearby genes during the nigericin damage.

On the other hand, previous reports have indicated that lncRNAs are capable of binding to a specific site or sequence, including TFs, to achieve trans-regulation functions. We constructed a TF-lncRNA binary network combined by these common differentially expressed lncRNAs with TFs. The network showed that 44 up-regulated lncRNAs were found to correspond

to 31 TFs, and 27 down-regulated lncRNAs corresponded to 12 TFs. Furthermore, we introduced target genes to build TF-lncRNA-gene ternary network. 10 up-regulated lncRNAs correspond to 3 TFs and 283 target genes, while 6 down-regulated lncRNAs were found to associate with 3 TFs and 125 target genes. Interestingly, up to 14 dys-regulated lncRNAs were regulated by 5 TFs, such as MYC, TAF1, E2F4, STAT1, and STAT2. Recent evidence strongly suggests that these 5 TFs potentially regulate the expression of target genes in PC or other cancers. For instance, Valenti et al. found that Mutp53 and E2F4 proteins formed a transcriptional repressive complex that assembled onto the regulatory regions of BRCA1 and RAD17 genes inhibiting their expressions in head and neck squamous cell carcinoma (48). Guerrero-Zotano et al. identified 18 of the 20 E2F4 target genes, and suggested a potential benefit of adjuvant CDK4/6 inhibitors in patients with ER^+ breast cancer who failed to respond to preoperative estrogen deprivation (49). STAT1, which is a member of the family of signal transducers and transcription activators, corresponded to lymph node metastasis, advanced stage, tumor dedifferentiation and poor prognosis in patients with PC (50). A study from Seshacharyulu et al. also confirmed STAT1 as a key regulator through down-regualtion of MUC4 in PC (51). Thus, our cis- and trans-regulation predictions might provide a deep insight into the involved lncRNAs in nigericin treatment.

Finally, a PPI network with common differentially expressed genes, in which 12 hub proteins were identified, including TOP2A, MYC, ANAPC1, FBXW7, KIF20A, MTOR, CREB1, EXO1, MELK, NEDD4L, RACGAP1, and HERC2. The most significant hub proteins were TOP2A and MYC. TOP2A could induce tumor development and progression in many cancer types, including PC (52), prostate cancer (53) and breast cancer (54). In 2016, a phase II study by Tarpgaard et al. found that metastatic colorectal cancer (mCRC) patients, who were refractory to treatment with oxaliplatin-based chemotherapy, had TOP2A gene amplification in their tumor cells (55). Similarly, human estrogen receptor-positive breast cancer cells typically displayed elevated levels of Myc protein due to overexpression of MYC mRNA (56). Other studies had also identified the abnormal expression of MYC-binding protein (MYCBP) during tumorigenesis in multiple types of cancer, such as gastric cancer (57), colon cancer (58), and PC (59). Therefore, this core PPI network exhibited the associations between these interested genes, which might provide useful clues for the mechanism analysis of nigericin in PC.

CONCLUSION

In summary, our experiments further investigated the anti-cancer properties of nigericin in PC. In light of the high-throughput RNA sequencing analysis, we comprehensively characterized the potential contributions of lncRNAs and mRNAs after nigericin exposure. Additionally, the bioinformatics analyses, including GO and KEGG analysis, coding and non-coding co-expression network, cis- and trans-regulation predictions and PPI network, were applied to annotate the potential regulatory mechanisms among these coding and non-coding RNAs during the nigericin anti-cancer process. Our data provided new insight into the molecular mechanism of nigericin toward cancer cells, and suggested a possible clinical application in PC.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra, PRJNA543685).

AUTHOR CONTRIBUTIONS

ZX conceived the project and wrote the manuscript. QZ and YX reviewed the manuscript. All authors participated in experiment and data analysis.

FUNDING

This study was supported by grants from the National Science Foundation of China (81802340, 81902805), the Suzhou Gusu Medical Youth Talent (GSWS2019032), the Project of Invigorating Health Care through Science, Technology and Education, Jiangsu Provincial Medical Youth Talent (QNRC2016723) and National Natural Science Foundation of China (8180100710).

REFERENCES

1. Benson AB III. Adjuvant therapy for pancreatic cancer: one small step forward. *JAMA*. (2007) 297:311–3. doi: 10.1001/jama.297.3.311
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin.* (2014) 64:9–29. doi: 10.3322/caac.21208
3. Laheru D, Jaffee EM. Immunotherapy for pancreatic cancer - science driving clinical progress. *Nat Rev Cancer.* (2005) 5:459–67. doi: 10.1038/nrc1630
4. Duffy MJ, Sturgeon C, Lamerz R, Haglund C, Holubec VL, Klapdor R, et al. Tumor markers in pancreatic cancer: a European Group on Tumor Markers (EGTM) status report. *Ann Oncol.* (2010) 21:441–7. doi: 10.1093/annonc/mdp332
5. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. *Lancet.* (2011) 378:607–20. doi: 10.1016/S0140-6736(10)62307-0
6. Pramanik KC, Fofaria NM, Gupta P, Ranjan A, Kim SH, Srivastava SK. Inhibition of β-catenin signaling suppresses pancreatic tumor growth by disrupting nuclear β-catenin/TCF-1 complex: critical role of STAT-3. *Oncotarget.* (2015) 6:11561–74. doi: 10.18632/oncotarget.3427
7. Gobbi PG, Federico M. What has happened to VBM (vinblastine, bleomycin, and methotrexate) chemotherapy for early-stage Hodgkin lymphoma? *Crit Rev Oncol Hematol.* (2012) 82:18–24. doi: 10.1016/j.critrevonc.2011.04.003
8. Altena R, Fehrmann RS, Boer H, de Vries EG, Meijer C, Gietema JA. Growth differentiation factor 15 (GDF-15) plasma levels increase during bleomycin- and cisplatin-based treatment of testicular cancer patients and relate to endothelial damage. *PLoS ONE.* (2015) 10:e0115372. doi: 10.1371/journal.pone.0115372
9. Kumar SM, Biswas S, Sarkar S, Mandal SS, Biswas J. Analytical study on the efficacy of neoadjuvant chemotherapy using a combination of methotrexate,

bleomycin, and cisplatin in the management of advanced squamous cell Carcinoma of the buccal mucosa. *Indian J Med Paediatr Oncol.* (2017) 38:345–8. doi: 10.4103/ijmpo.ijmpo_123_17
10. Zhi QM, Chen XH, Ji J, Zhang JN, Li JF, Cai Q, et al. Salinomycin can effectively kill ALDH(high) stem-like cells on gastric cancer. *Biomed Pharmacother.* (2011) 65:509–15. doi: 10.1016/j.biopha.2011.06.006
11. Zhou J, Li P, Xue XF, He SB, Kuang YT, Zhao H, et al. Salinomycin induces apoptosis in cisplatin-resistant colorectal cancer cells by accumulation of reactive oxygen species. *Toxicol Lett.* (2013) 222:139–45. doi: 10.1016/j.toxlet.2013.07.022
12. Yadav V, Varshney P, Sultana S, Yadav J, Saini N. Moxifloxacin and ciprofloxacin induces S-phase arrest and augments apoptotic effects of cisplatin in human pancreatic cancer cells via ERK activation. *BMC Cancer.* (2015) 15:581. doi: 10.1186/s12885-015-1560-y
13. Yadav V, Sultana S, Yadav J, Saini N. Gatifloxacin induces S and G2-phase cell cycle arrest in pancreatic cancer cells via p21/p27/p53. *PLoS ONE.* (2012) 7:e47796. doi: 10.1371/journal.pone.0047796
14. Shavit N, Dilley RA, San Pietro A. Ion translocation in isolated chloroplasts. uncoupling of photophosphorylation and translocation of k+ and h+ ions induced by nigericin. *Biochemistry.* (1968) 7:2356–63. doi: 10.1021/bi00846a043
15. Poole DT, Butler TC, Williams ME. The effects of nigericin, valinomycin, and 2,4-dinitrophenol on intracellular pH, glycolysis, and K + concentration of Ehrlich ascites tumor cells. *Biochim Biophys Acta.* (1972) 266:463–70. doi: 10.1016/0005-2736(72)90102-2
16. Mashima T, Okabe S, Seimiya H. Pharmacological targeting of constitutively active truncated androgen receptor by nigericin and suppression of hormone-refractory prostate cancer cell growth. *Mol Pharmacol.* (2010) 78:846–54. doi: 10.1124/mol.110.064790
17. Deng CC, Liang Y, Wu MS, Feng FT, Hu WR, Chen LZ, et al. Nigericin selectively targets cancer stem cells in nasopharyngeal carcinoma. *Int J Biochem Cell Biol.* (2013) 45:1997–2006. doi: 10.1016/j.biocel.2013.06.023
18. Zanke BW, Lee C, Arab S, Tannock IF. Death of tumor cells after intracellular acidification is dependent on stress-activated protein kinases (SAPK/JNK) pathway activation and cannot be inhibited by Bcl-2 expression or interleukin 1β-converting enzyme inhibition. *Cancer Res.* (1998) 58:2801–8.
19. Lu D, Choi MY, Yu J, Castro JE, Kipps TJ, Carson DA. Salinomycin inhibits Wnt signaling and selectively induces apoptosis in chronic lymphocytic leukemia cells. *Proc Natl Acad Sci USA.* (2011) 108:13253–7. doi: 10.1073/pnas.1110431108
20. Varnes ME, Bayne MT, Menegay HJ, Tuttle SW. Effect of the K+/H+ ionophore nigericin on response of A549 cells to photodynamic therapy and tert-butylhydroperoxide. *Free Radic Biol Med.* (1993) 15:395–405. doi: 10.1016/0891-5849(93)90039-W
21. Vaupel P. Tumor microenvironmental physiology and its implications for radiation oncology. *Semin Radiat Oncol.* (2004) 14:198–206. doi: 10.1016/j.semradonc.2004.04.008
22. Zhou HM, Dong TT, Wang LL, Feng B, Zhao HC, Fan XK, et al. Suppression of colorectal cancer metastasis by nigericin through inhibition of epithelial-mesenchymal transition. *World J Gastroenterol.* (2012) 18:2640–8. doi: 10.3748/wjg.v18.i21.2640
23. Xu ZH, Shen JQ, Hua SB, Wan DW, Chen Q, Han Y, et al. High-throughput sequencing of circRNAs reveals novel insights into mechanisms of nigericin in pancreatic cancer. *BMC Genomics.* (2019) 20:716. doi: 10.1186/s12864-019-6032-3
24. Guenzl PM, Barlow DP. Macro lncRNAs: a new layer of cis-regulatory information in the mammalian genome. *RNA Biol.* (2012) 9:731–41. doi: 10.4161/rna.19985
25. Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol.* (2010) 7:582–5. doi: 10.4161/rna.7.5.13216
26. Peng BY, Liu AF, Yu XW, Xu EW, Dai JB, Li MC, et al. Silencing of lncRNA AFAP1-AS1 suppressed lung cancer development by regulatory mechanism in cis and trans. *Oncotarget.* (2017) 8:93608–23. doi: 10.18632/oncotarget.20549
27. Yakisich JS, Azad N, Kaushik V, O'Doherty GA, Iyer AK. Nigericin decreases the viability of multidrug-resistant cancer cells and lung tumorspheres and potentiates the effects of cardiac glycosides. *Tumour Biol.* (2017) 39:1010428317694310. doi: 10.1177/1010428317694310
28. Liu F, Li W, Hua SB, Han Y, Xu ZH, Wan DW, et al. Nigericin exerts anticancer effects on human colorectal cancer cells by inhibiting wnt/β-catenin signaling pathway. *Mol Cancer Ther.* (2018) 17:952–65. doi: 10.1158/1535-7163.MCT-17-0906
29. Li L, Chen H, Gao Y, Wang YW, Zhang GQ, Pan SH, et al. Long noncoding RNA MALAT1 Promotes aggressive pancreatic cancer proliferation and metastasis via the stimulation of autophagy. *Mol Cancer Ther.* (2016) 15:2232–43. doi: 10.1158/1535-7163.MCT-16-0008
30. Li TF, Liu J, Fu SJ. The interaction of long non-coding RNA MIAT and miR-133 play a role in the proliferation and metastasis of pancreatic carcinoma. *Biomed Pharmacother.* (2018) 104:145–50. doi: 10.1016/j.biopha.2018.05.043
31. Wang YD, Ding XY, Hu H, He Y, Lu ZP, Wu PF, et al. Long non-coding RNA lnc-PCTST predicts prognosis through inhibiting progression of pancreatic cancer by downregulation of TACC-3. *Int J Cancer.* (2018) 143:3143–54. doi: 10.1002/ijc.31657
32. Li YF, Qian HL, Li X, Wang HJ, Yu J, Liu YJ, et al. Adenoviral-mediated gene transfer of Gadd45a results in suppression by inducing apoptosis and cell cycle arrest in pancreatic cancer cell. *J Gene Med.* (2009) 11:3–13. doi: 10.1002/jgm.1270
33. Paulson KE, Rieger-Christ K, McDevitt MA, Kuperwasser C, Kim J, Unanue V, et al. Alterations of the HBP1 transcriptional repressor are associated with invasive breast cancer. *Cancer Res.* (2007) 67:6136–45. doi: 10.1158/0008-5472.CAN-07-0567
34. Sun XH, Geng XL, Zhang J, Zhao HX, Liu Y. miR-155 promotes the growth of osteosarcoma in a HBP1-dependent mechanism. *Mol Cell Biochem.* (2015) 403:139–47. doi: 10.1007/s11010-015-2344-z
35. Yan ZY, Wang JP, Wang C, Jiao YB, Qi WG,s Che SS. miR-96/HBP1/Wnt/β-catenin regulatory circuitry promotes glioma growth. *FEBS Lett.* (2014) 588:3038–46. doi: 10.1016/j.febslet.2014.06.017
36. Lee MF, Hsieh NT, Huang CY, Li CI. All trans-retinoic acid mediates MED28/HMG box-containing protein 1 (HBP1)/β-catenin signaling in human colorectal cancer cells. *J Cell Physiol.* (2016) 231:1796–803. doi: 10.1002/jcp.25285
37. Chan CY, Huang SY, Sheu JJ, Roth MM, Chou IT, Lien CH, et al. Transcription factor HBP1 is a direct anti-cancer target of transcription factor FOXO1 in invasive oral cancer. *Oncotarget.* (2017) 8:14537–48. doi: 10.18632/oncotarget.14653
38. Donadelli M, Dalla Pozza E, Costanzo C, Scupoli MT, Piacentini P, Scarpa A, et al. Increased stability of P21(WAF1/CIP1) mRNA is required for ROS/ERK-dependent pancreatic adenocarcinoma cell growth inhibition by pyrrolidine dithiocarbamate. *Biochim Biophys Acta.* (2006) 1763:917–26. doi: 10.1016/j.bbamcr.2006.05.015
39. Taniuchi K, Furihata M, Saibara T. KIF20A-mediated RNA granule transport system promotes the invasiveness of pancreatic cancer cells. *Neoplasia.* (2014) 16:1082–93. doi: 10.1016/j.neo.2014.10.007
40. Stangel D, Erkan M, Buchholz M, Gress T, Michalski C, Raulefs S, et al. Kif20a inhibition reduces migration and invasion of pancreatic cancer cells. *J Surg Res.* (2015) 197:91–100. doi: 10.1016/j.jss.2015.03.070
41. Pei YF, Yin XM, Liu XQ. TOP2A induces malignant character of pancreatic cancer through activating β-catenin signaling pathway. *Biochim Biophys Acta.* (2018) 1864:197–207. doi: 10.1016/j.bbadis.2017.10.019
42. Tanno S, Tanno S, Mitsuuchi Y, Altomare DA, Xiao GH, Testa JR. AKT activation up-regulates insulin-like growth factor I receptor expression and promotes invasiveness of human pancreatic cancer cells. *Cancer Res.* (2001) 61:589–93. doi: 10.1016/S0016-5085(08)80193X
43. Murthy D, Attri KS, Singh PK. Phosphoinositide 3-kinase signaling pathway in pancreatic ductal adenocarcinoma progression, pathogenesis, and therapeutics. *Front Physiol.* (2018) 9:335. doi: 10.3389/fphys.2018.00335
44. Javle M, Li Y, Tan DF, Dong XQ, Chang P, Kar S, et al. Biomarkers of TGF-β signaling pathway and prognosis of pancreatic cancer. *PLoS ONE.* (2014) 9:e85942. doi: 10.1371/journal.pone.0085942
45. Zhang R, Xu J, Zhao J, Bai JH. Proliferation and invasion of colon cancer cells are suppressed by knockdown of TOP2A. *J Cell Biochem.* (2018) 5:113–20. doi: 10.1002/jcb.26916
46. Yi JM, Guzzetta AA, Bailey VJ, Downing SR, Neste LV, Chiappinelli KB, et al. Novel methylation biomarker panel for the early detection of pancreatic cancer. *Clin Cancer Res.* (2013) 19:6544–55. doi: 10.1158/1078-0432.CCR-12-3224

47. Fang Z, Zhou LN, Jiang SM, Cao LH, Yu L. UNC50 prompts G1/S transition and proliferation in HCC by regulation of epidermal growth factor receptor trafficking. *PLoS ONE.* (2015) 10:e0119338. doi: 10.1371/journal.pone.0119338
48. Valenti F, Ganci F, Fontemaggi G, Sacconi A, Strano S, Blandino G, et al. Gain of function mutant p53 proteins cooperate with E2F4 to transcriptionally downregulate RAD17 and BRCA1 gene expression. *Oncotarget.* (2015) 6:5547–66. doi: 10.18632/oncotarget.2587
49. Guerrero-Zotano AL, Stricker TP, Formisano L, Hutchinson KE, Stover DG, Lee KM, et al. ER(+) breast cancers resistant to prolonged neoadjuvant letrozole exhibit an E2F4 Transcriptional program sensitive to CDK4/6 inhibitors. *Clin Cancer Res.* (2018) 24:2517–29. doi: 10.1158/1078-0432.CCR-17-2904
50. Sun Y, Yang SS, Sun N, Chen JQ. Differential expression of STAT1 and p21 proteins predicts pancreatic cancer progression and prognosis. *Pancreas.* (2014) 43:619–23. doi: 10.1097/MPA.0000000000000074
51. Kunigal S, Ponnusamy MP, Momi N, Batra SK, Chellappan SP. Nicotine, IFN-γ and retinoic acid mediated induction of MUC4 in pancreatic cancer requires E2F1 and STAT-1 transcription factors and utilize different signaling cascades. *Mol Cancer.* (2012) 11:24. doi: 10.1186/1476-4598-11-24
52. Tsiambas E, Karameris A, Tiniakos DG, Karakitsos P. Evaluation of topoisomerase IIa expression in pancreatic ductal adenocarcinoma: a pilot study using chromogenic in situ hybridization and immunohistochemistry on tissue microarrays. *Pancreatology.* (2007) 7:45–52. doi: 10.1159/000101877
53. de Resende MF, Vieira S, Chinen LT, Chiappelli F, Fonseca FP, Guimaraes GC, et al. Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer. *J Transl Med.* (2013) 11:36. doi: 10.1186/1479-5876-11-36
54. An X, Xu F, Luo RZ, Zheng QF, Lu JB, Yang YH, et al. The prognostic significance of topoisomerase II alpha protein in early stage luminal breast cancer. *BMC Cancer.* (2018) 18:331. doi: 10.1186/s12885-018-4170-7
55. Tarpgaard LS, Qvortrup C, Nygard SB, Nielsen SL, Andersen DR, Jensen NF, et al. A phase II study of Epirubicin in oxaliplatin-resistant patients with metastatic colorectal cancer and TOP2A gene amplification. *BMC Cancer.* (2016) 16:91. doi: 10.1186/s12885-016-2124-5
56. Tian X, Chakrabarti A, Amirkhanov NV, Aruva MR, Zhang KJ, Mathew B, et al. External imaging of CCND1, MYC, and KRAS oncogene mRNAs with tumor-targeted radionuclide-PNA-peptide chimeras. *Ann N Y Acad Sci.* (2005) 1059:106–44. doi: 10.1196/annals.1339.038
57. Gong LJ, Xia YJ, Qian ZY, Shi J, Luo JG, Song GY, et al. Overexpression of MYC binding protein promotes invasion and migration in gastric cancer. *Oncol Lett.* (2018) 15:5243–9. doi: 10.3892/ol.2018.7944
58. Jung HC, Kim K. Identification of MYCBP as a β-catenin/LEF-1 target using DNA microarray analysis. *Life Sci.* (2005) 77:1249–62. doi: 10.1016/j.lfs.2005.02.009
59. Hata T, Suenaga M, Marchionni L, Das AM, Yu J, Shindo K, et al. Genome-wide somatic copy number alterations and mutations in high-grade pancreatic intraepithelial neoplasia. *Am J Pathol.* (2018) 188:1723–33. doi: 10.1016/j.ajpath.2018.03.012

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xu, Gao, Liu, Han, Dai, Wang, Wei, Kuang, Wan, Zhi and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fphar.2020.00881

Integrated PPI- and WGCNA-Retrieval of Hub Gene Signatures Shared Between Barrett's Esophagus and Esophageal Adenocarcinoma

Asma Sindhoo Nangraj[1], Gurudeeban Selvaraj[2], Satyavani Kaliamurthi[2], Aman Chandra Kaushik[1,3], William C. Cho[4] and Dong Qing Wei[1,2,5]**

[1] The State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [2] Center of Interdisciplinary Sciences-Computational Life Sciences, Henan University of Technology, Zhengzhou, China, [3] Wuxi School of Medicine, Jiangnan University, Wuxi, China, [4] Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, China, [5] Peng Cheng Laboratory, Shenzhen, China

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Esra Gov,
Adana Science and Technology University, Turkey
Shuyan Li,
Lanzhou University, China

***Correspondence:**
Gurudeeban Selvaraj
gurudeeb99@haut.edu.cn
Dong Qing Wei
dqwei@sjtu.edu.cn

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *05 January 2020*
Accepted: *28 May 2020*
Published: *31 July 2020*

Citation:
Nangraj AS, Selvaraj G, Kaliamurthi S, Kaushik AC, Cho WC and Wei DQ (2020) Integrated PPI- and WGCNA-Retrieval of Hub Gene Signatures Shared Between Barrett's Esophagus and Esophageal Adenocarcinoma. Front. Pharmacol. 11:881. doi: 10.3389/fphar.2020.00881

Esophageal adenocarcinoma (EAC) is a deadly cancer with high mortality rate, especially in economically advanced countries, while Barrett's esophagus (BE) is reported to be a precursor that strongly increases the risk of EAC. Due to the complexity of these diseases, their molecular mechanisms have not been revealed clearly. This study aims to explore the gene signatures shared between BE and EAC based on integrated network analysis. We obtained EAC- and BE-associated microarray datasets GSE26886, GSE1420, GSE37200, and GSE37203 from the Gene Expression Omnibus and ArrayExpress using systematic meta-analysis. These data were accompanied by clinical data and RNAseq data from The Cancer Genome Atlas (TCGA). Weighted gene co-expression network analysis (WGCNA) and differentially expressed gene (DEG) analysis were conducted to explore the relationship between gene sets and clinical traits as well as to discover the key relationships behind the co-expression modules. A differentially expressed gene-based protein–protein interaction (PPI) complex was used to extract hub genes through Cytoscape plugins. As a result, 403 DEGs were excavated, comprising 236 upregulated and 167 downregulated genes, which are involved in the cell cycle and replication pathways. Forty key genes were identified using modules of MCODE, CytoHubba, and CytoNCA with different algorithms. A dark-gray module with 207 genes was identified which having a high correlation with phenotype (gender) in the WGCNA. Furthermore, five shared hub gene signatures (SHGS), namely, pre-mRNA processing factor 4 (PRPF4), serine and arginine-rich splicing factor 1 (SRSF1), heterogeneous nuclear ribonucleoprotein M (HNRNPM), DExH-Box Helicase 9 (DHX9), and origin recognition complex subunit 2 (ORC2), were identified between BE and EAC. SHGS enrichment denotes that RNA metabolism and splicosomes play a key role in

esophageal cancer development and progress. We conclude that the PPI complex and WGCNA co-expression network highlight the importance of phenotypic identifying hub gene signatures for BE and EAC.

Keywords: bioinformatics analysis, Barrett's esophagus, hub gene signature, esophageal adenocarcinoma, weighted gene co-expression network analysis, protein–protein interaction

INTRODUCTION

Esophageal cancer is a deadly cancer considering its high mortality rate, with 572,034 newly diagnosed cases and 508,585 deaths in 2018 (Bray et al., 2018). Esophageal cancer is classified into two subcategories: esophageal adenocarcinoma (EAC; distal esophagus) and esophageal squamous cell carcinoma (ESCC; proximal esophagus). It starts from the esophageal epithelium, the innermost layer of the esophagus (Rustgi and El-Serag, 2014). Esophageal cancer is a very complex disease, as its various subtypes have different risk factors, time trends, and geographic patterns (Analysis et al., 2017) (Montgomery et al., 2014; Lordick et al., 2016). According to the geographic variation, EAC is more common in economically advanced regions than in low-income countries (Chai et al., 2019). The common risk factors of EAC are age, male sex, obese, gastroesophageal reflux disease (GERD), cigarette smoking, and diet (low in vegetables and fruits). Cook et al. (2014) report some common symptoms like vomiting/nausea and heartburn in EAC and GERD. Besides, Barrett's esophagus (BE) is considered as a precursor for EAC. BE is a metaplastic transformation from the normal squamous mucosa of the esophagus to a columnar lining; its presence conveys a 30–40-fold increased risk of EAC (Schneider and Corley, 2015). The tumor development is a step-by-step process that comprises constant changes from erosive esophagitis to non-dysplastic BE, low-grade dysplasia, high-grade dysplasia, adenocarcinoma *in situ*, and finally invasive adenocarcinoma (Anaparthy and Sharma, 2014). Due to poor prognosis, over 40% of patients are diagnosed with high-grade dysplasia. Additionally, the 5-year survival rate is less than 20% despite the advances in diagnosis and treatment (Tramontano et al., 2017). Certainly, surgical therapy has improved the patient's survival yet it is not suitable for advanced-stage cancer patients (Davies et al., 2014).

Thus, it is essential to discover biomarkers that can lead to the discovery of medication. Microarray analysis of gene expression profiles is a common practice for identifying key hub genes and key pathways (Wei et al., 2018; Sadhu and Bhattacharyya, 2019). In the current era of integrated bioinformatics, acquiring data is not an issue; rather, normalization seems to be a tough job (Campain and Yang, 2010). Considering all of these notions, we designed an integrated study to find key hub genes associated with BE and EAC. First, we extracted BE- and EAC-associated microarray datasets from the Gene Expression Omnibus (GEO) and ArrayExpress using systematic meta-analysis as well as RNA-seq data from TCGA. Preprocessing and normalization were conducted for further analysis. DEGs were identified using linear models for microarray data (LIMMA) algorithm. Meta-analysis was performed using a network analysis tool. We analyzed functional and pathway enrichment of DEGs. Additionally, a protein–protein interaction (PPI) network was constructed to study the associations between the DEGs and to recognize target genes using different modules of Cytoscape software. Weighted gene co-expression network analysis (WGCNA) was conducted by the construction of the co-expression network to find a correlation between modules and clinical traits. Furthermore, clinically significant modules were identified. Finally, key hub genes were identified and validated using immunohistochemistry and survival analysis.

MATERIALS AND METHODS

Data acquisition, Preprocessing, and Normalization

The microarray datasets were systematically extracted from the GEO[1] (Edgar et al., 2002) and the ArrayExpress[2] database (Brazma et al., 2003). The gene expression profiles based on RNA-sequencing were additionally obtained from The Cancer Genome Atlas (TCGA)[3] (Zhu et al., 2014). The framework of this study is shown in **Figure 1**. For microarray profiles, we selected four datasets (GSE26886, GSE1420, GSE37200, and GSE37201) available by October 2019 (Kimchi et al., 2005; Silvers et al., 2010; Wang et al., 2013; Lin et al., 2015). The GEO accession number, sample size, description, platform, expression data, and references are extracted from each identified dataset (**Table 1**). The TCGA portal was accessed in October 2019, 184 esophageal cancer samples were retrieved. The tab-delimited text (.txt) files of microarray datasets were obtained from the GEO database. The Network Analyst (NA) web interface for integrative biological network analysis was employed for background correction preprocessing, normalization, probe identification, and meta-analysis of the datasets (Xia et al., 2015). The input files were prepared as per the description of the tool (first line #Name (sample ID); second line #class (sample type); genes in the rows and samples in the columns). We applied two different methods to normalize the datasets: first, variance stabilizing normalization (VSN), which improves DEG detection and reduces false-positive errors (Konishi, 1985), and second, quantile normalization, which can make two distributions equal in statistical methods (Hansen et al., 2012). The processed datasets were used for subsequent microarray meta-analysis.

[1] http://www.ncbi.nlm.nih.gov/geo

[2] https://www.ebi.ac.uk/arrayexpress/

[3] https://portal.gdc.cancer.gov/

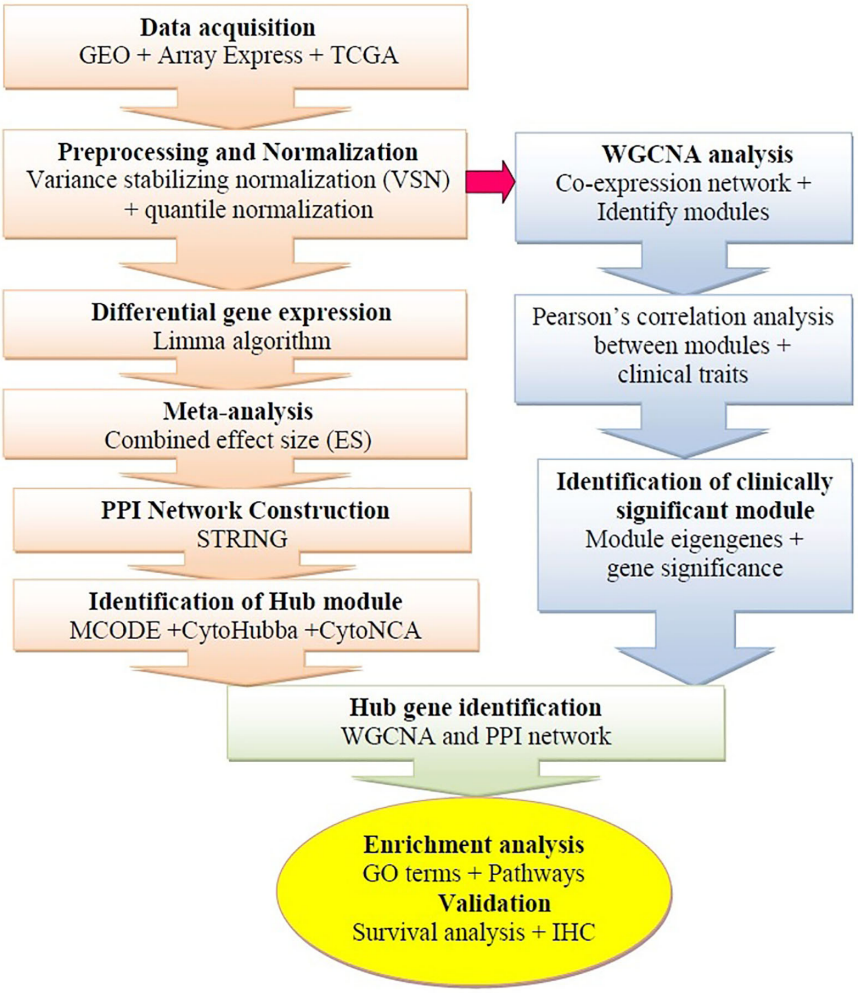


FIGURE 1 | Schematic flow diagram of the study. GEO, Gene Expression Omnibus; IHC, Immunohistochemistry; WGCNA, weighted gene co-expression network analysis; TCGA, The Cancer Genome Atlas; PPI, Protein-protein interaction.

DEG Identification and Meta-Analysis

Differential gene expression analysis was performed with the R package LIMMA (linear models for microarray data), which is embedded in NA (Ritchie et al., 2015). Each gene expression was calculated based on the false discovery rate (FDR; $p < 0.05$) using the Benjamini–Hochberg method and *t*-test. In addition, the microarray meta-analysis between EAC and BE samples was performed using combined effect size (ES). The combined ES is the difference between two group means divided by standard deviation, which is comparable across different studies. It can be calculated by two types of models, namely fixed-effect models (FEM) and random-effect models (REM). In FEM, the calculated effect size in each study

TABLE 1 | Relevant information about selected microarray datasets.

GSE Acc. No.	No. of Samples	Platform	Description	Country	PMID
GSE26886	21 vs 20	AHG-U133 Plus 2.0 Array	Gene expression profiling of Barrett's esophagus, adenocarcinoma, esophageal squamous epithelium, and squamous cell carcinoma	Germany	23514407
GSE1420	8 vs 8	AHG-U133A Array	Barrett's esophagus, Barrett's-associated adenocarcinomas and normal esophageal epithelium	USA	15833844
GSE37200	15 vs 31	AHG-U133A Array	Gene expression profiling of Barrett's esophageal tissues and esophageal adenocarcinoma	USA	26068949
GSE37201	22	AHG-U133A Array	Barrett's esophageal tissues and esophageal adenocarcinoma	USA	20332323

AHG, Affymetrix Human Genome.

is supposed to arise from an original true effect size plus measurement error. In REM, each study further contains a random effect that can incorporate unknown cross-study (different platforms) heterogeneities in the model. The FEM or REM can be chosen based on statistical heterogeneity estimated using Cochran's Q tests (Cochran, 1950). The method typically gives a lower number of DEGs but more confidence (Selvaraj et al., 2018).

Cochran's Q test equation:

$$\text{Cochran}'\text{s Q test equation}: \; T = k(k-1)\frac{\sum_{j=1}^{k}(\mathrm{X}\cdot_j - {}^{N}/_{K})^2}{\sum_{i-1}^{b}X_{i\cdot}(k-\mathrm{X}_i\cdot)}$$

where k is the number of samples; $X\cdot_j$ is the column total for the j^{th} sample; b is the number of genes; $X_i\cdot$ is the row total for the i^{th} gene; N is the grand total.

Gene Ontology and Pathway Enrichment Analysis

We used ClueGO v2.5.3, a Cytoscape[4] plugin, for function and pathway enrichment analysis of DEGs (Bindea et al., 2009; Kohl et al., 2011). A list of DEGs or hub genes were provided as input into ClueGO with select specific parameters, for example, species, such as *Homo sapiens,* ID type—Entrez gene ID, different enrichment functions—biological process or cellular component or molecular function or KEGG pathways, for the analysis. Each enrichment was calculated based on the Bonferroni method (kappa score 0.96; cutoff value $p < 0.005$).

PPI Network Construction and Module Extraction

The search tool for retrieval of interacting genes/proteins (STRING)[5] (Szklarczyk et al., 2017) is a database that is used to construct the PPI network. Currently, the database consists of 18,838 human proteins with 25,914,693 core network interactions. In this study, we constructed the PPI network from identified DEGs using the STRING interactome. The highest confidence interaction score was set to 0.9, which reduces the number of false-positive interactions (Bozhilova et al., 2019). Molecular complex detection (MCODE) is a Cytoscape plugin used to identify the finest clusters. MCODE calculates accurate correlation levels as well as identifying essential PPI network modules (Shannon et al., 2003). In addition, other add-ins of Cytoscape, namely, CytoHubba and CytoNCA, were employed to discover the highest linkage hub genes in the network (Chen et al., 2009; Tang et al., 2015).

WGCNA Analysis

The WGCNA package was employed to construct a gene co-expression network using a variant set of genes (12,701 genes). The analysis was performed based on the package instructions (Langfelder and Horvath, 2008). The connection strength between each pair of nodes was calculated using the adjacency matrix a_{ij}.

$$Z_{ij} = \left[cor\left(b_i, b_j\right)\right]a_{ij} = Z_{ij\beta}$$

While vectors (*bi* and *bj*) were expression values for genes, Pearson's correlation coefficient of gene i and j and a_{ij} were represented as the connection strength between genes. The soft-thresholding power of $\beta = 9$ was used to ensure scale-free topology. The hierarchical clustering of the weighting coefficient matrix was used to define the modules. The functional modules in the co-expression network with defined genes, the topological measure (TOM) indicating the concurrence in shared adjacent genes, was calculated as

$$TOMi, j = \frac{\Sigma_{K=1}^{N}A_{i,j}\cdot A_{k,j} + A_{i,j}}{min\left(K_i, K_j\right) + 1 - A_{i,j}}$$

where A is the weighted adjacency matrix described in the above formula. TOM-based dissimilarity measures with a minimum size of 100 for the gene dendrogram and average linkage hierarchical clustering were performed, and similar expression profiles were divided into the same gene modules using the dynamic tree cut package.

Identification of Clinically Significant Modules

Eigengene and gene significance methods were used to identify modules that were correlated with clinical traits of the GSE37200 microarray data set. The first principal component of each gene module and the expression of the module eigengene were defined as representative of the whole gene set and were described in the first eigengene module. The association between module eigengenes and clinical trait was used to calculate and identify the significant clinical module. Second, the gene significance was described as a mediated p-value of each gene in the linear regression between expression and clinical traits. Furthermore, the module significance was described as the average the gene significance of all genes associated with the module. The average absolute gene significance was defined as module significance. It was calculated to incorporate clinical traits into a co-expression network (Langfelder and Horvath, 2008).

Survival Analysis and Validation of SHGS

The SHGS were identified from the modules of WGCNA and the PPI network using an interactive Venn diagram. The R package *survival* was employed to calculate Kaplan–Meier (KM) survival plots with hazard ratio (HR) and log-rank tests of hubs, which was implemented in the *OSeac*[6] (consensus survival analysis for EAC) web interface. *OSeac* retrieved the gene expression profiles and clinical data including TNM (Stage I, II, III, and IV), gender (male and female), race (White, Black, and African American), and grade (G1, G2, G3, and GX) of 198 patients from TCGA and GEO. We analyzed the overall survival rate of the shared gene signature as an input and obtained the plot from the tool (Wang et al., 2020). The Human Protein Atlas[7] was used to validate the immunohistochemistry of SHGS (Uhlén et al., 2005; Uhlen et al., 2017).

[4] https://cytoscape.org/
[5] https://string-db.org/
[6] http://bioinfo.henu.edu.cn/EAC/EACList.jsp
[7] https://www.proteinatlas.org/

RESULTS

Physiognomies of Selected Studies

We collected a total of 682 studies from the GEO and ArrayExpress database up to October 2019. In all, 678 datasets/studies that did not satisfy the inclusion criteria were excluded. Finally, four potential studies were selected (**Supplementary Figure 1**). Among the four selected studies, three were conducted on the Affymetrix human genome U133A platform and one was performed on the Affymetrix human genome U133 plus 2.0 platform, which included 125 samples in total chosen in this study. In each study, EAC samples were compared with the adjacent BE samples. The dataset GSE37200 was used to construct a co-expression network with the relevant clinical trait information. After preprocessing and normalization, the GSE37200 dataset with 22,284 genes was further processed, and variant genes (12,701) were selected for WGCNA studies.

Identification of DEGs and Enrichment Analysis

In total, 403 DEGs were obtained through microarray meta-analysis, which include 169 downregulated and 234 upregulated genes. A heatmap is a simple yet effective way to compare the content of multiple major gene lists. Major DEGs across all the datasets were represented in red, orange, and yellow in a heatmap. Gray indicates that the respective gene is not present in the gene list (**Supplementary Figure 2**). **Table 2** illustrates the top 10 upregulated and downregulated DEGs. Monocyte differentiation antigen CD14 (CD14), ribose 5-phosphate isomerase A (RPIA), tumor necrosis factor superfamily member 11 (TNFSF11), plexin D1 (PLXND1), major histo-compatibility complex, class II DM beta (HLA-DMB), and spliceosome-associated factor 3, and U4/U6 recycling protein (SART3) were highly expressed upregulated genes, whereas fucosyltransferase 2 (FUT2), SECIS binding protein 2 like (SECISBP2L), COP9 signalosome subunit 4 (COPS4), gelsolin (GSN), and glutathione peroxidase 3 (GPX3) were highly expressed downregulated genes. According to the gene ontology (GO) terms BP, MF, and CC, downregulated genes were significantly enriched in the mitotic cell cycle process, sister chromatid segregation, antigen processing, presentation of peptide antigen *via* MHC class I, chromosomal region, and MHC class I protein binding, whereas retinol dehydrogenase activity and fucosyltransferase activity were highly enriched in upregulated genes associated with EAC (**Figures 2A–C**). In KEGG, pathway enrichment demonstrated that the upregulated genes were enriched for viral myocarditis, cell cycle, DNA replication, and AGE-RAGE signaling pathways in diabetic complications. Downregulated genes were associated with pathways involved in fatty acid degradation, glycosphingolipid biosynthesis, and amino sugar and nucleotide sugar metabolism (**Figure 2D**).

WGCNA and Clinically Significant Module Identification

A dendrogram of samples (GSE37200) with clinical trait was clustered using the average linkage method and Pearson's correlation method (**Figure 3A**). Co-expression analysis was carried out to construct the co-expression network. In this study, the power of $\beta = 9$ (scale-free $R^2 = 0.95$) was selected as the soft-thresholding parameter to ensure a scale-free network (**Figure 3B**). A dendrogram of all differentially expressed genes was clustered based on a dissimilarity measure (1-TOM) (**Supplementary Figure 3**). A total of 39 modules were identified through hierarchical

TABLE 2 | Top ten up- and downregulated genes.

S.No.	Gene	Gene name	Combined ES	P-value
Upregulated genes				
1	CD14	Monocyte differentiation antigen CD14	1.2948	4.98E-08
2	RPIA	Ribose 5-phosphate isomerase A	1.1376	4.63E-08
3	TNFSF11	Tumor necrosis factor super family member 11	1.1229	3.66E-08
4	PLXND1	Plexin D1	1.1223	3.60E-08
5	HLA-DMB	Major histo-compatibility complex, class II, DM beta	1.1168	3.97E-08
6	SART3	Spliceosome associated factor 3, U4/U6 recycling protein	1.1142	4.06E-08
7	OSBPL3	Oxysterol binding protein like 3	1.1119	4.79E-08
8	PRAF2	PRA1 domain family member 2	1.1117	4.61E-08
9	PILRB	Paired immunoglobin like type 2 receptor beta	1.1078	4.67E-08
10	RGS16	Regulator of G protein signaling 16	1.1064	4.79E-08
Downregulated genes				
1	FUT2	Fucosyltransferase 2	-1.1095	4.67E-08
2	SECISBP2L	SECIS binding protein 2 like	-1.111	4.52E-08
3	COPS4	COP9 signalosome subunit 4	-1.1148	4.79E-08
4	GSN	Gelsolin	-1.117	4.04E-08
5	GPX3	Glutathione peroxidase 3	-1.1185	3.87E-08
6	ADH1A	Alcohol dehydrogenase 1A (class I), alpha polypeptide	-1.1271	4.39E-08
7	CORO2A	Coronin 2A	-1.2074	4.67E-08
8	ACADS	Acyl-CoA dehydrogenase short chain	-1.2157	3.57E-08
9	RCAN2	Regulator of calcineurin 2	-1.2172	5.07E-08
10	CLEC3B	C-type lectin domain family 3 member B	-1.2316	3.73E-08

**Combined ES, Cochran's combination test of random effect model (REM) or effect size (ES).*

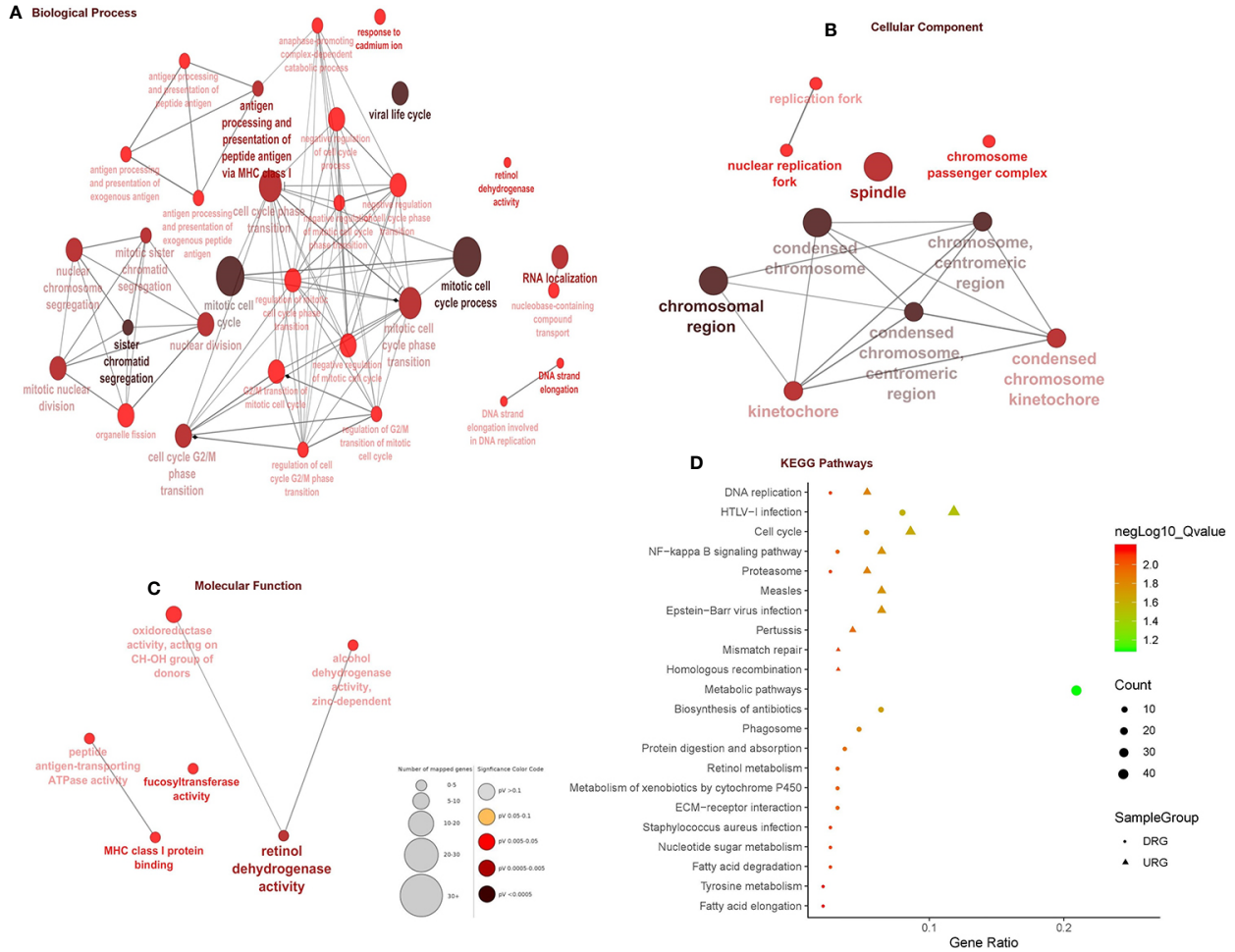


FIGURE 2 | Gene ontology and pathway enrichment analysis. **(A)** Biological process analysis. **(B)** Cellular component analysis. **(C)** Molecular function analysis. **(D)** KEGG pathway analysis.

clustering. Light green (eigengene value = 0.41), dark gray (eigengene value = 0.62) and Sienna3 (eigengene value = 0.46) modules appeared to have the highest association with age, gender, and ethnicity. There was no module–trait relationship associated with tumor stage, denoted as NA in **Figure 3C**. Therefore, the dark gray module having the highest association with gender was selected as the clinically significant module for further analysis. There were 207 phenotypic genes identified in the dark-gray module (**Figure 3D**). In **Supplementary Figure 4**, the hierarchical clustering dendrogram of the eigengene network represents the relationships among the modules and the clinical trait weight.

Identification and Validation of Hub Genes

The PPI network was constructed with 403 DEGs using the STRING database. The interactive relationships between the key genes in the whole network were determined using the Cytoscape plugins (MCODE, Cytoscape, and CytoHubba). There are two clusters: 82 nodes and 938 edges in cluster 1, and 20 nodes and 168 edges in cluster 2, which were identified from MCODE based on a scoring system (cutoff k-score = 12). In addition, the data were imported into another plugin, CytoHubba, which helped to identify 104 key genes through five different calculation methods, namely, EPC, MCC, DMNC, MNC, and Stress. Then, the two clusters were imported into the CytoNCA plugin, which helped to identify 40 key genes using five different algorithms, namely, betweeness, closeness, degree, eigenvector, and subgraph. We securely conceive that the key genes are the intersections between the PPI network and the dark-gray module with 207 genes (**Supplementary Table 1**) highly correlated with phenotype (gender) from the WGCNA analysis (**Figures 4A, B**). Finally, five SHGS, namely, pre-mRNA processing factor 4 (PRPF4), serine and arginine rich splicing factor 1 (SRSF1), heterogeneous nuclear ribonucleoprotein M (HNRNPM), DExH-box helicase 9 (DHX9), and origin recognition complex subunit 2 (ORC2), are identified between BE and EAC. Pathway enrichment demonstrated that all the SHGS are involved in the metabolism of RNA, and its molecular functional terms include cell cycle, DNA binding, DNA topoisomerase binding, pre-mRNA splicing, and RNA helicase activity (**Figure 5**).

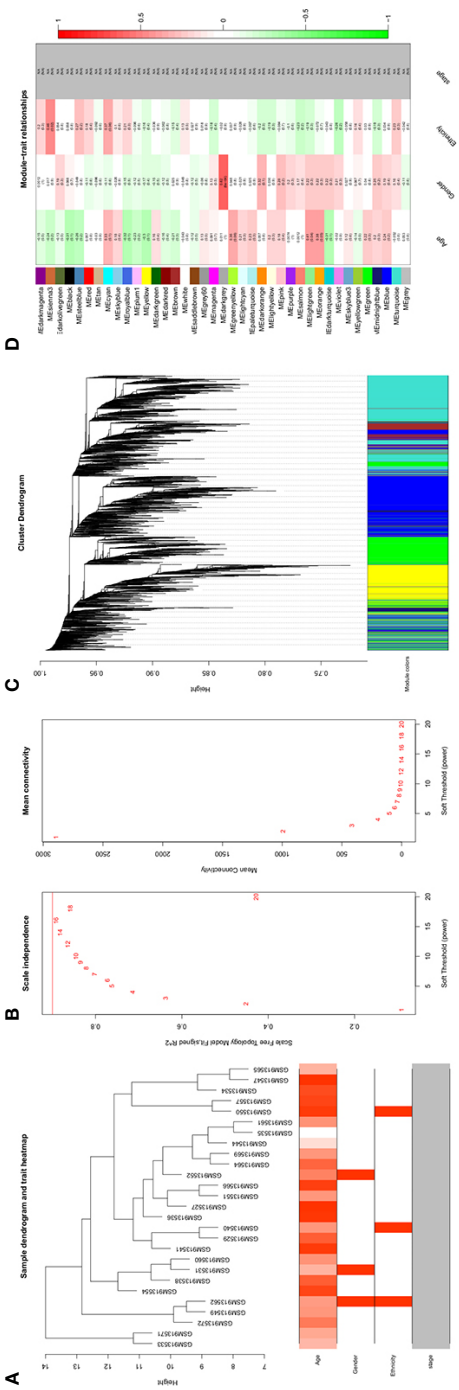


FIGURE 3 | (A) Clustering dendrogram of samples with trait heatmap. **(B)** Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis). **(C)** Clustering dendrogram of genes, with dissimilarity based on topological overlap, together with assigned module colors. **(D)** Module-trait associations: Each row corresponds to a module eigengene and each column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. Distribution of average gene significance and errors in the modules associated with gender and ethnicity of cancer patient.

Survival Analysis and Immunohistochemistry

Kaplan–Meier plots demonstrated that the prognostic impact of the SHGS was identified from modules of the PPI network complex and WGCNA. The results revealed that high expression of HNRNPM and SRSF1 was associated with poor overall survival of BE and EAC patients ($p < 0.05$). Moreover, high expression of PRPF4, DHX9, and ORC2 was correlated with longer overall survival of BE and EAC patients (**Figure 6**). In addition, we plotted a gender-based survival curve to determine the correlation of WGCNA modules. The hazard ratio (HR) and 95% confidence interval were as follows in males: PRPF4 (HR =1.08; 95%CI – 0.46 ± 2.48; $p = 0.865$); SRSF1 (HR =3.08; 95%CI – 1.49 ± 6.37; $p = 0.002$); HNRNPM (HR =3.295; 95%CI – 1.54 ± 7.02; $p = 0.002$); DHX9 (HR =1.39; 95%CI – 0.64 ± 2.48; $p = 0.404$); ORC2 (HR =1.25; 95%CI – 0.58 ± 2.72; $p = 0.564$). Further, in female cases PRPF4 (HR =0.39; 95%CI – 0.03 ± 3.89; $p = 0.421$); SRSF1 (HR =1.49; 95%CI – 0.20 ± 10.79; $p = 0.689$); HNRNPM (HR =8.06; 95%CI – 0.82 ± 79.01; $p = 0.073$); DHX9 (HR =0.38; 95%CI – 0.04 ± 3.89; $p = 0.424$); ORC2 (HR =3.24; 95% CI – 0.20 ± 51.91; $p = 0.4061$). The results clearly demonstrated that the high expression of SHGS correlated to the poor prognosis of male compared to female. Furthermore, immunohistochemical slides of the Human Protein Atlas database indicated that the protein expressions of SHGS were drastically higher in cancerous tissues compared with in adjacent normal tissues, as shown in **Figure 7**. Therefore, these SHGS were all key genes that play an initiative role and might have a tendency to co-express.

DISCUSSION

EAC is an obstinate type of cancer, which has a high mortality rate because of poor prognosis, metastatic rate, and treatment resistance (Tatarian and Palazzo, 2019). EAC usually arises from a premalignant variation in the lining of the esophagus known as BE (Thrift, 2016). Unfortunately, the treatment and diagnosis of EAC and BE are limited due to the lack of precise molecular targets. Therefore, we designed this study to explore SHGS between EAC and BE to improve the diagnosis and prognosis status of the patients. There are numerous advanced technologies that can quantify the enormous amount of transcripts in a parallel manner. Microarray and data mining are well-known approaches for cancer biomarker discovery (Selvaraj et al., 2019). Nevertheless, a single microarray dataset is not enough to deal with this obstinate disease. However, a comprehensive analysis of a number of microarray datasets with different platforms can assist with identifying the molecular mechanism of EAC and BE. Therefore, we selected four different microarray datasets to identify SHGS and the associated pathways between BE and EAC. Moreover, WGCNA is a powerful tool for searching effective biological mechanisms and key genes from gene expression microarrays. It provides module construction and correlation analysis within the gene expression data to determine the associations between genes. It also elucidates the biological significance of a gene module to provide insights into molecular and pathological characteristics in many diseases. All these characteristics make it a robust, reliable, and significant

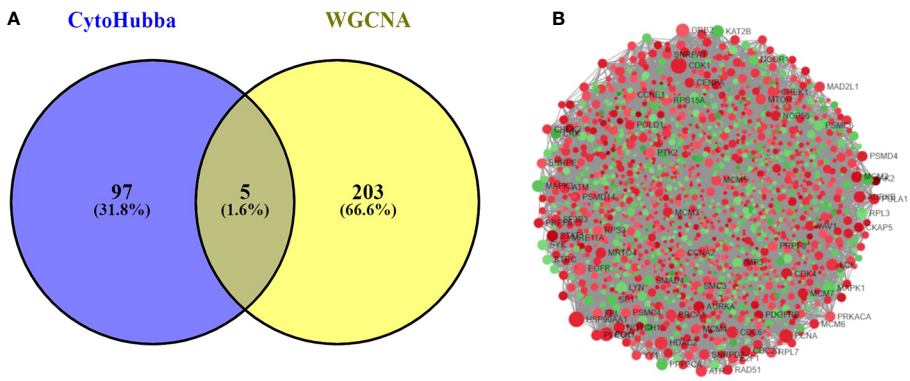


FIGURE 4 | (A) Venn diagram demonstrates overlapping genes of the DEG-PPI network and WGCNA. **(B)** DEG-PPI network complex (upregulated genes showed in green; downregulated genes showed in red).

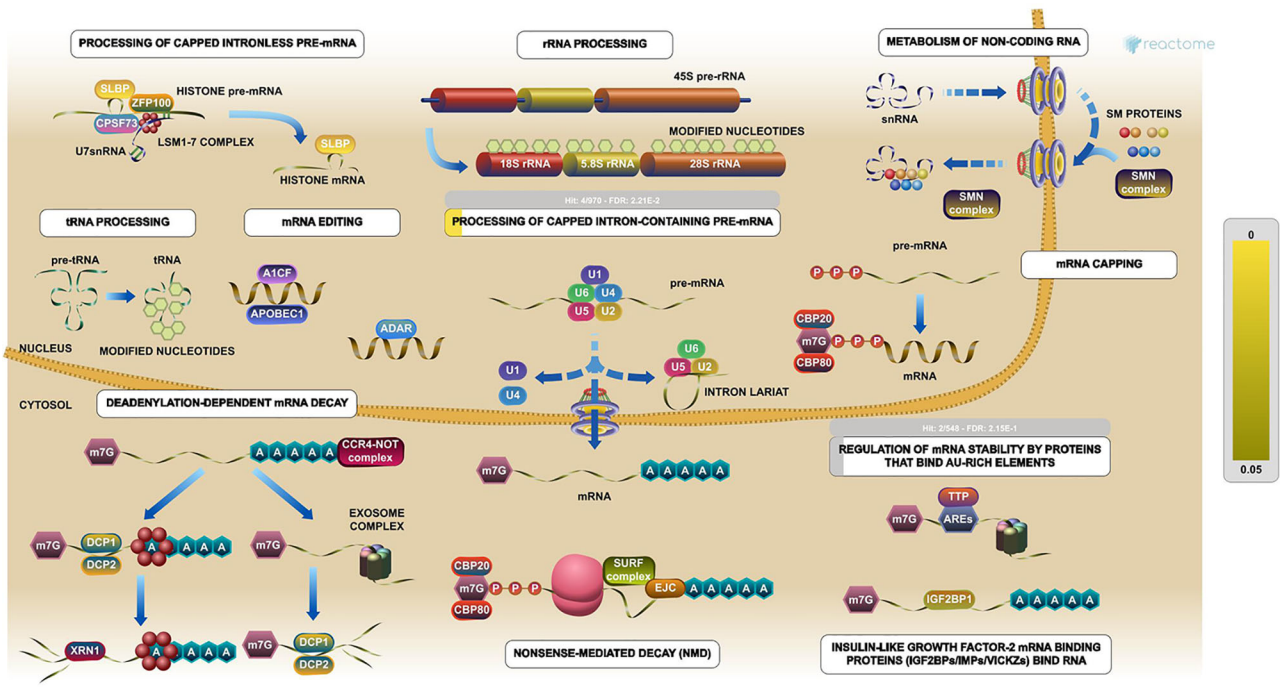


FIGURE 5 | Pathway enrichments of SHGS. PRPF4, SRSF1, HNRNPM, and DHX9 are keys genes in the RNA metabolic pathway. These genes especially are involved in the preprocessing of capped intron-containing pre-mRNA and regulation of mRNA stability by proteins that bind AU-rich elements. (Image extracted from the Reactome pathway analyzer).

method for analysis of large-scale data. There is no prior research employing WGCNA to do gene co-expression network analysis with BE and EAC. To explore SHGS, we decided to construct a gene co-expression network with relevant clinical trait information from the GSE37200 dataset.

Phenotype variants like age, gender, and ethnicity are factors that are intensively involved in the prognosis and diagnosis of BE and EAC (Ford et al., 2005; Runge et al., 2015). EAC usually appears at the later stage of life, but it may start at a young age in the form of BE. Earlier studies have reported that male patients with BE are at low risk of malignant progression and predominantly die due to causes other than EAC (Sikkema et al., 2010). There are studies reported that there is a marked male prevalence of EAC with a male-to-female ratio of 9:1 due to sex hormone factors. Androgen exposure may increase the risk of EAC compared to estrogen (Xie and Lagergren, 2016; Kim et al., 2016). Furthermore, geographically

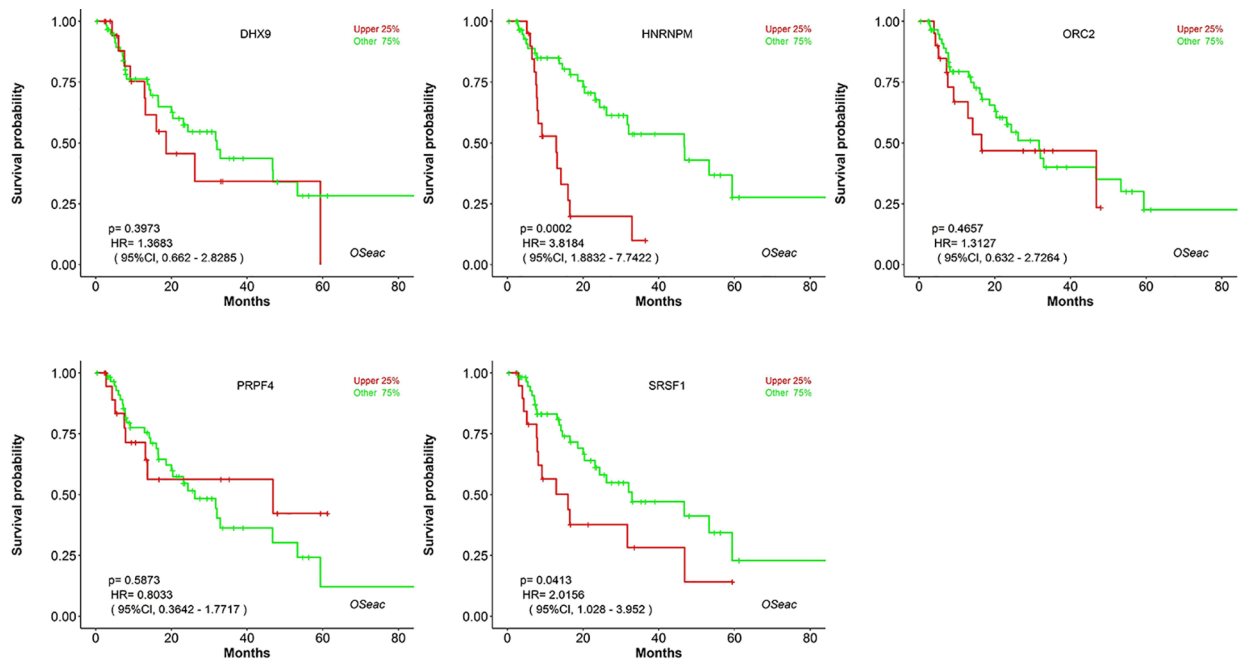


FIGURE 6 | The prognostic value of hub genes in BE and EAC patients (Kaplan–Meier plot).

White people, especially White Americans, are at higher risk than other ethnicities (Schneider and Corley, 2015). A comprehensive study from January 2006 to December 2017 reported that high risk of male patients with esophageal diseases in the province of Madinah in Saudi Arabia is due to a variety of factors, including inflammatory disorders, infection, and neoplastic condition (Albasri et al., 2019). In addition, genomic analysis by restriction fragment length polymorphism indicated that the highest frequencies of Y-chromosomal haplogroups are associated with BE and EAC in White males (Westra et al., 2020). Recent case–control studies demonstrate that gastroesophageal reflux disease in male patients is highly associated with the development of BE in Germany (Schmidt et al., 2020). These reports supported the present results, indicating that predicted dark gray modules with the highest association with gender must have a clinically significant module.

Two types of biological materials, namely, GO and KEGG pathway data, are key to understanding the disease mechanism. CD14 acts as a co-receptor with toll-like receptors (TLRs) to identify evading pathogens and improve the immune system. It is reported that TLRs 1-10 are expressed in the normal esophagus and that there is a high association of TLRs 4, 5, and 9 with BE and EAC (Kauppila and Selander, 2014). TNFSF11 is a key regulator of interactions between T cells and dendritic cells, which regulate the T-cell-dependent immune response and enhance bone-resorption in hypercalcemia of malignancy (Luan et al., 2012). Somja et al. (2013) observed that both metaplastic and malignant lesions of the esophagus are infiltrated by regulatory T cells. They concluded that soluble factors secreted by epithelial cells during the EAC or BE influence tumor progression through tolerogenic dendritic cells, which can be a potential therapeutic tool. In addition, different cohort studies have reported that GSN is a serum glycoprotein biomarker used as a diagnostic tool for EAC and BE (Shah et al., 2015; Shah et al., 2018). Glycosphingolipid biosynthesis is an important pathway that can produce cell-surface glycans. These glycans are altered in the development from BE into EAC, with specific changes in lectin binding patterns. This binding is a key marker in endoscopic visualization of high-grade dysplastic lesions (Bird-Lieberman et al., 2012). These reports suggest that the predicted GO terms and pathways of DEGs are highly associated with EAC and BE.

We have identified five different SHGS (PRPF4, SRSF1, HNRNPM, DHX9, and ORC2) between EAC and BE. PRPF4, SRSF1, and HNRNPM are U4/U6 small nuclear ribonucleoprotein Prp4, serine and arginine-rich splicing factor 1, and heterogeneous nuclear ribonucleoprotein M coding genes, respectively. These genes play an important role in pre-mRNA splicing and spliceosome assembly (Bertram et al., 2017). Pre-mRNA splicing is key to the pathology and has a substantial role in generating multiple oncogenic and tumor-suppressor proteins after the post-transcriptional process. Splicing is of different types such as amino acid addition, exon skipping, frame shift, intron retention, promoter usage, truncated C-terminus, and 5′-SS, which have various clinical applications including proliferation, metastasis, drug resistance, and radiotherapy (Guo et al., 2015; Di et al., 2019). In addition, there are studies reporting splicing signatures associated with the prognosis of

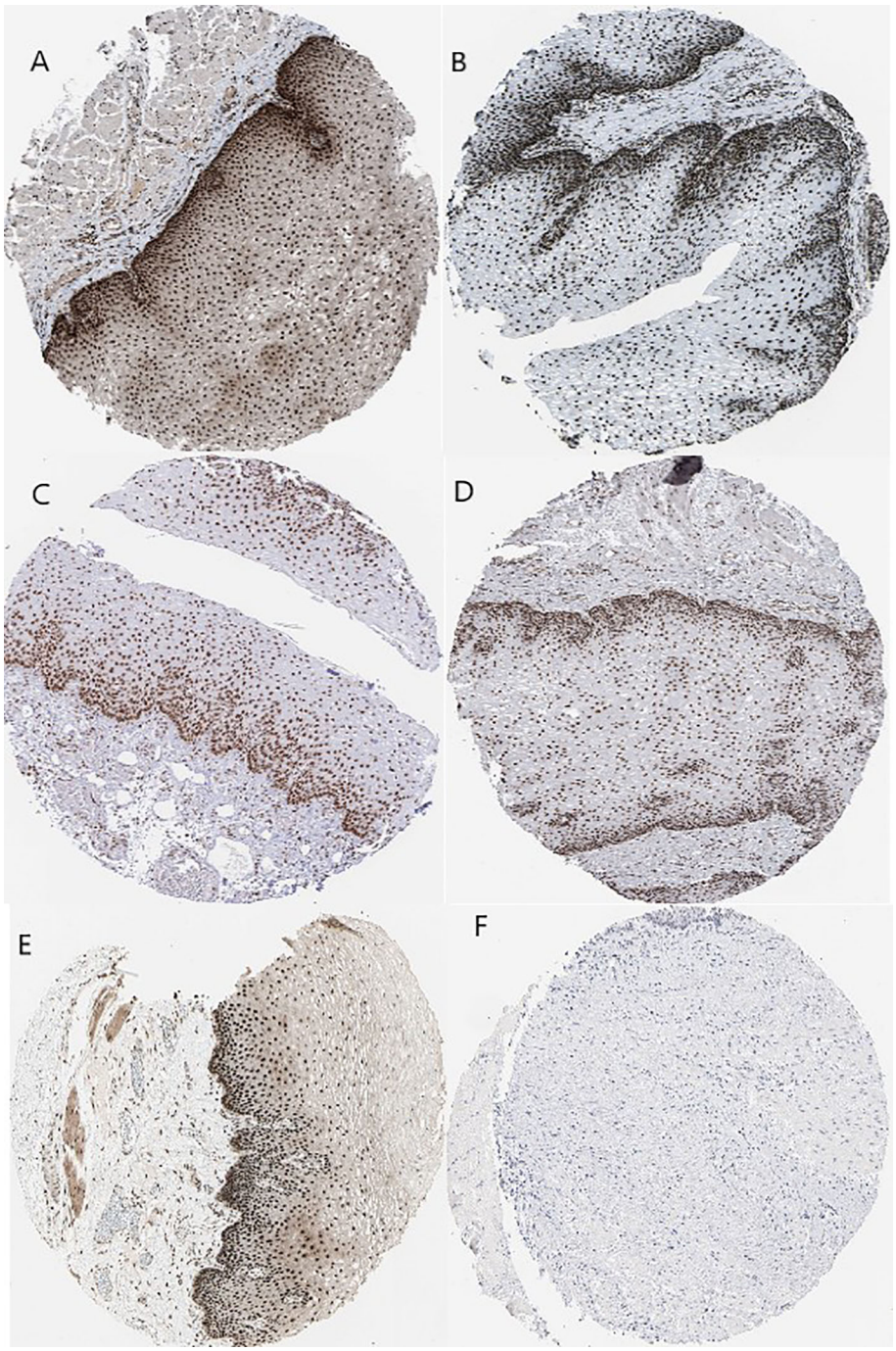


FIGURE 7 | Immunohistochemistry of the five hub genes based on the Human Protein Atlas. **(A)** Protein levels of PRPF4 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(B)** Protein levels of SRSF1 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(C)** Protein levels of SRSF1 in tumor tissue (staining: high; intensity: strong; quantity: >75%). **(D)** Protein levels of HNRNPM in normal tissue (staining: high; intensity: strong; quantity: >75%). **(E)** Protein levels of DHX9 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(F)** Protein levels of ORC2 in normal tissue (staining: not detected; intensity: low; quantity: <25%).

esophageal cancer (Lin et al., 2018; Mao et al., 2019). Through the splicing mechanism, PRPF4, SRSF1, and HNRNPM regulate the cell proliferation, migration, and invasion in different cancers, including lung cancer (Choi, 2012; Chang and Lin, 2019), breast cancer (Anczuków et al., 2015; Sun et al., 2017; Park et al., 2019), cutaneous squamous cell carcinoma (Zhang et al., 2018), hepatocellular carcinoma (Tu et al., 2019), esophagus dysplasia (Varghese et al., 2015; Fitzgerald et al., 2018), gastric cancer (Wu

et al., 2019), cervical cancer (Dong et al., 2019), and Ewing's sarcoma (Passacantilli et al., 2017).

DHX9 is an ATP-dependent RNA helicase A coding gene involved in DNA replication, transcriptional activation, post-transcriptional RNA regulation, mRNA translation, and RNA-mediated gene silencing (Capitanio et al., 2017). Knockdown of ATP-dependent RNA helicase inhibited the expression of β-catenin, c-Myc, and cyclin D1 in esophageal cancer cells through suppressing the Wnt/β-catenin signaling pathway (Ma et al., 2017). In addition, ATP-dependent RNA helicase was reported to dysregulate distinct steps of mRNA and pre-ribosomal RNA metabolism in cancer cells (Awasthi et al., 2018). ORC2 is an origin recognition complex subunit 2 coding gene binding origins of replication (Shen et al., 2012). It can bind to different histone trimethylation proteins and stabilize leucine-rich repeat and WD repeat-containing protein 1 (LRWD1) through protecting it from ubiquitin-mediated proteasomal degradation (Chan and Zhang, 2012). Studies demonstrated that increased expressions of certain histone-mediated proteins correlate with advanced TNM stages, tumor grade, metastatic potential, and decreased overall and disease-free survival of patients with esophageal cancer (Schizas et al., 2018). This supportive information enhances the understanding of why the predicted DHX9, HNRNPM, ORC2, PRPF4, and SRSF1 genes are highly correlated to EAC and BE progression and act as potential biomarkers for diagnosis as well as prognosis.

CONCLUSION

This network pharmacology-based study provides new insights into BE and EAC patients for their diagnosis and prognosis. The results of microarray dataset-based PPI networks and WGCNA exhibited that the dark-gray module had the maximum association with EAC and BE, with the identification of five SHGS, namely PRPF4, SRSF1, HNRNPM, DHX9, and ORC2. The WGCNA-based gene co-expression network indicated that the relationships between co-expressed genes and clinical trait (gender of the patient) were associated with the progression of esophageal cancer. SHGS enrichment denotes that the RNA metabolic and spliceosome pathways play an essential role in the development and progress of esophageal cancer. Survival analysis demonstrates that the high expression of HNRNPM and SRSF1 in esophageal cancer might be a poor prognostic marker. The co-expression modules were established to preserve a reliable expression relationship independent of phenotype and may share similar biological functions. This approach shares the limitations of other data mining methods: the results of WGCNA can technically be biased due to tissue contamination or artifacts. To enhance the reliability of the WGCNA results, immuno-histochemical data from the Human Protein Atlas were used for confirmation. However, we could not obtain all the related IHC data of tumor and adjacent normal samples for each gene due to the database constraint. These findings may support new therapeutic targets and potential useful for the advancement of prognostic biomarker evaluation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Publicly available datasets were analyzed in this study. This data can be found at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26886 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1420 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37200 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37201 https://portal.gdc.cancer.gov/exploration?filters=%7B%22op%22%3A%22and%22%2C%22content%22%3A%5B%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22cases.primary_site%22%2C%22value%22%3A%5B%22esophagus%22%5D%7D%7D%5D%7D https://www.proteinatlas.org/ENSG00000136875-PRPF4/tissue/esophagus.

https://www.proteinatlas.org/ENSG00000136450-SRSF1/tissue/esophagus https://www.proteinatlas.org/ENSG00000099783-HNRNPM/tissue/esophagus https://www.proteinatlas.org/ENSG00000135829-DHX9/tissue/esophagus

https://www.proteinatlas.org/ENSG00000115942-ORC2/tissue/esophagus.

AUTHOR CONTRIBUTIONS

DW, GS, and AN conceived and designed the experiment. AN, AK, and SK performed the meta-analysis and network analysis. AN and GS performed the WGCNA. AN, AK, and GS aided in preparation of the figures and supplemental materials. AN and SK performed Cytoscape analysis. AN, SK and GS compiled the manuscript. WC and DW provided expert advice in the study, performed editing and proofreading of the manuscript. All authors contributed to the study and approved the submitted version.

FUNDING

The authors are thankful for financial support from the Ministry of Science and Technology of China (Grant No.: 2016YFA0501703), the National Natural Science Foundation of China (Contract No. 61832019, 61503244), the Science and Technology Commission of Shanghai Municipality (Grant: 19430750600), Henan Natural Science (Grant No.:162300410060), as well as SJTU JiRLMDS Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14) to DQ.W; Doctoral Study Scholarship from Shanghai Jiao Tong University to ASN; China Postdoctoral Science Foundation (Grant No.: 2018M632766) to GS; Henan Province Postdoctoral Science Foundation (Grant No.: 001802029 & 001803035) to SK and GS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.00881/full#supplementary-material

REFERENCES

Albasri, A. M., Ansari, I. A., Anjum, S., Elsawaf, Z. M., and Alhujaily, A. S. (2019). Pattern of oesophageal diseases in Madinah region, Saudi Arabia: An 11 years experience. *J. Pak Med. Assoc.* 69 (9), 1365–1368.

Analysis, Working Group, University Asan and Cancer Genome Atlas Research Network (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature* 541 (7636), 169. doi: 10.1038/nature20805

Anaparthy, R., and Sharma, P. (2014). Progression of Barrett oesophagus: role of endoscopic and histological predictors. *Nat. Rev. Gastroenterol. Hepatol.* 11 (9), 525. doi: 10.1038/nrgastro.2014.69

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Chen, S., Shirole, N. H., et al (2015). SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell* 60 (1), 105–117. doi: 10.1016/j.molcel.2015.09.005

Awasthi, S., Verma, M., Mahesh, A., Khan, M. I. K., Gayathri,G., Rajavelu, A., et al. (2018). DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Res.* 46 (12), 6304–6317. doi: 10.1093/nar/gky231

Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., et al. (2017). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* 170 (4), 701–713. doi: 10.1016/j.cell.2017.07.011

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi: 10.1093/bioinformatics/btp101

Bird-Lieberman, E. L., Neves, AndréA., Lao-Sirieix, P., O'donovan, M., Novelli, M., Lovat, L. B., et al. (2012). Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nat. Med.* 182, 315. doi: 10.1038/nm.2616

Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., and Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinf.* 20 (1), 446. doi: 10.1186/s12859-019-3036-6

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. Clin.* 68 (6), 394–424. doi: 10.3322/caac.21492

Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., et al. (2003). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31 (1), 68–71. doi: 10.1093/nar/gkg091

Campain, A., and Yang, Y. H. (2010). Comparison study of microarray meta-analysis methods. *BMC Bioinf.* 111, 408. doi: 10.1186/1471-2105-11-408

Capitanio, J. S., Montpetit, B., and Wozniak, R. W. (2017). Human Nup98 regulates the localization and activity of DExH/D-box helicase DHX9. *Elife* 6, e18825. doi: 10.7554/eLife.18825

Chai, T., Shen, Z., Zhang, P., Lin, Y., Chen, S., Zhang, Z., et al. (2019). Comparison of high risk factors (hot food, hot beverage, alcohol, tobacco, and diet) of esophageal cancer: A protocol for a systematic review and meta-analysis. *Medicine* 98 (17), e15176. doi: 10.1097/MD.0000000000015176

Chan, K. M., and Zhang, Z. (2012). Leucine-rich repeat and WD repeat-containing protein 1 is recruited to pericentric heterochromatin by trimethylated lysine 9 of histone H3 and maintains heterochromatin silencing. *J. Biol. Chem.* 287 (18), 15024–15033. doi: 10.1074/jbc.M111.337980

Chang, H.-L., and Lin, J.-C. (2019). SRSF1 and RBM4 differentially modulate the oncogenic effect of HIF-1α in lung cancer cells through alternative splicing mechanism. *BiochimicaetBiophysicaActa (BBA)-Mol. Cell Res.* 1866 (12), 118550. doi: 10.1016/j.bbamcr.2019.118550

Chen, S.-H., Chin, C.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2009). "cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology", in *20th International Conference on Genome Informatics*. (Pacifico Yokohama, Japan: Imperial College Press).

Choi, S.-Y. (2012). Identification of PRPF4 as a novel cancer promoter through AKT signaling in lung cancer. *AACR; Cancer Res.* 72 (8 Suppl), 4174–4174. doi: 10.1158/1538-7445.AM2012-4174

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika* 373/4, 256–266. doi: 10.1093/biomet/37.3-4.256

Cook, M. B., Corley, D. A., Murray, L. J., Liao, L. M., Weimin Ye, F., Gammon, M. D., et al. (2014). Gastroesophageal reflux in relation to adenocarcinomas of the esophagus: a pooled analysis from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON). *PloS One* 9 (7), e103508. doi: 10.1371/journal.pone.0103508

Davies, A. R., Gossage, J. A., Zylstra, J., Mattsson, F., Lagergren, J., Maisey, N., et al. (2014). Tumor stage after neoadjuvant chemotherapy determines survival after surgery for adenocarcinoma of the esophagus and esophagogastric junction. *J. Clin. Oncol.* 32 (27), 2983–2990. doi: 10.1200/JCO.2014.55.9070

Di, C., Zhang, Q., Chen, Y., Wang, Y., Zhang, X., Liu, Y., et al. (2019). Function, clinical application, and strategies of Pre-mRNA splicing in cancer. *Cell Death Differ.* 267, 1181–1194. doi: 10.1038/s41418-018-0231-3

Dong, M., Dong, Z., Zhu, X., Zhang, Y., and Song, L. (2019). Long non-coding RNA MIR205HG regulates KRT17 and tumor processes in cervical cancer via interaction with SRSF1. *Exp. Mol. Pathol.* 111, 104322. doi: 10.1016/j.yexmp.2019.104322

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30 (1), 207–210. doi: 10.1093/nar/30.1.207

Fitzgerald, R. C., Varghese, S., Newton, R., and Wernisch, L. "Methods and means for dysplasia analysis." U.S. Patent Application 15/747,117, filed July 26, 2018.

Ford, A. C., Forman, D., Reynolds, P.D., Cooper, B. T., and Moayyedi, P. (2005). Ethnicity, gender, and socioeconomic status as risk factors for esophagitis and Barrett's esophagus. *Am. J. Epidemiol.* 162 (5), 454–460. doi: 10.1093/aje/kwi218

Guo, W., Wang, C., Guo, Y., Shen, S., Guo, X., Kuang, G., et al. (2015). RASSF5A, a candidate tumor suppressor, is epigenetically inactivated in esophageal squamous cell carcinoma. *Clin. Exp. Metastasis* 32 (1), 83–98. doi: 10.1007/s10585-015-9693-6

Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13 (2), 204–216. doi: 10.1093/biostatistics/kxr054

Kauppila, J. H., and Selander, K. S. (2014). Toll-like receptors in esophageal cancer. *Front. Immunol.* 5, 200. doi: 10.3389/fimmu.2014.00200

Kim, Y. S., Kim, N., and Gwang Ha, K. (2016). Sex and gender differences in gastroesophageal reflux disease. *J. Neurogastroenterol. Motil.* 22 (4), 575. doi: 10.5056/jnm16138

Kimchi, E. T., Posner, M. C., Park, J. O., Darga, T. E., Kocherginsky, M., Karrison, T., et al. (2005). Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. *Cancer Res.* 65 (8), 3146–3154. doi: 10.1158/0008-5472.CAN-04-2490

Kohl, M., Wiese, S., and Warscheid, B. (2011). "Cytoscape: software for visualization and analysis of biological networks." in *Data mining in proteomics* (Humana Press), 291–303.

Konishi, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations. *Ann. Inst. Stat. Math.* 37 (1), 87–94. doi: 10.1007/BF02481082

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9 (1), 559. doi: 10.1186/1471-2105-9-559

Lin, J., Myers, A. L., Wang, Z., Nancarrow, D. J., Ferrer-Torres, D., Handlogten, A., et al. (2015). Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. *Oncotarget* 6 (26), 22239. doi: 10.18632/oncotarget.4161

Lin, P., He, R. Q., Ma, F. C., Liang, L., He, Y., Yang, H., et al. (2018). Systematic analysis of survival-associated alternative splicing signatures in gastrointestinal pan-adenocarcinomas. *EBioMedicine* 34, 46–60. doi: 10.1016/j.ebiom.2018.07.040

Lordick, F., Mariette, C., Haustermans, K., Obermannová, R., and Arnold, D. (2016). Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 27, suppl_5, v50–v57. doi: 10.1093/annonc/mdw329

Luan, X., Lu, Q., Jiang, Y., Zhang, S., Wang, Q., Yuan, H., et al. (2012). Crystal structure of human RANKL complexed with its decoy receptor osteoprotegerin. *J. Immunol.* 189 (1), 245–252. doi: 10.4049/jimmunol.1103387

Ma, Z., Feng, J., Guo, Y., Kong, R., Ma, Y., Sun, L., et al. (2017). Knockdown of DDX5 inhibits the proliferation and tumorigenesis in esophageal cancer. *Oncol. Res. Featuring Preclinical Clin. Cancer Ther.* 25 (6), 887–895. doi: 10.3727/096504016X14817158982636

Mao, S., Li, Y., Lu, Z., Che, Y., Sun, S., Huang, J., et al. (2019). Survival-associated alternative splicing signatures in esophageal carcinoma. *Carcinogenesis* 40 (1), 121–130. doi: 10.1093/carcin/bgy123

Montgomery, E. A., Basman, F. T., Brennan, P., and Malekzadeh, R. (2014). Oesophageal cancer. *World Cancer Rep.* 15, 528–543.

Park, S., Han, S.-H., Kim, H.-G., Jeong, J., Choi, M., Kim, H.-Y., et al. (2019). PRPF4 is a novel therapeutic target for the treatment of breast cancer by influencing growth, migration, invasion, and apoptosis of breast cancer cells via p38 MAPK signaling pathway. *Mol. Cell. Probes* 47, 101440. doi: 10.1016/j.mcp.2019.101440

Passacantilli, I., Frisone, P., De Paola, E., Fidaleo, M., and Paronetto, M. P. (2017). hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. *Nucleic Acids Res.* 45 (21), 12270–12284. doi: 10.1093/nar/gkx831

Ritchie, M. E., Phipson, B., Wu, Di, Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47–e47. doi: 10.1093/nar/gkv007

Runge, T. M., Abrams, J. A., and Shaheen, N. J. (2015). Epidemiology of Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterol. Clinics* 44 (2), 203–231. doi: 10.1016/j.gtc.2015.02.001

Rustgi, A. K., and El-Serag, H. B. (2014). Esophageal carcinoma. *New Engl. J. Med.* 371 (26), 2499–2509. doi: 10.1056/NEJMra1314530

Sadhu, A., and Bhattacharyya, B. (2019). Common subcluster Mining in Microarray Data for molecular biomarker discovery. *Interdiscip. Sci.: Comput. Life Sci.* 11 (3), 348–359. doi: 10.1007/s12539-017-0262-3

Schizas, D., Mastoraki, A., Naar, L., Spartalis, E., Tsilimigras, D. I., Karachaliou, G.-S., et al. (2018). Concept of histone deacetylases in cancer: Reflections on esophageal carcinogenesis and treatment. *World J. Gastroenterol.* 2441, 4635. doi: 10.3748/wjg.v24.i41.4635

Schmidt, M., Ankerst, D. P., Chen, Y., Wiethaler, M., Slotta-Huspenina, J., Becker, K.-F., et al. (2020). Epidemiologic Risk Factors in a Comparison of a Barrett Esophagus Registry (BarrettNET) and a Case–Control Population in Germany. *Cancer Prev. Res.* 13 (4), 377–384. doi: 10.1158/1940-6207.CAPR-19-0474

Schneider, J. L., and Corley, D. A. (2015). A review of the epidemiology of Barrett's oesophagus and oesophageal adenocarcinoma. *Best Pract. Res. Clin. Gastroenterol.* 29 (1), 29–39. doi: 10.1016/j.bpg.2014.11.008

Selvaraj, G., Kaliamurthi, S., Kaushik, A. C., Khan, A., Wei, Y.-K., Cho, W. C., et al. (2018). Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. *J. Biomed. Inf.* 86, 120–134. doi: 10.1016/j.jbi.2018.09.004

Selvaraj, G., Satyavani, K., Lin, S., Gu, K., and Wei, D.-Q. (2019). Prognostic impact of tissue inhibitor of metalloproteinase-1 in non-small cell lung cancer: Systematic review and meta-analysis. *Curr. Med. Chem.* 26, 7694. doi: 10.2174/0929867325666180904114455

Shah, A. K., Lê Cao, K.-A., Choi, E., Chen, D., Gautier, Benoît, Nancarrow, D., et al. (2015). Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma. *Mol. Cell. Proteomics* 14 (11), 3023–3039. doi: 10.1074/mcp.M115.050922

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T, Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome. Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shah, A. K., Gunter, H., Brown, I., Winterford, C., Na, R., Lê Cao, K.-A., et al. (2018). Evaluation of Serum Glycoprotein Biomarker Candidates for Detection of Esophageal Adenocarcinoma and Surveillance of Barrett's Esophagus. *Mol. Cell. Proteomics* 17 (12), 2324–2334. doi: 10.1074/mcp.RA118.000734

Shen, Z., Chakraborty, A., Jain, A., Giri, S., Ha, T., Prasanth, K. V., et al. (2012). Dynamic association of ORCA with prereplicative complex components regulates DNA replication initiation. *Mol. Cell. Biol.* 32 (15), 3107–3120. doi: 10.1128/MCB.00362-12

Sikkema, M., De Jonge, P. J. F., Steyerberg, E. W., and Kuipers, E. J. (2010). Risk of esophageal adenocarcinoma and mortality in patients with Barrett's esophagus: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* 8 (3), 235–244. doi: 10.1016/j.cgh.2009.10.010

Silvers, A. L., Lin,L., Bass, A. J., Chen, G., Wang, Z., Thomas, D. G., et al. (2010). Decreased selenium-binding protein 1 in esophageal adenocarcinoma results from posttranscriptional and epigenetic regulation and affects chemosensitivity. *Clin. Cancer Res.* 16 (7), 2009–2021. doi: 10.1158/1078-0432.CCR-09-2801

Somja, J., Demoulin, S., Roncarati, P., Herfs, Michaël, Bletard, N., Delvenne, P., et al. (2013). Dendritic cells in Barrett's esophagus carcinogenesis: an inadequate microenvironment for antitumor immunity? *Am. J. Pathol.* 182 (6), 2168–2179. doi: 10.1016/j.ajpath.2013.02.036

Sun, H., Liu, T., Zhu, D., Dong, X., Liu, F., Liang, X., et al. (2017). HnRNPM and CD44s expression affects tumor aggressiveness and predicts poor prognosis in breast cancer with axillary lymph node metastases. *Genes Chromosomes Cancer* 56 (8), 598–607. doi: 10.1002/gcc.22463

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937

Tang, Y., Li, M., Wang, J., Pan, Yi, and Wu, F.-X. (2015). CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* 127, 67–72. doi: 10.1016/j.biosystems.2014.11.005

Tatarian, T., and Palazzo, Francisco (2019). "Epidemiology, Risk Factors, and Clinical Manifestations of Esophageal Cancer," in *Shackelford"s Surgery of the Alimentary Tract, 2 Volume Set* (Philadelphia, United States: Elesvier), 362–367.

Thrift, A. P. (2016). The epidemic of oesophageal carcinoma: where are we now? *Cancer Epidemiol.* 41, 88–955. doi: 10.1016/j.canep.2016.01.013

Tramontano, A. C., Sheehan, D. F., Yeh, J. M., Kong, C. Y., Dowling, E. C., Rubenstein, J. H., et al. (2017). The impact of a prior diagnosis of Barrett's esophagus on esophageal adenocarcinoma survival. *Am. J. Gastroenterol.* 1128, 1256. doi: 10.1038/ajg.2017.82

Tu, J., Chen, J., He, M., Tong, H., Liu, H., Zhou, B., et al. (2019). Bioinformatics analysis of molecular genetic targets and key pathways for hepatocellular carcinoma. *OncoTargets Ther.* 12, 5153. doi: 10.2147/OTT.S198802

Uhlén, M., Björling, E., Agaton, C., Al-KhaliliSzigyarto, C., Amini, B., Andersen, E., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* 4 (12), 1920–1932. doi: 10.1074/mcp.M500279-MCP200

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357 (6352), eaan2507. doi: 10.1126/science.aan2507

Varghese, S., Newton, R., Ross-Innes, C. S., Lao-Sirieix, P., Krishnadath, K. K., O'Donovan, M., et al. (2015). Analysis of dysplasia in patients with Barrett's esophagus based on expression pattern of 90 genes. *Gastroenterology* 149 (6), 1511–1518. doi: 10.1053/j.gastro.2015.07.053

Wang, Q., Ma, C., and Kemmner, W. (2013). Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma. *BMC Cancer* 13 (1), 137. doi: 10.1186/1471-2407-13-137

Wang, Q., Yan, Z., Ge, L., Li, N., Yang, M., Sun, X., et al. (2020). OSeac: An Online Survival Analysis Tool for Esophageal Adenocarcinoma. *Front. Oncol.* 10:3155. doi: 10.3389/fonc.2020.00315

Wei, D.-Q., Selvaraj, G., and Kaushik, A. C. (2018). Computational Perspective on the Current State of the Methods and New Challenges in Cancer Drug Discovery. *Curr. Pharm. Des.* 24 (32), 3725. doi: 10.2174/1381612824321901091053 39

Westra, W. M., Rygiel, A. M., Mostafavi, N., de Wit, G. M. J., Roes, A. L., Moons, L. M. G., et al. (2020). The Y-chromosome F haplogroup contributes to the development of Barrett's esophagus-associated esophageal adenocarcinoma in a white male population. *Dis. Esophagus*. doi: 10.1093/dote/doaa011

Wu, Z.-H., Liu, C.-C., Zhou, Y.-Q., Hu, L.-N., and Guo, W.-J. (2019). OnclncRNA-626 promotes malignancy of gastric cancer via inactivated the p53 pathway through interacting with SRSF1. *Am. J. Cancer Res.* 9 (10), 2249.

Xia, J., Gill, E. E., and Hancock, R. E. W. (2015). NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* 10 (6), 823. doi: 10.1038/nprot.2015.052

Xie, S.-H., and Lagergren, J. (2016). The male predominance in esophageal adenocarcinoma. *Clin. Gastroenterol. Hepatol.* 14 (3), 338–347. doi: 10.1016/j.cgh.2015.10.005

Zhang, L., Qin, H., Wu, Z., Chen, W., and Zhang, G. (2018). Pathogenic genes related to the progression of actinic keratoses to cutaneous squamous cell carcinoma. *Int. J. Dermatol.* 57 (10), 1208–1217. doi: 10.1111/ijd.14131

Zhu, Y., Qiu, P., and Ji, Y. (2014). TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* 11 (6), 599–600. doi: 10.1038/nmeth.2956

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nangraj, Selvaraj, Kaliamurthi, Kaushik, Cho and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

REVIEW
published: 12 August 2020
doi: 10.3389/fphar.2020.01177

Artificial Intelligence to Decode Cancer Mechanism: Beyond Patient Stratification for Precision Oncology

Sandip Kumar Patel[1,2], *Bhawana George*[3] *and Vineeta Rai*[4]*

[1] Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai, India, [2] Buck Institute for Research on Aging, Novato, CA, United States, [3] Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [4] Department of Entomology & Plant Pathology, North Carolina State University, Raleigh, NC, United States

The multitude of multi-omics data generated cost-effectively using advanced high-throughput technologies has imposed challenging domain for research in Artificial Intelligence (AI). Data curation poses a significant challenge as different parameters, instruments, and sample preparations approaches are employed for generating these big data sets. AI could reduce the fuzziness and randomness in data handling and build a platform for the data ecosystem, and thus serve as the primary choice for data mining and big data analysis to make informed decisions. However, AI implication remains intricate for researchers/clinicians lacking specific training in computational tools and informatics. Cancer is a major cause of death worldwide, accounting for an estimated 9.6 million deaths in 2018. Certain cancers, such as pancreatic and gastric cancers, are detected only after they have reached their advanced stages with frequent relapses. Cancer is one of the most complex diseases affecting a range of organs with diverse disease progression mechanisms and the effectors ranging from gene-epigenetics to a wide array of metabolites. Hence a comprehensive study, including genomics, epi-genomics, transcriptomics, proteomics, and metabolomics, along with the medical/mass-spectrometry imaging, patient clinical history, treatments provided, genetics, and disease endemicity, is essential. Cancer Moonshot℠ Research Initiatives by NIH National Cancer Institute aims to collect as much information as possible from different regions of the world and make a cancer data repository. AI could play an immense role in (a) analysis of complex and heterogeneous data sets (multi-omics and/or inter-omics), (b) data integration to provide a holistic disease molecular mechanism, (c) identification of diagnostic and prognostic markers, and (d) monitor patient's response to drugs/treatments and recovery. AI enables precision disease management well beyond the prevalent disease stratification patterns, such as differential expression and supervised classification. This review highlights critical advances and challenges in omics data analysis, dealing with data variability from lab-to-lab, and data integration. We also describe methods used in data mining and AI methods to obtain robust results for

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Manisha Sawhney,
University of Mary, United States
Gui Zhao Liang,
Chongqing University, China

***Correspondence:**
Vineeta Rai
vineetaiitbombay@gmail.com
orcid.org/0000-0003-4621-1260

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *25 April 2020*
Accepted: *20 July 2020*
Published: *12 August 2020*

Citation:
Patel SK, George B and Rai V (2020) Artificial Intelligence to Decode Cancer Mechanism: Beyond Patient Stratification for Precision Oncology. Front. Pharmacol. 11:1177. doi: 10.3389/fphar.2020.01177

precision medicine from "big" data. In the future, AI could be expanded to achieve ground-breaking progress in disease management.

Keywords: multi-omics, artificial intelligence (AI), data integration, cancer biomarkers, patient stratification

INTRODUCTION

Artificial intelligence (AI) is a branch of computer science with enhanced analytical or predictive capabilities to perform interdisciplinary tasks that otherwise require human intellect. AI has intensive problem-solving capabilities including prediction, data scalability, dimensionality, and integration, reasoning about their underlying phenomena and/or big data transformation into clinically actionable knowledge, based on the learning from model data sets. The learning capacity is maximized by improving the prediction task based on problem-specific measurements of performance. Particularly, machine learning (ML) and deep learning (DL)-based approaches were gaining recognition and emerged as key components in biomedical data analysis, driven by health care data availability and rapid progress of analytics techniques (Jiang et al., 2017; Saltz et al., 2018; Huang et al., 2020; Ibrahim et al., 2020). AI is currently used to automate the information extraction, summarize the electronic medical records or hand-written doctor notes, integrate health records, and store information in cloud scaling (big data storage) (Bedi et al., 2015; Chang et al., 2016; Miotto et al., 2016; Osborne et al., 2016; Garvin et al., 2018; Syrjala, 2018) AI has immense potentials to contribute significantly at every stage of cancer management ranging from reliable early detection, stratification, determination of infiltrative tumor margins during surgical treatment, response to drugs/therapy, tracking tumor evolution and potential acquired resistance to treatments over time, prediction of tumor aggressiveness, metastasis pattern, and recurrence (Bi et al., 2019).

Cancer is a major cause of death worldwide, accounting for an estimated 9.6 million deaths in 2018. Cancers can originate from various organs viz. lung, breast, kidney, represent phenotypic diversity like cell surface markers, molecular mutations (p53, PTEN, ER), demonstrate varied growth rate and apoptosis based on the cancer microenvironment and status of blood supply, and its aggressive nature. Also, cancer has a diverse disease progression mechanism and the effectors ranging from gene-epigenetics to a wide array of metabolites. Cancer/tumor being highly heterogeneous in terms of inter-tumor heterogeneity (cancers from different patients) and intra-tumor heterogeneity (within a single tumor) impose challenges for both detection, treatments, and recurrence. Medical decisions for cancer treatment should consider not only its variegated forms with the evolution of disease but also the individual patient's condition and their ability to receive and respond to treatment. Certain cancers, such as pancreatic and gastric cancers, are detected only after they have reached their advanced stages with frequent relapses. Integration of "multi-omics" (genomics, epi-genomics, transcriptomics, proteomics, and metabolomics), and "non-omics" (medical/mass-spectrometry imaging, patient clinical history, treatments, and disease endemicity) data could help overcome the challenges in the accurate detection, characterization, and monitoring of cancers. AI could play an immense role in the analysis of complex and heterogeneous data sets, particularly from multi-omics and inter-omics approaches and data integration to provide a holistic disease molecular mechanism, identification of novel dynamic diagnostic and prognostic markers and enable precision cancer management, well beyond the prevalent disease stratification patterns such as differential expression, and supervised classification (**Figure 1**). Advanced computational analyses could also augment a global interpretation and automation of the cancer patient radiographs that most commonly relies upon visual evaluations and hence differ in disease assessments. Cancer Moonsho℠ Research Initiatives by NIH National Cancer Institute aims to collect as many omics and non-omics information as possible from different regions of the world to create a national ecosystem for sharing and analyzing cancer data (Cancer Moonshot - National Cancer Institute, 2016). The project will help develop human tumor atlas, predict response to standard treatments, optimize guidelines for systematic cancer prediction and treatments, and identify ways to overcome drug resistance to improve (i) current understanding of cancer, (ii) enable new strategies/technologies for cancer characterization, (iii) early detection of tumors/cancer, and (iv) extend therapies to more patients in a personalized manner (Cancer Moonshot - National Cancer Institute, 2016). The large multidimensional biological data sets (including individual variability in genes, function, and environment) generated, and/or compiled for the fulfillment of this cross-border project require advanced computational analysis, and AI certainly could be one of the key plays.

Recently AI is successfully applied to tumor image segmentation, identify, and quantify the rate and amount of mitosis (Romo-Bucheli et al., 2017), screening mutations (Coudray et al., 2018), auto-detect and classify benign nuclei from cancer cells (Sirinukunwattana et al., 2016; Xu et al., 2016), protein alignments and spatial localization (Saltz et al., 2018), predicting unknown metabolites, precision medicine matching trials (Korbar et al., 2017; Coudray et al., 2018), drug repurposing (Aliper et al., 2016), liquid biopsies and pharmacogenomics based cancer screening/monitoring and predicting the patient outcomes (Cohen et al., 2018; Low et al., 2018), drug discovery (Abadi et al., 2017; Yu P. et al., 2017) and so on. AI has outperformed pathologists and dermatologists in diagnosing metastatic breast cancer (Low et al., 2018) and melanoma (Bejnordi et al., 2017). Conversely, multi-omics data has immense potentials to identify the caveat in the current

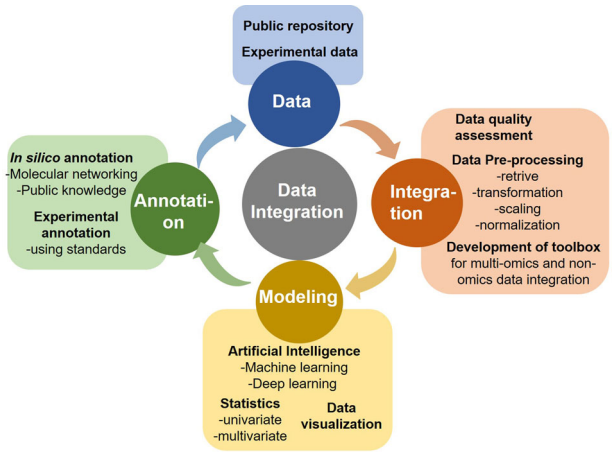


FIGURE 1 | Components of omics data analytics.

AI-based cancer diagnostics, stratification, mutant identification, treatment, and drug repurposing approaches, which could advance precision oncology research (Li et al., 2018). However, we have limited knowledge in the multi-omics and inter-omics data analysis and availability of algorithms (Buchhalter et al., 2014). This review highlights the current AI application in data integration, advancement, scope, and challenges in oncology research and clinical use. The reports mostly cover the articles published in the last two decades (2000–2020).

IMPLICATIONS OF ARTIFICIAL INTELLIGENCE IN CANCER MULTI-OMICS

Advancements in multidimensional "omics" technologies ranging from next-generation sequencing to the mass spectrometry have led to a plethora of information. AI mediated data integration obtained from different "-omics" platforms such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics enables the understanding of complex biological systems by describing nearly all biomolecules ranging from DNA to metabolites. Multi-omics researches have diverse applications in veterinary medicine (Li Q. et al., 2015), microbiology (Zhang et al., 2010), agriculture science (Van Emon, 2016), biofuel (Rai et al., 2016), and biomedical sciences (More et al., 2015; Hasin et al., 2017; Awasthi et al., 2018; Patel et al., 2019) including oncology (see **Table 1**).

Genomics

Genomics data analysis relies on the nucleotide sequences, including expressed sequence tags (ESTs), cDNAs, and gene arrangements on the respective chromosomes. Rapid advances in the next-generation sequencer (NGS) (Paolillo et al., 2016) and *in silico* computational algorithms have led to high-throughput data generation for whole genomes sequencing (WGS) and epigenomes. WGS comprehensively explores all types of genomic alterations in cancer and provides information on the repertoire of driver mutations and mutational signatures (including non-coding regions) in cancer genomes, which remain widely unexplored. Ley et al. reported the first-ever WGS analysis of cancer (cytogenetically normal acute myeloid leukemia, AML) (Ley et al., 2008), merely 6 months post-publication of the first human whole-genome sequence (Wheeler et al., 2008). Since then several cancer genomics databases and projects including The Cancer Genome Atlas (TCGA) (Wang Z. et al., 2016), the International Cancer Genome Consortium (ICGC) (Zhang J. et al., 2019), Catalog of Somatic Mutations in Cancer (COSMIC) (Forbes et al., 2015), Cancer Genomic Hub (CGHub) (Wilks et al., 2014), Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (Therapeutically Applicable Research to Generate Effective Treatments (TARGET)), cBioPortal (Gao et al., 2013), MethyCancer (He et al., 2008), UCSC Cancer Genomics Browser (Goldman et al., 2013) and moonshot project (Cancer Moonshot - National Cancer Institute, 2016) have surfaced (also see **Table 2**). Data accessibility has further led to the development of tools and resources to facilitate the rapid detection and analysis of biologically relevant genomic outcomes (Cerami et al., 2012; Gao et al., 2013; Gonzalez-Perez et al., 2013; Rubio-Perez et al., 2015; Chakraborty et al., 2018). WGS is thus a powerful tool to understand cancer genomics that typically contains unpredictable numbers of point mutations, fusions, and other aberrations. In contrast, targeted approaches like whole-exome sequencing (WES) are easier to analyze but miss out information of untranslated, intronic, and intergenic regions, which might have an impact on the molecular pathogenesis of cancer (Nik-Zainal et al., 2016). However, there are several associated limitations: (i) a vast majority of cancer genomics efforts remain focused around targeted approaches viz. WES (Morris et al., 2017) (ii) many of the genomics data reported lacks a comprehensive clinical annotation required for linking genomic events to specific cancer types, prognoses, and treatment responses (Robinson et al., 2017) (iii) most of the preliminary studies are performed on untreated cancers, and thus do not provide insight into the response to treatment regimens (Robinson et al., 2017). Integrating the cancer genomics data with clinical physiology data could, therefore, be expected to better define cancer biology and responses to treatments. Several studies have integrated genomics and non-omics cancer data (see **Table 1**). Histopathological images integration with genomics helps retrieves better information on cancer tissue architecture, which is generally compromised in molecular assays, rendering this rich information underused (López de Maturana et al., 2019). AI algorithms classify breast cancers using prognostic factors to quantitative image (Yuan et al., 2012) and the public data set (TCGA) (Yuan et al., 2012). AI algorithms to integrate (multi-) omics data with the pathology images has been successfully extended to develop predictive models for prostate cancer (Robinson et al., 2015), renal cell carcinoma (Schoof et al., 2019), low-grade glioma (Brat et al., 2015), and non-small cell lung cancer (Yu et al., 2016). Alongside integrating the multi-omics data from different platforms,

TABLE 1 | Comprehensive list of Artificial Intelligence-based omics and non-omics investigations in oncology.

#	Type of omics	Data type	AI	Tools/analysis	Type of cancer	References
1	Non-omics	Clinicopathological	**DL**	Genetic algorithm and Pearson's correlation coefficient	Oral	(Chang et al., 2011)
2		Clinicopathological	**DL**	Neural network	Colorectal	(Bottaci et al., 1997)
3		Clinicopathological	**DL**	Decision tree, artificial neural network (ANN), support vector machine (SVM) and logistic regression	Colorectal	(Wang et al., 2019)
4		Clinicopathological	**DL**	ANN and Cox regression	Gastric	(Zhu et al., 2013)
5		Clinicopathological	**DL**	Cox proportional hazard and ANN	Gastric	(Biglarian et al., 2011)
6		Sonographic images	**DL**	Deep convolutional neural network (DCNN)	Thyroid	(Li X. et al., 2019)
7		MR images	**DL**	Faster region-based convolutional neural networks (Faster R-CNN)	Metastatic lymph nodes	(Lu Y. et al., 2018)
8		Dermoscopic images	**DL**	Convolutional neural networks (CNN)	Melanoma	(Haenssle et al., 2018)
9		Digital Mammography DREAM	**DL**	Faster region-based convolutional neural networks (Faster R-CNN)	Breast	(Ribli et al., 2018)
10		Clinicopathological	**ML**	Neural networks, decision trees, and logistic regression	Breast	(Delen et al., 2005)
11		Clinicopathological	**ML**	ANN, SVM, and semi-supervised learning	Breast	(Park et al., 2013)
12		Clinicopathological	**ML**	Extreme Learning Machine (ELM), Neural networks and Genetic algorithm	Prostate	(Jović et al., 2017)
13		Clinicopathological	**ML**	Two-stage fuzzy neural network	Prostate	(Kuo et al., 2015)
14		Clinicopathological	**ML**	Linear regression, Decision Trees, Gradient Boosting Machines, and Support Vector Machines	Lung	(Lynch et al., 2017)
15		Radiomics	**ML**	Decision tree, AdaBoost, algorithm, RUSBoost algorithm, matthews correlation coefficient (MCC)	Gliomas	(Lu C. F. et al., 2018)
16		MR images & Clinicopathological	**ML**	SVM, bagged SVM, K-nearest neighbor (KNN), adaptive boosted trees (AdaBoost), random forest (RF), and gradient boosted trees (GBT)	Bladder	(Hasnain et al., 2019)
17	Single omics	Genomics	**DL**	Prognosis-enhanced neural networks (ENN), SVM, and probabilistic-enhanced NN (PENN)	Pan Cancer	(Vasudevan and Murugesan, 2018)
18		Proteomics	**DL**	SVM and C4.5	Breast	(Karsan et al., 2005)
19		Proteomics	**DL**	Deep Learning neural network (DLNN)	Myeloid Leukemia	(Liang et al., 2019)
20		Metabolomics	**DL**	multiple logistic regression (MLR) and alternative decision tree (ADTree)	Breast	(Murata et al., 2019)
21		Genomics	**ML**	SVM, genetic algorithm, log-rank test, and Cox hazard regression model	Ovarian	(Lu et al., 2019)
22		Genomics	**ML**	Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), and Pathway based Deep Clustering model (PACL)	GBM and ovarian cancer	(Mallavarapu et al., 2019)
23		Metabolomics	**ML**	SVM, Naive Bayes, Partial Least Square Discriminant Analysis (PLS-DA), LASSO, RF, KNN, and C4.5	Colonic	(Eisner et al., 2013)
24		Metabolomics	**ML**	RF, SVM, recursive partitioning and regression trees (RPART), linear discriminant analysis (LDA), prediction analysis for microarrays (PAM), and generalized boosted models (GBM)	Breast	(Alakwaa et al., 2018)
25	Non-omics and single omics	MR images and genomics	**DL**	Residual convolutional neural network (RCNN)	Gliomas	(Chang et al., 2018)
26		Clinicopathological and genomics	**DL**	DNN, Multi modal Deep Neural Network by integrating Multiulti-dimensional Data (MDNNMD) and receiver operating characteristic (ROC)	Breast	(Sun et al., 2018)
27		Clinicopathological and genomics	**ML**	Ensemble models-SVM, ANN, KNN, ROC, and calibration slope (CS).	Breast	(Zhao et al., 2018)
28		Clinicopathological and genomics	**ML**	SVM, and ROC	Prostate	(Zhang et al., 2017)
29		Histopathology images and proteomics	**ML**	RF and CNN	Kidney	(Azuaje et al., 2019)
30	Multi-omics	Epigenetics, genomics, and transcriptomics	**DL**	Hierarchical integration deep flexible neural forest framework (HI-DFNForest), KNN, SVM, RF, and multi-grained cascade forest (gcForest)	BRCA, GBM, and OV	(Xu et al., 2019)
31		Epigenetics and transcriptomics	**DL**	Unsupervised feed-forward, nonrecurrent neural network, Cox proportional hazards (Cox-PH) model, K-means clustering, SVM algorithm, concordance index, Log-rank P-value of Cox-PH regression, Brier score, and ANOVA test F values	Liver	(Chaudhary et al., 2018)
32		Epigenetics and transcriptomics	**DL**	OmiVAE, k-means clustering, support vector machine, Variational autoencoder (VAE), PCA, t-SNE, KPCA, and UMAP	Pan cancer	(Zhang X. et al., 2019)
33		Epigenetics and transcriptomics	**DL**	DeepProg, Autoencoder, Cox-PH model, Gaussian mixture model, concordance index, and Wilcoxon rank-sum test	Pan cancer	(Poirion et al., 2019)
34		Genomics, transcriptomics, and proteomics	**ML**	Generic model, gene-specific model, RF, Random Forest Regressor, and trans-tissue model, Wilcoxon signed-rank test	Breast and ovarian	(Li H. et al., 2019)

transcriptomics, proteomics, and metabolomics with genomics could consolidate molecular information. Wu P. et al., 2019 integrated the Clinical Proteomic Tumor Analysis Consortium (CPTAC) mass spectrometry-based proteomics data for selected breast, colon, and ovarian tumors with TCGA into the cBioPortal (cBioPortal for Cancer Genomics) to support easy exploration and integrative analysis of the proteomic data sets in the context of the clinical and genomics data from the same tumors (Wu P. et al., 2019). Considering the diversity of cancer genomes and phenotypes, cataloging and interpretation of the abundant mutation, particularly non-coding and structure variants, could be performed with confidence *via* integrating clinicopathological information along with transcriptomics, and epigenomics to decide the precise treatments that will produce the best results for the cancer patients.

Transcriptomics

Transcriptome denotes the active genes as well as long-noncoding RNA, short RNAs such as microRNAs, small nuclear RNAs in a defined physiological condition. The system-wide transcriptomic analysis evaluates overall transcripts in a metabolic process, while the targeted approach provides information regarding known genes. Differential expression of protein-coding RNA could provide insight into the disease mechanism, as well as integrated with genomics and proteomics to discover novel genes and their functional relevance. While non-coding RNAs have regulatory functions in several metabolic diseases, neurological disorders, and cancer. Transcriptome is directly co-related to any epigenomic change that manifests cancer, hence the integration of epigenomics and transcriptomics data could extend our understanding of cancer biology such studies are reported in breast (Robinson et al., 2015), prostate cancer (Varambally et al., 2002; Bhasin et al., 2015), head and neck squamous cell carcinoma (HNSCC) (Kelley et al., 2017). Also, the transcriptomics and epigenomics data integration approach opens-up avenues to know more about the promoter crosstalk through a shared enhancer (Eun et al., 2013) and dynamic switching of promoter and enhancer domains (Sohni et al., 2015). Moarii et al. used a large data set of 672 cancerous and healthy methylomes gene expression and copy number profiles from TCGA and performed a meta-analysis to clarify the interplay between promoter methylation and gene expression in normal and cancer samples (Moarii et al., 2015). Vantaku et al. demonstrated a novel approach for the unbiased integration of transcriptomics, metabolomics, lipidomics, and data to robustly predict high-grade patient survival and discovery of novel therapeutic targets in bladder cancer (Vantaku et al., 2019).

Proteomics

Proteomic profiles reveal cellular/molecular responses to (epi-) genomics, and environmental alterations, and their feedback responses. Post-translation modifications, including phosphorylation, glycosylation, ubiquitination, nitrosylation, enrich the protein repertoire (protein isoforms), and impacts protein functions like transport, enzymatic activity, and intracellular signaling pathways in cancer. Classifying specific protein isoforms provide unmatched clinical sensitivity and specificity. Various tissue and plasma proteomics studies are performed (Peng et al., 2018) to screen and diagnose cancers including colorectal (Tsai et al., 2012; Fayazfar et al., 2019; Thorsen et al., 2019), breast(Mishra et al., 2015), liver (Yang et al., 2013), oral (Lai et al., 2010) and so on. MS has applications beyond disease diagnostics and could be extended to monitor the feedback responses towards therapy, identify drug toxicity, and discovering new biomarkers. High-quality data sets are obligatory for clinical MS. Hence improvements in MS-instrument quality and robustness, automated sample processing, robust data analysis pipelines, and online automation (cloud computing) to synchronize results, data sets, and data portability have contributed to expanding the use and impact of MS in cancer research. Also, to deal with the variations in the proteomics data sets across the globe, Proteomics Standards Initiative (PSI) from the Human Proteome Organization (HUPO) has setup guidelines for sample collection viz. selecting appropriate disease controls, categorizing disease and sub-disease status (Maes et al., 2015), storage to rule-out pre-analytical variables (including patient and instrumental factors) that contribute to a large extent of variation, calibrating MS instrument for data-quality assurance, data reporting for untargeted (Martínez-Bartolomé et al., 2014) and targeted (Abbatiello et al., 2017) analysis. An amalgamation of proteomics data with (epi-)genomics, transcriptomics, metabolomics, and cancer histopathological images using AI gives confidence in the data or metabolic pathways identification. Proteomics investigation of breast cancer contoured more than 12,000 proteins and 33,000 phospho-sites. Proteogenomic analysis associated DNA mutations (data obtained from TCGA) to protein signaling to pinpoint the genetic drives of cancer, and revealed new signaling pathways for the breast cancer subtypes with specific mutations (PIK3CA and TP53) and identified two candidate markers (SKP1 and CETN3) in basal-like breast cancer (Mertins et al., 2016). Liu et al., integrated transcriptome (RNA-seq) and proteome (data-independent acquisition, DIA) data to co-relating RNA splicing links isoform expression with proteome diversity that may help for studying the perturbations associated with cancer (Liu et al., 2017). MS imaging (MSI) is yet another advancement in MS that enables visualization of tumor microenvironmental biochemistry and empowers tumor biology investigation to an entirely novel biochemical perspective, thereby potentially leading to the identification of a new pool of cancer biomarkers (Bi et al., 2019). High-throughput MSI analysis is a powerful tool for biomarker identification in a spatial manner, tracking drugs and its metabolites, imaging drug-response at cellular-level. MSI tool was used to identify unique region-of-interest–specific biomarkers (lipid signature) and therapeutic targets to classify colorectal cancer and subtyping in non-small cell lung cancer (Kriegsmann et al., 2016). MSI also finds application in the identification of prognostic signatures beyond classical histology. Proteins and protein isoforms associated with patient survival in four different high-grade sarcoma subtypes (Lou et al., 2017) and colorectal adenocarcinoma (Hinsch et al., 2017) were

TABLE 2 | List of cancer genomics databases.

#	Cancer genomic database name	Cancer alteration types	Organisms	Public data
1	The Cancer Genome Atlas (TCGA)	Copy number, mutation, methylation, gene expression, miRNA expression	Human	Yes
2	The International Cancer Genome Consortium (ICGC)	Mutation	Human	Yes
3	Catalog of Somatic Mutations in Cancer (COSMIC)	Mutation	Human	No
4	cBio Cancer Genomics Portal	Copy number, mutation, methylation, gene expression, miRNA expression, protein, phosphorylation	Human	Yes
5	MethyCancer	Methylation	Human	Yes
6	MutaGene	Mutation	Human	Yes
7	Moonshot project	Copy number, gene expression	Human	Yes
8	Integrative Oncogenomics Cancer Browser (IntOGen)	Copy number, mutation, gene expression	Human	Yes
9	Mouse Retrovirus Tagged Cancer Gene Database	Mutation	Mouse	Yes
10	Mouse Tumor Biology Database	Copy number, mutation, methylation, gene expression	Mouse	No
11	OncoDB.HCC	Copy number, gene expression, QTL	Human, mouse, rat	No
12	UCSC Cancer Genomics Browser	Copy number, mutation, gene expression, miRNA	Human, mouse, rat	Yes

identified. In gastric adenocarcinoma, native glycan fragments detected by MALDI-FT-ICR mass spectrometry imaging were linked to patient prognosis (Kunzke et al., 2017). Combining MSI with histology enables the extraction of molecular profiles from specific regions of tissue or histopathological entities, implying MSI can facilitate intelligent knife (iKnife) in sorting tumors during surgery with high sensitivity and specificity (Balog et al., 2013). Certainly, MS-based analysis, along with histopathological diagnosis, can show a stronger association with the clinical outcome (Huber et al., 2014). Recently MSI data is combined with other imaging data like fluorescence *in situ* hybridization, tissue microarrays, confocal Raman spectroscopy, and MRI, for example, MRI and MSI imaging data were collated to analyze brain pathophysiology (Porta Siegel et al., 2018). Combing vasculature staining (using an anti-CD31 antibody) and MSI could help attain a better picture of vascularization as well as vessel characteristics. However, with emerging MS technologies, there are still challenges in its clinical application including nonoptimized raw data preprocessing, imprecise image co-registration, and limited pattern recognition capabilities due to lack of reference spectra database (Addie et al., 2015). Nevertheless, efforts/measures are taken towards the successful implementation of MS technology for diagnosis of cancer biomarkers translatable to clinical setting. Additionally, the imaging data could be integrated with LC/GC-MS, the workhorse technique of proteomics workflow that includes the extraction of total proteins/peptides, fractionation, and deep proteomic analysis. Delcourt et al., combined MSI and top-down microproteomics to detect potential protein markers in serous ovarian cancer (Delcourt et al., 2017). Using LC-MS and peptide fractionation Kulak et al. achieved deep coverage of cellular proteomes with sub-microgram sample input (Liebl, 1967). Further the cancer signature biomarkers could be used to stratify patients according to subtype, metastatic risk, progression, recurrence, and treatment response. Lately single-cell proteomics is gaining importance to bring comprehensive insights into the cancer heterogeneity, clonality to metastasis or to capture information from rare/mutated cells (Doerr, 2019). Using a quantitative single-cell proteomics approach Schoof et al., characterized an acute myeloid leukemia hierarchy (Schoof et al., 2019).

Metabolomics

Metabolomics is a systematic analysis of small molecules (<1kD) within cells, biofluids, tissues, or organisms involved in primary or secondary metabolic processes. Metabolites (small molecules) are highly diverse classified into multiple categories: amino acids, lipids, nucleotides, carbohydrates, and organic acids. Metabolite repertoire changes significantly during the process of normal growth and development and/or exposure to stress, allergens, and disease conditions (Bertini et al., 2009; Lin et al., 2011; Veselkov et al., 2011), which relates strongly to the final clinical phenotype. Metabolomics thus enhances our molecular understanding of disease mechanisms, progression, response to drugs/treatments, and recurrence probability. Typical metabolomics analysis workflow comprises of metabolite extractions, separation by liquid/gas chromatography, capillary electrophoresis and ion mobility, detection by mass spectrometry (MS), or nuclear magnetic resonance (NMR) spectroscopy and data analysis. MS applications in metabolomics have increased exponentially since the discovery and development of soft ionization tools like electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). Several separation-free MS techniques including direct infusion-MS, MALDI-MS, mass spectrometry imaging (MSI), and direct analysis in real-time mass spectrometry are gaining popularity. The advantages of separation-free mass spectrometry are reduced sample volume requirements and minimization of the analytical variation. Untargeted metabolomics approaches are ideally used for hypothesis development, as it simultaneously identifies several unknown/known metabolites and quantifies. However, diverse physical and chemical properties and wide concentration ranges of the metabolites, biological variations (Heinemann et al., 2014), and identification of the unknown

compounds based on the MS/MS fragmentation patterns impose challenges for untargeted metabolomics. For a long time, researchers have identified the unknows in the biological samples by complementing the MS/MS fragmentation with public repository or standards, which leads to the identification of a very limited number of metabolites, while a majority of the potentially useful information in MS/MS data sets remains uncurated. Molecular networking like GNPS has proved to be very useful in cataloging the uncurated MS/MS data sets *via* a spectral correlation and visualization approach that can detect sets of spectra from related molecules even when the spectra themselves are not matched to any known compounds (Wang M. et al., 2016). ML in combination with data mining algorithms (supervised and unsupervised) like principal component analysis or hierarchical clustering has transformed metabolomics studies like analyzing several variables/treatments simultaneously (Duan et al., 2005; Bertini et al., 2009; Guan et al., 2009). Particularly unsupervised data mining allows extracting meaningful relationships between samples with less risk of human bias. Metabolomics is applied in biomarker identification for diagnosis, monitoring, and prognosis of several diseases(Alvarez et al., 2017; Chorell et al., 2017; Perng et al., 2017; Patel et al., 2019), particularly those impacting metabolic functions, such as cancer. Metabolomic biomarkers for several cancers including colorectal (Ma et al., 2012; Nishiumi et al., 2012; Manna et al., 2014; Yamazaki, 2015) pancreatic (Zhang et al., 2012), lungs (Koutros et al., 2013; Li et al., 2014; Zhuang et al., 2016), breast (Cui et al., 2016; Li et al., 2020), gastric (Ikeda et al., 2012), ovarian (Zhang et al., 2013), and prostate (Koutros et al., 2013; Mondul et al., 2014; Kelly et al., 2016) have been reported. Despite numerous ongoing studies, limited metabolomics biomarkers reach clinical trials, implying improvements in experimental designs, data analysis with reduced false discovery rates, pinpointing molecules accountable for metabolic aberrations, and data interpretation is needed. Besides, we also must overcome interlaboratory variability by generalizing the protocols that are robust and adaptable to enhance reproducibility. Indeed, MS-based metabolomics biomarker discoveries have entered the new realms of MSI that present intuitive metabolites distribution in tissues or cells. MSI is performed in two modes, namely, imaging (Stoeckli et al., 2001) to correlate with histology and profiling, to know the overall metabolites (Cornett et al., 2006). MSI alone or in conjunction with (immuno-)histochemistry (IHC) enhances our understanding of complex heterogeneous cancer metabolic reprogramming with spatial information and facilitate the discovery of potential metabolic vulnerabilities that might be targeted for tumor therapy. MSI suffers some technical limitations like area of detection limits, instrument sensitivity at the high spatial resolution, ion suppression, matrix effects, and data analysis, particularly normalization and background correction, but has tremendous potential to improve cancer diagnostics. Huang et al. developed a graphical data processing pipeline for MSI based spatially resolved metabolomics (Huang et al., 2019), which could achieve multivariate statistical results in an intuitive and simple way as well as discovery low-abundant but reliable biomarkers in heterogeneous tumors. MSI has been employed to different cancers including brain (Jarmusch et al., 2016; Clark et al., 2018), breast (Guenther et al., 2015; Abdelmoula et al., 2016; Angerer et al., 2016; Wang S. et al., 2016; Torata et al., 2018; Vidavsky et al., 2019), lung (Calligaris et al., 2015; Li T. et al., 2015; Carter et al., 2017; Holzlechner et al., 2018), ovarian (Dória et al., 2016; Briggs et al., 2019), prostrate (Wang et al., 2017), esophageal (Guo et al., 2014; Abbassi-Ghadi et al., 2016; Sun et al., 2019a), colon (Hiraide et al., 2016; Inglese et al., 2017), oral (Uchiyama et al., 2014; Bednarczyk et al., 2019), skin (Xu et al., 2017; Margulis et al., 2018), adrenal gland (Sun et al., 2019b) and gastrointestinal stromal tumors (Abu Sammour et al., 2019) for spatial metabolomics analysis. MSI is also used to determine the metabolite changes in the 3D osteosarcoma cell culture model upon drug treatments (Palubeckaitė et al., 2020). MSI has been used to investigate tumor biopsy tissues for hypoxia (Chughtai et al., 2013; Jiang et al., 2015), driver of tumor resistance to radiotherapy or chemotherapy, and lipid distributions (Inglese et al., 2017; Paine et al., 2019). Esteva et al., employed deep convolutional neural networks (CNNs)-representing a diverse class of multi-layer artificial neural networks, pre-trained on millions of images representing more than 1000 generic image classes to automate the classification of skin cancers (Esteva et al., 2017). The same approach could be extrapolated for analyzing the images captured by MSI for better cancer management. Inglese et al. recently developed a new computational multimodal pipeline Spatial Correlation Image Analysis (SPACiAL) to integrate MSI molecular imaging data with multiplex IHC. The pipeline allows comprehensive analyses of metabolic heterogeneity, thereby increasing the efficiency and precision for spatially resolved analyses of specific cell types (Inglese et al., 2017).

CONSIDERATION AND CHALLENGES FOR AI-MEDIATED MULTI-OMICS DATA INTEGRATION

AI-mediated clinical cancer research has attained new heights for its unpreceded learning capabilities to process complex data. ML and deep learning (DL) are the subset of artificial intelligence that enables computers to learn with data without being explicitly programmed. AI analytical skills are primarily due to image recognition, computer vision, data integration, decision making, and natural language processing. AI could thus self-adapt, synchronize qualitative and quantitative information, and validate clinical results obtained from multiple platforms. However, AI applications in oncology research still is infancy and must overcome several challenges (**Figure 2**).

Data Integration: A Major Challenge in Precision Oncology

A major challenge in precision oncology is to integrate data generated from multiple types of omics to predict biomarkers or phenotypic outcomes (tumor/normal, early/late stage, survival,

etc.). Machine learning tasks consist of three key steps in order to develop a computational model for biological data integration and analysis: (i) selection and pre-processing of data set, (ii) selection of algorithm and identify the ways to train it for development of a prediction model, and (iii) validation of the model in another data set (**Figure 3**).

Input Data Selection and Pre-Processing

Input data for most of the models consists of gene expression data, copy number alteration, epigenomics, proteomics, and single nucleotide mutations data sets. However, an integrated data analysis strategy combines various omics modalities, and this amalgamation of different types of data could help to develop promising prognostic models. Multi-platform data integration relies on (a) advances in sample extraction and processing technologies; (b) the availability of sufficiently large, matched, and carefully annotated data sets for multi-omics data; (c) molecular and physiologically characterized and graded tumor/cancer data set; (d) data sets with more informative images compared to present databases, e.g., the TCGA image collection (Yu K.-H. et al., 2017) for better 3D-fitting of *in vivo* imaging and *ex vivo* data. The first step in ML analysis is pre-processing of a defined data set(s). It requires normalization, noise filtration, and feature selection when more than one data sets are combined. Normalization becomes an essential step to eliminate biases during the analysis of different data sets that are merged. Selection of defined features is a critical phase in the success of an algorithm in classification, regression, and pattern recognition (Vougas et al., 2019).

Selection of Algorithm/Prediction Model and Data Integration

Algorithms are trained through optimizing the parameters to reach an ideal model. k-fold cross-validation (KF-CV) is widely used for optimizing without capturing the noise of data so that the results of statistical analysis can be generalized to an independent data set (Gao et al., 2019). Several studies on statistical methods and algorithms for data integration are reported (Huang et al., 2017; Perakakis et al., 2018; Zeng and Lumley, 2018; Wu C. et al., 2019). Standard machine learning techniques are supervised and unsupervised learning. Supervised learning requires algorithms to be provided with labeled inputs (e.g., omics data) and the desired output (e.g., the presence of a disease or not). In unsupervised learning, data are not labeled, and the algorithm is trained to look for naturally occurring pattern to correspond with the output. Another category that is more common in multi-omics studies is semi-supervised learning, where unlabelled data is used in conjunction with small labeled input. Briefly, multi-omics data integration consists of (a) dimension reduction: to reduce complexity, a number of factors are condensed to fewer variables (called components). (b) Clustering: Grouping input variables with common characteristics in same clusters, (c) density estimation to assess the distribution of input variables in specific space, and (d) regression to estimate the relationships among variables and for developing predictive models.

Testing the Prediction Models

Building a model that fits data beyond the current predictive model is the ultimate goal of training a candidate computational model. This can be tested by implementing a candidate predictive model to blind data sets. If the model is for developing tools to identify precision and personalized therapies for individual cancer patients, panels from clinical data sets should be preferentially used. A trained model that fails to generalize might be because of overfitting or underfitting (Dietterich and Bakiri, 1995). In the case of overfitting, noise, or random fluctuations are picked up in the training data, which negatively impact the model's ability to generalize. Overfitting of a trained model is a major issue in machine learning. In underfitting, the underlying structure of a particular data set is not captured in a set of *in silico* pipeline. The predictive model's capacity to make predictions understandable or interpretable to humans is another key requirement, i.e., the higher the complexity of the model, the more challenging interpretability becomes (black box models). This could be achieved at different levels of data processing and abstraction, however the development of methods for interpreting ML models is at a relatively early stage, particularly for precision oncology (Castelvecchi, 2016). Enhancing the interpretability will allow users to peer into the hidden layers of the model and determine how exactly the predictions are made on a case-to-case basis.

Deep Neural Networks: For Multi-Omics Data Integration

Deep neural networks (DNNs) are a subset of machine learning, which is gaining popularity in precision medicine. Today's complex multi-omics data might be challenging to analyze with traditional machine learning algorithms. DNNs algorithms can integrate multi-omics data with better sensitivity, specificity, and efficiency. Moreover, DNNs have the advantage of integrating other sources of information such as medical images or clinical health records, which is a pre-requisite for personalized medicine. Sakellaropoulos et al. designed the DNNs model, which could capture pathways that linked gene expression with drug response and showed that DNNs are better than other traditional machine learning algorithms. Also, DNNs predicted drug response and survival in a large clinical cohort (Sakellaropoulos et al., 2019). Deep learning is still an emerging area in biomedical field, their effectiveness is not always guaranteed. Cancer multi-omics data integration is done using various approaches: unsupervised cancer subtyping to show patient survival (Ramazzotti et al., 2018), graph-based integration to integrate copy number aberration, epigenome, and transcriptome data sets for ovarian cancer clinical outcome prediction (Kim et al., 2015) and integration DNA methylation and matched imaging data to predict glioblastoma disease progression (Klughammer et al., 2018). However, rigorous mathematical foundations for emerging DNNs

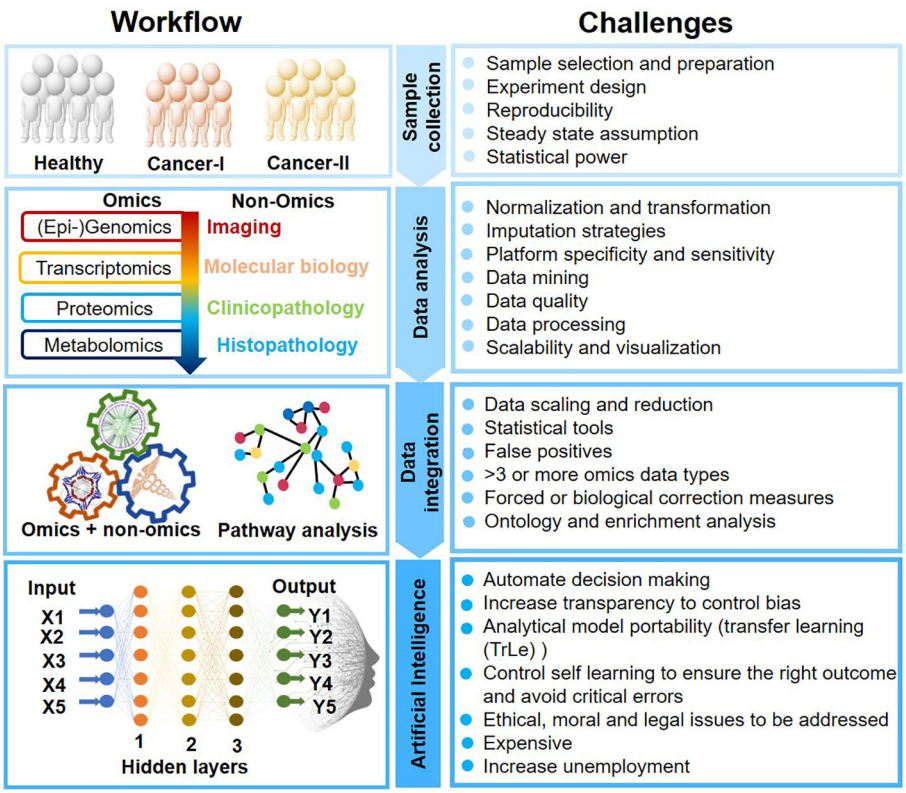


FIGURE 2 | Artificial Intelligence-mediated oncology workflow and challenges.

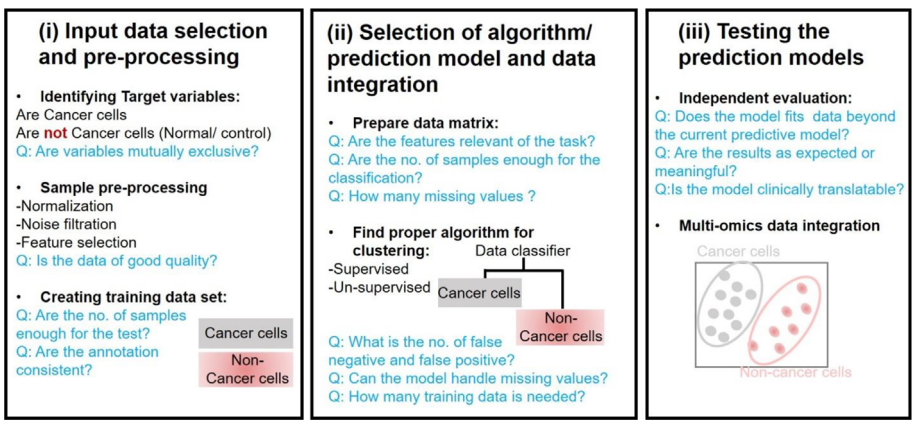


FIGURE 3 | Considerations for major stage of Machine Learning based analysis in oncology.

architectures are still lacking (Martorell-Marugán et al., 2019). One of the most challenging and futurist modules of the data-integration is combining multi-omics and non-omics data (imaging, biochemical/molecular profile data, clinical symptoms). Yu et al., associated omics data of lung cancer patients with the histopathology data to determine the patient survival rate (Yu K.-H. et al., 2017).

Machine Learning for Drug Response Prediction in Precision Oncology Applications

Identification of a panel of biomarkers that are associated with treatment responses is imperative for the precision oncology approach. Machine learning algorithms are being developed for prediction to drug response using response-predictive biomarkers

through integrative analysis of multi-omics data (Ali and Aittokallio, 2019). Drug sensitivity prediction models, which are entirely based on gene expression profile, are less trustworthy compared to those which are based on integrated multi-omics profiling. Input data type, noise ratio, dimensionality, data complexity, and heterogeneity, are the crucial factors for drug response prediction model development. Sometimes, it is difficult to understand prediction models due to the dominance of gene expression profile data sets, which can be decreased by a two-stage method, called TANDEM (Aben et al., 2016). Bayesian efficient multiple kernel learning (BEMKL) is another drug response prediction model based on multi-omics data. It was the top-performing model in the National Cancer Institute - Dialogue for Reverse Engineering Assessment and Methods (NCI-DREAM7) Drug Sensitivity Prediction Challenge (Costello et al., 2014). Currently, the majority of data in repositories that are publicly available represent a significant set of data that are derived using cell lines treated with different doses of drugs and a large number of compounds. Some of these widely used data sets are: (i) The genomics of Drug Sensitivity in Cancer (GDSC), (ii) Cancer Cell Line Encyclopedia (CCLE), and (iii) National Cancer Institute drug screening panel (NCI-60). It is essential to understand that data extracted from clinical samples are ideal for the development of favorable drug prediction models. Heterogeneous properties of cancers make *in silico* analysis for molecular matching using cancer cell lines challenging in clinical settings (Hanahan and Weinberg, 2011; Turajlic et al., 2019). Importantly, the interplay of tumor-microenvironment that determines cancer development and response to drug treatment cannot be recapitulated using cancer cell lines model, and therefore, molecular changes associated with clinical cancers are diverse than in cancer cell lines (Wu and Dai, 2017). Lack of reliable resources for input data set stalled the success of creating drug prediction models. There is an urgent need to evaluate *in silico* technologies like *transfer learning* (TrLe) methods employing different ML algorithms and applications that utilize predictive feature (very complex non-linear relationships between features) learned in cell line trained model to build a new model or leverage information from auxiliary data not directly belonging to the problem being handled, that can be used in real clinical settings. Several studies have executed TrLe approach and tested and trained machine learning model for data obtained from clinical samples (Daemen et al., 2013; Turki et al., 2018). Turki et al. used TrLe-based approach to transfer patterns learned in breast and lung cancer patient data sets to predict drug sensitivity of multiple myeloma patients (Turki et al., 2018). Daemen et al. used breast cancer cell lines data for training model and tested on clinical data sets derived from TCGA (Daemen et al., 2013). Similarly, the Geeleher group built a training model on gene expression data sets extracted from Cancer Genomics Project and tested them on TCGA data sets from non-small-cell-lung cancers (NSCLC) (Geeleher et al., 2014). Using an elastic net model on to B-cell lymphoma cell lines, Falgreen et al. identified gene signatures that are associated with the development of resistance to drug (cyclophosphamide, doxorubicin, and vincristine) in diffuse large B-cell lymphoma (Falgreen et al., 2015). Sevakula et al. transfer learning for molecular cancer classification using DNN (Sevakula et al., 2019).

Machine Learning in Biomarker Discovery and Patient Classification

The identification of the disease biomarkers from -omics data does not only facilitate the stratification of patient cohorts but also provides early diagnostic information to improve patient management and prevent adverse outcomes. Coudray et al. applied CNN on whole-slide images obtained from The Cancer Genome Atlas to accurately and automatically classify subtypes of lung cancer, namely adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) and normal lung tissue (Coudray et al., 2018). Likewise, Huttunen et al., automated classification of multiphoton microscopy images of ovarian tissue (Huttunen et al., 2018). Further, they reported a prediction performance comparable to that obtained by pathologists. Brinker et al., automated dermoscopic melanoma image classification using CNN and showed its superiority over both junior and board-certified dermatologists (Brinker et al., 2019). Molecular profiling of carcinoma using circulating cell-free DNA is another approach for sub-dividing patients in risk factors (Kaseb et al., 2019). It has the advantage of being a noninvasive panel of biomarkers based on the multi-omics approach to increase the accuracy compared to biomarker-based on single omics data. For instance, protein biomarkers found in small sample sizes in the discovery cohort may be prone to achieve over-fitting and overinterpretation of proteomic data. Combined analysis of genomics with proteomics data sets led to the identification of novel therapeutic targets such as altered PI3K pathway in hormone receptor-positive breast cancer (Stemke-Hale et al., 2008). Transcriptomics with proteomics data sets analysis leads to the identification of gonadotropin-releasing hormone (GnRH) signaling pathway in glioblastoma that was not interpreted with single omics data set (Jayaram et al., 2016). Similarly, integrated analysis of DNA copy number alteration, with gene expression data in breast cancer patients led to understand the biology of cancer type and promoted to identify novel therapeutic interventions (Curtis et al., 2012). Four unique urinary biomarkers were identified in an integrated transcriptomic and metabolomics data analysis that was more reliable than single omics data analysis (Nam et al., 2009). Integrated proteogenomic characterization of paired tumor and adjacent liver samples identified alterations of the liver-specific proteome and metabolism. Biomarkers and patients' subgroups with distinct features in metabolic reprogramming, microenvironment dysregulation, cell proliferation, and potential therapeutics were identified (Gao et al., 2019).

CONCLUDING REMARKS AND OUTLOOK

Cancer refers to a compendium of related diseases with uncontrolled dividing and spreading cells. More than 100s of different types of cancers are known. Cancer will be the leading

cause of mortality in developed countries by 2030 (Centre for Disease Control). Cancer treatments are challenging due to its heterogenicity (temporal and spatial), high recurrence, and low median survival rate causing millions of deaths every. The molecular understanding of tumor biology has notably changed cancer treatment paradigms during the past 15 years. Still, the success of cancer therapeutics in clinical trials is the lowest of all major diseases. Future cancer treatments thus vouch for tailoring personalized therapies and targeting components of the tumor microenvironment. Accurate early diagnosis and prognosis of cancer greatly increases the chances for successful treatment and patient's survival rate. Present cancer diagnosis relies on the clinician's judgment based on their knowledge and clinical experience, which certainly cannot be guaranteed accurate diagnosis. This aspect points to the variability of the human brain to integrate large amounts of sample data. AI (ML and deep learning) is extremely proficient at handling vast amounts of complex nonlinear data (multi-omics and non-omics) generated during cancer treatments and researches, fault tolerance, parallel distributed processing, learning, and decision-making capabilities to improve oncologic care. AI could thus not only integrate various aspects of the clinical diversity but also helps to address the current lack of objectivity and universality in expert systems. Various researches showed impressive diagnostic and prognosis performance of AI using ML (Esteva et al., 2017; Ferroni et al., 2019; Jiang and Xu, 2019). Yoon et al. showed the potential of AI models for personalized oncology treatments that can estimate individualized treatment effects based on the analysis of counterfactual clinical outcomes (Yoon et al., 2018). ML algorithms (supervised or unsupervised learning) guided by clinicians could unravel the hidden molecular patterns within the data sets (multi-omics and non-omics) to support discovery of biomarkers (diagnostic, prognostic, recovery, and recurrence), candidate therapeutic targets associated with a specific patient group, and clinically relevant subtypes without explicit programming in clinical setups. Clinicians' roles are inevitable in selecting the training data sets and multiple combinations of parameters necessary for building a classification ML model to address specific research questions. In turn, AI can help train junior physicians in clinical diagnosis and decision making. Expanding AI applications from pattern recognition capacity to dealing with multiple data modalities, insufficient data, evaluation of selective and predictive performance, guiding the learning process, and fine-tune models *via* feedback could revolutionize the cancer managements. Another step forward towards AI mediated clinical application is the development of ML pipelines that not only automate the design and evaluation of algorithms but also delineate the clinician the reasoning underlying the model predictions. This is a crucial step considering the fact although AI has learning potential but is in its infancy and cannot be left unattended. Yet another aspect is the extrapolation of the models generated using the cell line data to the patients, as the majority of the previous studies are performed on cell lines or limited small patient sample size, and the portability of the models generated in one cancer to another. AI has come long way but still it must achieve several landmarks: (a) non-reproducible results, (b) population heterogeneity, (c) instrument-variation, (d) lab-to-lab variation, (e) data normalization, (f) cross-compare results by different studies, (g) simulate results *in vitro* to clinics, (h) personalize, and (i) cost-effectiveness. Taken together, advancements in AI-based clinical cancer research will remarkably improve cancer prognosis and diagnosis with precision, resulting in enhanced prediction rates and patient survival.

AUTHOR CONTRIBUTIONS

VR and SKP conceived the idea, wrote the manuscript, made table, and figures. BG contributed to writing.

REFERENCES

Abadi, S., Yan, W. X., Amar, D., and Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput. Biol.* 13, e1005807. doi: 10.1371/journal.pcbi.1005807

Abbassi-Ghadi, N., Golf, O., Kumar, S., Antonowicz, S., McKenzie, J. S., Huang, J., et al. (2016). Imaging of Esophageal Lymph Node Metastases by Desorption Electrospray Ionization Mass Spectrometry. *Cancer Res.* 76, 5647–5656. doi: 10.1158/0008-5472.CAN-16-0699

Abbatiello, S., Ackermann, B. L., Borchers, C., Bradshaw, R. A., Carr, S. A., Chalkley, R., et al. (2017). New Guidelines for Publication of Manuscripts Describing Development and Application of Targeted Mass Spectrometry Measurements of Peptides and Proteins. *Mol. Cell. Proteomics* 16, 327–328. doi: 10.1074/mcp.E117.067801

Abdelmoula, W. M., Balluff, B., Englert, S., Dijkstra, J., Reinders, M. J. T., Walch, A., et al. (2016). Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of Mass spectrometry imaging data. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12244–12249. doi: 10.1073/pnas.1510227113

Aben, N., Vis, D. J., Michaut, M., and Wessels, L. F. A. (2016). TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. *Bioinformatics* 32, i413–i420. doi: 10.1093/bioinformatics/btw449

Abu Sammour, D., Marsching, C., Geisel, A., Erich, K., Schulz, S., Ramallo Guevara, C., et al. (2019). Quantitative Mass Spectrometry Imaging Reveals Mutation Status-independent Lack of Imatinib in Liver Metastases of Gastrointestinal Stromal Tumors. *Sci. Rep.* 9, 10698. doi: 10.1038/s41598-019-47089-5

Addie, R. D., Balluff, B., Bovée, J. V. M. G., Morreau, H., and McDonnell, L. A. (2015). Current State and Future Challenges of Mass Spectrometry Imaging for Clinical Research. *Anal. Chem.* 87, 6426–6433. doi: 10.1021/acs.analchem.5b00416

Alakwaa, F. M., Chaudhary, K., and Garmire, L. X. (2018). Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J. Proteome Res.* 17, 337–347. doi: 10.1021/acs.jproteome.7b00595

Ali, M., and Aittokallio, T. (2019). Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* 11, 31–39. doi: 10.1007/s12551-018-0446-z

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., and Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 13, 2524–2530. doi: 10.1021/acs.molpharmaceut.6b00248

Alvarez, J. A., Chong, E. Y., Walker, D.II, Chandler, J. D., Michalski, E. S., Grossmann, R. E., et al. (2017). Plasma metabolomics in adults with cystic

fibrosis during a pulmonary exacerbation: A pilot randomized study of high-dose vitamin D3 administration. *Metabolism* 70, 31–41. doi: 10.1016/j.metabol.2017.02.006

Angerer, T. B., Magnusson, Y., Landberg, G., and Fletcher, J. S. (2016). Lipid heterogeneity resulting from fatty acid processing in the human breast cancer microenvironment identified by GCIB-ToFSIMS imaging. *Anal. Chem.* 88, 11946–11954. doi: 10.1021/acs.analchem.6b03884

Awasthi, G., Tyagi, S., Kumar, V., Patel, S. K., Rojh, D., Sakrappanavar, V., et al. (2018). A Proteogenomic Analysis of Haptoglobin in Malaria. *Proteomics - Clin. Appl.* 12, e1700077. doi: 10.1002/prca.201700077

Azuaje, F., Kim, S.-Y., Perez Hernandez, D., and Dittmar, G. (2019). Connecting Histopathology Imaging and Proteomics in Kidney Cancer through Machine Learning. *J. Clin. Med.* 8:1535. doi: 10.3390/jcm8101535

Balog, J., Sasi-Szabó, L., Kinross, J., Lewis, M. R., Muirhead, L. J., Veselkov, K., et al. (2013). Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci. Transl. Med.* 5, 194ra93. doi: 10.1126/scitranslmed.3005623

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., et al. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* 1, 1–7. doi: 10.1038/npjschz.2015.30

Bednarczyk, K., Gawin, M., Chekan, M., Kurczyk, A., Mrukwa, G., Pietrowska, M., et al. (2019). Discrimination of normal oral mucosa from oral cancer by mass spectrometry imaging of proteins and lipids. *J. Mol. Histol.* 50, 1–10. doi: 10.1007/s10735-018-9802-3

Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - J. Am. Med. Assoc.* 318, 2199–2210. doi: 10.1001/jama.2017.14585

Bertini, I., Calabró, A., De Carli, V., Luchinat, C., Nepi, S., Porfirio, B., et al. (2009). The metabonomic signature of celiac disease. *J. Proteome Res.* 8, 170–177. doi: 10.1021/pr800548z

Bhasin, J. M., Lee, B. H., Matkin, L., Taylor, M. G., Hu, B., Xu, Y., et al. (2015). Methylome-wide Sequencing Detects DNA Hypermethylation Distinguishing Indolent from Aggressive Prostate Cancer. *Cell Rep.* 13, 2135–2146. doi: 10.1016/j.celrep.2015.10.078

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., et al. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA. Cancer J. Clin* 69, 127-157. doi: 10.3322/caac.21552

Biglarian, A., Hajizadeh, E., Kazemnejad, A., and Zali, M. R. (2011). Application of artificial neural network in predicting the survival rate of gastric cancer patients. *Iran. J. Public Health* 40, 80–86.

Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield, M. B., Farouk, R., Lee, P. W., et al. (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet (Lond. Engl.)* 350, 469–472. doi: 10.1016/S0140-6736(96)11196-X

Brat, D. J., Verhaak, R. G. W., Aldape, K. D., Yung, W. K. A., Salama, S. R., Cooper, L. A. D., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* 372, 2481–2498. doi: 10.1056/NEJMoa1402121

Briggs, M. T., Condina, M. R., Ho, Y. Y., Everest-Dass, A. V., Mittal, P., Kaur, G., et al. (2019). MALDI Mass Spectrometry Imaging of Early- and Late-Stage Serous Ovarian Cancer Tissue Reveals Stage-Specific N-Glycans. *Proteomics* 19, e1800482. doi: 10.1002/pmic.201800482

Brinker, T. J., Hekler, A., Enk, A. H., Berking, C., Haferkamp, S., Hauschild, A., et al. (2019). Deep neural networks are superior to dermatologists in melanoma image classification. *Eur. J. Cancer* 119, 11–17. doi: 10.1016/j.ejca.2019.05.023

Buchhalter, I., Hutter, B., Alioto, T. S., Beck, T. A., Boutros, P. C., Brors, B., et al. (2014). A comprehensive multicenter comparison of whole genome sequencing pipelines using a uniform tumor-normal sample pair. *Cold Spring Harbor Labs J.* 013177. doi: 10.1101/013177

Calligaris, D., Feldman, D. R., Norton, I., Brastianos, P. K., Dunn, I. F., Santagata, S., et al. (2015). Molecular typing of meningiomas by desorption electrospray ionization mass spectrometry imaging for surgical decision-making. *Int. J. Mass Spectrom.* 377, 690–698. doi: 10.1016/j.ijms.2014.06.024

Cancer Moonshot - National Cancer Institute (2016). *Natl. Cancer Inst.* Available at: https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative (Accessed April 19, 2020).

Carter, C. L., Jones, J. W., Farese, A. M., MacVittie, T. J., and Kane, M. A. (2017). Lipidomic dysregulation within the lung parenchyma following whole-thorax lung irradiation: Markers of injury, inflammation and fibrosis detected by MALDI-MSI. *Sci. Rep.* 7, 10343. doi: 10.1038/s41598-017-10396-w

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature* 538, 20–23. doi: 10.1038/538020a

cBioPortal for Cancer Genomics. Available at: http://www.cbioportal.org/ (Accessed March 24, 2020).

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi: 10.1158/2159-8290.CD-12-0095

Chakraborty, S., Hosen, M.II, Ahmed, M., and Shekhar, H. U. (2018). Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *BioMed. Res. Int.* 2018:9836256. doi: 10.1155/2018/9836256

Chang, S.-W., Kareem, S. A., Kallarakkal, T. G., Merican, A. F. M. A., Abraham, M. T., and Zain, R. B. (2011). Feature selection methods for optimizing clinicopathologic input variables in oral cancer prognosis. *Asian Pac. J. Cancer Prev.* 12, 2659–2664.

Chang, E. K., Yu, C. Y., Clarke, R., Hackbarth, A., Sanders, T., Esrailian, E., et al. (2016). Defining a Patient Population With Cirrhosis. *J. Clin. Gastroenterol.* 50, 889–894. doi: 10.1097/MCG.0000000000000583

Chang, K., Bai, H. X., Zhou, H., Su, C., Bi, W. L., Agbodza, E., et al. (2018). Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from mr imaging. *Clin. Cancer Res.* 24, 1073–1081. doi: 10.1158/1078-0432.CCR-17-2236

Chaudhary, K., Poirion, O. B., Lu, L., and Garmire, L. X. (2018). Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res.* 24, 1248–1259. doi: 10.1158/1078-0432.CCR-17-0853

Chorell, E., Hall, U. A., Gustavsson, C., Berntorp, K., Puhkala, J., Luoto, R., et al. (2017). Pregnancy to postpartum transition of serum metabolites in women with gestational diabetes. *Metabolism* 72, 27–36. doi: 10.1016/j.metabol.2016.12.018

Chughtai, K., Jiang, L., Greenwood, T. R., Glunde, K., and Heeren, R. M. A. (2013). Mass spectrometry images acylcarnitines, phosphatidylcholines, and sphingomyelin in MDA-MB-231 breast tumor models. *J. Lipid Res.* 54, 333–344. doi: 10.1194/jlr.M027961

Clark, A. R., Calligaris, D., Regan, M. S., Pomeranz Krummel, D., Agar, J. N., Kallay, L., et al. (2018). Rapid discrimination of pediatric brain tumors by mass spectrometry imaging. *J. Neurooncol.* 140, 269–279. doi: 10.1007/s11060-018-2978-2

Cohen, J. D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., et al. (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Sci. (80-.).* 359, 926–930. doi: 10.1126/science.aar3247

Cornett, D. S., Mobley, J. A., Dias, E. C., Andersson, M., Arteaga, C. L., Sanders, M. E., et al. (2006). A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol. Cell. Proteomics* 5, 1975–1983. doi: 10.1074/mcp.M600119-MCP200

Costello, J. C., Heiser, L. M., Georgii, E., Gönen, M., Menden, M. P., Wang, N. J., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* 32, 1202–1212. doi: 10.1038/nbt.2877

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24, 1559–1567. doi: 10.1038/s41591-018-0177-5

Cui, M., Wang, Q., and Chen, G. (2016). Serum metabolomics analysis reveals changes in signaling lipids in breast cancer patients. *Biomed. Chromatogr.* 30, 42–47. doi: 10.1002/bmc.3556

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346–352. doi: 10.1038/nature10983

Daemen, A., Griffith, O. L., Heiser, L. M., Wang, N. J., Enache, O. M., Sanborn, Z., et al. (2013). Modeling precision treatment of breast cancer. *Genome Biol.* 14, R110. doi: 10.1186/gb-2013-14-10-r110

Delcourt, V., Franck, J., Leblanc, E., Narducci, F., Robin, Y. M., Gimeno, J. P., et al. (2017). Combined Mass Spectrometry Imaging and Top-down Microproteomics Reveals Evidence of a Hidden Proteome in Ovarian Cancer. *EBioMedicine* 21, 55–64. doi: 10.1016/j.ebiom.2017.06.001

Delen, D., Walker, G., and Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artif. Intell. Med.* 34, 113–127. doi: 10.1016/j.artmed.2004.07.002

Dietterich, T. G., and Bakiri, G. (1995). *Solving Multiclass Learning Problems via Error-Correcting Output Codes* (AI Access Foundation and Morgan Kaufmann Publishers).

Doerr, A. (2019). Single-cell proteomics. *Nat. Methods* 16, 20. doi: 10.1038/s41592-018-0273-y

Dória, M. L., McKenzie, J. S., Mroz, A., Phelps, D. L., Speller, A., Rosini, F., et al. (2016). Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Sci. Rep.* 6. doi: 10.1038/srep39219

Duan, K. B., Rajapakse, J. C., Wang, H., and Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobiosci.* 4, 228–233. doi: 10.1109/TNB.2005.853657

Eisner, R., Greiner, R., Tso, V., Wang, H., and Fedorak, R. N. (2013). A Machine-Learned Predictor of Colonic Polyps Based on Urinary Metabolomics. *BioMed. Res. Int.* 2013, 11. doi: 10.1155/2013/303982

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056

Eun, B., Sampley, M. L., Good, A. L., Gebert, C. M., and Pfeifer, K. (2013). Promoter cross-talk via a shared enhancer explains paternally biased expression of Nctc1 at the Igf2/H19/Nctc1 imprinted locus. *Nucleic Acids Res.* 41, 817–826. doi: 10.1093/nar/gks1182

Falgreen, S., Dybkær, K., Young, K. H., Xu-Monette, Z. Y., El-Galaly, T. C., Laursen, M. B., et al. (2015). Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer* 15, 235. doi: 10.1186/s12885-015-1237-6

Fayazfar, S., Zali, H., Arefi Oskouie, A., Asadzadeh Aghdaei, H., Rezaei Tavirani, M., and Nazemalhosseini Mojarad, E. (2019). Early diagnosis of colorectal cancer via plasma proteomic analysis of CRC and advanced adenomatous polyp. *Gastroenterol. Hepatol. Bed Bench* 12, 328–339.

Ferroni, P., Zanzotto, F. M., Riondino, S., Scarpato, N., Guadagni, F., and Roselli, M. (2019). Breast cancer prognosis using a machine learning approach. *Cancers (Basel).* 11. doi: 10.3390/cancers11030328

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., et al. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 43, D805–D811. doi: 10.1093/nar/gku1075

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, pl1. doi: 10.1126/scisignal.2004088

Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., et al. (2019). Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* 179, 561–577.e22. doi: 10.1016/j.cell.2019.08.052

Garvin, J. H., Kim, Y., Gobbel, G. T., Matheny, M. E., Redd, A., Bray, B. E., et al. (2018). Automating quality measures for heart failure using natural language processing:a descriptive study in the department of veterans affairs. *J. Med. Internet Res.* 20, e5. doi: 10.2196/medinform.9150

Geeleher, P., Cox, N. J., and Huang, R. S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* 15, R47. doi: 10.1186/gb-2014-15-3-r47

Goldman, M., Craft, B., Swatloski, T., Ellrott, K., Cline, M., Diekhans, M., et al. (2013). The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Res.* 41, D949–D954. doi: 10.1093/nar/gks1008

Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M. P., Jene-Sanz, A., et al. (2013). IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* 10, 1081–1082. doi: 10.1038/nmeth.2642

Guan, W., Zhou, M., Hampton, C. Y., Benigno, B. B., Walker, L. D. E., Gray, A., et al. (2009). Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinf.* 10:259. doi: 10.1186/1471-2105-10-259

Guenther, S., Muirhead, L. J., Speller, A. V. M., Golf, O., Strittmatter, N., Ramakrishnan, R., et al. (2015). Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Res.* 75, 1828–1837. doi: 10.1158/0008-5472.CAN-14-2258

Guo, S., Wang, Y., Zhou, D., and Li, Z. (2014). Significantly increased monounsaturated lipids relative to polyunsaturated lipids in six types of cancer microenvironment are observed by mass spectrometry imaging. *Sci. Rep.* 4, 5959. doi: 10.1038/srep05959

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., et al. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* 29, 1836–1842. doi: 10.1093/annonc/mdy166

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18, 1–15. doi: 10.1186/s13059-017-1215-1

Hasnain, Z., Mason, J., Gill, K., Miranda, G., Gill, I. S., Kuhn, P., et al. (2019). Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients. *PLoS One* 14, e0210976.. doi: 10.1371/journal.pone.0210976

He, X., Chang, S., Zhang, J., Zhao, Q., Xiang, H., Kusonmano, K., et al. (2008). MethyCancer: The database of human DNA methylation and cancer. *Nucleic Acids Res.* 36, D836–D841. doi: 10.1093/nar/gkm730

Heinemann, J., Mazurie, A., Tokmina-Lukaszewska, M., Beilman, G. J., and Bothner, B. (2014). Application of support vector machines to metabolomics experiments with limited replicates. *Metabolomics* 10, 1121–1128. doi: 10.1007/s11306-014-0651-0

Hinsch, A., Buchholz, M., Odinga, S., Borkowski, C., Koop, C., Izbicki, J. R., et al. (2017). MALDI imaging mass spectrometry reveals multiple clinically relevant masses in colorectal cancer using large-scale tissue microarrays. *J. Mass Spectrom.* 52, 165–173. doi: 10.1002/jms.3916

Hiraide, T., Ikegami, K., Sakaguchi, T., Morita, Y., Hayasaka, T., Masaki, N., et al. (2016). Accumulation of arachidonic acid-containing phosphatidylinositol at the outer edge of colorectal cancer. *Sci. Rep.* 6, 29935. doi: 10.1038/srep29935

Holzlechner, M., Bonta, M., Lohninger, H., Limbeck, A., and Marchetti-Deschmann, M. (2018). Multisensor Imaging-From Sample Preparation to Integrated Multimodal Interpretation of LA-ICPMS and MALDI MS Imaging Data. *Anal. Chem.* 90, 8831–8837. doi: 10.1021/acs.analchem.8b00816

Huang, S., Chaudhary, K., and Garmire, L. X. (2017). More is better: Recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84. doi: 10.3389/fgene.2017.00084

Huang, L., Mao, X., Sun, C., Luo, Z., Song, X., Li, X., et al. (2019). A graphical data processing pipeline for mass spectrometry imaging-based spatially resolved metabolomics on tumor heterogeneity. *Anal. Chim. Acta* 1077, 183–190. doi: 10.1016/j.aca.2019.05.068

Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Lett.* 471, 61–71. doi: 10.1016/j.canlet.2019.12.007

Huber, K., Feuchtinger, A., Borgmann, D. M., Li, Z., Aichler, M., Hauck, S. M., et al. (2014). Novel approach of MALDI drug imaging, immunohistochemistry, and digital image analysis for drug distribution studies in tissues. *Anal. Chem.* 86, 10568–10575. doi: 10.1021/ac502177y

Huttunen, M. J., Hassan, A., McCloskey, C. W., Fasih, S., Upham, J., Vanderhyden, B. C., et al. (2018). Automated classification of multiphoton microscopy images of ovarian tissue using deep learning. *J. Biomed. Opt.* 23, 1. doi: 10.1117/1.jbo.23.6.066002

Ibrahim, A., Gamble, P., Jaroensri, R., Abdelsamea, M. M., Mermel, C. H., Chen, P.-H. C., et al. (2020). Artificial intelligence in digital breast pathology: Techniques and applications. *Breast* 49, 267–273. doi: 10.1016/j.breast.2019.12.007

Ikeda, A., Nishiumi, S., Shinohara, M., Yoshie, T., Hatano, N., Okuno, T., et al. (2012). Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *Biomed. Chromatogr.* 26, 548–558. doi: 10.1002/bmc.1671

Inglese, P., McKenzie, J. S., Mroz, A., Kinross, J., Veselkov, K., Holmes, E., et al. (2017). Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer††Electronic supplementary information (ESI) available. See DOI: 10.1039/c6sc03738kClick here for additional data file.Click here for additional data file.Click here for. *Chem. Sci.* 8, 3500–3511. doi: 10.1039/c6sc03738k

Jarmusch, A. K., Alfaro, C. M., Pirro, V., Hattab, E. M., Cohen-Gadol, A. A., and Cooks, R. G. (2016). Differential Lipid Profiles of Normal Human Brain Matter and Gliomas by Positive and Negative Mode Desorption Electrospray Ionization - Mass Spectrometry Imaging. *PLoS One* 11, e0163180. doi: 10.1371/journal.pone.0163180

Jayaram, S., Gupta, M. K., Raju, R., Gautam, P., and Sirdeshmukh, R. (2016). Multi-Omics Data Integration and Mapping of Altered Kinases to Pathways Reveal Gonadotropin Hormone Signaling in Glioblastoma. *OMICS* 20, 736–746. doi: 10.1089/omi.2016.0142

Jiang, N., and Xu, X. (2019). Exploring the survival prognosis of lung adenocarcinoma based on the cancer genome atlas database using artificial neural network. *Med. (Baltimore).* 98, e15642. doi: 10.1097/MD.0000000000015642

Jiang, L., Chughtai, K., Purvine, S. O., Bhujwalla, Z. M., Raman, V., Paša-Tolić, L., et al. (2015). MALDI-Mass Spectrometric Imaging Revealing Hypoxia-Driven Lipids and Proteins in a Breast Tumor Model. *Anal. Chem.* 87, 5947–5956. doi: 10.1021/ac504503x

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* 2, 230–243. doi: 10.1136/svn-2017-000101

Jović, S., Miljković, M., Ivanović, M., Šaranović, M., and Arsić, M. (2017). Prostate Cancer Probability Prediction By Machine Learning Technique. *Cancer Invest.* 35, 647–651. doi: 10.1080/07357907.2017.1406496

Karsan, A., Eigl, B. J., Flibotte, S., Gelmon, K., Switzer, P., Hassell, P., et al. (2005). Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin. Chem.* 51, 1525–1528. doi: 10.1373/clinchem.2005.050708

Kaseb, A. O., Sánchez, N. S., Sen, S., Kelley, R. K., Tan, B., Bocobo, A. G., et al. (2019). Molecular Profiling of Hepatocellular Carcinoma Using Circulating Cell-Free DNA. *Clin. Cancer Res.* 25, 6107–6118. doi: 10.1158/1078-0432.CCR1-18-3341

Kelley, D. Z., Flam, E. L., Izumchenko, E., Danilova, L. V., Wulf, H. A., Guo, T., et al. (2017). Integrated analysis of whole-genome ChIP-Seq and RNA-Seq data of primary head and neck tumor samples associates HPV integration sites with open chromatin marks. *Cancer Res.* 77, 6538–6550. doi: 10.1158/0008-5472.CAN-17-0833

Kelly, R. S., Heiden, M. G. V., Giovannucci, E., and Mucci, L. A. (2016). Metabolomic biomarkers of prostate cancer: Prediction, diagnosis, progression, prognosis, and recurrence. *Cancer Epidemiol. Biomarkers Prev.* 25, 887–906. doi: 10.1158/1055-9965.EPI-15-1223

Kim, D., Joung, J.-G., Sohn, K.-A., Shin, H., Park, Y. R., Ritchie, M. D., et al. (2015). Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction. *J. Am. Med. Inform. Assoc.* 22, 109–120. doi: 10.1136/amiajnl-2013-002481

Klughammer, J., Kiesel, B., Roetzer, T., Fortelny, N., Nemc, A., Nenning, K.-H., et al. (2018). The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat. Med.* 24, 1611–1624. doi: 10.1038/s41591-018-0156-x

Korbar, B., Olofson, A., Miraflor, A., Nicka, C., Suriawinata, M., Torresani, L., et al. (2017). Deep learning for classification of colorectal polyps on whole-slide images. *J. Pathol. Inform.* 8, 30. doi: 10.4103/jpi.jpi_34_17

Koutros, S., Meyer, T. E., Fox, S. D., Issaq, H. J., Veenstra, T. D., Huang, W.-Y., et al. (2013). Prospective evaluation of serum sarcosine and risk of prostate cancer in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Carcinogenesis* 34, 2281–2285. doi: 10.1093/carcin/bgt176

Kriegsmann, M., Casadonte, R., Kriegsmann, J., Dienemann, H., Schirmacher, P., Hendrik Kobarg, J., et al. (2016). Reliable Entity Subtyping in Non-small Cell Lung Cancer by Matrix-assisted Laser Desorption/Ionization Imaging Mass Spectrometry on Formalin-fixed Paraffin-embedded Tissue Specimens. *Mol. Cell. Proteomics* 15, 3081–3089. doi: 10.1074/mcp.M115.057513

Kunzke, T., Balluff, B., Feuchtinger, A., Buck, A., Langer, R., Luber, B., et al. (2017). Native glycan fragments detected by MALDI-FT-ICR mass spectrometry imaging impact gastric cancer biology and patient outcome. *Oncotarget* 8, 68012–68025. doi: 10.18632/oncotarget.19137

Kuo, R.-J., Huang, M.-H., Cheng, W.-C., Lin, C.-C., and Wu, Y.-H. (2015). Application of a two-stage fuzzy neural network to a prostate cancer prognosis system. *Artif. Intell. Med.* 63, 119–133. doi: 10.1016/j.artmed.2014.12.008

Lai, C.-H., Chang, N.-W., Lin, C.-F., Lin, C.-D., Lin, Y.-J., Wan, L., et al. (2010). Proteomics-based identification of haptoglobin as a novel plasma biomarker in oral squamous cell carcinoma. *Clin. Chim. Acta* 411, 984–991. doi: 10.1016/j.cca.2010.03.028

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72. doi: 10.1038/nature07485

Li, Y., Song, X., Zhao, X., Zou, L., and Xu, G. (2014). Serum metabolic profiling study of lung cancer using ultra high performance liquid chromatography/quadrupole time-of-flight mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 966, 147–153. doi: 10.1016/j.jchromb.2014.04.047

Li, Q., Freeman, L. M., Rush, J. E., Huggins, G. S., Kennedy, A. D., Labuda, J. A., et al. (2015). Veterinary Medicine and Multi-Omics Research for Future Nutrition Targets: Metabolomics and Transcriptomics of the Common Degenerative Mitral Valve Disease in Dogs. *OMICS* 19, 461–470. doi: 10.1089/omi.2015.0057

Li, T., He, J., Mao, X., Bi, Y., Luo, Z., Guo, C., et al. (2015). In situ biomarker discovery and label-free molecular histopathological diagnosis of lung cancer by ambient mass spectrometry imaging. *Sci. Rep.* 5. doi: 10.1038/srep14089

Li, Y., Wu, F. X., and Ngom, A. (2018). A review on machine learning principles for multi-view biological data integration. *Brief. Bioinform.* 19, 325–340. doi: 10.1093/bib/bbw113

Li, H., Siddiqui, O., Zhang, H., and Guan, Y. (2019). Joint learning improves protein abundance prediction in cancers. *BMC Biol.* 17, 107. doi: 10.1186/s12915-019-0730-9

Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., et al. (2019). Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol.* 20, 193–201. doi: 10.1016/S1470-2045(18)30762-9

Li, L., Zheng, X., Zhou, Q., Villanueva, N., Nian, W., Liu, X., et al. (2020). Metabolomics-Based Discovery of Molecular Signatures for Triple Negative Breast Cancer in Asian Female Population. *Sci. Rep.* 10, 370. doi: 10.1038/s41598-019-57068-5

Liang, C. A., Chen, L., Wahed, A., and Nguyen, A. N. D. (2019). Proteomics analysis of FLT3-ITD mutation in acute myeloid leukemia using deep learning neural network. *Ann. Clin. Lab. Sci.* 49, 119–126. doi: 10.1093/ajcp/aqx121.148

Liebl, H. (1967). Ion microprobe mass analyzer. *J. Appl. Phys.* 38, 5277–5283. doi: 10.1063/1.1709314

Lin, X., Wang, Q., Yin, P., Tang, L., Tan, Y., Li, H., et al. (2011). A method for handling metabonomics data from liquid chromatography/mass spectrometry: Combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection. *Metabolomics* 7, 549–558. doi: 10.1007/s11306-011-0274-7

Liu, Y., Gonzàlez-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., et al. (2017). Impact of Alternative Splicing on the Human Proteome. *Cell Rep.* 20, 1229–1241. doi: 10.1016/j.celrep.2017.07.025

López de Maturana, E., Alonso, L., Alarcón, P., Martín-Antoniano, I. A., Pineda, S., Piorno, L., et al. (2019). Challenges in the Integration of Omics and Non-Omics Data. *Genes (Basel).* 10. doi: 10.3390/genes10030238

Lou, S., Balluff, B., Cleven, A. H. G., Bovée, J. V. M. G., and McDonnell, L. A. (2017). Prognostic Metabolite Biomarkers for Soft Tissue Sarcomas Discovered by Mass Spectrometry Imaging. *J. Am. Soc Mass Spectrom.* 28, 376–383. doi: 10.1007/s13361-016-1544-4

Low, S.-K., Zembutsu, H., and Nakamura, Y. (2018). Breast cancer: The translation of big genomic data to cancer precision medicine. *Cancer Sci.* 109, 497–506. doi: 10.1111/cas.13463

Lu, C. F., Hsu, F. T., Hsieh, K. L. C., Kao, Y. C. J., Cheng, S. J., Hsu, J. B. K., et al. (2018). Machine learning–based radiomics for molecular subtyping of gliomas. *Clin. Cancer Res.* 24, 4429–4436. doi: 10.1158/1078-0432.CCR-17-3445

Lu, Y., Yu, Q., Gao, Y., Zhou, Y., Liu, G., Dong, Q., et al. (2018). Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. *Cancer Res.* 78, 5135–5143. doi: 10.1158/0008-5472.CAN-18-0494

Lu, T.-P., Kuo, K.-T., Chen, C.-H., Chang, M.-C., Lin, H.-P., Hu, Y.-H., et al. (2019). Developing a Prognostic Gene Panel of Epithelial Ovarian Cancer Patients by a Machine Learning Model. *Cancers (Basel).* 11. doi: 10.3390/cancers11020270

Lynch, C. M., Abdollahi, B., Fuqua, J. D., de Carlo, A. R., Bartholomai, J. A., Balgemann, R. N., et al. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int. J. Med. Inform.* 108, 1–8. doi: 10.1016/j.ijmedinf.2017.09.013

Ma, Y., Zhang, P., Wang, F., Liu, W., Yang, J., and Qin, H. (2012). An integrated proteomics and metabolomics approach for defining oncofetal biomarkers in the colorectal cancer. *Ann. Surg.* 255, 720–730. doi: 10.1097/SLA.0b013e31824a9a8b

Maes, E., Cho, W. C., and Baggerman, G. (2015). Translating clinical proteomics: the importance of study design. *Expert Rev. Proteomics* 12, 217–219. doi: 10.1586/14789450.2015.1041512

Mallavarapu, T., Hao, J., Kim, Y., Oh, J. H., and Kang, M. (2019). Pathway-based deep clustering for molecular subtyping of cancer. *Methods* 173, 24–31. doi: 10.1016/j.ymeth.2019.06.017

Manna, S. K., Tanaka, N., Krausz, K. W., Haznadar, M., Xue, X., Matsubara, T., et al. (2014). Biomarkers of coordinate metabolic reprogramming in colorectal tumors in mice and humans. *Gastroenterology* 146, 1313–1324. doi: 10.1053/j.gastro.2014.01.017

Margulis, K., Chiou, A. S., Aasi, S. Z., Tibshirani, R. J., Tang, J. Y., and Zare, R. N. (2018). Distinguishing malignant from benign microscopic skin lesions using desorption electrospray ionization mass spectrometry imaging. *Proc. Natl. Acad. Sci. U. S. A.* 115, 6347–6352. doi: 10.1073/pnas.1803733115

Martínez-Bartolomé, S., Binz, P.-A., and Albar, J. P. (2014). The Minimal Information about a Proteomics Experiment (MIAPE) from the Proteomics Standards Initiative. *Methods Mol. Biol.* 1072, 765–780. doi: 10.1007/978-1-62703-631-3_53

Martorell-Marugán, J., Tabik, S., Benhammou, Y., del Val, C., Zwir, I., Herrera, F., et al. (2019). "Deep Learning in Omics Data Analysis and Precision Medicine," in *Computational Biology*. Ed. H. Husi (Brisbane (AU): Codon Publications).

Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534, 55–62. doi: 10.1038/nature18003

Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep26094

Mishra, S., Srivastava, A. K., Suman, S., Kumar, V., and Shukla, Y. (2015). Circulating miRNAs revealed as surrogate molecular signatures for the early detection of breast cancer. *Cancer Lett.* 369, 67–75. doi: 10.1016/j.canlet.2015.07.045

Moarii, M., Boeva, V., Vert, J.-P., and Reyal, F. (2015). Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics* 16, 873. doi: 10.1186/s12864-015-1994-2

Mondul, A. M., Moore, S. C., Weinstein, S. J., Männistö, S., Sampson, J. N., and Albanes, D. (2014). 1-Stearoylglycerol is associated with risk of prostate cancer: Results from a serum metabolomic profiling analysis. *Metabolomics* 10, 1036–1041. doi: 10.1007/s11306-014-0643-0

More, T., RoyChoudhury, S., Gollapalli, K., Patel, S. K., Gowda, H., Chaudhury, K., et al. (2015). Metabolomics and its integration with systems biology: PSI 2014 conference panel discussion report. *J. Proteomics* 127, 73–79. doi: 10.1016/j.jprot.2015.04.024

Morris, V., Rao, X., Pickering, C., Foo, W. C., Rashid, A., Eterovic, K., et al. (2017). Comprehensive Genomic Profiling of Metastatic Squamous Cell Carcinoma of the Anal Canal. *Mol. Cancer Res.* 15, 1542–1550. doi: 10.1158/1541-7786.MCR-17-0060

Murata, T., Yanagisawa, T., Kurihara, T., Kaneko, M., Ota, S., Enomoto, A., et al. (2019). Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. *Breast Cancer Res. Treat.* 177, 591–601. doi: 10.1007/s10549-019-05330-9

Nam, H., Chung, B. C., Kim, Y., Lee, K., and Lee, D. (2009). Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. *Bioinformatics* 25, 3151–3157. doi: 10.1093/bioinformatics/btp558

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54. doi: 10.1038/nature17676

Nishiumi, S., Kobayashi, T., Ikeda, A., Yoshie, T., Kibi, M., Izumi, Y., et al. (2012). A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS One* 7. doi: 10.1371/journal.pone.0040459

Osborne, J. D., Wyatt, M., Westfall, A. O., Willig, J., Bethard, S., and Gordon, G. (2016). Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning. *J. Am. Med. Inf. Assoc.* 23, 1077–1084. doi: 10.1093/jamia/ocw006

Paine, M. R. L., Liu, J., Huang, D., Ellis, S. R., Trede, D., Kobarg, J. H., et al. (2019). Three-Dimensional Mass Spectrometry Imaging Identifies Lipid Markers of Medulloblastoma Metastasis. *Sci. Rep.* 9, 2205. doi: 10.1038/s41598-018-38257-0

Palubeckaitė, I., Crooks, L., Smith, D. P., Cole, L. M., Bram, H., Le Maitre, C., et al. (2020). Mass spectrometry imaging of endogenous metabolites in response to doxorubicin in a novel 3D osteosarcoma cell culture model. *J. Mass Spectrom.* 55, e4461. doi: 10.1002/jms.4461

Paolillo, C., Londin, E., and Fortina, P. (2016). Next generation sequencing in cancer: opportunities and challenges for precision cancer medicine. *Scand. J. Clin. Lab. Invest. Suppl.* 245, S84–S91. doi: 10.1080/00365513.2016.1210331

Park, K., Ali, A., Kim, D., An, Y., Kim, M., and Shin, H. (2013). Robust predictive model for evaluating breast cancer survivability. *Eng. Appl. Artif. Intell.* 26, 2194–2205. doi: 10.1016/j.engappai.2013.06.013

Patel, S. K., Rajora, N., Kumar, S., Sahu, A., Kochar, S. K., Krishna, C. M., et al. (2019). Rapid Discrimination of Malaria- and Dengue-Infected Patients Sera Using Raman Spectroscopy. *Anal. Chem.* 91, 7054–7062. doi: 10.1021/acs.analchem.8b05907

Peng, L., Cantor, D.II, Huang, C., Wang, K., Baker, M. S., and Nice, E. C. (2018). Tissue and plasma proteomics for early stage cancer detection. *Mol. Omi.* 14, 405–423. doi: 10.1039/c8mo00126j

Perakakis, N., Yazdani, A., Karniadakis, G. E., and Mantzoros, C. (2018). Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism* 87, A1–A9. doi: 10.1016/j.metabol.2018.08.002

Perng, W., Rifas-Shiman, S. L., McCulloch, S., Chatzi, L., Mantzoros, C., Hivert, M. F., et al. (2017). Associations of cord blood metabolites with perinatal characteristics, newborn anthropometry, and cord blood hormones in project viva. *Metabolism* 76, 11–22. doi: 10.1016/j.metabol.2017.07.001

Poirion, O. B., Chaudhary, K., Huang, S., and Garmire, L. X. (2019). Multi-omics-based pan-cancer prognosis prediction using an ensemble of deep-learning and machine-learning models. *medRxiv*, 19010082. doi: 10.1101/19010082

Porta Siegel, T., Hamm, G., Bunch, J., Cappell, J., Fletcher, J. S., and Schwamborn, K. (2018). Mass Spectrometry Imaging and Integration with Other Imaging Modalities for Greater Molecular Understanding of Biological Tissues. *Mol. Imaging Biol.* 20, 888–901. doi: 10.1007/s11307-018-1267-y

Rai, V., Karthikaichamy, A., Das, D., Noronha, S., Wangikar, P. P., and Srivastava, S. (2016). Multi-omics Frontiers in Algal Research: Techniques and Progress to Explore Biofuels in the Postgenomics World. *OMICS* 20, 387–399. doi: 10.1089/omi.2016.0065

Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S., and Sidow, A. (2018). Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat. Commun.* 9, 4453. doi: 10.1038/s41467-018-06921-8

Ribli, D., Horváth, A., Unger, Z., Pollner, P., and Csabai, I. (2018). Detecting and classifying lesions in mammograms with Deep Learning. *Sci. Rep.* 8, 4165. doi: 10.1038/s41598-018-22437-z

Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., et al. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228. doi: 10.1016/j.cell.2015.05.001

Robinson, D. R., Wu, Y.-M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., et al. (2017). Integrative clinical genomics of metastatic cancer. *Nature* 548, 297–303. doi: 10.1038/nature23306

Romo-Bucheli, D., Janowczyk, A., Gilmore, H., Romero, E., and Madabhushi, A. (2017). A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry A* 91, 566–573. doi: 10.1002/cyto.a.23065

Rubio-Perez, C., Tamborero, D., Schroeder, M. P., Antolín, A. A., Deu-Pons, J., Perez-Llamas, C., et al. (2015). In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell* 27, 382–396. doi: 10.1016/j.ccell.2015.02.007

Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., et al. (2019). A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Rep.* 29, 3367–3373.e4. doi: 10.1016/j.celrep.2019.11.017

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., et al. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* 23, 181–193.e7. doi: 10.1016/j.celrep.2018.03.086

Schoof, E. M., Rapin, N., Savickas, S., Gentil, C., Lechman, E., Haile, J. S., et al. (2019). A Quantitative Single-Cell Proteomics Approach to Characterize an Acute Myeloid Leukemia Hierarchy. *bioRxiv*, 745679. doi: 10.1101/745679

Sevakula, R. K., Singh, V., Verma, N. K., Kumar, C., and Cui, Y. (2019). Transfer Learning for Molecular Cancer Classification Using Deep Neural Networks. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 16, 2089–2100. doi: 10.1109/TCBB.2018.2822803

Sirinukunwattana, K., Ahmed Raza, S. E., Tsang, Y.-W., Snead, D. R. J., Cree, I. A., and Rajpoot, N. M. (2016). Locality Sensitive Deep Learning for Detection and

Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Trans. Med. Imaging* 35, 1196–1206. doi: 10.1109/TMI.2016.2525803

Sohni, A., Bartoccetti, M., Khoueiry, R., Spans, L., Vande Velde, J., De Troyer, L., et al. (2015). Dynamic Switching of Active Promoter and Enhancer Domains Regulates Tet1 and Tet2 Expression during Cell State Transitions between Pluripotency and Differentiation. *Mol. Cell. Biol.* 35, 1026–1042. doi: 10.1128/mcb.01172-14

Stemke-Hale, K., Gonzalez-Angulo, A. M., Lluch, A., Neve, R. M., Kuo, W.-L., Davies, M., et al. (2008). An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* 68, 6084–6091. doi: 10.1158/0008-5472.CAN-07-6854

Stoeckli, M., Chaurand, P., Hallahan, D. E., and Caprioli, R. M. (2001). Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. *Nat. Med.* 7, 493–496. doi: 10.1038/86573

Sun, D., Wang, M., and Li, A. (2018). A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 16, 841–850. doi: 10.1109/TCBB.2018.2806438

Sun, C., Li, T., Song, X., Huang, L., Zang, Q., Xu, J., et al. (2019a). Spatially resolved metabolomics to discover tumor-associated metabolic alterations. *Proc. Natl. Acad. Sci. U. S. A.* 116, 52–57. doi: 10.1073/pnas.1808950116

Sun, N., Kunzke, T., Sbiera, S., Kircher, S., Feuchtinge, A., Aichler, M., et al. (2019b). Prognostic relevance of steroid sulfation in adrenocortical carcinoma revealed by molecular phenotyping using high-resolution mass spectrometry imaging. *Clin. Chem.* 65, 1276–1286. doi: 10.1373/clinchem.2019.306043

Syrjala, K. L. (2018). Opportunities for improving oncology care. *Lancet Oncol.* 19, 449. doi: 10.1016/S1470-2045(18)30208-0

Therapeutically Applicable Research to Generate Effective Treatments (TARGET. Available at: https://ocg.cancer.gov/programs/target (Accessed March 24, 2020).

Thorsen, S. F., Gromova, I., Christensen, I. J., Fredriksson, S., Andersen, C. L., Nielsen, H. J., et al. (2019). Gel-Based Proteomics of Clinical Samples Identifies Potential Serological Biomarkers for Early Detection of Colorectal Cancer. *Int. J. Mol. Sci.* 20. doi: 10.3390/ijms20236082

Torata, N., Kubo, M., Miura, D., Ohuchida, K., Mizuuchi, Y., Fujimura, Y., et al. (2018). Visualizing energy charge in breast carcinoma tissues by MALDI mass-spectrometry imaging profiles of low-molecular-weight metabolites, in *Anticancer Research* (International Institute of Anticancer Research), 4267–4272. doi: 10.21873/anticanres.12723

Tsai, M.-H., Wu, C.-C., Peng, P.-H., Liang, Y., Hsiao, Y.-C., Chien, K.-Y., et al. (2012). Identification of secretory gelsolin as a plasma biomarker associated with distant organ metastasis of colorectal cancer. *J. Mol. Med. (Berl).* 90, 187–200. doi: 10.1007/s00109-011-0817-4

Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* 20, 404–416. doi: 10.1038/s41576-019-0114-6

Turki, T., Wei, Z., and Wang, J. T. L. (2018). A transfer learning approach via procrustes analysis and mean shift for cancer drug sensitivity prediction. *J. Bioinform. Comput. Biol.* 16:1840014. doi: 10.1142/S0219720018400140

Uchiyama, Y., Hayasaka, T., Masaki, N., Watanabe, Y., Masumoto, K., Nagata, T., et al. (2014). Imaging mass spectrometry distinguished the cancer and stromal regions of oral squamous cell carcinoma by visualizing phosphatidylcholine (16:0/16:1) and phosphatidylcholine (18:1/20:4). *Anal. Bioanal. Chem.* 406 (5), 1307–1316. doi: 10.1007/s00216-013-7062-3

Van Emon, J. M. (2016). The Omics Revolution in Agricultural Research. *J. Agric. Food Chem.* 64, 36–44. doi: 10.1021/acs.jafc.5b04515

Vantaku, V., Dong, J., Ambati, C. R., Perera, D., Donepudi, S. R., Amara, C. S., et al. (2019). Multi-omics Integration Analysis Robustly Predicts High-Grade Patient Survival and Identifies CPT1B Effect on Fatty Acid Metabolism in Bladder Cancer. *Clin. Cancer Res.* 25, 3689–3701. doi: 10.1158/1078-0432.CCR-18-1515

Varambally, S., Dhanasekaran, S. M., Zhou, M., Barrette, T. R., Kumar-Sinha, C., Sanda, M. G., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419, 624–629. doi: 10.1038/nature01075

Vasudevan, P., and Murugesan, T. (2018). Cancer Subtype Discovery Using Prognosis-Enhanced Neural Network Classifier in Multigenomic Data. *Technol. Cancer Res. Treat.* 17:1533033818790509. doi: 10.1177/1533033818790509

Veselkov, K. A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., et al. (2011). Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal. Chem.* 83, 5864–5872. doi: 10.1021/ac201065j

Vidavsky, N., Kunitake, J. A. M. R., Diaz-Rubio, M. E., Chiou, A. E., Loh, H. C., Zhang, S., et al. (2019). Mapping and Profiling Lipid Distribution in a 3D Model of Breast Cancer Progression. *ACS Cent. Sci.* 5, 768–780. doi: 10.1021/acscentsci.8b00932

Vougas, K., Sakellaropoulos, T., Kotsinas, A., Foukas, G. R. P., Ntargaras, A., Koinis, F., et al. (2019). Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining. *Pharmacol. Ther.* 203, 107395. doi: 10.1016/j.pharmthera.2019.107395

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597

Wang, S., Chen, X., Luan, H., Gao, D., Lin, S., Cai, Z., et al. (2016). Matrix-assisted laser desorption/ionization mass spectrometry imaging of cell cultures for the lipidomic analysis of potential lipid markers in human breast cancer invasion. *Rapid Commun. Mass Spectrom.* 30, 533–542. doi: 10.1002/rcm.7466

Wang, Z., Jensen, M. A., and Zenklusen, J. C. (2016). "A practical guide to The Cancer Genome Atlas (TCGA)," in *Methods in Molecular Biology* (Humana Press Inc.), 111–141. doi: 10.1007/978-1-4939-3578-9_6

Wang, X., Han, J., Hardie, D. B., Yang, J., Pan, J., and Borchers, C. H. (2017). Metabolomic profiling of prostate cancer by matrix assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry imaging using Matrix Coating Assisted by an Electric Field (MCAEF). *Biochim. Biophys. Acta - Proteins Proteomics* 1865, 755–767. doi: 10.1016/j.bbapap.2016.12.012

Wang, Y., Wang, D., Ye, X., Wang, Y., Yin, Y., and Jin, Y. (2019). A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction. *Inf. Sci. (NY).* 474, 106–124. doi: 10.1016/j.ins.2018.09.046

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876. doi: 10.1038/nature06884

Wilks, C., Cline, M. S., Weiler, E., Diehkans, M., Craft, B., Martin, C., et al. (2014). The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database (Oxford)* 2014. doi: 10.1093/database/bau093

Wu, T., and Dai, Y. (2017). Tumor microenvironment and therapeutic response. *Cancer Lett.* 387, 61–68. doi: 10.1016/j.canlet.2016.01.043

Wu, C., Zhou, F., Ren, J., Li, X., Jiang, Y., and Ma, S. (2019). A selective review of multi-level omics data integration using variable selection. *High-Throughput* 8. doi: 10.3390/ht8010004

Wu, P., Heins, Z. J., Muller, J. T., Katsnelson, L., de Bruijn, I., Abeshouse, A. A., et al. (2019). Integration and Analysis of CPTAC Proteomics Data in the Context of Cancer Genomics in the cBioPortal. *Mol. Cell. Proteomics* 18, 1893–1898. doi: 10.1074/mcp.TIR119.001673

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., et al. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imaging* 35, 119–130. doi: 10.1109/TMI.2015.2458702

Xu, C., Zhou, D., Luo, Y., Guo, S., Wang, T., Liu, J., et al. (2017). Tissue and serum lipidome shows altered lipid composition with diagnostic potential in mycosis fungoides. *Oncotarget* 8, 48041–48050. doi: 10.18632/oncotarget.18228

Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., and Dawood, H. (2019). A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data. *BMC Bioinf.* 20, 527. doi: 10.1186/s12859-019-3116-7

Yamazaki, Y. (2015). Metabolome Analysis of Human Serum: Implications for Early Detection of Colorectal Cancer. *Rinsho Byori.* 63, 328–335.

Yang, L., Rong, W., Xiao, T., Zhang, Y., Xu, B., Liu, Y., et al. (2013). Secretory/releasing proteome-based identification of plasma biomarkers in HBV-associated hepatocellular carcinoma. *Sci. China Life Sci.* 56, 638–646. doi: 10.1007/s11427-013-4497-x

Yoon, J., Jordon, J., and Van Der Schaar, M. (2018). Ganite: Estimation of individualized treatment effects using generative adversarial nets. in *6th*

International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings.

Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., et al. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* 7, 12474. doi: 10.1038/ncomms12474

Yu, K.-H., Berry, G. J., Rubin, D. L., Ré, C., Altman, R. B., and Snyder, M. (2017). Association of Omics Features with Histopathology Patterns in Lung Adenocarcinoma. *Cell Syst.* 5, 620–627.e3. doi: 10.1016/j.cels.2017.10.014

Yu, P., Wilhelm, K., Dubrac, A., Tung, J. K., Alves, T. C., Fang, J. S., et al. (2017). FGF-dependent metabolic control of vascular development. *Nature* 545, 224–241. doi: 10.1038/nature22322

Yuan, Y., Failmezger, H., Rueda, O. M., Raza Ali, H., Gräf, S., Chin, S. F., et al. (2012). Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Transl. Med.* 4, 157ra143–157ra143. doi: 10.1126/scitranslmed.3004330

Zeng, I. S. L., and Lumley, T. (2018). Review of statistical learning methods in integrated omics studies (An integrated information science). *Bioinform. Biol. Insights* 12. doi: 10.1177/1177932218759292

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple "omics" analysis for microbial biology: Application and methodologies. *Microbiology* 156, 287–301. doi: 10.1099/mic.0.034793-0

Zhang, L., Jin, H., Guo, X., Yang, Z., Zhao, L., Tang, S., et al. (2012). Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by 1H nuclear magnetic resonance-based metabonomic profiles. *Clin. Biochem.* 45, 1064–1069. doi: 10.1016/j.clinbiochem.2012.05.012

Zhang, T., Wu, X., Ke, C., Yin, M., Li, Z., Fan, L., et al. (2013). Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *J. Proteome Res.* 12, 505–512. doi: 10.1021/pr3009572

Zhang, S., Xu, Y., Hui, X., Yang, F., Hu, Y., Shao, J., et al. (2017). Improvement in prediction of prostate cancer prognosis with somatic mutational signatures. *J. Cancer* 8, 3261–3267. doi: 10.7150/jca.21261

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* 37, 367–369. doi: 10.1038/s41587-019-0055-9

Zhang, X., Zhang, J., Sun, K., Yang, X., Dai, C., and Guo, Y. (2019). "Integrated Multi-omics Analysis Using Variational Autoencoders: Application to Pan-cancer Classification," in *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, 765–769. doi: 10.1109/BIBM47256.2019.8983228

Zhao, M., Tang, Y., Kim, H., and Hasegawa, K. (2018). Machine Learning With K-Means Dimensional Reduction for Predicting Survival Outcomes in Patients With Breast Cancer. *Cancer Inform.* 17:1176935118810215. doi: 10.1177/1176935118810215

Zhu, L., Luo, W., Su, M., Wei, H., Wei, J., Zhang, X., et al. (2013). Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomed. Rep.* 1, 757–760. doi: 10.3892/br.2013.140

Zhuang, J., Tang, X., Du, Z., Yang, M., and Zhou, Y. (2016). Prediction of biomarkers of therapeutic effects of patients with lung adenocarcinoma treated with gefitinib based on progression-free-survival by metabolomic fingerprinting. *Talanta* 160, 636–644. doi: 10.1016/j.talanta.2016.08.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Patel, George and Rai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 August 2020
doi: 10.3389/fonc.2020.01545

Combined Metabolomic Analysis of Plasma and Tissue Reveals a Prognostic Risk Score System and Metabolic Dysregulation in Esophageal Squamous Cell Carcinoma

Zhongjian Chen[1,2,3,4], *Yalan Dai*[2,3,4], *Xiancong Huang*[2,3,4], *Keke Chen*[2,3,4,5], *Yun Gao*[2,3,4], *Na Li*[2,3,4], *Ding Wang*[2,3,4], *Aiping Chen*[2,3,4], *Qingxia Yang*[1], *Yanjun Hong*[6], *Su Zeng*[1]* and *Weimin Mao*[2,3,4]*

[1] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, [2] Institute of Cancer and Basic Medicine (ICBM), Chinese Academy of Sciences, Hangzhou, China, [3] Cancer Hospital of the University of Chinese Academy of Sciences, Hangzhou, China, [4] Zhejiang Cancer Hospital, Hangzhou, China, [5] School of Medicine, Imperial College London, London, United Kingdom, [6] Department of Chemistry, Hong Kong Baptist University, Hong Kong, China

OPEN ACCESS

Edited by:
Yu Zong Chen,
National University of Singapore, Singapore

Reviewed by:
Bo Li,
Chongqing Normal University, China
Panpan Wang,
Huanghuai University, China
Chai K. Lim,
Macquarie University, Australia

***Correspondence:**
Su Zeng
zengsu@zju.edu.cn
Weimin Mao
maowm1218@163.com

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Oncology

Received: 21 March 2020
Accepted: 20 July 2020
Published: 26 August 2020

Citation:
Chen Z, Dai Y, Huang X, Chen K, Gao Y, Li N, Wang D, Chen A, Yang Q, Hong Y, Zeng S and Mao W (2020) Combined Metabolomic Analysis of Plasma and Tissue Reveals a Prognostic Risk Score System and Metabolic Dysregulation in Esophageal Squamous Cell Carcinoma. Front. Oncol. 10:1545. doi: 10.3389/fonc.2020.01545

Background: Esophageal squamous cell carcinoma (ESCC) is a gastrointestinal malignancy with a poor prognosis. Although studies have shown metabolic reprogramming to be linked to ESCC development, no prognostic metabolic biomarkers or potential therapeutic metabolic targets have been identified.

Method: The present study investigated some circulating metabolites associated with overall survival in 276 curatively resected ESCC patients using liquid chromatography/mass spectrometry metabolomics and Kaplan-Meier analysis. Tissue metabolomic analysis of 23-paired ESCC tissue samples was performed to discover metabolic dysregulation in ESCC cancerous tissue. A method consisting of support vector machine recursive feature elimination and LIMMA differential expression analysis was utilized to select promising feature genes within transcriptomic data from 179-paired ESCC tissue samples. Joint pathway analysis with genes and metabolites identified relevant metabolic pathways and targets for ESCC.

Results: Four metabolites, kynurenine, 1-myristoyl-glycero-3-phosphocholine (LPC(14:0)sn-1), 2-piperidinone, and hippuric acid, were identified as prognostic factors in the preoperative plasma from ESCC patients. A risk score consisting of kynurenine and LPC(14:0)sn-1 significantly improved the prognostic performance of the tumor-node-metastasis staging system and was able to stratify risk for ESCC. Combined tissue metabolomic analysis and support vector machine recursive feature elimination gene selection revealed dysregulated kynurenine pathway as an important metabolic feature of ESCC, including accumulation of tryptophan, formylkynurenine, and kynurenine, as well as up-regulated indoleamine 2,3-dioxygenase 1 in ESCC cancerous tissue.

Conclusions: This work identified for the first time four potential prognostic circulating metabolites. In addition, kynurenine pathway metabolism was shown to be up-regulated tryptophan-kynurenine metabolism in ESCC. Results not only provide a metabolite-based risk score system for prognosis, but also improve the understanding of the molecular basis of ESCC onset and progression, and as well as novel potential therapeutic targets for ESCC.

Keywords: esophageal squamous cell carcinoma (ESCC), metabolomics, risk score, prognosis, diagnosis, indoleamine 2, 3-dioxygenase 1 (IDO1), artificial intelligence

INTRODUCTION

Esophageal cancer (EC), a common gastrointestinal malignancy, ranks as the sixth leading cause of cancer death worldwide (1–3). EC is the fifth most common cancer and the fourth leading cause of death in China (4). Esophageal squamous cell carcinoma (ESCC) is the predominant histological subtype of EC in China (5, 6). Surgical resection with lymphadenectomy is the main treatment for ESCC (7). However, despite advances in surgical management and multidisciplinary treatment of ESCC, prognosis remains poor (8). Currently, tumor-node-metastasis (TNM) staging system is used for ESCC prognosis, even though staging components such as lymph node metastasis, invasion depth, and differentiation are not obtained during surgery but commonly determined postoperatively. Therefore, there is an urgent need for non-invasive and convenient biomarkers that may assist the clinical decision-making and provide novel insights into tumorigenesis and biology of ESCC (9).

Metabolic reprogramming is an oncogene-driven mechanism that alters the metabolism of cancer cells. It supports tumor proliferation and anabolic growth and is considered as an essential hallmark of cancer (10). Metabolomics, in which small-molecule metabolites are identified and quantified, is the closest "omics" to phenotype (11). Compared to other wide-ranging forms of analysis, metabolomics is more sensitive to alterations of biochemical homeostasis, providing comprehensive and direct information regarding cell status and treatment response. Metabolomic analysis requires a little sample material and preparation time (11). A growing body of literature demonstrates that monitoring cellular metabolites may not only provide promising biomarkers, but may also help to identify involved biological processes (11–13). For example, diagnostic and prognostic cancer biomarkers have been recently investigated in lung, colorectal, and breast cancers by different metabolomic approaches (14–16). Metabolomics, the study of altered metabolites accompanying cancer-associated metabolic reprogramming, is an emerging field that can contribute to the identification of novel cancer biomarkers and the discovery of potential drugs for prevention and therapy (17).

Previous metabolomic studies have demonstrated various metabolic alterations in patients with ESCC including changes in amino acids, glucose, lipids, organic acids, nucleotides, and fatty acids (18–21). Though many promising serum/plasma metabolites have been found to be diagnostic biomarkers for ESCC (18, 22, 23), no metabolite with prognostic value has been identified, nor has a potential metabolic therapeutic target been recognized. Gu et al. found serum D-mannose to be a novel prognostic biomarker for patients with esophageal adenocarcinoma (the main histological subtype in the USA). Those results encouraged us to investigate potential ESCC prognostic circulating metabolites by a combination metabolomics and survival analysis (24).

Despite the advantages of metabolomics, limitations for clinical application need to be considered. Due to the dynamic and sensitive nature of the metabolome, clinical metabolomic studies in particular, must be designed based on a relatively large sample size to reduce unwanted excessive variability, and results validated by multiple models or sample types (25). Thus, the aim of the present study was to discover metabolic biomarkers and potential metabolic therapeutic targets for ESCC with the following design improvements. (1) Plasma metabolic profiling was conducted on a relatively large sample size ($n = 276$). (2) Prognostic metabolites were discovered by survival analysis. (3) Integrative bioinformatics and metabolomics were used to discover metabolic features and potential therapeutic target enzymes for ESCC (**Figure 1**). Our results will assist clinicians in management of ESCC patients, as well as contribute to an understanding of the mechanisms underlying ESCC tumorigenesis, and possibly offer novel therapeutic targets.

MATERIALS AND METHODS

Chemicals and Reagents

Acetonitrile (high-performance liquid chromatography (HPLC) grade) and Methanol (HPLC) were purchased from Tedia (Ohio, USA). Formic acid (HPLC) was purchased from Roe Scientific Inc. (Delaware, USA). Distilled water was from Wahaha Group Co., Ltd. (Hangzhou, China). L-kynurenine (purity > 98%), L-tryptophan (purity > 98%), and hippuric acid (purity > 98%) were purchased from Sigma-Aldrich (Missouri, USA). L-phenylalanine (purity > 98%), 2-piperidinone (purity > 98%), and LPC(14:0)sn-1 (purity > 98%) were purchased from Aladdin Reagent Co. Ltd. (Shanghai, China). Rabbit anti-indoleamine 2, 3-dioxygenase 1 (IDO1) polyclonal antibody (13268-1-AP) was purchased from Proteintech Group, Inc. (Hubei, China).

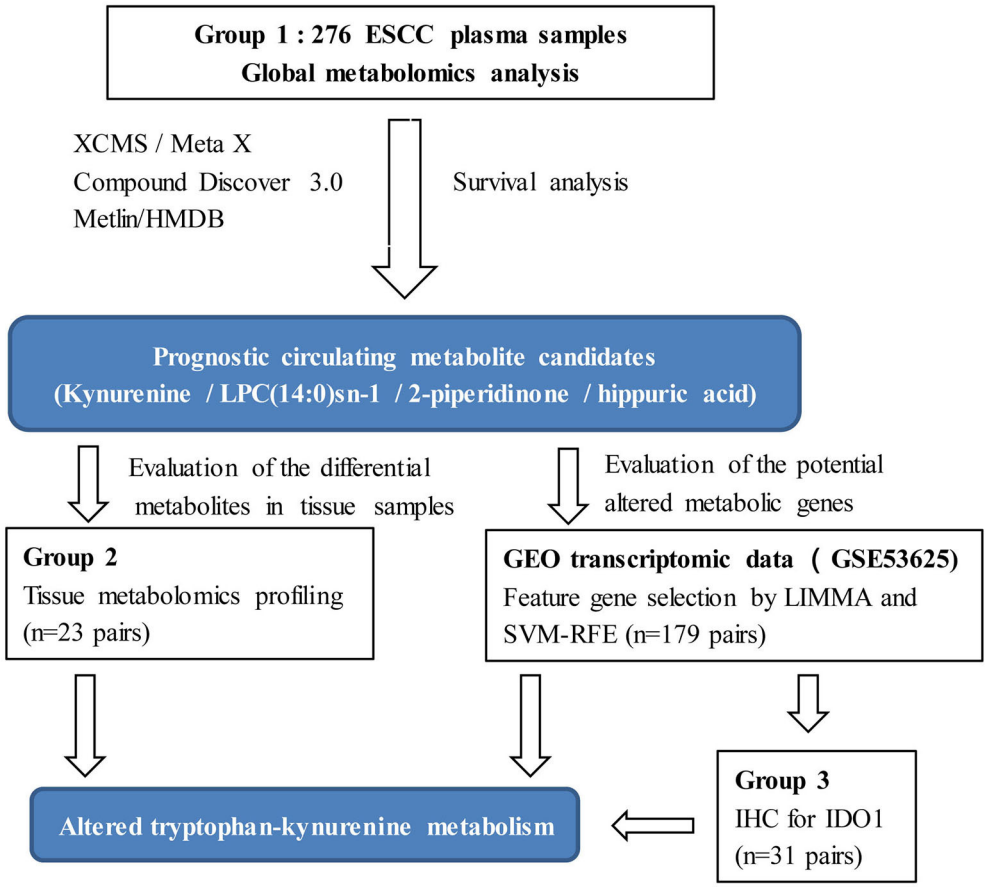


FIGURE 1 | Flowchart of this study.

Study Patients and Samples

Group 1: the preoperative fasting plasma samples, using the di-potassium salt of ethylenediaminetetraacetic acid as anticoagulant, were collected from 276 patients recruited after histopathologic confirmation of ESCC and radical resection at Zhejiang Cancer Hospital (Hangzhou, China), from May 2010 to December 2012. Clinicopathological features and preoperative biochemical parameters were obtained. Participants were followed until December 2017, evaluating overall survival (OS) from surgery to the date of death or of the last follow-up visit. Group 2: a total of 23-pairs of matched cancerous and normal tissue samples were used for tissue metabolomic study. Normal tissues were collected from the distal edge of the resected tissues, at more than 2 cm from the solid tumor border. Group 3: a total of 31-pairs of matched cancerous and normal tissue samples were assessed by immunohistochemistry (IHC). All samples collected in this study were stored at −80°C until analysis. Demographic and clinicopathologic characteristics of the patients are reported in **Table 1**.

The study protocol was performed in accordance with the declaration of Helsinki, approved by the Research Ethics Committee of Zhejiang Cancer Hospital, China, with written informed consent obtained from all individuals.

Plasma-Based Metabolomic Analysis

Sample Preparation

Plasma samples were from Group1 (**Table 1**). Plasma samples (50 μL) were thawed on ice and immediately mixed with 200 μL of ice-cold acetonitrile. After mixing by vortex for 1 min, the mixture was centrifuged at 16,200 g for 15 min at 4°C. The supernatant (150 μL) was transferred into a fresh tube and lyophilized. The residues were resuspended by adding 80 μL of 25% acetonitrile in water and mixed by vortex for 1 min. After centrifugation at 16,200 g for 15 min at 4°C, 60 μL of the supernatant was transferred into the sample bottle. A supernatant aliquot of 10 μL was used for liquid chromatography-mass spectrometry (LC-MS) analysis.

Quality control (QC) samples were prepared by pooling the re-dissolved sample with an equal amount (15 μL) and periodically analyzed throughout the complete analytical run to monitor signal drift.

LC-MS Analysis

LC-MS analysis was conducted as previously described (26). The Ultimate 3000 UHPLC system (Dionex, Idstein, Germany), linked to a Q Exactive orbitrap mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany), was used.

TABLE 1 | Demographic and clinicopathologic characteristics of study patients.

Parameters[a]	Group 1[b]	Group 2[b]	Group 3[b]
	n = 276	*n* = 23 (paired)	*n* = 31 (paired)
Sex			
Male	232	19	3
Female	44	4	28
Age			
Mean age, yr	60.8 ± 7.1	59.7 ± 7.3	62.0 ± 6.7
BMI			
Median(range)	22 (16-30)	22 (17-27)	22 (16-28)
Smoking habit			
No	79	8	10
Yes	197	15	21
Alcohol consumption			
No	88	10	10
Yes	188	13	21
Tumor thrombus			
No	191	16	20
Yes	85	7	11
Neural invasion			
No	153	16	21
Yes	123	7	10
Tumor grade			
Well	15	1	3
Moderately	186	14	16
Poorly	75	8	12
N stage			
0	101	11	12
1	99	8	13
2	49	3	5
3	27	1	1
T stage			
1	12	3	4
2	49	3	3
3	213	17	24
4	2	-	-
TNM stage			
1	23	4	4
2	85	7	9
3	144	11	17
4	24	1	1

[a]N stage, T stage, and TNM stage were determined according to the American Joint Committee on Cancer, 8th edition. [b]Group 1: Preoperative plasma samples from 276 patients with ESCC were used for plasma metabolomics. Group 2: 23-pairs of matched cancerous and normal tissue samples were used for tissue metabolomic study. Group 3: 31-pairs of matched cancerous and normal tissue samples were assessed by IHC analysis.

Separation was performed on an ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm × 1.8 μm, Waters, MA, USA) at 35°C, with a mobile phase consisting of 0.1% formic acid and acetonitrile. The gradient is reported in **Table S1**. Full mass scan mode was used for all the samples and data-dependent MS/MS acquisition mode was utilized for the identification of QC samples. Detailed parameters are listed in **Table S2**.

Metabolomic Analysis

Raw data were converted to mzXML format using MSconvert program (http://proteowizard.sourceforge.net/download.html). The R package *XCMS* (version 3.3.2) was utilized for data preprocessing, including retention time alignment, peak detection, and peak matching. R package *MetaX* (version 1.4.16) was used to remove peaks with more than 20% of zero values in all samples or the peaks with coefficient of variation values >30% in QC. Peaks were corrected with the QC-robust LOESS signal correction algorithm. All the detected ions in each sample were normalized to the sum of the peak area defined as 100,000 (27). Thermo Scientific Compound Discoverer 3.0 software (Thermo Fisher Scientific, USA) combined with the METLIN (http://metlin.scripps.edu) and the HMDB (http://www.hmdb.ca/) databases were used for metabolite annotation by comparison of MS fragmentation information. Standard substances were used to verify prognostic metabolites.

Survival Analysis With Regard to Circulating Metabolites

Once relative concentrations of circulating metabolites were obtained by metabolomic analysis, Kaplan-Meier curves were performed to identify associations between metabolite levels and OS, with median split and log-rank test. Cox proportional hazards regression test was also performed for each metabolite to calculate the hazard ratio (HR) value. Factors with *P*-values < 0.05 were considered to have prognostic significance. Multivariate Cox proportional hazards regression was analyzed to estimate independent and significantly prognostic circulating metabolites. With the independently prognostic metabolites, a risk score was derived by summation of each metabolite level multiplying their corresponding coefficient according to Li et al. (28).

Tissue-Based Metabolomic Analysis

Sample Preparation and LC-MS Analysis

Tissue samples were from Group 2 (**Table 1**). Approximately 20 mg of tissue was transferred into a 1.5 mL tube with immediate addition of 400 μL of ice-cold methanol and two steel balls (diameter: 2 mm). Homogenization was performed with a Tissuelyser (2 min, 30 Hz). After centrifugation at 16,200 g for 15 min at 4°C, 200 μL of the supernatant was transferred into a fresh tube, to which was added with 200 μL of water followed by lyophilization. Reconstitution, analysis protocols as well as QC sample preparation were conducted by the same methods used for plasma-based metabolomics. The LC-MS analysis and metabolomic analysis protocols were the same as that for plasma-based metabolomics.

Metabolomic Analysis

Relative concentrations of ion features were obtained from metabolomic data with the same protocol as that for plasma-based metabolomics. Unsupervised principal component analysis (PCA) was conducted to assess the trends for all

samples. Supervised partial least squares discriminant analysis (PLS-DA) was performed to identify the most discriminating ion features between ESCC cancerous tissues and non-cancerous counterparts based on variable importance in projection (VIP) values. Finally, ions with VIP > 1, Benjamini–Hochberg adjusted *p*-value (FDR) < 0.05, and |log2(Fold change)| > 0.585 were defined as differential ion features. Metabolite annotation was performed using the above method. Receiver operating characteristic (ROC) curve analysis was used to evaluate the diagnostic significance of metabolites, in order to distinguish ESCC cancerous and non-cancerous tissues.

Bioinformatic Analysis

The ESCC microarray dataset (GSE53625) was generated using the Agilent-038314 CBC *Homo sapiens* lncRNA+mRNA microarray V2.0 (http://www.genomics.agilent.com/) deposited in the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and processed as described previously (9). Briefly, probe re-annotation was performed based on the previous study (29). For genes with multiple probes, mean expression was calculated and used. LIMMA (R package version 3.38.3) was used to analyze differentially expressed genes (DEGs) and genes with $|\log_2(\text{Fold change})| > 1$ and FDR < 0.05 were considered to be DEGs. The differential DEGs were further ranked by support vector machine recursive feature elimination (SVM-RFE) algorithm proposed according the Huang et al. (30). Briefly, the SVM-RFE computes the ranking weights for all DEGs and sorts the DEGs according to weight vectors as the classification basis. SVM-RFE conducted an iteration process of the backward removal of DEGs as follows: (1) use the current dataset to train the classifier; (2) calculate the ranking weights for all DEGs; (3) eliminate the DEG with the smallest weight. The above iteration process repeats until there is only one DEG remained in the dataset, and the implementation result provides a list of DEGs in the descending order of weight. Top 500 genes were selected as features for further analysis. Additionally, the metabolic feature genes were obtained through searching on KEGG (31). Then, pathway analysis was conducted by searching genes and metabolites together by KEGG Mapper Search (https://www.genome.jp/kegg/mapper.html). The metabolism pathway/module for both hits from gene and metabolite was considered to be of potential importance.

IHC Analysis

Tissue sections (4 μm thick) were dewaxed and rehydrated through graded alcohols. IHC staining with IDO1 antibodies was performed according to the manufacturer's instructions. The results were analyzed using a semiquantitative method (32), with the immunohistochemical score calculated by multiplying the percentage of positive cells by staining intensity.

Statistical Analysis

Statistical analysis was performed using SPSS 16.0 software (USA) and R software (http://www.r-project.org/). Normality of the variables was tested by Shapiro-Wilk normality test. Unpaired Wilcoxon rank-sum test, and Kruskal-Wallis test were used for comparison of two or more than two groups of data. The correlation between circulating metabolites and other variables was calculated by Kendall rank or Spearman's rank correlation. ROC analysis and ROC comparisons were performed by R package *pROC* (version 1.15.3). A two-tailed *p* value < 0.05 was considered statistically significant.

RESULTS

Identification of Circulating Prognostic Metabolites for ESCC

For plasma-based metabolomic profiling study (Group 1), a total of 4,121 metabolic features in electrospray ionization positive mode and 3,046 in electrospray ionization negative mode were extracted from the metabolomic data. A total of 129 ion features were annotated with metabolites (**Tables S3**, **S4**). Survival analysis aided identification of four metabolites with *p* values < 0.05: kynurenine, LPC(14:0)sn-1, 2-piperidinone, and hippuric acid. Higher plasma levels of kynurenine, 2-piperidinone, and hippuric acid correlated with shorter survival, while higher levels of LPC(14:0)sn-1 correlated with longer survival (**Figures 2A–D**, **Table 2**). Moreover, Multivariate Cox regression analysis of the four metabolites indicated that kynurenine and LPC(14:0)sn-1 to be independent factors suitable for metabolite-based risk score calculation (**Table 2**).

These four prognostic metabolites were compared with chromatograms and spectra of reference substances. A representative identification of kynurenine is shown in **Figure S1**, while the identification of the other three metabolites is illustrated in **Figure S2**.

In order to clarify the potential influence of demographic factors on these prognostic circulating metabolites, multivariate Cox regression was performed for each of the four metabolites with age, sex, smoking habit, and alcohol consumption. Result demonstrated that kynurenine (HR: 1.37, $p = 0.040$), LPC(14:0)sn-1 (HR: 0.618, $p = 0.00229$), hippuric acid (HR: 1.423, $p = 0.021$) were independent prognostic factors, while 2-piperidinone (HR: 1.35, $p = 0.054$) was not an independent factor.

Potential Relationships Among Prognostic Metabolites, Clinicopathologic Features, and Biochemical Parameters

Analysis of the four circulating metabolites indicated that kynurenine was significantly positively correlated with the other three metabolites, while 2-piperidinone positively correlated with hippuric acid (**Table S5**). The four metabolites were assessed for correlations with clinicopathologic features including sex, age, body mass index (BMI), smoking habit, alcohol consumption, tumor grade, tumor thrombus, neural invasion, T stage, N stage, and TNM stage. The following biochemical parameters were also included in the correlation analysis: glycyl proline dipeptidyl aminopeptidase (GPDA), alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), prealbumin (PA), albumin (ALB), triglyceride (TG), total cholesterol (TC), low density lipoprotein cholesterol, high

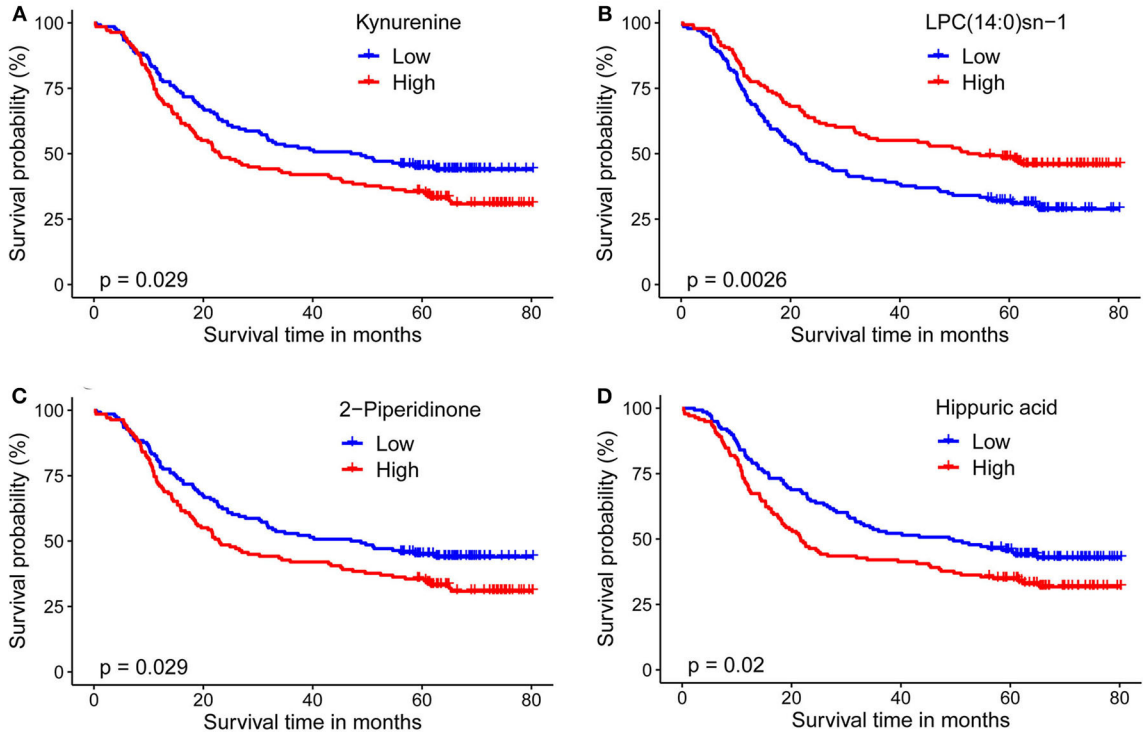


FIGURE 2 | Kaplan-Meier survival curves for ESCC patients stratified by the four circulating metabolites by a median-split. **(A)** Kynurenine; **(B)** LPC(14:0)sn-1; **(C)** 2-Piperidinone; **(D)** Hippuric acid. Log-rank test was used, $p < 0.05$ was considered significant.

TABLE 2 | Prognostic circulating metabolites in ESCC plasma.

Metabolites	Features in LC/MS			KM[d]	Cox regression[e]	
	m/z[a]	RT (min)[b]	Ion[c]	*p* value	*p* value	HR (95%CI)
2-Piperidinone	100.07629	3.9	Positive	0.035	0.2	1.22 (0.90-1.67)
Kynurenine	209.09222	3.9	Positive	0.029	0.035	1.40 (1.03-1.91)
LPC(14:0)sn-1	468.30879	8.9	Positive	0.0026	0.0019	0.61 (0.45-0.84)
Hippuric acid	178.05156	5.2	Negative	0.02	0.11	1.29 (0.95-1.76)

[a]m/z, mass/charge number; [b]RT, retention time; [c]mass spectrometer scan types; [d]KM, Kaplan-Meier; [e]Cox regression analysis of 2-piperidinone, kynurenine, LPC(14:0)sn-1 and hippuric acid, and HR, hazard ratio.

density lipoprotein cholesterol. The results of this analysis showed that: (1) Kynurenine levels were positively correlated with N stage and GGT levels, while a negative correlation was found with tumor grade and ALB levels. (2) LPC(14:0)sn-1 levels were positively correlated with GPDA, ALT, PA, and TG levels, and negatively correlated with age. (3) 2-piperidinone levels were positively correlated with GGT levels, BMI, and alcohol consumption, while it was negatively correlated with TC levels. (4) Hippuric acid levels were positively correlated with GPDA levels and negatively correlated with GGT, ALT, and PA levels (**Table S6**).

Metabolite-Based Risk Score Improves Prognostic Performance

Cox proportional hazards regression analysis showed kynurenine and LPC(14:0)sn-1 regression coefficients of 0.41 and −0.52, respectively. A risk score was attributed to each patient by adding the plasma level of each metabolite multiplied by the corresponding regression coefficient: risk score = (0.409 × level of kynurenine −0.522 × level of LPC (14:0)sn-1) (28). The risk score of all cases was calculated according to this formula and the patients were stratified into low-risk and high-risk groups, by applying the median-split method. Risk score efficiently stratified ESCC risk (**Figure S3**) independent of TNM or N stage (**Table S7**).

Area under the curve (AUC) of ROC curves for 5-year survival status prediction was calculated and compared using the method established by DeLong et al. (33). When combined with the risk score staging classification, the prediction accuracy of the conventional TNM stage and N stage was significantly improved from 0.650 (95% confidence interval (CI): 0.583-0.718) to 0.692 (95% CI: 0.628-0.756; $p = 0.015$), and from 0.665 (95% CI: 0.599-0.731) to 0.694 (95%CI: 0.630-0.750; $p = 0.042$), respectively (**Figures 3A,B**). Moreover, log-rank analysis of Kaplan-Meier curves related to the metabolite-based risk score groups demonstrated that the calculated risk score was able to significantly improve the prediction of clinical outcome in patients with ESCC, classified according to the stages TNM II ($p = 0.028$), TNM III ($p = 0.008$), N1 ($p = 0.024$), and

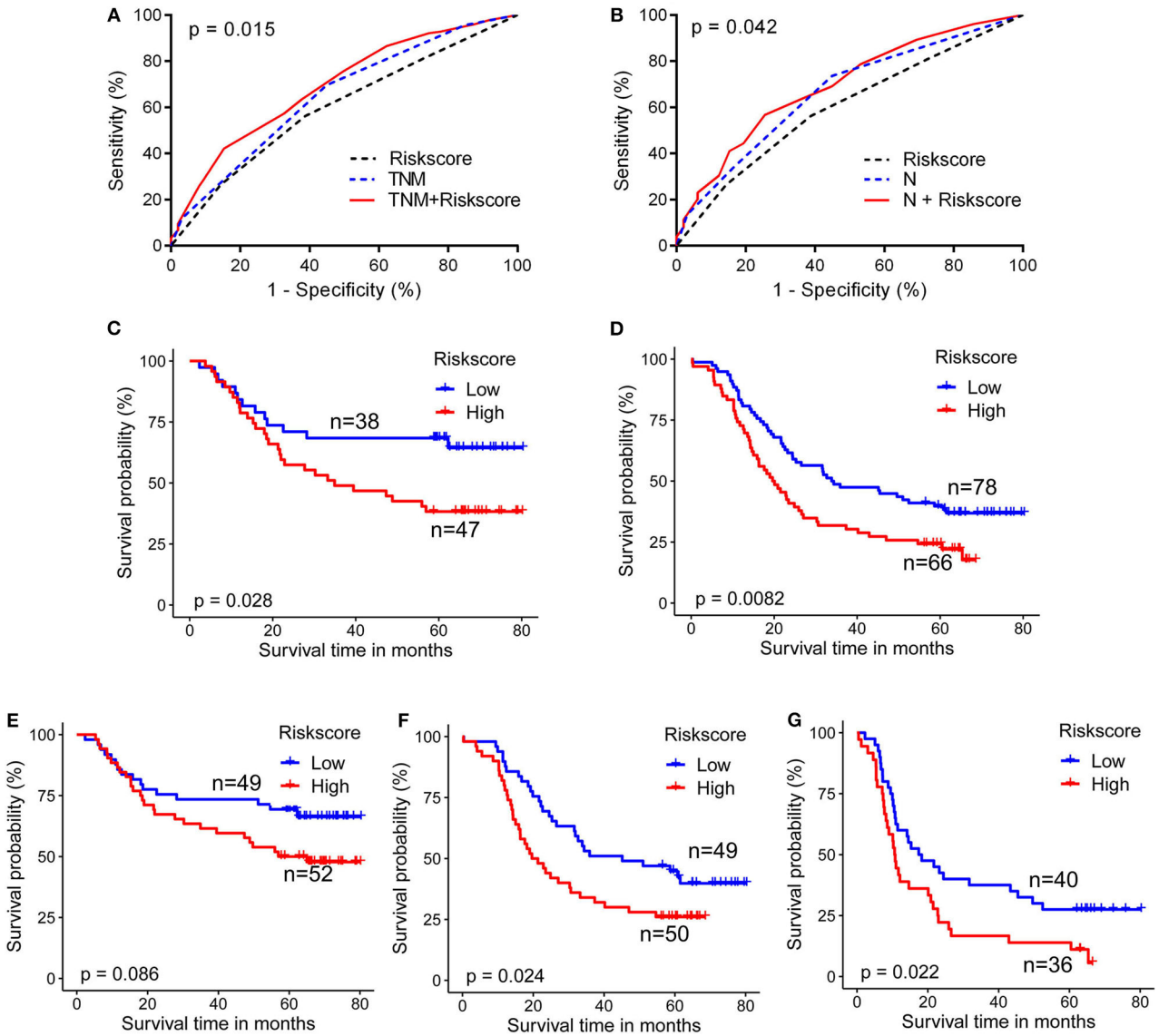


FIGURE 3 | Metabolite-based risk score improves prognostic performance for ESCC patients. **(A)** Combination of risk score with TNM stage **(B)**, or with N stage (G) significantly improved the predictive accuracy of 5-year survival rate. ESCC cases classified according to TNM II **(C)**, TNM III **(D)**, N1 **(F)**, and N2 **(G)** stages were significantly stratified by risk score. Cases from N0 **(E)** showed a trend for stratification by risk score. DeLong test for AUC in ROC curves and log-rank test for Kaplan-Meier survival curves were assessed, $p < 0.05$ was considered significant.

N2 ($p = 0.022$; **Figures 3C,D,F,G**), while for N0 classification a statistically non-significant ($p = 0.086$) trend to stratification based on the risk score was observed (**Figure 3E**).

Tissue-Based Metabolomics Reveals Altered Kynurenine Pathway in ESCC

Both PCA and PLS-DA analysis with the extracted 4,856 ion features showed a significant metabolic shift between ESCC cancerous and normal tissues (**Figures 4A,B**), with a total of 1,697 differential ion features were selected (**Figure 4C**). There were 26 differential metabolites were annotated that could significantly separate the cancerous and normal tissue samples (**Figure 4D**). However, analysis identified only kynurenine and LPC(14:0)sn-1 of the four prognostic circulating metabolites. Significantly higher levels of kynurenine and LPC(14:0)sn-1 were observed in cancerous tissues compared with normal counterparts (**Figures 4E,F**). Interestingly, when other molecules involved in tryptophan-kynurenine metabolism were investigated, the levels of tryptophan and formylkynurenine were also found to be higher in cancerous tissues (**Figures 4G,H**). Although hippuric acid was not detected by tissue metabolomic analysis, phenylalanine, a potential parent metabolite of hippuric acid, was found at higher levels in cancerous tissues (**Figure 4I**). ROC curve analysis showed that tissue formylkynurenine

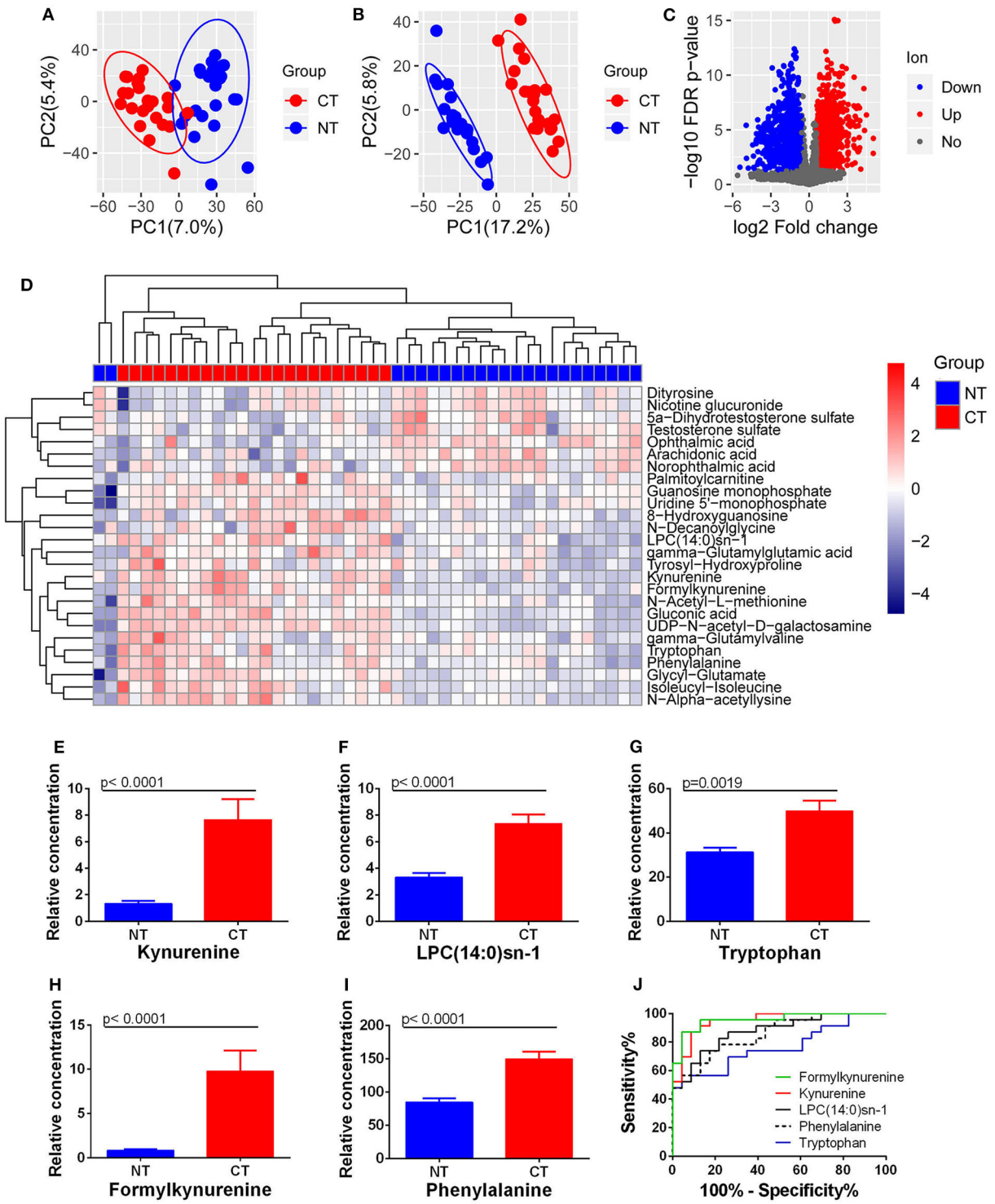


FIGURE 4 | Changes in metabolites between ESCC cancerous (CT) and normal tissues (NT). **(A)** PCA and **(B)** PLS-DA analysis with all ion features. **(C)** Differential ion features were defined as $|\log2(\text{Fold change})| > 0.585$ and FDR < 0.05. **(D)** Heatmap analysis with 26 differential metabolites. Accumulation of kynurenine **(E)**, LPC(14:0)sn-1 **(F)**, tryptophan **(G)**, formylkynurenine **(H)**, and phenylalanine **(I)** was observed in ESCC cancerous tissues compared to normal equivalents. ROC curves of these metabolites showed the potential diagnostic value for ESCC **(J)**.

levels had the best diagnostic performance (max AUC_{ROC}: 0.957), followed by kynurenine (AUC_{ROC}:0.947), LPC(14:0)sn-1 (AUC_{ROC}: 0.875), phenylalanine (AUC_{ROC}: 0.856), and tryptophan (AUC_{ROC}: 0.762) (**Figure 4J**). Taken together, tissue-based metabolomics inferred altered kynurenine pathway to be the most significant feature of ESCC.

IDO1 Up-Regulation in ESCC

After LIMMA differential expression analysis, a total of 2,856 DEGs were screened (**Figure 5A**). After SVM-RFE ranking, the top 500 genes were then selected as feature genes, including 46 metabolic genes (**Figure 5B**). The 46 metabolic feature genes were able to separate the cancerous and normal tissues (**Figure 5C**). Pathway analysis of both the 46 DEGs and the above 26 metabolites identified the top 10 metabolic pathways, in which fatty acid degradation ranked first in total hits/gene hits, and tryptophan metabolism had the most metabolite hits (**Figure 5D**). By module analysis, two potential modules, kynurenine pathway (metabolite: tryptophan, formylkynurenine, kynurenine, and gene: *IDO1*) and pyrimidine biosynthesis (metabolite: uridine monophosphate (UMP) and uridine diphosphate (UDP); gene: *CMPK2*), were identified by criteria (1) with both gene and metabolite hits, and (2) direct interaction between gene and metabolite (**Figure 5E**). Based on these results: (1) Kynurenine in plasma was associated with OS of patients with ESCC. (2) Significant accumulation of tryptophan, kynurenine, and formlykynurenine was found in ESCC cancerous tissue. (3) Upregulation of *IDO1* mRNA, which was the 209th feature in the 2856 DEGs by SVM-RFE ranking, and the 18th one in the 46 metabolic feature genes, was found in cancerous compared to normal tissues with a high level of statistical significance (**Figure 5F**). Consistent with these observations, IHC analysis demonstrated significantly higher levels of IDO1 in cancerous tissues compared to paired normal counterparts (**Figures 5G,H**). Collectively, kynurenine pathway is an important metabolic feature of ESCC, with IDO1 as a potential therapeutic target.

DISCUSSION

The present study revealed that circulating kynurenine, LPC(14:0)sn-1, 2-piperidinone, and hippuric acid were prognostic factors for ESCC. A kynurenine and LPC(14:0)sn-1 based risk score significantly improved the prediction accuracy of the current TNM staging system in ESCC. Up-regulated tryptophan pathway metabolism, including the accumulation of tryptophan, formylkynurenine, and kynurenine, as well as increased expression of IDO1, were identified as the most significant metabolic features of ESCC.

Circulating-metabolite-based prognostic models have previously been shown to have its promising clinical applications in several cancers, such as glioblastoma (34), non-small cell lung cancer (35), and esophageal adenocarcinoma (24). However, the previous metabolomic studies of ESCC solely focused on the diagnostic value of the metabolites and rarely assessed their prognostic significance. This study is the first to investigate the prognostic value of plasma metabolites in ESCC and found several metabolic biomarkers as well as established a metabolite-based risk score for ESCC. For this study, a combination of circulating metabolomic profiling and survival analysis was used to develop a prognostic approach for ESCC, which identified the kynurenine and LPC(14:0)sn-1 based risk score to have prognostic significance for ESCC. By use of the risk score, ESCC patients were stratified by risk within the same TNM stage (TNM II and III) or the same N stage (N1 and N2). As such, the risk score may assist clinical decision-making, leading to a better prognosis for ESCC patients. It is important to note that due to the limited sample size of TNM subgroups, TNM I (n =23) and TNM IV (n = 24), it was unavailable to evaluate its risk stratification for these subgroups. Further, for N0 classification, a trend based on risk score was observed that was not statistically significant (p = 0.086) (n = 101). Future analysis with larger cohorts is essential to determine the clinical significance of the risk score for these subgroups.

Kynurenine was the most interesting circulating biomarker identified by this study. It is one of the main metabolites of tryptophan metabolism, which is related to immune homeostasis, and is correlated with cancer initiation and development (12). Previously, Cheng et al. found the ratio of kynurenine/tryptophan in plasm to be significantly increased in ESCC, and correlated with lymph node metastasis. However, the relationship between kynurenine levels and their survival outcome were not considered (18). To the best of our knowledge, our study is the first to identify circulating kynurenine as a prognostic factor for ESCC, in which higher levels of kynurenine were correlated with poorer OS, and higher N stage and tumor grade levels. Our study and the previous study by Cheng at al. collectively indicate that circulating kynurenine is a promising unfavorable prognostic biomarker for ESCC. The negative correlation of kynurenine levels with survival outcomes is consistent with the immune suppressor role of kynurenine in cancers, in which many cancers enhance kynurenine levels by up-regulating IDOs activity, resulting in escape from immune clearance (36).

Our tissue-based metabolomics revealed up-regulated kynurenine pathway is a significant feature of ESCC, including the accumulation of kynurenine, tryptophan, and formylkynurenine. Significant accumulation of tryptophan and kynurenine in ESCC was reported previously by Tokunaga et al. (23) and Zhang et al. (22), respectively, with this study the first to identify up-regulated formylkynurenine in ESCC. The fold change of formylkynurenine in ESCC was 11.9 and displayed the best diagnostic performance in this study. Formylkynurenine is the direct metabolite of tryptophan mediated by IDO, with increased accumulation of formylkynurenine resulting in the production of the immune suppressive metabolite, kynurenine (fold change of 5.7), in ESCC cancer tissue. Therefore, up-regulated kynurenine pathway not only explained (at least partially) the increased serum kynurenine levels in ESCC patients reported by Cheng et al. (18), but also implied an important role for kynurenine in ESCC progression. In addition to the metabolite level, our bioinformatics analysis and IHC staining analysis demonstrated the key rate-limiting enzyme,

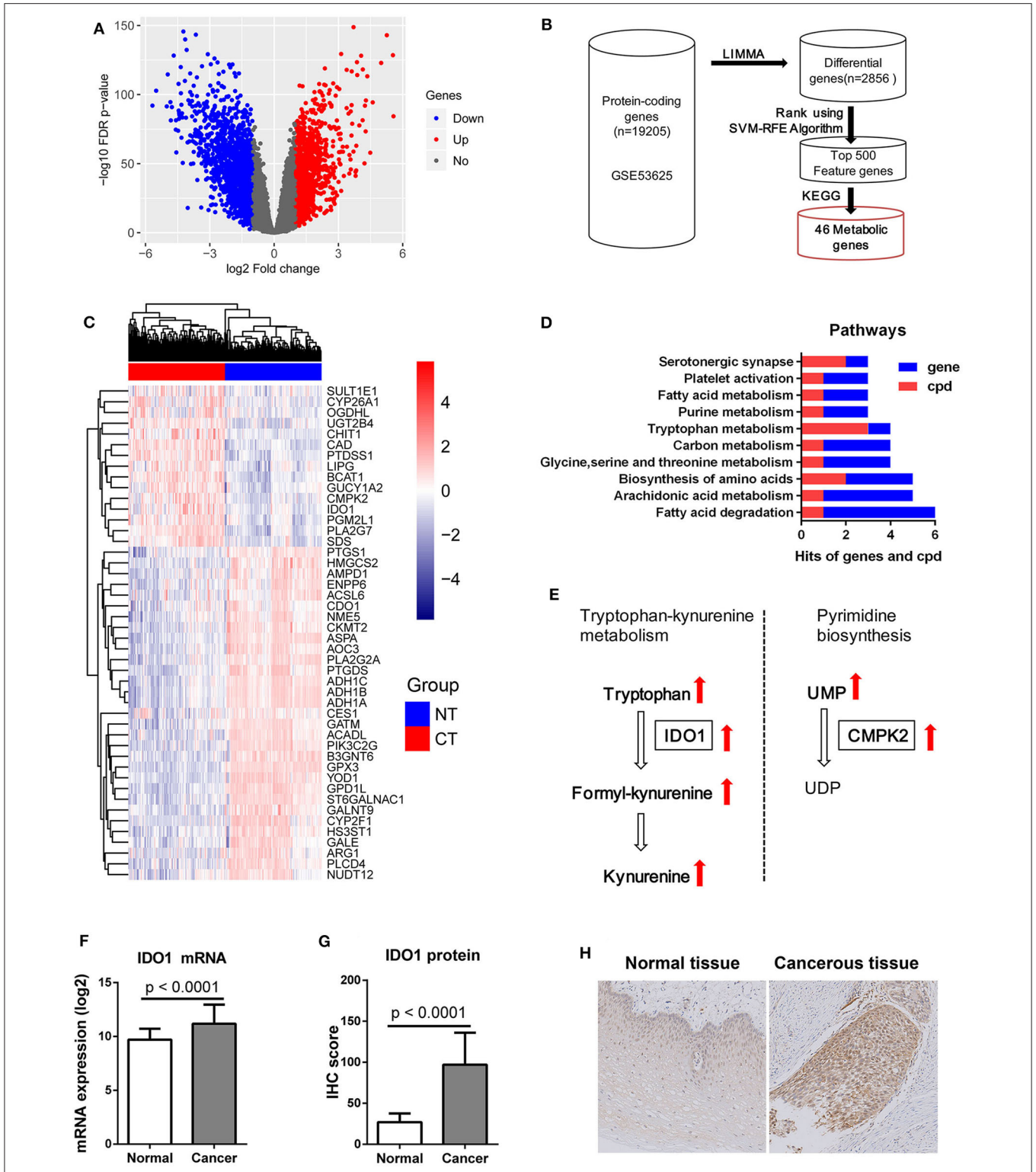


FIGURE 5 | IDO1 expression in ESCC. **(A)** Volcano plot of 2856 DEGs with $|\log_2(\text{Fold change})| > 1$ and FDR < 0.05 for transcriptomic data from 179 paired ESCC tissue samples from GEO microarray data (GSE53625), red dot: up-regulated gene, blue dot: down-regulated gene. **(B)** Schematic diagram for feature metabolic gene selection process. **(C)** Heatmap with 46 feature metabolic genes, NT: normal tissue, CT: cancerous tissue. **(D)** Top 10 metabolism pathways with both hits for gene and metabolite (cpd). **(E)** Significantly altered metabolic modules with both hits for gene and metabolite, UMP, uridine monophosphate; UDP, uridine diphosphate. **(F)** mRNA expression of IDO1 was significantly increased in ESCC cancerous tissues compared to normal counterparts. **(G)** IHC analysis of IDO1 protein expression showing significant up-regulation in cancerous tissues compared to normal counterparts. **(H)** Representative IHC staining of IDO1 in cancerous and matched equivalents, Magnification: 400×.

IDO1, to be significantly up-regulated in ESCC, which could be a potential therapeutic target for ESCC. Recently, Liu et al. found that kynurenine could up-regulate PD-1 expression on tumor infiltrating T cells through the IDO-kynurenine-AhR pathway (37). The significant accumulation of kynurenine in ESCC suggests that IDO1 inhibitors in combination with other immunotherapies (such as anti-PD-1/anti-PD-L1) may be useful as future therapeutics for ESCC.

LPC(14:0)sn-1, another prognostic circulating metabolite, is a form of lysophosphatidylcholine, in which a phosphorylcholine moiety occupies a glycerol substitution site. Since there are co-existence of LPC(14:0)sn-1 with myristic acid at the C-1 position (sn-1) and LPC(14:0)sn-2 with myristic acid at C-2 position (sn-2) in plasma and these two LPC(14:0)s have the same molecular weight and similar retention time by chromatography, it is crucial to identify LPC(14:0)sn-1 with standard compound. Xu et al. previously reported the down-regulation of plasma LPC(14:0) in ESCC, but they did not clarify the exact position of myristic acid in LPC(14:0) since the study only employed database searches. We believe our study is the first to identify LPC(14:0)sn-1 as a potential prognostic biomarker for ESCC. In the previous study by Xu et al. LPC(14:0) was shown to be a diagnostic plasma metabolite for ESCC. Our study for the first time identified the prognostic potential of circulating LPC(14:0)sn-1 for ESCC in that patients with higher levels had longer OS. Our tissue-based metabolomic analysis detected a significant accumulation of LPC(14:0)sn-1 in cancerous tissue, indicating up-regulated lipid metabolism in ESCC (at least within the lysophosphatidylcholine metabolism pathway). Kamphorst et al. demonstrated cancer cells to directly uptake and use lipids from circulation by macropinocytosis (38, 39). We propose that alterations in circulating LPC(14:0)sn-1 might be associated with enhanced lipid consumption by cancer cells. However, the reason why ESCC patients with lower LPC(14:0)sn-1 have poorer OS is unknown, and further study is needed to explore the potential biological functions of LPC(14:0)sn-1 in ESCC.

2-piperidinone and hippuric acid are the other two prognostic circulating metabolites only detected in plasma in this study. 2-piperidinone was previously found to be decreased in plasma of patients with ovarian cancer (40), but first identified in ESCC. Hippuric acid is formed by the conjugation of benzoic acid with glycine and it is an end-product of phenylalanine metabolism (41). Since increased uptake of phenylalanine was observed in ESCC cancerous tissues, a potential relationship between phenylalanine metabolism and changes in circulating hippuric acid can be postulated. However, neither 2-piperidinone nor hippuric acid were detected by tissue analysis, which suggests that the two metabolites may not have originated from cancer cells. The detailed origin and biological activity of these two metabolites requires further investigation.

Nevertheless, limitations of this study must be considered. First, the biological activity of the four circulating prognostic metabolites, in particular kynurenine and LPC(14:0)sn-1, has not been clarified. Additionally, there are limitations in metabolite annotation and identification, which is a common problem for all metabolomic studies. Future efforts are required to resolve this issue for the entire field.

In conclusion, after identification of potential candidates for circulating prognostic metabolites, and validation by risk score based on plasma levels and correlation coefficients, kynurenine and LPC(14:0)sn-1 were identified as two circulating metabolite biomarkers with prognostic potential. The identified risk score significantly improved prediction accuracy of the TNM staging system and allowed better stratification of ESCC clinical risk. This study demonstrated kynurenine pathway dysregulation in ESCC, which was accompanied by upregulation of IDO1. These observations provide novel insights into the molecular mechanisms of ESCC tumorigenesis and the possible identification of therapeutic targets for ECSS.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GEO/GSE53625/ https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53625.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Zhejiang Cancer Hospital, China. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZC, SZ, and WM conceived, designed the study, interpreted the data, wrote the first draft of the manuscript, and contributed to the final version of the manuscript. YD, XH, and YG performed the experiment. KC helped to interpret the data and write the manuscript. NL, DW, and AC helped to collect the clinical data and perform the statistical analysis. QY conducted the bioinformatics analysis. YH instructed the metabolomics analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the National Natural Science Foundation of China (No. 81672315, 81302840, 81802472), Zhejiang Provincial Natural Science Foundation of China (LQ18H160020), and the 1022 Talent Training Program of Zhejiang Cancer Hospital. Science, Technology and Innovation Commission of Shenzhen (JCYJ20160531193901593). Hong Kong Baptist University, Faculty Research Grants (2/17-18/088). Leading Talent of

Ten Thousand Plan—National High-Level Talents Special Support Plan.

ACKNOWLEDGMENTS

We thank Biobank in Zhejiang Cancer Hospital for providing all the samples in the study. We acknowledge technical support of Dr. Junzhou Wu (Core Facility, Cancer Hospital of the University of Chinese Academy of Sciences) and Dr. Jinjin Xu (Department of Pathology, Cancer Hospital of the University of Chinese Academy of Sciences).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01545/full#supplementary-material

REFERENCES

1. Global Burden of Disease Cancer Collaboration. Global, Regional, and National Cancer Incidence. Mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study. *JAMA Oncol.* (2019) 5:1749–68. doi: 10.1001/jamaoncol.2019.2996
2. Gupta B, Kumar N. Worldwide incidence, mortality and time trends for cancer of the oesophagus. *Eur J Cancer Prev.* (2017) 26:107–18.doi: 10.1097/CEJ.0000000000000249
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *Cancer J Clin.* (2016) 66:7–30.doi: 10.3322/caac.21332
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in china, 2015. *Cancer J Clin.* (2016) 66:115–32.doi: 10.3322/caac.21338
5. Chen W, Zheng R, Zeng H, Zhang S. The updated incidences and mortalities of major cancers in china, 2011. *Chin J Cancer.* (2015) 34:502–7.doi: 10.1186/s40880-015-0042-6
6. Malhotra GK, Yanala U, Ravipati A, Follet M, Vijayakumar M, Are C. Global trends in esophageal cancer. *J Surg Oncol.* (2017) 115:564–79.doi: 10.1002/jso.24592
7. Fujiwara Y, Yoshikawa R, Kamikonya N, Nakayama T, Kitani K, Tsujie M, et al. Neoadjuvant chemoradiotherapy followed by esophagectomy vs. surgery alone in the treatment of resectable esophageal squamous cell carcinoma. *Mol Clin Oncol.* (2013) 1:773–79.doi: 10.3892/mco.2013.128
8. Sohda M, Kuwano H. Current status and future prospects for esophageal cancer treatment. *Ann Thor Cardiovasc Surg.* (2017) 23:1–11.doi: 10.5761/atcs.ra.16-00162
9. Li Y, Lu Z, Che Y, Wang J, Sun S, Huang J, et al. Immune signature profiling identified predictive and prognostic factors for esophageal squamous cell carcinoma. *Oncoimmunology.* (2017) 6:e1356147.doi: 10.1080/2162402X.2017.1356147
10. Ward PS, Thompson CB. Metabolic reprogramming: a Cancer hallmark even warburg did not anticipate. *Cancer Cell.* (2012) 21:297–308.doi: 10.1016/j.ccr.2012.02.014
11. Guijas C, Montenegro-Burke JR, Warth B, Spilker ME, Siuzdak G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat Biotechnol.* (2018) 36:316–20.doi: 10.1038/nbt.4101
12. Bishnupuri KS, Alvarado DM, Khouri AN, Shabsovich M, Chen B, Dieckgraefe BK, et al. Ido1 and kynurenine pathway metabolites activate pi3k-Akt signaling in the neoplastic colon epithelium to promote cancer cell proliferation and inhibit apoptosis. *Cancer Res.* (2019) 79:1138–50.doi: 10.1158/0008-5472.CAN-18-0668
13. Rodriguez-Blanco G, Burgers PC, Dekker LJ, Vredenbregt-van den Berg MS, Ijzermans JN, Schenk-Braat EA, et al. Serum kynurenine/Tryptophan ratio is not a potential marker for detecting prostate cancer. *Clin Biochem.* (2014) 47:1347–8.doi: 10.1016/j.clinbiochem.2014.05.001
14. Mu Y, Zhou Y, Wang Y, Li W, Zhou L, Lu X, et al. Serum metabolomics study of nonsmoking female patients with non-small cell lung cancer using gas chromatography-Mass spectrometry. *J Prot Res.* (2019) 18:2175–84.doi: 10.1021/acs.jproteome.9b00069
15. Kim ER, Kwon HN, Nam H, Kim JJ, Park S, Kim YH. Urine-Nmr metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. *Sci Rep.* (2019) 9:4786.doi: 10.1038/s41598-019-41216-y
16. Jasbi P, Wang D, Cheng SL, Fei Q, Cui JY, Liu L, et al. Breast cancer detection using targeted plasma metabolomics. *J Chrom.* (2019) 1105:26–37.doi: 10.1016/j.jchromb.2018.11.029
17. Peng B, Li H, Peng XX. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell.* (2015) 6:628–37.doi: 10.1007/s13238-015-0185-x
18. Cheng J, Jin H, Hou X, Lv J, Gao X, Zheng G. Disturbed tryptophan metabolism correlating to progression and metastasis of esophageal squamous cell carcinoma. *Biochem Biophys Res Commun.* (2017) 486:781–87.doi: 10.1016/j.bbrc.2017.03.120
19. Sun C, Li T, Song X, Huang L, Zang Q, Xu J, et al. Spatially resolved metabolomics to discover tumor-Associated metabolic alterations. *Proc Natl Acad Sci USA.* (2019) 116:52–57.doi: 10.1073/pnas.1808950116
20. Xu J, Chen Y, Zhang R, He J, Song Y, Wang J, et al. Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging. *Sci Rep.* (2016) 6:35010.doi: 10.1038/srep35010
21. Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, et al. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. *PLoS ONE.* (2011) 6:e24143.doi: 10.1371/journal.pone.0024143
22. Zhang H, Wang L, Hou Z, Ma H, Mamtimin B, Hasim A, et al. Metabolomic profiling reveals potential biomarkers in esophageal cancer progression using liquid chromatography-Mass spectrometry platform. *Biochem Biophys Res Commun.* (2017) 491:119–25.doi: 10.1016/j.bbrc.2017.07.060
23. Tokunaga M, Kami K, Ozawa S, Oguma J, Kazuno A, Miyachi H, et al. Metabolome analysis of esophageal cancer tissues using capillary electrophoresis-time-of-flight mass spectrometry. *Int J Oncol.* (2018) 52:1947–58.doi: 10.3892/ijo.2018.4340
24. Gu J, Liang D, Pierzynski JA, Zheng L, Ye Y, Zhang J, et al. D-Mannose: a Novel prognostic biomarker for patients with esophageal adenocarcinoma. *Carcinogenesis.* (2017) 38:162–67.doi: 10.1093/carcin/bgw207
25. De Livera AM, Sysi-Aho M, Jacob L, Gagnon-Bartsch JA, Castillo S, Simpson JA, et al. Statistical methods for handling unwanted variation in metabolomics data. *Anal Chem.* (2015) 87:3606–15.doi: 10.1021/ac502439y
26. Yi X, Zhu J, Zhang J, Gao Y, Chen Z, Lu S, et al. Investigation of the reverse effect of danhong injection on doxorubicin-Induced cardiotoxicity in h9c2 cells: insight by lc-Ms based non-Targeted metabolomic analysis. *J Pharm Biomed Anal.* (2018) 152:264–70.doi: 10.1016/j.jpba.2018.02.012
27. Huang Q, Tan Y, Yin P, Ye G, Gao P, Lu X, et al. Metabolic characterization of hepatocellular carcinoma using nontargeted tissue metabolomics. *Cancer Res.* (2013) 73:4992–5002.doi: 10.1158/0008-5472.CAN-13-0308
28. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-Microrna signature. *Gut.* (2010) 59:579–85.doi: 10.1136/gut.2008.175497
29. Guo JC, Wu Y, Chen Y, Pan F, Wu ZY, Zhang JS, et al. Protein-Coding genes combined with long noncoding rna as a novel transcriptome molecular staging model to predict the survival of patients with esophageal squamous cell carcinoma. *Cancer Commun.* (2018) 38:4.doi: 10.1186/s40880-018-0277-0
30. Huang ML, Hung YH, Lee WM, Li RK, Jiang BR. Svm-Rfe based feature selection and taguchi parameters optimization for multiclass svm classifier. *Sci World J.* (2014) 2014:795624.doi: 10.1155/2014/795624

31. Ma B, Jiang H, Wen D, Hu J, Han L, Liu W, et al. Transcriptome analyses identify a metabolic gene signature indicative of dedifferentiation of papillary thyroid cancer. *J Clin Endocrinol Metab.* (2019) 104:3713–25.doi: 10.1210/jc.2018-02686
32. Moretti S, Menicali E, Voce P, Morelli S, Cantarelli S, Sponziello M, et al. Indoleamine 2, 3-Dioxygenase 1 (Ido1) is up-Regulated in thyroid carcinoma and drives the development of an immunosuppressant tumor microenvironment. *J Clin Endocrinol Metab.* (2014) 99:E832–40.doi: 10.1210/jc.2013-3351
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a Nonparametric approach. *Biometrics.* (1988) 44:837–45.
34. Shen J, Song R, Hodges TR, Heimberger AB, Zhao H. Identification of metabolites in plasma for predicting survival in glioblastoma. *Mol Carcinog.* (2018) 57:1078–84.doi: 10.1002/mc.22815
35. Tian Y, Wang Z, Liu X, Duan J, Feng G, Yin Y, et al. Prediction of chemotherapeutic efficacy in non-Small cell lung cancer by serum metabolomic profiling. *Clin Cancer Res.* (2018) 24:2100–09.doi: 10.1158/1078-0432.CCR-17-2855
36. Platten M, Wick W, Van den Eynde BJ. Tryptophan catabolism in cancer: beyond ido and tryptophan depletion. *Cancer Res.* (2012) 72:5435–40.doi: 10.1158/0008-5472.CAN-12-0569
37. Liu Y, Liang X, Dong W, Fang Y, Lv J, Zhang T, et al. Tumor-Repopulating cells induce pd-1 expression in cd8(+) T Cells by transferring kynurenine and ahr activation. *Cancer Cell.* (2018) 33:480–94 e7.doi: 10.1016/j.ccell.2018.02.005
38. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab.* (2016) 23:27–47.doi: 10.1016/j.cmet.2015.12.006
39. Kamphorst JJ, Cross JR, Fan J, de Stanchina E, Mathew R, White EP, et al. Hypoxic and ras-Transformed cells support growth by scavenging unsaturated fatty acids from lysophospholipids. *Proc Natl Acad Sci USA.* (2013) 110:8882–7.doi: 10.1073/pnas.1307237110
40. Zhang T, Wu X, Yin M, Fan L, Zhang H, Zhao F, et al. Discrimination between malignant and benign ovarian tumors by plasma metabolomic profiling using ultra performance liquid chromatography/mass spectrometry. *Clin Chim Acta.* (2012) 413:861–8.doi: 10.1016/j.cca.2012.01.026
41. Gruemer HD. Formation of hippuric acid from phenylalanine labelled with carbon-14 in phenylketonuric subjects. *Nature.* (1961) 189:63–4.doi: 10.1038/189063a0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Chen, Dai, Huang, Chen, Gao, Li, Wang, Chen, Yang, Hong, Zeng and Mao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 September 2020
doi: 10.3389/fonc.2020.01769

Evaluation of Artificial Intelligence in Participating Structure-Based Virtual Screening for Identifying Novel Interleukin-1 Receptor Associated Kinase-1 Inhibitors

Jinxin Che[1†], Ruiwei Feng[2,3†], Jian Gao[1], Hongyun Yu[2,3], Qinjie Weng[1], Qiaojun He[1,4,5], Xiaowu Dong[1,4,5], Jian Wu[2,3,4*] and Bo Yang[1,4*]*

[1] Hangzhou Institute of Innovative Medicine, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, [2] College of Computer Science and Technology, Zhejiang University, Hangzhou, China, [3] Real Doctor AI Research Centre, School of Medicine, Zhejiang University, Hangzhou, China, [4] Innovation Institute for Artificial Intelligence in Medicine, Zhejiang University, Hangzhou, China, [5] Cancer Center of Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Weiwei Xue, Chongqing University, China

Reviewed by:
Yun Lyna Luo, Western University of Health Sciences, United States
Amarjit Luniwal, North American Science Associates Inc., United States

***Correspondence:**
Xiaowu Dong
dongxw@zju.edu.cn
Jian Wu
wujian2000@zju.edu.cn
Bo Yang
yang924@zju.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Oncology

Received: *18 May 2020*
Accepted: *07 August 2020*
Published: *03 September 2020*

Citation:
Che J, Feng R, Gao J, Yu H, Weng Q, He Q, Dong X, Wu J and Yang B (2020) Evaluation of Artificial Intelligence in Participating Structure-Based Virtual Screening for Identifying Novel Interleukin-1 Receptor Associated Kinase-1 Inhibitors. Front. Oncol. 10:1769. doi: 10.3389/fonc.2020.01769

Interleukin-1 receptor associated kinase-1 (IRAK1) exhibits important roles in inflammation, infection, and autoimmune diseases; however, only a few inhibitors have been discovered. In this study, at first, a discriminatory structure-based virtual screening (SBVS) was employed, but only one active compound (compound **1**, IC_{50} = 2.25 μM) was identified. The low hit rate (2.63%) which derives from the weak discriminatory power of docking among high-scored molecules was observed in our virtual screening (VS) process for IRAK1 inhibitor. Furthermore, an artificial intelligence (AI) method, which employed a support vector machine (SVM) model, integrated information of molecular docking, pharmacophore scoring and molecular descriptors was constructed to enhance the traditional IRAK1-VS protocol. Using AI, it was found that VS of IRAK1 inhibitors excluded by over 50% of the inactive compounds, which could significantly improve the prediction accuracy of the SBVS model. Moreover, four active molecules (two of which exhibited comparative IC_{50} with compound **1**) were accurately identified from a set of highly similar candidates. Amongst, compounds with better activity exhibited good selectivity against IRAK4. The AI assisted workflow could serve as an effective tool for enhancement of SBVS.

Keywords: virtual screening, artificial intelligence, machine learning, IRAK1, inhibitors

INTRODUCTION

In the process of drug discovery, hunting for lead compounds is not only a starting point, but also a very challenging task. With the emergence of comprehensive chemical databases, high throughput screening (HTS), and virtual screening (VS) have been employed for finding lead compounds from known chemicals (1). As a complementary approach to HTS (2), VS filters chemicals through ligand- or structure-based approaches by taking advantages of high-performance computers,

overcomes some shortcomings of HTS, and remarkably reducing the time, money, and resources involved (3, 4). However, some problems still exist in the individual VS method. For example, the scoring functions of virtual screening are not accurate enough to predict the protein-ligand binding affinity and this leads to a high rate of false results, which needs combined strategies to improve prediction accuracy in a sequential or parallel manner (5, 6).

In recent years, artificial intelligence (AI) has offered new opportunities in drug discovery. The AI techniques that have displayed superior performances (7, 8) in finding new active chemicals include Naive Bayes, support vector machine (SVM), random forest (RF), feed-forward artificial neural networks (ANNs), and deep neural network approaches. It has been reported that the hit rate of VS can be significantly improved by combining AI methods. For example, Leong et al. developed an accurate ensemble docking scheme, which established a SVM based on combinatorial docking features and molecular descriptors to predict N-methyl-D-aspartate-receptor GluN1-ligand binding affinity (9). Tian et al. integrated ensemble molecular docking and complex-based pharmacophore searching using Naive Bayesian classification and recursive partitioning, which were of great significance in the discovery of novel ROCK inhibitors and increased the VS hit rate to 28.95% (10). Hence, the application of AI in VS seems promising.

Interleukin receptor associated kinase-1 (IRAK1) is a downstream member of the serine-threonine kinase interleukin receptor associated kinase (IRAK) family. Once the IL1-Rs and toll-like receptors (TLRs) are activated, IRAK4 is recruited to form the signaling complex with myeloid differentiation primary-response gene 88 (MyD88) and then IRAK1 is phosphorylated, which plays a crucial role in inflammation, infection, and autoimmune diseases (11). Researchers found that the suppression of IRAK1, either by inhibitors or RNAi, has potent activity against Waldenström's macroglobulinemia (12), myelodysplastic syndrome (13), and certain subtypes of acute myeloid leukemia (14). However, IRAK1 selective inhibitors are rare, and most of the compounds that inhibit IRAK1 are also IRAK4 inhibitors. The benzimidazole derivative 1 shown in **Figure 1** is commonly used in bio-experiment (15). In addition, researchers found that the anaplastic lymphoma kinase (ALK) inhibitor 2 (16), and Pacritinib (17), the JAK/FLT3 inhibitor 3, could be potent inhibitors of IRAK1.

As a continuation of our virtual screening work which identified new inhibitors targeting NIK, CHK1, Akt, etc. (18–24). In this study, we performed a traditional VS procedure to identify potential inhibitors of IRAK1. Using a well-designed screening process, only one compound (compound **1**, IC_{50} = 2.25 μM) was discovered, which was far from satisfactory. Considering the advantages of AI in VS, we further established a machine-learning model that combines multiple docking, complementary pharmacophore mapping, and molecular descriptors on the basis of a traditional VS workflow to increase the enrichment rate among high- scoring compounds. Training data consisted of IRAK1 inhibitors and decoys that were prepared for SVM, XGBOOST, and LGBM models. Finally, we used the SVM model that exhibited the best accuracy, to validate the activities of molecules in the post-docking stage and found that it significantly improved the performance of traditional VS and excluded over half of the false positive candidates which was predicted positive in VS but showed no activities in bioassay. Moreover, four other active compounds (compound **2**, IC_{50} = 2.32 μM; compound **3**, IC_{50} = 2.48 μM; compound **4** IC_{50} = 18.04 μM; and compound **5** IC_{50} = 23.75 μM) were identified from a series of highly similar compounds by utilizing this model, which demonstrated that the integrated VS strategy enhanced by AI was promising in the process of drug discovery.

MATERIALS AND METHODS

Evaluation of the Docking Method

Crystal structure of human wild type IRAK1 (PDB ID: 6BFN) with its inhibitor (DL1) was downloaded from RCSB Protein Data Bank. Each chain of the protein was prepared separately including the removal of water, alternate position of residues, the addition of hydrogens, the assignment of bond orders, the optimization of H-bonds and the restrained minimization of energy using Schrödinger's *Protein Preparation Wizard*. The receptor grids were generated using the *Receptor Grid Generation* module and all preparation parameters were set to default. The inhibitors were re-docked into their receptors to calculate the root mean square deviation (RMSD) value and compared with their original structure. Furthermore, 594 inhibitors with known IC_{50} values were collected from the ChEMBL database (**Supplementary Table S1**) and 150 potent inhibitors were selected after structural clustering to generate a decoy dataset using the DUD-E database (25). All 150 inhibitors and 9200 decoys were prepared using the *Ligprep* module to generate the possible ionization states and three-dimensional conformations. Lastly, the compounds were docked separately to the prepared chains of 6BFN using the SP and XP patterns in the *Ligand Docking* module. Similarly, the compounds and proteins were docked using AutoDock after preparation via *AutoDockTools* with a grid size of 54 points. The docking scores were collected to calculate the *p*-value after the docking process.

Pharmacophore Construction

The best conformation of 6BFN_B re-docking complex was applied to construct the pharmacophore model for ligand-based VS. The pharmacophore containing at most 7 features and a receptor-based excluded volume shell was created using *Develop Pharmacophore Model* in the Schrödinger module, with the method set to E-Pharmacophore.

Virtual Screening

A database containing 1.5 million compounds was downloaded from ChemDiv and was filtered using Lipinski's Rule-of-Five before commencing the process of VS. Molecules which passed the filter were further aligned to the established pharmacophore model in the *Ligand Screening* module of Schrödinger so that all compounds with similar structural features could be considered for the next step. After the preparation in *Ligprep*, the molecules preserved were docked in *Ligand Docking* module

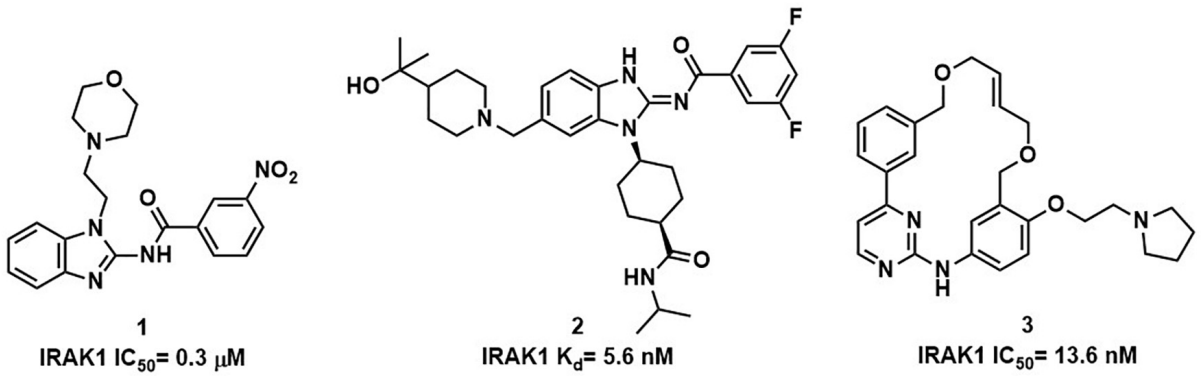


FIGURE 1 | Examples of compounds with IRAK1 inhibitory activity.

under standard precision (SP) and default parameters. The top-scoring compounds were subjected to extra precision (XP) docking in order to exclude the molecules that were low-ranking in the XP mode. The compounds which underwent the entire screening protocol were clustered based on their structural similarity and selected manually for the bioactivity test.

IRAK1 Bioactivity Test

Mobility shift assay was applied to test the bioactivity of the candidates at a concentration of 20 μM using Staurosporine as the positive control. All compounds were dissolved in 100% DMSO to yield a final concentration of 2 mM. IRAK1 kinase and kinase substrate (including ATP) were dissolved in 1 × Kinase buffer. A mixture of 250 nL of the test compound and 10 μL kinase solution were centrifuged at 1000 rpm for 30 s and incubated for 10 min at 25°C in a 384-well white plate. The plate was centrifuged at 1000 rpm and incubated for 60 min after the addition of 15 μL kinase solution (including ATP). Finally, 30 μL stop buffer was added to terminate the reaction and the conversion rate was evaluated using ELIASA (Caliper EZ Reader, Perkin Elmer) to calculate the rate of inhibition.

SVM-Assisted Selection and Bioactivity Test

Machine Learning Dataset

An inhibitor dataset and a decoy dataset made up the training set of the machine- learning model. The inhibitors of IRAK1 were collected from ChEMBL Database and 150 potent inhibitors were extracted after a structural clustering. For the decoy dataset, 150 chemicals were randomly selected from the single-target compounds database of ChEMBL. Compounds in the training set were prepared and docked into each chain of 6BFN using Glide-SP, Glide-XP, and AutoDock. All docking parameters used were the same as those in the evaluation process.

A group of novel pharmacophore models were generated using the *3D QSAR Pharmacophore Generation* module of Discovery Studio Client 2.5 using the 594 inhibitors found in the ChEMBL Database. H-bond acceptors, H-bond donors, hydrophobic molecules, and aromatic compounds were considered and the upper limit of each feature was set to 5. For every molecule, at most 255 conformations were generated to find the best conformation and the maximum pharmacophores was set to 10. Since the pharmacophore models were built, the best conformations of compounds in the training dataset were generated and mapped to the pharmacophores flexibly in the *Ligand Pharmacophore Mapping* module of Discovery Studio Client 2.5. The fit value of each molecule was extracted. The compounds that failed to map onto the pharmacophore model were assigned fit values that were set to 0 uniformly.

PaDEL v2.20 was used to calculate the 1D and 2D descriptors of the compounds in the inhibitor and decoy sets. Salts were removed and the nitro groups were standardized before the calculation.

Data Preprocessing

In this study, we conducted data pre-processing including index elimination and data normalization. We collected almost 1460 indices to represent each molecule, which is large for any machine learning model to analyze. Thus, we used several approaches to eliminate some indices. This process mainly consisted of the following 4 steps:

(1) Counting: We counted the value of each index. Some of the indices on the condition were excluded if more than 90% molecules had the same value which indicates that there is no crucial significance of such indices.
(2) Correlation analysis: The correlation values of all indices were calculated between any pairs. If the value was larger than 0.85, one of the pairs was excluded because of their high collinearity.
(3) *T*-test analysis: Then we calculated the *T*-test scores for the means of two indices. This test assumed that the populations had identical variances by default.
(4) Principal components analysis (PCA): Finally, we utilized PCA for dimensionality reduction.

After conducting these four steps, only 244 of the 1460 indices remained. Then, the normalization on the basis of mean and standard deviation of a batch was conducted.

Model Construction

In this paper, we constructed three traditional machine learning models including SVM, LGBM, and XGBoost using Scikit-learn package.

(1) Support vector machine (26) is one of the most commonly used binary classification model. Its basic principle is to defy a linear classifier in the feature space with the largest spacing. In this study, we set C as 1, using a kernel function of RBF. RBF was calculated as follows:

$$RBF = exp(-\gamma |u - v|^2)$$

Besides, we used "degree 3, coef 0" for the kernel function. We used the shrinking heuristic and probability methods. When training error was less than 1e-3, we stopped further training.

(2) LightGBM is a gradient boosting framework, based on a histogram decision tree algorithm. Using leaf-wise growth strategy with depth limitation, LightGBM model can yield better performance and prevent overfitting. In this study, we used the gbdt boosting method and set the learning rate as 0.1 in order to prevent overfitting. We constructed the classifier with 10 leaves, with "max depth 3" and "minimum child samples 31."

(3) XGBoost uses depth-wise strategy. It first ranks all features according to their values using a pre-sorted algorithm and then splits samples. However, pre-sorted algorithm may cause additional memory space. In this study, an XGBoost model with 150 estimators, max depth 3, and minimum child weight 1 was conducted. We randomly selected 80% of samples and 80% of features to build a decision tree and used binary logistic function to learn and update parameters.

Metrics

In this study, to compare the performance of different models, we utilized the area under the curve (AUC) and accuracy (ACC) as metrics. AUC was defined as the area of the receiver operating characteristic (ROC) Curve, which was defined by true positive rate (TPR) and false positive rate (FPR). The TPR and FPR are given by:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

Where TP, FP, TN, and FN were true positives, false positives, true negatives and false negatives, respectively. AUC and ACC were computed using the following equations.

$$AUC = \int_{x=0}^{1} TPR\left(FPR^{-1}(x)\right) dx$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Prediction and Bioassay

Similarity search was carried out using the MolPort database. Ten chemicals, whose Tanimoto similarity was larger than 0.8 compared with compound **1**, were collected along with their multiple docking scores, pharmacophore scores, and molecular descriptors. These data were processed using the best model to predict their inhibition toward IRAK1. Candidates in the last stage of virtual screening procedure were also predicted similarly.

Lastly, compounds derived using the similarity search were bought from TargetMol and their inhibition rate was tested using the method described above. The dose-effect curve (all compounds were dissolved in DMSO and diluted 10-fold from 100 μM to 1 nM) was fitted to calculate the IC_{50} value.

Molecular Dynamics Simulation

The binding mode of compound **1** was confirmed using molecular dynamics simulation. The parameter files of compound **1** were generated in *antechamber* module of *Amber* and was combined with 6BFN_B under ff99SB and gaff force field. After six chloride ions were added, the neutralized system was solvated in a tetrahedral box of TIP3P and the distance between box boundary and IRAK1 protein was set to 10.0 Å. At first, the energy of the entire system was relaxed using three steps: the hydrogen atoms, chloride ions, and water were optimized using 2500 steps of steepest descent minimization and 2500 steps of conjugate gradient minimization; the side chains of the protein, chloride ions, and waters were relaxed using 2500 steps of steepest descent minimization and 2500 steps of conjugate gradient minimization; the system was minimized using 2500 steps of steepest descent minimization and 2500 steps of conjugate gradient minimization. The system was heated to 300 K in 100 ps and equilibrated for 100 ps in NPT mode (temperature = 300 K; pressure = 1 atm) with the heavy atoms of the complex being restrained. Then, the whole system was equilibrated in NPT mode for 100 ps. Lastly, a five nanosecond MD simulation was carried out and the binding energy was decomposed.

RESULTS

Virtual Screening

In order to select the most effective screening method, we evaluated the performance of different proteins and molecular-docking software. For each monomer of the IRAK1 crystal structure (6BFN_A and 6BFN_B from protein data bank) (27), we docked the original ligand into the binding pocket to evaluate the reproducibility of several frequently-used docking methods. We then shortlisted 150 potent IRAK1 inhibitors and used 9200 random decoys to judge whether the docking methods and protein chains could distinguish the inhibitors from decoys effectively. As seen in **Table 1**, the RMSD value for each docking approach was less than 1, which indicated that all approaches could reproduce the structure of each complex accurately. The best capabilities of discrimination power were exhibited by 6BFN_B coupling with *Glide SP* and *Glide XP*, which reached the smallest *p*-value of 10^{-32} (**Supplementary Figure S1**).

TABLE 1 | Evaluation of docking methods.

PDB_ID	Method	RMSD	Docking score	*p*_value
6BFN_A	*AutoDock*	0.5755	−10.0	8.92E−27
6BFN_B	*AutoDock*	0.5665	−9.9	4.79E−26
6BFN_A	*Glide_SP*	0.4728	−9.8	2.24E−30
6BFN_B	*Glide_SP*	0.9543	−9.3	6.87E−32
6BFN_A	*Glide_XP*	0.4742	−6.8	5.27E−32
6BFN_B	*Glide_XP*	0.9438	−6.4	4.60E−32

The ROC curve and the AUC are shown in **Figure 2**. Since the performances of the docking methods used in our study were not adequately satisfactory, we established a pharmacophore model to filter out molecules that we did not wish to pursue beyond the docking process. According to the conformation of the 6BFN_B re-docking results, we analyzed the binding mode of the IRAK1 inhibitor (**Figure 3A**) and constructed a pharmacophore, which contained a hydrogen-bond donor, a hydrogen-bond acceptor, and an aromatic ring, using the *Develop Pharmacophore Model* of Schrödinger which is suitable for the preliminary virtual screening and saving the time costs of calculation (**Figure 3B**).

We collected 1.5 million chemicals from the ChemDiv database. Firstly, compounds in the database were pre-processed using a filter which could exclude chemicals that did not conform to the Lipinski's rule of five (28). Next, we screened the remaining 1.2 million compounds in the established pharmacophore model and 56,000 molecules having a structure similar to that of the IRAK1 inhibitor were preserved. *Glide* SP docking was applied and all docked compounds were ranked by their docking scores. We selected the first 506 molecules for *Glide* XP docking in order to eliminate the compounds with lower docking scores. Thus, we obtained 353 molecules that possessed both, high docking scores and different docking precision. These compounds were clustered according to their molecular fingerprints. The most representative 43 candidates (38 compounds were purchasable) from each cluster were selected for biological assay (**Figure 3C**).

Biological Evaluation

To test whether the selected molecules were active, kinase activity experiments were carried out using the mobility shift assay, in which staurosporine was chosen as the positive control. Each compound was dissolved in DMSO at a concentration of 20 μM and used in the assay (**Supplementary Table S2**). Compound **1** (Y041-8246) showed a moderate inhibition rate (85.5% at 20 μM) compared to staurosporine (IC_{50} = 59.29 nM; **Supplementary Figure S2**). The screening hit rate was 2.63% (1/38). The Tanimoto similarities between compound **1** and the known IRAK1 inhibitors were below 0.34, thus indicating that this was a new structure which bears benzofuran scaffold for IRAK1 inhibition. Moreover, structure clustering of IRAK1 inhibitors collected from ChEMBL was performed, the result indicated that benzofuran derivatives were different from any of these scaffolds (**Supplementary Figure S3**).

Establishment of Machine Learning Model

Considering the unsatisfactory performance of the traditional VS method with a hit rate of 2.63% in this study, which could not completely meet the demand of drug discovery, we attempted to find a better approach to discriminate lead compounds from a set of high-scored compounds. Ten molecules from the MolPort database[1] whose Tanimoto similarities were higher

[1]www.molport.com

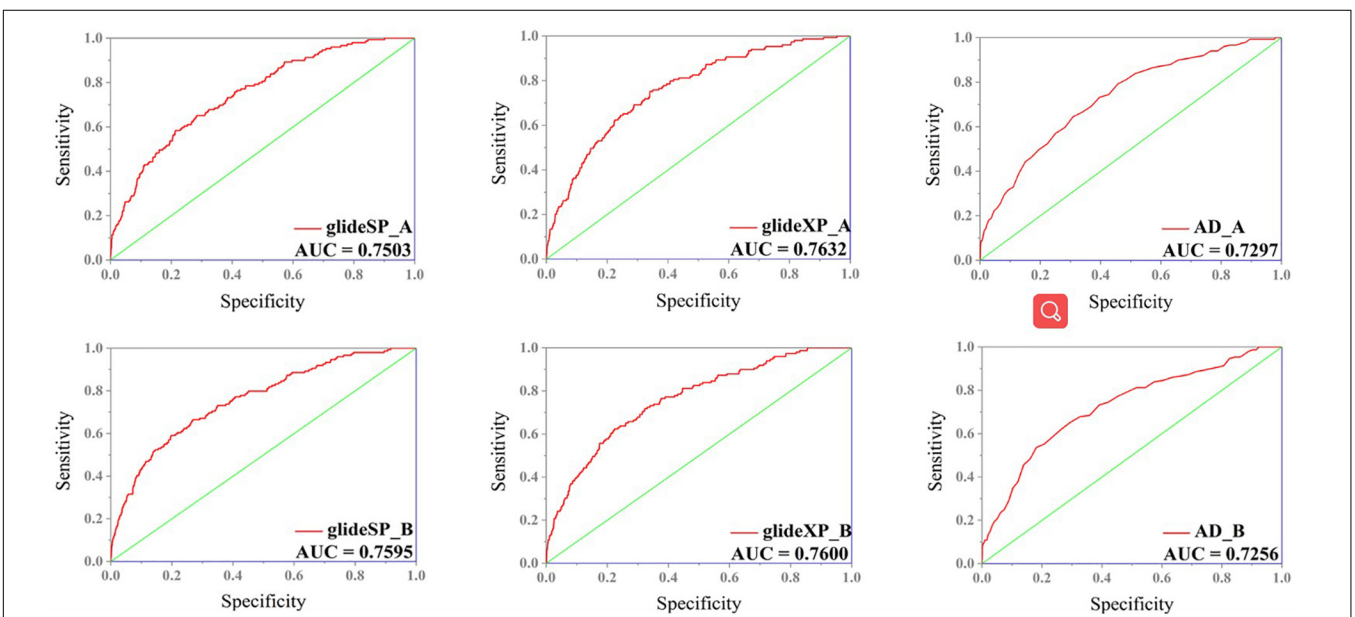


FIGURE 2 | ROC curve of different combination of receptors (6BFN_A, 6BFN_B) and docking software (Glide SP, XP, and AutoDock).

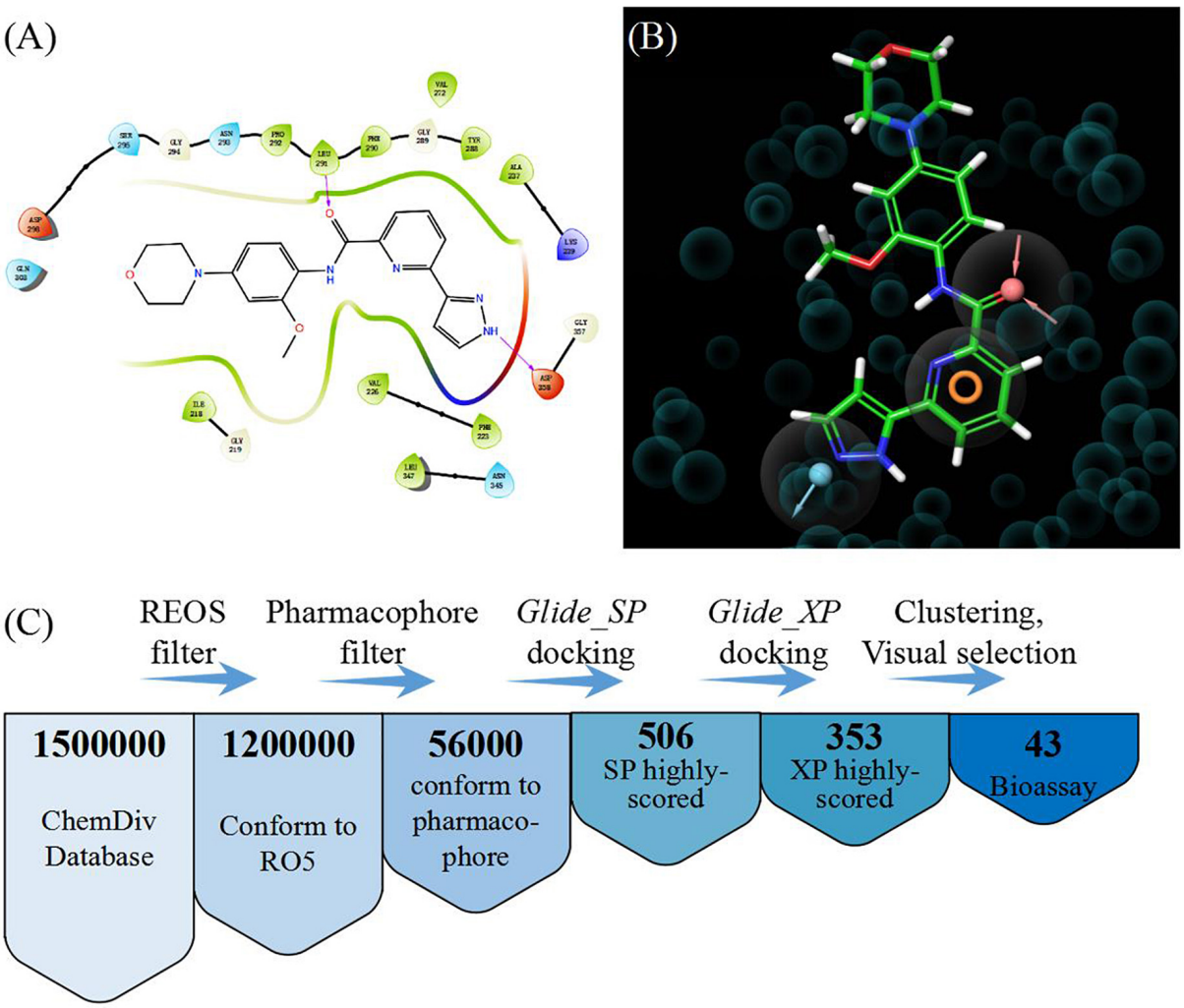


FIGURE 3 | (A) Binding mode of the ligand (DL1) in 6BFN_B; **(B)** Pharmacophore model; and **(C)** Work flow of virtual screening.

than 0.8, when compared with compound **1**, were chosen to be further studied. In order to take the pharmacophore and structural information into consideration, we calculated multiple docking scores, complementary pharmacophore mapping scores, and molecular descriptors (**Figure 4**). The molecules in the training set consisted of 150 potent IRAK1 inhibitors and 150 decoys (**Supplementary Table S3**). Their docking scores were obtained using AutoDock (29), Glide SP, and Glide XP toward 6BFN_A and 6BFN_B, respectively. In this model, since it was need to construct different pharmacophore models according to known ligands, Discovery Studio was employed and all selected molecules were then mapped to pharmacophores to calculate their fit values, which demonstrated their extent of conformity to the pharmacophore models. The one- and two-dimensional molecular descriptors were calculated using *PaDEL* (30) to complement the physicochemical properties. Lastly, 6 docking scores, 9 fit values, and 1444 molecular descriptors were obtained for each compound.

Based on the characteristics of the data, we employed several traditional machine learning methods for activity prediction. In this section, we describe how data pre-processing was performed and the three machine-learning models for activity prediction was constructed. Considering the inactivity of most molecules, we used ACC and AUC to measure the performance of each model. For a fair comparison, we divided the 300 molecules into 5 groups to conduct a 5-fold cross validation. For the final test, all parameters were selected based on the best ACC performance of models on validation sets. All models were developed using python 3.7.3 with TensorFlow deep learning library and all experiments were constructed on an NVIDIA GeForce GTX 1080Ti GPU.

In this study, we compared the performance of each model using data with different inputs. The mean AUC and ACC scores of different models on 5-fold validation sets are shown in **Table 2**, different inputs (docking scores, pharmacophore mapping scores, and molecular descriptors) are presented in the first column and simply denoted as DS, PS, and MD, respectively.

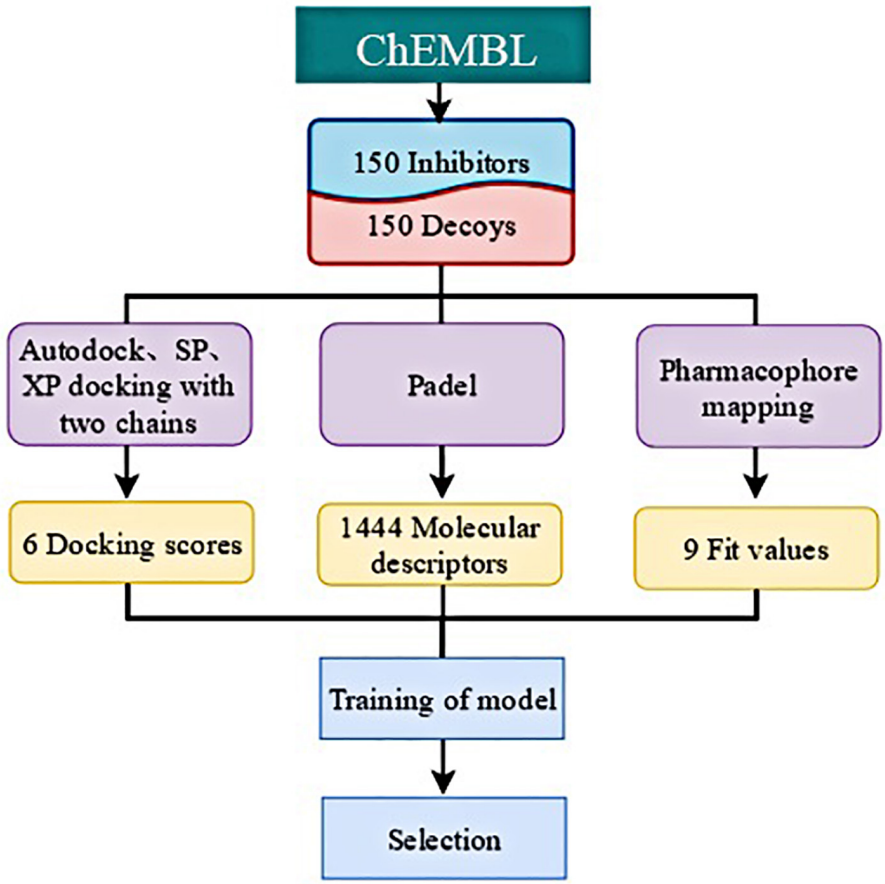


FIGURE 4 | Construction of training dataset.

That is to say, the AUC and ACC results of different models in the first row are obtained by models receiving docking scores, pharmacophore mapping scores and molecular descriptors at the same time, while results in the last row are obtained by models merely receiving molecular descriptor. Obviously, a slight improvement of AUC or ACC is observed when docking and pharmacophore information were taken into consideration. Besides, it is notable in each row that SVM outperforms all the other two models. Therefore, we use SVM with all the three kind of inputs as the final model, denoted as SVM (DS+PS+MD). **Table 2** and **Figure 5** illustrate the performance of each combination of model and input data.

TABLE 2 | Comparison of our three traditional machine-learning models for activity prediction.

Model	SVM		LGBM		XGBoost	
Data	AUC	ACC	AUC	ACC	AUC	ACC
DS+PS+MD	0.8889	0.78	0.8022	0.75	0.8522	0.77
PS+MD	0.8933	0.75	0.8044	0.72	0.8633	0.78
MD	0.8767	0.77	0.8267	0.73	0.8178	0.70

SVM-Model Can Identify Active Compounds Accurately

To verify the practicality of the SVM (DS+PS+MD) model, we predicted the activity of molecules at later stages of VS (**Figure 6**). Among the molecules exhibiting superior docking scores in Glide SP and Glide XP, our model identified 51.78% selected from Glide SP, and 56.66% from Glide XP docking as inactive. Among the 38 candidates whose inhibitory had been tested, 21 inactive chemicals were accurately identified. Although there also existed several inactive candidates that were classified as active, compound **1** was picked precisely.

Ten compounds in the MolPort database with a Tanimoto similarity greater than 0.8 as compared to compound **1** were collected. Since almost all multiple docking scores appeared similar and very close (**Table 3**), the docking software was unable to yield accurate results. When the training information of these compounds was fed into the SVM model, the results classified two compounds as inactive, although one of them had the best average docking score. The decrease in the exclusion rate from 50 to 20% indicated that this model was competent to identify the structural features of IRAK1 inhibitors, which could thus help computational chemists select lead compounds more accurately.

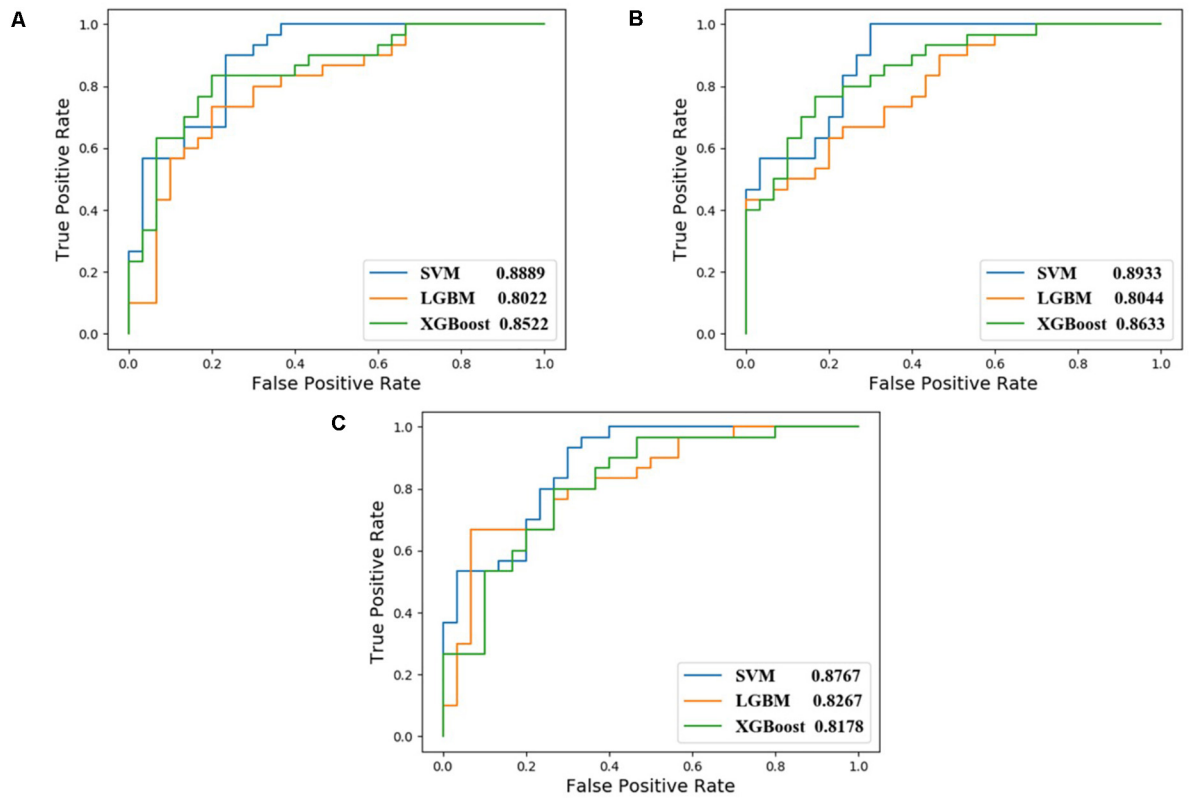


FIGURE 5 | ROC curves of different combination of models and input data. **(A)** corresponds to feed model with data containing DS, PS, and MD; **(B)** for PS and MD; and **(C)** for data with MD only.

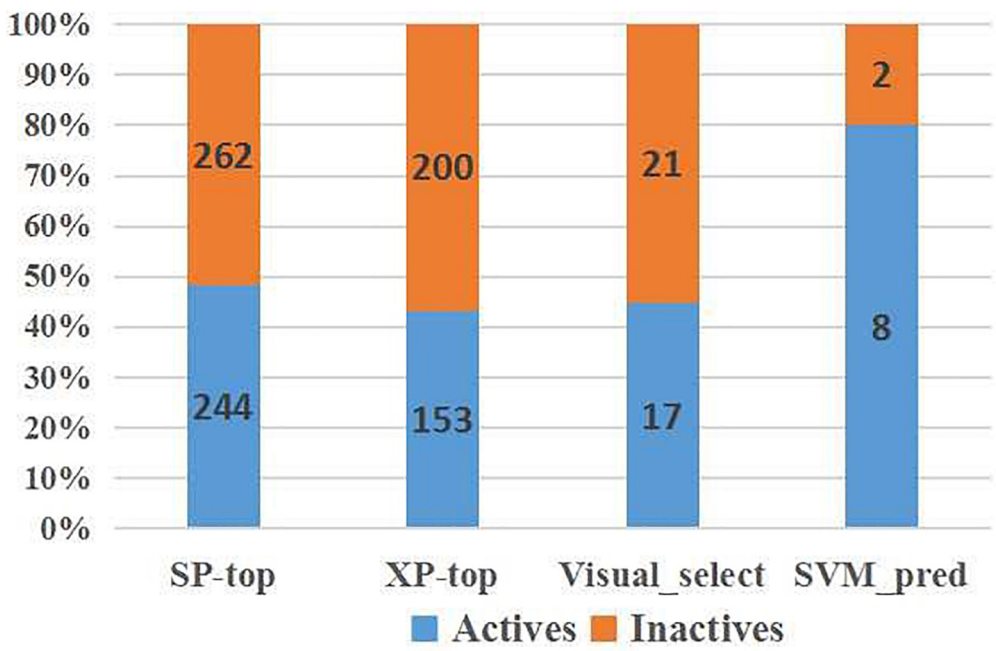


FIGURE 6 | SVM model performance on IRAK1 VS candidates (top scoring compounds derived from Glide SP docking, Glide XP docking, visual selection, and SVM prediction).

Next, the activities of these compounds were tested. Ten candidates were bought from TargetMol and their IC_{50}s were determined. The initial concentration was set to 100 μM and a serial 10-fold dilution was made to 1 nM. Compounds 2 (Y041-7950) and 3 (Y041-6433) exhibited similar IRAK1 inhibition and their IC_{50} values were determined to be approximately 2 μM. Additionally, two other chemicals showed weaker inhibition and their IC_{50} values were determined to be 18.04 μM (compound **4**: 11570480 from Otava Database) and 23.75 μM, (compound **5**: Y041-7951), respectively (**Figure 7** and **Supplementary Figure S2**). We further tested the activity against IRAK4 of the three most potent IRAK1 inhibitors such as compound **1**, **2**, and **3**. The result indicated that all the compounds showed extreme low inhibitory activities against IRAK4, which IC_{50}s were > 100 μM (**Supplementary Figure S4**). The compounds bearing benzofuran scaffold were demonstrated an over 40-fold selectivity to IRAK4, showing its potential to be developed as selective IRAK1 inhibitor. When these compounds were used in our model, it accurately predicted the results that were determined earlier in the similarity search results. The AI model proved to be an effective tool in this prediction.

Molecular Dynamics Simulation

Since the active compounds share same scaffold, the most potent compound **1** was chosen for the analysis of binding mode with IRAK1. In order to analyze the most stable binding pattern, we performed 5 ns molecular dynamics simulation of compound **1** using AmberTools (31). As seen in **Figure 8**, the RMSD value of the IRAK1 backbone and the ligand reached equilibration after 3 ns and 1 ns simulation, respectively. The average RMSF value of compound 1 was 2.47 Å, indicating a stable binding pattern. Compound 1 formed two hydrogen bonds with the H-bond

TABLE 3 | Detailed information of compounds predicted by SVM model (compound ID; docking scores from Glide SP, Glide XP, and AutoDock; average docking scores; predicted labels and IC_{50} values).

Comp_ID	SP_A	SP_B	XP_A	XP_B	AD_A	AD_B	AVG	AI_Pred	IC_{50}/μ M
2	−10.25	−10.39	−11.06	−11.07	−8.1	−8.1	−9.83	1	2.32
3	−10.22	−10.28	−10.92	−10.91	−9	−9	−10.06	1	2.48
4	−9.51	−7.27	−10.58	−10.68	−11.7	−11.7	−10.24	1	18.04
5	−10.32	−10.52	−11.25	−11.88	−9.4	−9.4	−10.46	1	23.75
6	−9.95	−10.12	−11.84	−11.03	−11.2	−11.2	−10.89	0	NA[a]
7	−9.9	−9.93	−11.81	−11.01	−7.9	−7.9	−9.74	0	NA
8	−9.26	−9.66	−10.26	−10.53	−9.9	−9.9	−9.92	1	NA
9	−9.66	−9.6	−11.26	−11.57	−10	−10	−10.35	1	NA
10	−9.7	−9.33	−10.95	−10.95	−10.1	−10.1	−10.19	1	NA
11	−9.36	−6.19	−8.65	−8.89	−9.4	−9.4	−8.65	1	NA

[a]*NA = No Activity.*

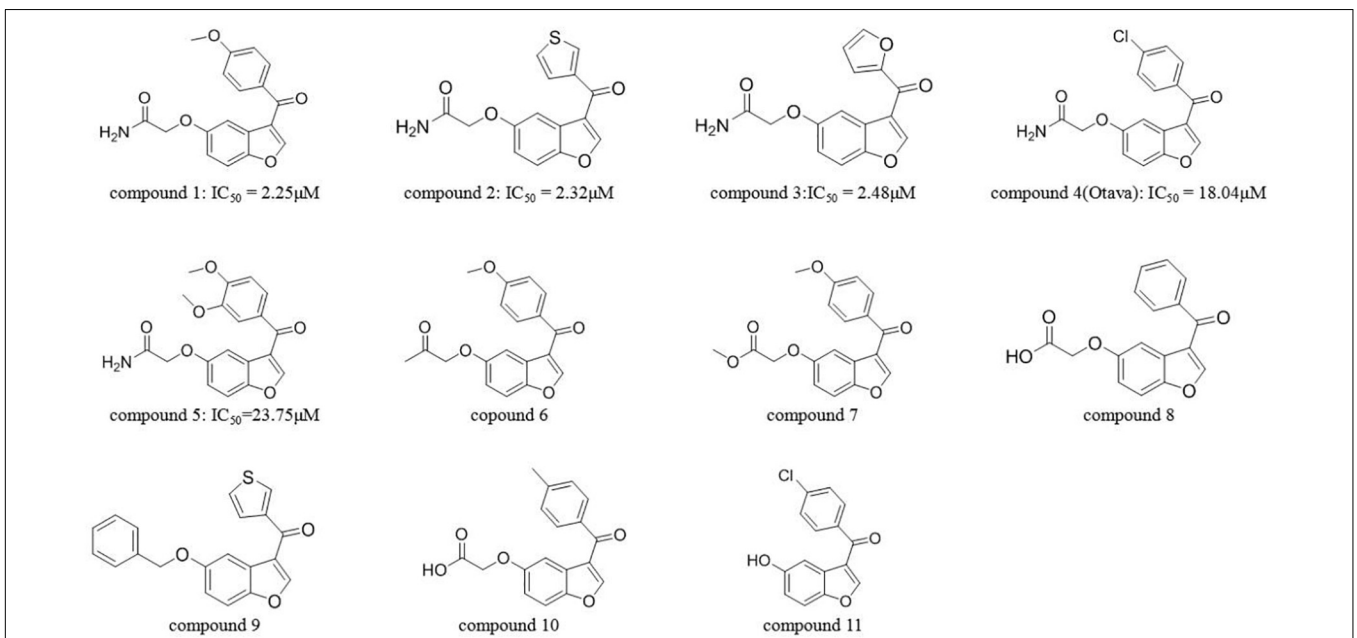


FIGURE 7 | Structure of compound **1** and those predicted using SVM (compounds 2–11).

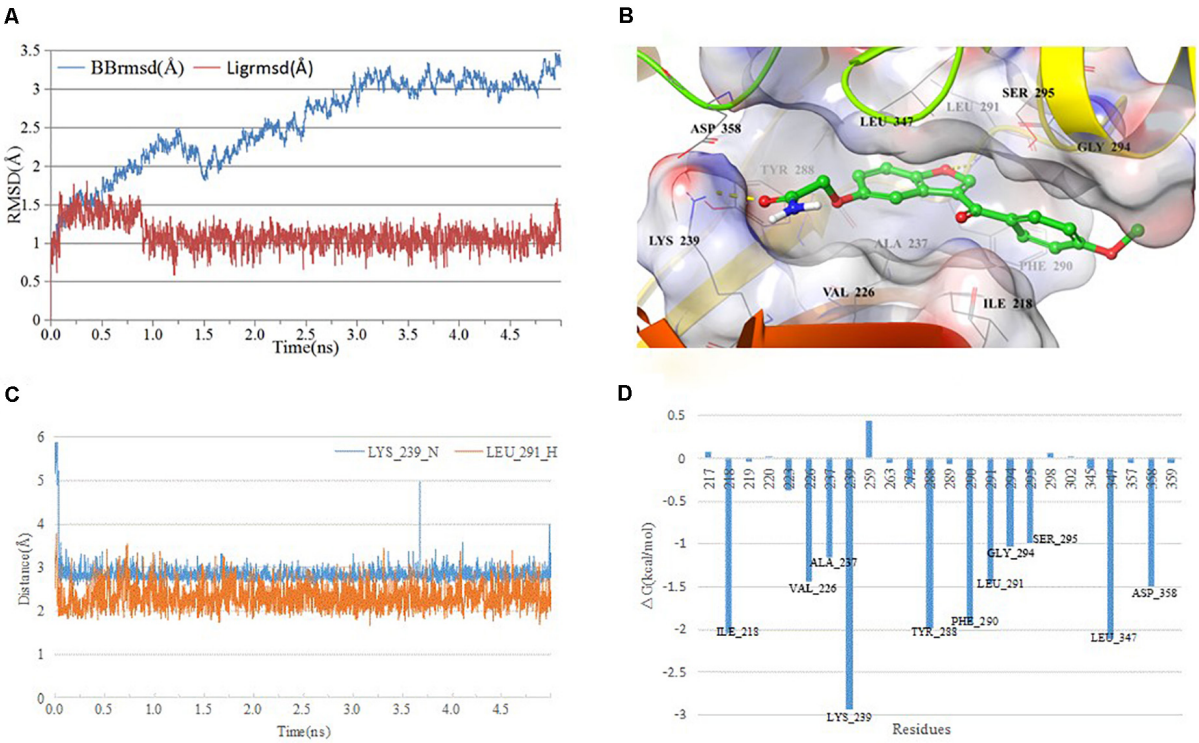


FIGURE 8 | Analysis of MD simulation. **(A)** root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of the IRAK1 backbone during MD simulation; **(B)** 3D-plot of the complex; **(C)** length of the hydrogen bond (compound **1**_O-Leu_291_H, compound **1**_C = O-Lys_239_N); and **(D)** plot of free energy decomposition.

donors from LYS_239 and LEU_291, and also formed stable Van der Waals interaction through hydrophobic or hydrophilic amino acid residues such as ILE_218, PHE_294, LEU_347, and ASP_358. The lengths of the H-bonds and MM-GBSA free energy confirmed that the hydrogen bond between compound **1** and LYS_239 was more stable and contributed significantly to the formation of the complex.

DISCUSSION AND CONCLUSION

In this study, we discovered a novel IRAK1 inhibitor (compound **1**) through traditional virtual screening and also obtained four similar compounds which exhibited good to moderate IRAK1 activity in the AI-aided selection process. As the screening power of docking-based virtual screening was significantly weak (hit rate = 2.63%) to select hit compounds from highly scored molecules, an AI-based discriminatory virtual screening protocol was conducted to assist the final selection procedure of virtual screening.

The major strength of this protocol is a comprehensive integration of activity-related factors through machine learning. Comparing with other machine learning studies (9, 10), this research integrated both protein-ligand binding information, ligand-based pharmacophore information and molecular physiochemical properties into an SVM classifier which can yield satisfactory performance without the requirement of a great deal of known inhibitors. Since diverse receptor structures and pharmacophores were considered, more reliable judgments can be made after dimension reduction and model training. Whereas, there still exists some limitations in the screening process. For example, since the virtual screening targeted a specific binding pocket, it was possible that some positive compounds showing different interaction mode would be excluded. It is necessary to further combine virtual screening with different AI methods for improving its prediction ability more accurately.

Experiments showed that this SVM model can effectively exclude over 50% of the inactive compounds in virtual screening and retain the most promising candidates, which can improve the hit rate prominently in the last phase of VS. Moreover, four molecules were successfully predicted using this model, from a set of compounds that were similar to compound **1**. The model displayed better discriminatory power among highly similar candidates.

The identified compounds all bear benzofuran scaffold, which was different from other IRAK1 inhibitors. The acetamide group seems important in maintaining the inhibitory activity of compounds, such as the IC_{50} values of compound **1** vs compound **7**, compound **2** vs compound **9**, and compound **4** vs compound **11**. Moreover, compound **1**, **2,** and **3** were demonstrated an over 40-fold selectivity to IRAK4, showing its potential to be developed as selective IRAK1 inhibitor. The results

provide valuable knowledge for further optimization and development of IRAK1 inhibitors and demonstrate that AI can assist VS strategy in a sequential manner for identifying new IRAK1 inhibitors.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JC, RF, JG, and HY processed the data, built the models, analyzed the results, and drafted the manuscript. QW and QH assisted in the design of this study. XD, JW, and BY conceived and designed this research. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by Key R&D Program of Zhejiang Province (2020C03010), National Natural Science Foundation of China (81673294 and 81973172), COVID-19 Emergency Scientific Research Project (2020XGZX010), and China Postdoctoral Science Foundation (2019M652123).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01769/full#supplementary-material

REFERENCES

1. Bajorath F. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov.* (2002) 1:882–94. doi: 10.1038/nrd941
2. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today.* (2006) 11:580–94. doi: 10.1016/j.drudis.2006.05.012
3. Rocha SFLD, Olanda CG, Fokoue HH, Sant'Anna CMR. Virtual screening techniques in drug discovery: review and recent applications. *Curr Top Med Chem.* (2019) 19:1751–67. doi: 10.2174/1568026619666190816101948
4. Bajusz D, Ferenczy GG, Keseru GM. Structure-based virtual screening approaches in kinase-directed drug discovery. *Curr Top Med Chem.* (2017) 17:2235–59. doi: 10.2174/1568026617666170224121313
5. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. *J Med Chem.* (2006) 49:5912–31. doi: 10.1021/jm050362n
6. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing scoring functions for protein-ligand interactions. *J Med Chem.* (2004) 47:3032–47. doi: 10.1021/jm030489h
7. Xie QQ, Zhong L, Pan YL, Wang XY, Zhou JP, Lei DW, et al. Combined SVM-based and docking-based virtual screening for retrieving novel inhibitors of c-Met. *Eur J Med Chem.* (2011) 46:3675–80. doi: 10.1016/j.ejmech.2011.05.031
8. Zhong FS, Xing J, Li XT, Liu XH, Fu ZY, Xiong ZP, et al. Artificial intelligence in drug design. *Sci China Life Sci.* (2018) 61:1191–204. doi: 10.1007/s11427-018-9342-2
9. Leong MK, Syu RG, Ding YL, Weng CF. Prediction of N-methyl-D-aspartate receptor GluN1-ligand binding affinity by a novel SVM-pose/SVM-score combinatorial ensemble docking scheme. *Sci Rep UK.* (2017) 7:40053. doi: 10.1038/Srep40053
10. Tian S, Sun HY, Li YY, Pan PC, Li D, Hou TJ. Development and evaluation of an integrated virtual screening strategy by combining molecular docking and pharmacophore searching based on multiple protein structures. *J Chem Inf Model.* (2013) 53:2743–56. doi: 10.1021/ci400382r
11. Rhyasen GW, Starczynowski DT. IRAK signalling in cancer. *Brit J Cancer.* (2015) 112:232–7. doi: 10.1038/bjc.2014.513
12. Ni HW, Shirazi F, Baladandayuthapani V, Lin H, Kuiatse I, Wang H, et al. Targeting myddosome signaling in waldenstrom's macroglobulinemia with the interleukin-1 receptor-associated kinase 1/4 Inhibitor R191. *Clin Cancer Res.* (2018) 24:6408–20. doi: 10.1158/1078-0432.CCR-17-3265
13. Rhyasen GW, Bolanos L, Fang J, Jerez A, Wunderlich M, Rigolino C, et al. Targeting IRAK1 as a therapeutic approach for myelodysplastic syndrome. *Cancer Cell.* (2013) 24:90–104. doi: 10.1016/j.ccr.2013.05.006
14. Hosseini MM, Kurtz SE, Abdelhamed S, Mahmood S, Davare MA, Kaempf A, et al. Inhibition of interleukin-1 receptor-associated kinase-1 is a therapeutic strategy for acute myeloid leukemia subtypes. *Leukemia.* (2018) 32:2374–87. doi: 10.1038/s41375-018-0112-2
15. Hynes J, Nair SK. Advances in the discovery of small-molecule IRAK4 inhibitors. *Annu Rep Med Chem.* (2014) 49:117–33. doi: 10.1016/B978-0-12-800167-7.00009-2
16. Lewis RT, Bode CM, Choquette DM, Potashman M, Romero K, Stellwagen JC, et al. The discovery and optimization of a novel class of potent, selective, and orally bioavailable anaplastic lymphoma kinase (ALK) inhibitors with potential utility for the treatment of cancer. *J Med Chem.* (2012) 55:6523–40. doi: 10.1021/jm3005866
17. Puri A, Mahmood S, Maertens B, Davare M, Kurtz SE, Tognon CE, et al. Pacritinib targets IRAK1 and shows synergy with HDAC and BET inhibitors in acute myeloid leukemia. *Blood.* (2016) 128:3514. doi: 10.1182/blood.V128.22.3514.3514
18. Cheng G, Mei XB, Yan YY, Chen J, Zhang B, Li J, et al. Identification of new NIK inhibitors by discriminatory analysis-based molecular docking and biological evaluation. *Archiv Pharmazie.* (2019) 352:1800374. doi: 10.1002/Ardp.201800374
19. Tong LX, Song PR, Jiang KL, Xu L, Jin TT, Wang PP, et al. Discovery of (R)-5-((5-(1-methyl-1H-pyrazol-4-yl)-4-(methylamino) pyrimidin-2-yl)amino)-3-(piperidin-3-yloxy)picolinonitrile, a novel CHK1 inhibitor for hematologic malignancies. *Eur J Med Chem.* (2019) 173:44–62. doi: 10.1016/j.ejmech.2019.03.062
20. Ying HZ, Xie JF, Liu XG, Yao TT, Dong XW, Hu CQ. Discriminatory analysis based molecular docking study for in silico identification of epigallocatechin-3-gallate (EGCG) derivatives as B-Raf(V600E) inhibitors. *Rsc Adv.* (2017) 7:44820–6. doi: 10.1039/c7ra04788f
21. Hu CQ, Li K, Yao TT, Hu YZ, Ying HZ, Dong XW. Integrating docking scores and key interaction profiles to improve the accuracy of molecular docking: towards novel B-Raf(V600E) inhibitors. *Medchemcomm.* (2017) 8:1835–44. doi: 10.1039/c7md00229g
22. Yao TT, Xie JF, Liu XG, Cheng JL, Zhu CY, Zhao JH, et al. Integration of pharmacophore mapping and molecular docking in sequential virtual screening: towards the discovery of novel JAK2 inhibitors. *Rsc Adv.* (2017) 7:10353–60. doi: 10.1039/c6ra24959k
23. Zhan WH, Li DQ, Che JX, Zhang LR, Yang B, Hu YZ, et al. Integrating docking scores, interaction profiles and molecular descriptors to improve the accuracy of molecular docking: toward the discovery of novel Akt1 inhibitors. *Eur J Med Chem.* (2014) 75:11–20. doi: 10.1016/j.ejmech.2014.01.019
24. Dong XW, Yan JY, Du LL, Wu P, Huang SF, Liu T, et al. Pharmacophore identification, docking and "in silico" screening for novel CDK1 inhibitors. *J Mol Graph Model.* (2012) 37:77–86. doi: 10.1016/j.jmgm.2012.04.003
25. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem.* (2012) 55:6582–94. doi: 10.1021/jm300687e
26. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *Acm T Intel Syst Tec.* (2011) 2:1–27. doi: 10.1145/1961189.1961199

27. Wang L, Qiao Q, Ferrao R, Shen C, Hatcher JM, Buhrlage SJ, et al. Crystal structure of human IRAK1. *Proc Natl Acad Sci USA*. (2017) 114:13507–12. doi: 10.1073/pnas.1714386114
28. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings (Reprinted from Advanced Drug Delivery Reviews, vol 23, pg 3-25, 1997). *Adv Drug Deliver Rev*. (2001) 46:3–26. doi: 10.1016/S0169-409x(00)00129-0
29. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem.* (2009) 30:2785–91. doi: 10.1002/jcc.21256
30. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem.* (2011) 32:1466–74. doi: 10.1002/jcc.21707
31. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, et al. *AMBER 2020*. San Francisco, CA: University of California (2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Che, Feng, Gao, Yu, Weng, He, Dong, Wu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fphar.2020.566058

Integrating Machine Learning-Based Virtual Screening With Multiple Protein Structures and Bio-Assay Evaluation for Discovery of Novel GSK3β Inhibitors

Jingyu Zhu[1†], Yuanqing Wu[1†], Man Wang[2†], Kan Li[1], Lei Xu[3], Yun Chen[1], Yanfei Cai[1] and Jian Jin[1*]*

[1] School of Pharmaceutical Sciences, Jiangnan University, Wuxi, China, [2] Jiangsu Institute of Hematology, National Clinical Research Center for Hematologic Diseases, NHC Key Laboratory of Thrombosis and Hemostasis, The First Affiliated Hospital of Soochow University, Suzhou, China, [3] School of Electrical and Information Engineering, Institute of Bioinformatics and Medical Engineering, Jiangsu University of Technology, Changzhou, China

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Ximing Xu,
Ocean University of China, China
Huiyong Sun,
China Pharmaceutical University, China
Danfeng Shi,
Lanzhou University, China

***Correspondence:**
Jingyu Zhu
jingyuzhu@jiangnan.edu.cn
Jian Jin
jianjin@jiangnan.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: 27 May 2020
Accepted: 14 August 2020
Published: 11 September 2020

Citation:
Zhu J, Wu Y, Wang M, Li K, Xu L, Chen Y, Cai Y and Jin J (2020) Integrating Machine Learning-Based Virtual Screening With Multiple Protein Structures and Bio-Assay Evaluation for Discovery of Novel GSK3β Inhibitors. Front. Pharmacol. 11:566058. doi: 10.3389/fphar.2020.566058

Glycogen synthase kinase-3β (GSK3β) is associated with various key biological processes, and it has been considered as a critical therapeutic target for the treatment of many diseases. However, it is a big challenge to develop ATP-competition GSK3β inhibitors because of the high sequence homology with other kinases. In this work, a novel parallel virtual screening strategy based on multiple GSK3β protein structures, integrating molecular docking, complex-based pharmacophore, and naive Bayesian classification, was developed to screen a large chemical database, the 50 compounds with top-scores then underwent a luminescent kinase assay, which led to the discovery of two GSK3β inhibitor hits. The high screening enrichment rate indicates the reliability and practicability of the integrated protocol. Finally, molecular docking and molecular dynamics simulation were employed to investigate the binding modes of the GSK3β inhibitors, and some "hot residues" critical to GSK3β affinity were highlighted. The present study may provide some valuable guidance for the development of novel GSK3β inhibitors.

Keywords: glycogen synthase kinase-3 beta inhibitor, GSK3β, virtual screening, molecular docking, pharmacophore, naive Bayesian classification, molecular dynamics simulation

INTRODUCTION

Glycogen synthase kinase-3 (GSK3) is an evolutionarily very conserved serine/threonine kinase that is ubiquitous in mammalian eukaryotic cells, and it presents a broad spectrum of cellular functions, like cell division, differentiation, transcription, apoptosis, and so on (Frame and Cohen, 2001; Hu et al., 2018). GSK3 contains two functionally distinct isoforms, α and β, and these two isoforms share a 98% sequence identity within the active domain (MacAulay and Woodgett, 2008). Among them, GSK3β has received more attention because of its central role in the regulation of many important metabolic and signaling proteins, structural proteins and transcription factors, that depicts GSK3β as a promising target for the treatment of cancer, neurodegenerative diseases,

neuropsychiatric diseases, and other major diseases (Cohen and Frame, 2001; Osolodkin et al., 2013; O'Leary and Nolan, 2015; Maqbool and Hoda, 2017; Abdul et al., 2018). Therefore, a certain number of GSK3β inhibitors with different kinds have been discovered in last decades (Wagman et al., 2004; Maqbool and Hoda, 2017), and some have been pushed into the clinical trials (Wu et al., 2019). GSK3β inhibitors can be roughly classified into two categories: non-ATP competition and ATP competition inhibitors (Force and Woodgett, 2009; Swinney et al., 2016; Dey et al., 2017; Saura et al., 2017). These two types of inhibitors represent two different focuses on the development of GSK3β inhibitors: affinity and selectivity. non-ATP inhibitors binding to the ATP outside area usually contain higher selectivity but lower affinity to GSK3β, but ATP inhibitors acting to ATP pocket are just the opposite. Nowadays, a large number of ATP competition GSK3β inhibitors have been developed, but no selective GSK3β inhibitor has been FDA approved so far (Sahin et al., 2019). As mentioned above, GSK3β show a 98% sequence identity to GSK3α, besides, GSK3β share a highly conserved sequence in many other kinases, such as cyclin-dependent kinases (CDK) (Bax et al., 2001), that makes it challenging to develop selective GSK3β inhibitors. Therefore, the application of appropriate methods for developing high selective GSK3β inhibitors has become critical and urgent.

As is known to all, traditional lead kinase inhibitors discovery based on experimental systems is an expensive, inefficient and lengthy process because of its screening against a broad panel of diverse kinases. As a counterpart to experimental high-throughput screening, virtual screening (VS) is able to virtually screen large compound databases and has become a good choice for the discovery of novel inhibitors (Hou and Xu, 2004; Amaro and Li, 2010). VS based on molecular docking has received more and more attention (Tanrikulu et al., 2013). For conventional VS, only one protein structure is used in most docking programs, and the protein almost performs the "rigid" structure to maintain optimum efficiency of VS. Thus, integrated strategies for VS in a parallel manner may be the most appropriate to balance the efficiency and precision of VS, and some successful cases in the field of kinase inhibitor development have been reported (Bajorath, 2002; Holliday et al., 2011; Tian et al., 2014; Fan and Huang, 2017; Zhou et al., 2018). Among them, naive Bayesian analysis, a machine learning algorithm, has been widely applied to drug discovery processes such as high-throughput VS, ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties evaluation, and SAR (structure-activity relationship) analysis (Liu, 2004; Chen et al., 2011; Ekins et al., 2019). Bayesian analysis is a well-known statistical algorithm and it could scale linearly with the number of samples comparing with traditional fitting methodologies, thus, Bayesian analysis could be fast and easily automated to process large amounts of data (Rogers et al., 2005; Tian et al., 2013a). Moreover, the Bayesian model focuses on the more important features of samples, and then assigns greater weight to the characteristics to distinguish the "good" samples from the large amounts of sample, that would significantly lead to a higher quality enrichment (Klon et al., 2004; Rogers et al., 2005). In this study, we present an efficient and reliable predictive strategy for parallel Bayesian machine learning-VS integrating molecular docking and complex-based pharmacophore based on multiple GSK3β proteins (the workflow is illustrated in **Figure 1**), 50 potential inhibitors of GSK3β were purchased and some compounds with potent GSK3β inhibitory activity were confirmed by a series of biochemical studies. Finally, the GSK3β binding mechanisms of these inhibitors were well analyzed through molecular docking and molecular dynamics

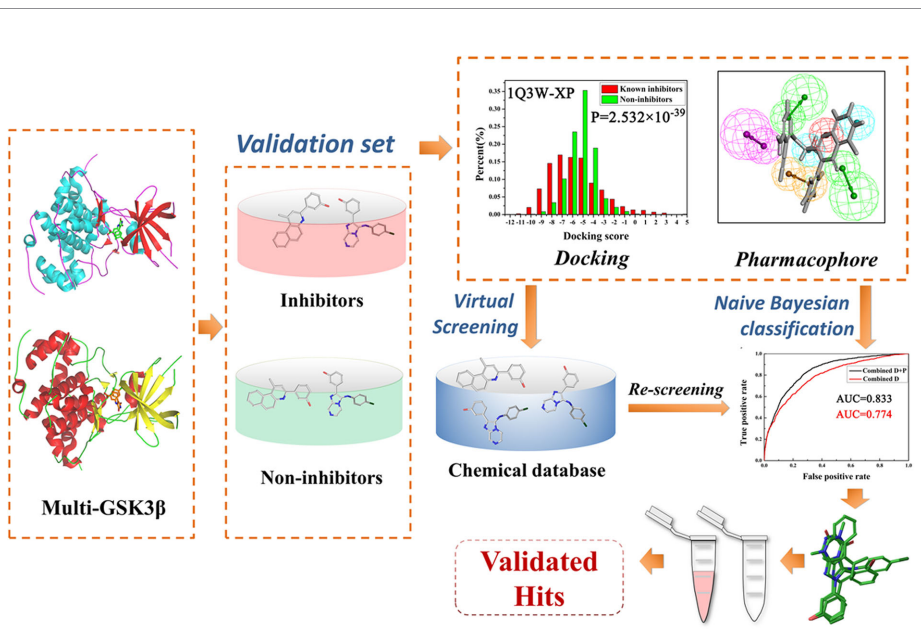


FIGURE 1 | The workflow of this study.

simulation. It could provide some important guidance for the discovery of promising GSK3β inhibitors from a huge chemical database for the treatment of related diseases.

METHODS AND MATERIALS

Preparation of GSK3β Complexes and Validation Dataset

A total of 54 crystallographic structures of GSK3β complexes were retrieved from the RCSB Protein Data Bank (PDB) (Berman et al., 2000). Firstly, the residues within 10 Å of each apo-ligand were reserved and structurally aligned with STAMP algorithm in VMD (**Figure S1**) (Humphrey et al., 1996; Tian et al., 2014). Secondly, the root-mean-square deviation (RMSD) values from the alignment were used to create a phylogenetic tree through the *phylogenetic tree* module of VMD (**Figure S2**). And then, these complexes could be classified eight types according to the differences of the GSK3β structures. Finally, eight representative complexes with the highest resolution within each type were chosen, namely, 1Q3W (Bertrand et al., 2003), 1Q4L (Bertrand et al., 2003), 2OW3 (Zhang et al., 2007), 3L1S (Arnost et al., 2010), 4B7T (Tahtouh et al., 2012), 4J71, 4NM3 (Stamos et al., 2014) and 4PTG (Sivaprakasam et al., 2015) (highlighted in red in **Figure S2**).

In order to evaluate the "screening power" of these eight protein structures, a validation dataset was built, including the known GSK3β inhibitors and non-inhibitors. The non-duplicated GSK3β inhibitors with definite biological activity were obtained from the BindingDB database (Liu et al., 2007), and the non-inhibitors were randomly selected from the ChemDiv database through the *Find Diverse Molecules* module in Discovery Studio 3.5 (DS3.5), and the ratio of non-inhibitors versus inhibitors was set to 1:20. At last, 800 inhibitors and 16,000 non-inhibitors were chosen for further study.

Molecular Docking Procedure

The crystallographic structures of eight GSK3β complexes were employed as initial receptors for VS. Molecular docking simulations were carried on by the *Glide* module in Schrodinger (Friesner et al., 2006). First, each complex was prepared using the *Protein Preparation Wizard* in Schrodinger, to remove crystallographic water molecules, add hydrogen atoms, assign protonated states, assign partial charges. Afterward, each complex was minimized with the OPLS-2005 force field until the RMSD reached a maximum value of 0.3 Å. The compounds in the validation set were processed through the *LigPrep* protocol in Schrodinger with the default parameters set. Finally, a bounding box of size 10 × 10 × 10 Å was generated with the co-crystallized ligand as centroid using the *Receptor Grid Generation* module for each system.

Subsequently, the "scoring power" was evaluated to estimate the docking precision of these eight systems. The crystallized ligand was first extracted from each GSK3β complex, and then re-docked into the corresponding binding site, the RMSD between the docking pose and crystallographic conformation of the inhibitors was calculated. On the other hand, the "screening power" was also investigated. All molecules in the validation dataset were docked into the binding site of each protein with the Standard Precision (SP) or Extra Precision (XP) scoring modes. The student 's t-test was used to measure the difference in the distribution of docking scores between active inhibitor and non-inhibitors.

Pharmacophore Procedure

The eight crystal GSK3β structures were also used as the initial receptors for complex-based pharmacophore models. First, the proteins were prepared through the *prepare protein* module in DS3.5 to add polar hydrogen, remove water molecules, repair the broken chain, and add the *CHARMm* field with the default parameters. Second, all molecules in the validation set were prepared through the *prepare ligands* module in DS3.5. Then, the *Receptor-Ligand Pharmacophore Generation* (RLPG) module was employed to generate the complex-based pharmacophore models for eight GSK3β complexes (Hou et al., 2012). A set of pharmacophoric features were identified from each system (Sutter et al., 2011), including hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (HYD), negative ionizable (NI), positive ionizable (PI), and ring aromatic (RA) features of ligands. And the parameters were kept as the default setting except for the minimum number of the pharmacophore features was set to 3.

The selectivity of the pharmacophore model is evaluated through genetic function approximation (GFA) scoring function (Rogers and Hopfinger, 1994; Meslamani et al., 2012). For each complex structure, the generated models are ranked according to their selectivity scores evaluated by the GFA model, and thus 10 top-ranked pharmacophore models would be produced, and feature(S) was defined as the pharmacophore features with the highest selectivity. Besides, in order to investigate the capability of each model to distinguish between inhibitors and non-inhibitors, a validation set containing 800 active GSK3β inhibitors and 16,000 non-inhibitors mentioned above was constructed. The pharmacophore features of the pharmacophore model with the best discrimination between inhibitors and non-inhibitors were set to feature(D), and the area under receiver operating characteristic (ROC) curve (AUC) was calculated to evaluate this discrimination capability. Moreover, the molecules in the validation set were all mapped onto the pharmacophore models to generate the fit values.

Naive Bayesian Classification

Consequently, naive Bayesian classification (NBC) based on multiple protein structures was employed to evaluate the screening accuracy. This machine learning approach has been proved to significantly increase the hit rate of virtual screening (Tian et al., 2013b; Li et al., 2014). First, the data matrix consists of above docking scores (from molecular docking) and fit values (from pharmacophore) were used as the independent variables (X); and 1/0 was used as the response variable (Y), in which 1 presents inhibitor and 0 means non-inhibitor. Then, NBC was developed using the *Create Bayesian Model* module in DS3.5 (Hou et al., 2012) to distinguish the inhibitors from

non-inhibitors. The prediction precision of each classifier was evaluated by the AUC value for discriminating the GSK3β inhibitors from non-inhibitors. Finally, the best Bayesian classifier was utilized to re-score the compounds.

Integrated Virtual Screening

The docking-VS was performed to screen the ChemDiv database (approximately 2,000,000 compounds). First, each compound was docked into the binding pockets of eight GSK3β with corresponding scoring modes. Then, 10,000 compounds with the best docking scores were obtained and then mapped onto the pharmacophore models generated from the GSK3β complexes. Afterward, 2,000 compounds with the best fit values were chosen and then to re-screening with the best Bayesian classifier. Last, the 100 compounds with top-ranked by the Bayesian scores were assessed the drug-likeness by *Filter by Lipinski and Veber Rules* module in DS3.5. Finally, 50 compounds were purchased from ChemDiv for subsequent GSK3β inhibitory assay *in vitro*.

Kinase Assay

GSK3β kinase activity was measured by ADP-Glo Kinase Assay system from Promega (Product code V1991) (Auld et al., 2009; Zegzouti et al., 2009; Cho et al., 2011), which is a non-radioisotope, homogeneous, ADP quantitative kit based on a luminescent kinase assay. Compounds were initially tested in triplicate at 20 μM, and then, those molecules which contain an inhibitory activity greater than 50% were chosen to further test in a nine-point dose curve with 2-fold serial dilution starting from 40 μM (three times). The assay protocol was listed as below: GSK3β protein was firstly incubated with 1μg/ml of the peptide substrate, second, the test compounds were added into the solution for 60 min at 25°C with a GSK3β reaction buffer, which contained 40 mM Tris (pH = 7.5), 20 mM $MgCl_2$, 0.1mg/ml BSA, 50 μM DTT, and 50 μM ATP. The assays were done in 384-well white plates with a total reaction volume of 5 μl per well. Then, the reactions were terminated through the introduction of 5 μl ADP-Glo reagent assay, and the assay plate was incubated for 40 min. Finally, 10 μl of kinase detection buffer was added to convert ADP to ATP and introduce luciferase and luciferin to detect ATP.

Promiscuity Assessment

The promiscuity of Cpd49 was evaluated by an online program, Badapple (http://pasilla.health.unm.edu/tomcat/badapple/badapple). the simplified molecular input line entry specification (SMILES) formats of Cpd49 was entered into the input box and the results would be generated automatically.

Molecular Docking and MD

The 3D structure of compound22 (Cpd22) and Cpd49 were sketched using maestro, LY2090314 was retrieved from the PubChem Compound database. All these inhibitors were prepared by the *Ligprep* module with the OPLS-2005 force field in Schrodinger. The crystal structure of GSK3β with the highest resolution, 1Q4L, and the crystal structure of GSK3α, 2DFM (Wang et al., 2019), were prepared as the initial receptors for the molecular docking through the *Protein Preparation Wizard* module. The grid file of the GSK3β protein was generated following the step in *Chapter 2.2*, and that of GSK3α was generated with the specified residues, Val101, Pro102, and Arg107 (hinge region), as centroids. Then Cpd22, Cpd49, and LY2090314 were respectively docked into the GSK3β protein, and Cpd49 was docked into the GSK3α protein using the *Glide* module (XP mode), and the best-docked structure of Cpd49 in the GSK3β-ligand system was chosen for the following MD simulation.

The Cpd49 with the best bioactivity was conducted to MD with *Desmond* package in Schrodinger (Chapon, 2014), and the docked GSK3β/Cpd49 complex was chosen as the initial system. All amino acid interactions in the protein were modeled with the OPLS-2005 force field. The simulation system was solvated in a 10 Å orthorhombic box with periodic boundary condition, and built with the 3-Point Transferable Inter-molecular Potential (TIP3P) water model. Afterward, the whole system was neutralized with salt concentrations of 0.15 M of Na^+ ions. Before MD simulation, the system was minimized with OPLS-2005 force field. And then, the 200-ns MD was performed using the NPT ensemble under a target temperature of 310 K and a target pressure of 1 atm, the energy and atomic coordinate trajectory recording interval were set to 20 ps. Thereafter, the RMSD and protein-ligand contacts were all calculated with the *Simulation Interaction Diagram* protocol in *Desmond* package.

RESULTS AND DISCUSSION

Validation of Docking-Based Virtual Screening

To investigate the prediction capability and reliability of molecular docking, the screening ability, including "scoring power" and "screening power", was evaluated. Scoring power shows the reliability of the docking program, namely, whether it can predict the real binding conformation between inhibitors and proteins (Shen et al., 2020). Herein, the RMSD between docking pose and crystallographic structures was calculated to reveal this "power" and RMSD ≤ 2.0 Å was used as a criterion. Firstly, the original ligands of eight GSK3β complexes were extracted and then re-docked into the corresponding binding pocket of the protein. The RMSD values were calculated and the values were summarized in **Table 1**. Generally, the docking program, whether SP mode or XP mode, basically reproduces the experimental conformation (all RMSD values ≤ 2.0 Å), indicating the eight chosen protein structures all satisfy the docking accuracy with *Glide* module.

Next, the screening power was estimated using the Student's t-test (P-value), which was conducted to assess the different distributions between the inhibitors and non-inhibitors under SP or XP score modes (Shen et al., 2020). As shown in **Table 1** and **Figure 2**, the molecular docking can effectively distinguish the inhibitors form non-inhibitors, the P-values of eight complexes all far lower than 0.05, except the SP mode of 4B7T. However, discrimination for certain structures still exhibits significant different sampling powers (**Figure 2**). For instance, the P-value of the SP mode for 2OW3 is 2.958×10^{-69}, while that for 4B7T is only

TABLE 1 | Scoring power and screening power of the eight GSK3β protein structures.

Complex	P-value		RMSD (Å)	
	SP	XP	SP	XP
1Q3W	2.079×10^{-23}	**2.532×10^{-39}**	0.72	0.58
1Q4L	0.016	**6.899×10^{-36}**	0.88	0.43
2OW3	2.958×10^{-69}	**1.841×10^{-76}**	1.25	0.93
3L1S	3.348×10^{-18}	**6.093×10^{-72}**	0.69	0.58
4B7T	0.172	**1.388×10^{-33}**	1.45	1.06
4J71	0.019	**3.203×10^{-22}**	0.62	0.43
4NM3	**1.502×10^{-41}**	0.040	1.33	1.14
4PTG	3.026×10^{-20}	**1.439×10^{-30}**	1.65	1.34

Bolded data: better discrimination.

0.172 (**Table 1**). In general, the XP scoring mode exhibited more accurate determination than the SP scoring mode (**Figure 2**). Therefore, it is necessary to choose an appropriate docking mode for a specific protein structure to ensure docking prediction accuracy. In summary, our molecular docking model based on the eight GSK3β structures with *Glide* could fulfill the requirement of satisfactory docking. Thus, the docking-based VS with the multiple structures would be a reliable tool for the development of potential GSK3β inhibitors.

Validation of Pharmacophore-Based Virtual Screening

Eight complexes of GSK3β were employed to generate the pharmacophore models, and the models were ranked through the selective scores detected by the GFA algorithm. The results were tabulated in **Table 2**. The prediction capacity of each pharmacophore model to distinguish the inhibitors and non-inhibitors was evaluated with the AUC values, and it indicates that the models contain satisfactory prediction power when AUC > 0.7 (Tian et al., 2013b). According to **Table 2**, feature (S) and feature(D) were analyzed and compared.

Generally speaking, a pharmacophore model with a higher selectivity score shows a stronger prediction power. But as shown in **Table 2**, we found that most models with the high selectivity score contain a poor discrimination power, such as 4NM3 (AUC < 0.7). Some models, especially the feature(D) of 2OW3, with a low selective score, show a fair inhibitor selectivity on the contrary. It seems that the pharmacophore models using feature (D) are more suitable for our VS protocol. Therefore, to simultaneously avoid a high rate of false positives and find more active GSK3β inhibitors, 1Q4L and 2OW3 models with feature(D) were chosen to perform pharmacophore-based VS (AUC > 0.7, **Table 2**).

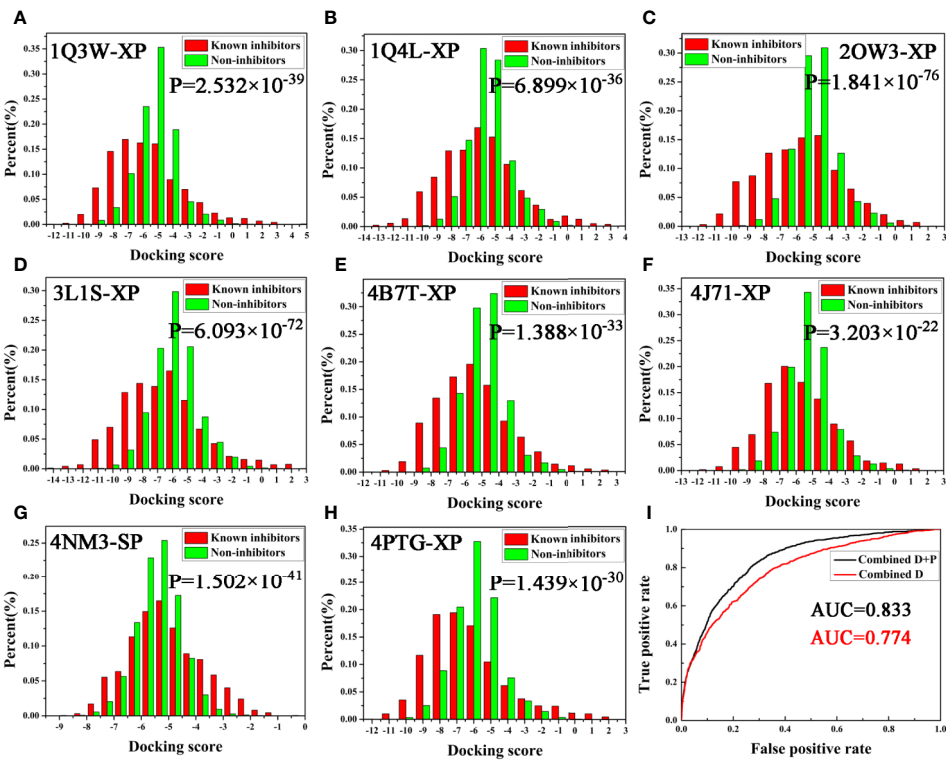


FIGURE 2 | (A–H) Eight GSK3β complexes have better ability to distinguish known inhibitors and non-inhibitors under SP or XP docking accuracy; **(I)** AUC value of Bayesian model constructed based on docking score or integrating docking scores and pharmacophore fit values.

TABLE 2 | The selectivity scores and AUC values of feature(S) and feature(D).

ID	Feature(S)[a]	Selectivity score	AUC	Feature(D)[a]	Selectivity score	AUC
1Q4L	AADHHN	11.701	0.612	ADHN	9.3532	**0.711**
4NM3	AADNNN	14.964	0.502	AADNNN	14.964	0.502
3L1S	AADHHH	10.447	0.595	AADHH	8.9325	0.657
2OW3	ADHHH	8.5895	0.604	DHH	5.5599	**0.709**
4B7T	AAHH	5.7496	0.514	AHH	4.2348	0.567
4PTG	ADH	5.3541	0.544	AHH	4.4406	0.564
4J71	ADH	4.8740	0.601	ADH	4.8740	0.601
1Q3W	AAD	5.0798	0.694	AAD	5.0798	0.694

[a]*A, hydrogen-bond receptor; D, hydrogen-bond donor; H, hydrophobic; N, negative charge.*
Bolded data: better discrimination.

Naive Bayesian Classifiers With Multiple GSK3β Structures

As described above, two protocols based on molecular docking and pharmacophore were respectively optimized for the following VS. In order to further improve the hit rate of screening, a parallel virtual screening based on multiple structures integrating molecular docking and pharmacophore was built and the prediction capabilities were evaluated by NBC model. The flow diagram of the model generation process is revealed in **Figure 1**. The data matrix consists of all the docking scores and fit values of the compounds in the validation dataset for each complex. According to the above results, the docking mode with a higher P-value and/or pharmacophore model with a fair AUC value should be selected as the independent variables to build the NBC models (bolded in **Tables 1** and **2**). Firstly, the prediction power based on the chosen docking score was estimated and the AUC = 0.774 (**Figure 2I**), this value was slightly better than that of the above pharmacophore. But when the model combined the docking scores and fit values, the accuracy was significantly improved, the classifier achieved an AUC of 0.833, indicating that the VS integrating a machine learning based on multi-conformational GSK3β could effectively highlight the active GSK3β inhibitors.

GSK3β Inhibition Assay

To determine the GSK3β inhibitory activity of those hit molecules from VS, we performed GSK3β inhibition assays based on ADP-Glo method, in which LY2090314 (Atkinson et al., 2015; Gray et al., 2015; Kunnimalaiyaan et al., 2018) was used as the reference compound. The 50 compounds were tested at the initial concentration of 20 μM. The results of the GSK3β inhibitory activities of these hits are shown in **Figure 3A**. Among the tested compounds, Cpd22 showed effective inhibiting GSK3β activity 57.6% and Cpd49 inhibiting GSK3β activity 58.5% at the concentration of 20 μM, respectively. Afterward, the GSK3β inhibition potency (IC_{50}) of Cpd22 and Cpd49 was identified,

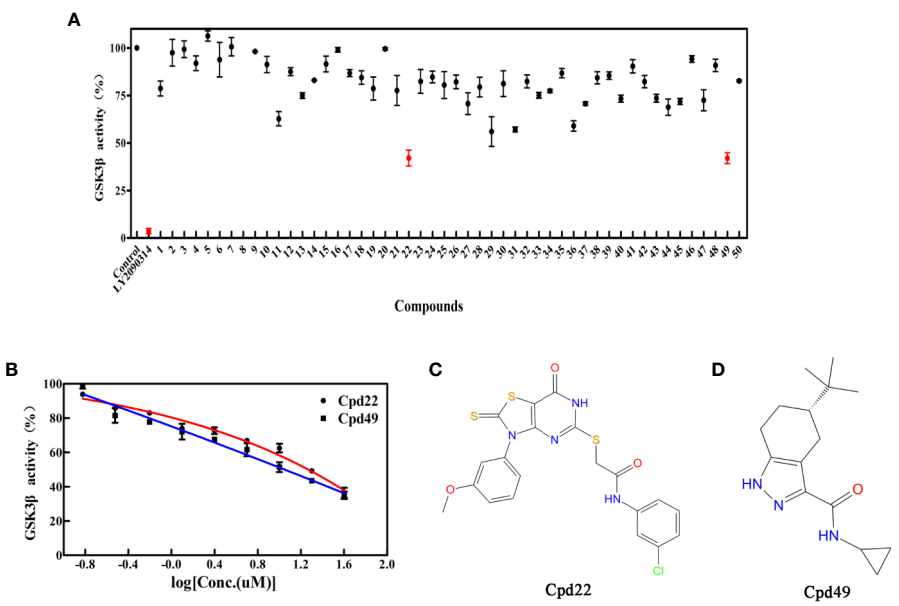
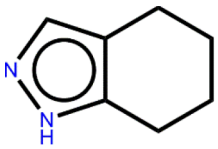
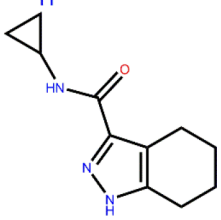


FIGURE 3 | (A) Preliminary screening results of 50 compounds with enzyme experiments; **(B)** Cpd22 and Cpd49 concentration gradient results; **(C, D)** 2D structure diagram of Cpd22 and Cpd49.

TABLE 3 | The pScore and inDrug values from Badapple prediction.

Compound	Scaffold Number	Scaffold Structure	pScore[a]	inDrug[b]
Cpd49	1		185	True
	2		83	False
	3		None	False

[a]*pScore values, <100 (low), no indication; 100–300 (moderate), weak indication of promiscuity; >300 (high), strong indication of promiscuity.*
[b]*inDrug, True, means it was found in the drug data base; False, means not found.*

both compounds could potentially inhibit the activity of GSK3β, with IC_{50} of 15.49 μM and 9.34 μM, respectively (**Figure 3B**). The 2D structures of the two inhibitors are illustrated in **Figures 3C, D**, and both compounds present new scaffolds.

PAINS Assessment for Cpd49

Cpd49 with the better bioactivity was then investigated whether it could be categorized as a PAINS compound. PAINS refer to a series of promiscuous compounds specifically binding to different macromolecular targets and leading to misleading false-positive results in experimental assays (Baell and Holloway, 2010). The assessment results from Badapple prediction were shown in **Table 3**. The cyclopropane of Cpd49 exhibits a moderate pScore value with 185 and shows a "True" inDrug result. While the tetrahydroindazole group presents a low pScore value with a "False" result inDrug database. Besides, the scaffold3 generated no pScore and the inDrug result is "False". These results indicate Cpd49 would be a GSK3β selective inhibitor.

Molecular Docking Analysis

To elucidate the binding mode of inhibition of GSK3β by Cpd22 and Cpd49, the two inhibitors were docked to the crystal structure 1Q4L. The docking scores were −9.032 kcal/mol for GSK3β/Cpd22, and −9.566 kcal/mol for GSK3β/Cpd49. The interactions between the inhibitors and GSK3β have respectively illustrated in **Figure 4**. The docking results indicated that Cpd22 and Cpd49 interact with GSK3β in the same position, in which residues Ile62, Val70, Ala83, Val110, Asp133, Tyr134, Val135, Thr138, Arg141, Leu188, Cys199, and Asp200 constructed the binding pocket and interacted with the inhibitors. Meanwhile, hydrophobic interactions were formed with residues Ile62, Val70, Ala83,

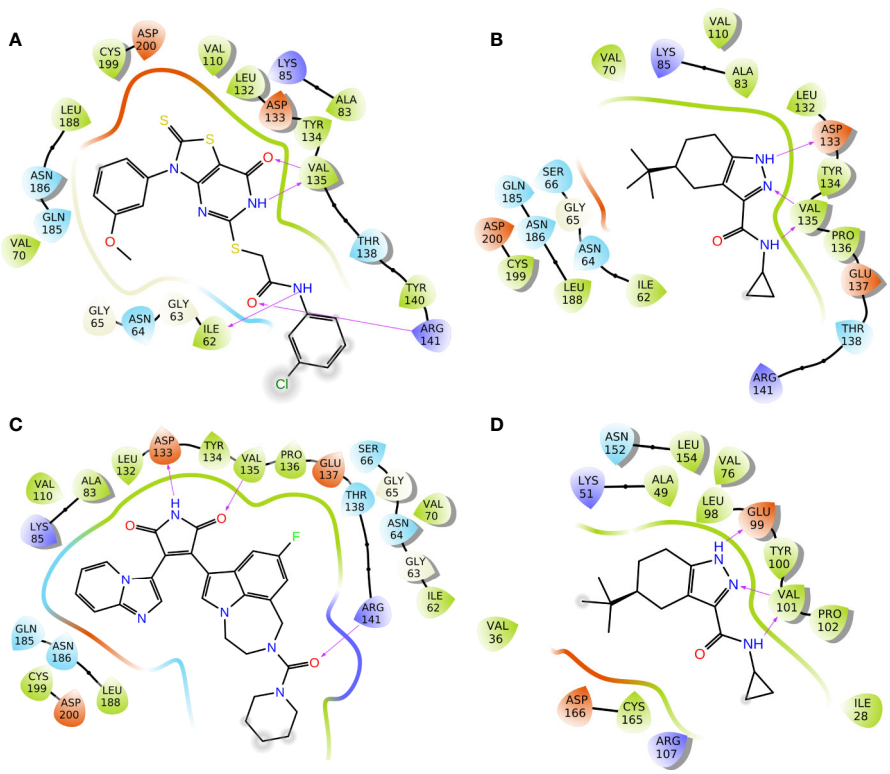


FIGURE 4 | 2D presentations of the interactions between GSK3β and **(A)** Cpd22; **(B)** Cpd49; **(C)** LY2090314; **(D)** 2D presentations of the interactions between GSK3α and Cpd49 (H-bonds colored in magenta).

Val110, Tyr134, Val135, Leu188, and Cys199 (shown in black lines in **Figures 4A, B**), which was consistent with the analytical results from our previous research. These residues are important to stabilize the binding affinity between the inhibitors and GSK3β (Zhu et al., 2019). In addition, there are several key hydrogen bonds (H-bond) were formed in both complexes. **Figure 4A** showed that Cpd22 formed four H-bonds with Ile62, Val135, and Arg141. For Cpd49, Val135 formed two H-bonds with the pyrazole and nitro group, and Asp133 also formed an H-bond. These H-bond interactions could maintain the specific binding interactions between the ligand and GSK3β (Zhao et al., 2017). Herein, LY2090314, as a positive control, was also docked into the GSK3β protein, and the result was illustrated in **Figure 4C**. Similar to Cpd49, LY2090314 formed the two H-bond interactions with Asp133 and Val135 of GSK3β (**Figure 4B**). Moreover, the carbonyl group of LY2090314 formed an H-bond with Arg141, similarly, the Arg141 of GSK3β also form H-bond with the carbonyl group of Cpd22 (**Figure 4A**). That may be the reason why LY2090314 contains more potent bioactivity than Cpd22 and Cpd49, because of the more H-bond interactions with GSK3β. Finally, Cpd49 was chosen to investigate the interaction with GSK3α, **Figure 4D** shows that the binding pose of Cpd49 in the binding site of GSK3α is almost same with that in GSK3β, it is not surprised because of the high sequence identity between GSK3α and β. Although Cpd49 formed three H-bonds with GSK3α, the docking score, -6.783 kcal/mol, was far lower than GSK3β, indicating that Cpd49 preferred to bind with GSK3β.

Molecular Dynamic Simulations Analysis

To explore the dynamic binding process of the studied inhibitor, Cpd49 with the best bioactivity was submitted to 200-ns MD. Firstly, the backbone RMSD of the complex was calculated and the result illustrated in **Figure S3**. The plot showed that the system reached equilibrium after 200 ns simulation. In order to understand the mechanism of ligand adaptation in binding space of GSK3β, the per-residue contributions in ligand binding were calculated and the key residues-inhibitor interactions were rendered in **Figure 5A**. As the appears at first glance, several residues formed the strong hydrophobic interactions with Cpd49, including Ile62, Ala83, Tyr134, and Leu188, these residues could form the strong non-polar interactions with GSK3β selective inhibitors, which make the dominating force for the high GSK3β affinity (Zhu et al., 2019), meanwhile, Asp133 and Val135 form strong H-bonds with Cpd49, these hydrogen bond interactions are significantly important for the specific GSK3β binding of the selective inhibitors (**Figure 5B**), which is consistent with the docking analysis discussed above. However, Val135 exhibits more favorable contributions to Cpd49 than Asp133. As shown in **Figure 5C**, the pyrazole group of Cpd49 was closer to Val135 and then formed two H-bonds with Val135. it can be perceived Val135 is a critical residue for selective binding to GSK3β. As shown in **Figure 5C**, Cpd49 was sandwiched between residues Ala83 and Leu188, and encompassed by residues Ala83 above and residues Leu188 below, that formed strong non-polar interactions which highly improved the binding affinity to GSK3β for Cpd49. Overall, the above results prove that Asp133, Val135, Ala83, and Leu188 may be the key residues for inhibitor binding.

Moreover, four X-ray crystallographic structures of GSK3β complexes, including AMP-PNP (PDB ID 1PYX) (Bertrand et al., 2003) and three inhibitors under clinical investigation (Xu et al., 2019), Alsterpaullone (PDB ID 1Q3W) (Bertrand et al., 2003), AZD2858 (PDB ID 4ACD) (Berg et al., 2012), and CHIR-99021 (PDB ID 5HLN) (Wagner et al., 2016), were chosen to compare to the binding configuration of Cpd49/GSK3β complex. The crystal structure of GSK3β/AMP-PNP was illustrated in **Figure 6A**. The adenine of AMP-PNP forms two H-bonds with Asp133 and Val135, respectively. Asp133 and Val135 are two key hinge residues, which plays a critical role in the GSK3β specific binding for a selective inhibitor (Pandey and DeGrado, 2016). As shown in **Figure 5C**, the tetrahydroindazole group of Cpd49 could mimic the adenine of ATP to form the key H-bonds with Asp133 and Val135. Similar H-bond interactions occur in AZD2858 and CHIR-99021 systems (**Figures 6B, C**). Besides, the oxygen atom of the ribose could hydrogen bond to Gln185, and the phosphate group of AMP-PNP could form the H-bonds with Lys85 and Lys183, respectively (**Figure 6A**) (Bertrand et al., 2003). The H-bond with Lys85 is also formed in AZD2858 system, which is caused by the internal hydrogen bond on the pyrazine ring (**Figure 6B**) (Berg et al., 2012), while these interactions all lose in Cpd49/GSK3β system, which may result in lower bioactivity of Cpd49. On the other hand, the hydrophobic interaction is important for the GSK3β affinity. In AMP-PNP system, these residues consist of Ile62, Val70, Ala83, Val110, Leu132, Tyr134, and Leu188 (Bertrand et al., 2003), and Cpd49 could form the similar non-polar interactions, especially

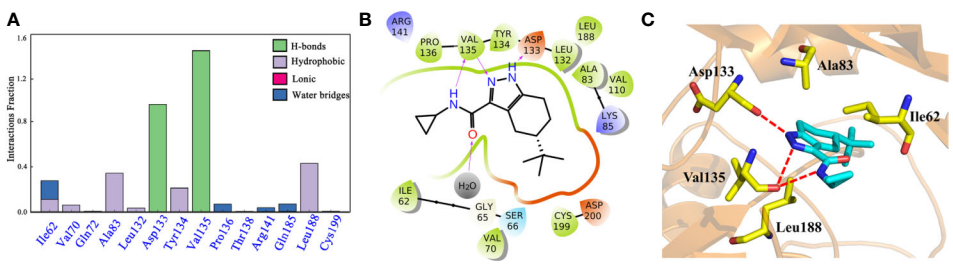


FIGURE 5 | (A) the Cpd49-GSK3β residues interaction spectrum; **(B)** 2D presentations of the MD simulation interactions between GSK3β and Cpd49; **(C)** 3D presentations of the MD simulation interactions between GSK3β and Cpd49.

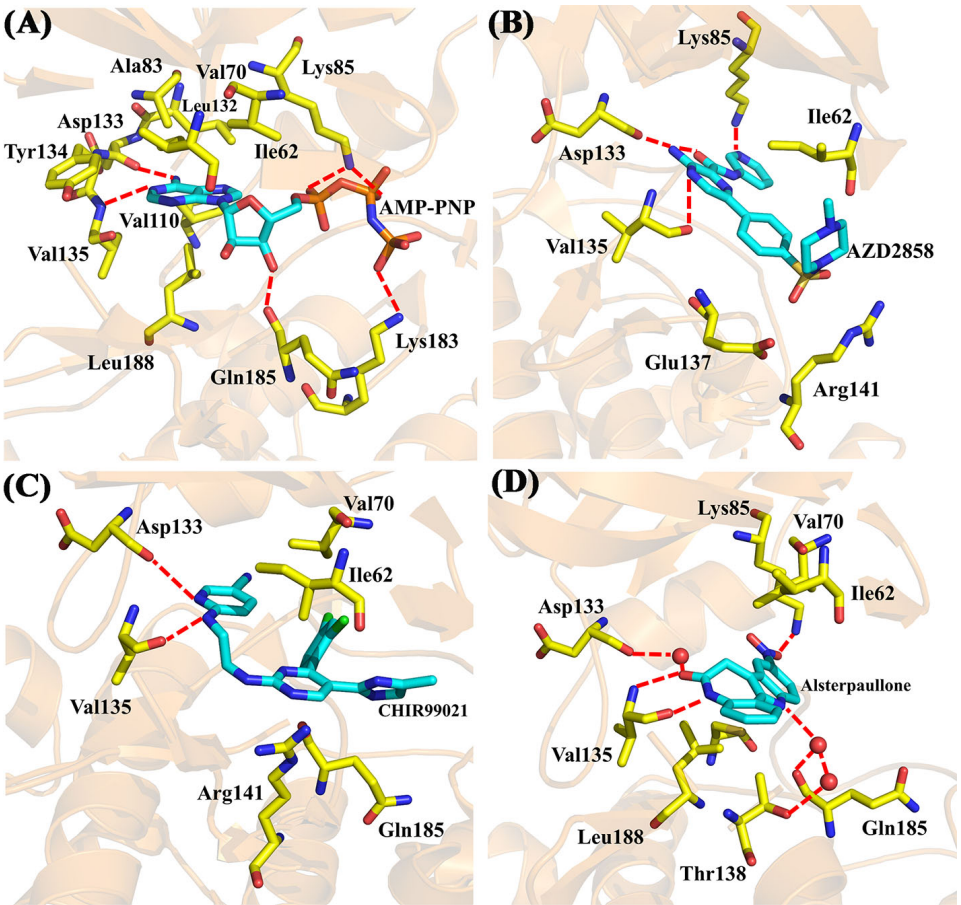


FIGURE 6 | X-ray crystal structures of **(A)** GSK3β/AMP-PNP (PDB ID 1PYX); **(B)** GSK3β/AZD2858 (PDB ID 4ACD); **(C)** GSK3β/CHIR-99021 (PDB ID 5HLN); **(D)** GSK3β/Alsterpaullone (PDB ID 1Q3W).

with Ile62, Ala83, Tyr134, and Leu188 (**Figure 5A**). The binding mode of Alsterpaullone/GSK3β shows in **Figure 6D**. First, Alsterpaullone forms two H-bonds with Val135, but unlike Cpd49 system, Alsterpaullone forms a water bridge with Asp133, and the other two water bridges also form between Alsterpaullone and Thr138/Gln185 (Bertrand et al., 2003). Similar to AMP-PNP and AZD2858, Alsterpaullone also could form a strong H-bond with Lys85 (**Figure 6D**). Second, Alsterpaullone also forms strong van der Waals interactions with GSK3β, especially Ile62 and Val70, so is in the CHIR-99021 system (**Figure 6C**) (Wagner et al., 2016), while Cpd49 loses that interaction with Val70 (**Figure 5A**). Thus, all these extra interactions discussed above may lead to the higher GSK3β binding affinity for Alsterpaullone than Cpd49

CONCLUSION

In the present study, a parallel VS strategy based on multiple GSK3β protein structures was developed to screen against a large chemical library, in which the NBC model combining molecular docking and pharmacophore show a reliable prediction capability. After a series of biochemical studies, two GSK3β inhibitor hits (Cpd22 and Cpd49) were identified from 50 virtual screened compounds, that highlights the high prediction accuracy and the robust reliability of the integrated machine learning-based VS. Besides, the binding modes between GSK3β and two inhibitors were identified by molecular docking, and some key residues critical to GSK3β selectivity were highlighted through MD simulation. We hope that this study would provide some guidance for the virtual screening or design of novel GSK3β inhibitors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

JZ and JJ developed the study concept and design. YW and KL performed the modeling studies and carried out the data analysis.

MW performed the bio-assay evaluation. LX provided software. JZ, YW, and YaC drafted the manuscript. YuC and JJ approved the manuscript.

FUNDING

The study was supported by the National Natural Science Foundation of China (No. 21807049 and 81803430), the Fundamental Research Funds for the Central Universities (JUSRP51703A), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19_1888), and the Natural Science Foundation of Jiangsu Province (No. BE2019650).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.566058/full#supplementary-material

REFERENCES

Abdul, A., De Silva, B., and Gary, R. K. (2018). The GSK3 kinase inhibitor lithium produces unexpected hyperphosphorylation of beta-catenin, a GSK3 substrate, in human glioblastoma cells. *Biol. Open* 7 (1), bio030874. doi: 10.1242/bio.030874

Amaro, R. E., and Li, W. W. (2010). Emerging methods for ensemble-based virtual screening. *Curr. Top. Med. Chem.* 10 (1), 3–13. doi: 10.2174/156802610790232279

Arnost, M., Pierce, A., ter Haar, E., Lauffer, D., Madden, J., Tanner, K., et al. (2010). 3-Aryl-4-(arylhydrazono)-1H-pyrazol-5-ones: Highly ligand efficient and potent inhibitors of GSK3beta. *Bioorg. Med. Chem. Lett.* 20 (5), 1661–1664. doi: 10.1016/j.bmcl.2010.01.072

Atkinson, J. M., Rank, K. B., Zeng, Y., Capen, A., Yadav, V., Manro, J. R., et al. (2015). Activating the Wnt/beta-Catenin Pathway for the Treatment of Melanoma–Application of LY2090314, a Novel Selective Inhibitor of Glycogen Synthase Kinase-3. *PloS One* 10 (4), e0125028. doi: 10.1371/journal.pone.0125028

Auld, D. S., Zhang, Y. Q., Southall, N. T., Rai, G., Landsman, M., MacLure, J., et al. (2009). A Basis for Reduced Chemical Library Inhibition of Firefly Luciferase Obtained from Directed Evolution. *J. Med. Chem.* 52 (5), 1450–1458. doi: 10.1021/jm8014525

Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53 (7), 2719–2740. doi: 10.1021/jm901137j

Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* 1 (11), 882–894. doi: 10.1038/nrd941

Bax, B., Carter, P. S., Lewis, C., Guy, A. R., Bridges, A., Tanner, R., et al. (2001). The structure of phosphorylated GSK-3beta complexed with a peptide, FRATtide, that inhibits beta-catenin phosphorylation. *Structure* 9 (12), 1143–1152. doi: 10.1016/s0969-2126(01)00679-7

Berg, S., Bergh, M., Hellberg, S., Hogdin, K., Lo-Alfredsson, Y., Soderman, P., et al. (2012). Discovery of novel potent and highly selective glycogen synthase kinase-3beta (GSK3beta) inhibitors for Alzheimer's disease: design, synthesis, and characterization of pyrazines. *J. Med. Chem.* 55 (21), 9107–9119. doi: 10.1021/jm201724m

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28 (1), 235–242. doi: 10.1093/nar/28.1.235

Bertrand, J. A., Thieffine, S., Vulpetti, A., Cristiani, C., Valsasina, B., Knapp, S., et al. (2003). Structural characterization of the GSK-3beta active site using selective and non-selective ATP-mimetic inhibitors. *J. Mol. Biol.* 333 (2), 393–407. doi: 10.1016/j.jmb.2003.08.031

Chapon, F. (2014). Interlacing property of zeros of eigenvectors of Schrodinger operators on trees. *J. Phys. A-Math Theor.* 47 (46), 465201. doi: 10.1088/1751-8113/47/46/465201

Chen, L., Li, Y., Zhao, Q., Peng, H., and Hou, T. (2011). ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Mol. Pharm.* 8 (3), 889–900. doi: 10.1021/mp100465q

Cho, G. W., Noh, M. Y., Kang, B. Y., Ku, I. W., Park, J., Hong, Y. H., et al. (2011). Development of a Cellular Tau Enzyme-Linked Immunosorbent Assay Method for Screening GSK-3 beta Inhibitors. *Assay Drug Dev. Technol.* 9 (5), 503–513. doi: 10.1089/adt.2010.0343

Cohen, P., and Frame, S. (2001). The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* 2 (10), 769–776. doi: 10.1038/35096075

Dey, A., Hao, S., Wosiski-Kuhn, M., and Stranahan, A. M. (2017). Glucocorticoid-mediated activation of GSK3beta promotes tau phosphorylation and impairs memory in type 2 diabetes. *Neurobiol. Aging* 57, 75–83. doi: 10.1016/j.neurobiolaging.2017.05.010

Ekins, S., Puhl, A. C., Zorn, K. M., Lane, T. R., Russo, D. P., Klein, J. J., et al. (2019). Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* 18 (5), 435–441. doi: 10.1038/s41563-019-0338-z

Fan, C., and Huang, Y. (2017). Identification of novel potential scaffold for class I HDACs inhibition: An in-silico protocol based on virtual screening, molecular dynamics, mathematical analysis and machine learning. *Biochem. Biophys. Res. Commun.* 491 (3), 800–806. doi: 10.1016/j.bbrc.2017.07.051

Force, T., and Woodgett, J. R. (2009). Unique and overlapping functions of GSK-3 isoforms in cell differentiation and proliferation and cardiovascular development. *J. Biol. Chem.* 284 (15), 9643–9647. doi: 10.1074/jbc.R800077200

Frame, S., and Cohen, P. (2001). GSK3 takes centre stage more than 20 years after its discovery. *Biochem. J.* 359 (Pt 1), 1–16. doi: 10.1042/0264-6021:3590001

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49 (21), 6177–6196. doi: 10.1021/jm051256o

Gray, J. E., Infante, J. R., Brail, L. H., Simon, G. R., Cooksey, J. F., Jones, S. F., et al. (2015). A first-in-human phase I dose-escalation, pharmacokinetic, and pharmacodynamic evaluation of intravenous LY2090314, a glycogen synthase kinase 3 inhibitor, administered in combination with pemetrexed and carboplatin. *Invest. New Drugs* 33 (6), 1187–1196. doi: 10.1007/s10637-015-0278-7

Holliday, J. D., Kanoulas, E., Malim, N., and Willett, P. (2011). Multiple search methods for similarity-based virtual screening: analysis of search overlap and precision. *J. Cheminform.* 3 (1), 29. doi: 10.1186/1758-2946-3-29

Hou, T., and Xu, X. (2004). Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* 10 (9), 1011–1033. doi: 10.2174/1381612043452721

Hou, T., Li, N., Li, Y., and Wang, W. (2012). Characterization of domain-peptide interaction interface: prediction of SH3 domain-mediated protein-protein interaction network in yeast by generic structure-based models. *J. Proteome Res.* 11 (5), 2982–2995. doi: 10.1021/pr3000688

Hu, X. L., Guo, C., Hou, J. Q., Feng, J. H., Zhang, X. Q., Xiong, F., et al. (2018). Stereoisomers of Schisandrin B Are Potent ATP Competitive GSK-3beta Inhibitors with Neuroprotective Effects against Alzheimer's Disease: Stereochemistry and Biological Activity. *ACS Chem. Neurosci.* 10 (2), 996–1007. doi: 10.1021/acschemneuro.8b00252

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14 (1), 33–38, 27-38. doi: 10.1016/0263-7855(96)00018-5

Klon, A. E., Glick, M., and Davies, J. W. (2004). Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J. Med. Chem.* 47 (18), 4356–4359. doi: 10.1021/jm049970d

Kunnimalaiyaan, S., Schwartz, V. K., Jackson, I. A., Clark Gamblin, T., and Kunnimalaiyaan, M. (2018). Antiproliferative and apoptotic effect of

LY2090314, a GSK-3 inhibitor, in neuroblastoma in vitro. *BMC Cancer* 18 (1), 560. doi: 10.1186/s12885-018-4474-7

Li, D., Chen, L., Li, Y., Tian, S., Sun, H., and Hou, T. (2014). ADMET evaluation in drug discovery. 13. Development of in silico prediction models for P-glycoprotein substrates. *Mol. Pharm.* 11 (3), 716–726. doi: 10.1021/mp400450m

Liu, T. Q., Lin, Y. M., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, D198–D201. doi: 10.1093/nar/gkl999

Liu, Y. (2004). A comparative study on feature selection methods for drug discovery. *J. Chem. Inf. Comput. Sci.* 44 (5), 1823–1828. doi: 10.1021/ci049875d

MacAulay, K., and Woodgett, J. R. (2008). Targeting glycogen synthase kinase-3 (GSK-3) in the treatment of Type 2 diabetes. *Expert Opin. Ther. Targets* 12 (10), 1265–1274. doi: 10.1517/14728222.12.10.1265

Maqbool, M., and Hoda, N. (2017). GSK3 Inhibitors in the Therapeutic Development of Diabetes, Cancer and Neurodegeneration: Past, Present and Future. *Curr. Pharm. Des.* 23 (29), 4332–4350. doi: 10.2174/1381612823666170714141450

Meslamani, J., Li, J., Sutter, J., Stevens, A., Bertrand, H. O., and Rognan, D. (2012). Protein-ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J. Chem. Inf. Model* 52 (4), 943–955. doi: 10.1021/ci300083r

O'Leary, O., and Nolan, Y. (2015). Glycogen Synthase Kinase-3 as a Therapeutic Target for Cognitive Dysfunction in Neuropsychiatric Disorders. *CNS Drugs* 29 (1), 1–15. doi: 10.1007/s40263-014-0213-z

Osolodkin, D. I., Palyulin, V. A., and Zefirov, N. S. (2013). Glycogen Synthase Kinase 3 as an Anticancer Drug Target: Novel Experimental Findings and Trends in the Design of Inhibitors. *Curr. Pharm. Des.* 19 (4), 665–679. doi: 10.2174/138161213804581972

Pandey, M. K., and DeGrado, T. R. (2016). Glycogen Synthase Kinase-3 (GSK-3)-Targeted Therapy and Imaging. *Theranostics* 6 (4), 571–593. doi: 10.7150/thno.14334

Rogers, D., and Hopfinger, A. J. (1994). Application of Genetic Function Approximation to Quantitative Structure-Activity-Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* 34 (4), 854–866. doi: 10.1021/Ci00020a020

Rogers, D., Brown, R. D., and Hahn, M. (2005). Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen* 10 (7), 682–686. doi: 10.1177/1087057105281365

Sahin, I., Eturi, A., De Souza, A., Pamarthy, S., Tavora, F., Giles, F. J., et al. (2019). Glycogen synthase kinase-3 beta inhibitors as novel cancer treatments and modulators of antitumor immune responses. *Cancer Biol. Ther.* 20 (8), 1047–1056. doi: 10.1080/15384047.2019.1595283

Saura, C., Roda, D., Rosello, S., Oliveira, M., Macarulla, T., Perez-Fidalgo, J. A., et al. (2017). A First-in-Human Phase I Study of the ATP-Competitive AKT Inhibitor Ipatasertib Demonstrates Robust and Safe Targeting of AKT in Patients with Solid Tumors. *Cancer Discovery* 7 (1), 102–113. doi: 10.1158/2159-8290.CD-16-0512

Shen, C., Wang, Z., Yao, X., Li, Y., Lei, T., Wang, E., et al. (2020). Comprehensive assessment of nine docking programs on type II kinase inhibitors: prediction accuracy of sampling power, scoring power and screening power. *Brief Bioinform.* 21 (1), 282–297. doi: 10.1093/bib/bby103

Sivaprakasam, P., Han, X., Civiello, R. L., Jacutin-Porte, S., Kish, K., Pokross, M., et al. (2015). Discovery of new acylaminopyridines as GSK-3 inhibitors by a structure guided in-depth exploration of chemical space around a pyrrolopyridinone core. *Bioorg. Med. Chem. Lett.* 25 (9), 1856–1863. doi: 10.1016/j.bmcl.2015.03.046

Stamos, J. L., Chu, M. L., Enos, M. D., Shah, N., and Weis, W. I. (2014). Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *Elife* 3, e01998. doi: 10.7554/eLife.01998

Sutter, J., Li, J. B., Maynard, A. J., Goupil, A., Luu, T., and Nadassy, K. (2011). New Features that Improve the Pharmacophore Tools from Accelrys. *Curr. Comput. Aided Drug Des.* 7 (3), 173–180. doi: 10.2174/157340911796504305

Swinney, Z. T., Haubrich, B. A., Xia, S., Ramesha, C., Gomez, S. R., Guyett, P., et al. (2016). A Four-Point Screening Method for Assessing Molecular Mechanism of Action (MMOA) Identifies Tideglusib as a Time-Dependent Inhibitor of Trypanosoma brucei GSK3beta. *PloS Negl. Trop. Dis.* 10 (3), e0004506. doi: 10.1371/journal.pntd.0004506

Tahtouh, T., Elkins, J. M., Filippakopoulos, P., Soundararajan, M., Burgy, G., Durieu, E., et al. (2012). Selectivity, cocrystal structures, and neuroprotective properties of leucettines, a family of protein kinase inhibitors derived from the marine sponge alkaloid leucettamine B. *J. Med. Chem.* 55 (21), 9312–9330. doi: 10.1021/jm301034u

Tanrikulu, Y., Kruger, B., and Proschak, E. (2013). The holistic integration of virtual screening in drug discovery. *Drug Discovery Today* 18 (7-8), 358–364. doi: 10.1016/j.drudis.2013.01.007

Tian, S., Li, Y., Li, D., Xu, X., Wang, J., Zhang, Q., et al. (2013a). Modeling compound-target interaction network of traditional Chinese medicines for type II diabetes mellitus: insight for polypharmacology and drug design. *J. Chem. Inf. Model* 53 (7), 1787–1803. doi: 10.1021/ci400146u

Tian, S., Sun, H., Li, Y., Pan, P., Li, D., and Hou, T. (2013b). Development and evaluation of an integrated virtual screening strategy by combining molecular docking and pharmacophore searching based on multiple protein structures. *J. Chem. Inf. Model* 53 (10), 2743–2756. doi: 10.1021/ci400382r

Tian, S., Sun, H., Pan, P., Li, D., Zhen, X., Li, Y., et al. (2014). Assessing an ensemble docking-based virtual screening strategy for kinase targets by considering protein flexibility. *J. Chem. Inf. Model* 54 (10), 2664–2679. doi: 10.1021/ci500414b

Wagman, A. S., Johnson, K. W., and Bussiere, D. E. (2004). Discovery and development of GSK3 inhibitors for the treatment of type 2 diabetes. *Curr. Pharm. Des.* 10 (10), 1105–1137. doi: 10.2174/1381612043452668

Wagner, F. F., Bishop, J. A., Gale, J. P., Shi, X., Walk, M., Ketterman, J., et al. (2016). Inhibitors of Glycogen Synthase Kinase 3 with Exquisite Kinome-Wide Selectivity and Their Functional Effects. *ACS Chem. Biol.* 11 (7), 1952–1963. doi: 10.1021/acschembio.6b00306

Wang, Y., Dou, X., Jiang, L., Jin, H., Zhang, L., Zhang, L., et al. (2019). Discovery of novel glycogen synthase kinase-3alpha inhibitors: Structure-based virtual screening, preliminary SAR and biological evaluation for treatment of acute myeloid leukemia. *Eur. J. Med. Chem.* 171, 221–234. doi: 10.1016/j.ejmech.2019.03.039

Wu, X., Stenson, M., Abeykoon, J., Nowakowski, K., Zhang, L., Lawson, J., et al. (2019). Targeting glycogen synthase kinase 3 for therapeutic benefit in lymphoma. *Blood* 134 (4), 363–373. doi: 10.1182/blood.2018874560

Xu, M., Wang, S. L., Zhu, L., Wu, P. Y., Dai, W. B., and Rakesh, K. P. (2019). Structure-activity relationship (SAR) studies of synthetic glycogen synthase kinase-3beta inhibitors: A critical review. *Eur. J. Med. Chem.* 164, 448–470. doi: 10.1016/j.ejmech.2018.12.073

Zegzouti, H., Zdanovskaia, M., Hsiao, K., and Goueli, S. A. (2009). ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases. *Assay Drug Dev. Technol.* 7 (6), 560–572. doi: 10.1089/adt.2009.0222

Zhang, H. C., Bonaga, L. V., Ye, H., Derian, C. K., Damiano, B. P., and Maryanoff, B. E. (2007). Novel bis(indolyl)maleimide pyridinophanes that are potent, selective inhibitors of glycogen synthase kinase-3. *Bioorg. Med. Chem. Lett.* 17 (10), 2863–2868. doi: 10.1016/j.bmcl.2007.02.059

Zhao, S., Zhu, J., Xu, L., and Jin, J. (2017). Theoretical studies on the selective mechanisms of GSK3beta and CDK2 by molecular dynamics simulations and free energy calculations. *Chem. Biol. Drug Des.* 89 (6), 846–855. doi: 10.1111/cbdd.12907

Zhou, W., Duan, M., Fu, W., Pang, J., Tang, Q., Sun, H., et al. (2018). Discovery of Novel Androgen Receptor Ligands by Structure-based Virtual Screening and Bioassays. *Genom. Proteom. Bioinf.* 16 (6), 416–427. doi: 10.1016/j.gpb.2018.03.007

Zhu, J., Wu, Y., Xu, L., and Jin, J. (2019). Theoretical Studies on the Selectivity Mechanisms of Glycogen Synthase Kinase 3beta (GSK3beta) with Pyrazine ATP-competitive inhibitors by 3D-QSAR, Molecular Docking, Molecular Dynamics Simulation and Free Energy Calculations. *Curr. Comput. Aided Drug Des.* 16 (1), 17–30. doi: 10.2174/1573409915666190708102459

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer HS declared a past co-authorship with one of the authors, LX, to the handling editor.

Copyright © 2020 Zhu, Wu, Wang, Li, Xu, Chen, Cai and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fphar.2020.579024

Discovery of Icotinib-1,2,3-Triazole Derivatives as IDO1 Inhibitors

Long-fei Mao[1,2†], Yu-wei Wang[3†], Jie Zhao[2], Gui-qing Xu[2], Xiao-jun Yao[4] and Yue-Ming Li[1*]*

[1] State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and Tianjin Key Laboratory of Molecular Drug Research, Nankai University, Tianjin, China, [2] School of Chemistry and Chemical Engineering, Henan Engineering Research Center of Chiral Hydroxyl Pharmaceutical, Henan Normal University, Xinxiang, China, [3] College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an-Xianyang New Economic Zone, Xianyang, China, [4] State Key Laboratory of Quality Research in Chinese Medicine/Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macau, China

Tumor immunotherapy is considered to be a highlight in cancer treatment in recent years. Indoleamine 2,3-dioxygenase 1 (IDO1) is closely related to the over expression of many cancers, and is therefore a promising target for tumor immunotherapy. To search for novel IDO1-targeting therapeutic agents, 22 icotinib-linked 1,2,3-triazole derivatives were prepared and evaluated for their inhibitory activity against IDO1. The structures of the prepared compounds were confirmed with ^{1}H NMR, ^{13}C NMR and HR MS. IDO1 inhibitory activity assay results indicated that 10 of those compounds showed remarkable inhibitory activity against IDO1, among which compound **a17** was the most potent with IC_{50}value of 0.37 μM. The binding model between the prepared compounds and IDO1 was studied with molecular modeling study. The current study suggested that icotinib-1,2,3-triazole derivatives could be used as potential inhibitors that preferentially bind to the ferrous form of IDO1 through the formation of coordinate bond with the haem iron.

Keywords: icotinib, 1,2,3-triazole, indoleamine 2,3-dioxygenase 1, inhibitor, immunotherapy

OPEN ACCESS

Edited by:
Weiwei Xue,
Chongqing University, China

Reviewed by:
Zhili Zuo,
Chinese Academy of Sciences, China
Che-Hsin Lee,
National Sun Yat-sen University, Taiwan

****Correspondence:***
Xiao-jun Yao
xjyao@must.edu.mo
Yue-Ming Li
ymli@nankai.edu.cn

[†]These authors have contributed equally to this work

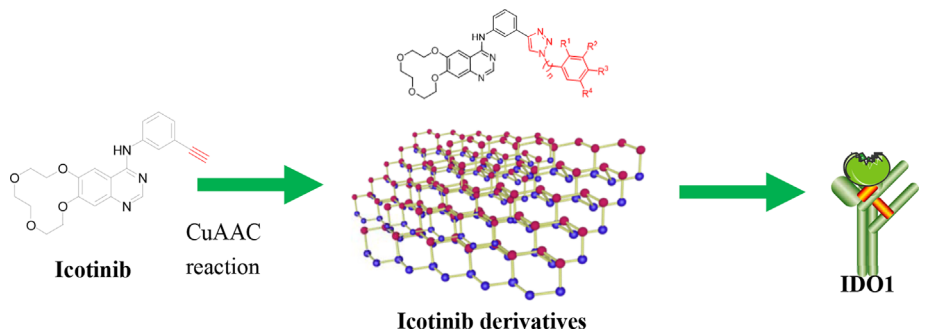
Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *01 July 2020*
Accepted: *28 August 2020*
Published: *30 September 2020*

Citation:
Mao L-f, Wang Y-w, Zhao J, Xu G-q, Yao X-j and Li Y-M (2020) Discovery of Icotinib-1,2,3-Triazole Derivatives as IDO1 Inhibitors. Front. Pharmacol. 11:579024. doi: 10.3389/fphar.2020.579024

INTRODUCTION

Tumor immunotherapy is an emerging field in tumor treatment. Studies show that indoleamine 2,3-dioxygenase 1 (IDO1) is the initial and rate-limiting enzyme that catalyzes the metabolism of tryptophan along the kynurenine pathway outside the human liver (Chen et al., 2019), and plays an important role in regulating the body's innate and adaptive immunity by catalyzing tryptophan metabolism (Takikawa et al., 1986; Takikawa, 2005). In the tumor microenvironment, tumor cells can induce IDO1 over expression, which causes the depletion of local tryptophan and the accumulation of metabolites such as kynurenine, thereby activating GCN2 and AHR signaling pathways, inhibiting T cell proliferation, and inducing apoptosis (Muller et al., 2005). Additionally, the original T cells are stimulated to differentiate into regulatory T cells, thus mediating tumor immune escape (Efimov et al., 2011). Over expression of IDO1 has been found in a variety of malignant tumors, such as ovarian cancer, pancreatic cancer, and non-small cell lung cancer. Therefore, IDO1 inhibitors once attracted considerable attention as potential agents for cancer treatment.



GRAPHICAL ABSTRACT |

Several candidates are currently undergoing clinical trials, but none of these has been approved so far, suggesting that the identification of potent and clinically useful IDO1 inhibitors is an open challenge. For example, epacadostat (**Figure 1**, **1**, INCB024360) (Yue et al., 2017), indoximod (**Figure 1**, **2**, 1-methyl-D-tryptophan) (Soliman et al., 2009), navoximod (**Figure 1**, **3**, NLG-919) (Kumar et al., 2019), EOS-200271 (**Figure 1**, **4**, PF-06840003) (Crosignani et al., 2017), and BMS-986205 (**Figure 1**, **5**) have been are currently being tested in human clinical trials. Epacadostat, developed by Incyte, is the first highly effective and highly selective oral IDO1 inhibitor (Morgan et al., 2008; Lin et al., 2016; Lewis-Ballester et al., 2017; Yue et al., 2017). It can effectively restore the anti-tumor immune response in human Hela cells treated with IFN-γ. *Via* reversing tumor-associated immunosuppression, it can effectively suppress kynurenineproduction. Epacadostat also increases IFN-γ production, promotes the growth of natural killer (NK) and T cells, and reduces the number of converted regulatory T cells (Tregs) (Dounay et al., 2015). Based on the promising results in Phase 1/2 studies, epacadostat proceeded to a Phase 3 trial (ECHO-301) in combination with pembroluzimab in the treatment of metastatic melanoma. Recent results coming from the pivotal Phase 3 trial of ECHO-301 have shown no indication that epacadostat provides an increased benefit compared to pembrolizumab alone, questioning the effectiveness of IDO1 inhibitors. This failure led to the interruption of other Phase III trials and the reconsideration of whether some elements had been neglected in the landscape of IDO1 inhibitors (Serafini et al., 2020). Study showed that the key group playing the active role in the epacadostat molecule was the oxadiazole structure. The epacadostat molecule entered the heme pocket of the IDO1, and the oxadiazole structure located directly above the Fe ion in the heme and then interacted with it. Based on this, a new type of IDO1 inhibitor which linked urea groups to the oxadiazole structure was developed (Wu et al., 2017; Song et al., 2020) (**Figure 1**, **6**) and the compounds showed submicromolar level of IC_{50} against IDO1.

1,2,3-Triazole, an *N*-heterocyclic building block, played a significant role in drug design and synthesis (Majeed et al., 2013). Many compounds containing the 1,2,3-triazole unit exhibited good activities against inflammation, cancer, and microbes (De Souza et al., 2020). Moreover, copper(I)-catalysedazide-alkyne

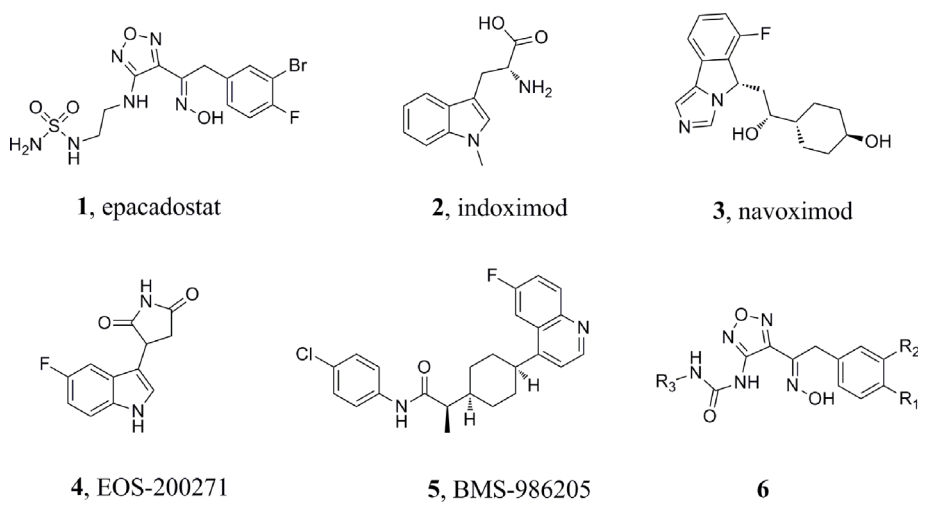


FIGURE 1 | Chemical structures of sixIDO1 inhibitors.

cycloaddition (CuAAC) reaction, a convenient and regiospecific approach to 1,4-disubstituted triazoles (Thomopoulou et al., 2015), has aroused great interest among the researchers and has been widely used in the preparation of different bioactive molecules (Hong et al., 2010). Compounds containing 1,2,3-triazole moiety showed good bioactivities such as antitumor or antibacterial activity (Röhrig et al., 2012; Mao et al., 2017). Furthermore, according to the literature, compounds containing 1,2,3-triazole possessed promising IDO1 inhibition (IC_{50} = 12.6 μM).

Encouraged by these results, we decided to study the bioactivity of compounds bearing different 1,2,3-triazole groups. At first, icotinib was chosen as the starting point. This compound has been clinically used in China for the treatment of NSCLC (Yang et al., 2017; Liang et al., 2018; Zhang et al., 2018). We envisioned that introducing 1,2,3-triazole structure into the molecule *via* conventional click reaction would give compounds with additional benefit by the 1,2,3-triazole group, and this twin drug approach will combine the advantages of both EGFR-TKI and IDO1 inhibitors. Herein, we wish to present our preliminary results on the preparation of the 1,2,3-triazole derivatives and their *in vitro* inhibitory activity against IDO1.

CHEMISTRY

The synthetic strategy for the preparation of the target compounds is illustrated in **Figure 2**. Copper(I)-catalysed azide–alkyne cycloaddition between icotinib and different azido compounds afforded the target compounds **a1**–**a22**. The reaction conditions of these steps were convenient and easy to control. The structures of some key intermediates and all target compounds were confirmed by nuclear magnetic resonance and high-resolution mass spectrometry experiments.

EXPERIMENTAL PROTOCOLS

Materials and Chemistry

Icotinib-1,2,3-triazole derivatives were in-house synthesized. Icotinib and Aryl-azido compounds were purchased from Acros Organics (Morris Plains, NJ, USA). All reagents and solvents obtained from commercially available source were used without further treatment. ^{1}H NMR and ^{13}C NMR spectra were acquired in DMSO-d_6 or $CDCl_3$ solution with a Bruker 600 spectrometer. Chemical shifts (δ) were given in parts per million with tetramethylsilane as internal reference and coupling constants were expressed in hertz. High-resolution mass spectra (HRMS) measurements were carried out using an Bruker MicrOTOF-Q II mass spectrometer.

Hela cell line, DMEM medium and fetal bovine serum were purchased from ATCC (Virginia, USA). Recombinant human IFN-γ was purchased from R&D systems (Emeryville, CA, USA). The 3.05 Ntrichloroacetic acid, 4-(dimethylamino)benzaldehyde and acetic acid were purchased from Sigma Aldrich (St. Louis, MI, USA).

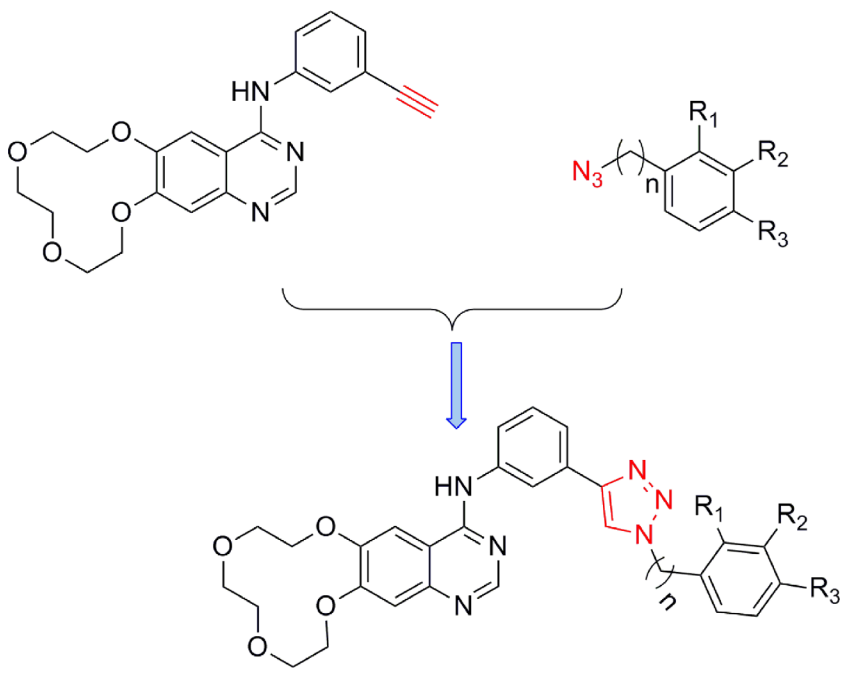


FIGURE 2 | The reaction routes to1,2,3-triazole-linkedicotinibderivatives.

General Procedure for the Synthesis of Analogues a1–a22

General Procedure for Preparation of Compound a1–a22

Aryl-azido (1.2 mmol) and icotinib (1.0 mmol) were added to 15 ml mixed solvent (water: *tert*-butanol =2:1). The reaction was carried out with copper sulfate pentahydrate (0.1 mmol) and sodium ascorbate (0.2 mmol) at 80°C. After completion of the reaction (monitored by TLC), the mixture was extracted with dichloromethane (15 ml×3). The combined organic phase was washed successively with water and brine, dried over sodium sulfate and concentrated *in vacuo*. The residue was purified wiht column chromatography (CH_2Cl_2/MeOH=20:1) to give the desired compound a.

*{3-[1-(3-Fluoro-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine(**a1**):* Yellow solid, Purity 96%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.66 (s, 1H), 9.38 (s, 1H), 8.56 (s, 1H), 8.44 (s, 1H), 8.23 (s, 1H), 7.98 (d, *J*=7.9 Hz, 1H), 7.91 (dd, J_1 = 19.0 *Hz*, J_2 = 9.0 *Hz*, 2H), 7.70 (dd, J_1 = 14.8 *Hz*, J_2 = 7.8 *Hz*, 1H), 7.65 (d, *J*=7.5 Hz, 1H), 7.53 (t, *J*=7.8 Hz, 1H), 7.39 (t, *J*=8.3 Hz, 1H), 7.33 (s, 1H), 4.32 (d, *J*=12.4 Hz, 4H), 3.79 (d, *J*=19.9 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 163.8, 162.1, 157.1, 156.5, 153.8, 150.3, 147.9, 140.6, 138.4, 132.4, 130.8, 129.7, 122.6, 121.0, 120.3, 119.3, 116.4, 115.8, 112.1, 110.7, 108.1, 107.9, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4N_6F$ $[M+H]^+$ 529.1994, found 529.2000.

*{3-[1-(4-Chloro-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a2**)*: Brown solid, Purity 97%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.37 (s, 1H), 8.59 (s, 1H), 8.44 (s, 1H), 8.30 (s, 1H), 8.04 (d, *J*=7.7 Hz, 2H), 7.95 (d, *J*=8.0 Hz, 1H), 7.74 (d, *J*=7.7 Hz, 2H), 7.67 (d, *J*=7.6 Hz, 1H), 7.53 (t, *J* = 7.8 Hz, 1H), 7.34 (s, 1H), 4.33 (d, *J*=13.2 Hz, 4H), 3.79 (d, *J* = 22.6 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.3, 156.7, 150.4, 147.9, 140.4, 135.9, 133.6, 130.8, 130.4, 129.6, 126.5, 125.7, 122.7, 122.1, 121.2, 119.5, 110.9, 73.4, 70.9, 70.5, 69.2, 68.9, 64.3, 45.9, 8.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4N_6Cl$ $[M+H]^+$ 545.1699, found 545.1703.

*{3-[1-(4-Fluoro-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a3**)*: White solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.70 (s, 1H), 9.31 (s, 1H), 8.62 (s, 1H), 8.43 (s, 1H), 8.25 (s, 1H), 8.04 (d, J_1 = 8.4 $_{Hz}$, J_2 = 4.7 *Hz*, 2H), 7.96 (d, *J*=7.9 Hz, 1H), 7.66 (d, *J*=7.5 Hz, 1H), 7.63-7.39 (m, 3H), 7.35 (s, 1H), 4.32 (s, 4H), 3.80 (d, *J*=21.6 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 162.9, 161.4, 157.2, 156.5, 150.3, 147.8, 140.5, 133.7, 130.9, 129.6, 122.9, 122.8, 122.6, 121.1, 120.5, 119.4, 117.4, 117.2, 111.9, 110.8, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcdfor $C_{28}H_{26}O_4N_6F$ $[M+H]^+$ 529.1994, found 529.2000.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(2-methoxy-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a4**)*: White solid, Purity 96%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.63 (s, 1H), 8.92 (s, 1H), 8.53 (s, 1H), 8.41 (s, 1H), 8.25 (s, 1H), 7.97 (d, *J*=8.0 Hz, 1H), 7.70 (d, *J*=7.8 Hz, 1H), 7.66 (d, *J*=7.6 Hz, 1H), 7.57 (t, *J*=7.9 Hz, 1H), 7.50 (t, *J*=7.8 Hz, 1H), 7.36 (d, *J*=8.4 Hz, 1H), 7.32 (s, 1H), 7.19 (t, *J*=7.6 Hz, 1H), 4.32 (d, *J*=13.1 Hz, 4H), 3.90 (s, 3H), 3.80 (d, *J*=20.7 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.4, 153.9, 152.3, 150.2, 148.1, 146.7, 140.6, 131.4, 131.2, 129.5, 126.4, 126.2, 123.9, 122.3, 121.4, 121.0, 119.2, 113.5, 112.1, 110.8, 110.3, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{29}H_{29}O_5N_6$ $[M+H]^+$ 541.2194, found 541.2197.

*{3-[1-(2-Chloro-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a5**)*: Yellow solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.63 (s, 1H), 9.07 (s, 1H), 8.53 (s, 1H), 8.45 (s, 1H), 8.22 (s, 1H), 7.97 (d, *J*=8.1 Hz, 1H), 7.82 (t, *J*=9.0 Hz, 2H), 7.69-7.63 (m, 3H), 7.51 (t, *J*=7.9 Hz, 1H), 7.32 (s, 1H), 4.31 (d, *J*=14.9 Hz, 4H), 3.80 (d, *J*=21.4 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.9, 150.2, 148.2, 147.0, 140.6, 131.1, 130.9, 129.6, 129.1, 129.0, 128.9, 124.2, 119.3, 112.2, 110.7, 110.2, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4N_6Cl$ $[M+H]^+$ 545.1699, found 545.1704.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-[3-(1-phenyl-1H-[1,2,3]triazol-4-yl)-phenyl]-amine (**a6**)*: Yellow solid, Purity 97.1%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.68 (s, 1H), 9.33 (s, 1H), 8.53 (s, 1H), 8.45 (s, 1H), 8.24 (s, 1H), 7.99 (t, *J*=10.1 Hz, 3H), 7.67-7.64 (m, 3H), 7.53 (q, *J*=7.5 Hz, 2H), 7.32 (s, 1H), 4.32 (d, *J*=13.6 Hz, 4H), 3.79 (d, *J*=20.2 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.2, 156.4, 153.9, 150.2, 148.1, 147.8, 140.6, 137.1, 131.0, 130.4, 129.6, 129.2, 122.5, 121.0, 120.5, 120.2, 119.3, 112.2, 110.7, 110.3, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4NaN_6$ $[M+Na]^+$ 533.1913, found 533.1915.

*{3-[1-(3-Chloro-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a7**)*: Yellow solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.70 (s, 1H), 9.41 (s, 1H), 8.68 (s, 1H), 8.44 (s, 1H), 8.28 (s, 1H), 8.13 (s, 1H), 7.99 (dd, J_1 = 27.4 *Hz*, J_2 = 8.0 *Hz*, 2H), 7.67 (dd, J_1 = 16.5 *Hz*, J_2 = 8.0 *Hz*, 2H), 7.60 (d, *J*=8.0 Hz, 1H), 7.53 (t, *J*=7.8 Hz, 1H), 7.37 (s, 1H), 4.32 (s, 4H), 3.79 (d, *J*=21.8 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.2, 156.5, 153.7, 150.3, 147.9, 140.5, 138.2, 134.7, 132.2, 130.8, 130.1, 129.7, 129.0, 122.6, 121.1, 120.4, 120.3, 119.4, 119.0, 112.0, 110.8, 93.3, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4N_6Cl$ $[M+H]^+$ 545.1699, found 545.1705.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(2-iodo-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a8**)*: White solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.65-9.60 (m, 1H), 8.99 (s, 1H), 8.58-8.46 (m, 2H), 8.32-8.14 (m, 2H), 7.96 (d, *J*=10.5 Hz, 1H), 7.68-7.65 (m, 2H), 7.57-7.40 (m, 2H), 7.32 (d, *J*=6.1 Hz, 1H), 5.70 (m, 1H), 4.31 (s, 4H), 3.79 (d, *J*=25.0 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.9, 150.3, 148.0, 147.0, 140.6, 140.3, 140.3, 140.0, 131.1, 130.2, 129.9, 129.6, 129.4, 128.6, 124.0, 122.3, 121.0, 119.2, 112.1, 110.8, 110.2, 96.5, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9, 58.0; HR MS (ESI) *m/z*: calcd for $C_{28}H_{26}O_4N_6I$ $[M+H]^+$ 637.1060, found 637.1065.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(3-methoxy-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a9**)*: Brown solid, Purity 97%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.69 (s, 1H), 9.39 (s, 1H), 8.58 (s, 1H), 8.47 (s, 1H), 8.27 (s, 1H), 8.02 (d, *J*=7.9 Hz, 1H), 7.70 (d, *J*=7.6 Hz, 1H), 7.63-7.56 (m, 4H), 7.37 (s, 1H), 7.14 (d, *J*=7.9 Hz, 1H), 4.36 (d, *J*=11.7 Hz, 4H), 3.94 (s, 3H), 3.84 (d, *J*=20.1 Hz, 4H), 3.70 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 160.7, 157.1, 156.5, 153.9, 150.3, 148.1, 147.8, 140.6, 138.2, 131.4, 131.0, 129.6, 122.5, 121.0, 120.2, 119.3, 114.9, 112.4, 112.2, 110.7, 110.2, 106.1, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9, 56.1; HR MS (ESI) *m/z*: calcd for $C_{29}H_{28}O_5N_6Na$ $[M+Na]^+$ 563.2013, found 563.2017.

*{3-[1-(4-Bromo-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a10**)*: White solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.63 (s, 1H), 9.36 (s, 1H), 8.53 (s, 1H), 8.44 (s, 1H), 8.22 (s, 1H), 7.97 (d, *J*=8.6 Hz, 3H), 7.86 (d, *J*=8.6 Hz, 2H), 7.65 (d, *J*=7.6 Hz, 1H), 7.52 (t, *J*=7.9 Hz, 1H), 7.32 (s, 1H), 4.32 (d, *J*=11.2 Hz, 4H), 3.79 (d, *J*=20.5 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.9, 150.3, 148.1, 148.0, 140.6, 136.3, 133.3, 130.8, 129.6, 122.5, 122.4, 121.8, 121.0, 120.2, 119.3, 112.2, 110.7, 110.2, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{25}O_4N_6BrNa$ $[M+Na]^+$ 611.1013, found 611.1020.

*{3-[1-(2-Bromo-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a11**)*: Yellow solid, Purity 97%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.64 (s, 1H), 9.04 (s, 1H), 8.54 (s, 1H), 8.45 (s, 1H), 8.22 (s, 1H), 7.97 (d, *J*=8.0 Hz, 2H), 7.77 (d, *J*=7.8 Hz, 1H), 7.71-7.65 (m, 2H), 7.60 (dd, J_1 = 17.2 *Hz*, J_2 = 9.5 *Hz*, 1H), 7.51 (t, *J*=7.8 Hz, 1H), 7.32 (s, 1H), 4.32 (d, *J*=10.4 Hz, 4H), 3.79 (d, *J*=21.8 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): *d* 157.1, 156.5, 153.9, 150.2, 148.1, 146.9, 140.6, 136.7, 134.1, 132.6, 130.9, 129.6, 129.5, 129.2, 124.2, 122.4, 121.0, 119.5, 119.3, 112.1, 110.8, 110.2, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{28}H_{25}O_4N_6BrNa$ $[M+Na]^+$ 611.1013, found 611.1021.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(4-trifluoromethyl-phenyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a12**)*: White solid, Purity 99%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.69 (s, 1H), 9.48 (s, 1H), 8.45 (s, 1H), 8.36 (s, 1H), 8.25 (d, *J*=8.0 Hz, 2H), 8.04 (d, *J*=8.0 Hz, 2H), 7.97 (d, *J*=7.5 Hz, 1H), 7.68 (d, *J*=7.4 Hz, 1H), 7.53 (t, *J*=7.6 Hz, 1H), 4.32 (s, 4 H), 3.79 (d, *J*=19.9 Hz, 4H), 3.64 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 156.7, 156.3, 150.4, 148.1, 140.7, 139.9, 130.7, 129.9, 129.3, 129.1, 128.8, 127.8, 125.2, 123.4, 122.6, 121.6, 121.1, 120.9, 120.4, 119.4, 110.9, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9; HR MS (ESI) *m/z*: calcd for $C_{29}H_{26}O_4N_6F_3$ $[M+H]^+$ 579.1962, found 579.1972.

*{3-[1-(2-Fluoro-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a13**)*: Yellow solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.99 (s, 1H), 9.05 (s, 1H), 8.93 (s, 1H), 8.73 (t, *J*=1.8 Hz, 1H), 8.61 (s, 1H), 8.36 (d, *J*=7.3 Hz, 1H), 7.99 (d, *J*=7.7 Hz, 1H), 7.89-7.82 (m, 3H), 7.73-7.66 (m, 3H), 6.16 (s, 2H), 4.73 (d, *J*=16.9 Hz, 4H), 4.23-4.18 (m, 4H), 4.07 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 161.8, 160.2, 157.5, 156.9, 154.3, 150.6, 148.5, 147.5, 140.9, 131.7, 131.7, 129.9, 125.8, 123.7, 122.6, 122.5, 121.2, 119.5, 116.6, 116.5, 112.6, 111.1, 110.6, 73.8, 71.4, 71.3, 70.9, 69.7, 69.3, 48.0; HRMS (ESI) *m/z*: calcd for $C_{29}H_{28}O_4N_6F$ $[M+H]^+$ 543.2151, found 543.2158.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(2-methyl-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a14**)*: Yellow solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.61 (s, 1H), 8.70 (s, 1H), 8.56 (s, 1H), 8.31 (s, 1H), 8.25 (s, 1H), 7.93 (d, *J*=8.0 Hz, 1H), 7.57 (d, *J*=7.6 Hz, 1H), 7.45 (t, *J*=7.9 Hz, 1H), 7.37 (s, 1H), 7.28-7.21 (m, 3H), 7.16 (d, *J*=7.5 Hz, 1H), 5.68 (s, 2H), 4.31 (s, 4H), 3.78 (d, *J*=24.3 Hz, 4H), 3.64 (s, 4H), 2.36 (s, 3H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.4, 153.7, 150.3, 147.0, 140.5, 136.8, 134.6, 131.4, 130.9, 129.5, 129.2, 128.8, 126.8, 122.1, 120.9, 119.1, 112.3, 110.8, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9, 51.7, 19.2; HRMS (ESI) *m/z*: calcd for $C_{30}H_{31}O_4N_6$ $[M+H]^+$539.2401, found 539.2406.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(4-methyl-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a15**)*: Brown solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.59 (s, 1H), 8.61 (s, 2H), 8.30 (s, 1H), 8.24 (s, 1H), 7.93 (d, *J*=8.0 Hz, 1H), 7.55 (d, *J*=7.6 Hz, 1H), 7.45 (t, *J*=7.9 Hz, 1H), 7.35 (s, 1H), 7.28 (d, *J*=7.7 Hz, 2H), 7.21 (d, *J*=7.7 Hz, 2H), 5.61 (s, 2H), 4.31 (s, 4H), 3.79 (d, J=23.8 Hz, 4H), 3.65 (s, 4H), 2.30 (s, 3H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.7, 150.3, 147.1, 140.5, 138.0, 133.5, 131.4, 129.8, 129.5, 128.5, 122.1, 121.9, 120.8, 119.1, 112.3, 110.8, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9, 53.3, 21.2; HRMS (ESI) *m/z*: calcd for $C_{30}H_{31}O_4N_6$ $[M+H]^+$ 539.2401, found 539.2408.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(2-iodo-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a16**)*: Yellow solid, Purity 96%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.61 (s, 1H), 8.57 (d, *J*=16.0 Hz, 2H), 8.32 (s, 1H), 8.21 (s, 1H), 7.95 (dd, J_1 = 13.5 Hz, J_2 = 8.0 Hz, 2H), 7.58 (d, *J*=7.6 Hz, 1H), 7.45 (dd, J_1 = 14.9 Hz, J_2 = 7.5 Hz, 2H), 7.32 (s, 1H), 7.17-7.13 (m, 2H), 5.70 (s, 2H), 4.31 (s, 4H), 3.79 (d, *J*=23.6 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.7, 150.2, 147.9, 147.0, 140.5, 140.0, 138.4, 131.3, 130.8, 130.2, 129.5, 129.4, 122.5, 122.1, 120.9, 119.1, 112.0, 110.8, 110.3, 99.7, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9, 58.0; HRMS (ESI) *m/z*: calcd for $C_{29}H_{28}O_4N_6I$ $[M+H]^+$ 651.1211, found 651.1220.

*{3-[1-(3-Bromo-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a17**)*: White solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.59 (s, 1H), 8.68 (s, 1H), 8.53 (s, 1H), 8.32 (s, 1H), 8.20 (s, 1H), 7.93 (d, *J*=8.0 Hz, 1H), 7.62 (s, 1H), 7.56 (d, *J*=6.9 Hz, 2H), 7.46 (t, *J*=7.9 Hz, 1H), 7.38 (d, *J*=4.4 Hz, 2H), 7.31 (s, 1H), 5.69 (s, 2H), 4.31 (s, 4H), 3.79 (d, *J*=22.7 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.5, 153.8, 150.3, 148.0, 147.2, 140.5, 139.1, 131.6, 131.5, 131.2, 129.5, 127.6, 122.4, 122.3, 122.1, 120.8, 119.1, 112.1, 110.7, 110.3, 73.4, 70.9, 70.8, 70.5, 69.3, 68.9, 52.7; HRMS (ESI) *m/z*: calcd for $C_{29}H_{28}O_4N_6Br$ $[M+H]^+$ 603.1350, found 603.1356.

*[3-(1-Benzyl-1H-[1,2,3]triazol-4-yl)-phenyl]-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b] naphthalen-4-yl)-amine (**a18**)*: Brown solid, Purity 95%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.60 (s, 1H), 8.66 (s, 1H), 8.51 (s, 1H), 8.32 (s, 1H), 8.21 (s, 1H), 7.93 (d, *J*=7.9 Hz, 1H), 7.56 (d, *J*=7.6 Hz, 1H), 7.47-7.35 (m, 6H), 7.31 (s, 1H), 5.67 (s, 2H), 4.31 (s, 4H), 3.81-3.76 (m, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.4, 153.9, 150.2, 148.1, 147.1, 140.5, 136.5, 131.4, 129.5, 129.3, 128.7, 128.4, 122.1, 122.1, 120.8, 119.1, 112.1, 110.7, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9, 53.5; HR MS (ESI) *m/z*: calcd for $C_{29}H_{29}O_4N_6$ $[M+H]^+$ 525.2245, found 525.2254.

*{3-[1-(2-Bromo-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a19**)*:White solid, Purity 98%; ^{1}H NMR (600 MHz, DMSO-d_6): δ9.59 (s, 1H), 8.62 (s, 1H), 8.51 (s, 1H), 8.32 (s, 1H), 8.20 (s, 1H), 7.94 (d, *J*=8.3 Hz, 1H), 7.73 (d, *J*=7.8 Hz, 1H), 7.57 (d, *J*=7.5 Hz, 1H), 7.45 (dd, J_1 = 14.4 *Hz*, J_2 = 7.4 *Hz*, 2H), 7.35 (t, *J*=7.7 Hz, 1H), 7.31 (s, 1H), 7.26 (d, *J*=7.7 Hz, 1H), 5.76 (s, 2H), 4.31 (s, 4H), 3.79 (d, *J*=23.7 Hz, 4H), 3.63 (s, 4H); HR MS (ESI) *m/z*: calcd for $C_{29}H_{27}O_4N_6BrNa$ $[M+Na]^+$ 625.1169, found 625.1178.

*(7,8,10,11,13,14-Hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-{3-[1-(3-methoxy-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-amine (**a20**)*: Yellow solid, Purity 97%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.72 (s, 1H), 8.66 (s, 1H), 8.50 (s, 1H), 8.35 (s, 1H), 8.31 (s, 1H), 7.95 (d, *J*=8.0 Hz, 1H), 7.55 (d, *J*=7.5 Hz, 1H), 7.44 (t, *J*=7.9 Hz, 1H), 7.35-7.28 (m, 2H), 6.98 (s, 1H), 6.93 (d, *J*=7.9 Hz, 2H), 5.63 (s, 2H), 4.32 (d, *J*=23.7 Hz, 4H), 3.80 (s, 2H), 3.76 (s, 5H), 3.64 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 159.9, 157.1, 156.4, 153.9, 150.2, 148.1, 147.1, 140.6, 137.9, 131.3, 130.5, 129.4, 122.1, 120.7, 120.5, 119.2, 114.3, 114.0, 112.2, 110.8, 110.3, 73.3, 70.9, 70.4, 69.2, 68.8, 55.6, 53.4; HR MS (ESI) *m/z*: calcd for $C_{30}H_{30}O_5N_6Na$ $[M+Na]^+$ 577.2170, found 577.2176.

*{3-[1-(3,5-Dibromo-benzyl)-1H-[1,2,3]triazol-4-yl]-phenyl}-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a21**)*: Yellow solid, Purity 97%; ^{1}H NMR (600 MHz, DMSO-d_6): δ9.62 (s, 1H), 8.71 (s, 1H), 8.51 (s, 1H), 8.33 (s, 1H), 8.22 (s, 1H), 7.93 (d, *J*=8.2 Hz, 1H), 7.86 (s, 1H), 7.64 (s, 2H), 7.57 (d, *J*=7.5 Hz, 1H), 7.46 (t, *J*=7.8 Hz, 1H), 7.31 (s, 1H), 5.70 (s, 2H), 4.32 (s, 4H), 3.78 (d, *J*=26.7 Hz, 4H), 3.65 (s, 4H); ^{13}C NMR (150 MHz, DMSO-d_6): 157.1, 156.4, 153.9, 150.2, 148.1, 147.2, 140.8, 140.5, 133.7, 131.2, 130.7, 129.5, 123.2, 122.4, 122.2, 120.8, 119.1, 112.2, 110.7, 110.2, 73.4, 70.9, 70.9, 70.5, 69.2, 68.9, 52.0; HR MS (ESI) *m/z*: calcd for $C_{29}H_{26}O_4N_6Br_2Na$ $[M+Na]^+$703.0274, found 703.0281.

*(3-{1-[2-(4-Fluoro-phenyl)-ethyl]-1H-[1,2,3]triazol-4-yl}-phenyl)-(7,8,10,11,13,14-hexahydro-6,9,12,15-tetraoxa-1,3-diaza-cyclododeca[b]naphthalen-4-yl)-amine (**a22**)*: Brown solid, Purity 95%; ^{1}H NMR (600 MHz, DMSO-d_6): δ 9.58 (s, 1H), 8.52 (s, 2H), 8.30 (s, 1H), 8.20 (s, 1H), 7.92 (d, *J*=7.9 Hz, 1H), 7.51 (d, *J*=7.5 Hz, 1H), 7.45 (t, *J*=7.8 Hz, 1H), 7.32 (s, 1H), 7.29-7.16 (m, 2H), 7.12 (t, *J*=8.7 Hz, 2H), 4.67 (t, *J*=7.2 Hz, 2H), 4.31 (s, 4H), 3.79 (d, *J*=21.0 Hz, 4H), 3.65 (s, 4H), 3.24 (t, *J*=7.2 Hz, 2H); ^{13}C NMR (150 MHz, DMSO-d_6): 162.4, 160.7, 157.1, 156.4, 153.9, 150.2, 148.1, 148.1, 146.6, 140.5, 134.3, 131.5, 131.1, 131.0, 121.9, 121.9, 120.7, 119.0, 115.7, 115.6, 112.2, 110.7, 110.2, 73.4, 70.9, 70.9, 70.5, 69.3, 68.9, 51.1, 35.2; HR MS (ESI) *m/z*: calcd for $C_{30}H_{29}O_4N_6FNa$ $[M+Na]^+$ 579.2127, found 579.2134.

IDO1 Enzymatic Inhibition Assay

To perform the Hela cell based IDO1 assay, Hela cells were seeded at 50,000 cells per well into 96-well microplate in 100 µl of DMEM 10% fetal bovine serum 1% Penicillin-Streptomycin. Cells were incubated at 37°C and 5% CO_2 overnight.

The next day 100 µl per well of diluted inhibitor in growth medium was added at a final concentration of 100 ng/ml human IFN-γ. Cells were incubated at 37°C in a CO_2 incubator for 18 h. The next day 140 µl of medium was moved into a new 96-well plate and 20 µl of 3.05 N trichloroacetic acid (TCA) was added. The plate was incubated at 50°C for 30 min to hydrolyze N-formylkynurenine. The plate was then centrifuged at 2,500 rpm for 10 min to remove sediments. 100 µl of supernatant per well was transferred to another 96-well plate and mixed with 100 µl of 2% (w/v) 4-(Dimethylamino)benzaldehyde in acetic acid. The plate was incubated at room temperature for 10 min, the yellow color derived from kynurenine was recorded by measuring absorbance at 480 nm using a microplate reader (PerkinElmer, USA).

Molecular Modeling

Molecular docking studies were performed with the *Glide*6.6 module in Schrödinger 2015, and the IDO1 complex with Amg-1 (PDB:4pk5) was used. The Protein preparation module in Maestro 10.1 was used to assign bond orders, add hydrogens, create zero-order bond to metals, create disulfide bonds, delete water molecules beyond 5 Å from het group, assign partial charge, assign protonation states, and minimize the structure with OPLS-2005 force field. The *Ligprep*3.3 module in Maestro 10.1 was used to generate stereoisomers, and the protonation states of ligands at pH 7.0 ± 2.0 were generated with *Epik*3.1. For the other parameters. the molecular interactions between ligand and receptor were visualized with Pymol software.

Cytotoxicity Assay

Cytotoxity of the chosen compounds was evaluated by the Cell Counting Kit-8 (CCK8, DOJINDO, Japan) assay. The cells were seeded at a density of 2,000 cells per well into 96-well microplate in 100 µl of growth medium. Cells were incubated at 37°C and 5% CO_2 overnight. The next day, 100 µl per well of diluted inhibitor in growth medium was added with the final concentration from 0.1nM to 100 µM. The cells were treated with DMSO as control. A series of dilutions were made in 0.1% DMSO in assay medium so that the final concentration of DMSO was 0.1% in all the treatments. Cells were incubated at 37°C and 5% CO_2 for 72 h. Then, 10 µl of CCK8 was added to each well. The plates were incubated at 37°C for 2 h, and the plates were recorded by measuring the absorbance at 450 nm with the reference wavelength of 630 nm using an EnVisionMultilabel Reader (PerkinElmer). The IC_{50} values were calculated and determined by the concentration causing a half-maximal percent activity. All assays were conducted with three parallel samples and three repetitions.

RESULTS AND DISCUSSION

IDO1 Inhibition Study

To investigate the IDO1 inhibition activities of the synthesized derivatives, all the new compounds and icotinib were screened *via* Hela cell-based functional assay using methods described in the literature (Yue et al., 2009; Malachowski et al., 2016; Qian et al., 2016). BMS-986205 was also used as a positive control and the IC_{50} value was tested as 0.62 nM, which is consistent with the results previously reported by Nelp et al. (IC_{50} = 0.5 nM) (Nelp et al., 2018).

As demonstrated in **Table 1**, IDO1 inhibitory activity showed that several compounds exhibited higher IDO1 inhibitory activity than icotinib, such as **a4** (IC_{50} = 1.32μM), **a6** (IC_{50} = 0.77μM), **a8** (IC_{50} = 2.50μM), **a9** (IC_{50} = 1.41μM), **a11** (IC_{50} = 1.00μM), **a14** (IC_{50} = 0.79μM), **a15** (IC_{50} = 0.59μM), **a16**(IC_{50} = 1.51μM), **a17** (IC_{50} = 0.37μM), **a18** (IC_{50} = 0.56μM), **a19** (IC_{50} = 1.50μM), **a20** (IC_{50} = 0.76μM), **a21** (IC_{50} = 0.68μM), and **a22** (IC_{50} = 0.81μM), suggesting that the inhibitory activity of the compounds against IDO1 could be enhanced upon introduction of triazolegroups, and the triazole ring might be used as an active group to interact with the IDO1. Carefully examining the results also suggested that when the benzene ring bore the same substituent, the compounds with benzyl-linked triazole groups were generally more active than those with phenyl-linked triazoles especially for those showing submicromolar levels of IC_{50} values against IDO1.

Molecular Docking Studies

Docking experiments were then carried out to explore the potential binding mode between the prepared compounds and IDO1. Compounds **a17** and compound **a18**, which showed the best activity against IDO1, were chosen as model compounds for the experiments. The docking results are depicted in **Figure 3**. The molecular docking results suggested that **a17** and **a18** could be docked into the hydrophobic site of IDO1with docking score of -8.41 and -8.19 kcal/mol, respectively. The docking experiments also suggested that compound **a17** and **a18** could bind to the binding pocket, the triazole structure mainly located on the top of the HEM, and one nitrogen atom tended to form a coordination bond with the iron in the heme. The triazole ring of a17 formed a π-π interaction with the benzene ring of phenylalanine 163. The benzyl group at the N3 position of 1,2,3-triazole could occupied the hydrophobic pocket containing cysteine 129 above the heme, the backbone amino group of glycine 261 formed a hydrogen bond with one oxygen of the hydroxyl group, and the main chain amino group of glycine 236 formed hydrogen bonds with the benzene ring. For compound **a18**, no intermolecular hydrogen bonds are formed except for the formation of coordination bonds. These results were generally in good agreement with previous study that compounds containing coordinating atoms would act as potential inhibitors of IDO1 (Röhrig et al., 2012; Tojo et al., 2014).

Compounds **a3** and **a12** should poor biological activity. These two compounds contained phenyl groups with strong electron withdrawing groups in the para position. Preliminary docking experiments in **Figure 4** suggested that due to the lack of methylene group, insertion of the phenyl groups into the hydrophobic pocket consisting L234 and Y126 residues was difficult. In addition, the distance between the triazolegroup and the HEM ion is 4.7Å and 6.5Å, respectively, which is also consistent with the low activity of **a3** and **a12**.

TABLE 1 | IDO1 inhibitory activities of compounds **a1–a22**.

Compd no.	n	R^1	R^2	R^3	R^4	IC_{50} (μM)
						IDO1
a1	0	H	F	H	H	4.25 ± 0.08
a2	0	H	H	Cl	H	2.90 ± 0.37
a3	0	H	H	F	H	18.19 ± 1.47
a4	0	OCH_3	H	H	H	1.32 ± 0.17
a5	0	Cl	H	H	H	4.88 ± 0.42
a6	0	H	H	H	H	0.77 ± 0.13
a7	0	H	Cl	H	H	2.82 ± 0.26
a8	0	I	H	H	H	2.50 ± 0.43
a9	0	H	OCH_3	H	H	1.41 ± 0.08
a10	0	H	H	Br	H	2.79 ± 0.46
a11	0	Br	H	H	H	1.00 ± 0.49
a12	0	H	H	CF_3	H	>100
a13	1	F	H	H	H	4.62 ± 0.19
a14	1	CH_3	H	H	H	0.79 ± 0.21
a15	1	H	H	CH_3	H	0.59 ± 0.05
a16	1	I	H	H	H	1.51 ± 0.11
a17	1	H	Br	H	H	0.37 ± 0.02
a18	1	H	H	H	H	0.56 ± 0.16
a19	1	Br	H	H	H	1.50 ± 0.45
a20	1	H	OCH_3	H	H	0.76 ± 0.07
a21	1	H	Br	H	Br	0.68 ± 0.44
a22	2	H	H	F	H	0.81 ± 0.35
icotinib						2.57 ± 0.44

Cytotoxicity Study

Next, several compounds with submicromolar level of IDO1 inhibitory activities were chosen for further study. The results are given in **Table 2**. Human renal epithelial cell (293T) was chosen for CCK-8 assay to evaluation the bioactivity of these compounds. As shown in **Table 2**, human renal epithelial cell 293T showed poor sensitivity to **a6** with IC_{50} values of 42.79 ± 11.81 μM, and icotinib-triazole derivatives such as **a15** and **a21** exhibited stronger killing effects on the cell line with IC_{50} values of 0.16 ± 0.05 μM and 0.74 ± 0.15 μM, respectively. Among these compounds, **a17** and **a18** showed medium cytotoxic effects on the 293T cell line, and the IC_{50} values of **a7** and **a18** on the cell line were 3.10 ± 0.20 μM and 3.08 ± 0.59 μM, respectively.

CONCLUSION

In summary, a series of icotinib derivatives containing 1,2,3-triazole ringsprepared and evaluated for the inhibition of IDO1. Most of the compounds exhibited better IDO1 inhibitory activities than the parent icotinib. For example, submicromolar levels of IC_{50} were observed for compound **a17** and **a18**, with the IC_{50} value of 0.37 μM and 0.56 μM, respectively. Docking experiments suggest that icotinib-1,2,3-triazole derivatives are potential IDO1 inhibitors that preferentially bind to the ferrous

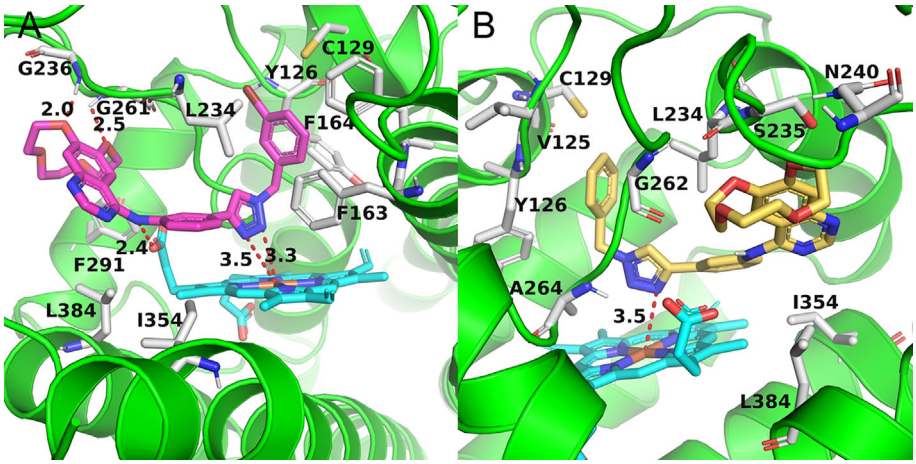


FIGURE 3 | The binding mode of compounds in complex with IDO1. Theprotein is represented by a green cartoon, while compound **a17** (pink, **A**) and compound **a18** (yellow, **B**) are represented as sticks. The hydrogen bonds are colored in red dash.

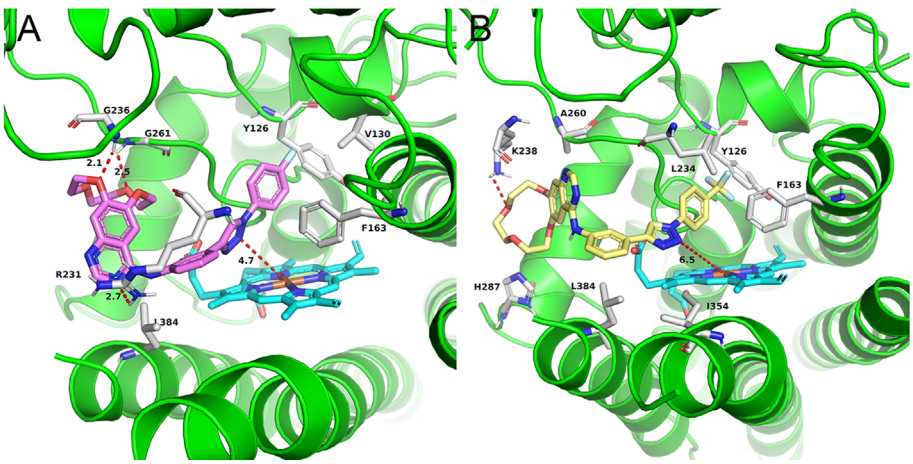


FIGURE 4 | The binding mode of compounds in complex with IDO1. Theprotein is represented by a green cartoon, while compound **a3** (pink, **A**) and compound **a12** (yellow, **B**) are represented as sticks. The hydrogen bonds are colored in red dash.

form of IDO1 by forming coordinate bond with the haem iron. However, considering the fact that several candidates are currently undergoing clinical trials but none of these has been approved so far, the identification of potent and clinically useful IDO1 inhibitors is still an open challenge. In addition, some toxicity problem arose when triazole functionality was introduced to icotinib, suggesting that one should be very careful when introducing additional pharmacophores into a known drug especially when the mode of interaction was altered after the introduction of the additional functional group. Ideally, the designed compounds should show

TABLE 2 | Cytotoxicity of selected compounds.

Compd no.	IC_{50} (μM)	
	IDO1	293T
a6	0.77 ± 0.13	42.79 ± 11.81
a14	0.79 ± 0.21	3.35 ± 0.89
a15	0.59 ± 0.05	0.16 ± 0.05
a17	0.37 ± 0.02	3.10 ± 0.20
a18	0.56 ± 0.16	3.08 ± 0.59
a20	0.76 ± 0.07	3.30 ± 1.30
a21	0.68 ± 0.44	0.74 ± 0.15
a22	0.81 ± 0.35	2.60 ± 0.17

significant toxicity against the cancer cell on one hand, and good safety against normal cells on the other. We are now designing new structures using the scaffold hopping strategy, and the results will be presented in due time.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Scientific and Technological Project of Henan Province (No. 192102310142), the Tianjin Research Innovation Project for Postgraduate Students (No. 2019YJSB077), the Shaanxi University of Chinese Medicine (No. 2020XG01), and the Subject Innovation Team of Shaanxi University of Chinese Medicine (No. 2019-PY02). Y-ML acknowledged the financial support from the National Natural Science Foundation of China (NSFC 21672106).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.579024/full#supplementary-material.

REFERENCES

Chen, S., Guo, W., Liu, X., Sun, P., Wang, Y., Ding, C., et al. (2019). Design, synthesis and antitumor study of a series of N-Cyclic sulfamoylaminoethyl substituted 1,2,5-oxadiazol-3-amines as new indoleamine 2, 3-dioxygenase 1 (IDO1) inhibitors. *Eur. J. Med. Chem.* 179, 38–55. doi: 10.1016/j.ejmech.2019.06.037

Crosignani, S., Bingham, P., Bottemanne, P., Cannelle, H., Cauwenberghs, S., Cordonnier, M., et al. (2017). Discovery of a novel and selective indoleamine 2,3-dioxygenase (IDO-1) inhibitor 3-(5-Fluoro-1H-indol-3-yl)pyrrolidine-2,5-dione (EOS200271/PF-06840003) and its characterization as a potential clinical candidate. *J. Med. Chem.* 60, 9617–9629. doi: 10.1021/acs.jmedchem.7b00974

De Souza, T. B., Caldas, I. S., Paula, F. R., Rodrigues, C. C., Carvalho, D. T., and Dias, D. F. (2020). Synthesis, activity, and molecular modeling studies of 1,2,3-triazole derivatives from natural phenylpropanoids as new trypanocidal agents. *Chem. Biol. Drug Des.* 95, 124–129. doi: 10.1111/cbdd.13628

Dounay, A. B., Tuttle, J. B., and Verhoest, P. R. (2015). Challenges and Opportunities in the Discovery of New Therapeutics Targeting the Kynurenine Pathway. *J. Med. Chem.* 58, 8762–8782. doi: 10.1021/acs.jmedchem.5b00461

Efimov, I., Basran, J., Thackray, S. J., Handa, S., Mowat, C. G., and Raven, E. L. (2011). Structure and reaction mechanism in the heme dioxygenases. *Biochemistry* 50, 2717–2724. doi: 10.1021/bi101732n

Hong, V., Steinmetz, N. F., Manchester, M., and Finn, M. G. (2010). Labeling live cells by copper-catalyzed alkyne–azide click chemistry. *Bioconjug. Chem.* 21, 1912–1916. doi: 10.1021/bc100272z

Kumar, S., Waldo, J. P., Jaipuri, F. A., Marcinowicz, A., Van Allen, C., Adams, J., et al. (2019). Discovery of clinical candidate (1R4r)-4-((R)-2-((S)-6-Fluoro-5H-imidazo[5,1-a]isoindol-5-yl)-1-hydroxyethy)cyclohexan-1-ol (navoximod), a potent and selective inhibitor of indoleamine 2,3-dioxygenase 1. *J. Med. Chem.* 62, 6705–6733. doi: 10.1021/acs.jmedchem.9b00662

Lewis-Ballester, A., Pham, K. N., Batabyal, D., Karkashon, S., Bonanno, J. B., Poulos, T. L., et al. (2017). Structural insights into substrate and inhibitor binding sites in human indoleamine 2,3-dioxygenase 1. *Nat. Commun.* 8, 1693–1693. doi: 10.1038/s41467-017-01725-8

Liang, S., Xu, Y., Tan, F., Ding, L., Ma, Y., and Wang, M. (2018). Efficacy of icotinib in advanced lung squamous cell carcinoma. *Cancer. Med.* 7, 4456–4466. doi: 10.1002/cam4.1736

Lin, S.-Y., Yeh, T.-K., Kuo, C.-C., Song, J.-S., Cheng, M.-F., Liao, F.-Y., et al. (2016). Phenyl Benzenesulfonylhydrazides Exhibit Selective Indoleamine 2,3-Dioxygenase Inhibition with Potent in Vivo Pharmacodynamic Activity and Antitumor Efficacy. *J. Med. Chem.* 59, 419–430. doi: 10.1021/acs.jmedchem.5b01640

Majeed, R., Sangwan, P. L., Chinthakindi, P. K., Khan, I., Dangroo, N. A., Thota, N., et al. (2013). Synthesis of 3-O-propargylated betulinic acid and its 1,2,3-triazoles as potential apoptotic agents. *Eur. J. Med. Chem.* 63, 782–792. doi: 10.1016/j.ejmech.2013.03.028

Malachowski, W. P., Winters, M., Duhadaway, J. B., Lewis-Ballester, A., Badir, S., Wai, J., et al. (2016). O-alkylhydroxylamines as rationally-designed mechanism-based inhibitors of indoleamine 2,3-dioxygenase-1. *Eur. J. Med. Chem.* 108, 564–576. doi: 10.1016/j.ejmech.2015.12.028

Mao, L.-F., Xu, G.-Q., Sun, B., Jiang, Y.-Q., Dong, W.-P., Zhang, S.-T., et al. (2017). Design, Synthesis and Antibacterial Evaluation of Novel 1,2,3-Triazole Derivatives Incorporating 3′-Deoxythymidine. *J. Chem. Res.* 41, 645–649. doi: 10.3184/174751917X15094552081189

Morgan, M. A., Parsels, L. A., Kollar, L. E., Normolle, D. P., Maybaum, J., and Lawrence, T. S. (2008). The combination of epidermal growth factor receptor inhibitors with gemcitabine and radiation in pancreatic cancer. *Clin. Cancer Res.* 14, 5142–5149. doi: 10.1158/1078-0432.CCR-07-4072

Muller, A. J., Duhadaway, J. B., Donover, P. S., Sutanto-Ward, E., and Prendergast, G. C. (2005). Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. *Nat. Med.* 11, 312–319. doi: 10.1038/nm1196

Nelp, M. T., Kates, P. A., Hunt, J. T., Newitt, J. A., Balog, A., Maley, D., et al. (2018). Immune-modulating enzyme indoleamine 2,3-dioxygenase is effectively inhibited by targeting its apo-form. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3249–3254. doi: 10.1073/pnas.1719190115

Qian, S., He, T., Wang, W., He, Y., Zhang, M., Yang, L., et al. (2016). Discovery and preliminary structure-activity relationship of 1H-indazoles with promising indoleamine-2,3-dioxygenase 1 (IDO1) inhibition properties. *Bioorg. Med. Chem.* 24, 6194–6205. doi: 10.1016/j.bmc.2016.10.003

Röhrig, U. F., Majjigapu, S. R., Grosdidier, A., Bron, S., Stroobant, V., Pilotte, L., et al. (2012). Rational design of 4-aryl-1,2,3-triazoles for indoleamine 2,3-dioxygenase 1 inhibition. *J. Med. Chem.* 55, 5270–5290. doi: 10.1021/jm300260v

Serafini, M., Torre, E., Aprile, S., Grosso, E. D., Gesu, A., Griglio, A., et al. (2020). Discovery of highly potent benzimidazole derivatives as indoleamine 2,3-dioxygenase-1 (IDO1) inhibitors: from structure-based virtual streening to in vivo pharmacodynamic activity. *J. Med. Chem.* 63, 3047–3065. doi: 10.1021/acs.jmedchem.9b01809

Soliman, H. H., Antonia, S., Sullivan, D., Vanahanian, N., and Link, C. (2009). Overcoming coming tumor antigen anergy in human malignancies using the novel indeolamine 2,3-dioxygenase (IDO) enzyme inhibitor, 1-methyl-D-tryptophan (1MT). *J. Clin. Oncol.*

Song, X., Sun, P., Wang, J., Guo, W., Wang, Y., Meng, L.-H., et al. (2020). Design, synthesis, and biological evaluation of 1,2,5-oxadiazole-3-carboximidamide derivatives as novel indoleamine-2,3-dioxygenase 1 inhibitors. *Eur. J. Med. Chem.* 189, 112059–112059. doi: 10.1016/j.ejmech.2020.112059

Takikawa, O., Yoshida, R., Kido, R., and Hayaishi, O. (1986). Tryptophan degradation in mice initiated by indoleamine 2,3-dioxygenase. *J. Biol. Chem.* 261, 3648–3653.

Takikawa, O. (2005). Biochemical and medical aspects of the indoleamine 2,3-dioxygenase-initiated L-tryptophan metabolism. *Biochem. Biophys. Res. Commun.* 338, 12–19. doi: 10.1016/j.bbrc.2005.09.032

Thomopoulou, P., Sachs, J., Teusch, N., Mariappan, A., Gopalakrishnan, J., and Schmalz, H.-G. (2015). New Colchicine-Derived Triazoles and Their Influence on Cytotoxicity and Microtubule Morphology. *ACS Med. Chem. Lett.* 7, 188–191. doi: 10.1021/acsmedchemlett.5b00418

Tojo, S., Kohno, T., Tanaka, T., Kamioka, S., Ota, Y., Ishii, T., et al. (2014). Crystal Structures and Structure–Activity Relationships of Imidazothiazole Derivatives as IDO1 Inhibitors. *ACS Med. Chem. Lett.* 5, 1119–1123. doi: 10.1021/ml500247w

Wu, Y., Xu, T., Liu, J., Ding, K., and Xu, J. (2017). Structural insights into the binding mechanism of IDO1 with hydroxylamidine based inhibitor INCB14943. *Biochem. Biophys. Res. Commun.* 487, 339–343. doi: 10.1016/j.bbrc.2017.04.061

Yang, J.-J., Zhou, C., Huang, Y., Feng, J., Lu, S., Song, Y., et al. (2017). Icotinib versus whole-brain irradiation in patients with EGFR-mutant non-small-cell lung cancer and multiple brain metastases (BRAIN): a multicentre, phase 3, open-label, parallel, randomised controlled trial. *Lancet Respir. Med.* 5, 707–716. doi: 10.1016/S2213-2600(17)30262-X

Yue, E. W., Douty, B., Wayland, B., Bower, M., Liu, X., Leffet, L., et al. (2009). Discovery of potent competitive inhibitors of indoleamine 2,3-dioxygenase with in vivo pharmacodynamic activity and efficacy in a mouse melanoma model. *J. Med. Chem.* 52, 7364–7367. doi: 10.1021/jm900518f

Yue, E. W., Sparks, R., Polam, P., Modi, D., Douty, B., Wayland, B., et al. (2017). INCB24360 (Epacadostat), a Highly Potent and Selective Indoleamine-2,3-dioxygenase 1 (IDO1) Inhibitor for Immuno-oncology. *ACS. Med. Chem. Lett.* 8, 486–491. doi: 10.1021/acsmedchemlett.6b00391

Zhang, S., Fu, Y., Wang, D., and Wang, J. (2018). Icotinib enhances lung cancer cell radiosensitivity in vitro and in vivo by inhibiting MAPK/ERK and AKT activation. *Clin. Exp. Pharmacol. Physiol.* 45, 969–977. doi: 10.1111/1440-1681.12966

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mao, Wang, Zhao, Xu, Yao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 11 November 2020
doi: 10.3389/fphar.2020.579768

Discovery of Novel IDH1 Inhibitor Through Comparative Structure-Based Virtual Screening

Yuwei Wang[1,2], Shuai Tang[3], Huanling Lai[2], Ruyi Jin[1], Xu Long[1], Na Li[1], Yuping Tang[1], Hui Guo[1], Xiaojun Yao[2]* and Elaine Lai-Han Leung[2]**

[1] College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an, China, [2] State Key Laboratory of Quality Research in Chinese Medicine, Macau University of Science and Technology, Macau, China, [3] State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academic of Sciences, Shanghai, China

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Wenjin Li,
Shenzhen University, China
Ying Li,
The University of Texas Health Science Center at San Antonio, United States

***Correspondence:**
Hui Guo
guohui@sntcm.edu.cn
Xiaojun Yao
xjyao@must.edu.mo
Elaine Lai-Han Leung
lhleung@must.edu.mo

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: 03 July 2020
Accepted: 14 September 2020
Published: 11 November 2020

Citation:
Wang Y, Tang S, Lai H, Jin R, Long X, Li N, Tang Y, Guo H, Yao X and Leung EL-H (2020) Discovery of Novel IDH1 Inhibitor Through Comparative Structure-Based Virtual Screening. Front. Pharmacol. 11:579768. doi: 10.3389/fphar.2020.579768

IDH1 mutations occur in about 20–30% of gliomas and are a promising target for the treatment of cancer. In the present study, the performance of aIDH1^{R132H} was verified *via* glide-docking-based virtual screening. On the basis of the two crystal structures (5TQH and 6B0Z) with the best discriminating ability to identify IDH1^{R132H} inhibitors from a decoy set, a docking-based virtual screening strategy was employed for identifying new IDH1^{R132H} inhibitors. In the end, 57 structurally diverse compounds were reserved and evaluated through experimental tests, and 10 of them showed substantial activity in targeting IDH1^{R132H} (IC_{50} < 50 μM). Molecular docking technology showed that L806-0255, V015-1671, and AQ-714/41674992 could bind to the binding pocket composed of hydrophobic residues. These findings indicate that L806-0255, V015-1671, and AQ-714/41674992 have the potential as lead compounds for the treatment of IDH1-mutated gliomas through further optimization.

Keywords: IDH1, gliomas, molecular docking, virtual screening, docking-based virtual screening

INTRODUCTION

Isocitrate dehydrogenase 1 (IDH1) is a critical metabolic enzyme involved in the tricarboxylic acid cycle. This enzyme catalyzes the oxidative decarboxylation of isocitrate acid to aketoglutaric (a-KG) in an $NADP^+$-dependent manner by using divalent magnesium ion (Jiao et al., 2016), which is related to the progression of various tumors, including acute myeloid leukemia, gliomas, and other solid tumors (Yan and Reitman, 2010; Yen et al., 2016).

Somatic mutations of IDH1 have been frequently identified in many types of cancer, including approximately 80% of grade II-III gliomas, nearly 45% of secondary glioblastoma multiforme (GBM), and 33%-50% of adult primitive neuroectodermal tumors (Dang et al., 2009; Wang et al., 2013). IDH1

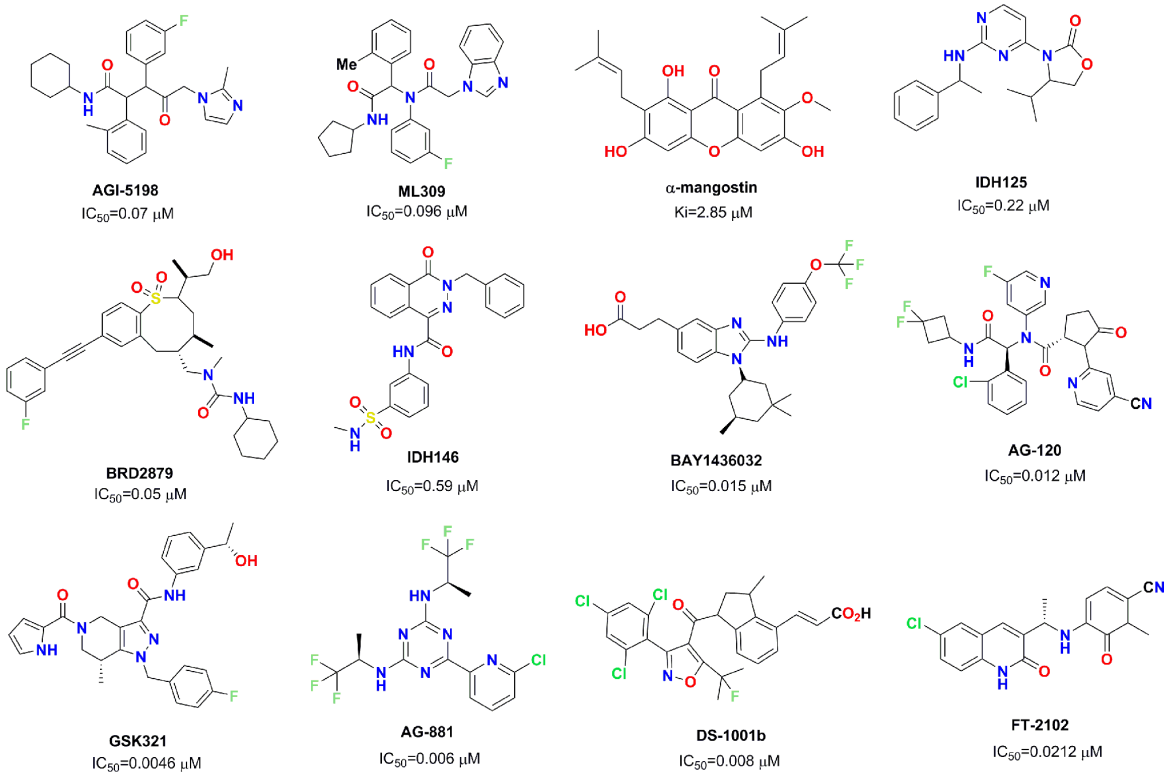


FIGURE 1 | Chemical skeleton of nine representative IDH1^{R132H} inhibitors.

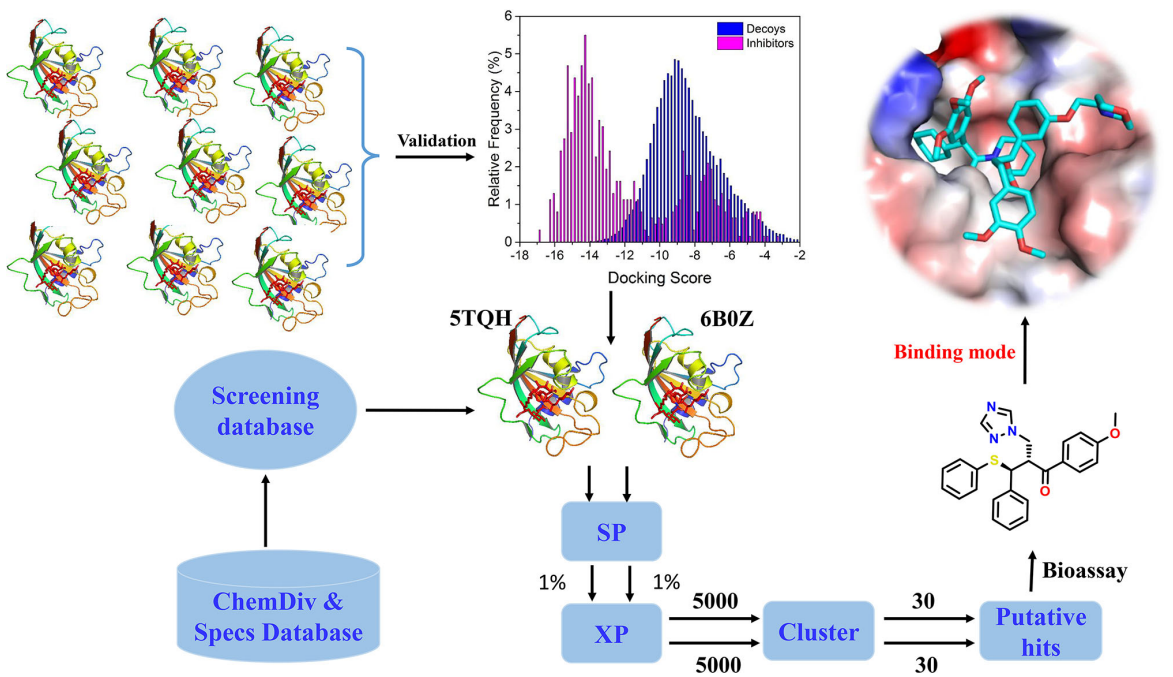


FIGURE 2 | The workflow of docking-based virtual screening and bioassay for IDH1^{R132H} inhibitor.

mutations have also been discovered in other cancers, such as colorectal cancer (Xu et al., 2011), acute myeloid leukemia (Parsons et al., 2008), and prostate cancer (Hartmann et al., 2009). Key amino acid residue Arg132 is the most common mutation in IDH1, which is located in the catalytic pocket (Dang et al., 2009). Specific mutations belong to heterozygous missense mutations and lead to a new form of IDH1 catalytic activity, which convert α-KG into an oncometabolite D2-hydroxyglutarate (Dang et al., 2009). The oncometabolite (D2-HG) is associated with tumorigenesis, which impairs hematopoietic differentiation and promotes leukemia by inducing the hypermethylation of histone and chromatin and preventing cell differentiation (Figueroa et al., 2010; Xu et al., 2011). Due to the IDH1 mutation, high levels of D2-HG are created that promote the occurrence and development of cancers, such as gliomas (Parsons et al., 2008) and acute myeloid leukemia (Mardis et al., 2009). Therefore, although the contribution of IDH1 mutants to carcinogenic properties has yet to be elucidated, IDH1 mutants have become therapeutic targets for cancer, especially AML.

Mutant IDH1 has become a very attractive therapeutic target in the field of antitumor drug discovery, and several pharmaceutical companies have attempted to develop novel small molecule inhibitors against mutant IDH1. So far, several small molecule inhibitors targeting mutant IDH1 enzymes have been developed (see **Figure 1**) (Rohle et al., 2013; Davis et al., 2014; Deng et al., 2015; Kim et al., 2015; Okoye-Okafor et al., 2015; Law et al., 2016; Chaturvedi et al., 2017; Xie et al., 2017; Popovici-Muller et al., 2018; Nakagawa et al., 2019; Caravella et al., 2020; Konteatis et al., 2020). Some of these have been studied in various preclinical models, and some are currently being evaluated in phase I/II clinical studies for different tumor pathologies with IDH1 enzyme mutations. AG-120 as the only mutant IDH1 inhibitor in clinic approved by the FDA that has shown encouraging clinical benefits with a total overall response rate of 42% for advanced hematological malignancies (Foran et al., 2019). In light of these encouraging finding, we employed docking-based virtual screening to identify active hits with novel skeleton for targeting mutant IDH1.

Structure-based virtual screening is now widely used in early-stage drug discovery (Sheisi et al., 2019), and has been applied to the discovery of IDH1 inhibitors. To date, there have been several attempts to identify potential IDH1 inhibitors by using structure-based virtual screening in terms of the reported crystal structures of the IDH1 complex (Zou et al., 2016; Zheng et al., 2017; Zou et al., 2018). In 2016, by using a docking-based virtual screening strategy (PDB: 4UMX), Zou et al. identified a series of IDH1 inhibitor FX-03 with IC_{50} values of 55.50 μM and 68.38 μM in HEK-293T cells transfected with IDH1 R132H and IDH1 R132C, respectively (Zou et al., 2016). Importantly, FX-03 exhibited significant selectivity between the $IDH1^{WT}$ and $IDH1^{R132H}$ mutants. In 2017, Zheng et al. discovered a natural product, clomifene, as an effective inhibitor against the $IDH1^{R132H}$ mutant with a K_d value of 18.45 μM by using docking-based virtual screening (PDB: 4UMX) (Zheng et al., 2017). They also proved that clomifene selectively inhibits mutant IDH1 activities *in vitro* and *in vivo* models. It should be noted that, although these studies have identified several $IDH1^{R132H}$ inhibitors, they used the same $IDH1^{R132H}$ crystal structure in structure-based virtual screening. Considering the difference in binding mode after the binding of various ligands, comparing the virtual screening capabilities of different $IDH1^{R132H}$ crystal structures based on docking-based virtual screening appears to a more reasonable strategy to discover potential $IDH1^{R132H}$ inhibitors.

In the present study, the performance of docking-based virtual screening for nine crystal structures of $IDH1^{R132H}$ were compared through a combination of docking power and screening power. Two best performing $IDH1^{R132H}$ complexes were employed to identify potential $IDH1^{R132H}$ inhibitors with diverse structures from ChemDiv (http://www.chemdiv.com) and Specs (http://www.specs.net) databases. Followed by further examination and verification, a series of compounds with novel skeleton were addressed and could be used as $IDH1^{R132H}$ inhibitors. The overall workflow was shown in **Figure 2**.

MATERIALS AND METHODS

Preparation of Crystal Structures and Data Sets

The crystal structures of the $IDH1^{R132H}$ in complex with an inhibitor were downloaded from the PDB database (http://www.rcsb.org), including 4UMX, 5L57, 5L58, 5LGE, 5SUN, 5SVF, 5TQH, 6ADG, and 6B0Z. For each complex, the *Protein Preparation Wizard* module in Schrödinger 2015 (Schrödinger, LLC, New York, NY, 2015) was applied to add hydrogen and missing side chains, remove all water molecules, assign protonation states and partial charges through OPLS2005 force field (Jorgensen et al., 1996), and minimize all heavy atoms until the root-mean-square deviation (RMSD) was reached ≤0.3 Å.

To evaluate the virtual screening capability of different crystal structures, 423 actives were directly extracted from the PubChem database and served as a validation data set (https://pubchem.ncbi.nlm.nih.gov/bioassay/1344832#section=Top), and their decoys, generated by DUD•E (Mysinger et al., 2012), were considered as a decoy data set. In total, 23,900 decoys were generated.

Evaluate the Performance of Each Structure

In order to discover the favorable crystal structure for virtual screening, the docking performance of each $IDH1^{R132H}$ structure

TABLE 1 | The summary of the docking power of molecular docking in glide for nine $IDH1^{R132H}$ crystal structures.

PDB	Ligand	SP		XP	
		Docking score	RMSD	Docking score	RMSD
4UMX	VVS	-7.13	2.16	-7.04	2.10
5L57	6N3	-8.50	2.32	-9.23	2.56
5L58	6MX	-9.38	1.56	-9.97	1.59
5LGE	6VN	-6.85	0.90	-7.18	1.79
5SUN	70Q	-10.31	5.12	-10.57	5.11
5SVF	70P	-9.74	1.06	-12.94	0.44
5TQH	7J2	-12.75	0.95	-16.12	0.60
6ADG	9UO	-5.74	0.66	-5.61	1.39
6B0Z	C81	-12.53	0.45	-15.52	1.07

was systematically evaluated. All actives and decoys were preprepared using the LigPrep (LigPrep, Schrödinger, LLC, New York, NY, 2015) module in the Schrödinger package. The possible ionized states of each compound were calculated by using Epik (Shelley et al., 2007) at pH= 7.0 ± 2.0. The chirality of the $IDH1^{R132H}$ inhibitors with 3D structures were preserved, while the chirality of the decoys was determined from 3D structures based on the different combinations. The stereoisomers for each

TABLE 2 | The summary of the screening power of molecular docking in glide for nine $IDH1^{R132H}$ crystal structures for validation set.

PDB ID	SP Precision								XP Precision							
	p value	AUC-ROC	RIE	$EF^{1\%}$	$EF^{2\%}$	$EF^{5\%}$	$EF^{10\%}$	$EF^{20\%}$	*p* value	AUC-ROC	RIE	$EF^{1\%}$	$EF^{2\%}$	$EF^{5\%}$	$EF^{10\%}$	$EF^{20\%}$
4UMX	$3.51x10^{-100}$	0.53	0.37	0.47	0.35	0.33	0.33	0.43	$1.16x10^{-45}$	0.88	7.5	13	14	9.3	6.2	3.9
5L57	$1.84x10^{-6}$	0.7	1.06	0.95	0.95	0.95	0.97	1.4	$6.15x10^{-8}$	0.76	2.15	0.71	1.3	2	2.5	2.4
5L58	$5.20x10^{-21}$	0.76	2.21	0.94	2.4	2.6	1.9	2	$1.81x10^{-15}$	0.82	4.52	6.8	7.3	5.3	4	2.9
5LGE	$1.25x10^{-3}$	0.72	1.31	0.71	0.83	1	1.4	1.8	$1.08x10^{-10}$	0.77	2.61	3.8	3.1	2.5	2.6	2.4
5SUM	$5.67x10^{-17}$	0.77	1.86	0.71	1.4	1.6	1.9	2.2	$2.50x10^{-16}$	0.84	3.48	0.71	1.5	3.5	4.2	3.6
5SVF	$7.33x10^{-87}$	0.91	10.38	31	24	12	7.2	4	$8.91x10^{-87}$	0.95	14.95	62	39	17	9	4.6
5TQH	$3.14x10^{-106}$	0.92	11.82	50	30	13	7.1	4.1	$6.09x10^{-102}$	0.96	15.76	74	43	18	9	4.6
6ADG	$1.12x10^{-50}$	0.9	9.4	31	22	11	6.4	4	$1.60x10^{-44}$	0.9	9.56	29	21	11	6.7	4
6B0Z	$2.42x10^{-135}$	0.96	12.88	46	30	15	8.3	4.7	$2.33x10^{-96}$	0.96	15.43	70	41	18	8.9	4.5

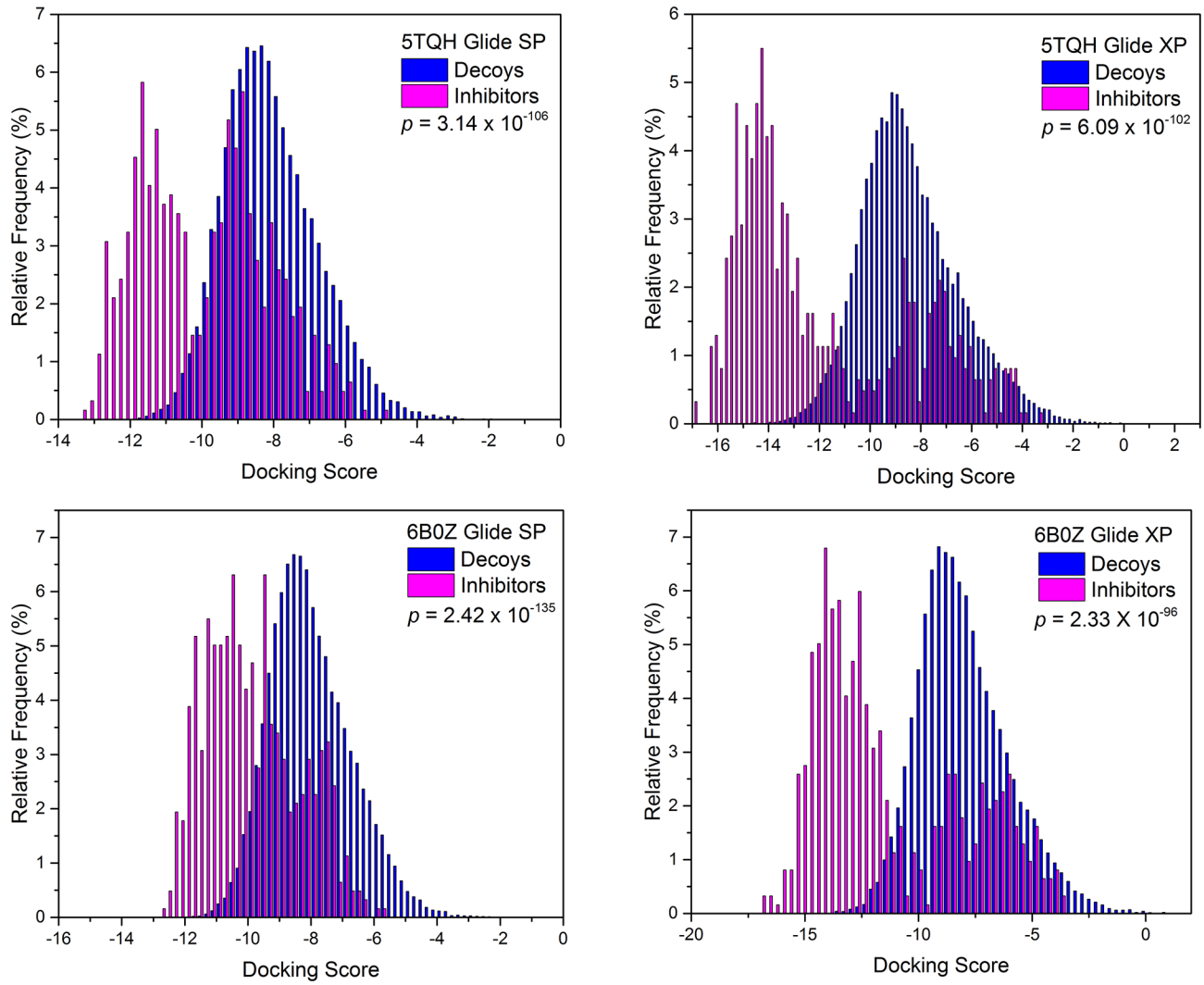


FIGURE 3 | The distributions of Glide docking scores of validation sets for two $IDH1^{R132H}$ crystal structures with the best screening power.

ligand generated, at most, 32, and the other parameters were set to default values. Subsequently, a grid box of each complex was generated by using the Receptor Grid Generation module of Schrödinger software, which was centered at the native ligand of the complex and defined as a similar size to the native ligand space. Finally, all chemicals in the validation set and decoy set were docked into the binding site of each $IDH1^{R132H}$ complex in turn and evaluated by using the standard precision (SP) and extra precision (XP) scoring function of Glide. In order to choose the best crystal structure of $IDH1^{R132H}$ for virtual screening, the enrichment factor (EF) (Halgren et al., 2004) was used to evaluate the virtual screening capability of each model, which was defined as the following Equation:

$$EF\frac{\frac{Hits_{set}}{n}}{\frac{Hits_{all}}{N}}$$

where $Hits_{set}$ is the number of actives in the selected subset n of the ranked database and $Hits_{all}$ is the total number of actives in the database. The model with the highest EF value was reserved and used to screen potential $IDH1^{R132H}$ inhibitors.

Docking-Based Virtual Screening

All compounds in the ChemDiv and Specs database were first preprocessed according to the method of the above step, and then screened by docking-based virtual screening against two $IDH1^{R132H}$ crystal structures (PDB ID: 5TQH and 6B0Z). After the possible ionized states and tautomer were calculated at pH=7.0 ± 2.0 by using Epik module, the chirality of each compound was determined from 3D structures; the stereoisomer for each ligand generated, at most, 32. The final virtual screening library was generated to include approximately 2 million compounds, and then initially filtered by Lipinski's Rule, removing ligands with reactive functional groups. Finally, docking-based virtual screening was employed by use of the high throughput virtual screening (HTVS) scoring function, SP scoring function, and XP scoring function of Glide in sequence. In the screening process at each step, 10% of the best compounds were reserved for further analysis.

Clustering Analysis

The reserved compounds after docking-based virtual screening were structurally clustered into 30 clusters by using K-means clustering on the MACCS structural keys in Canvas (Canvas, Schrödinger, LLC, New York, NY, 2015), and the compound in each cluster with the lowest docking score was selected. In the end, 60 chemicals were eventually submitted to purchase from Topscience Co., Ltd (https://www.tsbiochem.com).

Enzymatic Assay

The primary assay was carried out in 10 μL of base buffer (10 mM MgCl2, 20 mM Tris pH7.5, 150 mM NaCl, 0.05% (w/v) bovine serum albumin) containing 2.5 μL of the test compound, 5 μL of an enzyme solution (0.3 ng/μL mutant $IDH1^{R132H}$), and 2.5 μL of a substrate solution (4 mM α-KG, 16 μM NADPH). This assay added into a 384-well blank plate and then incubated at room temperature for 60 min. The secondary assay, with 5 μL of base

TABLE 3 | The molecular weight and docking score for putative hits.

ID	MW	Docking score (kcal/mol)
6470-0047	473.524	-15.53
G420-0655	460.55	-15.30
C798-1008	456.561	-15.25
E894-1127	469.539	-15.21
V004-0504	488.618	-15.17
L710-2843	447.49	-14.94
G389-1098	464.495	-14.72
S383-0082	412.438	-14.47
D103-1045	473.545	-14.45
C647-0812	484.551	-14.43
D491-0852	435.524	-14.40
V015-1671	491.426	-14.39
S733-2152	475.51	-14.37
V016-3750	453.515	-14.25
L710-0317	419.479	-14.20
L970-0181	487.529	-14.12
5782-4343	407.465	-14.11
V020-6264	478.931	-14.10
F019-2828	374.485	-14.09
M506-0358	404.44	-14.00
G741-1212	466.898	-13.96
S631-0764	421.513	-13.96
V022-0932	414.503	-13.95
D217-0418	416.454	-13.88
D336-7545	441.544	-13.78
V020-8255	472.54	-13.77
AQ-714/41674992	429.536	-13.62
M136-0372	474.949	-13.47
K781-3358	464.338	-13.39
3601-0061	426.452	-13.01
AQ-149/42126332	488.536	-15.34
V010-1281	478.555	-15.23
E867-1033	462.522	-15.14
V028-6550	490.53	-15.13
G800-0501	488.53	-15.09
C798-1007	476.979	-14.92
E894-1218	469.539	-14.81
V013-4787	435.524	-14.67
V025-9252	467.951	-14.51
AK-778/43465022	494.341	-14.45
V025-7538	496.485	-14.38
V003-2610	458.488	-14.18
K297-1090	474.576	-14.17
M136-0633	474.949	-14.11
8019-1512	410.398	-14.09
V001-8209	458.909	-14.05
F521-0664	486.526	-13.83
L487-0168	459.476	-13.78
G798-0506	434.534	-13.74
C647-0805	454.524	-13.68
J108-0614	432.478	-13.63
D349-0203	442.473	-13.62
L806-0255	457.842	-13.53
F815-0210	440.494	-13.53
C769-0129	438.54	-13.46
V020-4317	465.351	-13.40
G568-0082	454.973	-13.32
E867-0977	452.957	-13.27
V005-6943	477.534	-13.08

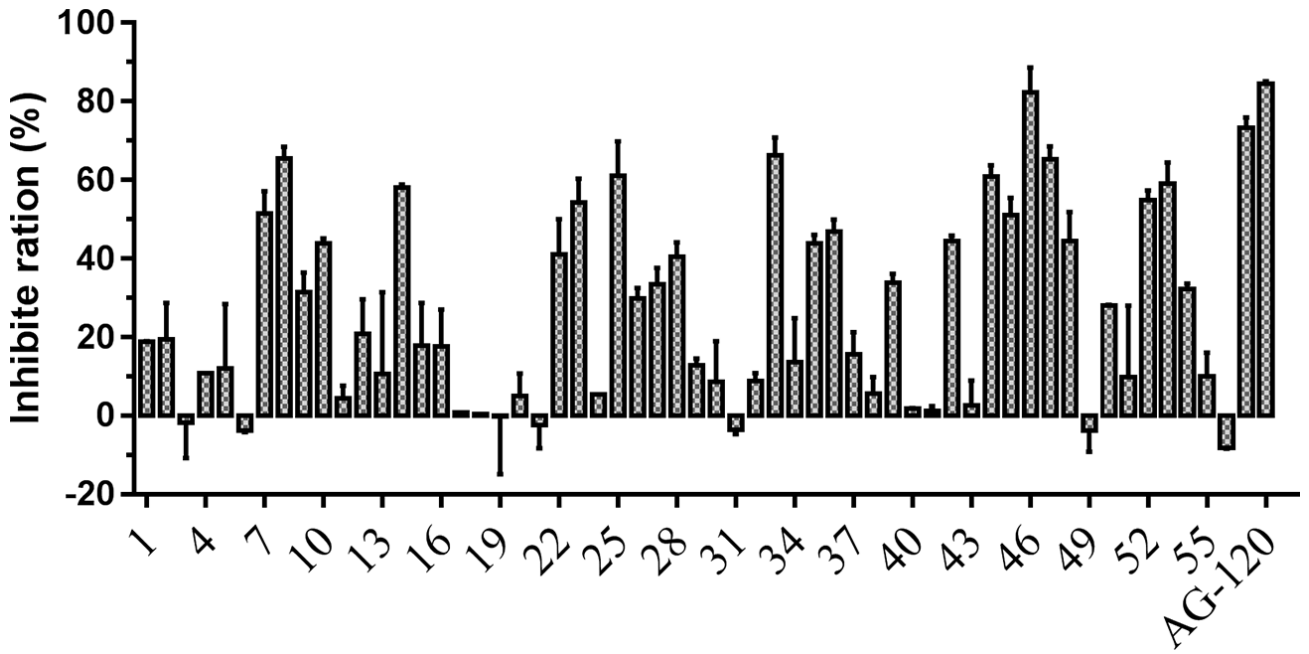


FIGURE 4 | Inhibitory activity of the 57 candidates at 50μM. The bars indicate the inhibitory activity of chemicals targeting $IDH1^{R132H}$. AG-120 at 100 nM was used as the positive control.

TABLE 4 | The summary of the inhibition ratio of 10 candidate compounds by using virtual screening.

No.	Database	PDB	Inhibition ratio (%) 50 (μM)	Enzymatic
				IC_{50} (μM)
C798-1007	ChemDiv	6B0Z	65.45±2.15	44.4±1.3
D491-0852	ChemDiv	5TQH	58.00±0.60	46.9±6.0
G568-0082	ChemDiv	6B0Z	54.30±4.30	41.9±8.0
G798-0506	ChemDiv	6B0Z	61.15±6.15	38.0±2.0
L806-0255	ChemDiv	6B0Z	66.25±3.25	28.3±2.5
V010-1281	ChemDiv	6B0Z	60.85±2.05	50.0±6.4
V015-1671	ChemDiv	5TQH	65.30±2.30	23.8±1.8
V016-3750	ChemDiv	5TQH	54.90±1.70	42.9±2.8
V025-9252	ChemDiv	6B0Z	59.15±3.75	45.5±3.1
AQ-714/41674992	Specs	5TQH	73.30±1.90	20.8±4.2
AG-120	–	–	84.40±0.50 (nM)	16.7±1.7 (nM)

buffer containing 15 μM resazurin and 0.01 unit diaphorase, was added to the entire plate and incubated at room temperature for 10 min. Florescence was read on a $SYNERGY^{H1}$ microplate reader (BioTek) at Ex 540 Em 590. Curve fitting for dose response IC_{50} was done using GraphPad Prism.

RESULTS AND DISCUSSION

Performance of the Nine $IDH1^{R132H}$ Complex

As a significant indicator of the docking reliability, docking power was used to reveal the binding pose of the experiment between small molecules and proteins, which was mainly evaluated after redocking with the RMSD value of the docking pose and native pose of the small molecule in the $IDH1^{R132H}$ complex. For each $IDH1^{R132H}$ complex, after the native inhibitor was separated from the corresponding complex and preprepared, it was redocked into the original binding site. The RMSD value between the native conformation of the inhibitor and the docked pose for each crystal structure was respectively computed, and RMSD ≤ 2.0 Å served as the evaluation standard to verify the docking reliability. It can be seen from **Table 1** that Glide docking could identify the near-native pose of most inhibitors in $IDH1^{R132H}$ crystal structures by using the XP or SP scoring function in Glide.

Next, screening power of glide docking was used to identify the reported inhibitors from the decoys in nine $IDH1^{R132H}$ complexes, and these were compared and calculated. In

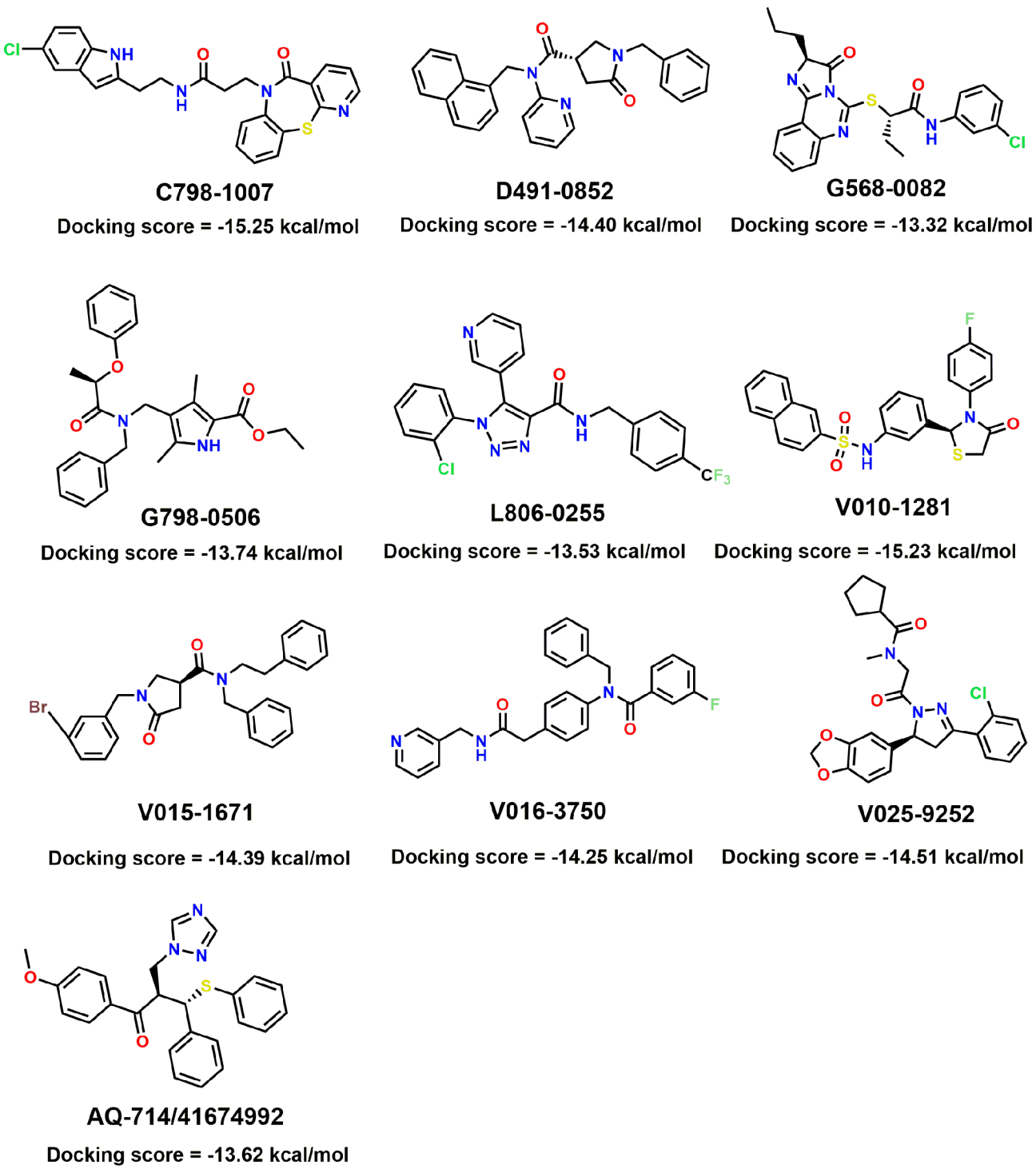


FIGURE 5 | Molecular structures of the 10 selected IDH1^{R132H} inhibitors by using based-docking virtual screening.

contrast with the docking power of glide docking, the screening power of each crystal structure is a more important index for the docking-based virtual screening process. Herein, we performed student's t test to evaluate the significant difference between the means of the two distributions of the Glide XP or SP scores for the known actives and decoys. It can be seen from **Table 2** that molecular docking of Glide can efficiently discriminate the IDH1^{R132H} inhibitors from the decoys in nine complexes of

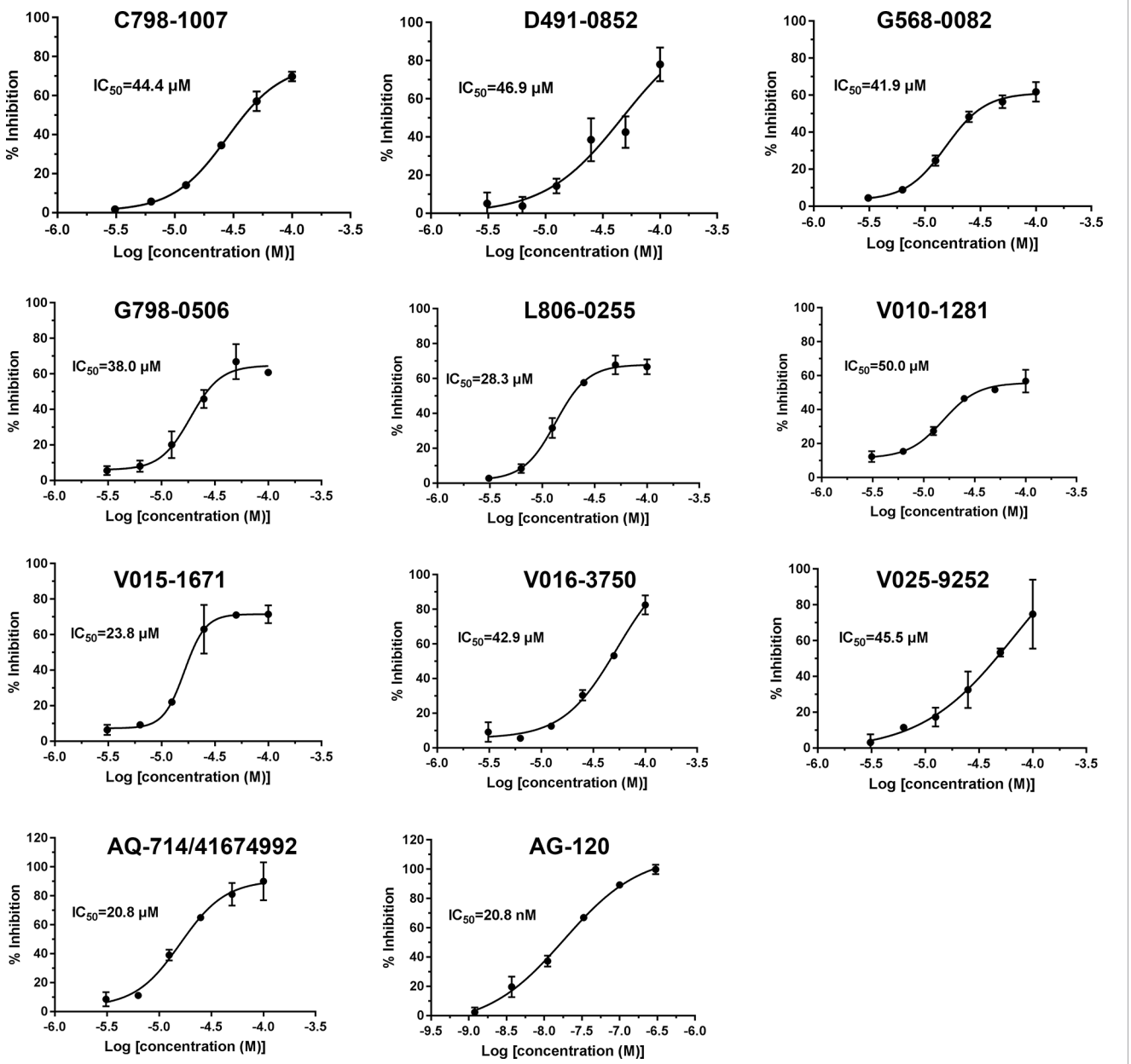


FIGURE 6 | IDH1^{R132H} enzymatic inhibition of 10 identified small molecule inhibitors.

IDH1^{R132H} based on the relatively low *p* value. The area under the receiver operating characteristic curve (AUC-ROC), EF, and Robust Initial Enhancement (RIE) were also employed to comprehensively evaluate the screening capabilities of each crystal structure. As shown in **Figure 3**, the best screening power (*p* value = 2.42×10^{-135}, AUC-ROC=0.96, and RIE= 12.88) was acquired by using SP scoring function and 6B0Z was reserved as the screening template. However, 5TQH exhibited the best screening power (*p* value = 6.09×10^{-102}, AUC-ROC=0.96, and RIE= 15.76) in XP scoring function, which was also retained as a screening complex. Our results suggest that that it is necessary to compare the performance of different complexes in the process of virtual screening.

Structure-Based Virtual Screening

The overall workflow of structure-based virtual screening was shown in **Figure 2**. The Specs and ChemDiv database, which consisted of more than 2,100,000 compounds, have been used for virtual screening of small molecule databases. Firstly, Lipinski's rules of five was employed to filter compounds that did not meet the criteria, and then these compounds containing PAINS substructures were also removed. A total of 1.46 million

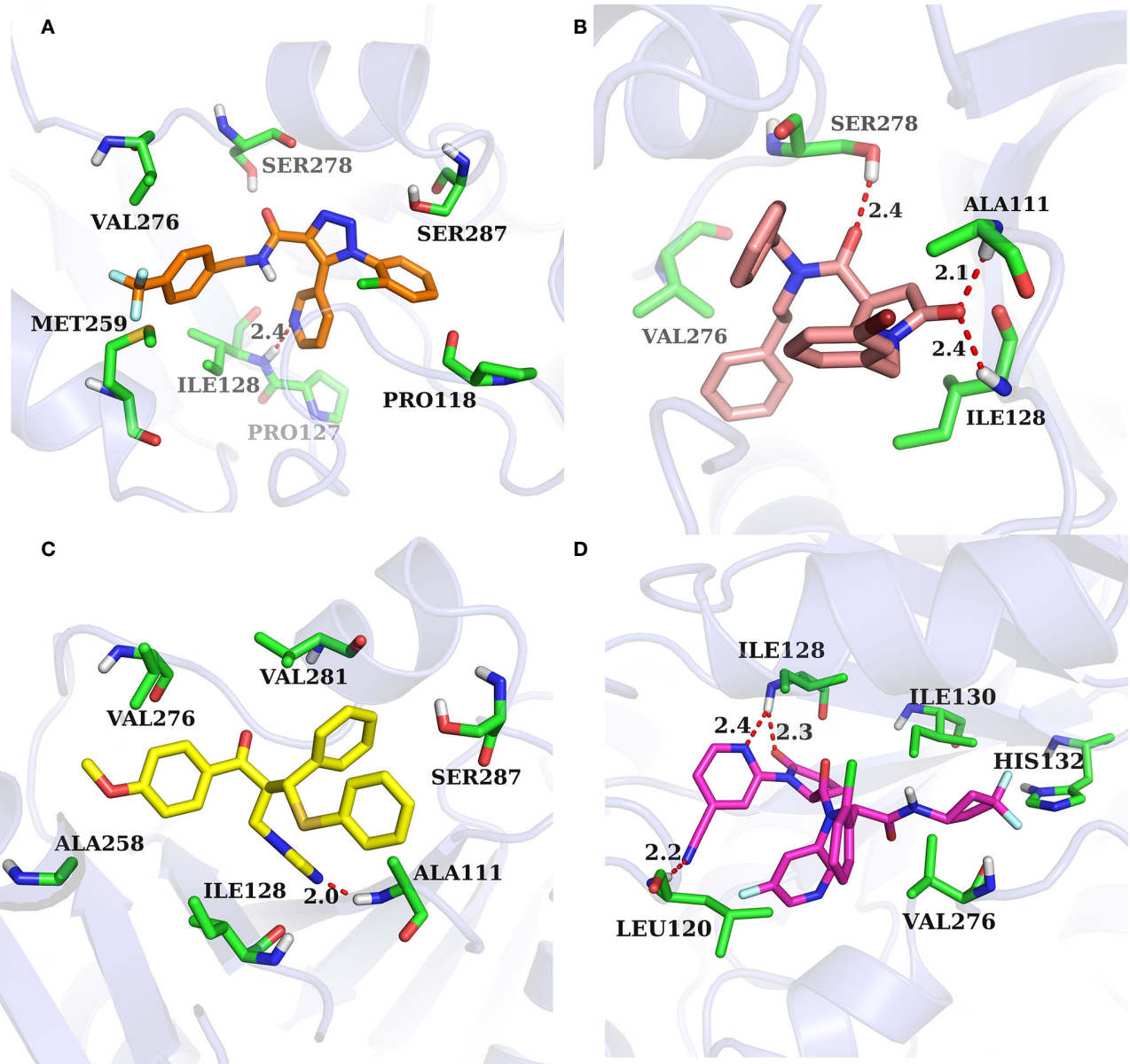


FIGURE 7 | Binding mode of identified IDH1^{R132H} inhibitor and AG-120. **(A)** L806-0255. **(B)** V015-1671. **(C)** AQ-714/41674992. **(D)**AG-120.

compounds were retained. According to the MACCS structural fingerprint, residual chemicals were structurally clustered in 30 clusters *via* K-means clustering in Canvas, and the chemical with the lowest docking score in each cluster was retained was retained (see **Table 3**). Finally, a total of 57 chemicals were purchased and tested based on the docking-based virtual screening on two IDH1 complex (5TQH and 6B0Z).

IDH1^{R132H} Enzymatic Assay

To verify the inhibitory activity of screening compounds targeting IDH1^{R132H}, enzyme activity assay was performed. As shown in **Figure 4**, we found that 12 compounds (7, 8, 14, 23, 25, 33, 44, 46, 47, 52, 53, and 57) exhibited over 50% inhibition at 50 μM. These 12 ligands were submitted to determine the IC_{50}. It can be seen from **Table 4** that 10 of them show $IC_{50} \leq 50$ μM. Molecular structures of the 10 selected compounds of IDH1^{R132H} are exhibited in **Figure 5**. The enzymatic curves and docking score for these 10 compounds against IDH1^{R132H} are depicted in **Figure 6**. Tanimoto coefficient (Tc) (Willett and Winterman, 1986; Willett et al., 1986), in terms of the ECFP4 fingerprint, was calculated to compare the structural similarity between 10 compounds and reported inhibitors. As shown in **Figure S1**, we can find that putative hits have low similarity with reported inhibitors (Tc < 0.2). Therefore, these compounds are

structurally new and have the potential to be promising leads for further optimizations.

Binding Mode Prediction

The binding pocket of IDH1 lies on the dimer interface and most of the reported compounds have been shown to bind to this allosteric site. In order to gain insight into the structural basis of the identified IDH1^{R132H} inhibitor, the binding mode of the three compounds was compared with AG-120. As shown in **Figure 7**, three molecules that could be docked into the binding pocket consisted of hydrophobic residues in a similar manner to AG-120, and formed intermolecular hydrogen bonds with key residues, which stabilized the complex. L806-0255 and V015-1671 form a key hydrogen bond with ILE128, which is consistent with AG-120. In addition, V015-1671 and AQ-714/41674992 also form a key hydrogen bond with ALA111. Moreover, the hydrophobic contacts formed between surrounded residues, such as VAL276, SER278, SER287, ILE128, PRO118, and compounds also contribute to enhanced binding of the small molecule inhibitor to IDH1 ^{R132H}. Therefore, the above results suggested that L806-0255, V015-1671, and AQ-714/41674992 could bind to IDH1^{R132H}.

CONCLUSIONS

In the present work, we first verified the performance of IDH1^{R132H} by using glide-docking-based virtual screening and discovered two crystal structures with the most credible screening ability. Based on the best performing crystal structure, docking-based virtual screening was performed to identify new IDH1^{R132H} inhibitors. A total of 57 potential hits were purchased and their activity against IDH1^{R132H} was addressed, and 10 of them exhibited anti-IDH1^{R132H} activity.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

HG, XY, and EL conceived this study and revised the manuscript. YW, ST, HL, RJ, and XL carried out the experiments and analyzed the experimental data. YW and XY wrote the manuscript. All authors reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (82003653), the Shaanxi University of Chinese Medicine (Project No. 2020XG01), the Subject Innovation Team of Shaanxi University of Chinese Medicine (Project No. 2019-PY02), and Macau Science and Technology Development Fund (Project Nos. 005/2014/AMJ, 082/2015/A3, 046/2016/A2 and 086/2015/A3).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.579768/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | The Tanimoto coefficient matrix based on ECFP4 fingerprint of identified IDH1^{R132H} inhibitor *vs* reported IDH1^{R132H} inhibitor.

REFERENCES

Caravella, J. A., Lin, J., Diebold, R. B., Campbell, A.-M., Ericsson, A., Gustafson, G., et al. (2020). Structure-Based Design and Identification of FT-2102 (Olutasidenib), a Potent Mutant-Selective IDH1 Inhibitor. *J. Med. Chem.* 63 (4), 1612–1623. doi: 10.1021/acs.jmedchem.9b01423

Chaturvedi, A., Herbst, L., Pusch, S., Klett, L., Goparaju, R., Stichel, D., et al. (2017). Pan-mutant-IDH1 inhibitor BAY1436032 is highly effective against human IDH1 mutant acute myeloid leukemia in vivo. *Leukemia* 31 (10), 2020–2028. doi: 10.1038/leu.2017.46

Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739. doi: 10.1038/nature08617

Davis, M. I., Gross, S., Shen, M., Straley, K. S., Pragani, R., Lea, W. A., et al. (2014). Biochemical, cellular, and biophysical characterization of a potent inhibitor of mutant isocitrate dehydrogenase IDH1. *J. Biol. Chem.* 289 (20), 13717–13725. doi: 10.1074/jbc.M113.511030

Deng, G., Shen, J., Yin, M., McManus, J., Mathieu, M., Gee, P., et al. (2015). Selective inhibition of mutant isocitrate dehydrogenase 1 (IDH1) via disruption of a metal binding network by an allosteric small molecule. *J. Biol. Chem.* 290 (2), 762–774. doi: 10.1074/jbc.M114.608497

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18 (6), 553–567. doi: 10.1016/j.ccr.2010.11.015

Foran, J. M., DiNardo, C. D., Watts, J. M., Stein, E. M., De Botton, S., Fathi, A. T., et al. (2019). Ivosidenib (AG-120) in Patients with IDH1-Mutant Relapsed/ Refractory Myelodysplastic Syndrome: Updated Enrollment of a Phase 1 Dose Escalation and Expansion Study. *Blood* 134 (Supplement_1), 4254–4254. doi: 10.1182/blood-2019-123946

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* 47 (7), 1750–1759. doi: 10.1021/jm030644s

Hartmann, C., Meyer, J., Balss, J., Capper, D., Mueller, W., Christians, A., et al. (2009). Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta Neuropathol.* 118 (4), 469–474. doi: 10.1007/s00401-009-0561-9

Jiao, C., Jie, Y., and Peng, C. (2016). The Evolving Landscape in the Development of Isocitrate Dehydrogenase Mutant Inhibitors. *Mini Rev. Med. Chem.* 16 (16), 1344–1358. doi: 10.2174/1389557516666160609085520

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* 118 (45), 11225–11236. doi: 10.1021/ja9621760

Kim, H. J., Fei, X., Cho, S. C., Choi, B. Y., Ahn, H. C., Lee, K., et al. (2015). Discovery of alpha-mangostin as a novel competitive inhibitor against mutant isocitrate dehydrogenase-1. *Bioorg. Med. Chem. Lett.* 25 (23), 5625–5631. doi: 10.1016/j.bmcl.2015.10.034

Konteatis, Z., Artin, E., Nicolay, B., Straley, K., Padyana, A. K., Jin, L., et al. (2020). Vorasidenib (AG-881): A First-in-Class, Brain-Penetrant Dual Inhibitor of Mutant IDH1 and 2 for Treatment of Glioma. *ACS Med. Chem. Lett.* 11 (2), 101–107. doi: 10.1021/acsmedchemlett.9b00509

Law, J. M., Stark, S. C., Liu, K., Liang, N. E., Hussain, M. M., Leiendecker, M., et al. (2016). Discovery of 8-Membered Ring Sulfonamides as Inhibitors of Oncogenic Mutant Isocitrate Dehydrogenase 1. *ACS Med. Chem. Lett.* 7 (10), 944–949. doi: 10.1021/acsmedchemlett.6b00264

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl. J. Med.* 361 (11), 1058–1066. doi: 10.1056/NEJMoa0903840

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* 55 (14), 6582–6594. doi: 10.1021/jm300687e

Nakagawa, M., Nakatani, F., Matsunaga, H., Seki, T., Endo, M., Ogawara, Y., et al. (2019). Selective inhibition of mutant IDH1 by DS-1001b ameliorates aberrant histone modifications and impairs tumor activity in chondrosarcoma. *Oncogene* 38 (42), 6835–6849. doi: 10.1038/s41388-019-0929-9

Okoye-Okafor, U. C., Bartholdy, B., Cartier, J., Gao, E. N., Pietrak, B., Rendina, A. R., et al. (2015). New IDH1 mutant inhibitors for treatment of acute myeloid leukemia. *Nat. Chem. Biol.* 11 (11), 878–886. doi: 10.1038/nchembio.1930

Parsons, D. W., Jones, S., Zhang, X., Lin, J. C.-H., Leary, R. J., Angenendt, P., et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science (New York N.Y.)* 321 (5897), 1807–1812. doi: 10.1126/science.1164382

Popovici-Muller, J., Lemieux, R. M., Artin, E., Saunders, J. O., Salituro, F. G., Travins, J., et al. (2018). Discovery of AG-120 (Ivosidenib): A First-in-Class Mutant IDH1 Inhibitor for the Treatment of IDH1 Mutant Cancers. *ACS Med. Chem. Lett.* 9 (4), 300–305. doi: 10.1021/acsmedchemlett.7b00421

Rohle, D., Popovici-Muller, J., Palaskas, N., Turcan, S., Grommes, C., Campos, C., et al. (2013). An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* 340 (6132), 626–630. doi: 10.1126/science.1236062

12Sheisi, F., Carolina, G. O., Harold, H. F., and Carlos, M. R. S. A. (2019). Virtual Screening Techniques in Drug Discovery: Review and Recent Applications. *Curr. Top. Med. Chem.* 19 (19), 1751–1767. doi: 10.2174/1568026619666190816101948

Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007). Epik: a software program for pKaprediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21 (12), 681–691. doi: 10.1007/s10822-007-9133-z

Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., et al. (2013). Targeted Inhibition of Mutant IDH2 in Leukemia Cells Induces Cellular Differentiation. *Science* 340 (6132), 622. doi: 10.1126/science.1234769

Willett, P., and Winterman, V. (1986). A Comparison of Some Measures for the Determination of Inter-Molecular Structural Similarity Measures of Inter-Molecular Structural Similarity. *Quant. Structure-Activity Relat.* 5 (1), 18–25. doi: 10.1002/qsar.19860050105

Willett, P., Winterman, V., and Bawden, D. (1986). Implementation of nearest-neighbor searching in an online chemical structure search system. *J. Chem. Inf. Comput. Sci.* 26 (1), 36–41. doi: 10.1021/ci00049a008

Xie, X., Baird, D., Bowen, K., Capka, V., Chen, J., Chenail, G., et al. (2017). Allosteric Mutant IDH1 Inhibitors Reveal Mechanisms for IDH1 Mutant and Isoform Selectivity. *Structure* 25 (3), 506–513. doi: 10.1016/j.str.2016.12.017

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* 19 (1), 17–30. doi: 10.1016/j.ccr.2010.12.014

Yan, H., and Reitman, Z. J. (2010). Isocitrate Dehydrogenase 1 and 2 Mutations in Cancer: Alterations at a Crossroads of Cellular Metabolism. *J. Natl. Cancer Inst.* 102 (13), 932–941. doi: 10.1093/jnci/djq187

Yen, K., Dang, L., and Attar, E. C. (2016). IDH mutations in cancer and progress toward development of targeted therapeutics. *Ann. Oncol.* 27 (4), 599–608. doi: 10.1093/annonc/mdw013

Zheng, M., Sun, W., Gao, S., Luan, S., Li, D., Chen, R., et al. (2017). Structure based discovery of clomifene as a potent inhibitor of cancer-associated mutant IDH1. *Oncotarget* 8 (27), 44255–44265. doi: 10.18632/oncotarget.17464

Zou, F., Pusch, S., Eisel, J., Ma, T., Zhu, Q., Deng, D., et al. (2016). Identification of a novel selective inhibitor of mutant isocitrate dehydrogenase 1 at allosteric site by docking-based virtual screening. *RSC Adv.* 6 (99), 96735–96742. doi: 10.1039/c6ra21617j

Zou, F., Pusch, S., Hua, J., Ma, T., Yang, L., Zhu, Q., et al. (2018). Identification of novel allosteric inhibitors of mutant isocitrate dehydrogenase 1 by cross docking-based virtual screening. *Bioorg. Med. Chem. Lett.* 28 (3), 388–393. doi: 10.1016/j.bmcl.2017.12.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Tang, Lai, Jin, Long, Li, Tang, Guo, Yao and Leung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Pharmacology

PERSPECTIVE
published: 24 March 2021
doi: 10.3389/fphar.2021.565748

Targeting Nrf2-Mediated Oxidative Stress Response Signaling Pathways as New Therapeutic Strategy for Pituitary Adenomas

Xianquan Zhan[1,2,3]*, *Jiajia Li*[2,4] *and Tian Zhou*[2,4]

[1]Shandong Key Laboratory of Radiation Oncology, Cancer Hospital of Shandong First Medical University, Jinan, China, [2]Science and Technology Innovation Center, Shandong First Medical University, Jinan, China, [3]Department of Oncology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China, [4]Department of Thoracic Surgery, Xiangya Hospital, Central South University, Changsha, China

Oxidative stress and oxidative damage are the common pathophysiological characteristics in pituitary adenomas (PAs), which have been confirmed with many omics studies in PA tissues and cell/animal experimental studies. Nuclear factor erythroid 2 *p*45-related factor 2 (Nrf2), the core of oxidative stress response, is an oxidative stress sensor. Nrf2 is synthesized and regulated by multiple factors, including Keap1, ERK1/2, ERK5, JNK1/2, *p*38 MAPK, PKC, PI3K/AKT, and ER stress, in the cytoplasm. Under the oxidative stress status, Nrf2 quickly translocates from cytoplasm into the nucleus and binds to antioxidant response element /electrophile responsive element to initiate the expressions of antioxidant genes, phases I and II metabolizing enzymes, phase III detoxifying genes, chaperone/stress response genes, and ubiquitination/proteasomal degradation proteins. Many Nrf2 or Keap1 inhibitors have been reported as potential anticancer agents for different cancers. However, Nrf2 inhibitors have not been studied as potential anticancer agents for PAs. We recommend the emphasis on in-depth studies of Nrf2 signaling and potential therapeutic agents targeting Nrf2 signaling pathways as new therapeutic strategies for PAs. Also, the use of Nrf2 inhibitors targeting Nrf2 signaling in combination with ERK inhibitors plus p38 activators or JNK activators targeting MAPK signaling pathways, or drugs targeting mitochondrial dysfunction pathway might produce better anti-tumor effects on PAs. This perspective article reviews the advances in oxidative stress and Nrf2-mediated oxidative stress response signaling pathways in pituitary tumorigenesis, and the potential of targeting Nrf2 signaling pathways as a new therapeutic strategy for PAs.

Keywords: pituitary adenoma, oxidative stress, Nrf2, signaling pathway, biomarker, therapeutic target and drug

OPEN ACCESS

Edited by:
Feng Zhu,
Zhejiang University, China

Reviewed by:
Chen Ling,
Fudan University, China
Qingxia Yang,
Nanjing University of Posts and Telecommunications, China

***Correspondence:**
Xianquan Zhan
yjzhan2011@gmail.com

Specialty section:
This article was submitted to Pharmacology of Anti-Cancer Drugs, a section of the journal Frontiers in Pharmacology

Received: *26 May 2020*
Accepted: *12 February 2021*
Published: *24 March 2021*

Citation:
Zhan X, Li J and Zhou T (2021) Targeting Nrf2-Mediated Oxidative Stress Response Signaling Pathways as New Therapeutic Strategy for Pituitary Adenomas. Front. Pharmacol. 12:565748. doi: 10.3389/fphar.2021.565748

INTRODUCTION

Pituitary adenoma (PA) is a common intracranial neoplasm that occurs in the central regulatory organ pituitary gland in the hypothalamic-pituitary-target organ axis system, which seriously affects human endocrine system and health. PAs account for 10–25% of all intracranial tumors, and are classified into benign (~65%), invasive (~35%), and malignant (carcinoma; only 0.1–0.2%) PAs according to the malignancy level (Stalla et al., 2019). PAs are divided into macroadenomas

(≥10 mm) and microadenomas (<10 mm) according to tumor size (Lopes, 2017). They are also divided into clinically functional and nonfunctional PAs (FPAs and NFPAs) according to the level of hormone secretion (Zhan et al., 2016). FPAs are hormone-secreting PAs, which result in hyperpituitarism, including acromegaly derived from growth hormone (GH)-secreting PAs, hyperprolactinemia derived from prolactin (PRL)-secreting PAs, and Cushing's syndrome derived from adrenocorticotropin (ACTH)-secreting PAs. NFPAs are non-hormone-secreting PAs (Qian et al., 2018). The main clinical symptoms of PAs include inappropriate hormone secretion syndrome, and compression of the neighboring tissues and structures such as headache, visual field defect, and increased intracranial pressure (Reimondo et al., 2019). PA is a multi-factor, multi-process, and multi-consequence complex disease, which is involved in a series of molecular alterations at the levels of genome, transcriptome, proteome, peptidome, metabolome, and radiome; and these molecules mutually associate and function in a molecular network system (Zhan and Desiderio, 2010b; Hu et al., 2013; Grech et al., 2015; Cheng and Zhan, 2017; Lu and Zhan, 2018). Thus, one must shift the research and practice strategy from a single-factor model to a multi-parameter systematic model for predictive, preventive, and personalized medicine in PAs (Hu et al., 2013; Grech et al., 2015; Cheng and Zhan, 2017). Multiomics is an effective approach to realize this multi-parameter systematic strategy model shift, which can establish signaling pathway systems for in-depth understanding of molecular mechanisms of PAs, identify molecular network-based biomarkers for prediction, diagnosis, and prognostic assessment of PAs, and discover signaling pathway network-based therapeutic targets for effective treatment of PAs (Grech et al., 2015; Cheng and Zhan, 2017; Lu and Zhan, 2018).

A series of omics analyses have been performed in PAs to reach our long-term goals that clarify molecular mechanisms and discover effective biomarkers and therapeutic targets for PAs (Zhan and Desiderio, 2010a; Long et al., 2019; Cheng et al., 2019; Wang Y. et al., 2019), including NFPA quantitative transcriptomics (differentially expressed genes, DEGs) (Moreno et al., 2005; Cheng et al., 2019), NFPA quantitative proteomics (differentially expressed proteins, DEPs) (Moreno et al., 2005), NFPA proteomic mapping (Zhan and Desiderio, 2003; Wang X. et al., 2015; Cheng et al., 2019), NFPA nitroproteomics (Zhan and Desiderio, 2006), invasive NFPA quantitative transcriptomics (Galland et al., 2010; Zhou et al., 2011; Wang Y. et al., 2019), invasive NFPA quantitative proteomics (Zhan et al., 2014b), control pituitary proteomic mapping (Beranova-Giorgianni et al., 2002; Giorgianni et al., 2003; Zhao et al., 2005), pituitary control nitroproteomics (Zhan and Desiderio, 2004; Zhan and Desiderio, 2007), control pituitary phosphoproteomics (Giorgianni et al., 2004; Beranova-Giorgianni et al., 2006), PRL-secreting adenoma proteomics and transcriptomics (Evans et al., 2008), and ACTH-secreting adenoma proteomics and metabolomics (Feng et al., 2018). Integrative analysis of these omics data has revealed some important signaling pathway network alterations in PA pathogenesis, including mitochondrial dysfunction, oxidative stress, cell cycle dysregulation, and mitogen-activated protein kinase (MAPK) signaling pathway alteration (Zhan and Desiderio, 2010a; Long et al., 2019). Mitochondrial dysfunction pathway network and mitochondrial dynamics (Li and Zhan, 2019), and MAPK signaling pathway-based drug therapeutic targets (Lu et al., 2019) have been discussed in detailed in PAs. It is well-known that mitochondria are the energy factories of the body, and mitochondrial metabolism is the source of reactive oxygen species (ROS). The imbalance between free radicals reactive oxygen/nitrogen species (ROS/RNS) and antioxidant system leads to oxidative stress, which plays an important role in diseases. Many studies focus on oxidative stress system as therapeutic strategy; for example, benfotiamine is an efficient antioxidant, which could prevent oxidative stress in the anterior tibialis muscle and heart of mice (Gonçalves et al., 2019). Another research shows that pancreatic oxidative damage in the diabetic state is caused by ROS, and scavenging the various ROS generated in the disease is one of effective ways to treat this disease (Afolabi et al., 2018). Studies have clearly demonstrated that mitochondrial dysfunction and oxidative stress pathway changes operate in PAs (Zhan and Desiderio, 2010a), and nuclear factor erythroid 2 p45-related factor 2 (Nrf2)-mediated oxidative stress response significantly impacts the pathogenesis of PAs and modulates the energy metabolism reprogramming for PAs (Sabatino et al., 2018). It is well-known that PAs can lead to abnormal hormone secretion, which might affect oxidative stress and Nrf2 signaling in PAs; for example, human growth hormone (hGH) can attenuate inflammation and oxidative stress attained by Cisplatin probably through inhibition of Nrf2/heme oxygenase 1 (HO-1) pathway (Mahran, 2020). More studies show that Nrf2 signaling and oxidative stress can be regulated by cortisol (Wu et al., 2019), thyroid hormone (Mishra et al., 2019), follicle-stimulating hormone (FSH) (Li et al., 2020), luteinizing hormone (LH) (Li et al., 2020), GH (Mahran, 2020), ACTH (Benlloch et al., 2016), and PRL (Ebokaiwe et al., 2020). These findings clearly demonstrate the importance of oxidative stress in PAs. This present review article will focus on oxidative stress response signing pathway network in PA pathogenesis.

REDOX HOMEOSTASIS AND NRF2 AS THE HEART OF OXIDATIVE STRESS RESPONSE

Oxidative stress is derived from the imbalance between the upload of free radicals ROS/RNS from *in vivo* and *in vitro* environmental approaches and the ability of endogenous antioxidants to detoxify these ROS/RNS (Prasad et al., 2016; Klaunig, 2018; Sajadimajd and Khazaei, 2018). It results in the injuries of multiple biomacromolecules such as DNAs, RNAs, proteins, and membrane lipids to significantly associate with a wide spectrum of diseases including cancers. Many studies demonstrate that the increased ROS/RNS productions promote carcinogenesis development (Kudryavtseva et al., 2016; Kruk and Aboul-Enein, 2017), and oxidative stress-medicated chronic inflammation is the risk factor of tumorigenesis (Reuter et al., 2010; Qian et al., 2019). The oxidative phosphorylation system in

mitochondrial respiratory chain is the central machine that generates ROS products such as superoxide radical ($O_2^{\cdot-}$). One study shows that ROS levels and signs of oxidative damage are significantly increased in PAs (Sabatino et al., 2018). One of the most important RNS, nitric oxide (NO), is generated by inducible nitric synthase (iNOS) in many pathogenesis conditions, which can rapidly react with superoxide radical ($O_2^{\cdot-}$) to generate more toxic peroxynitrite anion ($ONOO^-$) and highly reactive hydroxyl radical ($OH^{\cdot}$) to attack DNAs, RNAs, proteins, and membrane lipids. iNOS has been extensively found in rat and human pituitaries (Ceccatelli et al., 1993; Lloyd et al., 1995; Ueta et al., 1998; Kruse et al., 2002; Pawlikowshi et al., 2003) and has the elevated activities in PAs compared to those in controls (Vankelecom et al., 1997; Kruse et al., 2002). Another study shows that NO functions in the hypothalamic-pituitary-adrenocortical axis (Riedel, 2002) by promoting the release of follicle-stimulating hormone-releasing hormone (FSHRH) and luteinizing hormone-releasing hormone (LHRH) from hypothalamus (McCann et al., 2001; Pinilla et al., 2001; McCann et al., 2003), and regulating secretion of PRL (Duvilanski et al., 1995) and GH in pituitaries and PAs (Cuttica et al., 1997; Pinilla et al., 1999; Bocca et al., 2000). Peroxynitrite anion ($ONOO^-$) is a key factor *in vivo* that causes protein tyrosine nitration and alters protein functions. Nine nitrotyrosine-containing proteins have been identified in NFPA tissues, and tyrosine nitration occurs in important structural and functional domains to change protein functions (Zhan and Desiderio, 2006).

With the generation of ROS/RNS, the *in vivo* antioxidant detoxification system is correspondingly initiated to adapt against the increased ROS/RNS (Valko et al., 2006; Obrador et al., 2019). The endogenous antioxidant detoxification system is a very complex system, including i) enzymatic antioxidants such as superoxide dismutases (CuZnSOD and MnSOD), glutathione peroxidase, and catalase; ii) non-enzymatic antioxidants such as vitamin E, vitamin C, carotenoid, flavonoid, selenium, thiol antioxidant (thioredoxin, lipoic acid, and glutathione), and others; and iii) multiple regulatory factors [Nrf2, NF-kB (nuclear factor kB), and AP-1 (activator protein-1), etc.] that interact with antioxidants (Valko et al., 2006; Obrador et al., 2019). CuZnSOD exists in most parts of cells, while MnSOD is only found in mitochondrial matrix; and both of them are able to effectively scavenge $O_2^{\cdot-}$ and generate H_2O_2 (Li et al., 1995; Melov et al., 2001; Elchuri et al., 2005). H_2O_2 can be scavenged by GPX's (glutathione peroxidases) and peroxiredoxins (thioredox-independent peroxidases) (Chu et al., 2004; Kang et al., 2005). Studies have found that the levels of CuZnMOD and MnSOD are significantly lower in PAs compared to those of controls (Kurisaka et al., 2004; Yang et al., 2012; Ilhan et al., 2018). The abnormal activities of these antioxidant enzymes and non-enzymatic antioxidants are directly associated with carcinogenesis (Neumann et al., 2003; Chu et al., 2004; Harris et al., 2015). The transcription factor Nrf2 is pivotal to the antioxidant response, which is a sensor of oxidative stress in redox homeostasis, and is mainly located in the cytoplasm under basal conditions (Li and Kong, 2009; Furfaro et al., 2016a). When the upload of free radicals ROS/RNS is increased to cause oxidative stress, Nrf2 quickly translocates from cytoplasm into the nucleus to initiate the antioxidant response, protecting against oxidative/nitrative damages (Dhakshinamoorthy and Porter, 2004; Osburn et al., 2006; Mann et al., 2007; Pi et al., 2008). The Nrf2 signaling regulatory system contains at least four components, including Nrf2, Kelch-like ECH-associated protein 1 (Keap1), small musculoaponeurotic fibrosarcoma (Maf), and antioxidant response element (ARE) or electrophile responsive element (EpRE), which in combination are necessary for the antioxidant response (Kwak and Kensler, 2010; Furfaro et al., 2016; de la Vega et al., 2018). Nrf2 signaling pathways regulate multiple biological processes, including i) the expressions of antioxidant genes, ii) ubiquitin-proteasome system, iii) molecular chaperone/stress-response system, and iv) anti-inflammatory response (Kwak and Kensler, 2010; Furfaro et al., 2016). The accumulated evidence clearly demonstrates that Nrf2 signaling pathways are involved in 12 hallmarks of cancer, including sustained proliferative signaling, insensitivity to antigrowth signals, resistance to apoptosis, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, metabolic reprogramming, avoiding immune destruction, tumor-promoting inflammation, genome instability, altered redox homeostasis, and proteotoxic stress (de la Vega et al., 2018). Thereby, any decreased capability of the antioxidant protective system in the redox homeostasis might cause more susceptibility to carcinogen toxicity, tumor inflammatory response, oxidative stress, and carcinogenesis (Yates and Kensler, 2007).

MULTIUOMICS REVEALS OXIDATIVE STRESS-RELATED PATHWAY ALTERATIONS IN PAs

Our multiomics studies in PAs (Zhan and Desiderio, 2010a; Long et al., 2019) clearly demonstrate oxidative stress-related pathway changes in PAs. For example, i) Nrf2-mediated oxidative stress response pathway is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEPs: GST (glutathione S-transferase) or GSTM2 (glutathione S-transferase mu 2), and ERP29 (endoplasmic reticulum protein 29], and downregulation of key molecules [downregulated DEPs: HSP22 (heat shock protein 22), HSP27, and HSP90 or GRP94 (94 kD glucose-regulated protein)] in this pathway. ii) Mitochondrial dysfunction pathway is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEPs: NDUFS8 (NADH ubiquinone oxidoreductase core subunit S8), COX6B (cytochrome c oxidase subunit 6B), CAT (catalase), β-secret2, and ATP5B (ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit)], and downregulation of key molecules [downregulated DEPs: GPX4 (glutathione peroxidase 4), and ATP5A1] in this pathway. Mitochondrial dysfunction can increase ROS production in cancer cells to mediate tumor-related signaling pathways and activate pro-oncogenic signaling (Li and Zhan, 2019). iii) Oxidative phosphorylation pathway is significantly changed in NFPAs with evidence of upregulation of key molecules (upregulated DEPs: NDUFS8,

COX6B, and ATP5B) in this pathway. Mitochondrial oxidative phosphorylation system contains mitochondrial complexes I, II, III, IV, and V, which are the major sites that produce endogenous ROS such as $OH^{\cdot}$ and O_2^{-}; among these, complexes I, II, and III play a crucial role in the generation of mitochondrial ROS, because the electrons tend to be leaky at complexes I and III, which results in an incomplete reduction of oxygen and thus generates a free radical such as superoxide radical (Li and Zhan, 2019). iv) Glutathione redox reaction I pathway is significantly changed in NFPAs with evidence of downregulation of key molecule (downregulated DEP: GPX4) in this pathway. GPX's (glutathione peroxidases) are important components in the antioxidant defense system: the downregulation of GPX's can decrease the capability of the antioxidant defense system. v) The superoxide radical degradation pathway is significantly changed in NFPAs with evidence of upregulation of key molecule (upregulated DEP: CAT) in this pathway. vi) Aryl hydrocarbon receptor signaling is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEP: GST; upregulated DEGs: HSPCA (heat shock protein 90 alpha family class A member 1), HSPCB (heat shock protein 90 alpha family class B member 1), ESR1 (estrogen receptor 1), and Bax (BCL2 associated X, apoptosis regulator)], and downregulation of key molecules [downregulated DEPs: HSP27, HSP90 or GRP94, and TGM2 (transglutaminase 2); downregulated DEG: ESR2 (estrogen receptor 2)] in this pathway. vii) Glucocorticoid receptor signaling is significantly changed in NFPAs with evidence of upregulation of key molecule [upregulated DEG: PI3K (phosphatidylinositol 3 kinase)], and downregulation of key molecules [downregulated DEGs: HSP70, c-Fos, CCL2 (C-C motif chemokine ligand 2), BCL2, PRL, and POMC (proopiomelanocortin)] in this pathway. viii) Corticotropin-releasing hormone signaling is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEGs: CALM (calmodulin), and IP3R)], and downregulation of key molecules [downregulated DEGs: ACTH, Nur77 (NR4A1 nuclear receptor subfamily 4 group A member 1), and c-FOS)] in this pathway. ix) Melatonin signaling is significantly changed in NFPAs with evidence of nitration of key molecule (PKA) in this pathway. x) Methylglyoxal degradation III pathway is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEPs: aldose reductase or AKR1B1(aldo-keto reductase family 1 member B)] in this pathway. xi) AMPK signaling is significantly changed in NFPAs with evidence of upregulation of key molecules [upregulated DEGs: PP2C (putative protein phosphatase), and PFK (phosphofructokinase)], and downregulation of key molecules [downregulated DEGs: PI3K, PKA, and PDK1 (pyruvate dehydrogenase kinase 1)] in this pathway. Thereby, these signaling pathway changes clearly demonstrate that the disturbance in redox homeostasis, the imbalance between generation and detoxification of free radicals ROS/RNS, results in oxidative stress and damage in human PAs. Recently, these findings are also confirmed with experiments in cell models and animal models, which demonstrate that increased mitochondrial fusion results in bigger mitochondria, increased ROS levels, and oxidative damage in PAs, and that Nrf2 signaling pathway is activated in PAs as an antioxidant response (Sabatino et al., 2018). Thus, it suggests that Nrf2 is the master regulator of the cellular antioxidant response (de la Vega et al., 2018).

NRF2-MEDIATED OXIDATIVE STRESS RESPONSE SIGNALING PATHWAYS IN PAs

Nrf2 signaling pathway in response to oxidative stress is shown (**Figure 1**). Multiple *in vivo* and *in vitro* environmental factors, including inflammatory cytokines, prostaglandins, growth factors, low-density lipoproteins, bacterial and viral infection, heavy metals, ultraviolet (UV) radiation, ionizing radiation, drugs, xenobiotics, antioxidants, oxidants, and chemopreventive agents, cause the increased upload of free radicals ROS/RNS and electrophiles to result in oxidative stress (Hetland et al., 2020; Mehnati et al., 2020). The increased ROS or electrophiles will activate the Nrf2/Keap1 complex in the cytoplasm through ERK1/2, ERK5, JNK1/2, p38 MAPK, PKC, and PI3K-AKT signaling pathways, and these signaling pathways will communicate with each other (Roy Chowdhury et al., 2014; Tian et al., 2014; Wang K.-C. et al., 2019). The activated Nrf2 is phosphorylated and separated from Keap1 (Hambright et al., 2015; Sánchez-Martín et al., 2020). The separated and phosphorylated Nrf2 quickly translocates into the nucleus to interact with ARE or EpRE, which will initiate at least five types of gene expressions to exert the corresponding biological functions (Furfaro et al., 2016; Sánchez-Martín et al., 2020): i) reduction of the oxidative damage via antioxidant proteins such as NRF2, small MAF, ATF4, SQSTM1, HO-1, PRDX1, FTL, FTH1, CAT, GPX's, SOD, TXN, GSR, and TRXR1 (Sun et al., 2019; Saad El-Din et al., 2020; Yu et al., 2020); ii) detoxification and metabolism of xenobiotics to regulate cell survival, or production of reactive metabolites to promote tumorigenesis via phase I and II metabolizing enzymes such as CYP1A/2A/3A/4A/2C, FMO, GST, NQD, UGT, AFAR, EPHX1, GCLC, GCLN, CBR4, AKR, and AOX4 (Zhao et al., 2015; Huang et al., 2018); iii) transportation of xenobiotics and metabolites via phase III detoxifying proteins such as SR-B1 and MRP1 (Sivils et al., 2013; Lubelska et al., 2016); iv) repairment and removal of the damaged proteins via chaperone and stress response proteins such as HSP22/40/90, STIP1, PTPLAD1, HERPUD1, CCT7, CLPP, FKBP5, PPIB, and ERP29 (Niture and Jaiswal, 2010; Sahin et al., 2012); and v) repairment and removal of the damaged proteins via ubiquitination and proteasomal degradation proteins such as PSM, UB2R1, VCP, USP14, UBB, and HIP2 (Liu et al., 2019; Song et al., 2019). This clearly demonstrates that while the Nrf2-mediated oxidative stress response signaling pathways are regulated by multiple factors, Nrf2 is the essential component. In the cytoplasm, Keap1, the main regulator of Nrf2, is a substrate adaptor protein for the Cul3-Keap1-E3 ligase complex that ubiquinates Nrf2, marking it for proteasomal degradation in the cytoplasm under basal conditions (Baird and Yamamoto, 2020; Dayalan Naidu and Dinkova-Kostova, 2020). To reduce its inhibitory effects on Nrf2, Keap1 can be ubiquitinated for degradation, leading to

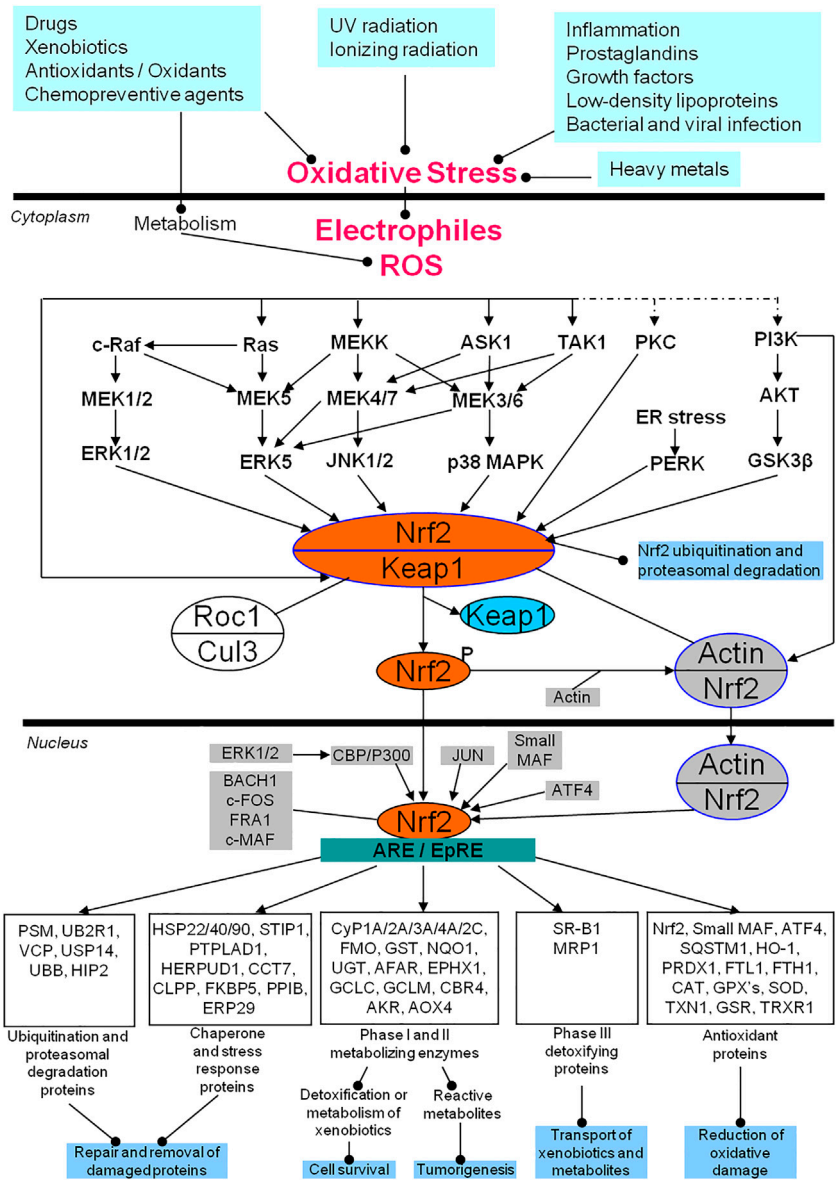


FIGURE 1 | Nrf2-mediated oxidative stress response signaling pathways in human pituitary adenomas. AKR, Palmitoyltransferase; AKT, Protein kinase B; AOX4, Aldehyde oxidase 4; ARE, Antioxidant response element; ASK1, Apoptosis signal-regulating kinase 1; ATF4, Activating transcription factor 4; BACH1, Transcription regulator protein BACH1; CAT, catalase; CBP, CREB-binding protein; CBR4, carbonyl reductase 4; CCT7, T-complex protein 1 subunit eta; c-FOS, Proto-oncogene protein c-FOS; CLPP, Caseinolytic protease; Cul3, Cullin 3-based ubiquitin E3 ligase complex; Cyp, cytochrome P; EPHX1, Epoxide hydrolase 1; EpRE, Electrophile responsive element; ER, endoplasmic reticulum; ERK: Extracellular signal-related kinase; ERP29: endoplasmic reticulum protein 29; FKBP5, FK506-binding protein 5; FMO, Dimethylaniline monooxygenase [N-oxide-forming]; FRA1, Fos-related antigen 1; FTH1, Ferritin heavy polypeptide 1; FTL1, ferritin light polypeptide; GCLC, glutamate-cysteine ligase catalytic subunit; GCLM, glutamate-cysteine ligase modifier subunit; GPX's, Glutathione peroxidases; GSK3β, glycogen synthase kinase 3β; GSR, glutathione reductase; GST, glutathione S-transferase; HERPUD1, Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 protein; HIP2, Ubiquitin-conjugating enzyme E2 K; HO-1, heme oxygenase 1; HSP22/40/90, heat shock proteins 22, 40 and 90; JNK, Jun N-terminal kinase; Keap1, Kelch-like ECH-associated protein 1; Maf, Musculoaponeurotic fibrosarcoma; MAPK, Mitogen-activated protein kinase; MEK, Mitogen-activated protein kinase kinase (MAPKK); MEKK, Mitogen-activated protein kinase kinase kinase (MAPKKK); MRP1, multidrug-resistant protein-1; NQO1, NAD(P)H:quinine oxidoreductase 1; Nrf2, Nuclear factor erythroid 2 p45-related factor 2; PERK: the double-stranded RNA (PKR)-activated protein kinase-like eukaryotic initiation factor 2 kinase; PI3K, Phosphatidylinositol 3 kinase; PKC, protein kinase C; PPIB, Peptidyl-prolyl cis-trans isomerase B; PRDX1, peroxiredoxin 1; PSM: multiple subunits of the 20S proteasome; PTPLAD1, 3-hydroxyacyl-CoA dehydratase 3; c-Raf, RAF proto-oncogene serine/threonine-protein kinase; Ras, GTPase Ras; ROS, reactive oxygen species; SOD, Superoxide dismutase; SQSTM1, sequestosome-1 protein; SR-B1, Scavenger receptor class B member 1; STIP1, stress induced phosphoprotein 1; TAK1, TGF beta-Activated Kinase 1; TXN1: thioredoxin; TRXR1, thioredoxin reductase 1; UBB, Polyubiquitin-B; UB2R1, Ubiquitin-conjugating enzyme E2 R1; UGT, UDP glucuronosyl transferase; USP14, ubiquitin-specific peptidase 14; and VCP, valosin-containing protein. Modified from Zhan X et al. (2010) (Zhan and Desiderio, 2010a), copyright permission from BioMed Central publisher open-access article, copyright 2010; and modified from Long Y et al. (2019) (Long et al., 2019), copyright permission from Frontiersin publisher open-access article, copyright 2019.

an increase in Nrf2 phosphorylation (activation) (Villeneuve et al., 2010). The phosphorylated Nrf2 can then interact with actin to form an Nrf2/actin complex that then translocates into the nucleus. After Nrf2 translocates into the nucleus, there are additional regulatory systems in place that include multiple factors such as ATF4, JUN, ERK1/2-CBP/P300, small MAF, BACH1, c-FOS, FRA1, and c-MAF, to influence the binding of Nrf2 and ARE/EpRE. The detailed regulatory mechanism system of Nrf2 has been extensively reviewed (Kwak and Kensler, 2010; Hybertson et al., 2011; Furfaro et al., 2016a; Lu et al., 2016; Menegon et al., 2016; Taguchi and Yamamoto, 2017; Bellezza et al., 2018; Chen and Maltagliati, 2018; de la Vega et al., 2018; Ryoo and Kwak, 2018; Sajadimajd and Khazaei, 2018; Cloer et al., 2019; Cuadrado et al., 2019; Qin et al., 2019) response signaling pathways have also been studied in pituitaries and PAs. One study shows that Nrf2, phosphorylated Nrf2 (p-Nrf2) protein, and mRNA expressions are increased in PAs, and the Nrf2 downstream effector HO-1 is also increased in PAs (Sabatino et al., 2018). This clearly demonstrates the activation of the Nrf2 signaling pathway, likely causing the extensive surviving capability of pituitary tumor cells. The Nrf2/PTEN-induced putative kinase protein 1 (PINK1)/Parkin pathway and mitophagy are activated in T-2 toxin-induced toxicities in rat pituitary GH3 cells (Deyu et al., 2018). Antioxidants N-acetylcysteine (NAC) and vitamin E can decrease the expressions of Nrf2 and HO-1 in rat pituitaries (Prevatto et al., 2017). Genetically induced Nrf2 overexpression in melanoma cells promotes tumor growth and increases antioxidant defense in malignant cells, which can be inhibited by anticancer agent pterostilbene (Pter, a natural dimethoxylated analog of resveratrol) through the downregulation of pituitary production of ACTH, plasma corticosterone, and the glucocorticoid receptor- and Nrf2-dependent antioxidant defense systems in growing melanomas (Benlloch et al., 2016). Irradiation can result in oxidative damage in C57/BL6 mice via activation of Nrf2 and HO-1 expressions, which can be blocked by antioxidant agent pituitary adenylate cyclase-activating polypeptide 38 (PACAP38) through inhibiting Nrf2 expression (Li et al., 2019). Chronic restraint stress (CRS) and acute restraint stress (ARS) can upregulate the mRNA expressions of oxidative stress molecules (gp91phox, iNOS, and Nrf2) and inflammation-related molecules (IL-1β, IL-6, TNFα, and TLR4) in the mouse hypothalamus, which can be alleviated by Iptakalim (Ipt), an ATP-sensitive potassium (K-ATP) channel opener (Zhao et al., 2017). The loss-of-function mutations of the aryl hydrocarbon receptor-interacting protein gene (AIP) are well-recognized in PAs (Hernández-Ramírez et al., 2018). The aryl hydrocarbon receptor signaling is also revealed by multiomics as an oxidative stress-related signaling pathway in PAs (Long et al., 2019). Further studies show that AIP interacts with antioxidants, chaperone and stress response-related proteins, and cytoskeletal proteins, including HSPA5, HSPA9, HSP90AA1, HSP90AB1, HSPA8, SOD1, TUBB, TUBB2A, and NME1; AIP variants show the impaired interaction of AIP with HSPA8, HSP90AB1, NME1, SOD1, TUBB, and TUBB2A; AIP-mutated PAs show the reduced expression of TUBB2A (Cuadrado et al., 2019). The levels of MnSOD and total antioxidant capability (TAC) are significantly decreased in GH-secreting PAs (Ilhan et al., 2018). The frequencies of micronuclei (MN), nucleoplasmic bridges, nuclear buds, apoptotic and necrotic cells, and plasma 8-hydroxy-2′-deoxyguanosine (8-OHdG) levels in peripheral blood lymphocytes are significantly increased in PRL-secreting PAs, which indicates the increased oxidative damage in PRL-secreting PAs (Bitgen et al., 2016). Oxidative stress and mitochondrial dysfunction have been revealed by multiple proteomics and nitroproteomics studies in human PAs (Zhan and Desiderio, 2010a; Zhan et al., 2013; Zhan et al., 2014a; Zhan et al., 2014b; Wang X. et al., 2015; Long et al., 2019). Also, tumor inflammation is an important pathophysiological characteristic in human PAs, which is always tightly associated with oxidative stress and chronic inflammation. The relationship among age-related disease, chronic inflammation, and oxidative stress has also been discussed (Pizza et al., 2011; Liguori et al., 2018; Qian et al., 2019). Oxidative stress is also involved in the processes of anti-proliferative effect and cell death induced by dopamine in the pituitary tumor cells via dopamine D2 receptors through p38 MAPK, and ERK pathways (An et al., 2003). Therefore, oxidative stress and antioxidative stress response extensively exist in PA pathogenesis. Nrf2, as the core of oxidative stress response, could be the novel target used to develop effective therapeutic agents for human PAs (Kwak and Kensler, 2010; Furfaro et al., 2016; de la Vega et al., 2018).

THERAPEUTIC STATUS TARGETING NRF2 SIGNALING PATHWAYS IN CANCERS

Nrf2 signaling, as the heart of oxidative stress response, is extensively related to cancer pathogenesis, which has attracted tremendous attention as possible anticancer therapeutic target. Nrf2 signaling-based anticancer therapeutic studies have been extensively carried out in multiple cancers, including acute myeloid leukemia, gallbladder cancer, renal carcinoma, pancreatic cancer, melanoma, hepatocellular carcinoma, lung cancer, colon cancer, ovarian cancer, breast cancer, esophageal cancer, and glioblastoma (**Table 1**). i) In acute myeloid leukemia, studies found that Nrf2 activators [dimethyl fumarate (DMF), tert-butylhydroquinone, or carnosic acid] and vitamin D derivatives can cooperatively induce acute myeloid leukemia cell differentiation to inhibit leukemia progression in a xenograft mouse model via activating the Nrf2/ARE signaling pathway (Nachliely et al., 2019). Novel pyrazolyl hydroxamic acid derivative (4f) can inhibit Nrf2 activity to induce apoptosis of human acute myeloid leukemia cells (Zhang et al., 2017). ii) In gallbladder cancer, one study found that atypical protein kinase Cι (aPKCι) can promote gallbladder tumorigenesis and chemoresistance of anticancer agent gemcitabine by competing with Nrf2 for binding to Keap1, implying that inhibiting the aPKC1-Keap1-Nrf2 axis might overcome drug resistance for the treatment of gallbladder cancer (Tian et al., 2019). iii) In renal carcinoma, one study found that the natural product chitosan oligosaccharide (COS) can inhibit human renal carcinoma cell proliferation *in vitro* and *in vivo* by promoting the expressions of Nrf2 and Nrf2 target genes such as HO-1, the modifier subunit of

TABLE 1 | Current research status of therapeutic potentials targeting Nrf2-mediated oxidative stress response signaling pathways in different cancers.

Cancer type	Experimental model	Chemical reagents or potential drugs	Possible mechanisms	References
Acute myeloid leukemia	Acute myeloid leukemia cells in a xenograft mouse model	Nrf2 activators: dimethyl fumarate (DMF), tert-butylhydroquinone, or carnosic acid	Cooperate with vitamin D derivatives to induce acute myeloid leukemia cell differentiation to inhibit leukemia progression in a xenograft mouse model via activating the Nrf2/ARE signaling pathway	Nachliely et al. (2019)
	Human acute myeloid leukemia cells	Novel pyrazolyl hydroxamic acid derivative (4f)	Inhibit Nrf2 activity to induce apoptosis of human acute myeloid leukemia cells	Zhang et al. (2019)
Gallbladder cancer	Gallbladder cancer cells	The aPKCι inhitors, Nrf2 activators, or gemcitabine	Atypical protein kinase Cι (aPKCι) can promote gallbladder tumorigenesis and chemoresistance of anticancer agent gemcitabine by competing with Nrf2 for binding to Keap1, implying that inhibiting the aPKCι-Keap1-Nrf2 axis might overcome drug resistance for the gallbladder cancer treatment	Tian et al. (2019)
Renal carcinoma	Human renal carcinoma cells	Chitosan oligosaccharide (COS)	Inhibit human renal carcinoma cell proliferation *in vitro* and *in vivo* by promoting the expressions of Nrf2 and Nrf2 target genes such as HO-1, the modifier subunit of glutamate cysteine ligase, solute carrier family 7 member 11, glucose-regulated protein 78, protein RNA-like endoplasmic reticulum kinase, and cytochrome C,etc.	Zhai et al. (2019)
Pancreatic cancer	Pancreatic cancer cells	Resveratrol	Enhance the sensitivity of pancreatic cancer cells to gemcitabine via suppressing NAF-1 expression, inducing ROS accumulation, and activating Nrf2 signaling pathways	Cheng et al. (2018)
Melanoma	Melanoma cells	Nrf2 inhibitor: Brusatol (BR)	The co-treatment of brusatol and UVA irradiation can effectively inhibit melanoma growth by regulating the AKT-Nrf2 pathway	Wang et al. (2018)
Hepatocellular carcinoma	Hepatocellular carcinoma (HCC) cells	Vitamin C (VC), all-trans retinoic acid (ATRA), ochratoxin A (OTA), bexarotene, flavonoids (including brusatol, luteolin, apigenin and chrysin), ruthenium (Ru) metal complexes, ursolic acid (UA), halofuginone, trigonelline, quercetin, and isoniazid	Sensitize chemotherapy drugs in hepatocellular carcinoma	Tian et al. (2018)
	Mouse hepatocellular carcinoma model	Cordycepin (CA)	Activate the Nrf2/HO-1/NF-κB pathway for its anti-hepatocarcinoma effect in N-nitrosodiethylamine (NDEA)-induced mouse hepatocellular carcinomas	Zeng et al. (2017)
	Hep3B (human hepatoma cell) and HL-7702 (normal human liver cell) cell lines	Novel indazolo[3,2-b] quinazolinone (IQ) derivatives: IQ-7 and IQ-12	Induce apoptosis and inhibit the Nrf2/ARE signaling pathway in Hep3B cells, and IQ-7 was suggested a degree of specificity against cancer cells.	Zhang et al. (2016)
	Liver injury mouse model	Dibenzoylmethane (DBM)	Protect against carbon tetrachloride (CCl4)-induced liver injury by activating Nrf2 signaling via JNK, AMPK, and calcium signaling	Cao et al. (2017)
Lung cancer	Lung cancer cells	The potent anticancer agent: Isodeoxyelephantopin	Induce protective autophagy in lung cancer cells via the Nrf2-p62-keap1 pathway	Wang et al. (2017)
	RAW 264.7 mouse macrophage-like cells, in VC1 lung cancer cells, and in the A/J model of lung cancer	Two clinically relevant classes of Nrf2 activators: DMF, and the synthetic oleanane triterpenoids –C-28 methyl ester of 2-cyano-3,12-dioxoolean-1,9-dien-28-oic acid (CDDO)-Imidazolide (CDDO-Im) and CDDO-Methyl ester (CDDO-Me)	Activate the Nrf2 pathway as well as regulate different subsets of Nrf2 target genes and Nrf2-independent genes	Chian et al. (2014) and To et al. (2015)

(Continued on following page)

TABLE 1 | (*Continued*) Current research status of therapeutic potentials targeting Nrf2-mediated oxidative stress response signaling pathways in different cancers.

Cancer type	Experimental model	Chemical reagents or potential drugs	Possible mechanisms	References
Colon cancer	SFN-treated human colon cancer cells and non-transformed colonic epithelial cells	Anticancer agent: Sulforaphane (SFN)	Regulate the activity of antioxidant and the detoxification of carcinogens via Nrf2 signaling to suppress human colon cancer	Johnson et al. (2017)
	1, 2-dimethyl hydrazine (DMH)-induced mouse colon model	Taxifolin (TAX)	Induce antioxidant response pathway, enhance level of Nrf2 proteins, and act as effective chemopreventive agent capable of modulating inflammatory	Manigandan et al. (2015)
Ovarian cancer	Human ovarian cancer cell lines: PEO4, OVCAR4, and SKOV3	Anti-HER2 drugs: Trastuzumab and Pertuzumab	HER2 targeting by antibodies inhibited growth in association with persistent ROS generation, glutathione (GSH) depletion, reduction in NRF2 levels, and inhibition of NRF2 function in ovarian cancer cell lines	Khalil et al. (2016)
	Human epithelial ovarian cancer (EOC) cell lines	Keap1 mutation reagent	Activation of Nrf2 pathway in EOC seems to be related to Keap1 mutations within highly conserved domains of the Keap1 gene; and Nrf2 may serve as an important therapeutic target for novel drugs capable of preventing or reversing resistance to chemotherapy in EOC	Konstantinopoulos et al. (2011)
Breast cancer	Breast cancer cells, and mouse model	Target antioxidant enzymes: GCLC and GCLM	Nrf2 serves as a key regulator in chemotherapeutic resistance under hypoxia through ROS-Nrf2-GCLC-GSH pathway, and can be a potential treatment for hypoxia-induced drug resistance in breast cancer cells.	Syu et al. (2016) and Song et al. (2011)
Esophageal cancer	Esophageal squamous cancer cells (ESCC): Ec109 and KYSE70 cells	CDDO-Me	Protects the cells against oxidative stress via inhibition of ROS generation, while CDDO-Me at low micromolar concentrations induces apoptosis by increasing ROS and decreasing intracellular glutathione levels	Wang X. et al. (2015)
Glioblastoma	Glioblastoma cells	Potential anti-cancer agents	Targeting Nrf2 signaling for chemotherapy and chemoresistance	Zhu et al. (2014)
Osteosarcoma	Human osteosarcoma 143B and MG63 cells	The bioengineered Nrf2-siRNA	Interfere with the Nrf2 signaling pathway to reduce the expression of NRF2-regulated oxidative enzymes and lead to higher intracellular ROS levels; knocking down NRF2 with bioengineered siRNA agent improves chemosensitivity of cancer cells, which is related to the suppression of NRF2-regulated efflux ABC transporters.	Li et al. (2018)
Other cancers	prostate cancer cell PC4-LN4; colon cancer cell HCT-116; breast cancer cells MB-MDA-231 and MB-MDA-231-ARE-Luc	PIM kinases inhibitors	Inhibit Nrf2 signaling and increase ROS to kill hypoxic tumor cells in a HIF-1-independent manner by controlling its cellular localization	Warfel et al. (2016)
	Mammalian cancer cells	Proteasome inhibitors	In response to proteasome inhibition, several responses are activated, such as the ALP, proteaphagy, the transcriptional upregulation of the autophagy Ubreceptor p62/SQSTM1, and proteasome genes, by Nrf1 and Nrf1/Nrf2 transcription factors, respectively.	Albornoz et al. (2019)
	Mouse epidermal cells (JB6 P+),	Gallic acid (GA), Z-ligustilide (LIG), and senkyunolide A (SA)	GA, LIG, and SA in Si-Wu-Tang (SWT) can individually or cooperatively target the Nrf2/ARE pathway to prevent cancer.	Liu et al. (2018)

ALP, Autophagic-Lysosomal Pathway; ATRA, All-trans retinoic acid; BR, Brusatol; CA, Cordycepin; CDDO, C-28 methyl ester of 2-cyano-3,12-dioxoolean-1,9-dien-28-oic acid; COS, Chitosan oligosaccharide; DBM, Dibenzoylmethane; DMF, dimethyl fumarate; GA, Gallic acid; IQ, Indazolo[3,2-b] quinazolinone; LIG, Z-ligustilide; OTA, Ochratoxin A; PIM, The Proviral Integration site for Moloney murine leukemia virus; Ru, Ruthenium; SA, Senkyunolide A; SFN, Sulforaphane; TAX, Taxifolin; UA, Ursolic acid; VC, vitamin C.

glutamate cysteine ligase, solute carrier family 7 member 11, glucose-regulated protein 78, protein RNA-like endoplasmic reticulum kinase, and cytochrome C. (Zhai et al., 2019). iv) In pancreatic cancer, one study found that anticancer agent resveratrol enhances the sensitivity of pancreatic cancer cells to gemcitabine via suppressing NAF-1 (nutrient-deprivation autophagy factor-1) expression, inducing ROS accumulation, and activating Nrf2 signaling pathways (Cheng et al., 2018). v) In melanoma, the co-treatment of Nrf2 inhibitor (brusatol, BR) and UVA irradiation can effectively inhibit melanoma growth by regulating AKT-Nrf2 pathway (Wang et al., 2018). vi) In hepatocellular carcinoma, one study found that potential Nrf2 inhibitors can sensitize chemotherapy drugs in hepatocellular carcinoma (Tian et al., 2018). Cordycepin (CA) can activate the Nrf2/HO-1/NF-κB pathway for its anti-hepatocarcinoma effect in N-nitrosodiethylamine (NDEA)-induced mouse hepatocellular carcinomas (Zeng et al., 2017). The novel indazolo[3,2-b]quinazolinone (IQ) derivatives, IQ-7 and IQ-12, can induce apoptosis of human hepatoma cells Hep3B and inhibit the Nrf2/ARE signaling pathway in Hep3B cells, and IQ-7 is suggested as a degree of specificity against cancer cells (Zhang et al., 2016). Also, dibenzoylmethane (DBM) can protect against carbon tetrachloride (CCl4)-induced liver injury by activating Nrf2 signaling via JNK, AMPK, and calcium signaling (Cao et al., 2017). vii) In lung cancer, one study found that the potent anticancer agent isodeoxyelephantopin can induce protective autophagy in lung cancer cells via the Nrf2-p62-keap1 pathway (Wang et al., 2017). The Nrf2 activators, DMF and the synthetic oleanane triterpenoids, activate the Nrf2 pathway as well as regulate different subsets of Nrf2 target genes and Nrf2-independent genes in lung cancer (Chian et al., 2014; To et al., 2015). viii) In colon cancer, one study found that anticancer agent sulforaphane (SFN) can activate Nrf2 signaling to suppress human colon cancer (Johnson et al., 2017). Also, taxifolin (TAX) can induce antioxidant response pathway and enhance level of Nrf2 protein, and act as effective chemopreventive agent capable of modulating inflammation in colon cancer (Manigandan et al., 2015). ix) In ovarian cancer, one study found that Nrf2 can mediate the response of cancer cells to the anti-HER2 drugs, trastuzumab and pertuzumab, in ovarian cancer cells (Khalil et al., 2016). Also, activation of Nrf2 pathway in ovarian cancer seems to be related to Keap1 mutations within highly conserved domains of Keap1 gene and that Nrf2 may serve as an important therapeutic target for novel drugs capable of preventing or reversing resistance to chemotherapy in ovarian cancer (Konstantinopoulos et al., 2011). x) In breast cancer, Nrf2 serves as a key regulator in chemotherapeutic resistance under hypoxia through ROS-Nrf2-GCLC-GSH pathway and can be a potential treatment for hypoxia induced drug resistance in breast cancer cells (Song et al., 2011; Syu et al., 2016). xi) In esophageal cancer, C-28 methyl ester of 2-cyano-3,12-dioxoolean-1,9-dien-28-oic acid (CDDO-Me) can protect the cells against oxidative stress via inhibition of ROS generation, while CDDO-Me at low micromolar concentrations induces apoptosis by increasing ROS and decreasing intracellular glutathione levels in esophageal squamous cancer cells (Wang Y. Y. et al., 2015). xii) In glioblastoma, there are many potent anti-cancer agents targeting Nrf2 signaling for chemotherapy and chemoresistance in glioblastoma (Zhu et al., 2014). xiii) In osteosarcoma, the bioengineered Nrf2-siRNA can effectively interfere with the Nrf2 signaling pathway to improve chemosensitivity of human cancer cells (Li et al., 2018). Moreover, the PIM (proviral integration site for moloney murine leukemia virus) kinase inhibitors can reduce Nrf2 signaling and increase ROS to kill hypoxic tumor cells such as prostate cancer cells (PC4-LN4), colon cancer cells (HCT-116), and breast cancer cells (MB-MDA-231 and MB-MDA-231-ARE-Luc) (Warfel et al., 2016). One study shows that proteasome biogenesis is dependent on the Nrf2 transcriptional factor, thus proteasome inhibitors have been actively developed as potential anticancer drugs (Albornoz et al., 2019). Gallic acid (GA), Z-ligustilide (LIG), and senkyunolide A (SA) can individually or cooperatively target Nrf2/ARE pathway to prevent cancer (Liu et al., 2018). Therefore, it can be said that Keap1-Nrf2 signaling pathways have different roles at different stages of cancer (Leinonen et al., 2014; Furfaro et al., 2016; de la Vega et al., 2018). Multiple Nrf2 or Keap1 inhibitors have been reported; and some of them are in the stages of pre- and clinical trial towards the Nrf2 signaling for cancers. For example, sulforaphane can target Nrf2 and the Nrf2 target genes NQO1 and GCLC to prevent oral cancer, and a preclinical trail has been performed to study its chemopreventive activity for oral cancer (Bauman et al., 2016). A single centre, single arm prospective phase II clinical trial has been performed for phytosome complex of curcumin targeting Nrf2 signaling as a the complementary therapy of gemcitabine on pancreatic cancer (Pastorelli et al., 2018). However, none of these Nrf2 or Keap1 inhibitors have currently entered into real clinical applications, which suggests that the sole inhibition of Nrf2 might not be sufficient for anticancer. A rational combination of Nrf2 inhibitors with other chemical agents would be a better strategy to treat cancers (Zhang et al., 2019).

POTENTIAL OF TARGETING NRF2 SIGNALING AS NEW THERAPEUTIC STRATEGY FOR PAs

As described above, many omics studies in human PA tissues and experimental studies in PA cells and animal models demonstrate that oxidative stress and oxidative damage is the important hallmark of PA pathogenesis. Nrf2-mediated oxidative stress response signaling pathways are at the heart of oxidative stress response, and many chemical agents targeting Nrf2 signaling pathways have been developed and tested as potential anticancer drugs for different cancers. This clearly demonstrates the potential of targeting Nrf2 signaling pathways as new therapeutic strategies for PAs. However, the use of Nrf2 signaling as a therapeutic target for PAs has not been studied. We strongly believe that the Nrf2-mediated oxidative stress response signaling pathways are the promising targets for novel therapeutic strategies for PAs. Furthermore, MAPK signaling pathways including ERK, JNK, and p38 MAPK clearly regulate Nrf2 signaling (**Figure 1**). Moreover, MAPK signaling pathways have been recognized as potential therapeutic targets for PAs (Lu et al., 2019). The combined

use of Nrf2 inhibitors targeting Nrf2 signaling and ERK inhibitors [e.g., somatostatin analogs pasireotide (SOM230) and octreotide (OCT), or dopamine] plus *p*38 activators (e.g., cabergoline, bromocriptine, and fulvestrant) or JNK activators (e.g., ursolic acid, UA) targeting MAPK signaling pathways (Lu et al., 2019) might produce better anti-tumor effects on PAs. In addition, oxidative stress is tightly associated with mitochondrial dysfunctions, both operate in PAs (Zhan and Desiderio, 2010a; Li and Zhan, 2019; Long et al., 2019). Some drugs targeting mitochondria are also recognized as a therapeutic strategy for PAs, including pyrimethamine, temozolomide, melatonin, melatonin inhibitors, gossypol acetate, 18 beta-glycyrrhetinic acid, T-2 toxin, Yougui pill, cyclosporine A, grifolic acid, paeoniflorin, and dopamine agonists (Li and Zhan, 2019). Therefore, the combined use of Nrf2 inhibitors targeting Nrf2 signaling and drugs targeting metochondria could be another way to generate better anti-tumor effects on PAs.

CONCLUSION

Pituitary adenoma (PA) is a common and important disease that occurs in the hypothalamic-pituitary-target organ axis system and seriously affects human endocrine system and health. The imbalance between oxidative stress and the antioxidant defense system is an important pathophysiological characteristic in PAs, which has been evidenced by many omics analysis in PA tissues and experimental studies in PA cells and animal models. Nrf2 signaling is at the heart of oxidative stress response signaling pathways. Multiple anticancer agents targeting Nrf2-mediated oxidative stress response pathways have been developed and tested as potential therapeutic drugs for different cancers. However, Nrf2 signaling and targeting Nrf2 signaling as a therapeutic strategy has not yet been extensively studied in PAs. We strongly recommend the emphasis on in-depth studies of Nrf2 signaling and potential therapeutic agents targeting Nrf2 signaling pathways in PAs. Furthermore, the combined use of Nrf2 inhibitors targeting Nrf2 signaling and ERK inhibitors plus *p*38 activators or JNK activators targeting MAPK signaling pathways, or drugs targeting mitochondria dysfunction pathway might produce better anti-tumor effects on PAs.

AUTHOR CONTRIBUTIONS

XZ conceived the concept, collected and analyzed literature, designed, coordinated, wrote and revised manuscript, and was responsible for its financial supports and the corresponding works. JL and TZ participated in literature collection and analysis, and prepared figures. All authors approved the final manuscript.

FUNDING

The financial supports from the Shandong First Medical University Talent Introduction Funds (to XZ), and the Hunan Provincial Hundred Talent Plan (to XZ).

ACKNOWLEDGMENTS

The authors acknowledge the financial supports from the Shandong First Medical University Talent Introduction Funds (to XZ), and the Hunan Provincial Hundred Talent Plan (to XZ). We acknowledge that Xiaohan Zhan from Wesleyan University in United States critically revised the manuscript.

REFERENCES

Afolabi, O. B., Oloyede, O. I., and Agunbiade, S. O. (2018). Inhibitory potentials of phenolic-rich extracts from Bridelia ferruginea on two key carbohydrate-metabolizing enzymes and Fe2+-induced pancreatic oxidative stress. *J. Integrat Med.* 16, 192–198. doi:10.1016/j.joim.2018.04.006

Albornoz, N., Bustamante, H., Soza, A., and Burgos, P. (2019). Cellular responses to proteasome inhibition: molecular mechanisms and beyond. *Int. J. Mol. Sci.* 20 (14), E3379. doi:10.3390/ijms20143379

An, J. J., Cho, S. R., Jeong, D. W., Park, K. W., Ahn, Y. S., and Baik, J. H. (2003). Anti-proliferative effects and cell death mediated by two isoforms of dopamine D2 receptors in pituitary tumor cells. *Mol. Cell Endocrinol.* 206 (1-2), 49–62. doi:10.1016/s0303-7207(03)00236-3

Baird, L., and Yamamoto, M. (2020). The molecular mechanisms regulating the KEAP1-NRF2 pathway. *Mol. Cell Biol.* 40 (13), e00099. doi:10.1128/MCB.00099-20

Bauman, J. E., Zang, Y., Sen, M., Li, C., Wang, L., Egner, P. A., et al. (2016). Prevention of carcinogen-induced oral cancer by sulforaphane. *Cancer Prev. Res.* 9 (7), 547–557. doi:10.1158/1940-6207.CAPR-15-0290

Bellezza, I., Giambanco, I., Minelli, A., Donato, R., Bellezza, I., et al. (2018). Nrf2-Keap1 signaling in oxidative and reductive stress. *Biochim. Biophys. Acta Mol. Cel. Res.* 1865 (5), 721–733. doi:10.1016/j.bbamcr.2018.02.010

Benlloch, M., Obrador, E., Valles, S. L., Rodriguez, M. L., Sirerol, J. A., Alcácer, J., et al. (2016). Pterostilbene decreases the antioxidant defenses of aggressive cancer cells *in vivo*: a physiological glucocorticoids- and Nrf2-dependent mechanism. *Antioxid. Redox Signal* 24 (17), 974–990. doi:10.1089/ars.2015.6437

Beranova-Giorgianni, S., Giorgianni, F., and Desiderio, D. M. (2002). Analysis of the proteome in the human pituitary. *Proteomics* 2, 534–542. doi:10.1002/1615-9861(200205)2:5<534::AID-PROT534>3.0.CO;2-K

Beranova-Giorgianni, S., Zhao, Y., Desiderio, D. M., and Giorgianni, F. (2006). Phosphoproteomic analysis of the human pituitary. *Pituitary* 9, 109–120. doi:10.1007/s11102-006-8916-x

Bitgen, N., Donmez-Altuntas, H., Bayram, F., Cakir, I., Hamurcu, Z., Diri, H., et al. (2016). Increased micronucleus, nucleoplasmic bridge, nuclear bud frequency and oxidative DNA damage associated with prolactin levels and pituitary adenoma diameters in patients with prolactinoma. *Biotech. Histochem.* 91 (2), 128–136. doi:10.3109/10520295.2015.1101163

Bocca, L., Valenti, S., Cuttica, C. M., Spaziante, R., Giordano, G., and Giusti, M. (2000). Nitric oxide biphasically modulates GH secretion in cultured cells of GH-secreting human pituitary adenomas. *Minerva. Endocrinol.* 25 (3-4), 55–59.

Cao, M., Wang, H., Guo, L., Yang, S., Liu, C., Khor, T. O., et al. (2017). Dibenzoylmethane protects against CCl4-induced acute liver injury by activating Nrf2 via JNK, AMPK, and calcium signaling. *AAPS J.* 19 (6), 1703–1714. doi:10.1208/s12248-017-0133-1

Ceccatelli, S., Hulting, A. L., Zhang, X., Gustafsson, L., Villar, M., and Hökfelt, T. (1993). Nitric oxide synthase in the rat anterior pituitary gland and the role of nitric oxide in regulation of LH secretion. *Proc. Natl. Acad. Sci. U.S.A.* 90 (23), 11292–11296. doi:10.1073/pnas.90.23.11292

Chen, Q. M., and Maltagliati, A. J. (2018). Nrf2 at the heart of oxidative stress and cardiac protection. *Physiol. Genomics* 50 (2), 77–97. doi:10.1152/physiolgenomics.00041.2017

Cheng, L., Yan, B., Chen, K., Jiang, Z., Zhou, C., Cao, J., et al. (2018). Resveratrol-induced downregulation of NAF-1 enhances the sensitivity of pancreatic cancer cells to gemcitabine via the ROS/Nrf2 signaling pathways. *Oxid Med. Cell Longev.* 2018, 9482018. doi:10.1155/2018/9482018

Cheng, T., Wang, Y., Lu, M., Zhan, X., Zhou, T., Li, B., et al. (2019). Quantitative analysis of proteome in non-functional pituitary adenomas: clinical relevance and potential benefits for the patients. *Front. Endocrinol.* 10, 854. doi:10.3389/fendo.2019.00854

Cheng, T., and Zhan, X. (2017). Pattern recognition for predictive, preventive, and personalized medicine in cancer. *EPMA J.* 8, 51–60. doi:10.1007/s13167-017-0083-9

Chian, S., Thapa, R., Chi, Z., Wang, X. J., and Tang, X. (2014). Luteolin inhibits the Nrf2 signaling pathway and tumor growth *in vivo*. *Biochem. Biophys. Res. Commun.* 447 (4), 602–608. doi:10.1016/j.bbrc.2014.04.039

Chu, F. F., Esworthy, R. S., Chu, P. G., Longmate, J. A., Huycke, M. M., Wilczynski, S., et al. (2004). Bacteria-induced intestinal cancer in mice with disrupted GPx1 and Gpx2 genes. *Cancer Res.* 64 (3), 962–968. doi:10.1158/0008-5472.can-03-2272

Cloer, E. W., Goldfarb, D., Schrank, T. P., Weissman, B. E., and Major, M. B. (2019). NRF2 activation in cancer: from DNA to protein. *Cancer Res.* 79 (5), 889–898. doi:10.1158/0008-5472.CAN-18-2723

Cuadrado, A., Rojo, A. I., Wells, G., Hayes, J. D., Cousin, S. P., Rumsey, W. L., et al. (2019). Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases. *Nat. Rev. Drug Discov.* 18 (4), 295–317. doi:10.1038/s41573-018-0008-x

Cuttica, C. M., Giusti, M., Bocca, L., Sessarego, P., De Martini, D., Valenti, S., et al. (1997). Nitric oxide modulates *in vivo* and *in vitro* growth hormone release in acromegaly. *Neuroendocrinology* 66 (6), 426–431. doi:10.1159/000127268

Dayalan Naidu, S., and Dinkova-Kostova, A. T. (2020). KEAP1, a cysteine-based sensor and a drug target for the prevention and treatment of chronic disease. *Open Biol.* 10 (6), 200105. doi:10.1098/rsob.200105.106

de la Vega, M. R., Chapman, E., and Zhang, D. D. (2018). NRF2 and the hallmarks of cancer. *Cancer Cell* 34 (1), 21–43. doi:10.1016/j.ccell.2018.03.022

Deyu, H., Luqing, C., Xianglian, L., Pu, G., Qirong, L., Xu, W., et al. (2018). Protective mechanisms involving enhanced mitochondrial functions and mitophagy against T-2 toxin-induced toxicities in GH3 cells. *Toxicol. Lett.* 295, 41–53. doi:10.1016/j.toxlet.2018.05.041

Dhakshinamoorthy, S., and Porter, A. G. (2004). Nitric oxide-induced transcriptional up-regulation of protective genes by Nrf2 via the antioxidant response element counteracts apoptosis of neuroblastoma cells. *J. Biol. Chem.* 279 (19), 20096–20107. doi:10.1074/jbc.M312492200

Duvilanski, B. H., Zambruno, C., Seilicovich, A., Pisera, D., Lasaga, M., Diaz, M. C., et al. (1995). Role of nitric oxide in control of prolactin release by the adenohypophysis. *Proc. Natl. Acad. Sci. U.S.A.* 92 (1), 170–174. doi:10.1073/pnas.92.1.170

Ebokaiwe, A. P., Obeten, K. E., Okori, S. O., David, E. E., Olusanya, O., Chukwu, C. J., et al. (2020). Co-administration of selenium nanoparticles and metformin abrogate testicular oxidative injury by suppressing redox imbalance, augmenting sperm quality and Nrf2 protein expression in streptozotocin-induced diabetic rats. *Biol. Trace Elem. Res.* 198, 544. doi:10.1007/s12011-020-02082-2

Elchuri, S., Oberley, T. D., Qi, W., Eisenstein, R. S., Jackson Roberts, L., et al. (2005). CuZnSOD deficiency leads to persistent and widespread oxidative damage and hepatocarcinogenesis later in life. *Oncogene* 24 (3), 367–380. doi:10.1038/sj.onc.1208207

Evans, C. O., Moreno, C. S., Zhan, X., Mccabe, M. T., Vertino, P. M., Desiderio, D. M., et al. (2008). Molecular pathogenesios f human prolactinomasi dentified by genee xpressionp rofiling, RT-qPCR, and proteomic analyses. *Pituitlry* 11, 231–245. doi:10.1007/s11102-007-0082-2

Feng, J., Zhang, Q., Zhou, Y., Yu, S., Hong, L., Zhao, S., et al. (2018). Integration of proteomics and metabolomics revealed metabolite-protein networks in ACTH-secreting pituitary adenoma. *Front. Endocrinol.* 9, 678. doi:10.3389/fendo.2018.00678

Furfaro, A. L., Traverso, N., Domenicotti, C., Piras, S., Moretta, L., Marinari, U. M., et al. (2016). The Nrf2/HO-1 axis in cancer cell growth and chemoresistance. *Oxid Med. Cell* 2016, 1958174. doi:10.1155/2016/1958174

Galland, F., Lacroix, L., Saulnier, P., Dessen, P., Meduri, G., Bernier, M., et al. (2010). Differential gene expression profiles of invasive and non-invasive non-functioning pituitary adenomas based on microarray analysis. *Endocr. Relat. Cancer* 17, 361–371. doi:10.1677/ERC-10-0018

Giorgianni, F., Beranova-Giorgianni, S., and Desiderio, D. M. (2004). Identification and characterization of phosphorylated proteins in the human pituitary. *Proteomics* 4, 587–598. doi:10.1002/pmic.200300584

Giorgianni, F., Desiderio, D. M., and Beranova-Giorgianni, S. (2003). Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry. *Electrophoresis* 24, 253–259. doi:10.1002/elps.200390021

Gonçalves, Á. d. C., Moreira, E. J. S., and Portari, G. V. (2019). Benfotiamine supplementation prevents oxidative stress in anterior tibialis muscle and heart. *J. Integr. Med.* 17, 423–429. doi:10.1016/j.joim.2019.07.001

Grech, G., Zhan, X., Yoo, B. C., Bubnov, R., Hagan, S., Danesi, R., et al. (2015). EPMA position paper in cancer: current overview and future perspectives. *EPMA J.* 6, 9. doi:10.1186/s13167-015-0030-6

Hambright, H. G., Meng, P., Kumar, A. P., and Ghosh, R. (2015). Inhibition of PI3K/AKT/mTOR axis disrupts oxidative stress-mediated survival of melanoma cells. *Oncotarget* 6 (9), 7195–7208. doi:10.18632/oncotarget.3131

Harris, I. S., Treloar, A. E., Inoue, S., Sasaki, M., Gorrini, C., Lee, K. C., et al. (2015). Glutathione and thioredoxin antioxidant pathways synergize to drive cancerinitiation and progression. *Cancer Cell* 27 (2), 211–222. doi:10.1016/j.ccell.2014.11.019

Hernández-Ramírez, L. C., Morgan, R. M. L., Barry, S., D'Acquisto, F., Prodromou, C., and Korbonits, M. (2018). Multi-chaperone function modulation and association with cytoskeletal proteins are key features of the function of AIP in the pituitary gland. *Oncotarget* 9 (10), 9177–9198. doi:10.18632/oncotarget.24183

Hetland, G., Tangen, J.-M., Mahmood, F., Mirlashari, M. R., Nissen-Meyer, L. S. H., Nentwich, I., et al. (2020). Antitumor, anti-Inflammatory and antiallergic effects of Agaricus blazei Mushroom extract and the related medicinal basidiomycetes mushrooms, hericium erinaceus and Grifola frondosa: a Review of Preclinical and Clinical Studies. *Nutrients* 12 (5), 1339. doi:10.3390/nu12051339

Hu, R., Wang, X., and Zhan, X. (2013). Multi-parameter systematic strategies for predictive, preventive and personalised medicine in cancer. *EPMA J.* 4, 2. doi:10.1186/1878-5085-4-2

Huang, C. S., Chen, H. W., Lin, T. Y., Lin, A. H., and Lii, C. K. (2018). Shikonin upregulates the expression of drug-metabolizing enzymes and drug transporters in primary rat hepatocytes. *J. Ethnopharmacol.* 216, 18–25. doi:10.1016/j.jep.2018.01.026

Hybertson, B. M., Gao, B., Bose, S. K., and McCord, J. M. (2011). Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. *Mol. Aspects Med.* 32 (4-6), 234–246. doi:10.1016/j.mam.2011.10.006

Ilhan, M., Turgut, S., Turan, S., Demirci Cekic, S., Ergen, H. A., Korkmaz Dursun, G., et al. (2018). The assessment of total antioxidant capacity and superoxide dismutase levels, and the possible role of manganese superoxide dismutase polymorphism in acromegaly. *Endocr. J.* 65 (1), 91–99. doi:10.1507/endocrj.EJ17-0300

Johnson, G. S., Li, J., Beaver, L. M., Dashwood, W. M., Sun, D., Rajendran, P., et al. (2017). A functional pseudogene, NMRAL2P, is regulated by Nrf2 and serves as a coactivator of NQO1 in sulforaphane-treated colon cancer cells. *Mol. Nutr. Food Res.* 61 (4). doi:10.1002/mnfr.201600769

Kang, S. W., Rhee, S. G., Chang, T. S., Jeong, W., and Choi, M. H. (2005). 2-Cys peroxiredoxin function in intracellular signal transduction: therapeutic implications. *Trends Mol. Med.* 11 (12), 571–578. doi:10.1016/j.molmed.2005.10.006

Khalil, H. S., Langdon, S. P., Goltsov, A., Soininen, T., Harrison, D. J., Bown, J., et al. (2016). A novel mechanism of action of HER2 targeted immunotherapy is explained by inhibition of NRF2 function in ovarian cancer cells. *Oncotarget* 7 (46), 75874–75901. doi:10.18632/oncotarget.12425

Klaunig, J. E. (2018). Oxidative stress and cancer. *Curr. Pharm. Des.* 24 (40), 4771–4778. doi:10.2174/1381612825666190215121712

Konstantinopoulos, P. A., Spentzos, D., Fountzilas, E., Francoeur, N., Sanisetty, S., Grammatikos, A. P., et al. (2011). Keap1 mutations and Nrf2 pathway activation in epithelial ovarian cancer. *Cancer Res.* 71 (15), 5081–5089. doi:10.1158/0008-5472.CAN-10-4668

Kruk, J., and Aboul-Enein, H. Y. (2017). Reactive oxygen and nitrogen species in carcinogenesis: implications of oxidative stress on the progression and development of several cancer types. *Mini Rev. Med. Chem.* 17 (11), 904–919. doi:10.2174/1389557517666170228115324

Kruse, A., Broholm, H., Rubin, I., Schmidt, K., and Lauritzen, M. (2002). Nitric oxide synthase activity in human pituitary adenomas. *Acta Neurol. Scand.* 106 (6), 361–366. doi:10.1034/j.1600-0404.2002.01138.x

Kudryavtseva, A. V., Krasnov, G. S., Dmitriev, A. A., Alekseev, B. Y., Kardymon, O. L., Sadritdinova, A. F., et al. (2016). Mitochondrial dysfunction and oxidative stress in aging and cancer. *Oncotarget* 7 (29), 44879–44905. doi:10.18632/oncotarget.9821

Kurisaka, M., Nakajo, T., and Mori, K. (2004). Immunohistochemical study on the expression of copper and zinc-superoxide dismutase (Cu, Zn-SOD) in human adenohypophysis and pituitary adenomas. *No To Shinkei* 46 (10), 948–954.

Kwak, M. K., and Kensler, T. W. (2010). Targeting NRF2 signaling for cancer chemoprevention. *Toxicol. Appl. Pharmacol.* 244 (1), 66–76. doi:10.1016/j.taap.2009.08.028

Leinonen, H. M., Kansanen, E., Pölönen, P., Heinäniemi, M., and Levonen, A. L. (2014). Role of the Keap1-Nrf2 pathway in cancer. *Adv. Cancer Res.* 122, 281–320. doi:10.1016/B978-0-12-420117-0.00008-6

Li, H., Cao, L., Yi, P. Q., Xu, C., Su, J., Chen, P. Z., et al. (2019). Pituitary adenylate cyclase-activating polypeptide ameliorates radiation-induced cardiac injury. *Am. J. Transl. Res.* 11 (10), 6585–6599. www.ajtr.org/ISSN:1943-8141/AJTR0102651.

Li, N., and Zhan, X. (2019). Mitochondrial dysfunction pathway networks and mitochondrial dynamics in the pathogenesis of pituitary adenomas. *Front. Endocrinol.* 10, 690. doi:10.3389/fendo.2019.00690

Li, P. C., Tu, M. J., Ho, P. Y., Jilek, J. L., Duan, Z., et al. (2018). Bioengineered NRF2-siRNA is effective to interfere with NRF2 pathways and improve chemosensitivity of human cancer cells. *Drug Metab. Dispos.* 46 (1), 2–10. doi:10.1124/dmd.117.078741

Li, W., Dong, S., Chen, Q., Chen, C., and Dong, Z. (2020). Selenium may suppress peripheral blood mononuclear cell apoptosis by modulating HSP70 and regulate levels of SIRT1 through reproductive hormone secretion and oxidant stress in women suffering fluorosis. *Eur. J. Pharmacol.* 878, 173098. doi:10.1016/j.ejphar.2020.173098

Li, W., and Kong, A. N. (2009). Molecular mechanisms of Nrf2-mediated antioxidant response. *Mol. Carcinog.* 48 (2), 91–104. doi:10.1002/mc.20465

Li, Y., Huang, T. T., Carlson, E. J., Melov, S., Ursell, P. C., Olson, J. L., et al. (1995). Dilated cardiomyopathy and neonatal lethality in mutant mice lacking mangese superoxide dismutase. *Nat. Genet.* 11 (4), 376–381. doi:10.1038/ng1295-376

Liguori, I., Russo, G., Curcio, F., Bulli, G., Aran, L., Della-Morte, D., et al. (2018). Oxidative stress, aging, and diseases. *Clin. Interv. Aging* 13, 757–772. doi:10.2147/CIA.S158513

Liu, M. M., Huang, K. M., Qian, L., Chatterjee, P., Zhang, S., Li, R., et al. (2018). Effects of bioactive constituents in the Traditional Chinese Medicinal formula Si-Wu-Tang on Nrf2 signaling and neoplastic cellular transformation. *Phytomedicine* 40, 1–9. doi:10.1016/j.phymed.2017.12.031

Liu, P., Tian, W., Tao, S., Tillotson, J., Wijeratne, E. M. K., Gunatilaka, A. A. L., et al. (2019). Non-covalent NRF2 activation confers greater cellular protection than covalent activation. *Cell Chem Biol.* 26 (10), 1427–1435. doi:10.1016/j.chembiol.2019.07.011

Lloyd, R. V., Jin, L., Qian, X., Zhang, S., and Scheithauer, B. W. (1995). Nitric oxide synthase in the human pituitary gland. *Am. J. Pathol.* 146 (1), 86–94.

Long, Y., Lu, M., Cheng, T., Zhan, X., and Zhan, X. (2019). Multiomics-based signaling pathway network alterations in human non-functional pituitary adenomas. *Front. Endocrinol.* 10, 835. doi:10.3389/fendo.2019.00835

Lopes, M. B. S. (2017). The 2017 World Health Organization classification of tumors of the pituitary gland: a summary. *Acta Neuropathol.* 134, 521–535. doi:10.1007/s00401-017-1769-8

Lu, M., Wang, Y., and Zhan, X. (2019). The MAPK pathway-based drug therapeutic targets in pituitary adenomas. *Front. Endocrinol.* 10, 330. doi:10.3389/fendo.2019.00330

Lu, M., and Zhan, X. (2018). The crucial role of multiomic approach in cancer research and clinically relevant outcomes. *EPMA J.* 9, 77–102. doi:10.1007/s13167-018-0128-8

Lu, M. C., Ji, J. A., Jiang, Z. Y., and You, Q. D. (2016). The Keap1-Nrf2-ARE pathway as a potential preventive and therapeutic target: an update. *Med. Res. Rev.* 36 (5), 924–963. doi:10.1002/med.21396

Lubelska, K., Wiktorska, K., Mielczarek, L., Milczarek, M., Zbroińska-Bregisz, I., and Chilmonczyk, Z. (2016). Sulforaphane regulates NFE2L2/Nrf2-Dependent xenobiotic metabolism phase II and phase III enzymes differently in human colorectal cancer and untransformed epithelial colon cells. *Nutr. Cancer* 68 (8), 1338–1348. doi:10.1080/01635581.2016.1224369

Mahran, Y. F. (2020). New insights into the protection of growth hormone in cisplatin-induced nephrotoxicity: the impact of IGF-1 on the keap1-nrf2/HO-1 signaling. *Life Sci.* 253, 117581. doi:10.1016/j.lfs.2020.117581

Manigandan, K., Manimaran, D., Jayaraj, R. L., Elangovan, N., Dhivya, V., and Kaphle, A. (2015). Taxifolin curbs NF-κB-mediated Wnt/β-catenin signaling via up-regulating Nrf2 pathway in experimental colon carcinogenesis. *Biochimie* 119, 103–112. doi:10.1016/j.biochi.2015.10.014

Mann, G. E., Rowlands, D. J., Li, F. Y., de Winter, P., and Siow, R. C. (2007). Activation of endothelial nitric oxide synthase by dietary isoflavones: role of NO in Nrf2-mediated antioxidant gene expression. *Cardiovasc. Res.* 75 (2), 261–274. doi:10.1016/j.cardiores.2007.04.004

McCann, S. M., Haens, G., Mastronardi, C., Walczewska, A., Karanth, S., Rettori, V., et al. (2003). The role of nitric oxide (NO) in control of LHRH release that mediates gonadotropin release and sexual behavior. *Curr. Pharm. Des.* 9 (5), 381–390. doi:10.2174/1381612033391766

McCann, S. M., Karanth, S., Mastronardi, C. A., Dees, W. L., Childs, G., Miller, B., et al. (2001). Control of gonadotropin secretion by follicle-stimulating hormone-releasing factor, luteinizing hormone-releasing hormone, and leptin. *Arch. Med. Res.* 32 (6), 476–485. doi:10.1016/s0188-4409(01)00343-5

Mehnati, P., Baradaran, B., Vahidian, F., and Nadiriazam, S. (2020). Functional response difference between diabetic/normal cancerous patients to inflammatory cytokines and oxidative stresses after radiotherapy. *Rep. Pract. Oncol. Radiother.* 25 (5), 730–737. doi:10.1016/j.rpor.2020.06.008

Melov, S., Doctrow, S. R., Schneider, J. A., Haberson, J., Patel, M., Coskun, P. E., et al. (2001). Lifepan extension and rescue of spongiform encephalopathy in superoxide dismutase 2 nullizygous mice treated with superoxide dismutasecatalase mimetics. *J. Neurosci.* 21 (21), 8348–8353. doi:10.1523/jneurosci.21-21-08348.2001

Menegon, S., Columbano, A., and Giordano, S. (2016). The dual roles of NRF2 in cancer. *Trends Mol. Med.* 22 (7), 578–593. doi:10.1016/j.molmed.2016.05.002

Mishra, P., Paital, B., Jena, S., Swain, S. S., Kumar, S., Yadav, M. K., et al. (2019). Possible activation of NRF2 by Vitamin E/Curcumin against altered thyroid hormone induced oxidative stress via NFκB/AKT/mTOR/KEAP1 signalling in rat heart. *Sci. Rep.* 9, 7408. doi:10.1038/s41598-019-43320-5

Moreno, C. S., Evans, C. O., Zhan, X., Okor, M., Desiderio, D. M., and Oyesiku, N. M. (2005). Novel molecular signaling and classification of human clinically nonfunctional pituitary adenomas identified by gene expression profiling and proteomic analyses. *Cancer Res.* 65, 10214–10222. doi:10.1158/0008-5472.CAN-05-0884

Nachliely, M., Trachtenberg, A., Khalfin, B., Nalbandyan, K., Cohen-Lahav, M., Yasuda, K., et al. (2019). Dimethyl fumarate and vitamin D derivatives cooperatively enhance VDR and Nrf2 signaling in differentiating AML cells *in vitro* and inhibit leukemia progression in a xenograft mouse model. *J. Steroid Biochem. Mol. Biol.* 188, 8–16. doi:10.1016/j.jsbmb.2018.11.017

Neumann, C. A., Krause, D. S., Carman, C. V., Das, S., Dubey, D. P., Abraham, J. L., et al. (2003). Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression. *Nature* 424 (6948), 561–565. doi:10.1038/nature01819

Niture, S. K., and Jaiswal, A. K. (2010). Hsp90 interaction with INrf2(Keap1) mediates stress-induced Nrf2 activation. *J. Biol. Chem.* 285 (47), 36865–36875. doi:10.1074/jbc.M110.175802

Obrador, E., Liu-Smith, F., Dellinger, R. W., Salvador, R., Meyskens, F. L., and Estrela, J. M. (2019). Oxidative stress and antioxidants in the pathophysiology of malignant melanoma. *Biol. Chem.* 400 (5), 589–612. doi:10.1515/hsz-2018-0327

Osburn, W. O., Wakabayashi, N., Misra, V., Nilles, T., Biswal, S., Trush, M. A., et al. (2006). Nrf2 regulates an adaptive response protecting against oxidative damage following diquat-mediated formation of superoxide anion. *Arch. Biochem. Biophys.* 454 (1), 7–15. doi:10.1016/j.abb.2006.08.005

Pastorelli, D., Fabricio, A. S. C., Giovanis, P., D'Ippolito, S., Fiduccia, P., Soldà, C., et al. (2018). Phytosome complex of curcumin as complementary therapy of advanced pancreatic cancer improves safety and efficacy of gemcitabine: results of a prospective phase II trial. *Pharmacol. Res.* 132, 72–79. doi:10.1016/j.phrs.2018.03.013

Pawlikowshi, M., Winczyk, K., and Jaranowska, M. (2003). Immunohistochemical demonstration of nitric oxide synthase (NOS) in the normal rat pituitary gland, estrogen-induced rat pituitary tumor and human pituitary adenomas. *Folia Histochem. Cytobiol* 41 (2), 87–90.

Pi, J., Zhang, Q., Woods, C. G., Wong, V., Collins, S., and Andersen, M. E. (2008). Activation of Nrf2-mediated oxidative stress response in macrophages by hypochlorous acid. *Toxicol. Appl. Pharmacol.* 226 (3), 236–243. doi:10.1016/j.taap.2007.09.016

Pinilla, L., González, L. C., Tena-Sempere, M., Bellido, C., and Aguilar, E. (2001). Effects of systemic blockade of nitric oxide synthases on pulsatile LH, prolactin, and GH secretion in adult male rats. *Horm. Res.* 55 (5), 229–235. doi:10.1159/000050001

Pinilla, L., Tena-Sempere, M., and Aguilar, E. (1999). Nitric oxide stimulates growth hormone secretion *in vitro* through a calcium- and cyclic guanosine monophosphate-independent mechanism. *Horm. Res.* 51 (5), 242–247. doi:10.1159/000023378

Pizza, V., Agresta, A., D'Acunto, C. W., Festa, M., and Capasso, A. (2011). Neuroinflammation and ageing: current theories and an overview of the data. *Rev. Recent Clin. Trials* 6 (3), 189–203. doi:10.2174/157488711796575577

Prasad, S., Gupta, S. C., Pandey, M. K., Tyagi, A. K., and Deb, L. (2016). Oxidative stress and cancer: advances and challenges. *Oxid Med. Cell Longev.* 2016, 5010423. doi:10.1155/2016/5010423

Prevatto, J. P., Torres, R. C., Diaz, B. L., Silva, P. M. R. E., Martins, M. A., and Carvalho, V. F. (2017). Antioxidant treatment induces hyperactivation of the HPA Axis by upregulating ACTH receptor in the adrenal and downregulating glucocorticoid receptors in the pituitary. *Oxid Med. Cell Longev.* 2017, 4156361. doi:10.1155/2017/4156361

Qian, S., Golubnitschaja, O., and Zhan, X. (2019). Chronic inflammation: key player and biomarker-set to predict and prevent cancer development and progression based on individualized patient profiles. *EPMA J.* 10, 365–381. doi:10.1007/s13167-019-00194-x

Qian, S., Yang, Y., Li, N., Cheng, T., Wang, X., Liu, J., et al. (2018). Prolactin variants in human pituitaries and pituitary adenomas identified with two-dimensional gel electrophoresis and mass spectrometry. *Front. Endocrinol.* 9, 468. doi:10.3389/fendo.2018.00468

Qin, J. J., Cheng, X. D., Zhang, J., and Zhang, W. D. (2019). Dual roles and therapeutic potential of Keap1-Nrf2 pathway in pancreatic cancer: a systematic review. *Cell Commun. Signal* 17 (1), 121. doi:10.1186/s12964-019-0435-2

Reimondo, G., Puglisi, S., Pia, A., and Terzolo, M. (2019). Autonomous hypercortisolism: definition and clinical implications. *Minerva. Endocrinol.* 44, 33–42. doi:10.23736/s0391-1977.18.02884-5

Reuter, S., Gupta, S. C., Chaturvedi, M. M., and Aggarwal, B. B. (2010). Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* 49 (11), 1603–1616. doi:10.1016/j.freeradbiomed.2010.09.006

Riedel, W. (2002). Role of nitric oxide in the control of the hypothalamic-pituitary-adrenocortical axis. *Z. Rheumatol.* 59 (Suppl. 2), 42. doi:10.1007/s003930070016

Roy Chowdhury, S., Sengupta, S., Biswas, S., Sinha, T. K., Sen, R., Basak, R. K., et al. (2014). Bacterial fucose-rich polysaccharide stabilizes MAPK-mediated Nrf2/Keap1 signaling by directly scavenging reactive oxygen species during hydrogen peroxide-induced apoptosis of human lung fibroblast cells. *PLoS One* 9 (11), e113663. doi:10.1371/journal.pone.0113663

Ryoo, I. G., and Kwak, M. K. (2018). Regulatory crosstalk between the oxidative stress-related transcription factor Nfe2l2/Nrf2 and mitochondria. *Toxicol. Appl. Pharmacol.* 359, 24–33. doi:10.1016/j.taap.2018.09.014

Saad El-Din, S., Rashed, L., Medhat, E., Emad Aboulhoda, B., Desoky Badawy, A., Mohammed ShamsEldeen, A., et al. (2020). Active form of vitamin D analogue mitigates neurodegenerative changes in Alzheimer's disease in rats by targeting Keap1/Nrf2 and MAPK-38p/ERK signaling pathways. *Steroids* 156, 108586. doi:10.1016/j.steroids.2020.108586

Sabatino, M. E., Grondona, E., Sosa, L. D. V., Mongi Bragato, B., Carreño, L., Juarez, V., et al. (2018). Oxidative stress and mitochondrial adaptive shift during pituitary tumoral growth. *Free Radic. Biol. Med.* 120, 41–55. doi:10.1016/j.freeradbiomed.2018.03.019

Sahin, K., Orhan, C., Akdemir, F., Tuzcu, M., Iben, C., and Sahin, N. (2012). Resveratrol protects quail hepatocytes against heat stress: modulation of the Nrf2 transcription factor and heat shock proteins. *J. Anim. Physiol. Anim. Nutr.* 96 (1), 66–74. doi:10.1111/j.1439-0396.2010.01123.x

Sajadimajd, S., and Khazaei, M. (2018). Oxidative stress and cancer: the role of Nrf2. *Curr. Cancer Drug Targets* 18 (6), 538–557. doi:10.2174/1568009617666171002144228

Sánchez-Martín, P., Sou, Y. S., Kageyama, S., Koike, M., Waguri, S., and Komatsu, M. (2020). NBR1-mediated p62-liquid droplets enhance the Keap1-Nrf2 system. *EMBO Rep.* 21 (3), e48902. doi:10.15252/embr.201948902

Sivils, J. C., Ancrum, T. M., and Bain, L. J. (2013). LOSS of Mrp1 alters detoxification enzyme expression in a tissue- and hormonal-status-specific manner. *J. Appl. Toxicol.* 33 (8), 766–773. doi:10.1002/jat.2727

Song, M. K., Lee, J. H., Ryoo, I. G., Lee, S. H., Ku, S. K., and Kwak, M. K. (2019). Bardoxolone ameliorates TGF-β1-associated renal fibrosis through Nrf2/Smad7 elevation. *Free Radic. Biol. Med.* 138, 33–42. doi:10.1016/j.freeradbiomed.2019.04.033

Song, N. Y., Kim, D. H., Kim, E. H., Na, H. K., Kim, N. J., Suh, Y. G., et al. (2011). Multidrug resistance-associated protein 1 mediates 15-deoxy-Δ(12,14)-prostaglandin J2-induced expression of glutamate cysteine ligase expression via Nrf2 signaling in human breast cancer cells. *Chem. Res. Toxicol.* 24 (8), 1231–1241. doi:10.1021/tx200090n

Stalla, G. K., Ciato, D., and Dimopoulou, C. (2019). "The adrenal gland: central relay in health and disease - current challenges and perspectives 2018" - Cushing's disease. *Exp. Clin. Endocrinol. Diabetes* 127, 147–155. doi:10.1055/a-0664-7632

Sun, T., Gao, J., Han, D., Shi, H., and Liu, X. (2019). Fabrication and characterization of solid lipid nano-formulation of astraxanthin against DMBA-induced breast cancer via Nrf-2-Keap1 and NF-kB and mTOR/Maf-1/PTEN pathway. *Drug Deliv.* 26 (1), 975–988. doi:10.1080/10717544.2019.1667454

Syu, J. P., Chi, J. T., and Kung, H. N. (2016). Nrf2 is the key to chemotherapy resistance in MCF7 breast cancer cells under hypoxia. *Oncotarget* 7 (12), 14659–14672. doi:10.18632/oncotarget.7406

Taguchi, K., and Yamamoto, M. (2017). The KEAP1-NRF2 system in Cancer. *Front. Oncol.* 7, 85. doi:10.3389/fonc.2017.00085

Tian, B., Lu, Z. N., and Guo, X. L. (2018). Regulation and role of nuclear factor-E2-related factor 2 (Nrf2) in multidrug resistance of hepatocellular carcinoma. *Chem. Biol. Interact* 280, 70–76. doi:10.1016/j.cbi.2017.12.014

Tian, H., Zhang, D., Gao, Z., Li, H., Zhang, B., Zhang, Q., et al. (2014). MDA-7/IL-24 inhibits Nrf2-mediated antioxidant response through activation of p38 pathway and inhibition of ERK pathway involved in cancer cell apoptosis. *Cancer Gene Ther.* 21 (10), 416–426. doi:10.1038/cgt.2014.45

Tian, L., Lu, Y., Yang, T., Deng, Z., Xu, L., Yao, W., et al. (2019). aPKCι promotes gallbladder cancer tumorigenesis and gemcitabine resistance by competing with Nrf2 for binding to Keap1. *Redox Biol.* 22, 101149. doi:10.1016/j.redox.2019.101149

To, C., Ringelberg, C. S., Royce, D. B., Williams, C. R., Risingsong, R., Sporn, M. B., et al. (2015). Dimethyl fumarate and the oleanane triterpenoids, CDDO-imidazolide and CDDO-methyl ester, both activate the Nrf2 pathway but have opposite effects in the A/J model of lung carcinogenesis. *Carcinogenesis* 36 (7), 769–781. doi:10.1093/carcin/bgv061

Ueta, Y., Levy, A., Powell, M. P., Lightman, S. L., Kinoshita, Y., Yokota, A., et al. (1998). Neuronal nitric oxide synthase gene expression in human pituitary tumours: a possible association with somatotroph adenomas and growth hormone-releasing hormone gene expression. *Clin. Endocrinol.* 49 (1), 29–38. doi:10.1046/j.1365-2265.1998.00399.x

Valko, M., Rhodes, C. J., Moncol, J., Izakovic, M., and Mazur, M. (2006). Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem. Biol. Interact* 160 (1), 1–40. doi:10.1016/j.cbi.2005.12.009

Vankelecom, H., Matthys, P., and Denef, C. (1997). Inducible nitric oxide synthase in the anterior pituitary gland: induction by interferon-r in a subpopulation of folliculostellate cells and in an unidentifiable population of nonhormone-secreting cells. *J. Histochem. Cytochem.* 45 (6), 847–857. doi:10.1177/002215549704500609

Villeneuve, N. F., Lau, A., and Zhang, D. D. (2010). Regulation of the Nrf2–Keap1 antioxidant response by the ubiquitin proteasome system: an insight into cullin-

ring ubiquitin ligases. *Antioxid. Redox Signal.* 13 (11), 1699–1712. doi:10.1089/ars.2010.3211

Wang, K.-C., Liu, Y.-C., El-Shazly, M., Shih, S.-P., Du, Y.-C., Hsu, Y.-M., et al. (2019). The antioxidant from ethanolic extract of Rosa cymosa fruits activates phosphatase and tensin homolog *in vitro* and *in vivo*: a new insight on its antileukemic effect. *Int. J. Mol. Sci.* 20 (8), 1935. doi:10.3390/ijms20081935

Wang, M., Shi, G., Bian, C., Nisar, M. F., Guo, Y., Wu, Y., et al. (2018). UVA irradiation enhances brusatol-mediated inhibition of melanoma growth by downregulation of the Nrf2-mediated antioxidant response. *Oxid Med. Cell Longev.* 2018, 9742154. doi:10.1155/2018/9742154

Wang, X., Guo, T., Peng, F., Long, Y., Mu, Y., Yang, H., et al. (2015). Proteomic and functional profiles of a follicle-stimulating hormone positive human nonfunctional pituitary adenoma. *Electrophoresis* 36, 1289–1304. doi:10.1002/elps.201500006

Wang, Y., Cheng, T., Lu, M., Mu, Y., Li, B., Li, X., et al. (2019). TMT-based quantitative proteomics revealed follicle-stimulating hormone (FSH)-related molecular characterizations for potentially prognostic assessment and personalized treatment of FSH-positive non-functional pituitary adenomas. *EPMA J.* 10, 395–414. doi:10.1007/s13167-019-00187-w

Wang, Y., Zhang, J., Huang, Z. H., Huang, X. H., Zheng, W. B., Yin, X. F., et al. (2017). Isodeoxyelephantopin induces protective autophagy in lung cancer cells via Nrf2-p62-keap1 feedback loop. *Cell Death Dis.* 8 (6), e2876. doi:10.1038/cddis.2017.265

Wang, Y. Y., Yang, Y. X., Zhao, R., Pan, S. T., Zhe, H., He, Z. X., et al. (2015). Bardoxolone methyl induces apoptosis and autophagy and inhibits epithelial-to-mesenchymal transition and stemness in esophageal squamous cancer cells. *Drug Des. Devel Ther.* 9, 993–1026. doi:10.2147/DDDT.S73493

Warfel, N. A., Sainz, A. G., Song, J. H., and Kraft, A. S. (2016). PIM kinase inhibitors kill hypoxic tumor cells by reducing Nrf2 signaling and increasing reactive oxygen species. *Mol. Cancer Ther.* 15 (7), 1637–1647. doi:10.1158/1535-7163.MCT-15-1018

Wu, C., Zhang, H., Zhang, J., Zhang, H., Zeng, Y., Fang, S., et al. (2019). Increased oxidative stress, inflammation and fibrosis in perirenal adipose tissue of patients with cortisol-producing adenoma. *Adipocyte* 8 (1), 347–356. doi:10.1080/21623945.2019.1690834

Yang, J., Zhang, C., Zhang, W., Shi, R., and Zhang, Z. (2012). Extracellular superoxide dismutase, a potential extracellular biomarker candidate for prolactinoma. *West Indian Med. J.* 61 (7), 665–669.

Yates, M. S., and Kensler, T. W. (2007). Chemopreventive promise of targeting the Nrf2 pathway. *Drug News Perspect.* 20 (2), 109–117. doi:10.1358/dnp.2007.20.2.108343

Yu, Z., Dai, Z. Y., Qin, G. X., Li, M. Y., and Wu, L. F. (2020). Alleviative effects of dietary microbial floc on copper-induced inflammation, oxidative stress, intestinal apoptosis and barrier dysfunction in Rhynchocypris lagowski Dybowski. *Fish Shellfish Immunol.* 106, 120–132. doi:10.1016/j.fsi.2020.07.070

Zeng, Y., Lian, S., Li, D., Lin, X., Chen, B., Wei, H., et al. (2017). Anti-hepatocarcinoma effect of cordycepin against NDEA-induced hepatocellular carcinomas via the PI3K/Akt/mTOR and Nrf2/HO-1/NF-κB pathway in mice. *Biomed. Pharmacother.* 95, 1868–1875. doi:10.1016/j.biopha.2017.09.069

Zhai, X., Yuan, S., Yang, X., Zou, P., Li, L., Li, G., et al. (2019). Chitosan oligosaccharides induce apoptosis in human renal carcinoma via reactive-oxygen-species-dependent endoplasmic reticulum stress. *J. Agric. Food Chem.* 67 (6), 1691–1701. doi:10.1021/acs.jafc.8b06941

Zhan, X., and Desiderio, D. M. (2003). A reference map of a human pituitary adenoma proteome. *Proteomics* 3, 699–713. doi:10.1002/pmic.200300408

Zhan, X., and Desiderio, D. M. (2006). Nitroproteins from a human pituitary adenoma tissue discovered with a nitrotyrosine affinity column and tandem mass spectrometry. *Anal. Biochem.* 354, 279–289. doi:10.1016/j.ab.2006.05.024

Zhan, X., and Desiderio, D. M. (2010a). Signaling pathway networks mined from human pituitary adenoma proteomics data. *BMC Med. Genomics* 3, 13. doi:10.1186/1755-8794-3-13

Zhan, X., and Desiderio, D. M. (2010b). The use of variations in proteomes to predict, prevent, and personalize treatment for clinically nonfunctional pituitary adenomas. *EPMA J.* 1, 439–459. doi:10.1007/s13167-010-0028-z

Zhan, X., and Desiderio, D. M. (2004). The human pituitary nitroproteome: detection of nitrotyrosyl-proteins with two-dimensional Western blotting, and amino acid sequence determination with mass spectrometry. *Biochem. Biophys. Res. Commun.* 325, 1180–1186. doi:10.1016/j.bbrc.2004.10.169

Zhan, X., Wang, X., and Cheng, T. (2016). Human pituitary adenoma proteomics: new progresses and perspectives. *Front. Endocrinol.* 7, 54. doi:10.3389/fendo.2016.00054

Zhan, X., Wang, X., and Desiderio, D. M. (2013). Pituitary adenoma nitroproteomics: current status and perspectives. *Oxid Med. Cel Longev.* 2013, 580710. doi:10.1155/2013/580710

Zhan, X., Wang, X., Long, Y., and Desiderio, D. M. (2014a). Heterogeneity analysis of the proteomes in clinically nonfunctional pituitary adenomas. *BMC Med. Genomics* 7, 69. doi:10.1186/s12920-014-0069-6

Zhan, X., Desiderio, D. M., Wang, X., Zhan, X., Guo, T., and Li, M. (2014b). Identification of the proteomic variations of invasive relative to non-invasive non-functional pituitary adenomas. *Electrophoresis* 35, 2184–2194. doi:10.1002/elps.201300590

Zhan, X., and Desiderio, D. M. (2007). Linear ion-trap mass spectrometric characterization of human pituitary nitrotyrosine-containing proteins. *Int. J. Mass. Spectrom.* 259, 96–104. doi:10.1016/j.ijms.2006.06.009

Zhang, B., Ma, Z., Tan, B., and Lin, N. (2019). Targeting the cell signaling pathway Keap1-Nrf2 as a therapeutic strategy for adenocarcinomas of the lung. *Expert Opin. Ther. Targets* 23 (3), 241–250. doi:10.1080/14728222.2019.1559824

Zhang, J., Su, L., Ye, Q., Zhang, S., Kung, H., Jiang, F., et al. (2017). Discovery of a novel Nrf2 inhibitor that induces apoptosis of human acute myeloid leukemia cells. *Oncotarget* 8 (5), 7625–7636. doi:10.18632/oncotarget.13825

Zhang, Y., Qiao, R., He, D., Zhao, Z., Yang, S., Zou, H., et al. (2016). Indazolo[3,2-b] quinazolinones attack hepatocellular carcinoma Hep3B cells by inducing mitochondrial-dependent apoptosis and inhibition of Nrf2/ARE signaling pathway. *Curr. Mol. Med.* 16 (9), 820–828. doi:10.2174/1566524016666161128114444

Zhao, L., Liu, Z., Jia, H., Feng, Z., Liu, J., and Li, X. (2015). Lipoamide acts as an indirect antioxidant by simultaneously stimulating mitochondrial biogenesis and phase II antioxidant enzyme systems in ARPE-19 cells. *PLoS One* 10 (6), e0128502. doi:10.1371/journal.pone.0128502

Zhao, X. J., Zhao, Z., Yang, D. D., Cao, L. L., Zhang, L., Ji, J., et al. (2017). Activation of ATP-sensitive potassium channel by iptakalim normalizes stress-induced HPA axis disorder and depressive behaviour by alleviating inflammation and oxidative stress in mouse hypothalamus. *Brain Res. Bull.* 130, 146–155. doi:10.1016/j.brainresbull.2017.01.026

Zhao, Y., Giorgianni, F., Desiderio, D. M., Fang, B., and Beranova-Giorgianni, S. (2005). Toward a global analysis of the human pituitary proteome by multiple gel-based technology. *Anal. Chem.* 77, 5324–5331. doi:10.1021/ac050354e

Zhou, W., Song, Y., Xu, H., Zhou, K., Zhang, W., Chen, J., et al. (2011). Nonfunctional pituitary adenomas, estrogen receptors and slug contribute to development of invasiveness. *J. Clin. Endocrinol. Metab.* 96, E1237–E1245. doi:10.1210/jc.2010-3040

Zhu, J., Wang, H., Fan, Y., Lin, Y., Zhang, L., Ji, X., et al. (2014). Targeting the NF-E2-related factor 2 pathway: a novel strategy for glioblastoma. *Oncol. Rep.* 32 (2), 443–450. doi:10.3892/or.2014.3259

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Copyright © 2021 Zhan, Li and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

ACTH Adrenocorticotropin

AFAR aldo-keto reductase family 7 member A2

AIP Aryl hydrocarbon receptor interacting protein gene

AKR Palmitoyltransferase AKT

AKR1B1 aldo-keto reductase family 1 member B

AKT Protein kinase B

AOX4 Aldehyde oxidase 4

AP-1 Activator protein-1

aPKCι Atypical protein kinase Cι

ARE Antioxidant response element

ARS Acute restraint stress

ASK1 Apoptosis signal-regulating kinase 1

ATF4 Activating transcription factor 4

ATP5A1 ATP synthase subunit alpha, mitochondrial

ATP5B ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit

BACH1 Transcription regulator protein BACH1

Bax BCL2 associated X, apoptosis regulator

BCL2 BCL2 apoptosis regulator

BR Brusatol

CALM calmodulin

CAT catalase

CBP CREB-binding protein

CBR4 carbonyl reductase 4

CCL2 C-C motif chemokine ligand 2

CCT7 T-complex protein 1 subunit eta

c-FOS Proto-oncogene protein c-FOS

CLPP Caseinolytic protease

COS Chitosan oligosaccharide

COX6B cytochrome c oxidase subunit 6B

c-MAF (MAF) MAF bZIP transcription factor

c-Raf RAF proto-oncogene serine/threonine-protein kinase

CRS Chronic restraint stress

Cul3 Cullin 3-based ubiquitin E3 ligase complex

Cyp cytochrome P

CYP1A cytochrome P450 family 1 subfamily A

CYP2A cytochrome P450 family 2 subfamily A

CYP2C cytochrome P450 family 2 subfamily C

CYP3A cytochrome P450 family 3 subfamily A

CYP4A cytochrome P450 family 4 subfamily A

DBM Dibenzoylmethane

DEG Differentially expressed gene

DEP Differentially expressed protein

DMF Dimethyl fumarate

EPHX1 Epoxide hydrolase 1

EpRE Electrophile responsive element

ER endoplasmic reticulum

ERK Extracellular signal-related kinase

ERK1/2 mitogen-activated protein kinase

ERK5 mitogen-activated protein kinase

ERP29 endoplasmic reticulum protein 29

ESR1 estrogen receptor 1

ESR2 estrogen receptor 2

FKBP5 FK506-binding protein 5

FMO Dimethylaniline monooxygenase [N-oxide-forming]

FPA Functional pituitary adenoma

FRA1 Fos-related antigen 1

FSH Follicle-stimulating hormone

FSHRH Follicle-stimulating hormone-releasing hormone

FTH1 Ferritin heavy polypeptide 1

FTL1 ferritin light polypeptide

GA Gallic acid

GCLC glutamate-cysteine ligase catalytic subunit

GCLM glutamate-cysteine ligase modifier subunit

GH Growth hormone

GPX4 glutathione peroxidase 4

GPX's Glutathione peroxidases

GRP94 94 kD glucose-regulated protein

GSK3β glycogen synthase kinase 3β

GSR glutathione reductase

GST glutathione S-transferase

GSTM2 glutathione S-transferase mu 2

HERPUD1 Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 protein

HIP2 Ubiquitin-conjugating enzyme E2 K

HO-1 heme oxygenase 1

HSP22/40/90 heat shock proteins 22, 40 and 90

HSP27 heat shock protein 27

HSP70 (HSPA4) heat shock protein family A member 4

HSP90 heat shock protein 90

HSP90AA1 heat shock protein 90 alpha family class A member 1

HSP90AB1 heat shock protein 90 alpha family class B member 1

HSP94 heat shock protein 94

HSPA5 heat shock protein family A (Hsp70) member 5

HSPA8 heat shock protein family A (Hsp70) member 8

HSPA9 heat shock protein family A (Hsp70) member 9

HSPCA (HSP90AA1) heat shock protein 90 alpha family class A member 1

HSPCB (HSP90AB1) heat shock protein 90 alpha family class B member 1

IL-1β interleukin 1 beta

IL-6 interleukin 6

iNOS Inducible nitric synthase

IP3R (ITPR1) inositol 1,4,5-trisphosphate receptor type 1

Ipt Iptakalim

JNK Jun N-terminal kinase

JNK1 (MAPK8) mitogen-activated protein kinase 8

JNK 2 (MAPK9) mitogen-activated protein kinase 9

JUN Jun proto-oncogene, AP-1 transcription factor subunit

K-ATP ATP-sensitive potassium

Keap1 Kelch-like ECH-associated protein 1

LH Luteinizing hormone

LHRH Luteinizing hormone-releasing hormone

LIG Z-ligustilide

Maf Musculoaponeurotic fibrosarcoma

MAPKs Mitogen-activated protein kinases

MEK Mitogen-activated protein kinase kinase (MAPKK)

MEKK Mitogen-activated protein kinase kinase kinase (MAPKKK)

MN Micronucleus

MRP1 multidrug-resistant protein-1

NAC N-acetylcysteine

NAF-1 Nutrient-deprivation autophagy factor-1

NDEA N-nitrosodiethylamine

NDUFS8 (NADH) ubiquinone oxidoreductase core subunit S8

NF-kB Nuclear factor kB

NFPA Nonfunctional pituitary adenoma

NME1 NME/NM23 nucleoside diphosphate kinase 1

NO Nitric oxide

NQO1 NAD(P)H:quinine oxidoreductase 1

Nrf2 Nuclear factor erythroid 2 p45-related factor 2

Nur77 (NR4A1) nuclear receptor subfamily 4 group A member 1

O2.- Superoxide radical

8-OHdG 8-hydroxy-2'-deoxyguanosine

OCT octreotide

OH hydroxyl radical

ONOO- Peroxynitrite anion

PA Pituitary adenoma

PACAP38 Pituitary adenylate cyclase-activating polypeptide 38

PDK1 pyruvate dehydrogenase kinase 1

PERK the double-stranded RNA (PKR)-activated protein kinase-like eukaryotic initiation factor 2 kinase

PFK phosphofructokinase

PI3K Phosphatidylinositol 3 kinase

PINK1 PTEN-induced putative kinase protein 1

PKA cAMP dependent protein kinase

PKC protein kinase C

POMC proopiomelanocortin

PPIB Peptidyl-prolyl cis-trans isomerase B

PP2C putative protein phosphatase

PRDX1 peroxiredoxin 1

PRL Prolactin

PSM multiple subunits of the 20S proteasome

Pter Pterostilbene

PTPLAD1 3-hydroxyacyl-CoA dehydratase 3

Ras GTPase Ras

RNS Reactive nitrogen species

ROS Reactive oxygen species

SA Senkyunolide A

SFN Sulforaphane

SOD Superoxide dismutase

SOD1 superoxide dismutase 1

SOM230 somatostatin analogs pasireotide

SQSTM1 sequestosome-1 protein

SR-B1 Scavenger receptor class B member 1

STIP1 stress induced phosphoprotein 1

TAC Total antioxidant capability

TAK1 TGF beta-Activated Kinase 1

TGM2 transglutaminase 2

TLR4 toll like receptor 4

TNFα tumor necrosis factor alpha

TRXR1 thioredoxin reductase 1

TUBB tubulin beta class I

TUBB2A tubulin beta 2A class IIa

TXN1 thioredoxin

UBB Polyubiquitin-B

UB2R1 Ubiquitin-conjugating enzyme E2 R1

UGT UDP glucuronosyl transferase

USP14 ubiquitin-specific peptidase 14

UV ultraviolet

VCP valosin-containing protein

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership